DOKTORANDSKÉ DNY 2016

sborník workshopu doktorandů FJFI
oboru Matematické inženýrství

4. a 11. listopadu 2016

P. Ambrož, Z. Masáková (editoři)

Doktorandské dny 2016
sborník workshopu doktorandů FJFI oboru Matematické inženýrství

P. Ambrož, Z. Masáková (editoři)
Kontakt petr.ambroz@fjfi.cvut.cz / 224 358 569

Vydalo České vysoké učení technické v Praze
Zpracovala Fakulta jaderná a fyzikálně inženýrská
Vytisklo Nakladatelství ČVUT-výroba, Zikova 4, Praha 6
Počet stran 201, Vydání 1.

Seznam příspěvků

Předmluva

Je to již deset let, co se na katedře matematiky FJFI poprvé konaly Doktorandské dny. Letos se tento workshop koná ve dnech 4. a 11. listopadu a bude opět věnován prezentacím doktorandů oboru Matematické inženýrství zajišťovaného katedrami matematiky, fyziky a softwarového inženýrství na Fakultě jaderné a fyzikálně inženýrské Českého vysokého učení technického v Praze. Příspěvky našich studentů pokrývají široký záběr aplikované matematiky.

Věříme, že i letošní ročník workshopu Doktorandské dny se bude těšit přízni nejen doktorandů a jejich školitelů, ale samozřejmě i odborné veřejnosti z řad akademických pracovníků na ČVUT i na spolupracujících ústavech AV ČR.

Tato konference se koná s tradiční podporou Studentské grantové soutěže na ČVUT v Praze (grant SVK 36/16/F4).

Editoři

Equivalent Formulations of the Riemann Hypothesis Based on Lines of Constant Phase

Iva Bezděková*

5th year of PGS, email: bezdekova.iva@gmail.com
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:
Igor Jex, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague
Martin Štefaňák, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. The proof of the Riemann hypothesis is a long-standing problem in mathematics. Whereas numerical techniques an analytic estimates are prevalent, geometrical approaches are rare. Here we argue in favour of a new geometric route towards this conjecture. We state the equivalence of three formulations of the Riemann hypothesis. The proof is based on lines of constant phase of a complex function $f(s)$, satisfying certain assumptions. Since the proof is quite long, here we only summarize some ideas and method important in the proof. Full proof is a part of the original paper listed below.

Keywords: Riemann $\zeta-$function, Riemann hypothesis, Riemann $\xi-$function, Newton flow, lines of constant phase, separatrix

Abstrakt. Důkaz Riemannovy hypotézy je v matematice dlouhodobým problémem. Při zkoumání tohoto problému převládají analytické odhady, zatímco geometrické přístupy jsou spíše výjimkou. Tento článek hovoří ve prospěch nového geometrického přístupu při studiu této hypotézy. Uvedeme tři ekvivalentní formulace Riemannovy hypotézy. Důkaz je založen na liniích konstantní fáze komplexní funkce $f(s)$, která vyhovuje určitým podmínkám. Jelikož je důkaz poměrně dlouhý, uvedeme zde pouze myšlenky a metody, které se u důkazu použijí. Celý důkaz je část originálního článku uvedeném v zápatí stránky.

Klíčová slova: Riemannova $\zeta-$funkce, Riemannova hypotéza, Riemannova $\xi-$funkce, Newtonův tok, linie konstatní fáze, separatrix

Full paper: W.P. Schleich, I. Bezděková, M.B. Kim, P.C. Abbott, H. Maier, H. Montgomery, J.W. Neuberger, Equivalent formulations of the Riemann hypothesis based on lines of constant phase, to be submitted.

1 Introduction

Riemann hypothesis [1, 2, 3, 4] describing a conjecture about zeros of the Riemann zeta function ζ was first stated by Bernhard Riemann 1859 in his seminal article "On the number of primes below a given number" [5].

[*] This work was supported by SGS13/217/OHK4/3T/14 and GAČR 13-33906S.

The Riemann $\zeta-$function is defined for $\sigma > 1$ by the expression

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}, \tag{1}$$

where $s = \sigma + i\tau$ and can be meromorphically continued to the entire complex plane. The only singularity is the simple pole at $s = 1$. Euler discovered connection of the $\zeta-$function with prime numbers, the equation (1) can be written as

$$\zeta(s) = \prod_{p \text{ prime}} (1 - p^{-s})^{-1}. \tag{2}$$

Riemann also found a functional equation for ζ which is most conveniently formulated by introducing a function

$$\xi(s) = \pi^{-s/2}(s-1)\Gamma\left(\frac{s}{2}+1\right)\zeta(s), \tag{3}$$

where Γ denotes Gamma function. Equation (3) leads to the relation

$$\xi(s) = \xi(1-s). \tag{4}$$

Therefore ξ is symmetric with respect to line where $\sigma = 1/2$. Equation (2) implies that ξ has no zeros for $\sigma > 1$ and from the functional equation Eq.(4) also no zeros for $\sigma < 0$. Region

$$0 < \sigma < 1$$

is called a critical strip. Moreover, ξ has no poles.

The original statement of the Riemann hypothesis reads:

All non-trivial zeros of ζ lie on the critical axis.

Where by critical axis we mean line with

$$\sigma = \frac{1}{2}.$$

Function ζ has also trivial zeros and it is known that they are located at negative even integers, $s = -2, -4, \ldots$. In the case of function ξ, Eq.(3), trivial zeros disappear due to poles of the γ function. Therefore, ξ function has only non-trivial zeros that are the same as the non-trivial zeros of the function ζ. Since ξ has no other zeros, we can reformulate the hypothesis as:

All zeros of ξ lie on the critical axis.

2 Method

We analyse the lines of constant phase of the function ξ, Eq.(3) from the point of view of the Newton flow in the complex plane [6, 7]. Here we identify the crucial role of special phase lines called separatrices. These lines divide the flow into the different domains, which means into the different zeros.

2.1 Newton flow

Is defined for complex functions $g(s)$ as

$$\dot{s(t)} = \frac{d}{dt}s = -\frac{g(s)}{g'(s)}, \tag{5}$$

where $t \in [0, b),\ b > 0$. The function $g : \mathbb{C} \to \mathbb{C}$ has to have continuous derivative on an open subset G, $s_0 \in G$, $g'(s) \neq 0$ for all $s \in G$. The initial condition is $g(s(0)) = g(s_0)$.

Further, the equation is equal to

$$\begin{aligned} g'(s)\dot{s(t)} &= -g(s) \\ g(\dot{s(t)}) &= -g(s) \end{aligned}$$

with solution

$$g(s(t)) = g(s_0)e^{-t}. \tag{6}$$

Note that solution in Eq.(6) does not change the phase of the complex function g during the time evolution. Therefore, it corresponds to the lines of constant phase of the function g.

This lines of constant phase have a source and a sink [7]. The source is pole of the function or, if there is no pole, then the source is infinity. The sink of the function is zero of the function. This is easy to see since

$$\lim_{t\to\infty} g(s(t)) = \lim_{t\to\infty} g(s_0)e^{-t} = 0.$$

If the function has no zero, than the lines of constant phase terminate in infinity.

2.2 Assumptions

Let us assume that the function $f(s)$ satisfy four assumptions:

(a1) f satisfies functional equation $f(s) = f(1 - s)$

(a2) the zeros of f and the zeros of its first derivative f' are simple zeros

(a3) f is free of any pole

(a4) for large positive values of σ, the phase Θ of f increases in a monotonic way as τ increases

It can be shown that the function ξ from Eq.(3) satisfy assumptions *(a1)-(a4)*.

2.3 Two cases violating the Riemann hypothesis

Due to the assumption *(a3)*, Newton flow of the ξ function consist of phase lines approaching from infinity. There exist special phase lines that divide the complex plane into several regions, where each of this regions is attracted by one zero. Therefore, we call this phase lines separatrices.

As you can see in the lower part of the figure (1), typical behaviour is that a zero of the function ξ alternates with a zero of its first derivative ξ'. Two cases that would results in zero of a function located off the critical axis are:

(c1) there exist three subsequent vanishing derivatives of the function located at the critical axis

(c2) vanishing derivative is not located at the critical axis

Both cases cause that the phase lines separated by separatrices cannot terminate at zero located on the critical axis. In the first case, there is no such zero between two separatrices. Nevertheless, the phase lines have o terminate in zero of the function and this zero will be located off the axis. In the second case the separatrix does not cross the critical axis at all and therefore has to automatically send the flow into a zero located off the critical axis. The situation and detailed description is depicted in figure (1).

3 Results

This geometrical approach result in three equivalent formulations of the Riemann hypothesis, defined for rather general class of functions f, which include function ξ, and satisfy assumptions $(a1) - (a4)$. The equivalent statements read:

(R1) all zeros of f are located on the critical axis

(R2) all lines of constant phase of f corresponding to $\pm\pi, \pm 2\pi, ...$ merge with the critical axis in zeros of f'

(R3) all points where f' vanishes are located on the critical axis and the phases of f at two consecutive zeros of f' differ by π

4 Conclusion

We have stated three equivalent formulations of the Riemann hypothesis that are based on the lines of constant phase. In this equivalent formulations, we have used properties of the special phase lines called separatrices. This equivalent formulation lead to the behaviour, where two extremal cases with three vanishing derivatives in a row and with vanishing derivative off the axis cannot occur. This three equivalences also helps to better understanding of the behaviour of the ξ function.

References

[1] K. Sabbagh, *The Riemann Hypothesis: The Greatest Unsolved Problem in Mathematics* (Farrar, Straus and Giroux, New York, 2003)

[2] D. Rockmore, *Stalking the Riemann hypothesis* (Pantheon Books, New York, 2005).

[3] M. Du Sautoy, *The music of the primes* (Harper Collins, New York, 2003).

[4] P. Borwein, S. Choi, B. Rooney, A. Weirathmueller, *The Riemann Hypothesis: A Resource for the Afficionado and Virtuoso Alike* in *CMS Books in Mathematics* series (2008).

[5] B. Riemann, Monatsberichte der Berliner Akademie (1859), transcribed German version and English translation by D. R. Willkins see `http://www.claymath.org/publications/riemanns-1859-manuscript`.

[6] J. W. Neuberger, Math. Intell. **21**, 18 (1999).

[7] J. W. Neuberger, C. Feiler, H. Maier, and W. P. Schleich, New J. Phys. **16**, 103023 (2014).

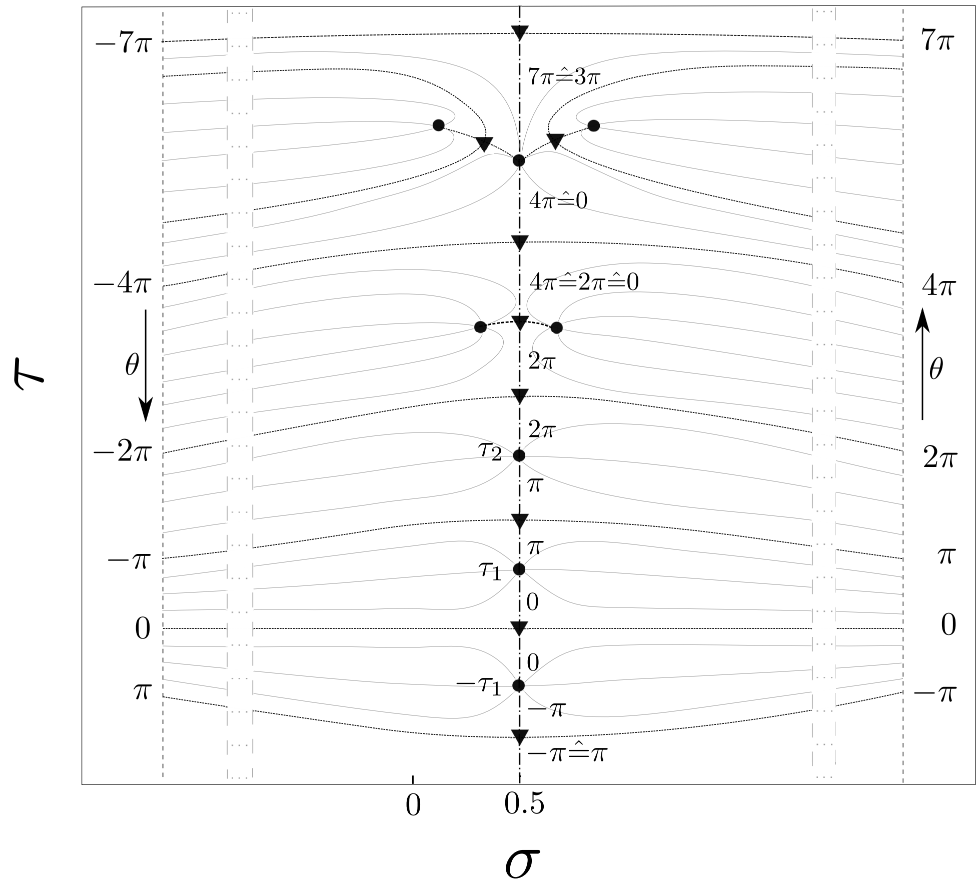


Figure 1: Consequences of the assumptions (*a1*)-(*a4*) on the separatrices, denoted by thicker dashed line, directing the flow of lines of constant phase of $f = f(s)$ in the complex plane and the distribution of zeros. Due to the functional equation, Eq. (4) of f, the phase θ is anti-symmetric with respect to the critical axis $\sigma = 1/2$ depicted by the dashed dot line. As a consequence, f is real on its and can only assume the phases $k\pi$ where k is integer. Moreover, complex analysis enforces an anti-symmetry of θ with respect to the real axis which is a line of constant phase with $\theta = 0$. At $s = 1/2$, the first derivative of f vanishes as indicated by the triangle. A simple zero on the critical axis with imaginary part τ_1, denoted by a dot and no zeros off the axis requires phase lines of a π-interval to approach from the right *and* from the left. If f enjoys more zeros their domains of attraction need to be fenced of by separatrices shown by the thicker dashed lines which start at infinity and merge with the critical axis at right angles where again f' vanishes. Since f is free of poles infinity is indeed the only source of phase lines. For a zero that is located off the critical axis the functional equation automatically enforces an additional zero which is its mirror image. When three consecutive points on the critical axis where f' vanishes are caught between two zeros there must be two zeros off-axis and the phase difference between two consecutive separatrices is 2π rather than π. Similarly when we have one zero on the critical axis and two off the axis then the phase difference between two consecutive separatrices can be a fraction of π. The phase $k\pi$ on the critical axis indicated in the figure is dictated by the distribution of zeros.

Multivariate generalizations of the Chebyshev Polynomials of the Second Kind*

Adam Brus

3rd year of PGS, email: brusadam@fjfi.cvut.cz
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jiří Hrivnák, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. The Chebyshev polynomials of the second kind are generalized with use of multivariate symmetric generalizations of trigonometric functions. For such orthogonal polynomials of several variables are shown the recurrence relations by use of generalized trigonometric identities. For dimension three the exact form of recurrence relations is obtained and then used to calculate first ten polynomials. Further the possibility of generalization which uses antisymmetric multivariate sine function and generalization of the Chebyshev polynomials of the fourth kind is discussed.

Keywords: Chebyshev Polynomials, Multivariate Trigonometric Functions, Orthogonal Polynomials

Abstrakt. Chebyshovy polynomy druhého druhu jsou zobecněny za užití zobecněných symetrických trigonometrických funkcí více proměnných. Pro tyto ortogonální polynomy více proměnných jsou ukázány rekurentní relace užitím zobecněných trigonometrických identit. Pro dimenzi tři je získán přesný tvar rekurentních relací a následně je spočítáno prvních deset polynomů. Dále možnost zobecnění Chebyshových polynomů druhého druhu za použití antisymetrické sinové funkce více proměmmých a zobecnění Chebyshových polynomů čtvrtého druhu je diskutována.

Klíčová slova: Chebyshovy polynomy, Trigonometrické funkce více proměnných, Ortogonální Polynomy

1 Introduction

Orthogonal polynomials [2] are used in many parts of mathematics and physics. Special case of such polynomials are the Chebyshev polynomials [4], [9] which are defined using trigonometric functions and are extensively used in mathematics. One can ask about existence of generalization of such a polynomials to polynomials of several variables [3] using the multivariate generalizations of trigonometric functions [6], [7], [8]. With the use of generalized multivariate trigonometric functions one can obtain eight classes of orthogonal polynomials of several variables.

*This work was supported by the Grant Agency of Czech Technical University in Prague, grant No. SGS13 /217/OHK4/3T/14

Chebyshev Polynomials of the first and the third kind and their generalizations by antisymmetric and symmetric multivariate cosine functions are studied in [5]. This generalizations lead to four classes of orthogonal polynomials of several variables. The possibility of such generalization leads to question of generalization of the Chebyshev polynomials of the second and the fourth kind using the antisymmetric and symmetric multivariate sine functions. In this paper we focus on symmetric multivariate generalization of the Chebyshev polynomials of the second kind.

Such a generalization uses symmetric multivariate trigonometric functions instead of classical trigonometric functions. This a functions follows lot of symmetries coming both from the definition of symmetric multivariate sine function and properties of classical sine function. These properties should be then applied to simplify the recurrence relations of generalized Chebyshev polynomials.

2 Chebyshev polynomials

The classical Chebyshev Polynomials of one variable are well known and extensively used class of orthogonal polynomials. There exist four kinds of the Chebyshev Polynomials defined as

$$\begin{aligned} \mathcal{P}_n^{I}(x) &= T_n(x) = \cos\left(n\theta\right), & \mathcal{P}_n^{III}(x) &= V_n(x) = \frac{\cos\left(\left(n+\frac{1}{2}\right)\theta\right)}{\cos\left(\frac{1}{2}\theta\right)}, \\ \mathcal{P}_n^{II}(x) &= U_n(x) = \frac{\sin\left(\left(n+1\right)\theta\right)}{\sin\left(\theta\right)}, & \mathcal{P}_n^{IV}(x) &= W_n(x) = \frac{\sin\left(\left(n+\frac{1}{2}\right)\theta\right)}{\sin\left(\frac{1}{2}\theta\right)}, \end{aligned} \tag{1}$$

with variable $x = \cos(\theta)$, $x \in [-1, 1]$.

For our purposes we will focus mainly on the Chebyshev polynomials of the second kind. From trigonometric definition one can see the orthogonality of these polynomials, i.e.,

$$\int_0^\pi \frac{\sin\left(n+1\right)\theta \sin\left(m+1\right)\theta}{\sin^2\theta} d\theta = 0, \quad n \neq m, \tag{2}$$

which in variable x takes form

$$\int_{-1}^{1} U_n(x)U_m(x)\left(1-x^2\right)^{\frac{1}{2}} dx = 0, \quad n \neq m. \tag{3}$$

Further we can obtain the first two polynomials by use of trigonometric formulas as:

$$U_1(x) = 1, \qquad U_2(x) = 2\cos\left(\theta\right) = 2x. \tag{4}$$

From theory of orthogonal polynomials we know that there exist recurrence relations connecting the three consecutive polynomials. Easiest way to obtain this formula is by using the trigonometric identity:

$$\sin\left((n+1)\theta\right) + \sin\left((n-1)\theta\right) = 2\cos\left(\theta\right)\sin\left(n\theta\right) \tag{5}$$

which leads to recurrence relations:

$$U_n(x) = 2xU_{n-1}(x) - U_{n-2}(x), \qquad n = 2, 3, \dots. \tag{6}$$

Together with knowledge of the first two polynomials this gives procedure for generating of polynomials. Such a relations can be obtained for all four kinds of the Chebyshev Polynomials [4].

3 Multivariate trigonometric functions

The symmetric and antisymmetric multivariate generalizations of trigonometric functions are defined and their properties detailed in [8]. The antisymmetric trigonometric functions $\cos^-_\lambda(x)$, $\sin^-_\lambda(x)$ and symmetric trigonometric functions $\cos^+_\lambda(x)$, $\sin^+_\lambda(x)$ of variable $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ with parameter $\lambda = (\lambda_1, \dots, \lambda_n)$ are defined as determinants and permanents of matrices with entries $\cos(\pi\lambda_i x_j)$ resp. $\sin(\pi\lambda_i x_j)$, i.e.

$$\begin{aligned} \cos^-_\lambda(x) &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma)\cos(\pi\lambda_{\sigma_1}x_1)\cos(\pi\lambda_{\sigma_2}x_2)\cdots\cos(\pi\lambda_{\sigma_n}x_n), \\ \sin^-_\lambda(x) &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma)\sin(\pi\lambda_{\sigma_1}x_1)\sin(\pi\lambda_{\sigma_2}x_2)\cdots\sin(\pi\lambda_{\sigma_n}x_n), \end{aligned} \tag{7}$$

for the antisymmetric trigonometric functions and

$$\begin{aligned} \cos^+_\lambda(x) &= \sum_{\sigma\in S_n} \cos(\pi\lambda_{\sigma_1}x_1)\cos(\pi\lambda_{\sigma_2}x_2)\cdots\cos(\pi\lambda_{\sigma_n}x_n), \\ \sin^+_\lambda(x) &= \sum_{\sigma\in S_n} \sin(\pi\lambda_{\sigma_1}x_1)\sin(\pi\lambda_{\sigma_2}x_2)\cdots\sin(\pi\lambda_{\sigma_n}x_n), \end{aligned} \tag{8}$$

for the symmetric trigonometric functions.

For our applications we will only consider functions $\cos^\pm_k(x)$ and $\sin^\pm_k(x)$ with integer parameter only, $k \in \mathbb{Z}^n$ and a shift $\rho = \left(\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}\right)$ and parameters k only lexicographically ordered due (anti)symmetries, i.e.,

$$k_1 \geq k_2 \geq \ldots \geq k_n. \tag{9}$$

Due to further properties we consider the functions only on closure of the fundamental domain $F(\widetilde{S}_n^{\text{aff}})$ of the form

$$F(\widetilde{S}_n^{\text{aff}}) = \{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n \mid 1 \geq x_1 \geq x_2 \geq \ldots \geq x_n \geq 0\}. \tag{10}$$

Because of additional properties discussed in [1] we can omit boundaries

- $x_i = x_{i+1}, i \in \{1, \dots, n-1\}$ for $\cos^-_k(x)$ and $\sin^-_k(x)$,
- $x_i = x_{i+1}, i \in \{1, \dots, n-1\}$ or $x_1 = 1$ for $\cos^-_{k+\rho}(x)$ and $\sin^-_{k+\rho}(x)$,
- $x_i = 1, i \in \{1, \dots, n\}$ for $\cos^+_{k+\rho}(x)$ and $\sin^+_{k+\rho}(x)$.

4 Multivariate symmetric generalization of the Chebyshev polynomials of the second kind

Using the multivariate generalization of trigonometric functions one can obtain multivariate versions of the Chebyshev polynomials. The generalization of the Chebyshev polynomials of the first and the third kind was done in [5]. For symmetric generalization of the Chebyshev polynomials of the second kind let us introduce the n functions $X_1, X_2, \dots, X_n$

$$X_1 = \cos^+_{(1,0,\dots,0)}, \quad X_2 = \cos^+_{(1,1,\dots,0)}, \quad \dots, \quad X_n = \cos^+_{(1,1,\dots,1)}, \tag{11}$$

We now consider the multivariate symmetric generalization of the Chebyshev polynomials of the second kind as:

$$\mathcal{P}^{III,+}_k (X_1, X_2, \dots, X_n) = \frac{\sin^+_{k+\rho_1}(x)}{\sin^+_{\rho_1}(x)}, \tag{12}$$

where $\rho_1 = (1,1,\dots,1)$. One can prove that these relations are valid for all points of interior of fundamental domain $F(\tilde{S}^{aff}_n), x \in F(\tilde{S}^{aff}_n)^\circ$.

We use ordering of polynomials from [5], we say that a polynomial $\mathcal{P}^{III,+}_k$ is greater than polynomial $\mathcal{P}^{III,+}_{k'}$, $k \neq k'$ if for all i, $k_i \geq k'_i$ and smaller if for all i, $k_i \leq k'_i$.

4.1 Recurrence relations

To obtain recurrence relations for polynomials $\mathcal{P}^{III,+}_{k'}$ we consider generalized trigonometric identity

$$\sin^+_k(x)\cos^+_l(x) = \frac{1}{2^n} \sum_{\sigma \in S_n} \sum_{\substack{a_i = \pm 1 \\ i=1,\dots,n}} \sin^+_{\left(k_1 + a_1 l_{\sigma(1)}, \dots, k_n + a_n l_{\sigma(n)}\right)}(x). \tag{13}$$

Using the special case where $l = \rho_1 = (1,1,\dots,1)$, i.e,

$$\sin^+_k(x)\cos^+_{\rho_1}(x) = \frac{n!}{2^n} \sum_{\substack{a_i = \pm 1 \\ i=1,\dots,n}} \sin^+_{(k_1+a_1,\dots,k_n+a_n)}(x), \tag{14}$$

we obtain recurrence relation:

$$\sin^+_k = \frac{2^n}{n!} \sin^+_{k-l_1-l_2-\dots-l_n} X_n - \sum_i^n \sin^+_{k-2l_i} - \sum_{\substack{i,j=1 \\ i<j}}^n \sin^+_{k-2l_i-2l_j} - \dots - \sin^+_{k-2l_1-2l_2-\dots-2l_n}. \tag{15}$$

where l_i is vector with 1 only on i-th coordinate.

Using this relation and properties of function $\sin^+_k$ each polynomial can be expressed as linear combination of lower polynomials and product of lower polynomial with X_n.

4.2 Three-dimensional polynomials

Relations (15) together with properties of generalized sine functions imply the following recurrence relations for $\mathcal{P}^{III,+}_{(k_1,k_2,k_3)}$. The first four polynomials can be obtained using trigonometric identities in form:

$$\mathcal{P}^{II,+}_{(0,0,0)} = 1, \quad \mathcal{P}^{II,+}_{(1,0,0)} = \frac{1}{3}X_1, \quad \mathcal{P}^{II,+}_{(1,1,0)} = \frac{2}{3}X_2, \quad \mathcal{P}^{II,+}_{(1,1,1)} = \frac{4}{3}X_3. \tag{16}$$

Following polynomials are then obtained using recurrence relations:

$$k_1 \geq 2, k_2 = k_3 = 0: \qquad \mathcal{P}^{II,+}_{(k_1,0,0)} = \mathcal{P}^{II,+}_{(k_1-1,0,0)}X_1 - \mathcal{P}^{II,+}_{(k_1-2,0,0)} - 2\mathcal{P}^{II,+}_{(k_1-1,1,0)}$$

$$\begin{aligned} k_1 - 1 > k_2 > k_3 = 0: \qquad \mathcal{P}^{II,+}_{(k_1,k_2,0)} &= \mathcal{P}^{II,+}_{(k_1-1,k_2,0)}X_1 - \mathcal{P}^{II,+}_{(k_1-2,k_2,0)} \\ &- \mathcal{P}^{II,+}_{(k_1-1,k_2+1,0)} - \mathcal{P}^{II,+}_{(k_1-1,k_2-1,0)} - \mathcal{P}^{II,+}_{(k_1-1,k_2,1)} \end{aligned}$$

$$\begin{aligned} k_1 - 1 >, k_2 = k_3 > 0: \qquad \mathcal{P}^{II,+}_{(k_1,k_2,k_2)} &= \mathcal{P}^{II,+}_{(k_1-1,k_2,k_2)}X_1 - \mathcal{P}^{II,+}_{(k_1-2,k_2,k_2)} \\ &- 2\mathcal{P}^{II,+}_{(k_1-1,k_2+1,k_2)} - 2\mathcal{P}^{II,+}_{(k_1-1,k_2,k_2-1)} \end{aligned}$$

$$\begin{aligned} k_1 - 1 >, k_2 > k_3 > 0: \qquad \mathcal{P}^{II,+}_{(k_1,k_2,k_3)} &= \mathcal{P}^{II,+}_{(k_1-1,k_2,k_3)}X_1 - \mathcal{P}^{II,+}_{(k_1-2,k_2,k_3)} - \mathcal{P}^{II,+}_{(k_1-1,k_2+1,k_3)} \\ &- \mathcal{P}^{II,+}_{(k_1-1,k_2-1,k_3)} - \mathcal{P}^{II,+}_{(k_1-1,k_2,k_3+1)} - \mathcal{P}^{II,+}_{(k_1-1,k_2,k_3-1)} \end{aligned}$$

$$\begin{aligned} k_1 - 1 = k_2 > k_3 = 0: \qquad \mathcal{P}^{II,+}_{(k_1,k_1-1,0)} &= \frac{1}{2}\mathcal{P}^{II,+}_{(k_1-1,k_1-1,0)}X_1 - \mathcal{P}^{II,+}_{(k_1-1,k_1-2,0)} \\ &- \frac{1}{2}\mathcal{P}^{II,+}_{(k_1-1,k_1-1,1)} \end{aligned}$$

$$\begin{aligned} k_1 - 1 = k_2 > k_3 > 0: \qquad \mathcal{P}^{II,+}_{(k_1,k_1-1,k_3)} &= \frac{1}{2}\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_3)}X_1 - \mathcal{P}^{II,+}_{(k_1-1,k_1-2,0)} \\ &- \frac{1}{2}\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_3+1)} - \frac{1}{2}\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_3-1)} \end{aligned}$$

$$k_1 - 1 = k_2 = k_3 > 0: \qquad \mathcal{P}^{II,+}_{(k_1,k_1-1,k_1-1)} = \frac{1}{3}\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-1)}X_1 - \mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-2)}$$

$$\begin{aligned} k_1 = k_2 = 2, k_3 = 0: \qquad \mathcal{P}^{II,+}_{(2,2,0)} &= 2\mathcal{P}^{II,+}_{(1,1,0)}X_2 - 2\mathcal{P}^{II,+}_{(1,0,0)}X_1 \\ &- \mathcal{P}^{II,+}_{(1,1,1)}X_1 + \mathcal{P}^{II,+}_{(0,0,0)} + 5\mathcal{P}^{II,+}_{(1,1,0)} \\ &+ \mathcal{P}^{II,+}_{(2,1,1)} \end{aligned}$$

$$\begin{aligned} k_1 = k_2 > 2, k_3 = 0: \qquad \mathcal{P}^{II,+}_{(k_1,k_1,0)} &= 2\mathcal{P}^{II,+}_{(k_1-1,k_1-1,0)}X_2 - 2\mathcal{P}^{II,+}_{(k_1-1,k_1-2,0)}X_1 \\ &- \mathcal{P}^{II,+}_{(k_1-1,k_1-1,1)}X_1 + \mathcal{P}^{II,+}_{(k_1-2,k_1-2,0)} + 3\mathcal{P}^{II,+}_{(k_1-1,k_1-1,0)} \\ &+ 2\mathcal{P}^{II,+}_{(k_1-1,k_1-2,1)} + 2\mathcal{P}^{II,+}_{(k_1-1,k_1-3,0)} + \mathcal{P}^{II,+}_{(k_1-1,k_1-1,2)} \end{aligned}$$

$$\begin{aligned}
k_1 = k_2 > k_3 + 2 > 2 : \quad \mathcal{P}^{II,+}_{(k_1,k_1,k3)} &= 2\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k3)} X_2 - 2\mathcal{P}^{II,+}_{(k_1-1,k_1-2,k3)} X_1 \\
&\quad - \mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_3+1)} X_1 - \mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_3-1)} X_1 \\
&\quad + \mathcal{P}^{II,+}_{(k_1-2,k_1-2,k_3)} + 2\mathcal{P}^{II,+}_{(k_1-1,k_1-2,k_3+1)} \\
&\quad + 2\mathcal{P}^{II,+}_{(k_1-1,k_1-2,k_3-1)} + 4\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_3)} \\
&\quad + 2\mathcal{P}^{II,+}_{(k_1-1,k_1-3,k_3)} + \mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_3+2)} \\
&\quad + \mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_3-2)}
\end{aligned}$$

$$\begin{aligned}
k_1 = k_2 = k_3 + 2 = 3 : \quad \mathcal{P}^{II,+}_{(3,3,1)} &= 2\mathcal{P}^{II,+}_{(2,2,2)} X_2 - 2\mathcal{P}^{II,+}_{(2,1,1)} X_1 - \frac{2}{3}\mathcal{P}^{II,+}_{(2,2,2)} X_1 \\
&\quad - \mathcal{P}^{II,+}_{(2,2,0)} X_1 + \mathcal{P}^{II,+}_{(1,1,1)} + 5\mathcal{P}^{II,+}_{(2,2,1)} \\
&\quad + 4\mathcal{P}^{II,+}_{(2,1,0)}
\end{aligned}$$

$$\begin{aligned}
k_1 = k_2 = k_3 + 2 > 3 : \quad \mathcal{P}^{II,+}_{(k_1,k_1,k_1-2)} &= 2\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-2)} X_2 - 2\mathcal{P}^{II,+}_{(k_1-1,k_1-2,k_1-2)} X_1 \\
&\quad - \frac{2}{3}\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-1)} X_1 - \mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-3)} X_1 \\
&\quad + \mathcal{P}^{II,+}_{(k_1-2,k_1-2,k_1-2)} + 5\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-2)} \\
&\quad + 4\mathcal{P}^{II,+}_{(k_1-1,k_1-2,k_1-3)} + \mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-4)}
\end{aligned}$$

$$\begin{aligned}
k_1 = k_2 = k_3 + 1 = 2 : \quad \mathcal{P}^{II,+}_{(2,2,1)} &= \frac{2}{3}\mathcal{P}^{II,+}_{(1,1,1)} X_2 - \mathcal{P}^{II,+}_{(1,1,0)} X_1 \\
&\quad + \mathcal{P}^{II,+}_{(1,0,0)} + \mathcal{P}^{II,+}_{(1,1,1)}
\end{aligned}$$

$$\begin{aligned}
k_1 = k_2 = k_3 + 1 > 2 : \quad \mathcal{P}^{II,+}_{(k_1,k_1,k_1-1)} &= \frac{2}{3}\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-1)} X_2 - \mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-2)} X_1 \\
&\quad + \mathcal{P}^{II,+}_{(k_1-1,k_1-2,k_1-2)} + \mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-1)} \\
&\quad + \mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-3)}
\end{aligned}$$

$$\begin{aligned}
k_1 = k_2 = k_3 = 2 : \quad \mathcal{P}^{II,+}_{(2,2,2)} &= \frac{4}{3}\mathcal{P}^{II,+}_{(1,1,1)} X_3 - 6\mathcal{P}^{II,+}_{(1,1,0)} X_2 + 3\mathcal{P}^{II,+}_{(1,0,0)} X_1 \\
&\quad + 2\mathcal{P}^{II,+}_{(1,1,1)} X_1 - \mathcal{P}^{II,+}_{(0,0,0)} - 6\mathcal{P}^{II,+}_{(1,1,0)}
\end{aligned}$$

$$\begin{aligned}
k_1 = k_2 = k_3 = 3 : \quad \mathcal{P}^{II,+}_{(3,3,3)} &= \frac{4}{3}\mathcal{P}^{II,+}_{(2,2,2)} X_3 - 6\mathcal{P}^{II,+}_{(2,2,1)} X_2 + 3\mathcal{P}^{II,+}_{(2,1,1)} X_1 \\
&\quad + 2\mathcal{P}^{II,+}_{(2,2,2)} X_1 + 3\mathcal{P}^{II,+}_{(2,2,0)} X_1 - \mathcal{P}^{II,+}_{(1,1,1)} \\
&\quad - 9\mathcal{P}^{II,+}_{(2,2,1)} - 6\mathcal{P}^{II,+}_{(2,1,0)}
\end{aligned}$$

$$
\begin{aligned}
k_1 = k_2 = k_3 > 3 : \qquad \mathcal{P}^{II,+}_{(k_1,k_1,k_1)} &= \frac{4}{3}\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-1)}X_3 - 6\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-2)}X_2 \\
&+ 3\mathcal{P}^{II,+}_{(k_1-1,k_1-2,k_1-2)}X_1 + 2\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-1)}X_1 \\
&+ 3\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-3)}X_1 - \mathcal{P}^{II,+}_{(k_1-2,k_1-2,k_1-2)} \\
&- 9\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-2)} - 6\mathcal{P}^{II,+}_{(k_1-1,k_1-2,k_1-3)} \\
&- 3\mathcal{P}^{II,+}_{(k_1-1,k_1-1,k_1-4)},
\end{aligned}
$$

which are obtained by the generalized trigonometric identity (15).

This gives the full set of recurrence relations which one can use to obtain exact form of polynomial with specific k. For $k \leq (2,2,2)$ the polynomials are:

$$
\begin{aligned}
\mathcal{P}^{II,+}_{(2,0,0)} &= \frac{1}{3}X_1^2 - \frac{4}{3}X_2 - 1, \\
\mathcal{P}^{II,+}_{(2,1,0)} &= \frac{1}{3}X_1X_2 - \frac{1}{3}X_1 - \frac{2}{3}X_3, \\
\mathcal{P}^{II,+}_{(2,1,1)} &= \frac{4}{9}X_1X_3 - \frac{2}{3}X_2, \\
\mathcal{P}^{II,+}_{(2,2,0)} &= \frac{2}{3}X_1^2 - \frac{4}{3}X_2^2 - \frac{4}{3}X_1X_2 + \frac{4}{9}X_1X_3 + \frac{8}{3}X_2 + 1, \\
\mathcal{P}^{II,+}_{(2,2,1)} &= -\frac{2}{3}X_1X_2 + \frac{8}{9}X_2X_3 + \frac{1}{3}X_1 + \frac{4}{3}X_1, \\
\mathcal{P}^{II,+}_{(2,2,2)} &= X_1^2 - 4X_2^2 + \frac{16}{9}X_3^2 + \frac{8}{3}X_1X_3 - 4X_2 - 1.
\end{aligned} \tag{17}
$$

From the knowledge of first ten polynomials one can see that the polynomial $\mathcal{P}^{II,+}_{(k_1,k_2,k_3)}$ is of order k_1 which can be proven generally.

5 Conclusion

We have shown the possibility of generalization of the Chebyshev polynomials of the second kind by use of symmetric multivariate sine function. Continuous orthogonality of these polynomial should follow the continuous orthogonality of generalized Chebyshev polynomials of the first and the third kind done in [5]. The antisymmetric generalization of the Chebyshev polynomials of the second kind and generalization of the Chebyshev polynomials of the fourth kind should follow similar procedure and is question of future work.

One should also be able to obtain cubature formulas for these new polynomials of several variables with use of generalized discrete multivariate trigonometric transforms. The cubature formulas for multivariate Chebyshev polynomials of the first and the third kind were already done in [5] with use of symmetric and antisymmetric discrete multivariate cosine transforms (SDMCT and ADMCT). The cubature formulas for multivariate Chebyshev polynomials of the second and the third kind is yet to be done.

References

[1] A. Brus, *Discrete Multivariate (Anti)Symmetric Trigonometric functions*, Doktorandské dny 2015, pp. 13-22, (2015).

[2] T. S. Chihara, *An Introduction to Orthogonal Polynomials*, Gordon and Breach, Science Publishers, Inc., 1978.

[3] C. F. Dunkel and Y. Xu, *Orthogonal Polynomials of Several Variables*, Encyclopedia Math. Appl. 81, Cambridge University Press, Cambridge, UK, 2001.

[4] D. C. Handscomb and J. C. Mason, *Chebyshev Polynomials*, Champman & Hall/CRC, Boca Raton, FL, 2003.

[5] J. Hrivnák and L. Motlochová, *Discrete Transforms and orthogonal polynomials of (anti)symmetric multivariate cosine functions*, SIAM J. Numer. Anal., Vol. 52, No 6, pp. 3021-3055 (2014).

[6] J. Hrivnák, L. Motlochová and J. Patera, *Two dimensional symmetric and antisymmetric generalization of sine functions*, J. Math. Phys., 51 (2010), 073509.

[7] J. Hrivnák and J. Patera, *Two dimensional symmetric and antisymmetric generalization of exponential and cosine functions*, J. Math. Phys., 51 (2010), 023515.

[8] A. Klimyk and J. Patera, *(Anti)symmetric multivariate trigonometric functions and corresponding Fourier transforms*, J. Math. Phys., 48 (2007), 093504.

[9] T. J. Rivlin, *The Chebyshev Polynomials*, John Wiley & Sons, New York, 1990.

Synchronization Effects in Pedestrian Dynamics on Complex Structures[*]

Marek Bukáček

4th year of PGS, email: marek.bukacek@fjfi.cvut.cz
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:
Milan Krbálek, Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague
Jaromír Kukal, Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. Microscopic studies recently published by our group [1][2][3][4][5] describe in detail many factors affecting motion of individuals though a bottleneck. The implementation of heterogeneous behavior into the most popular cellular automata model [6][7] closed this chapter and we focused to more complex structures.

The process of merging pedestrian streams emerging from simple rooms belongs to the most important parts of evacuation process. While the dynamic of motion inside a room is given mainly by individual preferences, the joining to the crowd and passing a bottleneck operates as a synchronization unit. Then, the stream of egressing pedestrian may be characterized by collective quantities until the individual motion due to the geometry or actual situation prevails the advantages of the cooperative process.

This phenomenon affects measurable quantities in several situations. The first insight was provided within the leave the room experiment (E4) [8][9], where the two meters long narrow corridor followed the exit with similar width. As implies from the mass conservation law, measurements of flow brought the same results along all corridor, but the time headways revealed lower variance with increasing distance from the exit.

The effect of synchronization on tree structure was tested under the egress experiment (E5) the spring 2016. Pedestrians' camming within one of four asynchronous streams arrived to one of two parallel bottlenecks that led to central room with one main exit. The observations confirmed expected effect, the decrease of variance was observed on each level of the tree structure [10].

This experiment is interested even from the general point of view. Similar experiment was organized even by our colleagues from Krakow and both geometries were implemented in to our variant of cellular automata model and into polish Social distance model. Detailed comparison of all result sets is summarized in [11], in general we can say that our model fits well both experimental data, even without non-standard calibration. Only the time step was calibrated to fit the door with.

The challenge of estimations on complex structure is the real time recalibration, i.e. the usage of measured performance in one segment for short time prediction of performance in consecutive node. Assuming the knowledge of time distance between the nodes and the constant capacity of bottlenecks, the estimator based on queuing theory was constructed. As shown the

[*] GAČR 15-15049S & SGS15/214/OHK4/3T/14

results evaluated for mentioned design E5, this simple model is more correct than expected [12]. The main drawdown is the constant capacity – as shown before [2], the capacity of bottleneck is increasing with the size of the crowd.

The concept of synchronization was tested even on the dataset collected within the large evacuation experiment of City Elephant train unit [13], where we contributed by data analysis under the developed cooperation with the Faculty of Civil Engineering CTU. The significantly higher exit door width enabled comfortable passing of two persons simultaneously, therefore the study focused on the merging the streams from top and bottom floor. With respect to the exit area geometry, the evacuation process was very asymmetric, the top floor passengers left the train the last bottom floor participants. This feature implies ineffective load of the main exit and therefore to higher evacuation time [10].

Even with many conclusions summarized in mentioned publications, this research is not finished. Upcoming work covers the application of advanced statistics and development of more sophisticated queuing theory model.

Keywords: pedestrian dynamics, egress experiments, synchronization

Abstrakt. Mikroskopické studie dynamiky davu, které jsme publikovali v posledních letech [1][2][3][4][5] do detailu popisují faktory, které ovlivňují pohyb různých jednotlivců jedním zúženým místem. Implementací heterogenního chování do nejčastěji využívaného celulárního modelu [6][7], jsme tuto kapitolu uzavřeli a přesunuli se na úroveň komplexnější infrastruktury.

Mezi typické prvky evakuačního procesu patří bezesporu proces slučování několika proudů chodců, kteří vycházení z jednotlivých místností. Zatímco dynamika je v těchto místnostech daná převážně individuálními parametry, zapojení se do davu a průchod zúženým místem působí jako synchronizační prvek. Výstupní proud chodců je pak charakterizovatelný kolektivními veličinami až do doby, než u jednotlivců z důvodů geometrie či aktuálních podmínek převáží soutěživé chován nad potenciální výhodou individuálního postupu.

Tento jev ovlivňuje měřitelné veličiny v různých situacích, prvotní pozorování jsme provedli v rámci experimentu opuštění místnosti (E4) [8][9] kde za úzkým průchodem následoval dvoumetrový koridor obdobné šířky. Jak vyplývá i zákona zachování hmoty, měření toku na různých místech koridoru přinášelo stejné výsledky, ale měřené časové odstupy vykazovaly nižší rozptyl s rostoucí vzdáleností od východu.

V rámci experimentu E5 realizovaném na jaře 2016 byl efekt synchronizace vyhodnocován na stromové struktuře. Chodci přicházeli ve čtyřech nesynchronizovaných proudech ke dvěma průchodům, které vedly do centrální místnosti s jedním východem. Pozorování potvrdilo očekávaný efekt, pokles rozptylu časových odstupů nenastal pouze při porovnání neuspořádaných vstupů čtyř proudů s výstupy první úrovně, rozptyl odstupů měřených na hlavním výstupu byl ještě nižší [10].

Tento experiment je zajímavý i z pohledu procesu evakuace obecně. Pro srovnání byl proveden obdobný experiment polskými spolupracovníky a obě řešené geometrie byly implementovány do našeho již dříve představeného CA modelu a polského modelu Social distance. Podrobné srovnání všech čtyř sad výsledků je shrnuté v článku [11], obecně lze říci, že náš model bez speciální kalibrace dobře odpovídá experimentálním datům. Jediným zásahem byla úprava časového kroku v závislosti na různé šířce dveří.

Výzvou při odhadování na komplexní struktuře je možnost využití pozorování předchozích uzlech pro predikci stavů na uzlech následujících. Za předpokladu známé časové vzdálenosti mezi jednotlivými uzly a konstantní kapacitě průchodů dané pouze jejich šířkou je možné zkonstruovat deterministický model založený na teorii front. Jak ukazují výsledky vyhodnocené pro výše popsaný design E5, tento jednoduchý model je překvapivě úspěšný [12]. Hlavním nedostatkem

je právě konstantní kapacita – jak jsme ukázali dříve [2], s velikostí davu dochází k nárůstu toku.

Koncept synchronizace při slučování toků byl testován i na datech z rozsáhlého experimentu evakuace vlakové jednotky City Elefant [13], na jehož vyhodnocení jsme se podíleli v rámci navázaná spolupráce s FS ČVUT. Výrazně vyšší šířka výstupních dveří umožňovala pohodlný průchod dvou osob současně, proto byla studie zaměřena více na dynamiku slučování skupin z horního a dolního patra. Vzhledem ke geometrii prostoru kolem výstupu probíhala evakuace velmi asymetricky, v některých scénářích se první zástupci horní skupiny dostali ven až poté, co vlak opustili poslední spodní chodci. Tato varianta vede k neefektivnímu využívání východu a tedy vyššímu evakuačnímu času [10].

Přes mnoho závěrů shrnutých v uvedených publikacích ještě není výzkum tohoto tématu ukončen, další práce zahrnuje aplikaci pokročilejších statistik a vývoj pokročilejšího modelu založeného na teorii front.

Klíčová slova: pohyb chodců, evakuační experimenty, synchronizace

References

[1] M. Bukáček, P. Hrabák and M. Krbálek, *Experimental Analysis of Two-Dimensional Pedestrian Flow in front of the Bottleneck*, In: 'Traffic and Granular Flow '13' (2014), 93–101.

[2] M. Bukáček, P. Hrabák and M. Krbálek, *Experimental Study of Phase Transition in Pedestrian Flow*, In: 'PED 2014 Proc.', Transportation Research Procedia **2** (2014), 105–113.

[3] P. Hrabák, M. Bukáček and M. Krbálek, *Cellular Model of Room Evacuation Based on Occupancy and Movement Prediction, Comparison with Experimental Study*, JCA **8** (2013), 383–395.

[4] M. Bukáček, P. Hrabák and M. Krbálek, *Cellular Model of Pedestrian Dynamics with Adaptive Time Span*, In: 'PPAM 2013 Proc.', LNCS **8385** (2014), 669–678.

[5] M. Bukáček and P. Hrabák, *Boundary Induced Phase Transition in Cellular Automata Models of Pedestrian Flow*, JCA **11/4** (2016), 327–338.

[6] M. Bukáček and P. Hrabák, *Conflict Solution According to Aggressiveness of Agents in Floor-Field-Based Model*, In: 'PPAM 2015 Proc.', LNCS **9574** (2016), 507–516.

[7] P. Hrabák and M. Bukáček, *Influence of Agents Heterogeneity in Cellular Model of Evacuation*, JCS, accepted. Available at: http://dx.doi.org/10.1016/j.jocs.2016.08.002

[8] P. Hrabák, M. Bukáček and M. Krbálek, *Individual Microscopic Results Of Bottleneck Experiments*, In: 'Traffic and Granular Flow '15', accepted. Available at: https://arxiv.org/abs/1603.02019

[9] M. Bukáček, P. Hrabák and M. Krbálek, *Microscopic Travel Time Analysis of Bottleneck Experiments*, Transportmetrica A, under revisions.

[10] M. Bukáček, H. Najmanová and P. Hrabák, *The Effects of Synchronization of Pedestrian Flow through Multiple Bottlenecks* – Train Egress Study, In: 'PED 2016 Proc.', accepted.

[11] P. Hrabák, J. Porzicky, M. Bukáček at all, *Advanced CA Crowd Models of Multiple Consecutive Bottlenecks*, In: 'ACRI 2016 Proc.', LNCS **9863** (2016), 396–404.

[12] J. Porzicky, P. Hrabák and M. Bukáček at all, *Data driven method of pedestrian flow estimation for evacuation scenario using queuing model*, In: 'EG-ICE 2016 Proc.', accepted.

[13] H. Najmanová, P. Hejtmánek and M. Bukáček. Počrní bezpečnost osobních kolejových vozidel: Analýza evakuace osob z dvoupodlažní jednotky CityElefant, In: 'Požární ochrana 2016' (2016), 294–301.

On the Investigation of Julia Sets Using Rotational Spectrum*

Martin Dlask

1st year of PGS, email: martindlask@centrum.cz
Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jaromír Kukal, Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. The correlation dimension is one of many types of fractal dimension. It is usually estimated from finite number of points from fractal set using correlation sum and regression in log-log plot. However, this traditional approach requires a large number of data and often leads to a biased estimate. Our novel approach employs the spectrum of point set which is averaged via rotation of power pattern. Resulting spectral characteristic was proven to have suitable properties in infinite-dimensional space. The theoretical results can be directly applied to uniformly distributed samples from given point set. The efficiency of proposed method was tested using Monte Carlo simulation on the sets with known correlation dimension. Additionally, the correlation dimension of Julia sets is experimentally calculated.

Keywords: point set, correlation dimension, power spectrum, rotation, Monte Carlo

Abstrakt. Korelační dimenze představuje jeden z mnoha typů fraktální dimenzí. Nejčastěji se odhaduje z konečného počtu datových bodů z fraktální množiny pomocí korelační sumy a regrese v log-log grafu. Tento tradiční přístup vyžaduje velké množství dat a často vede k vychýleným odhadům dimenze. Nový prezentovaný přístup využívá výkonového spektra bodové množiny které je průměrováno přes všechny možné rotace. Prokázalo se, že výsledná spektrální charakteristika má vhodné vlastnosti v případě rotace v nekonečně-rozměrném prostoru. Efektivnost metody byla testována metodou Monte Carlo na množinách se známou korelační dimenzí. Dále byla metodika použita pro experimentální odhad korelační dimenze Juliových množin.

Klíčová slova: bodová možia, korelační dimenze, výkonové spektrum, rotace, Monte Carlo

1 Introduction

Correlation dimension D_2 is a popular tool for fractal dimension estimation and belongs to the family of entropy-based fractal dimensions such as capacity dimension D_0, information dimension D_1 and their generalization Renyi dimension D_α for $\alpha \geq 0$.

Traditional approach of correlation dimension estimation is based on Grassberger and Procaccia algorithm and is widely used in biomedicine for EEG signal analysis [13, 12] or in cardiology [7]. Recently, new approaches of correlation dimension estimation were presented using weighting function [9] or methods suitable for high-dimensional

*This work has been supported by the grant SGS14/208/OHK4/3T/14.

signals [10]. The linear regression model, on which the majority of methods are based, provides often biased estimate of fractal dimension, and therefore there were some efforts of improving this procedure in [6].

In this work, we present a novel approach of correlation dimension estimation that is based on rotation of the power spectrum of point set. The proposed method is stable even for small amount of points and the resulting characteristic has smooth development.

2 Correlation Dimension

The correlation dimension, introduced by Grassberger and Procaccia [5, 4] employs in its definition measuring the distance between all pairs of points in the investigated set. For arbitrary set $\mathcal{F} \subset \mathbb{R}^n$, the *correlation sum* is defined for $r > 0$ as the limit case

$$\mathrm{C}(r) = \lim_{N\to\infty} \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \mathrm{I}(\|\boldsymbol{x_i} - \boldsymbol{x_j}\| \leq r) \tag{1}$$

where $\|.\|$ denotes Euclidean norm that is rotation invariant, I is the indicator function and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ are vectors from $\mathcal{F}$. Due to the fact, that the correlation dimension expresses the relative amount of points whose distance is less than r, the correlation sum can be rewritten as

$$\mathrm{C}(r) = \mathop{\mathrm{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\mathrm{U}(\mathcal{F})} \mathrm{I}(\|\mathbf{x} - \mathbf{y}\| \leq r) = \mathop{\mathrm{prob}}_{\boldsymbol{x},\boldsymbol{y}\sim\mathrm{U}(\mathcal{F})} (\|\boldsymbol{x} - \boldsymbol{y}\| \leq r)\,, \tag{2}$$

for $\boldsymbol{x}, \boldsymbol{y}$ that are uniformly distributed on $\mathcal{F}$. Therefore $\mathrm{C}(r)$ is cumulative distribution function of random variable $r = \|\boldsymbol{x} - \boldsymbol{y}\|$. The correlation dimension D_2 of set $\mathcal{F}$ is based on correlation sum and is defined as

$$D_2 = \lim_{r\to 0^+} \frac{\ln \mathrm{C}(r)}{\ln r}, \tag{3}$$

if the limit exists.

3 Continuous Spectrum of Point Set

The Fourier transform of n-dimensional set $\mathcal{F} \subset \mathbb{R}^n$ is defined using operator of expected value [3] as

$$\mathrm{F}(\boldsymbol{\omega}) = \mathop{\mathrm{E}}_{\boldsymbol{x}\sim\mathrm{U}(\mathcal{F})} \exp(-\mathrm{i}\boldsymbol{\omega} \cdot \boldsymbol{x}) \tag{4}$$

for angular frequency $\boldsymbol{\omega} \in \mathbb{R}^n$ and for $\boldsymbol{x}$ uniformly distributed on $\mathcal{F}$. The *power spectrum* of set $\mathcal{F}$ equals $\mathrm{P}(\boldsymbol{\omega}) = |\mathrm{F}(\boldsymbol{\omega})|^2 = \mathrm{F}(\boldsymbol{\omega}) \cdot \mathrm{F}^*(\boldsymbol{\omega})$, where F^* is complex conjugate of F. Moreover, it can be expressed as

$$\mathrm{P}(\boldsymbol{\omega}) = \mathop{\mathrm{E}}_{\boldsymbol{x}\sim\mathrm{U}(\mathcal{F})} \mathop{\mathrm{E}}_{\boldsymbol{y}\sim\mathrm{U}(\mathcal{F})} \exp(-\mathrm{i}\boldsymbol{\omega} \cdot \boldsymbol{x}) \exp(\mathrm{i}\boldsymbol{\omega} \cdot \boldsymbol{y}) = \mathop{\mathrm{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\mathrm{U}(\mathcal{F})} \exp(-\mathrm{i}\boldsymbol{\omega} \cdot (\boldsymbol{x} - \boldsymbol{y})), \tag{5}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are iid from $\mathcal{F}$. The power spectrum is frequently used for fractal set investigation [14, 15, 1]. When the research is physically motivated, it is usual to denote the angular frequency as $\omega = 2\pi/\lambda$ for wavelength λ of x-ray or light beam.

4 Rotational Spectrum

The goal of novel method is to obtain one-dimensional function as a derivate of the power spectrum that is useful in fractal analysis. The procedure is inspired by Debye [2] and his x-ray diffraction method that is often referred as Debye-Scherrer method. We denote $\mathrm{SO}(n)$ as the group of all rotations in $\mathbb{R}^n$ around origin. Because any rotation $\mathrm{R} \in \mathrm{SO}(n)$ is linear transform, the following equation holds

$$\mathrm{R}(\boldsymbol{x}) - \mathrm{R}(\boldsymbol{y}) = \mathrm{R}(\boldsymbol{x} - \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| \cdot \boldsymbol{\xi}, \tag{6}$$

where $\boldsymbol{\xi}$ is direction vector satisfying $\|\boldsymbol{\xi}\| = 1$ and $\boldsymbol{\xi} \in \mathcal{S}_{n-1}$ for n-dimensional sphere $\mathcal{S}_{n-1} = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| = 1\}$. Using factorization of angular frequency $\boldsymbol{\omega} = \Omega \cdot \boldsymbol{\psi}$ for $\Omega \in \mathbb{R}_0^+$ and normalization vector $\boldsymbol{\psi} \in \mathcal{S}_{n-1}$, we can define *rotational spectrum* as

$$\mathrm{S}(\Omega) = \mathop{\mathrm{E}}_{\mathrm{R} \in \mathrm{SO}(n)} \mathop{\mathrm{E}}_{\boldsymbol{\psi} \in \mathcal{S}_{n-1}} \mathop{\mathrm{E}}_{\boldsymbol{x}, \boldsymbol{y} \sim \mathrm{U}(\mathcal{F})} \exp(-\mathrm{i}\Omega \boldsymbol{\psi} \mathrm{R}(\boldsymbol{x} - \boldsymbol{y})), \tag{7}$$

which can be expressed explicitly.

Theorem 1. *Rotational spectrum can be expressed as*

$$\mathrm{S}(\Omega) = \mathop{\mathrm{E}}_{\boldsymbol{x}, \boldsymbol{y} \sim \mathrm{U}(\mathcal{F})} \mathrm{H}_n(\Omega \|\boldsymbol{x} - \boldsymbol{y}\|) \tag{8}$$

where

$$\mathrm{H}_n(q) = \frac{2^{\frac{n-2}{2}} \cdot \Gamma\left(\frac{n}{2}\right)}{q^{\frac{n-2}{2}}} \mathrm{J}_{\frac{n-2}{2}}(q). \tag{9}$$

Proof. Due to the fact that every rotation is linear transform, we can rewrite the rotational spectrum as

$$\mathrm{S}(\Omega) = \mathop{\mathrm{E}}_{\boldsymbol{x}, \boldsymbol{y} \sim \mathrm{U}(\mathcal{F})} \mathop{\mathrm{E}}_{\boldsymbol{\psi}, \boldsymbol{\xi} \in \mathcal{S}_{n-1}} \exp(-\mathrm{i}\Omega \|\boldsymbol{x} - \boldsymbol{y}\| \boldsymbol{\psi} \cdot \boldsymbol{\xi}). \tag{10}$$

The angle ν between vectors $\boldsymbol{\psi}$ and $\boldsymbol{\xi}$ satisfies $\cos\nu = \boldsymbol{\psi} \cdot \boldsymbol{\xi}$. Without loss of generality, we can set $\boldsymbol{\xi} = (1, 0, 0, \ldots, 0)$ and rewrite the rotational spectrum as

$$S(\Omega) = \mathop{\mathrm{E}}_{\boldsymbol{x}, \boldsymbol{y} \in \mathcal{F}} H_n\left(\Omega \|\boldsymbol{x} - \boldsymbol{y}\|\right) \tag{11}$$

where function $\mathrm{H}_n : \mathbb{R} \mapsto \mathbb{C}$ is defined as

$$\mathrm{H}_n(q) = \mathop{\mathrm{E}}_{\substack{\boldsymbol{\psi} \in \mathcal{S}_{n-1} \\ \psi_1 = \cos\nu}} \exp(-\mathrm{i}q \cos\nu) \tag{12}$$

For $n = 1$ we obtain a degenerated rotation together with $\nu \in \{0, \pi\}$, therefore the kernel function H_1 equals

$$\mathrm{H}_1(q) = \frac{\exp(-\mathrm{i}q) + \exp(\mathrm{i}q)}{2} = \cos q. \tag{13}$$

In case $n \geq 2$ we can express the kernel function using integral formula

$$\mathrm{H}_n(q) = \frac{\mathrm{I}_1(q)}{\mathrm{I}_2(q)} = \frac{\int_0^\pi \exp(-\mathrm{i}q \cos\nu) \sin^{n-2}\nu \,\mathrm{d}\nu}{\int_0^\pi \sin^{n-2}\nu \,\mathrm{d}\nu}. \tag{14}$$

Poisson integral [8] formula for Bessel function $\mathrm{J}_p(q)$ of first kind in the form

$$\mathrm{J}_p(q) = \frac{\left(\frac{q}{2}\right)^p}{\Gamma\left(p+\frac{1}{2}\right)\sqrt{\pi}} \int_0^{\pi} \exp(-\mathrm{i}q\cos\nu)\sin^{2p}\nu\,\mathrm{d}\nu \tag{15}$$

allows to rewrite the integral in nominator as

$$\mathrm{I}_1(q) = \frac{\mathrm{J}_p(q)\Gamma\left(p+\frac{1}{2}\right)\sqrt{\pi}}{\left(\frac{q}{2}\right)^p}, \tag{16}$$

whereas the integral in denominator is a limit case of Poisson formula

$$\mathrm{I}_2(q) = \lim_{q\to 0} \frac{\mathrm{J}_p(q)\Gamma\left(p+\frac{1}{2}\right)\sqrt{\pi}}{\left(\frac{q}{2}\right)^p} = \frac{\Gamma\left(p+\frac{1}{2}\right)\sqrt{\pi}}{\Gamma\left(p+1\right)}. \tag{17}$$

For $p = \frac{n-2}{2}$ we obtain the final form of kernel function expressed by Bessel function $\mathrm{J}_p(q)$ as

$$\mathrm{H}_n(q) = \frac{2^{\frac{n-2}{2}}\cdot\Gamma\left(\frac{n}{2}\right)}{q^{\frac{n-2}{2}}}\mathrm{J}_{\frac{n-2}{2}}(q). \tag{18}$$

Applying $\mathrm{H}_n(q)$ for $n = 1$ we obtain $\mathrm{H}_1(q) = \cos q$ as a particular case which extends the formula (18) range to $n \in \mathbb{R}$. □

The rotation can be performed in any space which dimension n is not less than the dimension m of original space of $\mathcal{F}$. When the dimension of rotation is greater than m, any vector $\boldsymbol{x} \in \mathcal{F}$ is completed with zeros for the remaining $n - m$ coordinates to have sufficient length. The most valuable result can be obtained in the case of rotation in infinite-dimensional space.

Theorem 2. *The scaled limit case of kernel function* H_n *is the Gaussian function i.e.*

$$\lim_{n\to\infty} \mathrm{H}_n(t\sqrt{n}) = \exp\left(-\frac{t^2}{2}\right). \tag{19}$$

Proof. For the investigation of behavior of kernel function when $n \to \infty$ we use Taylor expansion of $H_n(q)$ centered at $q_0 = 0$

$$\mathrm{H}_n(q) = \sum_{k=0}^{\infty} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2}+k)k!}\left(-\frac{q^2}{4}\right)^k, \tag{20}$$

using substitution $q = t\sqrt{n}$ it can be transformed into

$$\mathrm{H}_n(t\sqrt{n}) = \sum_{k=0}^{\infty} \frac{1}{k!}\left(-\frac{t^2}{2}\right)^k \frac{\Gamma(\frac{n}{2})n^k}{\Gamma(\frac{n}{2}+k)2^k}. \tag{21}$$

For every $k \in \mathbf{N}$ it holds that

$$\lim_{n\to\infty} \frac{\Gamma(\frac{n}{2})n^k}{\Gamma(\frac{n}{2}+k)2^k} = 1 \tag{22}$$

therefore the limit case of kernel function equals

$$\lim_{n\to\infty} \mathrm{H}_n(t\sqrt{n}) = \exp\left(-\frac{t^2}{2}\right) \tag{23}$$

□

For simplicity, we use this notation in the following sections

$$\mathrm{H}_\infty(q) = \exp\left(-\frac{q^2}{2}\right). \tag{24}$$

5 Relationship to Correlation Dimension

In this section we discuss the relationship between the rotational spectrum for limit kernel H_∞ and the correlation dimension. The correlation sum is cumulative distribution function of the distances between the points in the fractal set, therefore the rotational spectrum can be written as Stieltjes integral

$$\mathrm{S}(\Omega) = \int_0^\infty \mathrm{H}_\infty(\Omega r)\mathrm{dC}(r) = \int_0^\infty \exp\left(-\frac{\Omega^2 r^2}{2}\right)\mathrm{dC}(r) \tag{25}$$

After the application of integration by parts one obtains

$$\mathrm{S}(\Omega) = \int_0^\infty \Omega^2 r \exp\left(-\frac{\Omega^2 r^2}{2}\right)\mathrm{C}(r)\Omega \mathrm{d}r \tag{26}$$

and substituting $\xi = \Omega r$ we get the integral formula for rotational spectrum

$$\mathrm{S}(\Omega) = \int_0^\infty \xi \cdot \mathrm{C}\left(\frac{\xi}{\Omega}\right)\exp\left(-\frac{\xi^2}{2}\right)\mathrm{d}\xi \tag{27}$$

Theorem 3. *Let $\mathcal{F} \subset \mathbb{R}^n$ be arbitrary set with rotational spectrum*

$$\mathrm{S}(\Omega) = \mathop{\mathrm{E}}_{\boldsymbol{x},\boldsymbol{y}\sim \mathrm{U}(\mathcal{F})} \mathrm{H}_\infty(\Omega\, \|\boldsymbol{x} - \boldsymbol{y}\|) \tag{28}$$

and correlation dimension D_2 (3) exists. Than it holds that

$$\lim_{\Omega\to\infty} \frac{\ln \mathrm{S}(\Omega)}{\ln \Omega} = -D_2. \tag{29}$$

As a general remark, we could consider other kernel function instead of H_∞. For any non-increasing function $\Phi : \mathbb{R}_0^+ \mapsto [0;1]$ satisfying $\Phi(0) = 1$ and $\Phi(\infty) = 0$ whose first derivative $\Phi'(\xi)$ exists for any $\xi > 0$, we consider the rotational spectrum in more general form as

$$S(\Omega) = \mathop{\mathrm{E}}_{\boldsymbol{x},\boldsymbol{y}\sim \mathrm{U}(\mathcal{F})} \Phi\left(\Omega\|\boldsymbol{x} - \boldsymbol{y}\|\right). \tag{30}$$

The Ψ function is defined as

$$\Psi(\alpha) = -\int_0^\infty \xi^\alpha \Phi'(\xi))\mathrm{d}\xi \tag{31}$$

and the existence of limit (29) is guaranteed only if both $\Psi(D_2+\epsilon)$ and $\Psi(D_2-\epsilon)$ are finite for arbitrary $\epsilon \in (0;\epsilon_0)s$. Another examples of kernel functions could be the generalized exponential kernel

$$\Phi_1(\xi) = \exp\left(-\frac{\xi^\beta}{\beta}\right) \tag{32}$$

for $\beta > 0$ or inverse polynomial kernel

$$\Phi_2(\xi) = \frac{1}{\mathrm{P}(\xi)} \tag{33}$$

where $\mathrm{P}(\xi)$ represents polynomial of order $M > D_2 + 1$.

6 Method of Estimation

The spectrum $S(\Omega)$ is studied only for H_∞ kernel. The simulation of rotational spectrum is based on generating point pairs using Monte Carlo approach. The points are independently and uniformly sampled from analysed set $\mathcal{F}$. For $N \in \mathbb{N}$ fixed and $\boldsymbol{x}_i, \boldsymbol{y}_i \sim \mathrm{U}(\mathcal{F})$, the rotational spectrum is estimated as

$$\widehat{\mathrm{S}}(\Omega) = \frac{1}{N}\sum_{j=1}^{N} \mathrm{H}_\infty\left(\Omega\|\boldsymbol{x}_j - \boldsymbol{y}_j\|\right) \tag{34}$$

including variance estimate

$$\widehat{\mathrm{var}\,\mathrm{S}}(\Omega) = \frac{1}{N-1}\sum_{j=1}^{N}\Big(\mathrm{H}_\infty\left(\Omega\|\boldsymbol{x}_j - \boldsymbol{y}_j\|\right) - \widehat{\mathrm{S}}(\Omega)\Big)^2 \tag{35}$$

To take advantage of linear dependence between logarithm of rotational spectrum and the logarithm of distance, it is reasonable to consider model

$$\ln \mathrm{S}(\Omega) = A - D_2 \cdot \ln\, \Omega + \epsilon. \tag{36}$$

The estimation of parameter D_2 is based on the maximum likelihood method using L_p regression with minimization criterion

$$CRIT = \sum_{k=1}^{N} |y_k - \mathrm{f}(x_k, \boldsymbol{a})|^p \tag{37}$$

for $p > 1$ and general model formulated as $y = \mathrm{f}(x_k, \boldsymbol{a})$. In our case, the minimization criteria satisfy

$$CRIT_1 = \sum_{k=1}^{N} \left|\ln\, \widehat{S}(\Omega_k) - A + D_2 \ln\, \Omega_k\right|^p. \tag{38}$$

7 Application to simulated data

The main feature of proposed methodology is its smoother dependence of spectrum on Ω. We test this property on parametrized Cantor set with well-known Hausdorff dimension. Moreover it is possible to compare the correlation dimension estimate from rotational spectrum approach and traditional correlation sum. At first, the linear regression with least squares minimization criterion was used to fit the model. However, the results were biased for larger number of data points as can be seen from table 1, where *sd* is the standard deviation of the estimate. To avoid the bias, we decided to use L_p regression for rotational spectrum fitting using maximum likelihood method. Numerical experiments have proven that any order $p \geq 4$ is appropriate to fit the model. Therefore we consider L_4 regression for the estimation of correlation dimension. Table 2 presents the results for different number of point pairs M. The estimates of $\widehat{D_2}$ based on L_4 regression are unbiased both for correlation sum and rotational spectrum. However, the variance of spectrum based estimates rapidly decreases with M.

Table 1: Cantor dust analysis using linear least squares fitting.

	correlation sum			rotational spectrum		
M	$\widehat{D_2}$	sd	p-value	$\widehat{D_2}$	sd	p-value
10^3	1.2254	0.0648	0.2868	1.2501	0.0323	0.3579
10^4	1.2392	0.0202	0.1310	1.2689	0.0183	0.3502
10^5	1.2513	0.0039	0.0034	1.2592	0.0030	0.1915
10^6	1.2599	0.0005	$4.44 \cdot 10^{-5}$	1.2601	0.0003	$1.54 \cdot 10^{-7}$

Table 2: Cantor dust analysis using L_4.

	correlation sum			rotational spectrum		
M	$\widehat{D_2}$	sd	p-value	$\widehat{D_2}$	sd	p-value
10^3	1.2941	0.1178	0.3922	1.2378	0.1010	0.4059
10^4	1.2937	0.0803	0.3459	1.3019	0.0470	0.1971
10^5	1.2341	0.0574	0.3143	1.2618	0.0100	0.4976
10^6	1.2654	0.0474	0.4702	1.2609	0.0076	0.4498

It is also possible to estimate the rotational spectrum for finite rotation using kernel functions H_n for $n \in \mathbb{N}$. The comparison of kernels that can be used for rotation of power spectrum is displayed in figure 1 for H_2, H_3,H_4 and H_∞. The traditional Sierpinski carpet was used for this simulation.

8 Julia set correlation dimension

Julia set [11] is two dimensional set dependent on parameter $c \in \mathbb{R}$. For each complex number z it is possible to define sequence $\{\mathrm{f}_n(c,z)\}_{n=0}^{+\infty}$ in the following way

$$\mathrm{f}_0(c,z) = z, \tag{39}$$

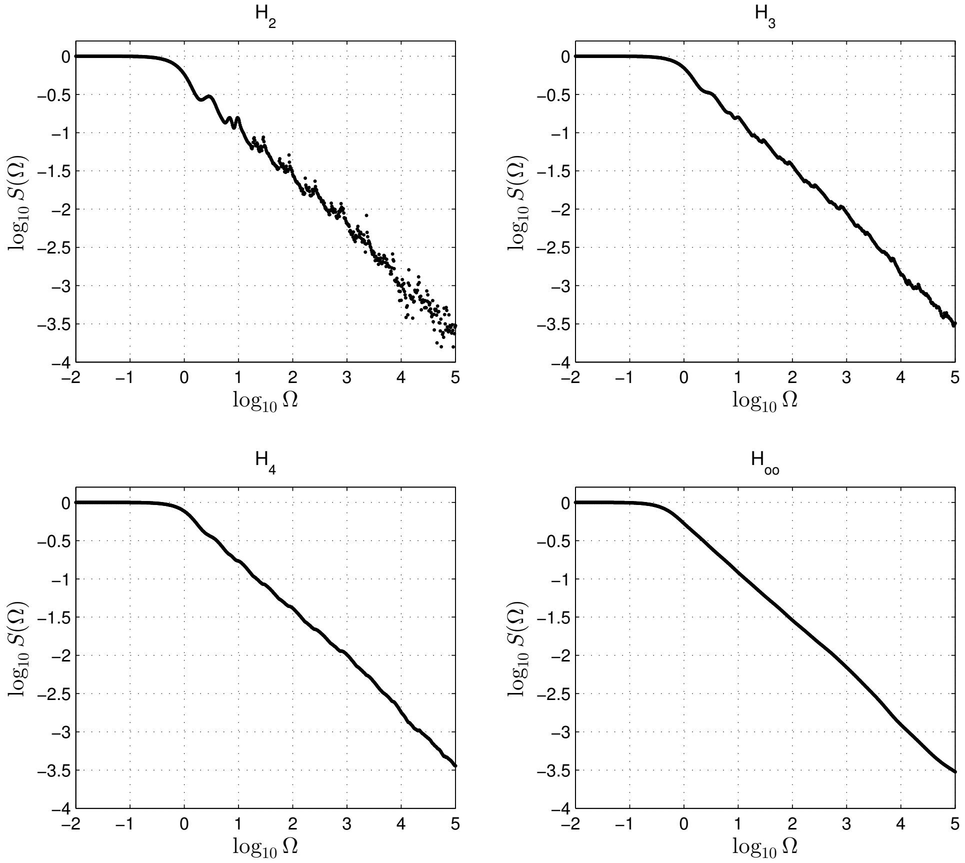


Figure 1: Rotational spectra of Sierpinski carpet.

$$\mathrm{f}_{n+1}(c, z) = \mathrm{f}_n^2(c, z) + c \tag{40}$$

for $n \in \mathbb{N}_0$. Respective Julia set $\mathcal{J}_c$ is the set of points satisfying

$$\mathcal{J}_c = \left\{\vec{x} \in \mathbf{R}^2 : x_1 = \operatorname{Re} z, x_2 = \operatorname{Im} z, z \in \partial\mathcal{H}_c\right\}, \tag{41}$$

where ∂H_c is the boundary of

$$\mathcal{H}_c = \left\{z \in \mathbf{C} : \lim_{n\to\infty} |\mathrm{f}_n(c, z)| < \infty\right\}. \tag{42}$$

There is no explicit formula for the Hausdorff dimension of Julia set $\dim_{\mathcal{H}} \mathcal{J}_c$, however, for certain parameters c is the dimension theoretically or numerically calculated with high accuracy. The correlation dimension was estimated for selected parameters c and the results are shown in Tab. 3. Parameters $\Omega_{\min}$ and $\Omega_{\max}$ are the lower and upper boundary for the regression, respectively.

Based on numerical results, the simulation proved that the estimation of correlation dimension of Julia set doesn't agree with the theoretical Hausdorff dimension for any dimension greater than one. Additionally, there is no interval in which the regression line would have slope with theoretical dimension. This experiment was performed for

Table 3: Correlation dimension of Julia set.

c	$\dim_{\mathcal{H}} \mathcal{J}_c$	$\widehat{D_2}$	sd	$\log_{10} \Omega_{\min}$	$\log_{10} \Omega_{\max}$
1/4	1.0812	0.9949	0.0282	0.5	1.5
−1	1.2683	1.0416	0.0504	0.0	1.0
i/4	1.0232	1.0031	0.0341	0.5	1.5
−5	0.4848	0.4652	0.0542	0.0	2.0
−20	0.3185	0.3364	0.0741	3.0	7.0
−3/2 + 2i/3	0.9038	0.8931	0.0755	3.0	4.0

$M = 10^5$, however further investigations for $M = 10^6$ and $M = 10^7$ led only to the decrease of standard deviation and the estimate remained with unit value. This result does not necessarily have to be in conflict with theory, because there is no proof that the Julia set satisfies the open set condition. It this condition is not met, than the theoretical correlation dimension can be lower than Hausdorff dimension. Based on these results one can hypothesize that the correlation dimension of Julia set fulfils

$$D_2 = \min\{1, \dim_{\mathcal{H}} \mathcal{J}_c\}. \tag{43}$$

9 Conclusion

Asymptotic behaviour of rotational spectrum was investigated under the assumption of D_2 existence. Rotation in infinitely-dimensional space is recommended for the correlation dimension estimation which is based on Monte Carlo simulation. As stated previously, there is a significant difference between traditional correlation integral behaviour and rotational spectrum that can be seen on the basis of log-log plot. The effect of spectrum stabilization for $n \to \infty$ is also useful for D_2 estimation from relative small uniform samples. However, the proposed method has one disadvantage in experimental choice of frequency range for regression as in the case of traditional approach. In the end, the Julia set was investigated and based on numerical results, new hypothesis for correlation dimension was presented, stating that the correlation dimension cannot exceed unit value.

References

[1] J. H. Churnside and J. J. Wilson. *Power spectrum and fractal dimension of laser backscattering from the ocean.* Journal of the Optical Society of America A **23** (nov 2006), 2829.

[2] P. Debye. *Zerstreuung von röntgenstrahlen.* Ann. Phys. **351** (1915), 809–823.

[3] L. Grafakos. *Classical Fourier Analysis: Graduate Texts in Mathematics.* Springer, (2014).

[4] P. Grassberger and I. Procaccia. *Characterization of strange attractors.* Phys. Rev. Lett. **50** (jan 1983), 346–349.

[5] P. Grassberger and I. Procaccia. *Measuring the strangeness of strange attractors.* Physica D: Nonlinear Phenomena **9** (oct 1983), 189–208.

[6] Y. Hongying and J. Duanfeng. *Mathematical Modelling: v. 2: Proceedings of First International Conference on Modelling and Simulation.* World Academic Union Ltd, (2008).

[7] A. Kalauzi, A. Vuckovic, and T. Bojić. *Topographic distribution of EEG alpha attractor correlation dimension values in wake and drowsy states in humans.* International Journal of Psychophysiology **95** (mar 2015), 278–291.

[8] S. G. Krantz. *Handbook of Complex Variables.* Birkhauser, (1999).

[9] Y. Liu, Z. Yu, M. Zeng, and S. Wang. *Dimension estimation using weighted correlation dimension method.* Discrete Dynamics in Nature and Society **2015** (2015), 1–10.

[10] K. P. Michalak. *How to estimate the correlation dimension of high-dimensional signals?* Chaos: An Interdisciplinary Journal of Nonlinear Science **24** (sep 2014), 033118.

[11] H.-O. Peitgen, P. H. Richter, and P. H. Richter. *The Beauty of Fractals: Images of Complex Dynamical Systems.* Springer-Verlag, (1987).

[12] K. Rawal, B. S. Saini, and I. Saini. *Adaptive correlation dimension method for analysing heart rate variability during the menstrual cycle.* Australas Phys Eng Sci Med **38** (aug 2015), 509–523.

[13] F. Shayegh, S. Sadri, R. Amirfattahi, and K. Ansari-Asl. *A model-based method for computation of correlation dimension, lyapunov exponents and synchronization from depth-EEG signals.* Computer Methods and Programs in Biomedicine **113** (jan 2014), 323–337.

[14] M. Talebinejad, A. D. Chan, A. Miri, and R. M. Dansereau. *Fractal analysis of surface electromyography signals: A novel power spectrum-based method.* Journal of Electromyography and Kinesiology **19** (oct 2009), 840–850.

[15] H. Wen and Z. Liu. *Separating fractal and oscillatory components in the power spectrum of neurophysiological signal.* Brain Topogr **29** (aug 2015), 13–26.

Composite $\mathfrak{gl}(2|1)$-Invariant Generalised Model: Bethe Vectors*

Jan Fuksa

5th year of PGS, email: fuksajan@fjfi.cvut.cz
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:
Čestmír Burdík, Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague
Alexey Petrovich Isaev, Bogoliubov Laboratory of Theoretical Physics
Joint Institute for Nuclear Research Dubna

Abstract. The last years were marked by considerable progress in finding the Bethe vectors for the higher rank (super)algebras. The Bethe vectors for the superalgebras $\mathfrak{gl}(2|1)$ and $\mathfrak{gl}(1|2)$ were found in [6] and used to calculate the form factors of the monodromy matrix elements [4]. This gives an opportunity to calculate the correlation functions for models with such the supersymmetry, like the t-J model known from the condensed matter physics. The next necessary step is to find the correlation functions for local operators. In models for which is known the solution of the quantum inverse scattering problem [5], such the correlation functions are reduced to the calculation of the scalar products of Bethe vectors. Unfortunately the solution is known only for a very specific class of models. However for models which do not possess this property, it is possible to obtain the correlation functions of some local operators with the help of the composite model [3]. Our article [1] investigates the composite models with $\mathfrak{gl}(2|1)$- and $\mathfrak{gl}(1|2)$-supersymmetry. The main idea is that the interval $[0, L]$, on which the original model is defined, is divided into two subintervals $[0, x]$ and $]x, L]$. Consequently the graded space $\mathcal{H}$ of the complete model is divided into two graded subspaces $\mathcal{H}^{(1)}$ and $\mathcal{H}^{(2)}$ corresponding to $[0, x]$ and $]x, L]$. Similarly the monodromy matrix $T(u)$ is divided into the partial monodromy matrices $T(u) = T^{(1)}(u)T^{(2)}(u)$. Nontrivial fact is that the Bethe vectors $\mathbb{B} \in \mathcal{H}$ of the full model can be represented as bilinear combinations of the partial Bethe vectors $\mathbb{B}^{(1)} \in \mathcal{H}^{(1)}$ and $\mathbb{B}^{(2)} \in \mathcal{H}(2)$. We explicitly describe such the representation for the Bethe vectors $\mathbb{B} \in \mathcal{H}$ as well as for their dual vectors $\mathbb{C} \in \mathcal{H}^*$ in [1]. Such the representation allows to compute the form factors of the partial monodromy matrix elements $T^{(\ell)}_{ij}(u)$, $\ell = 1, 2$, in the basis of the Bethe vectors of the full model, which is the object of our next publication [2]. Based on this the form factors and correlation functions of local operators can be investigated.

Keywords: quantum integrable systems; composite model; supersymmetry

Abstrakt. Poslední roky byly poznamenány značným pokrokem při hledání Betheho vektorů pro (super)algebry vyšších řádů. Betheho vektory pro superalgebry $\mathfrak{gl}(2|1)$ a $\mathfrak{gl}(1|2)$ byly nalezeny v [6] a použity pro výpočty form-faktorů matice monodromie [4]. To umožňuje výpočet korelačních funkcí v modelech s takovouto supersymetrií, jako t-J model známý z teorie pevných látek. Dalším nezbytným krokem je nalezení korelačních funkcí lokálních operátorů. Pro modely, v kterých je známo řešení problému kvantového inverzního rozptylu [5], se takovéto korelační

*This work has been supported by the grant SGS15/215/OHK4/3T/14.

funkce zjednodušují na skalární součiny Betheho vektorů. Bohužel modely tohoto druhu se řadí do velice úzké třídy. Pro ostatní modely je nicméně možné získat korelační funkce některých lokálních operátorů za pomoci složeného modelu [3]. V našem článku [1] studujeme složený model se supersymetrií $\mathfrak{gl}(2|1)$ a $\mathfrak{gl}(1|2)$. Hlavní myšlenka spočívá v rozdělení původního intervalu $[0, L]$, na kterém je model definován, do podintervalů $[0, x]$ a $]x, L]$. Gradovaný stavový prostor $\mathcal{H}$ celého modelu je rozdělen do dvou gradovaných podprostorů $\mathcal{H}^{(1)}$ a $\mathcal{H}^{(2)}$ odpovídajících $[0, x]$ a $]x, L]$. Matice monodromie $T(u)$ se podobně rozpadne do dvou částečných matic monodromie $T(u) = T^{(1)}(u)T^{(2)}(u)$. Betheho vektory $\mathbb{B} \in \mathcal{H}$ celého modelu jsou reprezentovány jako bilineární kombinace částečných Betheho vektorů $\mathbb{B}^{(1)} \in \mathcal{H}^{(1)}$ a $\mathbb{B}^{(2)} \in \mathcal{H}(2)$, což není vůbec triviální skutečnost. V [1] explicitně popisujeme takovouto reprezentaci Betheho vektorů $\mathbb{B} \in \mathcal{H}$ i jejich duálních vektorů $\mathbb{C} \in \mathcal{H}^*$. Tato reprezentace umožňuje výpočet form-faktorů pro prvky částečné matice monodromie $T^{(\ell)}_{ij}(u)$, $\ell = 1, 2$, v bázi Betheho vektorů celého modelu, což je předmětem naší následující publikace [2]. Díky tomu mohou být studovány form-faktory a korelační funkce lokálních operátorů.

Klíčová slova: kvantové integrabilní systémy; složený model; supersymetrie

References

[1] J. Fuksa. *Composite $\mathfrak{gl}(2|1)$-invariant generalised model: Bethe vectors.* to be submitted to an impacted journal.

[2] J. Fuksa, N. A. Slavnov. in preparation.

[3] A. G. Izergin, V. E. Korepin. *The quantum inverse scattering method approach to correlation functions.* Comm. Math. Phys., 94(1):67–92, 1984.

[4] A. Hutsalyuk, A. Liashyk, S. Z. Pakuliak, E. Ragoucy, N. A. Slavnov. *Form factors of the monodromy matrix entries in $gl(2|1)$-invariant integrable models.* Nucl. Phys. B, 2016.

[5] N. Kitanine, J. M. Maillet, V. Terras. *Correlation functions of the XXZ Heisenberg spin-$\frac{1}{2}$ chain in a magnetic field.* Nucl. Phys. B, 567(3):554–582, 2000.

[6] S. Z. Pakuliak, E. Ragoucy, N. A. Slavnov. *Bethe vectors for models based on the super-Yangian $Y(\mathfrak{gl}(m|n))$.* preprint arXiv:1604.02311[math-ph], 2016.

Podpora Intel Xeon Phi v TNL

Vít Hanousek

1. ročník PGS, email: hanouvit@fjfi.cvut.cz
Katedra matematiky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

školitel: Tomáš Oberhuber, Katedra matematiky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

Abstract. This paper presents the basic support of the Intel Xeon Phi co-processor in the Template Numerical Library. This library is developed at the Department of mathematic at the FNSPE. Firstly, two approaches how to copy an object to the co-processor are shortly presented. Then, the heat equation problem is introduced. Finally, computation times of the heat equation problem for a single core of processor and for the co-processor are measured and compared.

Keywords: Intel Xeon Phi, MIC, Offload, TNL

Abstrakt. V této práci prezentujeme úspěšné základní kroky v přidání podpory koprocesoru Intel Xeon Phi do numerické knihovny Tempate Numerical Library, která je vyvíjená na katedře matematiky FJFI. Krátce zde popisujeme dva způsoby kopírování objektu na tento koprocesor. Jejich rychlost v porovnáním s rychlostí na jednom jádře procesoru je změřena na úloze vedení tepla, která je zde také představena.

Klíčová slova: Intel Xeon Phi, MIC, Offload, TNL

1 Úvod

Intel Xeon Phi je moderní koprocesor vyvinutý firmou Intel. Jeho architektura je nazvána *Many Integrated Core* (MIC), tak se často zkráceně označuje i tento koprocesor. Hardwarově je tento koprocesor rozšiřující karta do PCIe 16x slotu a obsahuje 60 výpočetních jader s čtyřnásobným hyperthreadingem, disponuje tedy 240 výpočetními vlákny. Dále obsahuje 8 GB rychlé operační paměti RAM připojené vícekanálovým řadičem umožnujícím přenosové rychlosti až 320 GB/s [1]. Softwarově tato karta funguje jako samostatný počítač s minimalistickým operačním systémem GNU/Linux, který je k hostitelskému počítači připojený přes virtuální síťové rozhraní.

Cílem naší práce je přidat podporu tohoto koprocesoru do Template Numerical Library (TNL), numerické knihovny vyvíjené na katedře matematiky na FJFI. Tato knihovna v současné době podporuje procesory (CPU) a grafické karty firmy nVidia s technologií CUDA (GPU). Pomocí šablon jsou implementovány základní i pokročilé objekty pro různý hardware, což umožňuje pouhou změnou šablonového parametru změnit hardware, na kterém úloha bude počítána, bez dalších zásahů do kódu.

Experimentálně jsme implementovali základní datové struktury pro MIC a dále explicitní Eulerův řešič obyčejných diferenciálních rovnic ve dvou dimenzích. Součástí knihovny TNL je testovací úloha řešící rovnici vedení tepla. Na této úloze jsme ověřili, že je potřeba jen minimum změn v kódu využívajícím tuto knihovnu, a porovnali jsme výkon této jednoduché úlohy napsané pomocí knihovny TNL na CPU a MIC.

2 Testovací úloha

Testovací úloha, která je součástí knihovny TNL jakožto příklad, je rovnice vedení tepla. Zde ji budeme definovat ve dvou dimenzích. Nechť $\Omega = (0, X_{max}) \times (0, X_{max})$ je čtvercová dvourozměrná oblast a J je časový interval od času 0 do koncového času t_f. Problém vedení tepla můžeme definovat jako rovnici (1) s Dirichletovými okrajovými podmínkami (2) – (5). Počáteční podmínka je tvaru (6).

$$\begin{aligned}
\frac{\partial u}{\partial t} - \Delta u &= 0 && \text{v} \quad \Omega \times \mathcal{J}, && (1)\\
u|_{x=0} &= 0 && \text{na} \quad (0, Y_{max}) \times \mathcal{J}, && (2)\\
u|_{y=0} &= 0 && \text{na} \quad (0, X_{max}) \times \mathcal{J}, && (3)\\
u|_{x=X_{max}} &= 0 && \text{na} \quad (0, Y_{max}) \times \mathcal{J}, && (4)\\
u|_{y=Y_{max}} &= 0 && \text{na} \quad (0, X_{max}) \times \mathcal{J}, && (5)\\
u|_{t=0} &= u_{ini} && \text{v} \quad \Omega. && (6)
\end{aligned}$$

Na prostorovém intervalu Ω definujme čtvercovou síť $N \times N$ uzlů s prostorovým krokem $h = X_{max}/N$. Dále definujeme množinu vnitřních uzlů sítě ω jako

$$\omega = \{v_{i,j} \mid i = 1, \dots, N-1, \quad j = 1, \dots, N-1\}.$$

Prostorové souřadnice bodu $v_{i,j}$ jsou $[ih, jh]$. Označme τ časový krok a k časovou hladinu. Čas k-té časové hladiny je $t_k = k\tau$. Definujme množinu všech časových hladin jako

$$I = \left\{ k\tau \;\middle|\; k = 0, \dots, \left\lfloor \frac{t_f}{\tau} \right\rfloor + 1 \right\}.$$

Dále zavedeme síťovou funkci $\tilde{u} : \bar{\omega} \times I \to \mathbb{R}$ aproximující funkci u a označíme

$$u^k_{i,j} = \tilde{u}(v_{i,j}, t_k).$$

Za těchto podmínek má Eulerův řešič rovnic (1)–(5) tvar

$$u^{k+1}_{ij} = u^k_{ij} + \tau \frac{1}{h^2} \left(u^k_{i+1,j} + u^k_{i-1,j} + u^k_{i,j+1} + u^k_{i+1,j-1} - 4u^k_{ij} \right) \qquad \text{v} \quad \omega \times I, \qquad (7)$$

$$\begin{aligned}
u^k_{0,j} &= 0 && \text{na} \quad (0, \dots, N) \times I, && (8)\\
u^k_{i,0} &= 0 && \text{na} \quad (0, \dots, N) \times I, && (9)\\
u^k_{N,j} &= 0 && \text{na} \quad (0, \dots, N) \times I, && (10)\\
u^k_{i,N} &= 0 && \text{na} \quad (0, \dots, N) \times I. && (11)
\end{aligned}$$

Počáteční podmínku načítá knihovna TNL ze souboru. Pro generování souboru počátečních podmínek poskytuje knihovna generátory několika základních funkcí. Pro následující experimenty jsme zvolili počáteční podmínku tvaru $u_{ini} = \sin(x)\cos(y)$.

3 Implementace

Přestože koprocesor Intel Xeon Phi softwarově tvoří samostatný počítač, firma Intel vyvinula rozšíření jazyka C/C++, které umožňuje spouštět označené části kódu na tomto koprocesoru, i když zbytek běží na hlavním procesoru. Tyto části se nazývají *offload* [2]. Pro vytváření offloadů existují dvě syntaxe, jedna je vyvinuta ryze firmou Intel, druhá je součástí později vydaného standardu OpenMP 4.0. V rámci portování knihovny TNL na MIC jsme využili syntaxi firmy Intel. Narazili jsme však na problém s kopírováním objektů.

Tato syntaxe umožňuje kopírovat na koprocesor pouze *bytewise copyable* proměnné, což jsou proměnné základních typů, pole těchto proměnných a struktury složené z těchto proměnných. Případně dynamicky alokované pole těchto proměnných. Rozhodně nelze kopírovat struktury či objekty obsahující ukazatele nebo objekty které mají konstruktor či destruktor.

Knihovna TNL potřebuje kopírovat objekty, které určují typ sítě či síťových entit, na koprocesor. Nutno dodat, že knihovně TNL stačí mělká (bitová) kopie objektu. Pro tento problém jsme navrhli postupně dvě řešení. První řešení přetypuje ukazatel na libovolný objekt, který je potřeba zkopírovat do paměti koprocesoru, na ukazatel na pole proměnných velkých jeden byte (*uint8_t*). Jako délka pole se následně uvede velikost kopírovaného objektu. Druhá metoda vytvoří pole stejné velikosti jako je kopírovaný objekt. Na CPU tento objekt zkopíruje do paměti, kde je toto nové pole, a následně toto pole překopíruje na koprocesor. Na koprocesoru pak vytvoří ukazatel na danou třídu a naplní jej adresou pole na koprocesoru. Ač se tato novější metoda jeví jako náročnější, tak dosahuje rychlejších výsledků. Závěrem tohoto odstavce dodejme, že toto kopírování objektů funguje pouze v případě, že je objekt stejně velký na procesoru i koprocesoru. Protože se architektura procesoru a koprocesoru není shodná, nemusí to být zaručeno. Zároveň však jsou si tyto architektury tak podobné, že v našich testech jsme na problematický případ nenarazili.

4 Výkonnostní test

Po úspěšné implementaci jsme provedli výkonnostní test. Jako testovací sítě jsme zvolili sítě rozměru 64 × 64, 128 × 128, 256 × 256, 512 × 512, 1024 × 1024, 2048 × 2048, 4096 × 4096 a 8192 × 8192. Časový krok τ byl zvolen 0,00005 s a finální fyzikální čas 0,04 s. Provedli jsme tři měření, první je sériový běh na jednom jádře procesoru, druhé a třetí je paralelní kód na koprocesoru Intel Xeon Phi ve dvou dříve popsaných implementacích. Měření probíhala na následujícím hardwaru: Intel Xeon E5-2630 v3 @ 2,4 GHz, který při běhu na jednom jádře zvýší frekvenci na 3,2 GHz a na Intel Xeon Phi 5110P s 60 jádry na 1 GHz a 8 GB RAM. Výsledky měření jsou uvedeny v tabulce 1 a grafu 1.

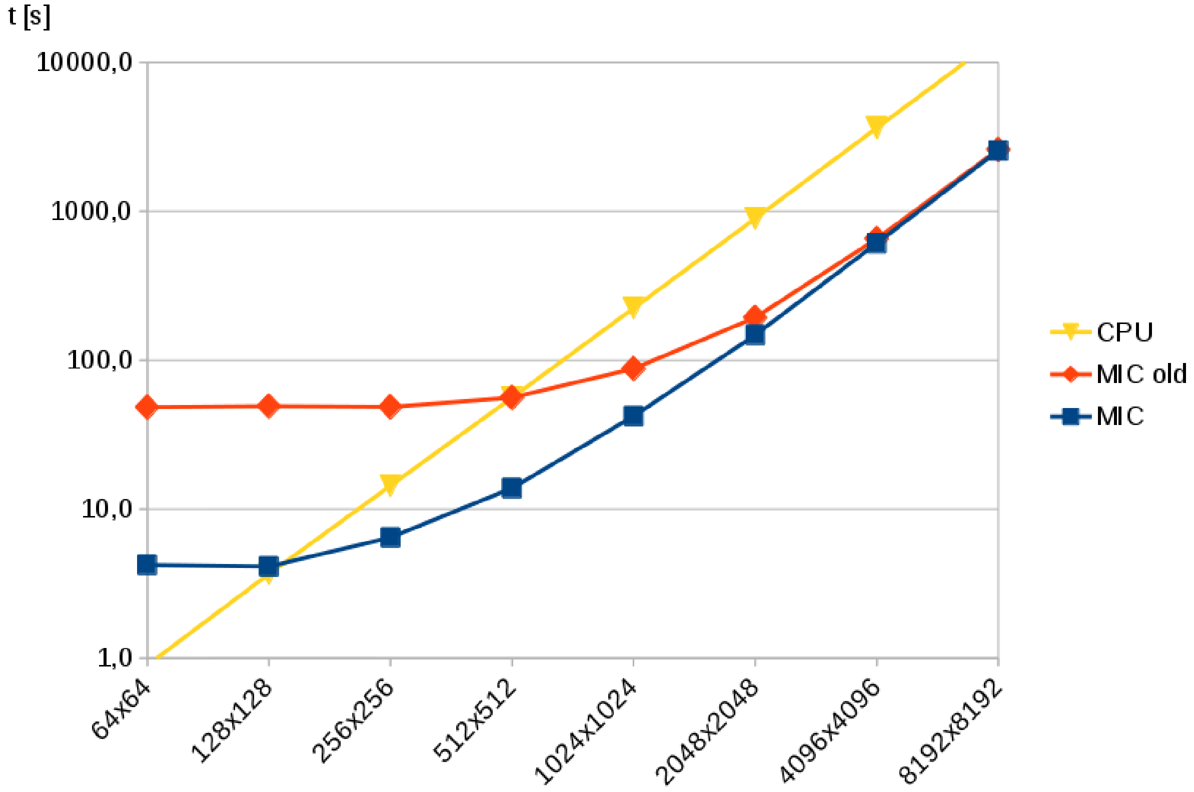
Z tabulky 2 vidíme, že dosahujeme u větších úloh zhruba šestinásobného urychlení proti jednomu jádru procesoru. Toto urychlení se však s rostoucí úlohou již nezlepšuje. To může být způsobeno velkým množstvím kopírování malých objektů, nebo samotnou vlastností dané úlohy. Zajímavým závěrem tohoto měření je zvýšení výkonu při použití nové implementace kopírování objektů na koprocesor. Při kopírování polí, která jsou dynamicky alokovaná, tedy jsou předávána do offloadu ukazatelem, provádí systém několik

		Doba běhu [s]		
		CPU	MIC	MIC old
Velikost sítě	64 × 64	0,9	4,2	48,4
	128 × 128	3,7	4,1	49,1
	256 × 256	14,4	6,5	48,6
	512 × 512	56,1	13,8	56,1
	1024 × 1024	223,9	42,2	87,8
	2048 × 2048	895,6	147,9	193,6
	4096 × 4096	3637,1	611,1	654,7
	8192 × 8192	14693,8	2567,9	2604,6

Tabulka 1: Doba běhu testovací aplikace na jednom jádře procesoru (CPU), na Intel Xeon Phi v nové implementaci (MIC) a staré implementaci (MIC old) pro různě velké sítě. Doba běhu je v sekundách.

		Urychlení	
		MIC	MIC old
Velikost sítě	64 × 64	0,21	0,02
	128 × 128	0,89	0,07
	256 × 256	2,23	0,30
	512 × 512	4,05	1,00
	1024 × 1024	5,31	2,55
	2048 × 2048	6,06	4,63
	4096 × 4096	5,95	5,56
	8192 × 8192	5,72	5,64

Tabulka 2: Urychlení běhu programu použitím koprocesoru Intel Xeon Phi vůči jednomu jádru procesoru pro obě implementace.



Obrázek 1: Doba běhu (vyjádřeno v sekundách) testovací aplikace na jednom jádře procesoru (CPU), na Intel Xeon Phi v nové implementaci (MIC) a staré implementaci (MIC old) pro různě velké sítě.

operací navíc, jejichž význam se nám zatím nepodařilo dohledat.

5 Závěr

Po úspěšném portování některých částí knihovny TNL na nový koprocesor jsme porovnali výkon této knihovny na procesoru a tomto koprocesoru. Bohužel ve výsledcích nenabízíme porovnání s GPU a s využitím vícejádrového procesoru. Verze TNL, ze které implementace pro MIC vychází, způsobovala při paralelizaci na procesoru pád překladače a výkon explicitního řešiče na GPU byl velmi špatný. Současná verze TNL obchází chybu překladače OpenMP a přináší změny v architektuře knihovny, které mají pozitivní dopad na rychlost explicitních řešičů na koprocesorech. Tyto změny však ještě nebyly zahrnuty do větve projektu podporujícího MIC. Dále nenabízíme porovnání s jednoduchou implementací těchto řešičů v jazyce C, která by poskytla cenné informace o této knihovně. Toto celkové porovnání bude předmětem dalšího výzkumu.

Literatura

[1] Intel co. *Intel® Xeon Phi^{TM} Coprocessor 5110P.* `http://ark.intel.com/products/71992/Intel-Xeon-Phi-Coprocessor-5110P-8GB-1_053-GHz-60-core`. [Accessed 30.9.2016].

[2] Kevin D. (Intel co.) *Effective Use of the Intel Compiler's Offload Features.* `https://software.intel.com/en-us/articles/effective-use-of-the-intel-compilers-offload-features`. [Accessed 30.9.2016].

Example of 2D State Transfer

Antonín Hoskovec

5th year of PGS, email: antonin.hoskovec@fjfi.cvut.cz
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Igor Jex, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. An example of how state transfer in two dimensions can be achieved is presented. The method relies on factorization of recurrence relations and the related orthogonal polynomials. These arise naturally from the form that the system Hamiltonian takes.

Keywords: perfect state transfer, quantum communication, quantum information

Abstrakt. Příspěvek ukazuje příklad přenosu stavu ve dvou rozměrech. Výpočet je založen na faktorizaci rekuretních relací a souvisejících ortogonálních polynomů. Obojí je výsledkem tvaru Hamiltoniánu systému, který popisujeme.

Klíčová slova: přenos kvantového stavu, kvantová komunikace, kvantová informace

1 Introduction

Perfect state transfer is a topic actively researched in the recent years [1]. Its purpose is faithful placement of a quantum state into a given position in the network consisting of qubits. The qubits in quantum network can be arranged into different topologies and have a wide variety of interactions.

What we show in this short article is an example of how state transfer can be achieved on a 2D square lattice of qubits, where qubits interact only with their closest neighbors. Our case is a generalization of some previously known concepts [1].

We rely on method presented in [2] and some simple ideas as we will show in the following sections. First we will summarize the method and then use it on the 2D lattice.

2 State Transfer on a Linear Qubit Chain

Hamiltonian of a linear qubit chain with nearest-neighbor interactions can be written in the form of [2]

$$H = \frac{1}{2}\sum_{i=0}^{N-1}\left[I_{i+1}\left(\sigma_i^x\sigma_{i+1}^x + \sigma_i^y\sigma_{i+1}^y\right) + B_i\left(\sigma_i^z + 1\right)\right], \tag{1}$$

where $\sigma_i^x, \sigma_i^y, \sigma_i^z$ are Pauli matrices acting on ith qubit in the chain. $N+1$ qubits in the chain are numbered $0 \dots N$. I_{i+1} denotes interaction strengths and finally B_i are the magnetic fields acting on each qubit. The Hamiltonian is defined on the Hilbert space $\mathscr{H} = (\mathbb{C}^2)^{\otimes(N+1)}$. In most cases the choice $B_i = B$ is sufficient.

Since the Hamiltonian preserves the number of excitations in the system, it is enough to focus on the single excitation subspace spanned by the basis vectors [2]

$$|i\rangle = (0,\ldots,1,\ldots,0)\,,\; i = 0,\ldots,N. \tag{2}$$

The matrix representation of the Hamiltonian in this basis is then

$$H = \begin{pmatrix} B_0 & I_1 & 0 & & & \\ I_1 & B_1 & I_2 & & & \\ 0 & I_2 & B_2 & I_3 & & \\ & & \ddots & \ddots & & \\ & & & I_N & B_N & \end{pmatrix}. \tag{3}$$

Applying this matrix on basis vectors gives

$$H\,|i\rangle = I_{i+1}\,|i+1\rangle + B_i\,|i\rangle + I_i\,|i-1\rangle, \tag{4}$$

$$I_0 = I_{N+1} = 0. \tag{5}$$

Because the matrix of H is Hermitian, there exist $N+1$ vectors such that

$$H\,|s\rangle = x_s\,|s\rangle,\; s = 0,1,\ldots,N. \tag{6}$$

The eigenvalues x_s are all real and nondegerate. The transition between the two bases can be written as

$$|s\rangle = \sum_{i=0}^{N} V_i(s)\,|i\rangle, \tag{7}$$

and the inverse is also true [2]

$$|i\rangle = \sum_{s=0}^{N} V_i(s)\,|s\rangle, \tag{8}$$

From (4) it can be seen that the expansion coefficients must satisfy

$$I_{i+1}V_{i+1}(s) + B_iV_i(s) + I_iV_{i-1}(s) = x_sV_i(s). \tag{9}$$

In order for state transfer to happen between sites i,k after time T, we require

$$\langle i|e^{-iTH}\,|k\rangle = e^{i\varphi}, \tag{10}$$

for some φ. This is the standard condition.

In [2] is described a procedure of choosing coupling strengths and magnetic field strengths for achieving exactly this. This equation can be expanded to

$$\sum_s V_i(s)\,\langle s|e^{-iTH}\sum_u V_k(u)\,|u\rangle = e^{i\varphi}, \tag{11}$$

which can be further simplified using the orthonormality of the basis eigen-vectors to

$$\sum_s V_i(s)V_k(s)e^{-iTx_s} = e^{i\varphi}. \tag{12}$$

To summarize, from the equation (4) the couplings and magnetic field strengths can be chosen so that the equation (12) holds [2].

3 State Transfer on a Square Qubit Lattice

Here we would like to show how to employ the 1D formalism to construct a 2D network that transfers state between the bottom left $(0,0)$ qubit and the top right (N,M) on a 2D square lattice of qubits of rectangular shape ($(N+1)\times(M+1)$ qubits).

In complete analogue to the 1D situation, let us index the sites with a tuple of integers $(i,j)=(0,\dots,N;0,\dots,M)$. Then the Hamiltonian can be written as

$$\begin{aligned} H &= \frac{1}{2}\sum_{i,j=0}^{N-1,M-1}\left[I_{i+1,j}\left(\sigma^x_{ij}\sigma^x_{i+1,j}+\sigma^y_{ij}\sigma^y_{i+1,j}\right)\right. \\ &\quad +J_{i,j+1}\left(\sigma^x_{i,j}\sigma^x_{i,j+1}+\sigma^y_{i,j}\sigma^y_{i,j+1}\right) \\ &\quad \left. B_{ij}\left(\sigma^z_{ij}+1\right)\right], \end{aligned} \tag{13}$$

where I_{ij} are the horizontal couplings and J_{ij} are the vertical couplings between neighboring sites.

This matrix is also Hermitian and therefore again its eigenvectors can be found, let us denote them

$$H\,|s,t\rangle = x_{st}\,|s,t\rangle, \tag{14}$$

and again perform transition between these two bases

$$|s,t\rangle = \sum_{i,j} W_{i,j}(s,t)\,|i,j\rangle, \tag{15}$$

$$|i,j\rangle = \sum_{s,t} W_{i,j}(s,t)\,|s,t\rangle. \tag{16}$$

Now in complete analogue to (9), these expansion coefficients must satisfy

$$\begin{aligned} x_{st}W_{i,j}(s,t) &= I_{i+1,j}W_{i+1,j}(s,t)+J_{i,j+1}W_{i,j+1}(s,t) \\ &\quad +B_{ij}W_{i,j}(s,t)+I_{ij}W_{i-1,j}(s,t)+J_{ij}W_{i,j-1}(s,t). \end{aligned} \tag{17}$$

Let us assume that both these couplings were chosen so that I_{ij} are independent of j and similarly J_{ij} are independent of i. Furthermore that both the couplings and magnetic field strengths B_i and C_j have been chosen from equations very similar to (9), namely

$$I_{i+1,j}V_{i+1}(s)+B_iV_i(s)+I_{ij}V_{i-1}(s) = x_sV_i(s), \tag{18}$$

$$J_{i,j+1}W_{j+1}(t)+C_jW_j(t)+J_{ij}W_{j-1}(t) = y_tW_j(t). \tag{19}$$

Where $V_i(s)$ and $W_j(t)$ were calculated so that the two equations analogous to (12) hold

$$\sum_s V_i(s)V_k(s)e^{-iTx_s} = e^{i\varphi}, \tag{20}$$

$$\sum_t W_j(t)W_l(t)e^{-iTy_t} = e^{i\eta}, \tag{21}$$

for some φ,η. Which can be done with exactly the same procedure as before.

If we choose

$$\begin{aligned} W_{i,j}(s,t) &\equiv V_i(s)W_j(t), \\ B_{i,j} &\equiv B_i + C_j, \\ x_{st} &\equiv x_s + y_t, \end{aligned}$$

then the equation (17) holds as well.

But more importantly, using the equations (20) and (21), and orthonormality of the basis vectors $|s,t\rangle$, we can show:

$$\begin{aligned} \langle i,j|e^{-iTH}\,|k,l\rangle &= \sum_{s,t} W_{ij}(s,t)\,\langle s,t|e^{-iTH}\sum_{u,v} W_{k,l}(u,v)\,|u,v\rangle \\ &= \sum_{s,t} W_{ij}(s,t)W_{kl}(s,t)e^{-iT(x_s+y_t)} \\ &= \sum_{s,t} V_i(s)W_j(t)V_k(s)W_l(t)e^{-iT(x_s+y_t)} \\ &= \underbrace{\sum_{s} V_i(s)V_k(s)e^{-iTx_s}}_{e^{i\varphi}}\underbrace{\sum_{t} W_j(t)W_l(t)e^{-iTy_t}}_{e^{i\eta}} \\ &= e^{i(\varphi+\eta)}. \end{aligned}$$

Therefore state transfer between the sites (i,j) and (j,k) takes place.

4 Conclusions

We have shown that if we choose horizontal and vertical couplings independently of each other just like we would choose them in the 1D case, state transfer will take place between the corners o the network.

This property was previously only known for one specific protocol [1]. Our case works for any known 1D protocols even if the horizontal protocol is different form the vertical one.

What remains to be seen is if the factorization is necessary for 2D state transfer or if there is some other fundamental way of transferring quantum state on a 2D quantum network.

References

[1] G.M. Nikolopoulos and I. Jex. *Quantum State Transfer and Network Engineering.* Springer, Berlin, 2014.

[2] L. Vinet and A. Zhedanov. *How to construct spin chains with perfect state transfer.* Phys. Rev. A **85** (2012), 012323.

Unions of Hidden Classes as Optimization Task*

Radek Hřebík

1st year of PGS, email: Radek.Hrebik@seznam.cz
Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:
Josef Jablonsky, Department of Econometrics
Faculty of Informatics and Statistics, University of Economics, Prague
Jaromir Kukal, Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. Cluster analysis is a traditional tool for multi-varietal data processing. Using K-means method we can split pattern set into given number of clusters. They can be used for final classification to known output classes. The paper is focused on various approaches which can be used for optimal union of hidden classes. Resulting tasks are binary programming or convex optimization ones. Presented techniques are demonstrated on crisis prediction based on clustering of macroeconomical indicators and cluster unions.

Keywords: classification, cluster analysis, binary programming, convex programming, cluster union, crisis prediction

Abstrakt. Shluková analýza představuje tradiční nástroj pro zpracování vícerozměrných dat. Použitím algoritmu k-means lze rozdělit množinu vzorů do daného počtu shluků. Shluky mohou být použity pro finální klasifikaci známých tříd. Příspěvek se zaměřuje na různé přístupy pro optimální sjednocení skrytých tříd. Řešení představuje binární programování nebo se jedná o konvexní úlohy. Představené techniky jsou prezentovány na úloze předpovídání krize, která je založena na shlukování makroekonomických ukazatelů a sjednocování shluků.

Klíčová slova: Klasifikace, shluková analýza, binární programování, konvexní programování, sjednocení shluků, předpověď krize

1 Introduction

Novel methodology is based on cluster analysis, commonly used in case of data mining and statistical data analysis. [11, 12] Akaike information criterion (AIC) and and Bayesian information criterion (BIC) [8, 2] can be used to select optimum number of clusters. The interrelation between two systems of classes is represented by contingency table [10] which is used frequently in statistics and displays the frequency of events. [7, 4] Biased estimate of adequate probabilities can be improved by using Bayesian approach [14] for

*This work has been supported by the research grants SGS14/208/OHK4/3T/14 and SGS14/209/OHK4/3T/14

bias reduction. Individual tasks can be solved by binary programming [15] or convex programming techniques [5].
The aim of research is to design a new method for optimum unions of hidden classes and apply such method to real macroeconomic data.

2 Classification Primer

Let $\mathscr{S} = \{d_1, ..., d_m\}$, $\mathscr{C}_i \subset \mathscr{S}$ for $i = 1, ..., N$, $\mathscr{H}_j \subset \mathscr{S}$ for $j = 1, ..., H$ be pattern set, disjoint system of non-empty classes, and disjoint system of hidden non-empty groups where m, N, H be numbers of patterns, classes and hidden groups. The relation between the classes and the hidden groups is declared via contingency table $\mathbb{F} \in \mathbb{N}_0^{N \times H}$ where $f_{i,j} = \text{card}\{k : d_k \in \mathscr{C}_i \bigcap \mathscr{H}_j\}$ is result of pattern counting as join frequency which can be relativized as

$$q_{i,j} = \frac{f_{i,j}}{\sum_{k=1}^{H} f_{i,k}} \tag{1}$$

where $i = 1, ..., N$, $j = 1, ..., H$.
An example of data partition for $N = 3$, $H = 5$ is illustrated in Tab.1.

Table 1: Contingency Table as $\mathbb{F}$

	$\mathscr{H}_1$	$\mathscr{H}_2$	$\mathscr{H}_3$	$\mathscr{H}_4$	$\mathscr{H}_5$
$\mathscr{C}_1$	3	98	7	11	0
$\mathscr{C}_2$	7	4	31	10	1
$\mathscr{C}_3$	1	5	27	9	0

The paper is focused on the optimum unions of hidden classes for the best classification performance using various approaches.

3 Deterministic Case

Strict classifier is defined here as mapping

$$\text{c} : \mathscr{L}_H \rightarrow \mathscr{L}_N$$

from the set $\mathscr{L}_H$ of hidden class indicies to the set $\mathscr{L}_N$ of final class indicies where $\mathscr{L}_n = \{1, ..., n\}$. This mapping can be expressed via matrix $\mathbb{X} \in \{0, 1\}^{N \times H}$ where $x_{i,j} = 1$ iff $d_k \in \mathscr{H}_j \Rightarrow d_k \in \mathscr{C}_i$ with uniqueness conditions $\sum_{i=1}^{N} x_{i,j} = 1$ for $j = 1, ..., H$. There are many quantitative measures of classification efficiency. First, the *accuracy* [13] of classification can be expressed as

$$acc = \frac{1}{m} \sum_{i=1}^{N} \sum_{j=1}^{H} f_{i,j} x_{i,j} \tag{2}$$

and will be subject of maximization.

Using concept of *class sensitivity* [1] as relative frequency of true classification, we can calculate it as

$$se_i = \sum_{j=1}^{H} q_{i,j} x_{i,j} \tag{3}$$

for $i = 1, ..., N$.
Average sensitivity can be defined as

$$ase = \frac{1}{N} \sum_{i=1}^{N} se_i. \tag{4}$$

The lower estimate of class sensitivity is declared as *critical sensitivity* [9]

$$se^* = \min\{se_i : i = 1, ..., N\} \tag{5}$$

Now, we can formulate several linear programming tasks related to optimum classifier design using *planning matrix* $\mathbb{X} \in \{0,1\}^{N \times H}$ where $x_{i,j} = 1$ indicates $\mathscr{H}_j$ as a part of $\mathscr{C}_i$.

3.1 Accuracy Maximization

Let $s^* \in [0, se^*]$ be minimum acceptable class sensitivity. We maximize

$$acc = \frac{1}{M} \sum_{i=1}^{N} \sum_{j=1}^{H} f_{i,j} x_{i,j}, \tag{6}$$

subject to

$$\sum_{i=1}^{N} x_{i,j} = 1 \text{ for } j = 1, ..., H,$$

$$\sum_{j=1}^{H} q_{i,j} x_{i,j} \geq \mathrm{s}^* \text{ for } i = 1, ..., N,$$

$$x_{i,j} \in \{0, 1\}$$

which is binary programming task. [15] Having no prior knowledge of se^* we can start with $s^* = 0$.
In the case of frequency matrix presented in Tab. 1, $s^* = 0$ and accuracy maximization, the class $\mathscr{C}_1$ is formed by $\mathscr{H}_2$ and $\mathscr{H}_4$, class $\mathscr{C}_2$ is formed by $\mathscr{H}_1$, $\mathscr{H}_3$ and $\mathscr{H}_5$, and $\mathscr{C}_3$ is empty. The value of *acc* reached 0.6916 but $ase = 0.5506$ and $se^* = 0$ which is the main disadvantage of accuracy maximization without prior knowledge of s^*.

3.2 Mean Sensitivity Maximization

Using minimum acceptable sensitivity s* again, we maximize

$$ase = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{H} q_{i,j} x_{i,j} \tag{7}$$

subject to

$$\sum_{i=1}^{N} x_{i,j} = 1 \text{ for } j = 1, ..., H,$$

$$\sum_{j=1}^{H} q_{i,j} x_{i,j} \geq s^* \text{ for } i = 1, ..., N,$$

$$x_{i,j} \in \{0, 1\}$$

as another binary programming task.
In the case of frequency matrix presented in Tab. 1, $s^* = 0$ and mean sensitivity maximization, the class $\mathscr{C}_1$ is formed by $\mathscr{H}_2$, class $\mathscr{C}_2$ is formed by $\mathscr{H}_1$ and $\mathscr{H}_5$, class $\mathscr{C}_3$ is formed by $\mathscr{H}_3$ and $\mathscr{H}_4$. The value of ase reached 0.6105 together with $acc = 0.6105$ and $se^* = 0.1509$. Therefore previous two approaches offered relative low values of se^* which is their main disadvantage.

3.3 Maximization of se^*

In the case of class equity, we can use minimax approach and maximize se^*. Adequate non-linear optimization task is

$$se^* = \min\{se_i : i = 1, ..., N\} = \max \tag{8}$$

$$\sum_{i=1}^{N} x_{i,j} = 1 \text{ for } j = 1, ..., N,$$

$$x_{i,j} \in \{0, 1\}.$$

This task can be converted to linear one

$$se^* = \max \tag{9}$$

subject to

$$\sum_{i=1}^{N} x_{i,j} = 1 \text{ for } j = 1, ..., H,$$

$$\sum_{j=1}^{H} q_{i,j} x_{i,j} - se^* \geq 0 \text{ for } i = 1, ..., N,$$

$$x_{i,j} \in \{0, 1\},$$

$$se^* \in [0, 1].$$

In the case of frequency matrix presented in Tab. 1 and maximization of se^* class $\mathscr{C}_1$ is formed by $\mathscr{H}_2$, class $\mathscr{C}_2$ is formed by $\mathscr{H}_1$, $\mathscr{H}_4$ and $\mathscr{H}_5$, class $\mathscr{C}_3$ is formed by $\mathscr{H}_3$. The value of se^* reached 0.3396 together with $acc = 0.6105$ and $ase = 0.6020$. This approach is preferred in experimental part and can be also used for s^* determination in acc and ase maximization tasks.

4 Bayesian Approach

Relative join frequencies $q_{i,j}$ can be interpreted as biased estimate of

$$p_{i,j} = \text{prob}(d \in \mathscr{H}_j | d \in \mathscr{C}_i). \tag{10}$$

Using natural Bayesian approach [14] we supposed uniform prior probabilities and therefore calculate posterior probabilities as

$$q_{i,j}^{\text{BAY}} = \frac{k_{i,j}^{\text{BAY}}}{\sum f_{i,k}^{\text{BAY}}} \tag{11}$$

where

$$f_{i,j}^{\text{BAY}} = f_{i,j} + 1. \tag{12}$$

This approach is preferred in experimental part as improvement of previous methods.

5 Mixed Strategy

In particular cases, we can randomized the union of hidden groups $\mathscr{H}_j$. In this case the planning matrix $\mathbb{X} \in [0, 1]^{N \times H}$ consists of probabilities

$$x_{i,j} = \text{prob}(d \in \mathscr{C}_i | d \in \mathscr{H}_j) \tag{13}$$

that $\mathscr{H}_j$ forms $\mathscr{C}_i$. Using exponent $\alpha \geq 1$ we can design optimization task

$$Q = \sum_{i=1}^{N} (1 - se_i)^{\alpha} = \min \tag{14}$$

subject to

$$\sum_{i=1}^{N} x_{i,j} = 1 \text{ for } j = 1, ..., N,$$

$$se_i \geq s^* \text{ for } i = 1, ..., N,$$

$$0 \leq x_{i,j} \leq 1$$

which is convex programming [5] one. This approach is included for completeness and improves ase maximization for $\alpha = 1$. It can be converted to linear programming ones for $\alpha = 1$ and $\alpha \to \infty$, and also to quadratic programming one for $\alpha = 2$.

6 Case Study: Crisis Prediction

The new method is demonstrated on real data about EU crisis prediction based on macroeconomical indicators. The indicators were evaluated from European Commission statistical data[3]. Main nine indicators were selected based on our previous research [6] and are included in Tab. 2. Annual data of 28 EU countries from 1993 to 2017 period was proceeded by logarithmic transform and resulting pattern consist of 9 logarithmic differences of corresponding indicators. Each state was represented by 24 patterns and extreme values of the differences are collected in Tab. 3. The main idea for crisis prediction was to perform cluster analysis into hidden classes according to pattern properties for each state, first. We defined two output classes presenting the economic indicators before crisis (1993-2008) and after crisis (2010-2017). Optimum union of hidden classes into these two classes was main subject of following numerical experiments.

Table 2: List of Descriptors

i	Variable	Explanation
1	TP	Total population
2	UR	Unemployment rate
3	GDP	Gross domestic product at current market prices
4	PFC	Private final consumption expenditure at current prices
5	GFC	Gross fixed capital formation at current prices
6	DD	Domestic demand including stocks at current prices
7	E	Exports of goods and services at current prices
8	I	Imports of goods and services at current prices
9	GNS	Gross national saving

For each state we analyse the results of se^* maximization for various number of hidden classes for $H = 2, ..., 20$ which corresponds to twenty three patterns per state. The number of hidden classes was selected to maximize the critical sensitivity. As an alternative, we used AIC minimization which also determined optimal number of hidden classes but without knowledge of output. The BIC was not optimal for our propose because generated too low number of classes with small critical sensitivity. The AIC generated similar results as in the case of se^* maximization as demonstrated in Tab. 4. The pair H_{opt} and se^*_{opt} represents the result of se^* maximization with knowledge of output for each state. Minimizing AIC we obtain the pair H_{AIC} and se^*_{AIC} without previous knowledge of crisis status. These to approaches slightly differs in number of hidden classes but obtain very similar values of critical sensitivity. Therefore, the crisis prediction can be based on the cluster analysis with AIC minimization without loosing prediction quality.

7 Conclusions

Novel method of optimal cluster union was designed and tested. The main advantage of this approach is in maximization of critical sensitivity or its control at least. It was shown that this method in combination with cluster analysis can be helpful in case of crisis prediction. Optimal number of clusters was found for each state. We identified three groups of states as side effect of our study.

The first group is represented by states which indicators can easily serve for crisis prediction. These are represented by the states with $se^* = 1$, namely Spain, Cyprus, Latvia, Portugal, Bulgaria, Czech Republic, Hungary and Romania. The second group is represented by states in which can be more difficult to predict the upcoming crisis. In this case

Table 3: Maximal Values of Absolute Logarithmic Differences

State	TP	UR	GDP	PFC	GCF	DD	E	I	GNS
Belgium	0.00869	0.20854	0.07289	0.02172	0.07893	0.03373	0.13982	0.16728	0.18158
Germany	0.00913	0.18805	0.06353	0.03727	0.08158	0.02683	0.14045	0.15143	0.11197
Estonia	0.02148	0.89794	0.34200	0.07250	0.31805	0.08263	0.21123	0.23667	0.23448
Ireland	0.02999	0.62861	0.20250	0.06591	0.17480	0.04896	0.11461	0.11983	0.24273
Greece	0.00828	0.34320	0.09539	0.03934	0.19416	0.02679	0.20830	0.22314	0.59250
Spain	0.01953	0.46000	0.08392	0.01942	0.18369	0.03781	0.13634	0.24476	0.06328
France	0.00750	0.20679	0.05324	0.01815	0.07020	0.01228	0.12833	0.14253	0.14175
Italy	0.00771	0.24201	0.14034	0.01829	0.06863	0.02592	0.18232	0.18160	0.09579
Cyprus	0.02633	0.40968	0.10805	0.09861	0.22839	0.07445	0.09359	0.15101	0.34260
Latvia	0.02123	0.82098	0.50905	0.17530	0.35091	0.14066	0.45519	0.25068	0.87294
Lithuania	0.02253	0.86681	0.45046	0.07411	0.37330	0.09289	0.39897	0.38724	0.29069
Luxembourg	0.02474	0.37949	0.13948	0.07262	0.13778	0.08004	0.12675	0.15509	0.17939
Malta	0.01072	0.14058	0.13630	0.05389	0.23159	0.06185	0.16801	0.17165	0.37205
Netherlands	0.00757	0.26028	0.07916	0.03598	0.07809	0.01561	0.13036	0.13084	0.10970
Austria	0.00967	0.25672	0.07098	0.03019	0.04960	0.02368	0.16962	0.15653	0.14660
Portugal	0.00707	0.20661	0.08123	0.02292	0.15234	0.03711	0.13767	0.18232	0.19307
Slovenia	0.00984	0.29335	0.23349	0.06612	0.19730	0.03993	0.14850	0.20493	0.19259
Slovakia	0.00591	0.26176	0.21145	0.06321	0.22688	0.12197	0.16796	0.18160	0.24914
Finland	0.00488	0.24784	0.16361	0.05873	0.06782	0.02425	0.21706	0.18814	0.18924
Bulgaria	0.02831	0.41522	0.32243	0.19842	0.76461	0.11754	0.32568	0.40767	2.83741
Czech Republic	0.01031	0.42050	0.15415	0.03737	0.13782	0.03815	0.19777	0.15858	0.14781
Denmark	0.00610	0.56798	0.07405	0.03341	0.14477	0.03038	0.14153	0.17680	0.18831
Croatia	0.02888	0.24039	0.34817	0.04841	0.24512	0.05836	0.17959	0.19691	0.34877
Hungary	0.00523	0.28117	0.17981	0.07504	0.11123	0.06198	0.43393	0.23402	0.28104
Poland	0.00950	0.37013	0.17792	0.03649	0.14505	0.03987	0.14818	0.14680	0.29663
Romania	0.03321	0.44183	0.26711	0.07002	0.38996	0.06283	0.20107	0.21092	0.30956
Sweden	0.02112	0.29171	0.17546	0.05454	0.08589	0.02428	0.11253	0.11692	0.18172
United Kingdom	0.00897	0.30538	0.21514	0.01603	0.11155	0.01298	0.07884	0.08140	0.26905

Table 4: Optimum Number of Hidden Classes for Crisis Prediction

State	H_{opt}	se^*_{opt}	H_{AIC}	se^*_{AIC}
Belgium	20	0.8750	20	0.8750
Germany	14	0.8750	14	0.8750
Estonia	16	0.8750	20	0.8750
Ireland	19	0.9333	16	0.8666
Greece	15	0.9333	17	0.9333
Spain	15	1.0000	20	1.0000
France	20	0.8750	18	0.8666
Italy	12	0.9333	20	0.9333
Cyprus	13	1.0000	20	1.0000
Latvia	20	1.0000	20	1.0000
Lithuania	19	0.8750	19	0.8750
Luxembourg	15	0.9333	19	0.9333
Malta	16	0.9333	20	0.9333
Netherlands	19	0.9333	20	0.9333
Austria	20	0.9333	20	0.9333
Portugal	19	1.0000	19	1.0000
Slovenia	20	0.9333	19	0.8666
Slovakia	20	0.9333	17	0.8750
Finland	19	0.8750	19	0.8750
Bulgaria	20	1.0000	20	1.0000
Czech Republic	20	1.0000	20	1.0000
Denmark	9	0.8666	19	0.8666
Croatia	17	0.9333	19	0.9333
Hungary	15	1.0000	19	1.0000
Poland	10	0.9333	19	0.9333
Romania	19	1.0000	20	1.0000
Sweden	16	0.9333	13	0.8666
United Kingdom	19	0.8750	20	0.8666

the value of se^* is lower than 0.875 as for Belgium, Germany, Estonia, France, Lithuania, Finland, Denmark, and United Kingdom. The third group is a compromise between the first and the second group.

The states of first group seems to be very sensitive to crisis origins and macroeconomical symptoms meanwhile the states forming the second group are not too sensitive. Our hypothesis is that they have their own stabilisation mechanism against economical crisis which could explain the lower predictability of macroeconomical behaviour.

Selection of the hidden class number was primary based on se^* maximization as left maximum position. The second possible way was to employ the information criteria AIC and BIC. The values of BIC was not recommended because of small number of clusters with very low values of critical sensitivity. In most cases AIC suggests the optimal number of hidden classes with the highest values of critical sensitivity. Using AIC to select the optimal number of hidden classes, the error of critical sensitivity was up to 0.0667.

References

[1] L. Chang and W. Slikker. *Neurotoxicology: Approaches and Methods.* Elsevier Science, (1995).

[2] J. DiStefano. *Dynamic Systems Biology Modeling and Simulation.* Elsevier Science, (2015).

[3] D. G. for Economic and F. A. (ECFIN). Statistical annex to european economy. autumn 2015. Technical report, European Commission, (2015).

[4] R. Harshbarger and J. Reynolds. *Mathematical Applications for the Management, Life, and Social Sciences.* Cengage Learning, (2015).

[5] J. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms I: Fundamentals.* Springer Berlin Heidelberg, (1996).

[6] R. Hrebik and J. Kukal. *Multivarietal data whitening of main trends in economic development.* In 'Mathematical Methods in Economics', 279–284, Plzeň, (2015). University of West Bohemia.

[7] M. Kateri. *Contingency Table Analysis: Methods and Implementation Using R.* Statistics for Industry and Technology. Springer New York, (2014).

[8] S. Konishi and G. Kitagawa. *Information Criteria and Statistical Modeling.* Springer Series in Statistics. Springer, (2008).

[9] K. Novakova. *Application of Transforms in Object Recognition (in Czech).* PhD thesis, FNSPE, CTU in Prague, (2008).

[10] L. O'Brien. *The statistical analysis of contingency table designs.* Concepts and techniques in modern geography. Environmental Publications, University of East Anglia, (1989).

[11] E. Santi, D. Aloise, and S. J. Blanchard. *A model for clustering data from heterogeneous dissimilarities.* European Journal of Operational Research **253** (2016), 659–672.

[12] G. Shi. *Data Mining and Knowledge Discovery for Geoscientists.* Elsevier Science, (2013).

[13] J. Taylor. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements.* A series of books in physics. University Science Books, (1997).

[14] J. Wang, Y. Ma, L. Ouyang, and Y. Tu. *A new bayesian approach to multi-response surface optimization integrating loss function with posterior probability.* European Journal of Operational Research **249** (2016), 231–237.

[15] G. Weber. *A solution technique for binary integer programming using matchings on graphs.* Cornell University, May, (1978).

Dynamics of Dislocations Described as Evolving Curves Interacting with Obstacles*

Miroslav Kolář

4th year of PGS, email: kolarmir@fjfi.cvut.cz
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:
Michal Beneš, Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague
Jan Kratochvíl, Department of Physics
Faculty of Civil Engineering, CTU in Prague

Abstract. In this paper we describe the model of glide dislocation interaction with obstacles based on the planar curve dynamics. The dislocations are represented as smooth curves evolving in a slip plane according to the mean curvature motion law, and are mathematically described by the parametric approach. We enhance the parametric model by employing so called tangential redistribution of curve points to increase the stability during numerical computation. We developed additional algorithms for topological changes (i.e. merging and splitting of dislocation curves) enabling a detailed modelling of dislocation interaction with obstacles. The evolving dislocations are approximated as a moving piece-wise linear curves. The obstacles are represented as idealized circular areas of a repulsive stress. Our model is numerically solved by means of semi-implicit flowing finite volume method. We present results of qualitative and quantitative computational studies where we demonstrate the topological changes and discuss the effect of tangential redistribution of curve points on computational results.

Keywords: dislocation, precipitate, parametric approach, tangential redistribution, topological changes

Abstrakt. V tomto článku se zabýváme popisem interakcí dislokací s překážkami s využitím dynamiky planárních křivek. Dislokace jsou parametricky popsány jako hladké křivky pohybující se v příslušné skluzové rovině, přičemž pohyb dislokací je řízen jejich střední křivostí. Parametrický model je z důvodu stability modifikován o tangenciální redistribuci diskretizačních bodů. V článku jsou prezentovány algoritmy pro topologické změny (tzn. spojování a rozpojování křivek), které umožňují vytvořit detailní model interakce dislokace s překážkou. Při numerických výpočtech se jednotlivé dislokace aproximují jako po částech lineární křivky a překážky jsou reprezentovány idealizovanou představou kruhových oblastí odpuzujících dislokační křivku. Pro vlastní výpočty je pak použito semiimplicitní schéma založené na metodě plovoucích konečných objemů. V článku jsou představeny kvalitativní a kvantitativní výsledky provedených výpočetních studií, které demonstrují vliv algoritmů topologických změn a tangenciální redistribuce.

Klíčová slova: dislokace, precipitát, parametrický přístup, tangenciální redistribuce, topologické změny

*This work was supported by the project P108/12/1463 of the Grant Agency of the Czech Republic.

Full paper: P. Pauš, M. Beneš, K. Kolář and J. Kratochvíl, *Dynamics of dislocations described as evolving curves interacting with obstacles*, Modelling Simul. Mater. Sci. Eng. 24 (2016) 035003 (34pp).

Higher Order Proton Lifetime Estimates in Grand Unified Theories

Helena Kolešová*

6th year of PGS, email: helena.kolesova@fjfi.cvut.cz
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Michal Malinský, Institute of Particle and Nuclear Physics
Faculty of Mathematics and Physics, CU

Abstract. Since the main experimentally testable prediction of grand unified theories is the instability of the proton, precise determination of the proton lifetime for each particular model is desirable. Unfortunately, the corresponding computation usually involves theoretical uncertainties coming e.g. from ignorance of the mass spectrum or from the Planck-suppressed higher-dimensional operators [1]. We show that in general this may result in errors in the proton lifetime estimates stretching up to several orders of magnitude. On the other hand, we present a model based on SO(10) gauge group [2] which is subsequently broken by a scalar adjoint representation, where the leading Planck-suppressed operator is absent, hence the two-loop precision may be achieved.

The effort to evaluate all possible errors in proton lifetime estimates is continued in [3] where we focus on the uncertainties coming from the ignorance of the flavour structure of the given theory. Possible suppression of the decay widths is analogous to Cabbibo suppression in Standard Model and was evaluated analytically e.g. in the review [4]. In contrast to this work, we perform also a numerical analysis for different models and different assumptions on the Yukawa sector revealing that the flavour structure of the theory may influence the proton decay rates even more than expected by the analytical approach. Moreover, the effect of Planck-suppressed effective operators on the flavour structure is studied.

Keywords: Grand Unified Theories, proton decay, Planck-suppressed operators

Abstrakt. Hlavní a někdy jedinou experimentálně ověřitelnou předovědí teorií velké unifikace je možnost rozpadu protonu, přesný výpočet doby života protonu je proto žádoucím výstupem pro každý konkrétní model. Tento výpočet však naráží na řadu teoretických nejistot například kvůli neznalosti hmot těžkých částic anebo kvůli operátorům vyšší dimenze potlačeným Planckovou energií [1]. Ukazujeme, že v obecném případě mohou výsledné chyby v určení doby života protonu dosahovat několika řádů. Naproti tomu představujeme model založený na kalibrační grupě SO(10) [2], později narušené adjungovanou reprezentací, kde ve vedoucím řádu Planckovsky potlačené operátory nejsou přítomny a lze tedy dosáhnout při výpočtu doby života protonu dvousmyčkové přesnosti.

Ve snaze vyčíslit možné chyby v odhadech doby života protonu pokračujeme člákem [3], kde se zaměřujeme na nejistoty plynoucí z neznalosti flavourové struktury daného modelu. Možné potlačení rozpadové šířky je analogické s tzv. Cabbibovským potlačením ve standardním modelu a analyticky bylo vyčísleno např. v článku [4]. Narozdíl od této práce provádíme i numerický

*The work of H.K. has been supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS13/217/OHK4/3T/15

výpočet pro různé modely a za různých předpokladů o Yukawovském sektoru a zjišťujeme, že flavourová struktura dané teorie může ovlivnit rozpadové šířky ještě více, než naznačoval analytický výpočet. Dále studujeme vliv Planckovsky potlačených operátorů na flavourovou strukturu modelu.

Klíčová slova: teorie velké unifikace, rozpad protonu, oprátory potlačené Planckovou energií

Full paper: H. Kolešová. *Higher order proton lifetime estimates in grand unified theories.* In 'Proceedings, 50th Rencontres de Moriond Electroweak interactions and unified theories', 511–514, (2015).

References

[1] X. Calmet, S. D. Hsu, and D. Reeb. *Grand unification and enhanced quantum gravitational effects.* Phys.Rev.Lett. **101** (2008), 171802.

[2] H. Kolešová and M. Malinský. *Proton lifetime in the minimal SO(10) GUT and its implications for the LHC.* Phys.Rev. **D90** (2014), 115001.

[3] H. Kolešová and M. Malinský. *Flavour structure of grand unified theories and related errors in proton lifetime estimates.* To appear in Physics Letters B (2016).

[4] P. Nath and P. Pérez. *Proton stability in grand unified theories, in strings and in branes.* Physics Reports **441** (2007), 191–317.

Invariants of Vector Fields from Gaussian–Hermite Moments*

Jitka Kostková

2nd year of PGS, email: kostkjit@fjfi.cvut.cz
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jan Flusser, Department of Image Processing
Institute of Information Theory and Automation, CAS

Abstract. Invariants of vector fields with respect to total rotation constructed from Gaussian-Hermite moments are introduced. Their numerical stability is shown to be better than that of the invariants published so far. The application in template matching of vector field is demonstrated.

Keywords: Rotation invariants, Gaussian–Hermite moments, Template matching.

Abstrakt. V tomto článku uvedeme invarianty vektorového pole vůči totální rotaci zkonstruované pomocí Gaussových–Hermiteových momentů. Ukážeme, že jejich numerická stabilita je vyšší než u doposud publikovaných invariantů. Aplikace je demonstrována na vyhledávání vzorů ve vektorových polích.

Klíčová slova: Rotační invarianty, Gaussovy–Hermiteovy momenty, hledání vzorů.

1 Introduction

In the last decade, an increasing attention has been paid to *vector field* (VF) images and to the tools for their analysis. The images of vector fields arise in mechanical engineering, fluid dynamics, computer vision, meteorology, etc. They visualize particle velocity, wind velocity, optical/motion flow, image gradient, and other phenomena. They may show e.g. flowing water in a pipe, an air flow around an aircraft wing or around a coachwork, or a wind velocity map. They may be obtained as a result of computer processing of standard digital images or video, numerical solution of Navier–Stokes equation, or from real physical measurements (see Fig. 1).

A 2D vector field $\mathbf{f}(\mathbf{x})$ can be mathematically described as a pair of scalar fields (images) $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}))$. At each point $\mathbf{x} = (x, y)$, the value of $\mathbf{f}(\mathbf{x})$ shows the orientation and the magnitude of certain vector. A common task in vector field analysis is a detection of various patterns such as sinks, vortexes, and saddle points. For engineers and designers, it is very important to identify these singularities in the flow, because they increase the friction, decrease the speed of the medium and consequently increase the power and cost which is necessary to transport the medium through the pipe or the object through the air or water. The detection of singularities is typically accomplished by template matching. Sample templates of these patterns are stored in the template

[*] This work has been supported by grants No. GA15–16928S and SGS15/214/OHK4/3T/14.

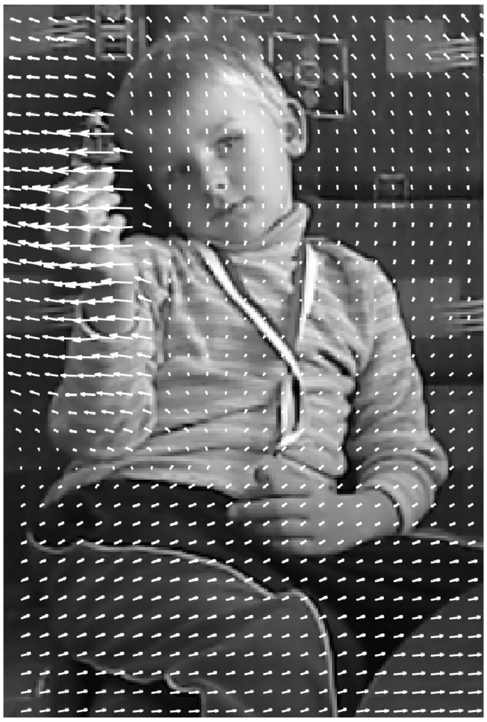


Figure 1: Optical flow field in a video. The field is depicted by arrows, which show the direction and velocity of the movement between adjacent frames.

database and searched in the given field. The search algorithm must be primarily *rotation invariant*, because the particular orientation of the template is unknown, irrelevant.

Many template-matching techniques have been developed for scalar images. The key point to avoid a brute-force search is to find a rotation-invariant template descriptors. The matching is then performed by a search of all possible template locations (which may be sped-up by a pyramidal representation of the image) and the matching position is determined as that one which minimizes certain "distance" (usually derived from ℓ_2-norm) in the space of descriptors. The first method of this kind was proposed by Goshtasby [8], who used rotation moment invariants as the descriptors.

The invariant descriptors originally designed for scalar images cannot be directly applied to vector fields because the behavior of a vector field under rotation is different. Rotation of scalar image f by angle α is described as $f'(\mathbf{x}) = f(\mathbf{R}_{-\alpha}\mathbf{x})$, where

$$\mathbf{R}_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

is a rotation matrix. This *inner rotation* affects the spatial coordinates only.

However, when rotating a vector field, the vectors rotate inversely to the in-plane rotation such that their relative orientation to the image content stays constant. The underlying model, which is called *total rotation*, is $\mathbf{f'(x)} = \mathbf{R}_\alpha \mathbf{f}(\mathbf{R}_{-\alpha}\mathbf{x})$. Let us illustrate the total rotation in Fig. 2 for $\alpha = 22.5°$. Each arrow is rotated around the image center to the new position and its direction is also rotated by the same angle. If a vector field is scaled by factor s, the underlying transformation is called *total scaling* $\mathbf{f'(x)} = s\mathbf{f}(\mathbf{x}/s)$.

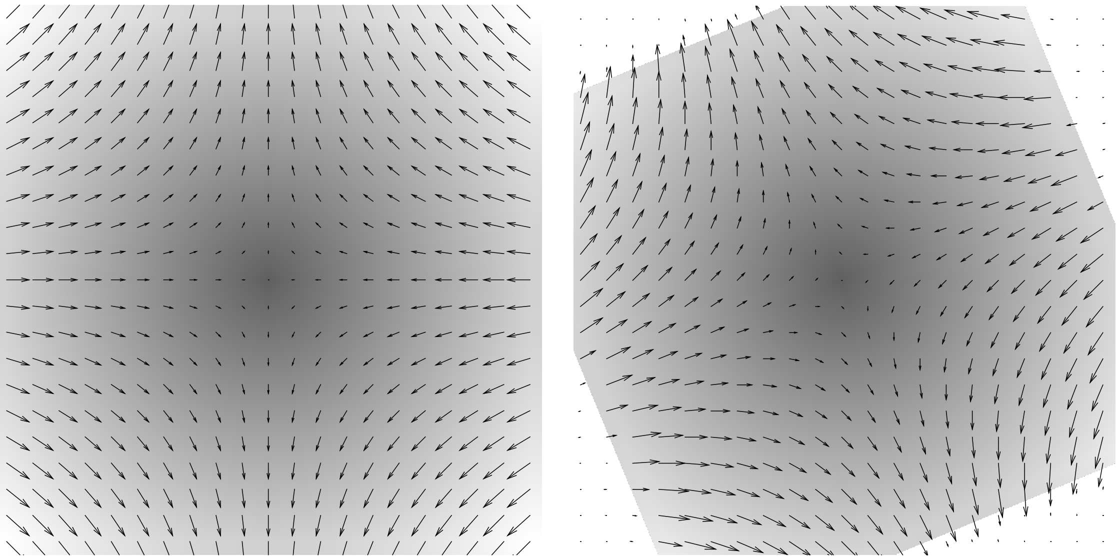


Figure 2: The original vector field (a) and its total rotation (b).

In order to implement rotation-invariant template matching, we need at first to find descriptors which are invariant to total rotation of a vector field. This problem was addressed for the first time by Schlemmer et al. [12] who adapted original scalar moment invariants proposed by Mostafa and Psaltis [1] and Flusser [5, 6], and designed invariants composed of complex moments of the field. Schlemmer et al. used these invariants to detect specific patterns in a turbulent swirling jet flow. The Schlemmer's method has found several applications. Liu and Ribeiro [10] used it, along with a local approximation of the vector field by a polynomial, to detect singularities on meteorological satellite images where the respective field was a wind velocity map. Basically the same kind of rotation invariants was used by Liu and Yap [9] for indexing and recognition of fingerprint images. Bujack et al. [3, 2] derived essentially the same system of invariants by means of the field normalization approach. These authors demonstrated the use of the invariants in template matching, where the template vortexes were searched in the image showing the Karman vortex street simulation.

In all of the above-mentioned papers, the invariants are based on standard geometric moments. It is well known from many studies of scalar images, that the geometric (and consequently the complex) moments have rather poor numerical properties, in particular they cannot be calculated in a stable way up to high orders [7]. This is caused by non-orthogonality of their basis functions $x^p y^q$. In scalar image analysis, this finding led to the design of invariants from orthogonal moments and from other orthogonal projections. However, nothing like that has been published for vector fields so far. In this paper, we introduce vector field invariants w.r.t. total rotation from orthogonal Gaussian-Hermite moments. We demonstrate they have better numerical properties than the Schlemmer's invariants and they can be advantageously used in the vector field template matching tasks.

2 Gaussian-Hermite polynomials and moments

Hermite polynomial of the n-th degree is defined as

$$H_n(x) = (-1)^n e^{x^2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2} \ . \tag{1}$$

Their three-term recurrence relation, which is used for their fast and stable evaluation, is

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= 2x, \\ H_n(x) &= 2xH_{n-1}(x) - 2(n-1)H_{n-2}(x) \ . \end{aligned} \tag{2}$$

Hermite polynomials are orthogonal on $(-\infty\,,\,\infty)$ with the weight $w(x) = e^{-x^2}$. If they are not modulated, they have a high dynamic range and poor localization, which makes them difficult to use directly for image description. Therefore, we modulate Hermite polynomials with a Gaussian function and scale them to yield *Gaussian-Hermite (GH) polynomials*

$$H_n(x,\sigma) = H_n(x/\sigma)e^{-\frac{x^2}{2\sigma^2}} \ . \tag{3}$$

In most cases, we work with *orthonormal* GH polynomials $\hat{H}_n$

$$\hat{H}_n(x,\sigma) = \frac{1}{\sqrt{n!2^n\sigma\sqrt{\pi}}} H_n(x,\sigma) \ . \tag{4}$$

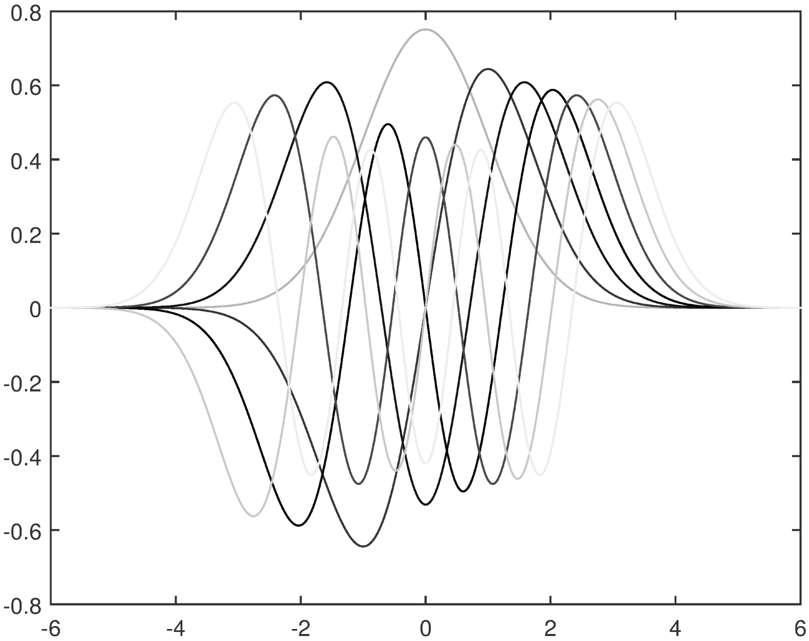


Figure 3: The graphs of the Gaussian-Hermite polynomials up to degree 6 with $\sigma = 1$.

As can be seen in Figure 3, the GH polynomials have the range of values inside $(-1, 1)$. Although they are formally defined on $(-\infty, \infty)$, they are effectively localized into a small neighborhood of the origin controlled by σ.

2D Gaussian-Hermite moments of function $f(x, y)$ are defined as

$$\hat{\eta}_{pq} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \hat{H}_p(x,\sigma)\hat{H}_q(y,\sigma)f(x,y)\mathrm{d}x\mathrm{d}y\,. \tag{5}$$

The GH moments were introduced into the image analysis area by Shen [13, 14] and were proved to be very robust w.r.t. additive noise comparing to other common moments, which is a remarkable advantage . They were employed in several successful applications, such as in detection of moving objects in a video [15], in licence plate recognition [11], in image registration as landmark descriptors [19], in fingerprint recognition [9], in face recognition [4],and as directional feature extractors [17].

3 Gaussian-Hermite rotation invariants of scalar images

Yang et al. [16, 18] proved that the 2D Hermite polynomials $H_{pq}(x, y) \equiv H_p(x)H_q(y)$ change under an in-plane rotation by angle α in the same way as do the monomials x^py^q. This property propagates from polynomials to moments, which leads to the assertion of the *Yang's theorem*: If there exist a rotation invariant of geometric moments $I(m_{p_1q_1}$, $m_{p_2q_2}$, …, $m_{p_dq_d})$, the same function of the corresponding Hermite moments $I(\eta_{p_1q_1}$, $\eta_{p_2q_2}$, …, $\eta_{p_dq_d})$ is also a rotation invariant.

The Gaussian weighting and scaling do not violate this theorem provided that the scale parameter σ is the same for x and y and that the weighting coefficient is modified

$$\hat{\eta}_{pq} = \frac{1}{\sigma\sqrt{\pi(p+q)!2^{p+q}}}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} e^{-\frac{(x^2+y^2)}{2\sigma^2}}H_p(x/\sigma)H_q(y/\sigma)f(x,y)\mathrm{d}x\mathrm{d}y\ . \tag{6}$$

Under these assumptions the Yang's theorem holds well and the functional $I(\hat{\eta}_{p_1q_1}$, $\hat{\eta}_{p_2q_2}$, …, $\hat{\eta}_{p_dq_d})$ is a rotation invariant of the Gaussian-Hermite moments.

4 Gaussian-Hermite rotation invariants of vector fields

In this section, we use the Yang's theorem for constructing rotation invariants of vector fields. We can treat the VF as a field of complex numbers $\mathbf{f}(x, y) = f_1(x, y) + if_2(x, y)$ which allows us to use standard definition of moments. It holds, for any moment M_{pq},

$$M_{pq}^{(\mathbf{f})} = M_{pq}^{(f_1)} + iM_{pq}^{(f_2)}\,.$$

Any moment M_{pq} is changed under outer rotation (i.e. the rotation of the vector values) as $M'_{pq} = e^{-i\alpha}M_{pq}$. Hence, the Yang's theorem is valid also for total rotation of vector fields. Its practical applicability depends on our ability to find a set (preferably complete) of rotation invariants of vector fields composed of geometric moments.

It is, however, well known in the theory of geometric moments of scalar images that the moment functions[1]

$$c_{pq} = \sum_{k=0}^{p} \sum_{j=0}^{q} \binom{p}{k} \binom{q}{j} (-1)^{q-j} i^{p+q-k-j} m_{k+j,p+q-k-j} \tag{7}$$

change under the inner rotation by angle α as $c'_{pq} = e^{-i(p-q)\alpha} c_{pq}$. Under a total rotation $c_{pq}^{(\mathbf{f})}$ is changed as $c_{pq}^{(\mathbf{f}')} = e^{-i\alpha} e^{-i(p-q)\alpha} \cdot c_{pq}^{(\mathbf{f})} = e^{-i(p-q+1)\alpha} \cdot c_{pq}^{(\mathbf{f})}$. Thanks to the Yang's theorem, replacing c_{pq}'s by corresponding functions of GH moments

$$d_{pq} = \sum_{k=0}^{p} \sum_{j=0}^{q} \binom{p}{k} \binom{q}{j} (-1)^{q-j} i^{p+q-k-j} \hat{\eta}_{k+j,p+q-k-j}. \tag{8}$$

must preserve the behavior under a total rotation. Hence, $d_{pq}^{(\mathbf{f}')} = e^{-i(p-q+1)\alpha} \cdot d_{pq}^{(\mathbf{f})}$.

By a multiplication of proper powers we can cancel the rotation parameter and obtain an invariant. It is desirable to work with an independent subset (basis) of rotation invariants. The simplest possible basis can be obtained as

$$\Phi(p,q) \equiv d_{pq} d_{q_0 p_0}^{p-q+1} \tag{9}$$

where $p_0 - q_0 = 2$. To get a complete system, we set by definition $\Phi(q_0, p_0) \equiv |d_{q_0 p_0}|$.

5 Experiments

The goal of the experimental section is to compare GH invariants of VFs (9) to the Schlemmer's invariants [12] composed of geometric/complex moments. In the first experiment, we demonstrate high numerical stability and low precision loss even for high-order GH invariants. The second experiment illustrates their application in template matching.

5.1 Numerical precision

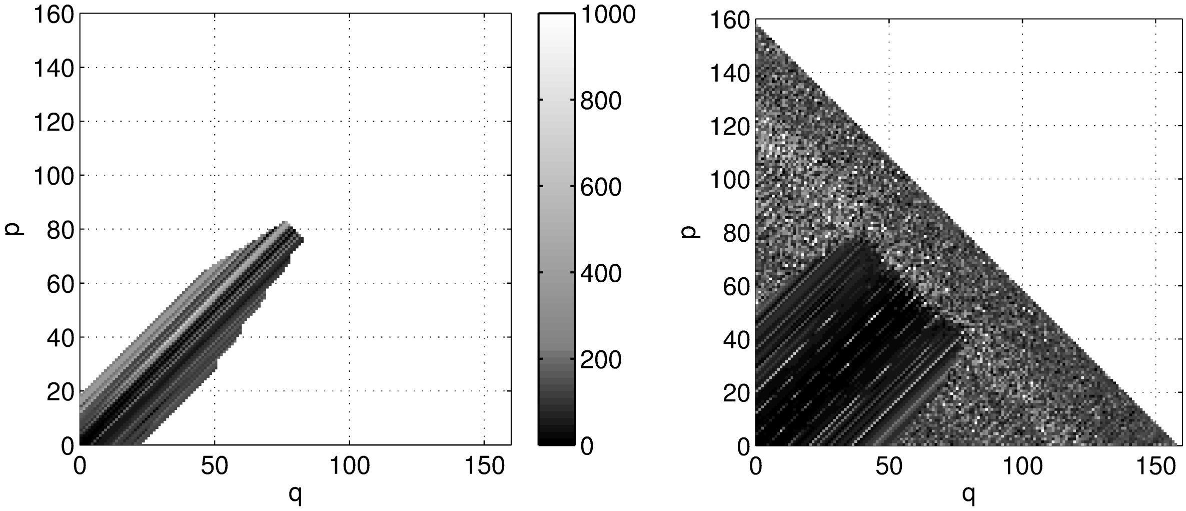
In this experiment, we evaluated numerical properties of both GH and Schlemmer's invariants up to the order $p + q = 160$. It can be expected that high-order Schlemmer's invariants lose precision because they calculate with very high and very low numbers. Since the GH moments can be calculated by recurrent relation (2), the overflow and underflow effects should be less significant or even not present at all.

The evaluation is done by measurement of a relative error of each invariant. We took a sample VF, rotated it by $\pi/4$ (total rotation), and calculated the relative error as

$$\varepsilon = 100 \frac{|\Phi(p,q) - \Phi'(p,q)|}{\Phi(p,q)} \ [\%]$$

where $\Phi'(p,q)$ stands for the invariant of the rotated field. Theoretically it should hold $\varepsilon = 0$ for any p and q; the non-zero values are caused solely by the field resampling and by

[1] Function c_{pq} is in moment theory called the *complex moment.*



(a) (b)

Figure 4: Relative errors of the Schlemmer's invariants (a). White area corresponds to NaN values of the invariants. And relative errors of the Gaussian-Hermite invariants (b).

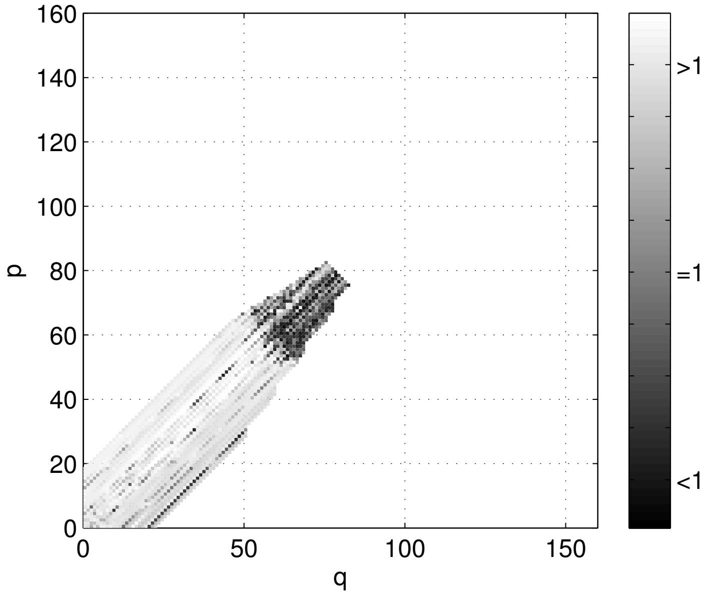


Figure 5: The ratio of the relative errors (10).

numerical errors. We used rotation by $\pi/4$, since the errors are greater than for any other angle and allow to observe the differences between the both types of invariants clearly.

The relative errors of the Schlemmer's invariants are visualized in Fig. 4(a). It is worth noting that the invariants are well defined only in a strip along the diagonal $p = q$. Outside the gray area, the Matlab code yielded NaN values when calculated the invariants. This illustrates the limited possibility of working with the Schlemmer's invariants if $p - q > 20$ and $p, q > 80$ (the particular numbers depend on the given vector field).

The relative errors of the GH invariants are visualized in the same way in Fig. 4(b). The main difference is that all investigated invariants are valid (there have been no NaN's). To compare the relative errors in the valid region, we calculated the ratio

$$r = \frac{\varepsilon(Schlemmer)}{\varepsilon(GH)} \tag{10}$$

and visualized it in Fig. 5. To keep the same range on both sides of colorbar, the values of $r > 1$ were displayed as $2 - 1/r$). We can observe that in vast majority of indexes (p, q) (precisely in 85 %) the value of r is greater than 1, i.e. the error of the Schlemmer's invariants is higher than that of the GH invariants. The mean value of $r(p, q)$ is 7.3 and the median equals 4.3, which clearly illustrates better stability of the GH invariants.

5.2 Template matching

In this experiment we demonstrate the use of the GH invariants in template matching.As a vector field, we used the gradient of the picture of hair (see Fig. 6).

We selected randomly 9 circular templates of the gradient field, rotated them by 5 degrees and matched them against the original field. The matching was carried out by searching for the minimum ℓ_2-distance in the space of the GH invariants of orders $p + q \leq 4$ between the template and all field patches of the same size. Eight templates were found in their exact location, one was matched with a localization error 1 pixel (see Fig. 6). We repeated this experiment with template rotations 23, 41, 59, and 77 degrees, respectively. The results were always exactly the same as depicted in Fig. 6.

6 Conclusion

In this paper we extended the theory of Gaussian-Hermite moment in- variants with respect to total rotation of vector fields. We demonstrated the high numerical stability and low precision loss even for high-order GH invariants unlike Schlemmer's invariants and the application of the GH invariants in template matching. We can conclude that the performance of the GH invariants in template matching is very good, regardless of the actual template content and of the template rotation.

References

[1] Y. S. Abu-Mostafa and D. Psaltis. *Recognitive aspects of moment invariants.* IEEE Transactions on Pattern Analysis and Machine Intelligence **6** (1984), 698–706.

[2] R. Bujack, M. Hlawitschka, G. Scheuermann, and E. Hitzer. *Customized TRS invariants for 2D vector fields via moment normalization.* Pattern Recognition Letters **46** (2014), 46–59.

[3] R. Bujack, I. Hotz, G. Scheuermann, and E. Hitzer. *Moment invariants for 2D flow fields using normalization.* In 'Pacific Visualization Symposium, PacificVis'14', 41–48. IEEE, (2014).

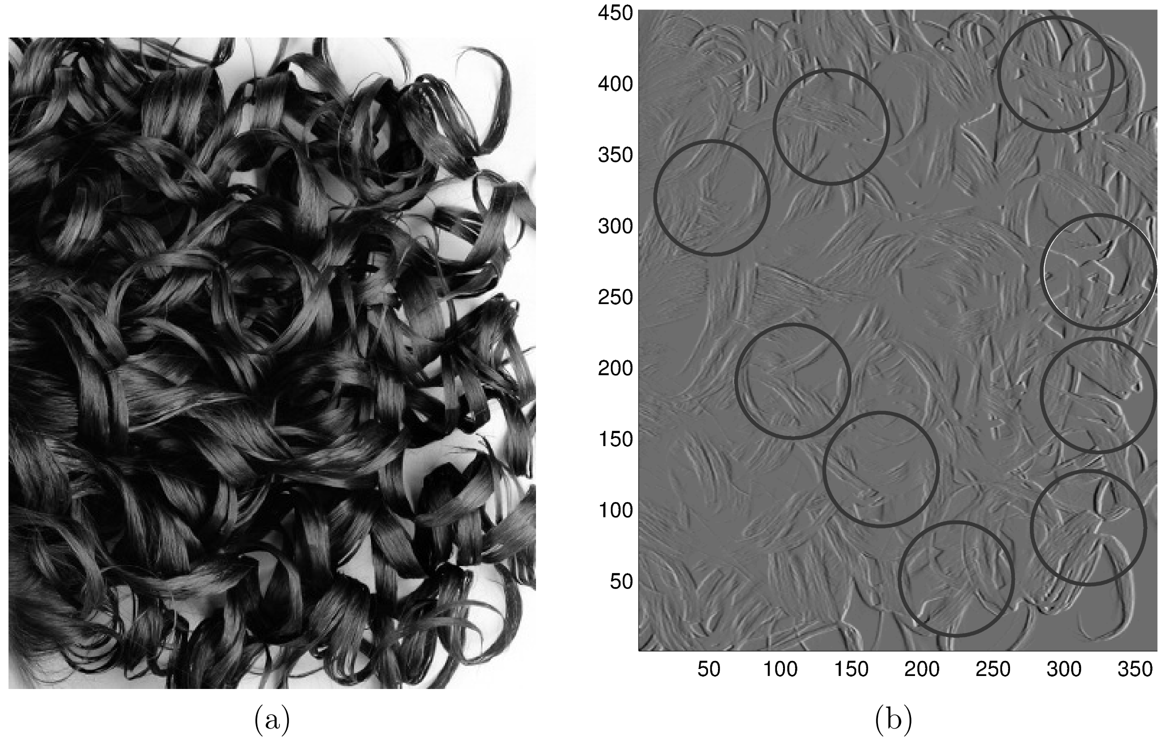


Figure 6: The original picture of hair (a) and magnitudes of a gradient field (b). Ground-truth template positions (white) and the positions localized by the GH invariants (black) differ from one another only for one template by 1 pixel.

[4] S. Farokhi, U. U. Sheikh, J. Flusser, and B. Yang. *Near infrared face recognition using Zernike moments and Hermite kernels.* Information Sciences **316** (2015), 234–245.

[5] J. Flusser. *On the independence of rotation moment invariants.* Pattern Recognition **33** (2000), 1405–1410.

[6] J. Flusser. *On the inverse problem of rotation moment invariants.* Pattern Recognition **35** (2002), 3015–3017.

[7] J. Flusser, T. Suk, and B. Zitová. *Moments and Moment Invariants in Pattern Recognition.* Wiley, Chichester, U.K., (2009).

[8] A. Goshtasby. *Template matching in rotated images.* IEEE Transactions on Pattern Analysis and Machine Intelligence **7** (1985), 338–344.

[9] M. Liu and P.-T. Yap. *Invariant representation of orientation fields for fingerprint indexing.* Pattern Recognition **45** (2012), 2532–2542.

[10] W. Liu and E. Ribeiro. *Detecting singular patterns in 2-D vector fields using weighted Laurent polynomial.* Pattern Recognition **45** (2012), 3912–3925.

[11] X. Ma, R. Pan, and L. Wang. *License plate character recognition based on Gaussian–Hermite moments.* In 'Second International Workshop on Education Technology and Computer Science ETCS'10', volume 3, 11–14. IEEE, (2010).

[12] M. Schlemmer, M. Heringer, F. Morr, I. Hotz, M.-H. Bertram, C. Garth, W. Kollmann, B. Hamann, and H. Hagen. *Moment invariants for the analysis of 2D flow fields.* IEEE Transactions on Visualization and Computer Graphics **13** (2007), 1743–1750.

[13] J. Shen. *Orthogonal Gaussian–Hermite moments for image characterization.* In 'Intelligent Robots and Computer Vision XVI: Algorithms, Techniques, Active Vision, and Materials Handling', D. P. Casasent, (ed.), volume 3208, 224–233. SPIE, (1997).

[14] J. Shen, W. Shen, and D. Shen. *On geometric and orthogonal moments.* International Journal of Pattern Recognition and Artificial Intelligence **14** (2000), 875–894.

[15] Y. Wu and J. Shen. *Properties of orthogonal Gaussian–Hermite moments and their applications.* EURASIP Journal on Applied Signal Processing (2005), 588–599.

[16] B. Yang and M. Dai. *Image analysis by Gaussian–Hermite moments.* Signal Processing **91** (2011), 2290–2303.

[17] B. Yang, J. Flusser, and T. Suk. *Steerability of Hermite kernel.* International Journal of Pattern Recognition and Artificial Intelligence **27** (2013), 1–25.

[18] B. Yang, G. Li, H. Zhang, and M. Dai. *Rotation and translation invariants of Gaussian–Hermite moments.* Pattern Recognition Letters **32** (2011), 1283–1298.

[19] B. Yang, T. Suk, M. Dai, and J. Flusser. *2D and 3D image analysis by Gaussian–Hermite moments.* In 'Moments and Moment Invariants - Theory and Applications', G. A. Papakostas, (ed.), 143–173. Science Gate Publishing, (2014).

Predicting Tennis Match Outcomes Using Logistic Regression

Tomáš Kouřim

3rd year of PGS, email: `kourim@outlook.com`
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Petr Volf, Department of Stochastic Informatics
Institute of Information Theory and Automation, CAS

Abstract. The demand for proper sport match prediction tools is constantly increasing together with the amount of money put into sports betting. A log-linear tennis result prediction model based on player rankings, past performance, current form and bookmaker's odds is developed in this paper and tested on ATP and WTA Grand Slam matches in the 2011-2016 seasons.

Keywords: tennis result forecast, log-linear regression, Grand Slam

Abstrakt. Celosvětově vzrůstající množství prostředků vložených do sportovních sázek stupňuje poptávku po kvalitních nástrojích k predikci sportovních výsledků. V tomto článku je odvozen log-lineární model predikce výsledků tenisových utkání založený na postavení hráčů v žebříčku, jejich historických výsledcích, současné formě a kurzech bookmakerů. Výsledky jsou testovány na zápasech série turnajů Grand Slam mužů i žen v letech 2011-2016.

Klíčová slova: predikce tenisových výsledků, loglineární regrese, Grand Slam

1 Introduction

The demand for reliable sport outcomes prediction methods has been constantly raising over the past several years. Such a rise can be explained be many causes. Not surprisingly, prediction models are of highest importance for sports betting companies as well as bettors, who want to obtain (or maintain) a competitive advantage over their rivals. But they are of use for team managers, coaches or players themselves as they can help to point out weaknesses or strong elements of the game. Last but not least there are useful for legal authorities across institutions as they can help to reveal illegal betting, result manipulation, bribery and corruption in general. The European Sports Security Association regularly reports on suspicious betting activities. The latest report (Q2-2016) contained 41 cases of suspicious betting activity, 34 (83 %) of which occurred in tennis only[1].

In this paper the tennis match results of the major "Grand Slam" tournaments are predicted using logistic regression models with player ranking and points, their past performance on the same tournament, their current form a and also bookmaker's winning odds as independent variables. Several approaches are introduced and their performance is compared on different sets of data from both men and women professional tennis.

The remainder of this paper is organized as follows. Section 2 gives brief overview of some other work regarding tennis and sports results prediction. Section 3 describes the data that was used in this paper, section 4 gives the description of how the models were derived. Their performance on different sets of data is described in section 5, which also concludes this paper.

2 Related work

There are several different approaches in modeling and simulating tennis games. The most common ones use Markov chains as the baseline of the model, creating Markov-like chains usually from one particular part of the game - set by set, game by game, point by point or even rally by rally [2, 7, 11, 13]. Other approaches use some sort of regression (logistic, probit) [5, 3, 4, 6, 8] or econometric methods [12]. Comparison of some of the existing methods can be found in Kovalchik [10].

The methods can be also divided into those focusing on the match result itself [5, 8] and those focusing on the partial results during a match - in play probabilities[9, 7].

3 Data description

Tennis competitions are organized by three major organizations. The International Tennis Federation (ITF), the Association of Tennis Professionals (ATP) and the Women Tennis Association (WTA). ATP covers the most prestigious men tournaments, WTA does the same for women tennis. ATP tournaments are divided into 4 main levels[1] according to their importance and the number of ranking points the winner gets. The lowest level is ATP 250, where the winner gets 250 points, then comes ATP 500 with 500 points for the winner, Masters 1000 with the winner gaining 1000 points and finally the most prestigious Grand Slam tournaments (Australian Open, French Open, Wimbledon and US Open), where the winner increases his point account by 2000 points. WTA has three levels of tournaments. The lowest level, International, awards 280 points to the winner. Premier tournament winner gets up to 1000 points and Grand Slam winner gains 2000 points. ITF organizes lower level tournaments for both men and women and also cover international competitions such as the Davis Cup, Fed Cup or the Olympic games.

There is a very complex tennis database available at www.tennis-data.co.uk. It contains data about all ATP and WTA matches since 2000 and 2007 respectively. Among others, there is the information about what tournament the game was part of, who did participate, what the participants' ranking was at the time of the game as well as their current ranking points, the result of the game and the winning odds of each player from up to five different bookmakers. It also has the information whether the match was finished regularly or whether on of the players retired. In this paper only the matches were considered that were finished regularly, had complete information about tournament, player rankings and ranking points, result and had the winning odds from at least one bookmaker available. I also omitted the older results and only worked with the matches from the 2010 season and newer.

[1] The Olympic games, The Masters Cup and the Davis Cup also count for ATP level tournaments.

year \ tournament	ATP 250	ATP 500	Masters 1000	Grand Slam	Total
2010	64.37 %	65.78 %	63.87 %	74.95 %	66.48 %
2011	66.90 %	68.62 %	63.15 %	74.95 %	67.89 %
2012	62.75 %	69.74 %	69.39 %	74.85 %	67.45 %
2013	62.34 %	64.97 %	66.54 %	75.26 %	66.14 %
2014	65.55 %	66.12 %	69.76 %	74.22 %	68.26 %
2015	65.05 %	66.09 %	71.69 %	76.24 %	68.85 %
2016	64.55 %	65.20 %	68.00 %	76.19 %	68.06 %
Total	64.41 %	67.01 %	67.65 %	75.16 %	67.63 %

Table 1: Success rates for predicting ATP match results: Higher ranked player wins.

year \ tournament	ATP 250	ATP 500	Masters 1000	Grand Slam	Total
2010	66.63 %	71.66 %	71.32 %	79.50 %	70.82 %
2011	72.07 %	73.14 %	67.59 %	79.30 %	72.65 %
2012	68.17 %	72.91 %	73.28 %	78.97 %	72.06 %
2013	67.79 %	70.59 %	67.47 %	78.39 %	70.18 %
2014	70.12 %	67.48 %	71.22 %	77.13 %	71.34 %
2015	68.21 %	69.78 %	75.00 %	78.92 %	72.05 %
2016	68.50 %	67.84 %	72.04 %	80.74 %	71.95 %
Total	68.81 %	70.66 %	70.86 %	78.76 %	71.49 %

Table 2: Success rates for predicting ATP match results: Bookmaker’s favorite wins.

4 Experiments

Three different approaches were used in order to predict tennis match results. The first two basic approaches should serve as a benchmark, they simply predict the higher ranked player or the bookmaker’s favorite[2] to win the match. The results of the two approaches for each year[3] and each tournament type are shown in tables 1, 2, 3 and 4. Several assumptios can be made according to these results. They show that Grand Slam tournaments are better predictable using both methods, that women tennis is harder to predict in general and finally that the bookmaker’s odds is a better result predictor than the official player ranking.

The third approach follows the results of [5], where a logistic regression model is derived to predict the results of Grand Slam tournaments. The model uses a combination of player ranks and previous year results of the players on the same tournament. The model shows remarkable results, but it is impossible to verify these results, because some details of the model derivation are omitted. It is not explained whether the positive or

[2] The average odds from all available bookmakers was used in order to determine the favorite.

[3] The 2016 season still does not have all the lower level tournaments finished.

year \ tournament	International	Premier	Grand Slam	Total
2010	66.37 %	64.78 %	73.17 %	67.24 %
2011	65.49 %	64.72 %	70.22 %	66.21 %
2012	64.26 %	66.44 %	69.15 %	66.11 %
2013	65.48 %	66.74 %	70.18 %	66.95 %
2014	63.24 %	65.65 %	69.01 %	65.32 %
2015	62.25 %	64.44 %	68.33 %	64.39 %
2016	63.41 %	62.52 %	70.22 %	64.75 %
Total	64.56 %	65.14 %	70.27 %	66.02 %

Table 3: Success rates for predicting WTA match results: Higher ranked player wins.

year \ tournament	International	Premier	Grand Slam	Total
2010	70.02 %	71.15 %	73.58 %	71.18 %
2011	69.72 %	70.32 %	74.65 %	70.98 %
2012	67.02 %	71.20 %	73.39 %	69.89 %
2013	69.54 %	69.20 %	74.75 %	70.54 %
2014	64.62 %	70.56 %	74.04 %	68.75 %
2015	66.07 %	65.61 %	73.55 %	67.45 %
2016	66.17 %	66.44 %	70.42 %	67.31 %
Total	67.98 %	69.67 %	73.35 %	69.66 %

Table 4: Success rates for predicting WTA match results: Bookmaker's favorite wins.

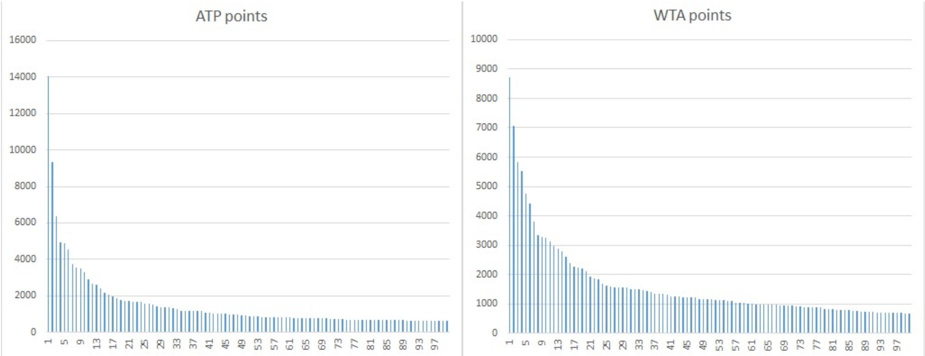


Figure 1: Ranking points distribution of top 100 ATP and WTA players. Source: www.atpworldtour.com and www.wtatennis.com, as available on Sep 29th 2016.

negative value of player rank difference is considered or if the round reached by a player on the same tournament last year is a categorical or a numerical variable.

In this paper, the following regressors were considered. The official ranks of both players together with their ranking points. Picture 1 shows that the points difference between player #1 and player #5 is significantly bigger than that between players #50 and #55, suggesting that the difference between their strengths could differ as well and that ranking points could contain additional information. The next regressors were the players' performances on the same tournament last year. This regressor serves as a surrogate variable for the tournament surface, as all the tournaments are played on a different surface, which can play a crucial role for the performance of certain players. This approach was also suggested in [5]. It takes a maximal value of 7 when the player reached the finals last year and minimal value of 0 when the player did not participate last year. Finally, two regressors were considered reflecting the current form of both players. The form of the $j-th$ player was calculated as follows

$$form_j = \sum_{i=1}^{N} R_j(i)\, V(i)\,,$$

where $R(i)$ stands for the result of the $i-th$ match preceding the current match (in which player j participated), taking the values of -1 and 1, $V(i)$ is the value of a tournament (where the $i-th$ match took place) according to its points prize pool (the values are shown in table 5[4]) and N is a memory parameter. In this paper, N was set to 6, that is the value $form$ reaches its maximal value of 12 for the Grand Slam finals. In this case, the 6 consecutive Grand Slam wins are considered for each finalist, suggesting that they are both in a top form. This is in accordance with the common sense for such an important tennis match.

This regressors were considered for the first basic log-linear (logit) model. Average bookmaker's winning odds were added for the second model. The third model only considered the four ranking and points related regressors and the last model only considered the four regressors representing the current and last year performance. All the models were derived using the *glm* (Generalize linear model) function in the R software, which

[4] For Premier tournaments the value corresponding with the average points awarded was considered.

Tournament	Value
Grand Slam	2
Masters 1000	1
ATP 500	0.5
ATP 250	0.25
Premier	0.673
International	0.28

Table 5: Tournament values.

Year\ prediction (in %)	Basic	Odds	Rank only	Performance only
2011	74.53	77.64	74.74	74.95
2012	75.26	78.35	74.64	75.05
2013	74.01	77.96	75.47	73.80
2014	74.01	76.92	74.22	73.60
2015	75.62	78.51	75.41	76.44
2016	76.81	79.30	75.57	78.05

Table 6: Prediction power of derived model, ATP Grand Slams, learning from previous year.

uses maximal likelihood to estimate model parameters, and are of the form

$$\ln(\frac{p}{1-p}) = \alpha + \beta\vec{x},$$

where p is the probability that a higher ranked player wins the match, $\vec{x}$ is the vector of independent variables,α is intercept value and β are model coefficients. The results are discussed further in Section 5.

5 Results and future work

All the derived models were tested on all Grand Slam matches since 2010. The Grand Slam matches were chosen because they are the most prestigious ones and are also the only tournament type held every year under same conditions (actually for much longer period of time than since 2010). The model was first trained on training data and then tested on new, previously unseen testing data. This was done on a yearly basis, that is for every year (starting 2011) the data from that year was used as a testing set. Two cases of learning were considered. First, only data from one previous year were used as learning set, then data from all previous years were used. The results can be seen in tables 6, 7, 8 and 9.

Several interesting observations can be concluded from these results. The comparison of respective tables shows that adding training data from further in the past does not improve the performance of the models. This makes sense from the tennis point of view as well as from data mining point of view. Tennis is a game that is constantly developing (new materials, techniques, ...) and thus older games might not be as similar to the

Year\ prediction (in %)	Basic	Odds	Rank only	Performance only
2011	74.53	77.64	74.74	74.95
2012	74.23	78.35	75.05	74.01
2013	73.18	77.96	74.84	75.88
2014	73.80	76.72	73.80	73.39
2015	75.86	79.34	76.24	75.82
2016	75.57	79.92	74.74	76.60

Table 7: Prediction power of derived model, ATP Grand Slams, learning from all previous years.

Year\ prediction (in %)	Basic	Odds	Rank only	Performance only
2011	70.62	72.03	70.62	70.22
2012	69.15	72.58	70.97	68.55
2013	69.38	74.16	67.79	70.78
2014	70.82	72.43	70.02	69.21
2015	67.13	71.94	67.33	67.94
2016	69.22	70.02	72.23	68.21

Table 8: Prediction power of derived model, WTA Grand Slams, learning from previous year.

Year\ prediction (in %)	Basic	Odds	Rank only	Performance only
2011	70.62	72.03	70.62	70.22
2012	69.96	72.38	70.16	68.96
2013	69.58	73.96	71.45	70.38
2014	71.63	73.64	70.02	69.22
2015	68.94	73.15	69.14	68.34
2016	70.22	70.42	71.28	69.42

Table 9: Prediction power of derived model, WTA Grand Slams, learning from all previous years.

new ones as one could think. Mathematically speaking, there are over 500 Grand Slam matches every year which is enough to build a model. The additional matches from previous years can bring some new information, but would bring in some noise as well; the results show that the new information does not exceed the noise contained in the older data.

The performance levels of the basic model are similar to the discriminatory power of the rank itself. This suggests that the players ranks carry the core information. This is further accentuated by the results of the model with only ranking related parameters, which results are again very similar to those of rank only. However, the model based on only the performance parameters delivers similar, sometimes even better results than then one with rank information available. The only model that shows better results is the one that uses bookmaker's odds as its additional regressors. Here the results are similar to those when predicting the bookmaker's favorite as the match winner. This suggests that most if not all possible information about a tennis game is exploited by the bookmaker's odds and that logistic regression models cannot add much more to this information. It also suggests that tennis matches are only predictable up to a certain level. This is in accordance with the reality, that is the fact that one tennis match is an encounter between two people, two individuals, and as such their future behavior is only hardly predictable.

Dziedzic presented different results [5]. The best presented performance of 90.01 % is presented for the 2014 Women US Open. However, this performance was achieved using only 74 matches, but there were 127 matches on that tournament. The same holds for the other cases presented in [5] as well, for example only 876 matches from 2009-2013 WTA Grand Slams were used for training, instead of all 2540 that took place during that period. The paper suggests that all the data comes from www.tennis-data.co.uk (which contains data about all matches), but does explain how the subsets of matches were selected. The presented performance is thus unverifiable.

The results and methods presented in this paper suggest that it is not an easy task to find a predicting method more powerful than bookmaker's odds. There are still many options that are yet to be investigated but so far, the gamblers dream - to beat bookmaker's odds - still remains untrue.

References

[1] European Sports Security Association. Essa q2 2016 integrity report. http://www.eu-ssa.org/wp-content/uploads/Draft-ESSA-Q2-2016.pdf, 2016. Accessed: 2016-09-30.

[2] Tristan Barnett and Stephen R Clarke. Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*, 16(2):113–120, 2005.

[3] Bryan L Boulier and Herman O Stekler. Are sports seedings good predictors?: an evaluation. *International Journal of Forecasting*, 15(1):83–91, 1999.

[4] Julio Del Corral and Juan Prieto-Rodriguez. Are differences in ranks good predictors for grand slam tennis matches? *International Journal of Forecasting*, 26(3):551–563, 2010.

[5] Edyta Dziedzic and Gordon Hunter. Proceedings of the 5th international conference on mathematics in sport. In Kay Anthony, editor, *Predicting the Results of Tennis and Volleyball Matches Using Regression Models, and Applications to Gambling Strategies*, pages 32–37. Univ. of Loughborough, GB, 2015.

[6] Keith F Gilsdorf and Vasant A Sukhatme. Testing rosen's sequential elimination tournament model incentives and player performance in professional tennis. *Journal of Sports Economics*, 9(3):287–303, 2008.

[7] Gordon J.A. Hunter and Krzysztof Zienowicz. Can markov models accurately simulate lawn tennis rallies? In *Proceedings of the Second International Conference on Mathematics in Sport*, volume 1, pages 69–75. Univ. of Groningen, NL, 2009.

[8] Franc JGM Klaassen and Jan R Magnus. Forecasting the winner of a tennis match. *European Journal of Operational Research*, 148(2):257–267, 2003.

[9] Tomáš Kouřim. Mathematical models of tennis matches applied on real life odds. *Doktorandské dny FJFI*, 2015.

[10] Stephanie A. Kovalchik. Proceedings of the 5th international conference on mathematics in sport. In Kay Anthony, editor, *Comparative performance of models to forecast match wins in professional tennis: Is there a GOAT for tennis prediction?*, pages 91–96. Univ. of Loughborough, GB, 2015.

[11] Paul K Newton and Kamran Aslam. Monte carlo tennis: A stochastic markov chain model. *Journal of Quantitative Analysis in Sports*, 5(3), 2009.

[12] J James Reade, Sachiko Akie, et al. Using forecasting to detect corruption in international football. In *Proceedings of the 4th International Conference on Mathematics in Sport*, 2013.

[13] Demetris Spanias and William J Knottenbelt. Predicting the outcomes of tennis matches using a low-level point model. *IMA Journal of Management Mathematics*, page dps010, 2012.

Numerická analýza Turingovy nestability pro Schnackenbergův model s prostorově závislou kinetikou*

Michal Kozák

4. ročník PGS, email: michal.kozak@fjfi.cvut.cz
Katedra matematiky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

školitel: Václav Klika, Katedra matematiky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

Abstract. Self-organised spatial pattern formation is one of the main focus of Mathematical Biology and the Turing reacion-diffusion model is widely studied for the case of constant coefficients. The spatial dependence yields to significant complications, which will be task of this paper. Here, we will follow the analysis performed for model with spatial dependence in the coefficient at the linear term of the activator kinetics and using series of simulations for Schnackenberg kinetics we will analyse relation between the earlier obtained conditions and the behaviour of the non-linear system. The results indicate where mentioned situations do not match. This will be helpful to find more precise conditions, if the emergence of patterns occurs or not.

Keywords: Schnackenberg kinetics, reaction-diffusion model, non-homogenous Turing's model, pattern formation.

Abstrakt. Samovolný vznik prostorových vzorů je jedním z hlavních zájmů Matematické biologie a Turingův reakčně-difuzní model je v případě konstantních koeficientů velmi studovaným problémem. Závislost koeficientů na prostoru přináší znatelné ztížení problému, což bude předmětem této práce. Konkrétně, navážeme na analýzu takového modelu v případě závislosti v koeficientu u lineárního členu kinetiky aktivátoru a pomocí série simulací pro Schnackenbergovy kinetiky prozkoumáme vztah nalezených podmínek s chováním nelineárního systému. Výsledky ukazují na případy, kdy si uvedené případy neodpovídají, což napomáhá k přesnějšímu nalezení podmínek, za kterých v daném systému dochází či nedochází ke vzniku prostorových struktur.

Klíčová slova: Schnackenbergovy kinetiky, reakčně-difuzní systém, nehomogenní Turingův model, pattern formation.

1 Úvod

Vznik prostorových struktur je jedním z nejdůležitějších jevů v mnoha nerovnovážných systémech, počínaje vývojovou biologií přes růst krystalů v tuhnoucích slitinách, konče plasmou nebo polovodiči. Základním mechanismem k narušení symetrie je nestabilita způsobená difuzí (diffusion driven instability; Turingova nestabilita [4]) reakčně-difuzních systémů (RD-systém). Turing ukázal, že malé perturbace homogenního systému autokata-

*Tato práce byla podpořena grantem SGS15/215/OHK4/3T/14.

lyticky a inhibičně difundujících druhů mohou způsobit nestabilitu, která vede ke vzniku prostorových struktur.

Výsledné vzory Turingova modelu jsou prostorově periodické a velmi symetrické, což nedostačuje k popsání skutečných vzorů v přírodě. Některé prostorové odlišnosti bychom sice mohli odůvodňovat nutným zjednodušením modelu oproti skutečnosti, například u skvrn jaguára, jindy jsou ale rozdíly vzorku v prostoru důležité. Příkladem je rozložení myších nebo kočičích vousků [5], střídající se tenké a tlusté pruhy u korálové rybičky Lionfish [8] nebo růst prstů na končetinách [6]. Z úvodních numerických simulací (například [8]) vyplývá, že jedním ze způsobů dosažení takových výsledků je přidání prostorové závislosti do parametrů úlohy; a analýzou takového modelu se budeme zabývat.

Analýza stability řešení systému s nekonstatními koeficienty je analyticky velmi obtížná, už jen proto, že samotný stacionární stav je často prostorově nekonstantní. I přes to se již v literatuře setkáme s částečnými výsledky; analýza tzv. spikes pro Gierer-Meinhardtův model [7] nebo analýza stability v případech speciálního tvaru prostorové závilosti koeficientů: v absolutním členu kinetiky uvažující ε-řady kolem homogenního stacionárního stavu [10] a [11] nebo ten samý případ závislosti ve tvaru skokové funkce [9]. My se budeme zabývat analýzou modelu v jedné prostorové dimenzi (interval $[0, L]$) s prostorovou závislostí koeficientu u lineárního členu kinetiky aktivátoru

$$\begin{aligned} \partial_t u &= d_1 \partial_{xx} u + f_1(u,v) + h(x)u \\ \partial_t v &= d_2 \partial_{xx} v + f_2(u,v) \end{aligned} \quad \text{v } (0,\infty) \times (0,L) \tag{1}$$

s Neumannovými okrajovými podmínkami

$$\frac{\partial u}{\partial n}(0) = \frac{\partial u}{\partial n}(L) = 0, \quad \frac{\partial v}{\partial n}(0) = \frac{\partial v}{\partial n}(L) = 0. \tag{2}$$

Abychom se mohli soustředit na důsledky samotné prostorové závislosti, budeme uvažovat co nejjednodušší formu závislosti – funkci konstatní téměř všude. To je motivováno představou spojení dvou systémů (tkání/prostředí) s mírně odlišnými parametry. Pro dostatečně velké oblasti vzhledem ke změně těchto parametrů by se dalo čekat, že vzorky na obou koncích oblasti se nebudou výrazně lišit od vzorů samostatných systémů. Zajímavý případ nastane, pokud malá změna v koeficientech způsobí díky nelineárním kinetikám velmi rozdílné vzory.

Vhodnými funkcemi $h(x)$ tedy může být skoková funkce

$$h(x) = \begin{cases} 0 & x \in [0,\xi), \\ s & x \in [\xi, L] \end{cases}$$

nebo obdobné hladké funkce

$$h_\eta(x) = h(x) * \eta_\delta(x), \qquad h_\delta(x) = \frac{s}{2}\Big(1 + \tanh\frac{x-\xi}{\delta}\Big),$$

kde $\eta_\delta(x)$ značí regularizátor. V numerických simulacích jsou tyto funkce pro dostatečně malé parametry δ v počítači reprezentovány stejně, vlastnosti celých systémů s těmito funkcemi by tedy měly být velmi podobné. Pro analytický přístup je nejvhodnější prvně jmenovaná, tu budeme také nadále používat.

V následující kapitole nejdříve vypíšeme základní fakta o klasickém Turingově modelu, na která se budeme v textu odkazovat. Dále shrneme postup a výsledky analýzy modelu s prostorovou závislostí v koeficientu u lineárního členu kinetiky aktivátoru prezentované minulý rok ([1] a [2]; uceleněji v připravované publikaci [3]). To nám poslouží jako motivace a odůvodnění pro třetí kapitolu, ve které prezentujeme výsledky simulací systému se Schnackenbergovými kinetikami

$$f_1(u,v) = a - u + u^2 v, \quad f_2(u,v) = b - u^2 v \tag{3}$$

a dle nich porovnáme platnost podmínek vzniku vzorků z Kapitoly 2. Práce pak končí s kapitolou se shrnutím a závěrečnou diskusí.

2 Analýza modelu se skokovou funkcí

Uvažme tedy na chvíli klasický Turingův model, což je systém (1) s $h(x) \equiv 0$. Nechť tento systém má homogenní stacionární stav a označme $\mathcal{A}$ Jacobiho matici ze zobrazení (f_1, f_2) v tomto bodě. Pak Turingovy podmínky pro vznik vzorku jsou tvaru

$$\operatorname{tr}\mathcal{A} < 0, \quad \det\mathcal{A} > 0, \quad a_{11}d_2 + a_{22}d_1 > 0, \quad (a_{11}d_2 + a_{22}d_1)^2 > 4d_1d_2\det\mathcal{A}. \tag{4}$$

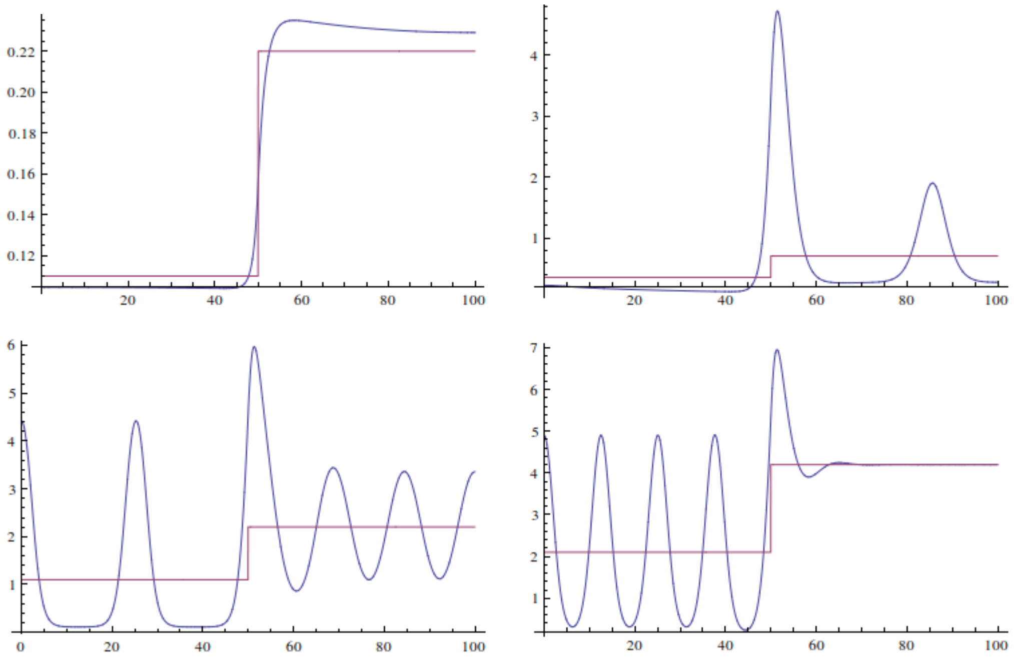
Dále již uvažme systém (1) se skokovou funkcí $h(x)$. Nejdříve se pro lepší představu podívejme na několik výsledků simulací pro konkrétní kinetiky a parametry, viz Obrázek 1. Zde, vedle toho, že dostáváme vzory velmi zajímavé pro aplikace v biologii (Obr. 1C, tedy vzorek s odlišnými frekvencemi na koncích intervalu), čímž se tady zabývat nebudeme, vidíme, že až na znatelné vychýlení v místě skoku funkce $h(x)$ se vzorek buď rychle zatlumuje, nebo tvoří periodické struktury podobně jako v klasickém případě. Tyto grafy nás vedou k zavedení následujícího označení; situaci s rychle se zatlumujícím řešením označme, že vzorek nevzniká, ostatní řešení pak nazvěme vzorem. Důvod tohoto odlišení vyniká, pokud si představíme měřítko malého skoku vůči velké velikosti oblasti. Dále označme Turingovým prostorem množinu parametrů úlohy (1), které vedou ke vzniku vzoru.

Nejdříve jsme zkoumali lineární systém, tedy úlohu

$$\begin{aligned} 0 = \partial_t u &= d_1\partial_{xx}u + b_{10} + (h(x) + b_{11})u + b_{12}v \\ 0 = \partial_t v &= d_2\partial_{xx}v + b_{20} + b_{21}u + b_{22}v \end{aligned} \quad \text{v } (0, L). \tag{5}$$

s Neumannovými okrajovými podmínkami. Takový systém má stacionární řešení, které lze analyticky spočítat pomocí metody použité v [9]: Stacionární systém se nejdříve rozdělí na dva podsystémy, „levý“ nad intervalem $[0, \xi]$, „pravý“ nad intervalem $[\xi, L]$ a doplní se o navazující okrajové podmínky v bodě ξ. Systémy se pak transformují na systémy dvou nezávislých eliptických rovnic, které nejsou těžké v případě jedné prostorové dimenzi spočítat. Více než samotný tvar řešení je zajímavé, že výsledek je dvojího tvaru, jejichž volba závisí na podmínkách

$$\begin{gathered} (d_2b_{11} + d_1b_{22})^2 - 4d_1d_2(b_{11}b_{22} - b_{12}b_{21}) > 0, \\ (d_2(b_{11} + s) + d_1b_{22})^2 - 4d_1d_2((b_{11} + s)b_{22} - b_{12}b_{21}) > 0 \end{gathered} \tag{6}$$



Obrázek 1: Vyobrazení grafu koncentrace aktivátoru u blízkému stacionárnímu stavu úlohy (1) se Schnackenbergovými kinetikami (3) na intervalu $[0, 100]$ s Neumannovými okrajovými podmínkami a parametry: $d_1 = 1$, $d_2 = 100$, $s = 0.5$, $\xi = L/2$, $a = 0.1$ a A) $b = 0.01$, B) $b = 0.25$, C) $b = 1$, D) $b = 2$. Číslování obrázků je v celém textu stejné – po řádcích, zleva doprava.

Je jednoduché nahlédnout, že tyto podmínky mají stejný tvar jako podmínka pro Turingovu nestabilitu (4).

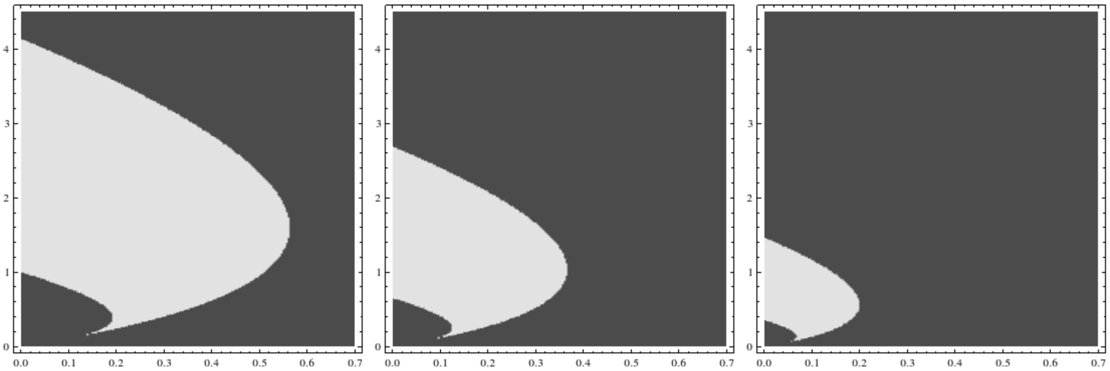
Při vyšetřování asymptotického chování řešení nelineárních reakčně-difuzních systémů se používá znalost chování příslušného linearizovaného systému, tedy lineárního systému co nejblíže aproximujícího chování původního nelineárního. To provedeme i zde; znovu systém rozdělíme na levý a pravý podsystém (označení vysvětleno výše) s přidanými navazujícími podmínkami, čímž dostaneme dva systémy s konstantními koeficienty, u kterých lehce vyšetříme stabilitu příslušných stacionárních stavů (složeninu těchto stacionárních stavů označme zkratkou PSS). Ty pak vedou k podmínkám analogickým k (6):

$$(d_2 b_{11}^j + d_1 b_{22}^j)^2 - 4 d_1 d_2 (b_{11}^j b_{22}^j - b_{12}^j b_{21}^j) > 0, \quad j \in \{L, R\}, \tag{7}$$

kde b_{kl}^j jsou prvky Jacobiho matice vektoru funkcí (f_1, f_2) vzhledem k jednotlivým podsystémům (systémy zvlášť nad intervalem $[0, \xi]$ a nad intervalem $[\xi, L]$) spočtených v bodech příslušného homogenního starionárního stavu daného podsystému. Jelikož je pravdivost těchto podmínek lehce zjistitelná z parametrů systému, slouží jako rychlě rozpoznání chování systému. Poznamenejme, že jelikož používáme informaci z lineárního systému u nelineárního, nemůžeme čekat úplnou shodu podmínek se skutečnou existencí či neexistencí vzorů – toto pravidlo platí u každé takové analýzy, tedy i klasických podmínek Turingovy nestability. Více se k tomu vrátíme v závěru.

3 Výsledky

Úkolem této části je prozkoumat na příkladech, jak jsou podmínky (7) zmíněné v minulé kapitole použitelné pro odlišení parametrů modelů, u kterých vzniká nebo nevzniká vzor, případně který ze tří typů uvedených na Obrázku 1. Opět vezmeme Schnackenbergovy kinetiky a budeme zkoumat příslušný Turingův prostor. Pro přehlednost výsledků volme dva parametry, a to a a b; zbylé parametry fixujme. Volme velikost intervalu $L = 500$, skok v polovině intervalu $\xi = 250$. Poslední parametr velikosti skoku s volme malý (vzhledem k ostatním koeficientům) libovolně vhodně tak, aby se množiny parametrů příslušné kombinacím splnění podmínek (7) daly viditelně odlišit. Ilustračně se podívejme na Turingovy prostory klasického systému se Schnackenbergovými kinetikami $f_1 = a + (s-1)u + u^2v$, $f_2 = b - u^2v$) a pro parametry $s = 0$, $s = 0.25$ a $s = 0.5$, viz Obrázek 2. Výše uvedené kritérium splňuje například volba $s = 0.25$, jak je vidět později i na Obrázku 3. Za počáteční podmínku volíme PSS, tedy funkci velmi blízkou stacionárnímu řešení.



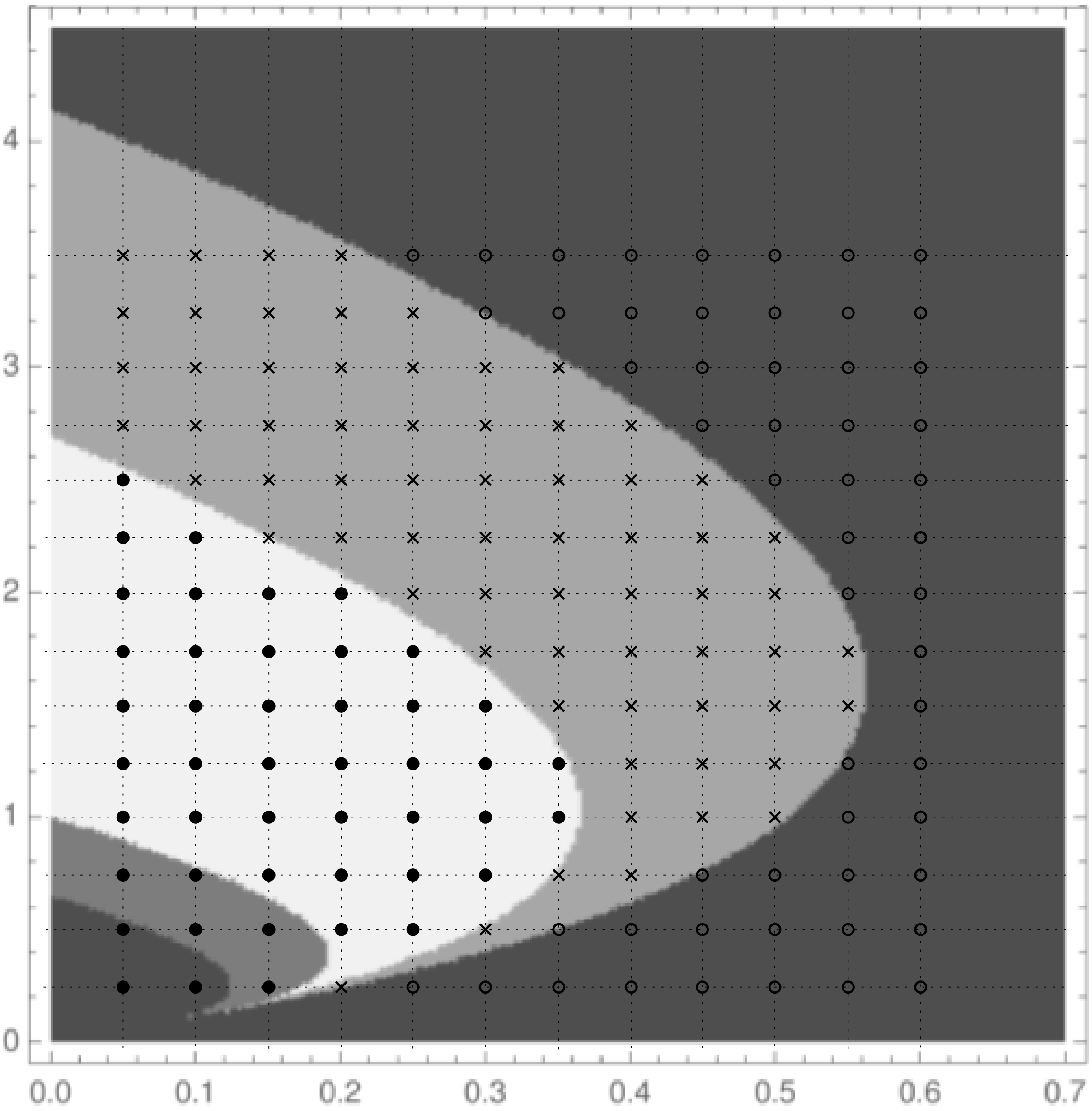
Obrázek 2: Vyobrazení Turingova prostoru pro Schnackenbergův model ($f_1 = a + (s-1)u + u^2v$, $f_2 = b - u^2v$) na intervalu $[0, 100]$ s parametry $d_1 = 1$, $d_2 = 100$ a A) $s = 0$, B) $s = 0.25$, C) $s = 0.5$. Na ose x je parametr a, na ose y parametr b. Světlá barva značí ty parametry, pro které dochází k Turingové nestabilitě, tmavá opak.

Pomocí softwaru Wolfram Mathematica jsme provedli simulace pro parametry $a = 0.05i$ pro $i = 1, 2, \ldots, 12$ a $b = 0.25j$ pro $j = 1, 2, \ldots, 14$. Zajímalo nás, pro které kombinace parametrů a a b vyjde který typ vzoru zobrazených na Obrázku 1. A na Obrázku 3 můžeme vidět výsledek.

Výsledky simulace se s predikcí shodují po velké části zkoumaného Turingova prostoru, až překvapivě přesně i na hranicích jednotlivých podoblastí. Na Obrázku 4 můžeme vidět, že na těchto hranicích tvar příslušných ustálených řešení v sebe postupně přechází.

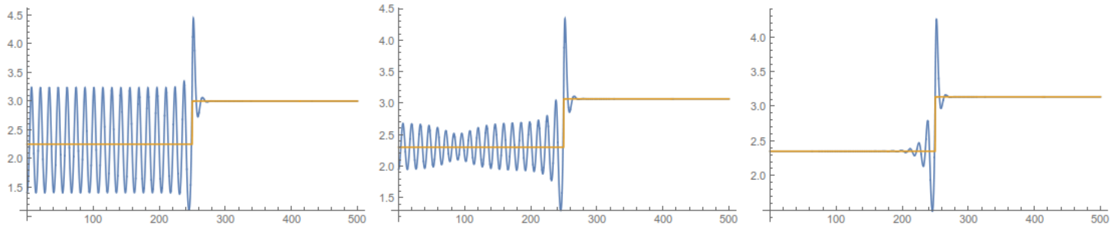
Výjimku tvoří oblast parametrů $(a, b) \in [0, 0.2] \times [0, 1]$. Z grafu koncentrace v na Obrázku 5A,B vidíme, že jsme sice vzor získali, je ale patrné, že je už mimo naši predikci, neboť je již velmi vzdálen od oblasti kolem PSS, kde platí stabilita lineární úlohy a tedy podmínky (7). Důvodem je, že v této oblasti parametrů roste díky nelinearitám velikost skoku v PSS pro koncentraci v nadevšechny meze. Skutečně, tento skok je roven

$$\frac{b(s-2)s}{(a+b)^2},$$



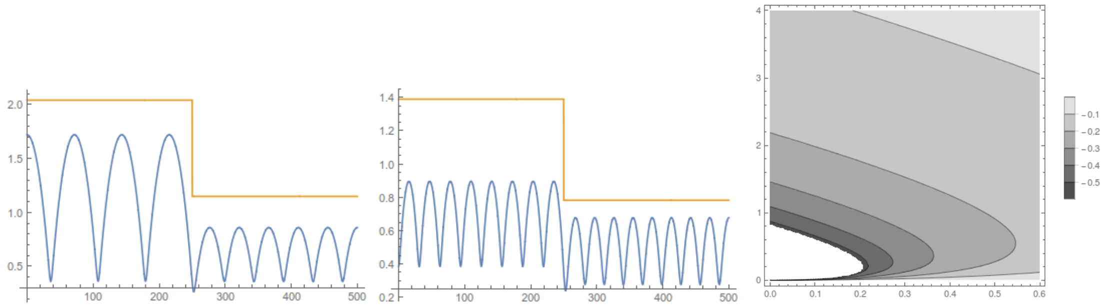
Obrázek 3: Zobrazení výsledků simulací systému se Schnackenbergovými kinetikami se skokem v bodě ξ ($f_1 = a + (h(x) - 1)u + u^2v$, $f_2 = b - u^2v$) pro parametry $L = 500$, $\xi = L/2$ a $s = 0.25$. Na pozadí je zakreslen průnik Turingových prostorů z Obrázku 2 pro $s = 0$ a $s = 0.25$; tedy predikce chování systému dle podmínek (7) – světlá barva značí predikovaný vznik vzoru (Obr. 1C), tmavá barva žádný vzor (Obr. 1A), barvy na škále mezi nimi – světlejší vzor s nehomogenitou vpravo (Obr. 1B), tmavší vzor s nehomogenitou vlevo (Obr. 1D). Značkami jsou pak ve vrcholech mříže označeny výsledky simulací: $\circ$ pro žádný vzor (Obr. 1A), $\bullet$ pro vzor na obou stranách (Obr. 1C) a $\times$ pro vzor jen na levé straně (Obr. 1D). Na ose x je parametr a, na ose y parametr b.

jeho graf můžeme vidět na Obrázku 5C, ze kterého vidíme, že podoblast Turingova prostoru, kde predikce neseděla, odpovídá oblasti, pro kterou je skok příliš velký. Tento



Obrázek 4: Vyobrazení grafu koncentrace aktivátoru u blízkému stacionárnímu stavu úlohy (1) se Schnackenbergovými kinetikami (3) na intervalu $[0, 500]$ s Neumannovými okrajovými podmínkami a parametry: $d_1 = 1$, $d_2 = 100$, $\xi = L/2$, $s = 0.25$, $b = 1.75$ a A) $a = 0.5$, B) $a = 0.55$, C) $a = 0.6$.

výsledek se zdá být v souladu s počáteční úvahou, že by měl být skok malý vzhledem k velikosti oblasti, aby efekt skoku byl stále zanedbatelný vůči efektu samotného systému.



Obrázek 5: Ilustrace místa neshody výsledků simulací s predikcí. Na prvních dvou obrázcích je graf koncentrace inhibitoru v se Schnackenbergovými kinetikami (3) na intervalu $[0, 500]$ s Neumannovými okrajovými podmínkami a parametry: $d_1 = 1$, $d_2 = 100$, $\xi = L/2$, $s = 0.25$, $a = 0.1$ a $b = 0.25$, respektive $b = 0.5$. Na třetím obrázku je zobrazení funkce skoku PSS u inhibitoru v v závistlosti na parametrech a a b.

4 Závěr

V této práci jsme navázali na předešlou práci, která se zabývala hledáním podmínek rozhodující o vzniku či nevzniku prostorově nekonstatních vzorků obecného RD systému s prostorovou závislostí v koeficientu u lineárního členu kinetiky ve formě skokové funkce $h(x)$. Cílem bylo vyzkoušet platnost těchto podmínek na příkladu úplného nelineárního systému a ilustrovat tak souvislost mezi predikcí získané z lineárního systému s chováním nelineárního systému. Provedli jsme tedy sérii simulací pro Schnackenbergovy kinetiky a výsledky jsme viděli v předchozí kapitole.

Nalezená míra shody se dá považovat za úspěšnou, potvrzující naši domněnku; ale zároveň poukazuje na omezenost platnosti této analýzy. Podobně jako u ostatních kvalitativních analýzách založených na získání informace z vedoucí části systému a zanedbávající méně důležité části (například linearizace) je třeba nezapomínat, že jsou tyto predikce platné jen velmí blízko výchozím stavům, a tedy někdy zavádějící. Ukazuje to ale, že daná analýza je správná cesta a že tyto podmínky dávají za určitých dodatečných předpokladů (které obdobně jako u linearizace nemusí být explicitně napsatelné) kýžené výsledky.

Literatura

[1] Kozák, M., *Stability analysis of reaction-diffusion-advection equations.* Study of the dissertation thesis, (2015).

[2] Kozák, M., *Pattern formation in Turing reaction-diffusion models with spatially dependent coefficient.* Doktoradnské dny 2015, (2015),93–96.

[3] Kozák, M. a Klika, V. a Gaffney, E. A., *Pattern formation in Turing reaction-diffusion models with spatially dependent coefficient.* will be submitted in Physical Review E.

[4] Turing, A., *The chemical basis of morphogenesis.* Phil. Trans. R. Soc. Lond. B, **237** (1952), 37–72.

[5] Painter, K. J. a Hunt, G. S. a Wells, K. L. a Johansson, J. A. a Headon, D. J., *Towards an integrated experimetal–theoretical approach for assesing the mechanistic basis of hair and feather morphogenesis.* Interface Focus **2** (2012), 433–450.

[6] Economou, A. D. a Green, J. BA., *Thick and thin fingers point out Turing waves.* Genome Biology **14:101** (2013).

[7] Wei, J. a Winter, M., *Mathematical Aspects of Pattern Formation in Biological Systems.* Volume 189 of Applied Mathematical Sciences (London, 2014), *Dev. Math.*, Springer London (2014), 149–174.

[8] Page, K. M. a Maini, P. K. a Monk, N. A.M., *Complex pattern formation in reaction-diffusion systems with spatially varying parameters.* Physica D **202** (2005), 95–115.

[9] Page, K. M. a Maini, P. K. a Monk, N. A.M., *Pattern formation in spatially heterogeneous Turing reaction-diffusion models.* Physica D **181** (2002), 80–101.

[10] Glim, T. a Zhang, J a Shen Y, *Interaction of Turing patterns with an external linear morphogen gradient.* Nonlinearity **22(10)** (2009), 2541–2560.

[11] Glim, T. a Zhang, J a Shen Y a Newmann S. A., *Reaction-Diffusion Systems and External Morphogen Gradients: The Two-Dimensional Case, with an Application to Skeletal Pattern Formation.* Bull Math Biol **74** (2012), 666–687.

One-Qubit and Two-Qubit Dynamics of Chaotic Purification Protocol*

Martin Malachov

2nd year of PGS, email: martin.malachov@fjfi.cvut.cz
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:

Igor Jex, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Tamás Kiss, Wigner Research Centre for Physics
Hungarian Academy of Sciences

Abstract. Quantum entanglement is an important resource for quantum computation. Therefore, entanglement purification is essential for quantum computation and communication. Purification protocol which uses measurement-based selection may induce nonlinear dynamics with exponential sensitivity of the state evolution to initial conditions. We aim to study the action of one particular protocol on a pair of qubits. This should lead to understanding of how the protocol purifies the entanglement. In contrast to previous work which studied only pure states, we also work with mixed states which are relevant for practical purposes. In this paper, we focus on evolution of mixed one-qubit states as it plays an important role in the two-qubit evolution. All new findings give us deeper insight into the complex dynamics of iterated maps.

After a brief introduction to the topic we describe the regime of asymptotic behaviour of one-qubit evolution. We discuss this new type of dynamics as it is different than previously described chaos. Then we examine the role of this chaos in the two-qubit dynamics. In the end, we conclude all new findings on the two-qubit entanglement purification.

Keywords: qubit, quantum entanglement, chaos

Abstrakt. Kvantové provázání je důležitým zdrojem pro kvantové počítání a komunikaci, proto je důležitá jeho purifikace. Purifikační protokoly využívají selekci stavů na základě výsledků měření, a mohou tak vyvolávat nelineární vývoj systému s exponenciální citlivostí k počátečnímu stavu. Cílem našeho výzkumu je studium jednoho takového protokolu působícího na dvojici qubitů. Tím se zjistí vhodnost protokolu k purifikaci provázání. Na rozdíl od předchozích článků, které se věnovaly pouze čistým stavům, zde se berou v potaz i stavy smíšené, což je relevantní pro praktické použití protokolu. Zejména je studována akce protokolu na jednoqubitové smíšené stavy, jejichž vývoj je významnou součástí vývoje stavů dvouqubitových.

Po stručném úvodu do problematiky je detailně popsán režim asympotické evoluce jednoqubitových smíšených stavů. Tato nová dynamika je jiná, než dříve popsaný chaos. Posléze je rozvedena úloha této dynamiky ve vývoji dvouqubitových systémů. Všechny nové poznatky o evoluci provázání dvou qubitů jsou shrnuty v závěru.

Klíčová slova: qubit, kvantové provázání, chaos

*This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS16/241/OHK4/3T/14 and he Hungarian National Research, Development and Innovation Office under Contract No. NN109651.

1 Introduction

Quantum information, computation, and communication are promising branches of modern physics. New algorithms using the very quantum properties of our world can offer advantages to classical algorithms. One of the most important resources that is essentially missing in classical physics is the counterintuitive phenomenon of *quantum entanglement*.

The entanglement was first noticed as a paradox in the quantum physics by Albert Einstein in the famous EPR paper [3]. Entanglement is a mathematical consequence of axiomatic definition of quantum-world description: Consider a system composed of two subsystems. It is than described as a tensor product of corresponding Hilbert spaces. When choosing bases of the systems, we can find states that cannot be written as a product of some states of the subsystems. I.e. the state of the system cannot be viewed as a composition of substates but it is an indivisible entity.

As an example, consider two photons in superposition $\frac{1}{\sqrt{2}}(|\uparrow\uparrow\rangle+|\downarrow\downarrow\rangle)$. This means that both photons have the same polarisation but we do not know which one. When measuring one of the photons, the information about the other's polarisation is also obtained without any restriction on its position. So we seemingly may acquire the information about its state at instant, i.e. with the speed higher than the speed of light. However, this EPR paradox must be viewed in the following way. These two photons form a single entangled system. In the quantum information theory, entanglement is viewed as an amount of information that is shared between the particles. In our example, when measuring one photon, we obtain complete information about the other. For this reason, this example state is one of the maximally entangled states (so called Bell states) which are crucial for applications like quantum teleportation.

Quantum computation offers interesting effective algorithms which makes this branch very promising. However, one of the most important difficulties for successful quantum communication is still being struggled with. It is the environment that in principle cannot be rid of and causes state decoherence. Together with the state, its entanglement also decays. One way to fight this decoherence is the use of purification protocols - processes that can repair the state at the cost of sacrificing of some of its copies.

In this article we consider one particular protocol proposed by Bechmann-Pasquinucci et.al. and improved by Alber et.al. [2, 1]. It uses measurement-based state selection to modify density operators in a very simple nonlinear manner. The nonlinearity of the protocol induces exponential sensitivity to the input state. We aim to study action of the protocol on a pair of qubit intending to prove its purification capabilities. However, even such a simple system is too complicated to study the chaotic behaviour. The biggest complication is the lack of mathematical apparatus for multidimensional functions.

Previous papers focused on pure one-qubit states or some special subsets of pure two-qubit states. We aim to investigate the protocol action generally on mixed two-qubit states. At least two qubits are of course needed to study the evolution of entanglement. As we mentioned, this system is already too complicated to understand and describe arbitrary state evolution at this moment. Therefore, we use some observations, e.g. about the behaviour of product states and study in detail evolution of a mixed single-qubit state. This knowledge should help us to obtain insight into global dynamics for arbitrary initial states.

2 Theoretical background

2.1 Protocol iteration

The protocol [1] is physically realised in four steps: 1) the input states (all are the same) are divided into control and target qubits 2) XOR gate is then applied to the pair of control and target qubits 3) the control qubits are measured and when 1 is measured (in computational basis), the pair is discarded 4) the remaining qubit is modified using a twirling operator.

The protocol is based on the application of XOR gate and the measurement based selection which together form a nonlinear operator. The cost of this nonlinear operation is paid with loosing considerable amount of qubits which are projected to $|1\rangle$.

Mathematically, the action of the protocol has a simple form (in computational basis) as one protocol iteration squares density matrix elements and renormalises them. We denote this nonlinear operation with S. This works with the one-qubit states as well as the two-qubit states. The matrix dimensions will not be distinguished when using S; the action of S can also be written as an elementwise product of the matrix with itself.

This protocol can be enriched with a twirling operator, we use the Hadamard gate ($H = H^T = H^{-1}$ for one-qubit or $H \otimes H$ for two-qubit protocol). For a single qubit states, the result after one iteration of protocol is

$$\rho \to \rho' = HS(\rho)H \,, \tag{1}$$

$$S(\rho) = \rho \odot \rho = \begin{pmatrix} \rho_{11}^2 & \rho_{12}^2 \\ \rho_{21}^2 & \rho_{22}^2 \end{pmatrix} \; ; \quad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} . \tag{2}$$

2.2 Chaos description

Briefly, chaos is sensitivity to initial conditions. So, the system may evolve in a very different way even for the slightest disturbation of the original state. A mathematical theory, where chaos is well described concerns functions of a single complex variable,[6]. For such function f and initial state $z \in \mathrm{Dom}_f \subset \mathbb{C}$, f is sensitive in z when:

$$(\exists \varepsilon > 0)(\forall \delta > 0)(\exists y \in \mathbb{C})(|z - y| < \delta \wedge |f^n(z) - f^n(y)| \geq \varepsilon) \,. \tag{3}$$

Numbers z satisfying previous condition form so called Julia set of f. The rest of points from the domain of the function forms so called Fatou set. In Fatou set, there are attractive states where other states asymptotically converge to. Because the theory is very extensive, we refer to [6] where the reader can find much more details.

Since we will generalise previous results in this paper, let us briefly summarise findings from [4]: we can parameterise pure one-qubit states with $\frac{1}{1+|z|^2}(|0\rangle + z|1\rangle)$, $z \in \mathbb{C}$. This state evolves into $\frac{1}{1+|f(z)|^2}(|0\rangle + f(z)|1\rangle)$ with evolution function

$$f(z) = \frac{1 - z^2}{1 + z^2} \,. \tag{4}$$

This $f(z)$ is a rational polynomial function of degree 2. Therefore, its Julia set is nonempty and we can look for attractor cycles by checking the critical points $f'(z) = 0$.

This leads to a superattractive length two cycle $0 \leftrightarrow 1$. All considered states $z \in \mathbb{C}$ are divided into three sets - the Julia set with states that evolve chaotically, and two subsets of the Fatou set which converge to the superattractive cycle with different parity (converging to 0 in even or odd number of iterations).

Physical consequences for z from the Julia set (which in this case has empty interior) is that the corresponding states evolve (seemingly) randomly, chaotically. Therefore, the result cannot be predicted and is useless for quantum computation. On the other hand, the states corresponding to the points from Fatou set (almost all states) are purified towards the cycle $|0\rangle \leftrightarrow \frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$. If one of this state would be needed for computation purposes, the protocol works as we wish.

The same evolution function drives the behaviour in a subset of pure two-qubit states parametrised $\{\frac{1}{1+|z|^2}(|00\rangle + z|11\rangle)|z \in \mathbb{C}\}$. The superattractive cycle here contains the Bell state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ which is the reason to use the protocol to purify entanglement.

In the further text we consider evolution in more variables, namely in 3 real numbers characterising a mixed one-qubit state. There is almost no theoretical background for such multidimensional functions. Therefore, we try to simplify the situation finding some special subsets which allow at least numerical examination.

Formally, the evolution we study goes beyond the theory in [6]. We should not use terms like Julia set or attractive cycles. However, the dynamics seems to exhibit similar features so we stick to these terms in rather informal way for the purposes of this paper. We also should not speak of chaos since it is proven to be contained only on the Bloch sphere. At the moment, we cannot present analytical proof for the sensitivity in some states inside the ball, although it seems to be true based on numerical analysis.

3 Mixed one-qubit state evolution

While paper [4] studies the action of the protocol [1] and its modifications on pure single-qubit states and [5] studies the protocol acting on some of pure two-qubit states, we now present the study of the protocol action on a mixed single-qubit state. For the general applicability of the protocol, it is vital to understand its action on all states, not only the pure ones. Therefore, in this paper, we go beyond borders of formerly used theories.

All considered qubits form a ball - in Pauli matrix representation:

$$\rho = \frac{1}{2}\begin{pmatrix} 1+a & b+ic \\ b-ic & 1-a \end{pmatrix} \; ; \; a,b,c \in \mathbb{R} : a^2+b^2+c^2 \leq 1\,. \tag{5}$$

We would like to stress that these matrices are characterised by three real parameters. What is now essentially different is the presence of an additional dimension compared to the pure states $\frac{1}{1+|z|^2}(|0\rangle + z|1\rangle)$ - they depend on a complex parameter, i.e. two real numbers only. The pure states form the border of the sphere with $a^2 + b^2 + c^2 = 1$. The interior of the sphere contains mixed states, we also consider important to find whether and how the purity of the state changes during the evolution, for this purpose we can measure the purity with

$$\mathfrak{Pur}(\rho) = \mathrm{Tr}(\rho^2) = \frac{1+a^2+b^2+c^2}{2}\,. \tag{6}$$

After one protocol application, the new density operator is parameterised by

$$a' = \frac{b^2 - c^2}{1 + a^2}\,, \quad b' = \frac{2a}{1 + a^2}\,, \quad c' = \frac{-2bc}{1 + a^2}\,. \tag{7}$$

The dynamics is thus given by a vector function of three real variables. Later, we will give reason for why we have not succeeded in an attempt to write it as a (real-)parameter-dependent function of a complex variable, e.g. $f_a(b + ic)$. Such a form would allow us to perform analyses like in previous paragraph along a scale of the parameter.

Let us make notion on the symmetries that arise from the evolution formulas: density operators determined by $(a, b, c), (a, -b, -c), (-a, -b, -c)$ always happen to end in the same state after two iterations. This means that the asymptotic evolution is symmetrical with respect to the centre of the ball and also with respect to the plane $a = 0$.

To examine and visualize the evolution, we decided to slice the ball of the mixed states and numerically estimate the evolution. We remind that the numerical approach is the only one we have at hand. Instead of slicing the ball in planes $a =$ const. etc. we decided to stratify the ball into spheres for two reasons: the spheres are formed by states with the same purity; the dynamics should collapse into formerly studied dynamics of 4 for the border (Bloch) sphere.

So, the first parameter determining a state is its purity, which is the same for a sphere of states. This sphere is than projected via stereographic projection onto a plane with coordinates x, y. Thanks to the symmetries, it does not matter which pole we take for the projection, the resulting pictures are symmetric with respect to the centre and also with respect to the axes x, y. The parameterisations are connected:

$$\mathfrak{Pur} = \frac{1 + a^2 + b^2 + c^2}{2}\,, \quad x = \frac{b}{1 + a}\,, \quad y = \frac{c}{1 + a}\,. \tag{8}$$

To acquire the asymptotical evolution numerically, we construct a dense equidistant grid in a plane, which is then folded onto a sphere of a chosen purity using inverse map

$$a = \sqrt{2\mathfrak{Pur} - 1}\frac{1 - x^2 - y^2}{1 + x^2 + y^2}, \; b = \sqrt{2\mathfrak{Pur} - 1}\frac{2x}{1 + x^2 + y^2}, \; c = \sqrt{2\mathfrak{Pur} - 1}\frac{2y}{1 + x^2 + y^2}\,. \tag{9}$$

These numbers are then evolved using sufficient number of iterations 7 (usually 40).

After calculating the evolution on the grid of states (pixels of the pictures below), a colour is assigned to each initial state according to its asymptotical limit. From the computed behaviour we find reason for why we could not compose variables a, b, c or $\mathfrak{Pur}, x, y$ into a single complex number like for the pure states; the evolution tears and mixes the ball interior in a very complicated (chaotic) manner leaving no obvious invariant hyperplanes of complex dimension 1, the only possible exception being the Bloch sphere.

Let us mention now two special sets - the ball axes $b = c = 0$, $a = c = 0$. They are invariant on two iterations of the protocol; after one iteration, the axes are mapped into each other. Inside these sets, the evolution (for two protocol iterations) is determined by a single function of a real variable:

$$a \to f_a(a) = \left(\frac{2a}{1 + a^2}\right)^2, \text{ resp.} \quad b \to f_b(b) = \frac{2b^2}{1 + b^4}\,. \tag{10}$$

Checking these functions one can see that it is sufficient to examine only the evolution on positive semiaxes, $a, b > 0 = c$. There is a point on each axis which is a fixed state. These two fixed states together form a repulsive length two cycle, numerically they are

$$\rho_a = \frac{1}{2}\begin{pmatrix} 1 & 0.543689 \\ 0.543689 & 1 \end{pmatrix} \leftrightarrow \frac{1}{2}\begin{pmatrix} 1+0.295598 & 0 \\ 0 & 1-0.295598 \end{pmatrix} = \rho_b\,. \tag{11}$$

States closer to the centre of the ball converge to the centre, states further from the centre than the fixed points converge to the pure state $a = 1, b = c = 0$, resp. $b = 1, a = c = 0$. These two states together form an attractive cycle of the protocol. It is exactly the cycle $z = 0 \leftrightarrow z = 1$ mentioned in paragraph 2.2.

Back to all mixed states, asymptotic dynamics is naturally different compared to analysis in [4] not only because we study more-dimensional object. From numerical results, new attractor is found - the centre of the ball which is the completely mixed state. In total, there is one attractive (pure) cycle and one attractive fixed (mixed) state

$$\rho_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \leftrightarrow \rho_2 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \ ; \quad \rho_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\,. \tag{12}$$

Because of the length two cycle, we perform even number of iterations in our numerical calculations. In such case, they say that from all the states, ca. 48.66% converge to ρ_1, 23.64% to ρ_2 which means 72.3% of all states converge to the pure cycle. (In case of odd number of iterations, the areas of convergence to ρ_1, ρ_2 swap.) Ca. 27.42% of states was estimated to converge to the mixed attractor ρ_3. The rest corresponds to the states that behave chaotically or converge too slowly to be successfully assigned to any attractor. Based on the theory in [6], we consider very probable that the 'Julia set' has empty interior and fractal dimension $\in (2, 3)$.

From the picture below, one can see that chaotic features are present even inside the ball. However, the fractal patterns decay with the decreasing purity and for low-purity states we obtain simple shapes (see the case of initial purity 0.65). The border of the white/grey/black regions which we suppose to form the 'Julia set' seems to satisfy some simple polynomial function. This is possible for the Julia set, remember function $f(z) = z^2$ with the unity circle forming the Julia set.

However, in our case we cannot guarantee that the points from the 'Julia set' satisfy the condition 3, which is the core of the Julia set definition. We remind that all presented findings are based on numerical analysis. Although we support them with very careful and detailed observations, we cannot exclude the existence of some degenerate cases violating the definition of attractiveness/repulsiveness of the states.

4 Mixed two-qubit states evolution

To purify entanglement one of course has to have at least two qubits. One of the reasons to investigate the protocol action on a single qubit is hidden in following observation. In words, *separable states form an invariant set.*

$$(H \otimes H)S(\rho_1 \otimes \rho_2)(H \otimes H) = (HS(\rho_1)H) \otimes (HS(\rho_2)H) \tag{13}$$

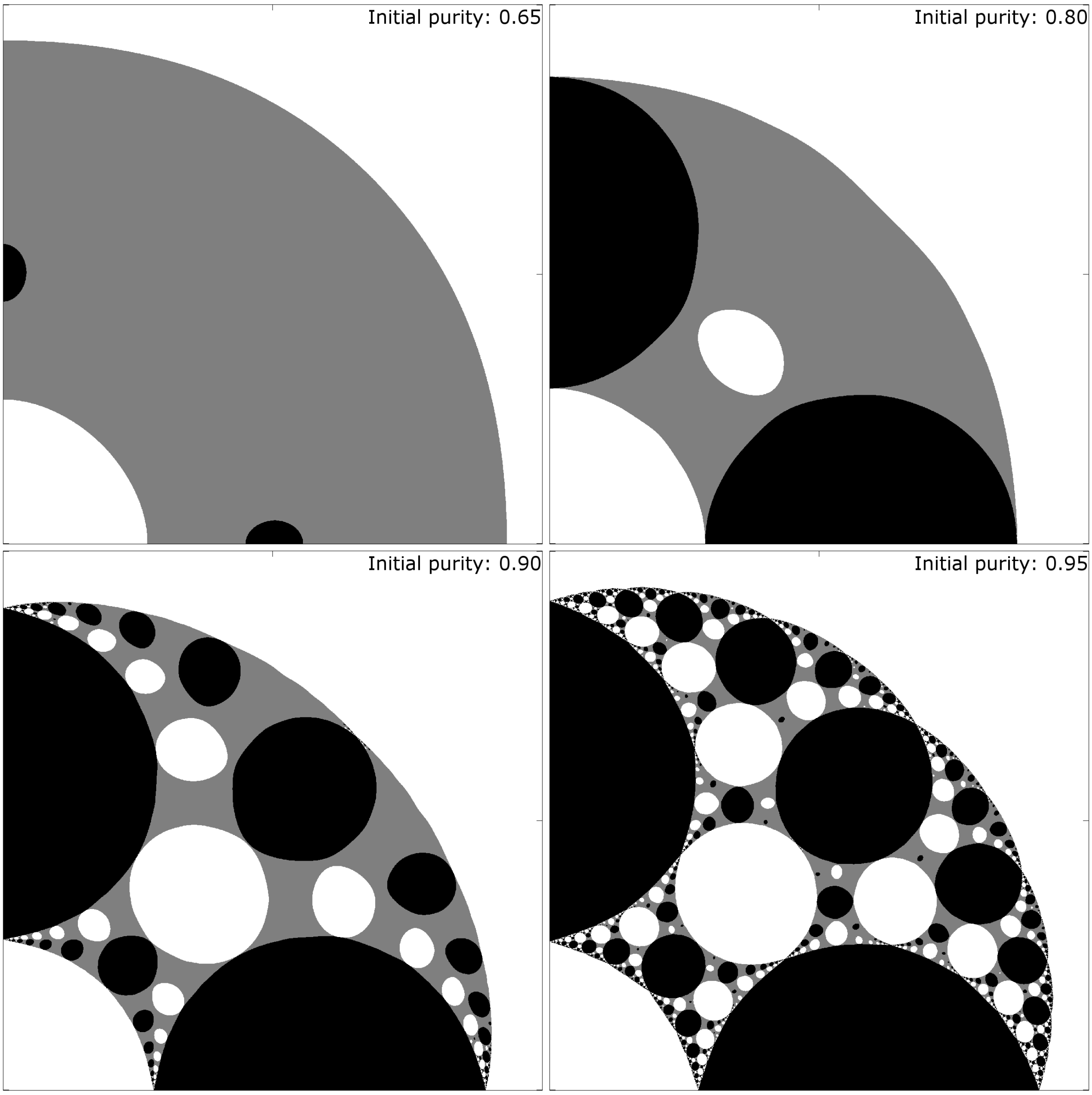


Figure: Convergence on the spheres of equal initial purity - cuts represent areas of $x, y \in (0,2) \times (0,2)$. For even number of iterations, white colour represents convergence to ρ_1, black to ρ_2. Grey colour stands for the states converging to the mixed attractor ρ_3.

It can also be written as

$$(H \otimes H)[(\rho_1 \otimes \rho_2) \odot (\rho_1 \otimes \rho_2)](H \otimes H) = [H(\rho_1 \odot \rho_1)H] \otimes [H(\rho_2 \odot \rho_2)H]\,. \tag{14}$$

This results into very simple dynamics factorisation - the subsystems are independent. *The dynamics of the product states is a cartesian product of the one-qubit dynamics.*

Thanks to this, we find all attractive cycles in the product states at instant, no other separable attractor can exist. The attractive cycles are formed by:

$$\rho_i \otimes \rho_j\,, \quad i, j \in \{1, 2, 3\} \tag{15}$$

The attractiveness of these states is of course meant in following way:

$$(\forall i, j \in \{1,2,3\})(\exists \mathcal{U}_{i,j} \text{neighbourhoods of } \rho_{i,j})(\forall \sigma_{i,j} \in \mathcal{U}_{i,j})(\sigma_i \otimes \sigma_j \xrightarrow{asympt.} \rho_i \otimes \rho_j)\,. \tag{16}$$

We cannot argue that these states also attract other states from the nonseparable states.

Consider now the cycle $\rho_3 \otimes \rho_a \leftrightarrow \rho_3 \otimes \rho_b$. This cycle is attractive in the first subsystem but repulsive in the other subsystem, which is the behaviour known for saddle points in the differential equations theory. Furthermore, e.g. cycle $\rho_a \otimes \rho_a \leftrightarrow \rho_a \otimes \rho_a$ is repulsive in both subsystems. We believe this effect when a state is attractive for some perturbation while repulsive for another perturbation is present in the evolution of general density matrices. Even more complicated behaviour might happen. That is exactly what makes the investigation of general state evolution so difficult - it is unclear, what more types of chaos may emerge. We mention that from the aspect of the theory of chaos, both examples are chaotic states and would belong to 'Julia set'. However, we believe it is important that the nature of the chaos is different depending on the direction in multiple dimensions allowing behaviour of Fatou set points as well as the chaotic features.

Surprisingly, the product states are not the only invariant set. At last, we present the most important set of states we have found. It is invariant on two iterations of the protocol and contains mixed, entangled states - it is the generalisation of the set of pure states $\left\{\frac{1}{1+|z|^2}(1,0,0,z)\right\}$ from [5].

$$\frac{1}{1+|z|^2}\begin{pmatrix} 1 & 0 & 0 & pz \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ pz & 0 & 0 & z \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1+a & 0 & 0 & b-ic \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ b+ic & 0 & 0 & 1-a \end{pmatrix} \tag{17}$$

The latter parameterisation is again based on use of Pauli matrices. We use it because after even number of protocol iterations, these parameters evolve identically with the evolution of a, b, c in mixed one-qubit states (after the same number of iterations, of course). Therefore, the evolution in this set of states is already examined with the results discussed in the previous paragraph. Only now, the attractors $a = 1\,,\, b = c = 0\,;\, a = c = 0\,,\, b = 1\,;\, a = b = c = 0$ represent another states: the pure states - the Bell state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and separable $|00\rangle$, the third attractive state is mixed with $\mathfrak{Pur} = 1/2$ and unfortunately contains no entanglement.

We conclude that when even number of iterations is performed, ca. 23.64% of states 17 converge to the Bell state. Moreover, the area of convergence seems to form quite a big neighbourhood of the Bell state. This makes the protocol very useful for the purification. Although we lack the analysis of the attractiveness in a general neighbourhood of the Bell state, each new dimension is an important step to understand the multidimensional chaos.

The attractiveness of the Bell state cycle still cannot be guaranteed concerning general evolution. It still can be some degenerate case like the attractiveness in the sense mentioned with the cycle $\rho_3 \otimes \rho_{a,b}$. However, this result extending the attractiveness to another dimension is still very important because of the complicated nature of the multidimensional nonlinear dynamics.

5 Tribonacci constant in entanglement purification

Golden ratio is a famous number known for its overextensive appearance in Nature. It is the only real root of polynomial $x^2 - x - 1$ and it also is a limit of two consecutive members of the famous Fibonacci sequence. We can construct so called tribonacci sequence [7]:

$$T_0 = T_1 = 0, T_2 = 1\,; \quad \forall n \in \mathbb{N} : T_{n+2} = T_{n+1} + T_n + T_{n-1} \tag{18}$$

The limit of two consecutive numbers exists and is called *tribonacci constant*$= \tau$; it can be shown that this number is the only real root of $x^3 - x^2 - x - 1$. This number can be found in our protocol: Pure two-qubit states $\frac{1}{1+|z|^2}(1, 0, 0, z)$ were found to be subject to evolution function 4, [5]. Looking for fixed states, i.e. such z that $f(z) = z$, we get equation $z^3 + z^2 + z - 1 = 0$. Putting $\frac{1}{z}$ instead of z into the polynomial and multiplying it with z^3, we get exactly the polynomial for tribonacci constant. Therefore, the states with z equal to τ^{-1} or the reciprocal value of the conjugated roots are the fixed states.

Nevertheless, this is not the only role of the τ. We mentioned a repulsive length-two cycle of mixed one-qubit states 11, we now see it is

$$\rho_a = \frac{1}{2}\begin{pmatrix} 1 & \tau^{-1} \\ \tau^{-1} & 1 \end{pmatrix} \leftrightarrow \frac{1}{2}\begin{pmatrix} 1 + \tau_*^{-1} & 0 \\ 0 & 1 - \tau_*^{-1} \end{pmatrix} = \begin{pmatrix} \tau_\times & 0 \\ 0 & 1 - \tau_\times \end{pmatrix} = \rho_b\,. \tag{19}$$

Newly appearing real numbers $\tau_*, \tau_\times$ have been found numerically; they have following properties: Solving $x^3 - 3x^2 - x - 1 = 0$ we obtain a root τ_* and the other two roots have the real part of their reciprocal values equal to $-\tau_\times$. They also appear in the fixed pure two-qubit states, e.g. $\frac{1}{N}(1, \tau^{-1}, \tau^{-1}, \tau_*^{-1})$. Computing N we can check $\tau_* = \tau^2$ which has its hidden reason in the evolution formulas. We finish our numerical observations with the fact that the minimal polynomials for the τ^{-1}, τ^{-2} are similar to the minimal polynomials of τ, τ^2. They are $x^3 + x^2 + x - 1$, $x^3 + x^2 + 3x - 1$ respectively.

Based on these observations and further numerical calculations we have found that all one-qubit states with purity smaller than τ^{-1} converge to ρ_3 while there is a state with purity τ^{-1} that does not. Also, pure two-qubit states $(1, 0, 0, z)$ with $|z| < \tau^{-1}$ must converge to $z = 0$ after even number of iterations not being so for $|z| \geq \tau^{-1}$.

All these very fortunate numerical findings are very surprising and we suggest deeper studies of this circumstances. For example, we set few questions: What is the reason for the mentioned polynomial $x^3 - 3x^2 - x - 1 = 0$? What is exact connection between polynomials $x^3 + x^2 + 3x - 1$, $x^3 + x^2 + x - 1$ (determining coefficients in fixed states) and functions $\frac{1-z^2}{1+z^2}, \frac{1-z^2}{1+3z^2}$ (found to determine behaviour in pure two-qubit states)? Is there an other useful relation between general pure two-qubit states, mixed one-qubit states and some two-qubit mixed states?

6 Conclusion

After brief introduction, we have described the evolution in the set of mixed one-qubit states. We have used numerical calculation to estimate the nature of chaotic behaviour that is present inside the ball of the considered states. A new attractor has been found - the completely mixed state. Chaotic features are present in the mixed state dynamics, although the fractal patterns disappear with the lower purity.

In the product states, the dynamics of the subsystems is separated and so the evolution is simply a cartesian product of the one-qubit dynamices. This set has no use for purification as all attractors are separable, thus with no entanglement.

Thanks to the found relation between mixed one-qubit states and a particular set of mixed two-qubit states, we have also described the evolution inside this set. The set contains entangled states and so we get to the original aim of our protocol - does it purify the entanglement? The formerly known cycle containing the Bell state has been shown to attract much more states than before. Indeed, mixed two-qubit states can also be purified to the Bell state. We conclude the protocol purifies the entanglement in much more states than previously shown.

Although we have numerically shown that most states of form 17 converge to some attractive cycles, we cannot give any analytical proof for the behaviour in the 'Julia set'. Numerical estimates support the chaotic behaviour but we have to realise that the finite precision of the computation inherently makes the chaos calculations imprecise. We consider important to change our attitude to chaos in the 15-dimensional space of mixed two-qubit states as some states/cycles may be attractive in some directions but repulsive in others. Although such states are considered repulsive in total, we believe this type of chaos should be treated differently.

An interesting observation has been made revealing that the tribonacci constant is deeply connected to the special states being purified by the protocol. While some of the reasons for the appearance of this number have been explained, we still feel that the connection is much more involved.

References

[1] G. Alber, A. Delgado, N. Gisin, I. Jex. J. Phys. A: Math. Gen. **34** (2001), 8821–8833

[2] H. Bechmann-Pasquinucci et. al. Phys. Lett. A **242** (1998), 198–204

[3] A. Einstein, B. Podolsky, N. Rosen. Phys. Rev. **47** (1935), 777

[4] T. Kiss, I. Jex, G. Alber, S. Vymětal. Phys. Rev. A **74** (2006), 040301(R).

[5] T. Kiss, S. Vymětal, L.D. Tóth, A. Gábris, I. Jex, G. Alber. Physical Review Letters **107** (2011), 100501.

[6] J.W. Milnor. *Dynamics in One Complex Variable*, 3^{rd} edition. Princeton University Press (2000).

[7] J. Sharp. Mathematical Gazette **82** (1998), 203–214.

Modified Definition of Coined Quantum Walks for Arbitrary Graphs

Jan Mareš

3rd year of PGS, email: `maresj23@fjfi.cvut.cz`
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:
Igor Jex, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague
Jaroslav Novotný, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. Quantum walks have established their place in quantum information sciences and many of their aspects have been studied. Here we present an alternative definition of a quantum walk, which is more general and contains the standard definition. We modify both the Hilbert space of the system and the time evolution.

An essential part of a the time evolution of a quantum walk is the shift operation S moving the walker among vertices of a graph that the walk is defined on. In simple cases like line graphs or square lattices, there is a natural shift operator. Nevertheless, there are always multiple options and this becomes much more apparent for more complex graph. Our approach allows for convenient classification of possible shift operations, which we consider very relevant when in we want to use quantum walks for simulation of physical systems.

The formalism itself and its benefits are demonstrated by investigating quantum walks with dynamical bond percolation on honeycomb graphs. The results show a crucial dependence of the resulting asymptotic state of the quantum walk on the choice of the shift operation S.

Keywords: quantum walks, definition, arbitrary graphs, honeycomb lattice

Abstrakt. Kvantové procházky mají své pevné místo v kvantové informatice a mnohé jejich aspekty byly intenzivně zkoumány. Zde prezentujeme alternativní definici kvantové procházky, která je obecnější a zahrnuje v sobě definici standardní. Nová definice upravuje jak Hilbertův prostor systému, tak jeho časový vývoj.

Zásadní součástí časového vývoje kvantové procházky je "operátor přesunu" S, které přesunuje chodce mezi vrcholy grafu, na kterém je kvantová procházka definována. V jednoduchých, jako jsou přímkové grafy nebo čtvercové mřížky, existuje přirozená volba operátoru S. Nicméně vždy existuje řada možností, jak tento operátor zvolit, což je daleko zřejmější u složitějších grafů. Náš přístup umožňuje pohodlnou klasifikaci možných operátorů přesunu, což považujeme za zásadní, pokud chceme používat kvantové procházky k simulacím fyzikálních systémů.

Samotný formalismus a jeho výhody jsou demonstrovány při zkoumání kvantových procházek s dynamickou perkolací hran na šestiúhelníkových mřížkách. Výsledky ukazují zásadní závislost výsledného asymptotického stavu kvantové procházky na volbě operátoru S.

Klíčová slova: kvantové procházky, definice, libovolné grafy, šestiúhelníková mřížka

1 Introduction

Quantum walks have established their place in quantum information sciences. Apart from existence of specific quantum walk algorithms, they can serve as a universal quantum computer. Quantum walks have already been realised experimentally in various physical systems, which also demonstrates their usability for quantum systems simulations.

The basis of a quantum walk is an undirected graph. There are some simple cases of quantum walks with well established definitions like probably the simplest one: a quantum walk on a line graph. Nevertheless, when we move to some more complex graphs or more general situations, it may be unclear, how to define the quantum walk or which one of the possible variants to choose. Examples of such problematic situations may be: graphs that are not regular (finite cuts of lattices), graphs with various directions of edges in different vertices (honeycomb lattice), dynamical changes of the graph in different steps of the walk (dynamical percolation of edges) and so on.

We present a formalism for defining quantum walks in all cases described above and many others. This formalism allows to address the ambiguity of the definition and therefore allows to either examine multiple options or choose the appropriate one.

2 Perfect (Non-Percolated) Quantum Walk

Let us have an arbitrary undirected graph $G(V,E)$ with the set of vertices V and the set of edges E, both of which are at most countably infinite. We call G the structure graph of the quantum walk. In general, the structure graph is not assumed to be simple neither connected.

2.1 The Hilbert Space

The Hilbert space $\mathcal{H}$ of our quantum walk is spanned by states corresponding to directed edges of a directed graph $G^{(d)}(V,E^{(d)})$, which we will call the state graph. The state graph $G^{(d)}(V,E^{(d)})$ has the same set of vertices V as the structure graph $G(V,E)$ and its set of directed edges $E^{(d)}$ consists of two subsets: $E^{(d)} = E_p^{(d)} \cup E_u^{(d)}$. Edges from the fist subset $E_p^{(d)}$ will be called paired and are derived from the structure graph $G(V,E)$. For every undirected edge $e \in E$ we have two directed edges $e_1^{(d)}, e_2^{(d)} \in E_S^{(d)}$ going in opposite directions and connecting the same two vertices as e. Corresponding to these paired edges we have paired states $|e_1^{(d)}\rangle, |e_2^{(d)}\rangle$. Edges in the other subset $E_u^{(d)}$ are called unpaired and are independent of $G(V,E)$. Unpaired edges are loops – edges beginning and ending in the same vertex. Adding loops allows us to arbitrarily increase degrees in vertices of our choice. (There may be some other loops originating from loops in the structure graph $G(V,E)$, but those are still paired edges.) Overall, for every directed edge $e^{(d)} \in E^{(d)}$, there is a base state $|e^{(d)}\rangle$ in $\mathcal{H}$. The state $|e^{(d)}\rangle$ represents a walker standing in the initial vertex of $e^{(d)}$ facing the direction of the terminal vertex of $e^{(d)}$.

Traditionally, coined quantum walks are described by a position Hilbert space $\mathcal{H}_p$ represented by vertices of an undirected graph and so called coin Hilbert space $\mathcal{H}_c$ representing some internal degree of freedom of the walker. The overall Hilbert space is then $\mathcal{H} = \mathcal{H}_p \otimes \mathcal{H}_c$ [5]. Often, the graph G is regular and the dimension of $\mathcal{H}_c$ corresponds to the degree of a vertex, in which case the correspondence is straightforward. Sometimes, the degree of the vertex is lower than the dimension of the coin space. It may be for example on borders of finite graphs [2] or if there is some state representing no movement of the walker [4]. For those cases, we have the possibility to introduce loops in our state graph $G^{(d)}(V, E^{(d)})$ that appropriately increase the dimension of the Hilbert space.

Even though the Hilbert space $\mathcal{H}$ of a quantum walk according to our definition does not have to be of the tensor product form $\mathcal{H} = \mathcal{H}_p \otimes \mathcal{H}_c$, it can always be written as a direct sum of vertex subspaces: $\mathcal{H} = \bigoplus_{v\in V} \mathcal{H}_v$, where $\mathcal{H}_v$ is a subspace spanned by states corresponding to edges originating in $v \in V$.

2.2 The Time Evolution

Here, we only deal with discrete-time quantum walks. One step of the evolution from time t to $t+1$ is realised by application of a unitary evolution operator U:

$$|\psi(t+1)\rangle = U\,|\psi(t)\rangle = U^{t+1}\,|\psi(0)\rangle\,.$$

In analogy with classical random walks, the evolution operator of a quantum walk is a product of two unitary operators: $U = CS$. The operator C is called the coin operator and S is called the shift operator. The coin operator C represents some local unitary evolution in vertices (subspaces $\mathcal{H}_v$ for $v \in V$ are invariant under the operation C). Therefore C can be represented by a block-diagonal matrix and can be written as $C = \bigoplus_{v\in V} C_v$, where C_v is the operator of the action of C restricted to the subspace $\mathcal{H}_v$ for $v \in V$. In the special (but common) case of a regular graph and one coin C_v in all vertices, the coin operator is a tensor product $C = I_p \oplus C_v$, where I_p is the identity on a Hilbert space spanned by states corresponding to vertices of the structure graph.

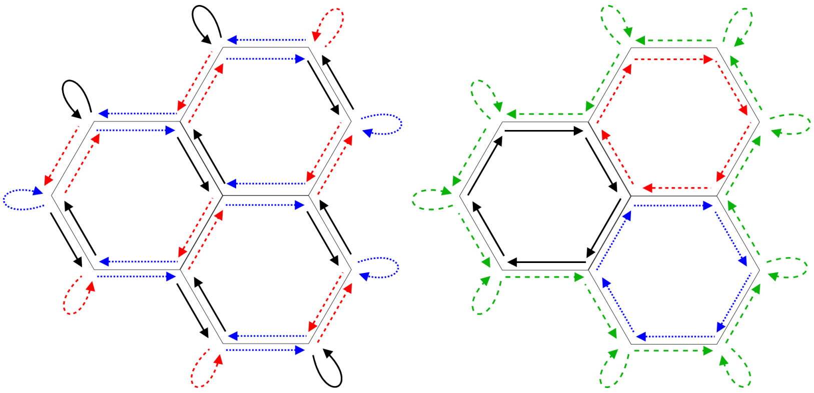
The shift operator S realises the movement of the walker among different vertices. The shift operator has to respect the structure of the state graph $G^{(d)}(V, E^{(d)})$. In particular, from an edge $e^{(d)} \in E^{(d)}$ going from $v_1 \in V$ to $v_2 \in V$, the walker moves to some edge beginning in v_2. There is one canonical way of defining the shift operator. We will denote this particular shift operator by R and refer to it as a reflecting shift operator. The action of R is defined as follows: If we have an undirected edge $e \in E$ with two corresponding directed paired edges $e_1^{(d)}, e_2^{(d)} \in E_p^{(d)}$, then $R\,|e_1^{(d)}\rangle = |e_2^{(d)}\rangle$ and $R\,|e_2^{(d)}\rangle = |e_1^{(d)}\rangle$. Any unpaired state $|l\rangle$ for $l \in E_u^{(d)}$ is mapped to itself, so $R\,|l\rangle = |l\rangle$. Obviously, such shift operator is defined on any graph in consideration. An example for a finite honeycomb lattice is given in figure 1(a). It is good to note that the operator R is its own inversion and therefore is also Hermitian ($R = R^{-1} = R^{\dagger}$).

Any other shift operator S can be composed of the action of the reflecting operator R and some local permutation P (P only permutes states locally in vertices.) Therefore, we can write the evolution operator as:

$$U = CS = CPR.$$

3 Percolated Quantum Walk

Let us now describe what we call a bond percolation in quantum walks – scenarios, in which some edges of the original structure graph $G(V, E)$ are broken (closed/missing).



(a) The reflecting shift operation. (b) The cycling shift operation.

Figure 1: Examples of the actions of two different shift operators S on a honeycomb graph. Colours and line types indicate the action S.

The generic way of introducing percolation is to choose some probability $p \in (0, 1)$ and make every edge open with the probability p and closed with the probability $1 - p$. A single realisation of this process gives rise to some percolation graph $G_K(V, K)$, where only the edges in $K \subset E$ remained open.

The modification of a quantum walk for the percolated version is very simple in our formalism. Just by choosing some configuration of the percolation graph, i.e. choosing a subset of open edges $K \subset E$, the whole dynamics of a percolated quantum walk is determined. For a given percolation structure graph $G_K(V, K)$ we just modify the state graph $G^{(d)}(V, E^{(d)})$ into $G_K^{(d)}(V, K^{(d)})$. Simply, if some edge $e \in E$ is broken ($e \notin K$), we replace the two corresponding directed edges $e_1^{(d)}, e_2^{(d)} \in E^{(d)}$ with two loops. These two edges are still paired, but we will now call them broken paired edges. Broken edges result in a natural change of the reflecting shift operator R to the operator R_K. The difference is that states corresponding to broken paired edges are also mapped to themselves in R_K, as it is for unpaired edges.

The coin operation C is not altered by the percolation at all and also the local permutation P stays the same.

4 Asymptotic States of Quantum Walks with Dynamical Percolation on Finite Graphs

From now on, a quantum walk is assumed to have a finite state graph $G^{(d)}(V, E^{(d)})$ with finite number of vertices $\#V$ and finite number of edges $\#E^{(d)}$. The probability of occurrence π_K for a configuration K with $\#K$ open edges is $\pi_K = p^{\#K}(1-p)^{\#E-\#K}$, where $\#E$ is obviously the number of edges in the structure graph $G(V, E)$.

The term dynamical percolation refers to the situation, in which a new percolation graph $G_K(V, K)$ is generated for every step of the quantum walk. Since such percolation introduces a classical uncertainty to the evolution of the quantum walk, we use the description of the state by a density matrix. We have to take into account all possible configurations of the percolation graph and therefore one step of the walk must be described as:

$$\rho(t+1) = \sum_{K\subset E} \pi_K U_K \rho(t) U_K^\dagger,$$

where U_K is the evolution operator with the modified reflecting shift operator R_K corresponding to the particular structure graph $G_K(V,K)$ for $K \subset E$.

This kind of time evolution is referred to as a random unitary operation. A procedure for determining the asymptotic behaviour of a system governed by this kind of evolution has been suggested in [3]. The asymptotic regime of such system is determined by so called attractors – solutions of the set of equations:

$$U_K X_\lambda U_K^\dagger = \lambda X_\lambda\,, \quad \text{for all } K \in 2^E, \tag{1}$$

for some given λ fulfilling $|\lambda| = 1$.

The asymptotic state (the limit for infinitely many steps) of a percolated quantum walk is than given as [3]:

$$\rho_{t\to\infty}(t) = \sum_{\lambda,i} \lambda^t \mathrm{Tr}\Big(\rho(0) X_{\lambda,i}^\dagger\Big) X_{\lambda,i},$$

where i distinguishes different attractors for the eigenvalue λ in the orthonormal basis of the solutions of (1) and $\rho(0)$ is the initial state of the quantum walk.
We can now use the procedure described in [1] using the special structure of U_K. It is possible to rewrite the set of equations (1) into:

$$R_K X R_K^\dagger = \lambda (CP)^\dagger X (CP)\,, \quad \text{for all } K \subset 2^E. \tag{2}$$

The right-hand side is independent of the actual configuration of the percolation graph K. Therefore, we can solve this set of equations in two steps. First we choose $K = \emptyset$ (the configuration with all edges closed). In that case we have $R_\emptyset = I$, where I is the identity operator. The equation (2) becomes:

$$CPX(CP)^\dagger = \lambda X. \tag{3}$$

The operators C and P do not mix states in different vertex subspaces $\mathcal{H}_v$, so the matrix CP is block-diagonal (assuming appropriate ordering of the basis). We can then split the attractor matrix X into blocks $X_{v_2}^{v_1}$ corresponding to pairs of vertices $v_1, v_2 \in V$ and solve the equation (3) locally:

$$(C_{v_1} P_{v_1}) X_{v_2}^{v_1} (C_{v_2} P_{v_2})^\dagger = \lambda X_{v_2}^{v_1}, \tag{4}$$

where C_{v_1}, C_{v_2} and P_{v_1}, P_{v_2} are blocks of operators C and P respectively acting on subspaces $\mathcal{H}_{v_1}, \mathcal{H}_{v_v}$ for vertices $v_1, v_2 \in V$. Since the right-hand side of the set of equations 2 is the same for all configurations, the values of the left-hand sides must be mutually equal for all configurations, so we obtain:

$$R_K X R_K^\dagger = R_L X R_L^\dagger\,, \quad \text{for all } K, L \in 2^E. \tag{5}$$

We call this a shift condition and it gives the restrictions for binding the blocks $X_{v_2}^{v_1}$ together into one attractor matrix X.

4.1 Pure Eigenstates Ansatz

A method for finding attractors using pure common eigenstates of all unitary operators U_K has been proposed in [2]. If we have common eigenstates $\{|\phi_{\alpha,i}\rangle\}_{\alpha,i}$ fulfilling

$$U_K\,|\phi_{\alpha,i}\rangle = \alpha\,|\phi_{\alpha,i}\rangle\ , \quad \text{for all } K \subset 2^E, \tag{6}$$

corresponding to eigenvalue α (i distinguishes different common eigenstates corresponding to α) then

$$Y_\lambda = \sum_{\alpha\beta^*=\lambda} A^{\alpha,i}_{\beta,j}\,|\phi_{\alpha,i}\rangle\,\langle\phi_{\beta,j}|$$

is an attractor corresponding to the superoperator eigenvalue $\lambda = \alpha\beta^*$. We will call attractors of this type p-attractors.
All p-attractors trivially fulfil a larger set of equations

$$U_K Y_{\lambda,i} U_L^\dagger = \lambda Y_{\lambda,i}. \tag{7}$$

When compared to (1), the operators U_K and U_L can be different here. A non-trivial result [2] is that any solution of this set of equations is a p-attractor.
The set of equations (6) can be rewritten as:

$$R_K\,|\phi_{\alpha,i}\rangle = \alpha(CP)^\dagger\,|\phi_{\alpha,i}\rangle \quad \text{for all } K \subset 2^E,$$

with only the left-hand side being dependent on K. We again start by solving the equation for the empty configuration $K = \emptyset$:

$$CP\,|\phi_{\alpha,i}\rangle = \alpha\,|\phi_{\alpha,i}\rangle \tag{8}$$

and then apply the shift condition:

$$R_K\,|\phi_{\alpha,i}\rangle = R_L\,|\phi_{\alpha,i}\rangle\ , \quad \text{for all } K, L \in 2^E. \tag{9}$$

4.2 The Shift Condition

Let us have a closer look at the shift condition for p-attractors (9) and for general attractors (5). We introduce the following notation: for a paired edge $i \in E_p^{(d)}$, let $\tilde{i}$ be the other member of the pair and for an unpaired edge $i \in E_u^{(d)}$ let $\tilde{i} = i$. First note that $\tilde{\tilde{i}} = i$ in all cases. Further note that this notation is not related to any configuration of the percolation graph K. The relation to the shift operator is that for the graph with all edges open we have $R_E\,|i\rangle = |\tilde{i}\rangle$.

Let us chose two configurations of the percolation graph K and L. The operator R_K ($R_K = R_K^\dagger$) can be written as: $R_K = \sum_{i\in E^{(d)}} |k(i)\rangle\,\langle i| = \sum_{i\in E^{(d)}} |i\rangle\,\langle k(i)|$, where we use k as a map acting on directed paired edges as $k(i) = \tilde{i}$ if i is open in K and $k(i) = i$ otherwise and $k(i) = i = \tilde{i}$ if i is an unpaired edge. Similarly we use a map l for the configuration L.

If we write $|\phi\rangle$ as $|\phi\rangle = \sum_{j\in E^{(d)}} \phi_j \,|j\rangle$ then the shift condition for p-attractors (9) is $\phi_{k(i)} = \phi_{l(i)}\,,$ for all $i \in E^{(d)},\ K, L \in 2^E$, which can be elegantly summarised independently of percolation graph configurations as

$$\phi_i = \phi_{\tilde{i}}\,, \quad \text{for all } i \in E^{(d)}. \tag{10}$$

The equality (10) is trivial for unpaired edges and for every paired edge we just need one configuration K with the corresponding undirected edge open and one configuration L with this edge closed.

A similar argument results in the shift condition for general attractors in the form

$$X^{k(i)}_{k(j)} = X^{l(i)}_{l(j)}\,, \quad \text{for all } \ i, j \in E^{(d)}, K, L \in 2^E.$$

In the case $i \neq j$ and $i \neq \tilde{j}$ we have the strongest condition:

$$X^i_j = X^{\tilde{i}}_j = X^i_{\tilde{j}} = X^{\tilde{i}}_{\tilde{j}}, \tag{11}$$

because there are configurations with both relevant edges open, both closed and both configuration with just one edge open. (Alternatively, i or j may be unpaired edges, in which case the equality is trivial.) For $i = j$ and $j = \tilde{i}$ the conditions are only

$$X^i_i = X^{\tilde{i}}_{\tilde{i}}, \quad X^i_{\tilde{i}} = X^{\tilde{i}}_i.$$

A crucial point in practical search for the set of attractors will be the difference of the shift condition for the p-attractors and the general attractors. The general attractors may differ from p-attractors only in the shift condition, because since the operator CP is unitary, the matrix solution of (3) is just a combination of solutions of (8).

The only cases where the shift conditions may differ are $i = j$ and $i = \tilde{j}$. Because the configurations K and L may differ in the p-attractor equation (7), the p-attractors must also fulfil the equality:

$$X^i_i = X^i_{\tilde{i}}. \tag{12}$$

Therefore, every non-p-attractor must violate (12) in at least one case. If we show that all diagonal elements corresponding to paired edges of an attractor are equal to the non-diagonal ones from (12), that attractor is a p-attractor.

5 Percolated Grover Quantum Walk on a Honeycomb Lattice

Let us now focus on a particular example of our general quantum walk. As the structure graph $G(V, E)$ we take finite honeycomb lattices of various shapes with loops added in border vertices. Therefore, the Hilbert space $\mathcal{H}_v$ in every vertex is three-dimensional and we may use the coin operation:

$$C_v \equiv G_3 = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$$

in every vertex $v \in V$. For simplicity of the argument, we will assume that the structure graph is always connected.

Let us introduce a notation for directions in our graph so that we may denote any directed edge by a vertex symbol and a symbol for direction. Let us place the graph so that some edges are horizontal. All horizontal edges will be denoted by direction H. The edges going between top-left corner and bottom-right corner will be denoted D ("diagonal") and the remaining ones A ("anti-diagonal"). The computational basis is chosen in the order H, A, D.

We will be considering two different shift operators S (or two different local permutations P) for our Grover walk on honeycomb lattices.

Reflecting Shift Operator

The first one will be called a reflecting walk, where $P = I$ and we have the situation shown in figure 1(a). This shift operator is very local - without the action of the coin operator, the walker would not leave one edge.

When searching for p-attractors, we simply solve the eigenvalue problem for the coin (8) in a particular vertex: $G_3 |\phi_v\rangle = \alpha |\phi_v\rangle$. There are only two eigenvalues of the Grover matrix and those are 1 with an eigenvector $|\phi_v^1\rangle = [1, 1, 1]^T$ and -1, where eigenvectors form a plane orthogonal to the eigenvector $|\phi_v^1\rangle$.

The shift condition for p-attractors is

$$\phi_{v_1,\delta} = \phi_{v_2,\delta}, \tag{13}$$

where $v_1, v_2 \in V$, v_1 and v_2 are connected by an edge and δ is the direction of the connecting edge (δ is H, A or D). Finding common eigenvectors for a whole graph for the eigenvalue 1 is trivial - all components of the vector must be equal.

There are two rules for constructing the common eigenstates for the eigenvalue -1. First, the sum of elements in a given vertex must be equal to zero (form of the eigenvector space). Second, the two elements corresponding to the same edge $e \in E$ must be the same (the shift condition). Every vertex in V brings two free parameters and every edge in E poses one restriction. Since all equations are clearly independent, we will have $N \equiv 2\#V - \#E$ common eigenvectors on the whole graph. There are many ways how a (non-orthogonal) basis of these whole-graph common eigenstates can be chosen.

We have found one common eigenstate corresponding to the eigenvalue 1 and N common eigenstates corresponding to -1. That results in $2 \times N$ p-attractors corresponding to the eigenvalue -1 and $N^2 + 1$ p-attractors corresponding to 1. Let us now search for the remaining non-p-attractors. We start with the equation (4):

$$G_3 \Xi G_3^\dagger = \lambda \Xi,$$

where Ξ represents general form of blocks X_v^u $(u, v \in V)$ of the whole attractor X.

For the eigenvalue -1, the general form of the one-vertex block is:

$$\Xi = \begin{bmatrix} \alpha & \beta & \alpha + \beta + \gamma \\ -\beta - \gamma + \delta & -\alpha - \gamma + \delta & \delta \\ -\beta - \delta & -\alpha - \delta & \gamma - \delta \end{bmatrix}. \tag{14}$$

We will show that there are no non-p-attractors for the eigenvalue -1. We will use the fact that a non-p-attractor has to violate some of the equations (12). This also means that we only have to care about matrix blocks X_u^u for all $u \in V$ and their relations to blocks X_u^v, X_v^u and X_v^v, where $v \in V$ and u are connected by some undirected edge. Let us assume that vertices u and v are connected by a horizontal edge. (Due to the symmetry, the result for diagonal and anti-diagonal edges will be similar.) For clarity let us use indices 1 and 2 instead of u and v. The shift condition among the matrix blocks in question is:

$$X_{1H}^{1H} = X_{2H}^{2H}, X_{2H}^{1H} = X_{2H}^{1H}, X_{1A,1D,2A,2D}^{1H} = X_{1A,1D,2A,2D}^{2H}, X_{1H}^{1A,1D,2A,2D} = X_{2H}^{1A,1D,2A,2D}. \tag{15}$$

Multiple indices are just a short-hand for multiple equalities.

Every one of the blocks X_1^1, X_1^2, X_2^1 and X_2^2 has the form (14), so let us denote parameters of each block by the corresponding vertices. If the equation (12) is to be violated by these blocks, it must be $X_{1H}^{1H} \neq X_{2H}^{1H}$ and therefore $\alpha_{11} \neq \alpha_{12}$.

In terms of parameters, the shift condition (15) results in $\alpha_{11} = \alpha_{21}$ and therefore there are no other attractors apart from the p-attractors for the eigenvalue -1.

For the eigenvalue 1, the general form of the attractor in one vertex is:

$$\Xi = \begin{bmatrix} \alpha & \gamma+\delta & \beta+\epsilon \\ \gamma+\epsilon & \beta & \alpha+\delta \\ \beta+\delta & \alpha+\epsilon & \gamma \end{bmatrix}. \tag{16}$$

Let us have three vertices and let us denote them just by numbers $1, 2$ and 3 for clarity of equations. We assume that vertices 1 and 2 are connected by a horizontal edge and 2 and 3 are connected by an anti-diagonal edge. The part of the shift condition related to these three vertices result in the equality: $\alpha_{12} - \alpha_{22} = \beta_{23} - \beta_{22}$.

Thanks to the symmetry, similar restriction will result from the shift condition for any other orientation of the vertices. Overall, due to the shift condition, if the equation (12) is not violated in one case, it can not be violated in any other (e.g. $\alpha_{12} = \alpha_{22} \Rightarrow \beta_{23} = \beta_{22}$).

We know that there is at least the identity operator, which is certainly an attractor and is not a p-attractor. Let us now assume that we have two different non-p-attractors X_1, X_2. If there is a linear combination $z_1X_1 + z_2X_2 = Y$ for some complex numbers z_1, z_2 such that Y is a p-attractor, then we only have one independent non-p-attractor. The other non-p-attractor (e.g. X_2) is just a linear combination of the first non-p-attractor X_1 and some p-attractor Y and therefore we do not want to add X_2 to the set of attractors.

There clearly exist z_1, z_2 such that $z_1\alpha1_u^u + z_2\alpha2_u^u = z_1\alpha1_v^u + z_2\alpha2_v^u$. (Here $u, v \in V$ denote vertices and 1 and 2 distinguish the two attractors.) Since the equality will hold in all other elements that could violate (12), the resulting linear combination of X_1 and X_2 must be a p-attractor. In conclusion, there is only one non-p-attractor for the eigenvalue 1 in any honeycomb lattice graph and it can be chosen as an identity matrix.

Cycling Shift Operator

The walker may be cycling either clockwise or counter-clockwise in all hexagons. We will examine for example the clockwise variant, which is shown in figure 1(b). The action of a local permutation P for a clock-wise cycling walk is a cyclic permutation $H \to D \to A \to H$.

Since we have started by the most problematic case of the reflecting shift operator, we will now just replicate the procedure for finding the asymptotic regime. In order to solve (8) we start by solving: $G_3 P_v^{CW} |\phi_v\rangle = \alpha |\phi_v\rangle$. The eigenvalues are $\alpha_1 = 1, \alpha_2 = \mathrm{e}^{i\pi/3}$ and $\alpha_3 = \mathrm{e}^{-i\pi/3}$ with corresponding eigenvectors

$$|\phi_v^1\rangle = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, |\phi_v^2\rangle = \begin{bmatrix} \mathrm{e}^{-i\frac{2\pi}{3}} \\ \mathrm{e}^{i\frac{2\pi}{3}} \\ 1 \end{bmatrix}, |\phi_v^3\rangle = \begin{bmatrix} \mathrm{e}^{i\frac{2\pi}{3}} \\ \mathrm{e}^{-i\frac{2\pi}{3}} \\ 1 \end{bmatrix}.$$

The shift condition (13) immediately results in three vectors $|\phi^1\rangle, |\phi^2\rangle$ and $|\phi^3\rangle$ that are just composed of identical blocks - one block for every vertex independently of the particular structure of the graph. We have a 3-dimensional subspace of p-attractors $|\phi^1\rangle\langle\phi^1|, |\phi^2\rangle\langle\phi^2|, |\phi^3\rangle\langle\phi^3|$ corresponding to the eigenvalue 1, 2-dimensional subspace $|\phi^1\rangle\langle\phi^3|, |\phi^2\rangle\langle\phi^1|$ corresponding to the eigenvalue $\mathrm{e}^{i\pi/3}$ and it's conjugate space and 1-dimensional subspace $|\phi^2\rangle\langle\phi^3|$ corresponding to the eigenvalue $\mathrm{e}^{i2\pi/3}$ and it's conjugate space.

When searching for the remaining attractors, a similar but simpler investigation as in the case of the reflecting walks leads to the same conclusion that there is only one non-p-attractor that can be chosen as an identity matrix.

6 Conclusion

In the examples above, we have seen that different choices of the shift operation may have major influence on the dynamics of the system. (Even the number of attractors is very different for the two presented variants.) Our definition of a quantum walk gives us a good overview of all possibilities and those can then be addressed. Further we have seen that with this definition, we can investigate quantum walks on any graph we want.

References

[1] B. Kollár, T. Kiss, J. Novotný, I. Jex. *Asymptotic dynamics of coined quantum walks on percolation graphs.* Physical Review Letters **108** (2012), 230505.

[2] B. Kollár, J. Novotný, T. Kiss, I. Jex. *Percolation induced effects in 2D coined quantum walks: analytic asymptotic solutions.* New Journal of Physics **16** (2014), 023002.

[3] J. Novotný, G. Alber, I. Jex. *Asymptotic evolution of Random Unitary Operations.* Central European Journal of Physics **8** (2010), 1001–1014.

[4] M. Štefaňák, I. Bezděková, I. Jex, M. Barnett. *Stability of point spectrum for three-state quantum walks on a line.* Quantum Information & Computation **14** (2014), 1213–1226.

[5] S. E. Venegas-Andraca. *Quantum walks: a comprehensive review.* Quantum Information Processing **11(5)** (2012), 1015–1106.

Asymptotics of Quantum Markov Dynamical Semigroups*

Jiří Maryška

4th year of PGS, email: maryska.jiri@gmail.cz
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:

Igor Jex, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Jaroslav Novotný, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. We investigate asymptotic dynamics of quantum markov dynamical semigroups equiped with so-called faithful state. We derive the relation between the asymptotics generated by the map describing time evolution of the states and the asymptotics of the conjugated map, which describes time evolution of observables. This relation enables us to show that the stationary states can be written in a form of so-called Gibbs-like states, which resemble of the Gibbs states in statistical physics.

Keywords: markovian systems, asymptotic evolution, integrals of motion, stationary state

Abstrakt. Studujeme asymptotický vývoj kvantových markovovských dynamických semigrup, pro které existuje takzvaný věrný stav. Odvodíme vztah mezi asymtotikou kvantových stavů a asymptotikou pozorovatelných. Tento vztah nám umožňuje napsat stacionární stavy ve formě, která připomíná Gibbsovy stavy používané ve statistické fyzice.

Klíčová slova: markovovské systémy, asymptotický vývoj, integrály pohybu, stacionární stavy

1 Introduction

Generally, the quantum mechanics is able to analytically describe the time evolution of a closed quantum system. A closed quantum system is characterized by unitary time evolution, resulting in a semigroup of unitary maps, which describe the time evolution in an arbitrary time $t > 0$. As real systems are alway interacting with their environment, a closed quantum system is an idealization and it can be used in a very limited number of cases, where interactions between system and its environment can be neglected. Any quantum system, whose time evolution is not described by an unitary map is called an open quantum system. Openness of a quantum system can result from more than an interaction with an environment. For instance, we may not be able to fully describe the quantum system itself, neglecting some degrees of freedom. This introduces randomness in a time evolution of the quantum system, which is impossible to describe with an unitary operator. For the reasons stated above, the theory of open quantum systems is

*Tato práce byla podpořena grantem SGS16/241/OHK4/3T/14.

important for both practical applications and for better understanding of the physical laws governing microscopic systems.

The theory of open quantum systems is mainly focused on so-called quantum markov processes [1, 2]. These are the systems, which meet the markov condition, which states that the evolution of the state of the system depends only on the state of the system at present time and not on the whole history of the system. From microscopic point of view, this condition requires that the influence of the interaction between the system and its environment on the environment is quickly dissolved in the environment. This is usually true for systems, whose environment is much larger than the system itself. Systems of these kind are called markovian.

There are two main approaches towards the time evolution of markovian systems - iterative discrete approach and the continuous approach. In discrete approach, we choose a time interval of length Δt which then defines our time resolution. The outcome of the evolution during this time step is then described by a completely possitive map Φ. By iterating this map, we get an arbitrary long time evolution of an open quantum system. Within this document, we focus on the continuous approach, in which we describe the time evolution by a certain dynamical semigroup $\mathcal{P}_t$, with $t > 0$. The markovianity of the evolution then implies certain properties, which need to be met by the generator $\mathcal{L}$ of the semigroup $\mathcal{P}_t$, which needs to be so-called conditionally completely positive [3].

Only a limited number of physically relevant cases can be succesfully solved without many unphysical restrictions. The mathematical difficulty results from a fact that the generator responsible for the time evolution is often not normal and thus a diagonalization in some orthogonal basis is not guaranteed. However, it can be showed that the asymptotic part (i.e. for $t \to \infty$) of the generator is always diagonalizable in case of finite systems and thus it can be solved analytically. This enables us not only to study the asymptotics of the system, but also a general features as stationary states and integrals of motion corresponding to the system.

The purpose of this document is to sum up known facts concerning asymptotics of finite markovian systems, to present newest development in a field and to show applications of the discussed theory on finite quanutm networks. For this purpose, this document is divided into following sections. In section 2, we define the quantum markov dynamical semigroup and we hint the procedure, which leads to derivation of the asymptotic evolution of finite markovian systems. We put stress on a fact that to be able to obtain asymptotics, given system must be equipped with so-called faithful state, which is crucial in obtaining of the set of asymptotic states. Furthermore we discuss the relations of the asymptotic evolution in different picutres of quantum mechanics. We show that the Schrödinger picture and the Heisenberg picture are related by the faithful state, should it exist. In part 3 we give mathematical description of stationary states and integrals of motion corresponding to the given markovian system and we show that stationary states can be written in a form which closely resembles of a well-known Gibbs states. The results and outlook is discussed in section 4.

2 Asymptotics of Quantum Markov Processess

A quantum markov dynamical semigroup is a continuous one-parameter family of maps $\mathcal{P}_t$, $t \geq 0$, which fulfils $\mathcal{P}_0 = I$. The evolution of quantum state ρ is given as

$$\rho(t) = \mathcal{P}_t(\rho(0)).$$

Generally, $\mathcal{P}_t$ can be written as [2]

$$\mathcal{P}_t = \exp[\mathcal{L}t], \tag{1}$$

where so-called generator $\mathcal{L}$ can be written in a following form:

$$\mathcal{L}(\rho) = -i[H, \rho] + \sum_j \left(L_j \rho L_j^\dagger - \frac{1}{2}\{L_j^\dagger L_j, \rho\} \right), \tag{2}$$

with $[\cdot,\cdot]$ and $\{\cdot,\cdot\}$ being commutator and anticommutator respectivelly and $H = H^\dagger$ being usually interpreted as the Hamiltonian corresponding to the system under consideration.The adjoint QMDS generated by a map $\mathcal{L}^\dagger$ describes the time evolution of observables. This can be seen from the relation for mean value of observable $A = A^\dagger$:

$$\langle A \rangle_{\rho(t)} = (A, \mathcal{P}_t(\rho))_{HS} = (\mathcal{P}_t^\dagger(A), \rho). \tag{3}$$

Using this equation yields the result for the generator of the adjointed map $\mathcal{P}_t^\dagger$:

$$\mathcal{L}^\dagger(A) = i[H, A] + \sum_j \left(L_j^\dagger A L_j - \frac{1}{2}\{L_j^\dagger L_j, A\} \right).$$

We can notice that the map $\mathcal{L}^\dagger$ fulfils the property

$$\mathcal{L}^\dagger(I) = 0.$$

Such maps $\mathcal{L}^\dagger$ are called unital maps and they form an important class with many neat properties. An adjoint to the unital map is called trace-preserving map. As a result, $\mathcal{L}$ of form (2) is always trace-preserving map. Note that the property $\mathcal{L}(I) = 0$ implies $\mathcal{P}_t(I) = I$ and thus unital maps leave the maximally mixed state undisturbed.

As mentioned in the section 1, generators of QMDS $\mathcal{L}$ are usually not normal and thus the diagonalization in some orthonormal basis is not possible. However, it turns out that the asymptotic part of the evolution is always diagonalizable and thus we can obtain it analytically. This is mainly due to the spectral properties of the generator $\mathcal{L}$. In this section, we review these properties and we indicate the procedure which leads towards obtaining the asymptotic dynamics. Then we discuss the asymptotic dynamics itself. Since the discrete and quantum case are very similar, we discuss only the continuous quantum markov processess. For the sake of transparency, we divide this section in the following subsections. First, we show that the asypmtotic part of the evolution of QMDS can be diagonalized. In the next subsection, we present equations fulfiled by the operators, which are responsible for asymptotic dynamics. Similar derivation for QMC was done in [6].

2.1 Diagonalization of the asympotic part of the evolution

As the generator $\mathcal{L}$ is not diagonalizable, we are forced to use its Jordan canonical form, which reads

$$\mathcal{L} = R \bigoplus_i J_i(\lambda_i) R^{-1},$$

with

$$J(\lambda_i) = \begin{pmatrix} \lambda_i & 1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & 0 & \dots & 0 \\ 0 & 0 & \lambda_i & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & \lambda_i \end{pmatrix}$$

being the i-th diagonal block of the Jordan form and R being the Jordan transformation. The time evolution is then generated by a map, which in Jordan form reads

$$\mathcal{P}_t = \exp[\mathcal{L}t] = R \exp\left[t \bigoplus_i J_i(\lambda_i) \right] R^{-1} = \\ = R \bigoplus_i \exp\left[t J_i(\lambda_i) \right] R^{-1}.$$

Clearly, we can write

$$J(\lambda_i) = \lambda_i I_i + N_i,$$

where N_i is nilpotent matrix satisfying $N_i^{\dim(J(\lambda_i))} = 0$. We thus have

$$\mathcal{P}_t = R \bigoplus_i \left(\exp[\lambda_i t] \left[\sum_{k=0}^{\dim(J(\lambda_i))-1} \frac{t^k N_i^k}{k!} \right] \right) R^{-1}.$$

The operator

$$D_i(t) = \sum_{k=0}^{\dim(J(\lambda_i))-1} \frac{t^k N_i^k}{k!}$$

has a polynomially increasing norm with exception of the case $\dim(J(\lambda_i)) = 1$, in which the norm is constant in time. Considering the latter case, we must have $\lambda_i = \mathrm{Im}(\lambda_i)$. If the real part of the eigenvalue λ_i was greater than zero, we would have contradiction with the existence of the finite bound of the map $\mathcal{P}_t$. If the real part of eigenvalue λ_i is lesser than zero, the part $\exp[\lambda_i t] D_i(t)$ converges towards zero operator and thus it does not contribute to the asymptotic dynamics. Applying similar reasoning for the case $\dim(J(\lambda_i)) > 1$ we conclude that we must have $\mathrm{Re}(\lambda_i) < 0$ as otherwise we would have contradiction with the existence of the finite bound of the map $\mathcal{P}_t$. The norm of the operator $\exp[\lambda_i t] D_i(t)$ thus converges towards zero value. We thus make the following conclusions.

- All eigenvalues of the generator $\mathcal{L}$ have non-positive real parts

- Jordan blocks corresponding to the eigenvalues with negative real parts do not contribute to asympotitic dynamics

- Jordan blocks corresopnding to the purely imaginary eigenvalues determine the asymptotics. All these blocks are one-dimensional and thus the asymptotic regime can be diagonalized

The asymptotic part of the evolution is thus confined to the subspace $\mathrm{Atr}(\mathcal{P}_t)$ defined as

$$\mathrm{Atr}(\mathcal{P}_t) = \bigoplus_{\lambda \in \sigma_{\mathrm{asm}}} \mathrm{Ker}\,(\mathcal{L} - \lambda I)\,, \tag{4}$$

with σ_{asm} being the purely imaginary part of the spectrum of $\mathcal{L}$. However, it is worth to notice that the elements of the attractor space are generally not density matrices as they do not need to be possitive or even selfadjointed. They are operators $X \in \mathcal{B}(\mathcal{H})$ and they are just a tool to construct the corresponding asymptotic state. This construction is done in the following subsection.

2.2 Faithful state and attractor equations

To be able to describe the asymptotics, we need a QMDS $\mathcal{P}_t$ to be equipped with so-called faithful state ρ [2, 6]. It is defined as an arbitrary strictly possitive invariant state, thus it meets the requirements $\rho > 0,\ \mathcal{P}_t(\rho) = \rho$ or equivalently $\rho > 0,\ \mathcal{L}(\rho) = 0$. In the rest of the text, the symbol ρ will be reserved for faithful state and a general state of the system will be denoted as σ. The existence of a faithful state is not guaranteed as any system has an invariant state, but it needs not to be strictly possitive. A system can have more than one faithful state, in which case it has infinit amount of faithful states a any convex combination of faithful states can again be a faithful state. In that case, to obtain the asymptotics we can use an arbitrary faithful state.

One can imidiatelly notice that problem of the existence of a faithful invariant state is trivial for unital quantum markov processess, as identity operator is always faithful state in case of unital quantum markov processess.

Suppose the set $\{X_{\lambda,i} | i \in \mathcal{I}_\lambda\}$ forms an orthonormal basis of the subspace $\mathrm{Ker}\,(\mathcal{L} - \lambda I)$. Then starting from the initial state $\sigma(0)$, the asymptotic state $\sigma(t \gg 0)$ takes form [7]

$$\sigma(t \gg 0) = \sum_{\substack{\lambda \in \sigma_{\mathrm{asm}},\\ j \in \mathcal{I}_\lambda}} e^{i\lambda t} \mathrm{Tr}\left[X^{\lambda,j\dagger} \sigma(0)\right] X_{\lambda,j}, \tag{5}$$

with $X^{\lambda,j}$ being dual operators to operators $X_{\lambda,j}$ satisfying relations

$$\mathrm{Tr}\left[X^{\lambda,j\dagger} X_{\lambda',j'}\right] = \delta_{\lambda\lambda'}\delta_{jj'}.$$

Using generalized Schwartz inequalities [5], one can prove that apart from normalization, following relation holds [6]:

$$X^{\lambda,j} \sim X_{\lambda,j}\rho^{-1}.$$

Furthermore, by using the linear contractive map $\mathcal{V}(X) = \mathcal{P}_t^\dagger(X\rho^{-1/2})\rho^{1/2}$ and its adjoint, we obtain following relations linking attractors of maps $\mathcal{P}_t$ and $\mathcal{P}_t^\dagger$ [6]:

$$\begin{aligned}
&X \in \mathrm{Ker}(\mathcal{L}-\lambda I) \Leftrightarrow X\rho^{-1} \in \mathrm{Ker}\left(\mathcal{L}^\dagger - \bar{\lambda} I\right),\\
&X \in \mathrm{Ker}(\mathcal{L}-\lambda I) \Leftrightarrow \rho^{-1}X \in \mathrm{Ker}\left(\mathcal{L}^\dagger - \bar{\lambda} I\right),\\
&X \in \mathrm{Ker}(\mathcal{L}-\lambda I) \Leftrightarrow \rho X\rho^{-1} \in \mathrm{Ker}\left(\mathcal{L} - \lambda I\right),\\
&X \in \mathrm{Ker}(\mathcal{L}-\lambda I) \Leftrightarrow \rho^{-1}X\rho \in \mathrm{Ker}\left(\mathcal{L} - \lambda I\right),\\
&X \in \mathrm{Ker}(\mathcal{L}-\lambda I) \Leftrightarrow \rho^{-1/2}X\rho^{-1/2} \in \mathrm{Ker}\left(\mathcal{L}^\dagger - \bar{\lambda} I\right).
\end{aligned} \tag{6}$$

These relation are of a significant importance. They map the attractor space generated by map $\mathcal{P}_t$ to the attractor space generated by map $\mathcal{P}_t^\dagger$ and vice versa. Furthermore, they allow us to find equations which must be satisfied by attractors. We note that for instance first of these equations imply for $\lambda = ia$ following relation:

$$\mathcal{L}(X) = ibX \Leftrightarrow \mathcal{L}^\dagger(X\rho^{-1}) = -ibX\rho^{-1},$$

or equivalently

$$\mathcal{P}_t(X) = e^{ibt}X \Leftrightarrow \mathcal{P}_t^\dagger(X\rho^{-1}) = e^{-ibt}X\rho^{-1}.$$

Using this we can prove the following result [7]:
Let $\mathcal{P}_t$ be a QMDS generated by a map $\mathcal{L}$ equipped with a faithful state ρ. Then $X \in \mathcal{B}(\mathcal{H})$ is an attractor corresponding to the eigenvalue $\lambda = ib$ iff the following set of equations hold:

$$\begin{gathered}
\left[L_i, X\rho^{-1}\right] = \left[L_i^\dagger, X\rho^{-1}\right] = \\
\left[L_i, \rho^{-1}X\right] = \left[L_i^\dagger, \rho^{-1}X\right] = 0,\\
[X\rho^{-1}, H] = bX\rho^{-1},\ [\rho^{-1}X, H] = b\rho^{-1}X.
\end{gathered} \tag{7}$$

In this part, we presented the form of the asymptotic state of QMDS $\mathcal{P}_t$ provided that the QMDS is equipped with so-called faithful state ρ. Next, we have shown that attractor spaces corresponding to QMDS $\mathcal{P}_t$ and $\mathcal{P}_t^\dagger$ are closely related by the faithful state ρ. Finally, we derived the equations governing the attractor space, so-called attractor equations. It is easy to see that the problem of solving attractor equations is considerably less complicated for the QMDS $\mathcal{P}_t^\dagger$ as its faithful state has a simple form $\rho \sim I$, which reduces attractor equations significantly.

3 Gibbs-like states

Within the statistical physics, one is usualy provided by a set of mean values corresponding to physical observables A_j of the system under consideration. According to quantum mechanics, these observables form a full set of integrals of motion $\mathcal{I}$. If the equilibrium is reached, the state of the system is then described by so-called Gibbs state σ_G [4], given by

$$\sigma_G = \frac{1}{Z} \exp\left[-\sum_j \lambda_j A_j \right], \tag{8}$$

with λ_j being Lagrangian multipliers, which need to be determined from the systems physics and Z is so-called partition sum, defined as

$$Z = \mathrm{Tr}\left[\exp\left[-\sum_j \lambda_j A_j \right] \right].$$

The Gibbs state is a result of a physical axiom called the maximal entropy principle [4], which states that the state of the system in a equilibrium maximalizes von Neumann entropy $S = -\mathrm{Tr}[\sigma \mathrm{Log}[\sigma]]$ while satisfying constraints given by mean values of physical observables.

A typical example of such state is so-called canonical ensembe. Suppose we are provided by mean value of the energy $\langle H \rangle$, the equilibrium (or stationary) state then takes the form [4]:

$$\sigma_G = \frac{1}{Z} \exp\left[-\beta H \right].$$

In this section, we show that QMDS $\mathcal{P}_t$, which is equipped with a faithful state ρ has a corresponding set of stationary states, which can be written in a form, which closely resembles of a Gibbs states. For this similarity, we call them Gibbs-like states [8]. For better clarity, we divide the rest of this section into two subsections. In the first subsection, we sum up important properties of the set of stationary states and the set of integrals of motion. In the second subsection, we derive the form of Gibbs-like states.

3.1 Stationary states and integrals of motion

Stationary states of QMDS $\mathcal{P}_t$ are closely linked to its attractor space [7]. As they are defined by requirement $\mathcal{P}_t(\sigma) = \sigma$, $\forall t > 0$ or equivalently $\mathcal{L}(\sigma) = 0$, we see that they are confined to subspace $\mathrm{Ker}(\mathcal{L}) \subset \mathrm{Atr}(\mathcal{P}_t)$. However, as before we stress that elements of subspace $\mathrm{Ker}(\mathcal{L})$ are not states, but operators $X \in \mathcal{B}(\mathcal{H})$ which are not necessarily selfadjointed and/or positive. The subset of stationary states $\mathcal{S}(\mathcal{L})$ thus satisfies

$$\mathcal{S}(\mathcal{L}) \subset \mathrm{Ker}(\mathcal{L}).$$

Similarly, the subspace $\mathrm{Ker}(\mathcal{L}^\dagger)$ contains the set of integrals of motion $\mathcal{I}(\mathcal{L})$ [7], which are defined by requirement $\mathcal{P}_t^\dagger(A) = A$ or equivalently $\mathcal{L}^\dagger(A) = 0$. As in the previous case of stationary states, elements of the subspace $\mathrm{Ker}(\mathcal{L}^\dagger)$ are necessarily selfadjointed and thus this subspace contains operators which cannot be interpreted as integrals of motion resulting in relation

$$\mathcal{I}(\mathcal{L}) \subset \mathrm{Ker}(\mathcal{L}^\dagger).$$

Relations (6) provide us an important connection between elements of subsets $\mathcal{S}(\mathcal{L})$ and $\mathcal{I}(\mathcal{L})$. According to (6) the following statement holds:

$$0 \leq A \in \mathcal{I}(\mathcal{L}) \Rightarrow \rho^{\frac{1}{2}} A \rho^{\frac{1}{2}} \in \mathcal{S}(\mathcal{L}). \tag{9}$$

This property is crucial for derivation of Gibbs-like form of stationary states corresponding to QMDS $\mathcal{P}_t$.

Unitality of the QMDS $\mathcal{P}_t^\dagger$ provides a neat algebraic properties of subspaces $\mathrm{Ker}(\mathcal{L}^\dagger - \lambda I)$. Suppose we have operators X_1, X_2 such that $X_j \in \mathrm{Ker}(\mathcal{L}^\dagger - \lambda_j I)$. As these operators are the solutions of equations (7) with $\rho \sim I$, one can easily verify that we must have

$$X_1 X_2 \in \mathrm{Ker}(\mathcal{L}^\dagger - (\lambda_1 + \lambda_2) I).$$

As QMDS $\mathcal{P}_t$ is generally not unital, it lacks such a property. However, by relations (6) we can uncover that if we have X_1, X_2 such that $X_j \in \mathrm{Ker}(\mathcal{L} - \lambda_j I)$, then for instance $X_1 X_2 \rho^{-1} \in \mathrm{Ker}(\mathcal{L} - (\lambda_1 + \lambda_2) I)$.

A subspace $\mathrm{Ker}(\mathcal{L}^\dagger)$ has thus following properties:

$$\begin{aligned} &\mathrm{Ker}(\mathcal{L}^\dagger) \subset \mathrm{Atr}(\mathcal{P}_t^\dagger) \subset \mathcal{B}(\mathcal{H}), \\ &X \in \mathrm{Ker}(\mathcal{L}^\dagger) \Rightarrow X^\dagger \in \mathrm{Ker}(\mathcal{L}^\dagger), \\ &X_1, X_2 \in \mathrm{Ker}(\mathcal{L}^\dagger) \Rightarrow X_1 X_2 \in \mathrm{Ker}(\mathcal{L}^\dagger) \end{aligned} \tag{10}$$

These properties imply, that the subspace $\mathrm{Ker}(\mathcal{L}^\dagger)$ forms a C*-algebra. One can thus choose an orthogonal basis in such way that

$$\mathrm{Ker}(\mathcal{L}^\dagger) = \mathrm{span}\{I, A_1, \ldots, A_n\},\ A_j = A_j^\dagger. \tag{11}$$

Having chosen the orthogonal basis consisting of selfadjointed operators, the set of integrals of motion $\mathcal{I}(\mathcal{L})$ can be mathematically described as a real space with basis (11). As a result, the subset $\mathcal{I}(\mathcal{L})$ itself also forms a C*-algebra. The fact is of a significant importance, as it means that for any $A \in \mathcal{I}(\mathcal{L})$ and any complex analytic function f we also have $f(A) \in \mathcal{I}(\mathcal{L})$.

3.2 Derivation of form of Gibbs-like states

An important example of an analytic function is exponential, defined by its Taylor series as

$$\exp[X] = \sum_{n=0}^{\infty} \frac{X^n}{n!}.$$

For selfadjointed operator $X = X^\dagger$ we have $\exp[X] > 0$. Thus, if σ represents a quantum state, the operator τ_σ defined as

$$\tau_\sigma = \frac{1}{\mathrm{Tr}[\exp[\sigma]]} \exp[\sigma]$$

represents a strictly positive quantum state. However if σ represents a stationary state of QMDS $\mathcal{P}_t$, then τ_σ is not a strictly positive stationary state apart from the case of unital $\mathcal{P}_t$, as the subset $\mathcal{S}(\mathcal{L})$ is not a C*-algebra and thus it is not closed with respect to the operator multiplication.

The subspace $\mathcal{I}(\mathcal{L})$ forms a C* algebra, thus if we have $A \in \mathcal{I}(\mathcal{L})$, then for T_A defined as

$$T_A = \exp[A]$$

satisfies thanks to unitality of $\mathcal{P}_t^\dagger$ $T_A \in \mathrm{Ker}(\mathcal{L}^\dagger)$ and thus it is also an integral of motion. Being provided by an integral of motion A, we can thus define a strictly positive stationary state τ_A as

$$\tau_A = \frac{1}{\mathrm{Tr}[\exp[A]\rho]}\rho^{\frac{1}{2}}\exp[A]\rho^{\frac{1}{2}}.$$

There are two important problems concerning such a map. We ask if for each strictly positive stationary state $\sigma \in \mathrm{Ker}(\mathcal{L})$ exists an integral of motion $A \in \mathrm{Ker}(\mathcal{L}^\dagger)$ such that $\sigma = \tau_A$ and in case of positive answer, we would like to investigate, if this result can be extended to any stationary state.

To answer these questions, we will stude the inverse map of exponential, which is the logarithm function. Generally, operator $\mathrm{Log}[A]$ is not uniquelly defined, however for $A > 0$ is is uniquelly defined by its Taylor expansion as

$$\mathrm{Log}[A] = \mathrm{Log}[I + (A - I)] = \sum_{n=0}^{\infty}\frac{(-1)^{n+1}}{n}(A - I)^n.$$

Now suppose we have a strictly positive stationary state σ. Then the operator $A_\sigma = \rho^{-\frac{1}{2}}\sigma\rho^{-\frac{1}{2}}$ is an strictly positive integral of motion and thus we must have $\mathrm{Log}[A_\sigma] \in \mathrm{Ker}(\mathcal{L}^\dagger)$ which according to (11) means that

$$\mathrm{Log}[A_\sigma] = \alpha I - \sum_j \lambda_j A_j$$

with λ_j real. Consequently, for the state σ we have

$$\begin{aligned}\sigma =& \rho^{\frac{1}{2}}\exp[\mathrm{Log}[A_\sigma]]\rho^{\frac{1}{2}} = \\ & \rho^{\frac{1}{2}}\exp\left[\alpha I - \sum_j \lambda_j A_j\right]\rho^{\frac{1}{2}} = \\ & \frac{1}{Z}\rho^{\frac{1}{2}}\exp\left[-\sum_j \lambda_j A_j\right]\rho^{\frac{1}{2}},\end{aligned} \tag{12}$$

with

$$Z = \mathrm{Tr}\left[\exp\left[-\sum_j \lambda_j A_j\right]\rho\right].$$

We call this state a Gibbs-like state, as it resembles the Gibbs states of statistical physics. For an arbitrary strictly positive stationary state σ and the set of integrals of motion (11) we can find a set of real valued parameters λ_j such that (12) holds.

Can this result to be extended to any stationary state? Suppose we have stationary state ω such that $0 \in \sigma(\omega)$ and thus ω is not strictly positive. In this case, for an arbitrary strictly positive stationary state σ exists a real parameter s such that the quantum state

$$\omega_{s,\sigma} = \frac{1}{\mathrm{Tr}[\omega + s\sigma]}(\omega + s\sigma)$$

is strictly positive and thus it can be written in Gibbs-like form

$$\omega_{s,\sigma} = \frac{1}{Z(s)}\rho^{\frac{1}{2}} \exp\left[-\sum_j \lambda_j(s) A_j\right] \rho^{\frac{1}{2}}.$$

The original state ω can be then retrieved as a limit

$$\omega = \lim_{s\to 0} \frac{1}{Z(s)}\rho^{\frac{1}{2}} \exp\left[-\sum_j \lambda_j(s) A_j\right] \rho^{\frac{1}{2}}. \tag{13}$$

Since ω is not strictly positive, some of the parameters λ_j must be necessarily divergent. This is analogous situation to the zero-temperature limit known from a statistical physics. However, as we have freedom in choosing a strictly positive state σ in definition of $\omega_{s,\sigma}$, the operator inside the limit (13) is not uniquelly defined.

Contrary to Gibbs states, Gibbs-like states do not maximalize the von Neumann entropy. However, a more general result can be derived concerning so-called relative entropy $S(\omega_1|\omega_2)$ [1]. For quantum states ω_1 and ω_2 such that $\mathrm{supp}(\omega_1 \subset \mathrm{supp}(\omega_2)$, the relative entropy of the state ω_1 with respect to the state ω_2 is defined as

$$S(\omega_1|\omega_2) = \mathrm{Tr}[\omega_1(\mathrm{Log}[\omega_2] - \mathrm{Log}[\omega_1])]. \tag{14}$$

It can be shown that the relative entropy is monotonically decreasing under positive maps [9]. Applying this result to a general quantum state σ and the faithful state ρ, we have

$$S(\sigma|\rho) \geq S(\mathcal{P}_t(\sigma)|\rho).$$

Thus, the stationary states, which have form of Gibbs-like states must minimize the relative entropy with respect to any faithful state ρ. In fact, this principle is a generalization of the principle of von Neumann entropy maximalization. If QMDS $\mathcal{P}_t$ is unital, then $\rho \sim I$ and (14) reduces to the negative of the von Neumann entropy and we obtain original maximal entropy principle.

4 Conclusion

In the preceeding text, we have investigated the asymptotic dynamics of quantum markov dynamical semigroups for finite quantum systems equipped with so-called faithful state. We have shown that the asymptotic part of the generator of the time evolution is diagonalizable, the asymtptotic evolution is confined to so-called attractor space and we have presented the equations which determine elements of the attractor space. Furthermore, we discussed the relation of the Schrödinger picture and the Heisenberg picture, which allow us to switch from one to another.

Next, we have shown the relation between the set of stationary states and the set of integrals of motion. The stationary states can be written in a form which closely resembles the Gibbs state used in statistical physics. For this similarity, we call these states Gibbs-like states. Also, Gibbs-like states are determined by a principle of minimalization of relative quantum entropy, which is a generalization of the maximal von Neumann entropy principle for nonunital quantum channels.

A natural generalization of the presented theory would be the one for quantum markov processes with no faithful state. Another direction is to generalize presented results for trace nonincreasing quantum channels. This has been done for QMC [6] but the situation is more complicated for QMDS as it requires more general form of the generator $\mathcal{L}$. These problems will be adressed in the future research.

References

[1] Nielsen A., Chuang I. *Quantum Computation and Quantum Information* Cambridge University Press, 2000

[2] Alicki R., Lendi K. *Quantum Dynamical Semigroups and Applications* Springer, 2007

[3] Attal S., Joye A., Pillet C. *Open Quanutm Systems II: The Markovian Approach* Springer-Verlag Berlin, 2006

[4] Balian R. *From Macrophysics to Microphysics* Springer-Verlag Berlin, 2007

[5] Paulsen V. *Completely Bounded Maps and Operator Algebras* Cambridge University Press, 2002

[6] Novotný J., Alber G., Jex I. *Asymptotic properties of quantum markov chains* J. Phys. A, **45** 485301, 2012

[7] Novotný J., Maryška J., Jex I. In preparation (concerning connection between integrals of motion and stationary states)

[8] Novotný J., Maryška J., Jex I. In preparation (concerning Gibbs-like states)

[9] Pétz D. *Monotonicity of quantum relative entropy revisited* arXiv:quant-ph/0209053, 2002

Competitive Rank Based Integer Optimization Heuristic with Lévy Flights[*]

Matej Mojzeš

6th year of PGS, email: mojzemat@fjfi.cvut.cz
Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:

Jaromír Kukal, Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Quang Van Tran, Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. Novel evolutionary integer optimization heuristic yields from the theory of Mean Field Annealing and is based on population quality rank instead of objective function values. This way population center and covariance matrix are estimated for given temperature and then used as directional correction of Lévy Flight mutation. Similarly to Competitive Differential Evolution, the heuristic is of competitive nature. Here, nine Lévy Flight mutations compete and are selected according to their success. Resulting heuristic has four parameters: population size, regularization factor, annealing temperature and Lévy Flight temperature. This heuristic is suitable integer optimization tasks with many local extremes. One such task is the Clerc's Zebra-3, discrete optimization benchmark problem, which is used for evaluation of novel heuristic.

Keywords: Heuristic, Mean Field Annealing, Lévy Flights, Rank, Integer Optimization

Abstrakt. Nová evolučná celočíselná heuristika čerpá z teórie žíhania stredného poľa a je založená na poradí kvality populácie namiesto pôvodných hodnôt účelovej funkcie. Týmto spôsobom odhaduje stred populácie a kovariančnú maticu pre danú teplotu, ktoré následne použije pre smerovú korekciu mutácie pomocou Lévyho letov. Podobne ako napr. kompetitívna diferenciálna evolúcia, aj táto heuristika je založená na koncepte súťaživosti. V tomto prípade súťaží deväť mutačných operátorov a sú vyberané podľa úspešnosti. Výsledná heuristika má štyri parametre: veľkosť populácie, regularizačný faktor, teplotu žíhania a teplotu Lévyho letov. Heuristika je vhodná pre úlohy celočíselnej optimalizácie s mnohými lokálnymi extrémami. Jednou takou úlohou je Clercova Zebra-3, záťažová úloha pre diskrétnu optimalizáciu, ktorá je použitá pre posúdenie výkonnosti a spoľahlivosti novej heuristiky.

Kľúčové slová: Heuristika, žíhanie stredného poľa, Lévyho lety, poradie, celočíselná optimalizácia

1 Introduction

There is a variety of optimization methods and meta-heuristics including integer and binary ones. Let $n \in \mathbb{N}$ be task dimension, $\mathbf{x}$, $\mathbf{a}$, $\mathbf{b} \in \mathbb{Z}^n$ be independent variable, its

[*] This paper is supported by grant SGS14/208/OHK4/3T/14 CTU in Prague.

lower and upper bounds satisfying $\mathbf{a} < \mathbf{b}$. The domain of integer optimization is

$$\mathcal{D} = \{ \ \mathbf{x} \in \mathbb{Z}^n | \ \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\} \tag{1}$$

as a frame of objective function

$$\mathrm{f} \colon \mathcal{D} \to \mathbb{R} \tag{2}$$

minimization. This task can be enriched by threshold

$$f^* \geq \min_{\mathbf{x} \in \mathcal{D}} \mathrm{f}(\mathbf{x}). \tag{3}$$

Then the goal set

$$\mathcal{G} = \{ \ \mathbf{x} \in \mathcal{D} \mid \mathrm{f}(\mathbf{x}) \leq f^*\} \tag{4}$$

also contains global optimum of objective function. Finding any $\mathbf{x}_{\mathrm{opt}} \in \mathcal{G}$ is called here as sub-optimization task to be solved. It is useful to define range vector as $\mathbf{d} = \mathbf{b} - \mathbf{a} + \mathbf{1}$ and then denote

$$M^* = \mathrm{card}(\mathcal{G}) \geq 1, \tag{5}$$

$$M = \mathrm{card}(\mathcal{D}) = \prod_{k=1}^{n} d_k \geq 2^n \tag{6}$$

as number of goal states and total number of states. Meta-heuristic approach to sub-optimization task plays main role in the case of NP-hard problems [5], e.g.: set covering [3], or travelling salesman [1] problems.

Traditional integer optimization meta-heuristics include many discrete variants of Genetic Optimization [6], Simulated Annealing [7], Fast Simulated Annealing [12], Random Descent with Lévy Flights [8], Cuckoo Search [15], [9], Modified Cuckoo Search [13], and many others.

2 Rank Mean Field Integer Flight

The novel integer optimization heuristic is motivated by Evolutionary Search (ES), Mean Field Annealing (MFA), Parzen Estimate (PE) [11] of Probability Density Function (PDF), Lévy Flight Mutation (LFM), and competitive approach.

The *Rank Mean Field Integer Flight* (RMFIF) is population based heuristic but the population of $N \in \mathbb{N}$ vectors is unsorted. Denoting $f_k = \mathrm{f}(\mathbf{x}_k)$ we form the population as N-tuple of pairs in **ascending** order, sorted by objective function values,

$$\mathbf{P} = ((\mathbf{x}_1, f_1), \ldots, (\mathbf{x}_N, f_N)) \ . \tag{7}$$

Traditional MFA is based on the partition function

$$Z = \sum_{k=1}^{N} \exp(-f_k/T_{\mathrm{MFA}}) \tag{8}$$

over all N states for annealing temperature $T_{\mathrm{MFA}} > 0$. Resulting steady state probabilities of MFA are directly

$$p_k = \exp(-f_k/T_{\mathrm{MFA}})/Z \tag{9}$$

for $k = 1, \dots, N$. The MFA estimate of global minimum is the mean value

$$\mathbf{e} = \sum_{k=1}^{N} p_k \mathbf{x}_k. \tag{10}$$

The efficiency of RMFIF is based on rank approach when the values f_k are substituted by k for $k = 1, \dots, N$. Therefore, the probabilities are

$$p_k = \exp(-k/T_{\mathrm{MFA}})/Z = \frac{1-Q}{1-Q^N} \cdot Q^{K-1} \tag{11}$$

where $Q = \exp(-T_{\mathrm{MFA}}^{-1})$. Novel RMFIF heuristics also employs Parzen Estimate (PE) of the probability density function (PDF) for given population $\mathbf{P}$ of fixed size N, fixed width $\sigma > 0$, and Gaussian kernel in the form

$$\mathrm{q}(\mathbf{x}) = \frac{1}{N} \cdot \sum_{k=1}^{N} \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_k\|^2}{2\sigma^2}\right). \tag{12}$$

Steady state probabilities of MFA are directly used for *Weighted Density Estimate* (WDE) as

$$\mathrm{g}(\mathbf{x}) = \sum_{k=1}^{N} \frac{p_k}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_k\|^2}{2\sigma^2}\right). \tag{13}$$

Very important particular cases are:

- When $T_{\mathrm{MFA}} \to +\infty$ then $\mathrm{g} \to \mathrm{q}$ which is the Parzen Estimate.
- When $\mathbf{x}_{\min}$ is unique minimum from given population $\mathbf{P}$ and $T_{\mathrm{MFA}} \to 0+$ then $\mathrm{g} \to \mathrm{N}(\mathbf{x}_{\min}, \sigma^2\mathbf{I})$ which is Gaussian distribution centered in the best population point.

Using characteristic function

$$\psi(\mathbf{t}) = \mathrm{E}\,\exp(\jmath \mathbf{x}^{\mathrm{T}}\mathbf{t}) \tag{14}$$

we explicitly obtained

$$\psi(\mathbf{t}) = \sum_{k=1}^{N} p_k \exp\left(\jmath \mathbf{x}_k^{\mathrm{T}}\mathbf{t} - \mathbf{t}^{\mathrm{T}}(\sigma^2\mathbf{I})\mathbf{t}/2\right) \tag{15}$$

which is useful for direct calculation of moment characteristics. The first moment is the mean value of sampled population which is well known from the MFA theory as

$$\mathbf{e} = \mathrm{E}\mathbf{x} = \sum_{k=1}^{N} p_k \mathbf{x}_k \tag{16}$$

in formal agreement with (10) meanwhile the covariance matrix is composed from two terms as

$$\mathbf{C} = \mathrm{E}(\mathbf{x}-\mathbf{e})(\mathbf{x}-\mathbf{e}^{\mathrm{T}}) = \mathbf{C}_{\mathrm{raw}} + \sigma^2\mathbf{I} \tag{17}$$

where

$$\mathbf{C}_{\text{raw}} = \sum_{k=1}^{N} p_k (\mathbf{x}_k - \mathbf{e})(\mathbf{x}_k - \mathbf{e})^{\mathrm{T}} \tag{18}$$

is obvious covariance matrix of sampled population and the second term of (17) is resulting effect of Parzen estimation as a kind of statistical regularization.

The main idea behind RMFIF method is in directional mutation based on $\mathbf{C}$ of $\mathbf{P}$ but driven by Lévy distribution. Novel *Rank Mean Field Integer Flight Mutation* (RFM) consists of several steps:

- Calculate $\mathbf{C}_{\text{raw}}$ for given population $\mathbf{P}$ and temperature $T_{\text{MFA}} > 0$,
- Generate random n-dimensional Gaussian vector $\mathbf{y} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$,
- Calculate directional vector $\mathbf{z} = (\mathbf{C}_{\text{raw}} + \sigma^2 \mathbf{I})^{1/2} \cdot (\mathbf{y}/\|\mathbf{y}\|_2)$,
- Generate $d \sim \text{Levy}(\beta)$ using Lévy distribution with $0 < \beta < 2$,
- Real unlimited mutation of $\mathbf{x} \in \mathcal{D}$ using mutation temperature $T_{\text{mut}} > 0$ produces $\mathbf{r} = \mathbf{x} + T_{\text{mut}} \cdot d \cdot \mathbf{z}$,
- Using component-wise rounding and perturbation via mirroring we calculate $\mathbf{x}_{\text{new}} = \mathrm{P}([\mathbf{r}], \mathbf{a}, \mathbf{b})$.

This novel type of mutation operator yields from the properties of given population $\mathbf{P}$ applying MFA theory to obtain the direction of mutation as vector $\mathbf{z}$. Final mutation of $\mathbf{x}$ is realized as perturbed directional integer Lévy flight with dimensionless mutation temperature T_{mut}.

The RFM operator has four tuning parameters:

- $T_{\text{MFA}} > 0$ for Mean Field Annealing,
- $T_{\text{mut}} > 0$ for Lévy flights,
- $\beta \in (0, 2)$ for Lévy distribution,
- $\delta > 0$ for Parzen estimate.

3 Basic Frame of RMFIF

First we set $N \in \mathbb{N}$, $H \in \mathbb{N}, H \geq 2, n_0 \in \mathbb{N}, \delta > 0$, f^*, $N_{\max}$ as population size, mutation portfolio size, initial counter value, threshold, final value and maximal number of evaluations. Then we introduce mutation family

$$\mathcal{F} = \{\mathrm{RFM}_1(\mathbf{x}), \mathrm{RFM}_2(\mathbf{x}), \ldots, \mathrm{RFM}_{\mathrm{H}}(\mathbf{x})\}\,. \tag{19}$$

The algorithm of RMFIF is described in detail in Algorithm 1.

Individual RFM_j is described by parameters $(\mathrm{T}_{\mathrm{MFA}}, \mathrm{T}_{\mathrm{mut}}, \beta, \sigma)_j$ for $i = 1, \ldots, H$. General suggestion is to use fixed $\sigma \in (0, 1)$ for all mutations in the portfolio due to

Algorithm 1 RMFIF

```
Set counters $\mathbf{n} = n_0\mathbf{1} \in \mathbb{N}^H$, and mutation portfolio.
Init population $\mathcal{P}$ of size $N$ by uniform sampling from $\mathcal{D}$.
**while** $f_{\text{best}} > f^*$ and $neval < N_{\max}$: **do**
    Sort population in ascending order
    Using systematic selection strategy find $\mathbf{x}_k \in \mathcal{P}$
    **for** $k = 1, \ldots, N$ **do**
        Generate randomly index $j$ according to mutation probabilities $p_j = n_j / \, ||\mathbf{n}||_1$
        Perform $\mathbf{x}_{\text{new}} = \text{RFM}_j(\mathbf{x})$
        Evaluate $f_{\text{new}} = \text{f}(\mathbf{x}_{\text{new}})$.
        **if** $f_{\text{new}} < f_{(k)}$ **then**
            Update $n_j = n_j + 1$, $\mathbf{x}_k = \mathbf{x}_{\text{new}}$, $f_{(k)} = f_{\text{new}}$
            **if** $\min_{i=1,\ldots,H} p_i = \frac{\min n_i}{||\mathbf{n}||_1} < \frac{\delta}{H}$ **then**
                Reset counters as $\mathbf{n} = n_0\mathbf{1}$
            **end if**
        **end if**
    **end for**
**end while**
```

integer nature of searching domain $\mathcal{D}$. The parameter of Lévy distribution can be also set to fixed value $\beta = 1$ for the first experiments. Therefore, the competition of mutations is only about adaptive changing of temperature $T_{\text{mut},i}$ for $i = 1, \ldots, H$, for given N, T_{MFA}.

4 Experimental Results

Our testing task will be Clerc's Zebra-3, which is a non-trivial binary optimization problem ($\mathbf{a} = \mathbf{0}$, $\mathbf{b} = \mathbf{1}$) and part of discrete optimization benchmark problems [2]. Zebra-3 function is defined for $n = 3\,d^*$, $d^* \in \mathbb{N}$ as

$$\text{z}(\mathbf{x}) = \sum_{k=1}^{d^*} \text{z}_{1+\text{mod}(k-1,2)}(\boldsymbol{\xi}_k) \tag{20}$$

where $\boldsymbol{\xi}_k = (x_{3\,k-2}, \ldots, x_{3\,k})$ and

$$\text{z}_1(\boldsymbol{\xi}) = \begin{cases} 0.9 & \text{for } ||\boldsymbol{\xi}||_1 = 0 \\ 0.6 & \text{for } ||\boldsymbol{\xi}||_1 = 1 \\ 0.3 & \text{for } ||\boldsymbol{\xi}||_1 = 2 \\ 1.0 & \text{for } ||\boldsymbol{\xi}||_1 = 3 \end{cases} \tag{21}$$

$$\text{z}_2(\boldsymbol{\xi}) = \begin{cases} 0.9 & \text{for } ||\boldsymbol{\xi}||_1 = 3 \\ 0.6 & \text{for } ||\boldsymbol{\xi}||_1 = 2 \\ 0.3 & \text{for } ||\boldsymbol{\xi}||_1 = 1 \\ 1.0 & \text{for } ||\boldsymbol{\xi}||_1 = 0 \end{cases} \tag{22}$$

Zebra-3 function is a subject of maximization with the maximum value of $n/3$. Therefore we will minimize

$$\text{f}(\mathbf{x}) = \frac{n}{3} - \text{z}(\mathbf{x}) \tag{23}$$

for $n = 30$ with $f_{\text{opt}} = 0$. This binary task consists of $2^n = 2^{30} = 1,073,741,824$ states with single global optimum and $2^{d^*} = 2^{10} = 1,024$ local minima. Simulations were performed for $N_{\max} = 100,000$.

Table 4 summarizes performance of the novel heuristic compared to referential methods (Fast Simulated Annealing [12] and Simulated Annealing [7]) using basic heuristic performance measures [10]:

- MNE as *mean number of objective function evaluations* until optimal solution was found,
- SNE as *standard deviation of the number of evaluations*,
- REL – *reliability*, ratio between the number of successful runs (when f^* was reached) and total number of runs (100).

Heuristic	**Parameters**	MNE	SNE	REL
SA	$T_0 = 0.00005, n_0 = 0, T_{\text{Gauss}} = 0.5$	13,402	9,497	100%
FSA	$T_0 = 0.01, n_0 = 1, T_{\text{Gauss}} = 0.5$	12,221	8,468	100%
RMFIF	$N = 100, n_0 = 1, \delta = 0.2, T_{\text{MFA}} = 2, \sigma = 0.001$	7,898	1,721	100%

Table 1: Performance comparison of RMFIF with referential methods

All heuristic parameters were empirically tuned to the given problem and from this perspective we can conclude that RMFIF can outperform referential methods:

- it is clearly faster (having lower MNE) and
- it exhibits much better ratio of SNE/MNE, which describes heuristics stability.

On the other hand, RMFIF, as all heuristics [14], does not always perform optimally. Figure 1 demonstrates its performance for different temperatures T_{MFA} using Feoktistov criterion [4], defined as

$$FEO = MNE/REL \ . \tag{24}$$

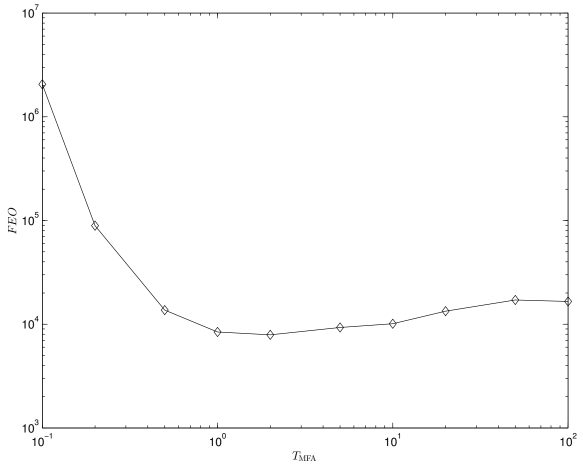


Figure 1: RMFIF performance measured by FEO criterion for different temperatures T_{MFA}

Figure 1 provides evidence on importance of correct setup of the T_{MFA} temperature. Mainly, very low temperature causes the algorithm to overestimate quality of the best population solution and thus become unreliable and slow. Based on our experience, good range for T_{MFA} selection is

$$\frac{1}{100} \leq \frac{T_{\mathrm{MFA}}}{N} \leq 2\ . \tag{25}$$

5 Conclusions

We have contributed to family of evolutionary heuristic method with a novel population based integer optimization heuristic that is of competitive nature and yields from the theory of Mean Field Annealing and reputable performance of Lévy Flights. The most important parameter is T_{MFA} that needs to be adjusted mainly according to population size and thanks to the rank correction, the selection of temperature does not need to reflect expected objective function ranges. As we have shown on Zebra-3 benchmark function, the novel heuristic has promising performance. However, more research and development is in progress, also because of the *No Free Lunch Theorem for optimization* [14].

References

[1] V. Cerny. *Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm.* J of optimization theory and applications **45** (1985), 41–51.

[2] H. Chen, Y. Zhu, K. Hu, and X. He. *Hierarchical swarm model: a new approach to optimization.* Discrete Dynamics in Nat and Society **2010** (2010).

[3] V. Chvatal. *A greedy heuristic for the set-covering problem.* Mathematics of operations research **4** (1979), 233–235.

[4] V. Feoktistov. Differential evolution–in search of solutions, (2006).

[5] M. R. Gary and D. S. Johnson. Computers and intractability: A guide to the theory of np-completeness, (1979).

[6] D. E. Golberg. *Genetic algorithms in search, optimization, and machine learning.* Addion wesley **1989** (1989), 102.

[7] S. Kirkpatrick. *Optimization by simulated annealing: Quantitative studies.* Journal of statistical physics **34** (1984), 975–986.

[8] M. Klimt, J. Kukal, and M. Mojzes. *Lévy flights in binary optimization.* Archives of Control Sciences **4** (2013), 447–454.

[9] J. Kukal, M. Mojzes, Q. V. Tran, and J. Bostik. *Integer cuckoo search.* In 'Proceedings of Mendel 2012 Soft Computing Conference', 298–303, (2012).

[10] M. Mojzes, J. Kukal, Q. V. Tran, and J. Jablonsky. *Performance comparison of heuristic algorithms via multi-criteria decision analysis.* In 'Proceedings of Mendel 2011 Soft Computing Conference', 244–251, (2011).

[11] E. Parzen. *On estimation of a probability density function and mode.* The annals of mathematical statistics **33** (1962), 1065–1076.

[12] H. Szu and R. Hartley. *Fast simulated annealing.* Physics letters A **122** (1987), 157–162.

[13] S. Walton, O. Hassan, K. Morgan, and M. Brown. *Modified cuckoo search: a new gradient free optimisation algorithm.* Chaos, Solitons & Fractals **44** (2011), 710–718.

[14] D. H. Wolpert and W. G. Macready. *No free lunch theorems for optimization.* Evolutionary Computation, IEEE Transactions on **1** (1997), 67–82.

[15] X.-S. Yang and S. Deb. *Engineering optimisation by cuckoo search.* International Journal of Mathematical Modelling and Numerical Optimisation **1** (2010), 330–343.

Small Non-Differentiable Perturbations of Crandall-Rabinowitz Bifurcation

Josef Navrátil [*]

4. ročník PGS, email: navrajos@fjfi.cvut.cz
Katedra fyziky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

školitelé:

Milan Kučera, Matematický ústav, Akademie věd ČR, v.v.i.

Martin Väth, Freie Universität Berlin

Abstract. The paper concerns abstract equations of the type

$$F(\lambda, u) = \tau G(\tau, \lambda, u). \tag{1}$$

Here U and V are Banach spaces, $F\colon \mathbb{R} \times U \to V$ and $G\colon \mathbb{R}^2 \times U \to V$ are maps, and it is supposed that

$$F(\lambda, 0) = G(\tau, \lambda, 0) = 0 \quad \text{for all } \tau, \lambda \in \mathbb{R}. \tag{2}$$

Hence, for all τ and λ there exists the so-called trivial solution $u = 0$ to (1) and local bifurcation of nontrivial solutions to (1) from the trivial solution is described. Under certain assumptions and if $\tau = 0$, the set of all solutions to (1) close to zero is described by the celebrated Crandall-Rabinowitz Theorem, see [1]. A typical field of applications of the Crandall-Rabinowitz Theorem are elliptic boundary value problems. The abstract setting (1) of the present paper is initiated by elliptic boundary value problems with non-smooth nonlinearities.

$$\begin{aligned} -\operatorname{div} A(x, \lambda, u, \nabla u) + f(x, \lambda, u, \nabla u) &= \tau h(x) g(x, \tau, \lambda, u)^+ && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega \end{aligned}$$

with $\tau \in \mathbb{R}$, $h \in L^p(\Omega)$ with $p > n$ and $g\colon \Omega \times \mathbb{R}^3 \to \mathbb{R}$ such that $g(x, \tau, \lambda, 0) = 0$ for all x, τ and λ. Here $g(x, \tau, \lambda, u)^+ := \max\{g(x, \tau, \lambda, u), 0\}$ is the positive part of $g(x, \tau, \lambda, u)$. The right-hand side of this equation is a so-called non-invasive perturbation, because it does not disrupt the existence of the trivial solution $u = 0$, but it may stabilize or destabilize this solution. In many applications this is just the reason to introduce this term. The perturbation works only in those points x, for those parameters τ and λ and for those states u for which $g(x, \tau, \lambda, u(x))$ is above the threshold zero.

Another field of applications of main abstract result are reaction-diffusion systems exhibiting a Turing diffusion driven instability:

$$\begin{aligned} &d_1 \Delta u_1 + f(u_1, u_2) = 0 \quad \text{in } \Omega, \\ &d_2 \Delta u_2 + g(u_1, u_2) + \tau\Big(\big[g_-(x, u_2) u_2\big]^- - \big[g_+(x, u_2) u_2\big]^+\Big) = 0 \quad \text{in } \Omega, \\ &\qquad \frac{\partial u_1}{\partial \nu} = \frac{\partial u_2}{\partial \nu} = 0 \quad \text{on } \partial\Omega, \end{aligned}$$

[*] Práce autora byla finančně podpořena grantem č. SGS16/239/OHK4/3T/14

with f and g being C^1 functions and satisfying additional sign conditions

$$\begin{aligned}
&f(0,0) = g(0,0) = 0,\\
&\frac{\partial f}{\partial u_1}(0,0) > 0 > \frac{\partial g}{\partial u_2}(0,0),\ \operatorname{tr} J = \frac{\partial f}{\partial u_1}(0,0) + \frac{\partial g}{\partial u_2}(0,0) < 0\\
&\det J = \frac{\partial f}{\partial u_1}(0,0)\frac{\partial g}{\partial u_2}(0,0) - \frac{\partial f}{\partial u_2}(0,0)\frac{\partial g}{\partial u_1}(0,0) > 0.
\end{aligned}$$

In this case the main results of the paper means a contribution to a study of domains of diffusion parameters for which spatial patterns, i.e. stationary spatially nonhomogeneous solutions of the corresponding evolution problem, exist.

Keywords: Crandall-Rabinowitz Theorem, Turing patterns, unilateral sources

Abstrakt. Článek se zabývá rovnicemi typu

$$F(\lambda, u) = \tau G(\tau, \lambda, u), \tag{3}$$

kde U a V jsou Banachovy prostory, $F\colon \mathbb{R} \times U \to V$ a $G\colon \mathbb{R}^2 \times U \to V$ jsou zobrazení, o kterých se dále předpokládá, že

$$F(\lambda, 0) = G(\tau, \lambda, 0) = 0 \quad \text{pro všechna } \tau, \lambda \in \mathbb{R}. \tag{4}$$

Potom pro všechny hodnoty parametrů τ a λ zde existuje tzv. triviální řešení $u = 0$ rovnice (3), a dále je popsána lokální bifurkace netriviálních řešení rovnice (3) z tohoto triviálního řešení. Pokud $\tau = 0$, tak za dalších předpokladů je množina všech řešení rovnice (3) blízkých nule popsána slavnou Crandall-Rabinowitzovou větou, viz [1]. Typickou oblastí aplikace Crandall-Rabinowitzovy věty jsou okrajové úlohy pro parciální diferenciální rovnice eliptického typu. Abstraktní formulace (3) v tomto článku je motivována právě těmito rovnicemi, ke kterým se navíc přidává malá nediferencovatelná nelinearita, celý systém pak vypadá následovně:

$$\begin{aligned}
-\operatorname{div} A(x, \lambda, u, \nabla u) + f(x, \lambda, u, \nabla u) &= \tau h(x) g(x, \tau, \lambda, u)^+ \quad \text{v } \Omega,\\
u &= 0 \quad \text{na } \partial\Omega,
\end{aligned}$$

kde $h \in L^p(\Omega)$ s $p > n$ a $g\colon \Omega \times \mathbb{R}^3 \to \mathbb{R}$ je takové, že $g(x, \tau, \lambda, 0) = 0$ pro všechna x, τ a λ. Výraz $g(x, \tau, \lambda, u)^+ := \max\{g(x, \tau, \lambda, u), 0\}$ označuje kladnou část $g(x, \tau, \lambda, u)$. Pravá strana v této eliptické rovnici je tzv. neinvazivní porucha, neboť nenarušuje existenci triviálního řešení $u = 0$, ale může toto řešení destabilizovat. V mnoha aplikacích je právě toto důvodem pro zavedení takového členu. Porucha působí jen v bodech x, jen pro parametry τ a λ a jen pro příslušné hodnoty funkce u, pro které je hodnota funkce $g(x, \tau, \lambda, u(x))$ kladná.

Další oblastí aplikace hlavního abstraktního výsledku tohoto článku jsou systémy reakce-difuze vykazující Turingovu difuzí řízenou nestabilitu:

$$\begin{aligned}
&d_1 \Delta u_1 + f(u_1, u_2) = 0 \quad \text{v } \Omega,\\
&d_2 \Delta u_2 + g(u_1, u_2) + \tau\Big(\big[g_-(x, u_2)u_2\big]^- - \big[g_+(x, u_2)u_2\big]^+\Big) = 0 \quad \text{v } \Omega,\\
&\frac{\partial u_1}{\partial \nu} = \frac{\partial u_2}{\partial \nu} = 0 \quad \text{na } \partial\Omega,
\end{aligned}$$

s funkcemi f a g třídy C^1 a splňujícími navíc následující podmínky:

$$\begin{aligned}
&f(0,0) = g(0,0) = 0\\
&\frac{\partial f}{\partial u_1}(0,0) > 0 > \frac{\partial g}{\partial u_2}(0,0),\ \operatorname{tr} J = \frac{\partial f}{\partial u_1}(0,0) + \frac{\partial g}{\partial u_2}(0,0) < 0\\
&\det J = \frac{\partial f}{\partial u_1}(0,0)\frac{\partial g}{\partial u_2}(0,0) - \frac{\partial f}{\partial u_2}(0,0)\frac{\partial g}{\partial u_1}(0,0) > 0.
\end{aligned}$$

Pro takový případ je hlavním výsledkem tohoto článku příspěvek ke studiu oblastí difúzních parametrů, pro které existují prostorové vzorky, tj. stacionární prostorově nehomogenní řešení příslušného evolučního problému.

Klíčová slova: Crandall-Rabinowitzova věta, Turingova nestabilita, jednostranné zdroje

Plná verze: Tato práce byla částečně prezentována na “Self-assembly in soft matter and biosystems” v Bad Honnefu a je součástí stejnojmenného článku [2], který byl odeslán k publikaci do sborníku konference “Patterns of Dynamic”, který vyjde v časopise ”Springer Proceedings in Mathematics & Statistics”.

Literatura

[1] Crandall, M. G. and Rabinowitz, P. H., *Bifurcation from simple eigenvalues*, J. Funct. Anal. **8** (1971), 321–340.

[2] Recke L., Väth M., Kučera M., Navrátil J. *Small Non-Differentiable Perturbations of Crandall-Rabinowitz Bifurcation*, submitted.

Numerical Modeling of Non-Isothermal Gas Flow and NAPL Vapor Transport in Soil*

Ondřej Pártl

5. ročník PGS, email: ondrej.partl@fjfi.cvut.cz
Katedra matematiky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

školitel: Michal Beneš, Katedra matematiky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

Abstract. In our paper, we introduce a mathematical model for the description of non-isothermal compressible flow of gas mixtures in heterogeneous porous media and we derive an efficient semi-implicit time-stepping numerical scheme for the solution of the governing equations. We experimentally estimate the order of convergence of the scheme in spatial variables and we present several computational studies that demonstrate the ability of the numerical scheme.

Keywords: non-isothermal flow, heterogeneous porous medium, semi-implicit scheme

Abstrakt. V našem článku navrhujeme matematický model pro popis neizotermického proudění směsi dvou plynů v heterogenním porézním prostředí a odvozujeme výkonné numerické schéma, které je semiimplicitní v čase, pro řešení systému rovnic, z nichž model sestává. Dále prezentujeme výsledky experimentálních odhadů řádu konvergence odvozeného schématu a rovněž několik výpočetních studií, které demonstrují schopnosti a možnosti tohoto schématu.

Klíčová slova: neizotermický proudění, heterogenní porézní prostředí, semiimplicitní schéma

Plná verze: O. Pártl, M. Beneš, P. Frolkovič, T. Illangasekare, K. Smits. *Numerical modeling of non-isothermal gas flow and NAPL vapor transport in soil.* Computer Physics Communications **202** (2016), 175–187.

[*] Tato práce byla podpořena grantem Development and Validation of Porous Media Fluid Dynamics and Phase Transitions Models for Subsurface Environmental Application Kontakt II LH14003 Ministerstva školství, mládeže a tělovýchovy České republiky.

On Uniqueness of T–duality with Spectators[*]

Filip Petrásek

3rd year of PGS, email: petrafil@fjfi.cvut.cz
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Ladislav Hlavatý, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. We investigate the dependence of non-Abelian T–duality on various identification of the isometry group of target space with its orbits, i.e. with respect to the location of the group unit on manifolds invariant under the isometry group. We show that T–duals constructed by isometry groups of dimension less than the dimension of the (pseudo)-Riemannian manifold may depend not only on the initial metric but also on the choice of manifolds defining positions of group units on each of the sub-manifold invariant under the isometry group. We investigate whether this dependence can be compensated by coordinate transformation.

Keywords: sigma model, string duality, non-Abelian T–duality, isometry group, spectator, coordinate transformation

Abstrakt. Zkoumáme závislost neabelovské T–duality na různém ztotožnění grupy isometrií cílového prostoru s jejími orbitami, tj. s ohledem na pozici grupové jednotky na varietách invariantních vzhledem ke grupě isometrií. Ukazujeme, že T–duály konstruovány pomocí grup isometrií dimenze menší než dimenze (pseudo)-Riemannovy variety můžou záviset nejen na počáteční metrice, ale také na volbě variet definujících pozice grupových jednotek na každé z podvariet invariantních vzhledem ke grupě isometrií. Vyšetřujeme, zda lze tuto závislost kompenzovat pomocí transformace souřadnic.

Klíčová slova: sigma model, strunová dualita, neabelovská T–dualita, grupa isometrií, přihlížeč, transformace souřadnic

Full paper: L. Hlavatý and F. Petrásek. *On uniqueness of T–duality with spectators.* Int. J. Mod. Phys. A **31** (2016) 1650143. arXiv:1606.02522 [hep-th].

[*] This work has been supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS16/239/OHK4/3T/14.

Modelování fázového chování směsi bitumenu a lehkých uhlovodíků nebo CO_2*

Tereza Petříková

4. ročník PGS, email: jindrter@fjfi.cvut.cz
Katedra matematiky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

školitel: Jiří Mikyška, Katedra matematiky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

Abstract. As the world supply of conventional light crudes decreases, the production from heavy oils and bitumens can supplement the societal energy needs. Knowledge of phase behavior of mixtures of heavy oils and bitumens with various light normal alkanes and CO_2 is important in efficient production from heavy petroleum fluids, especially bitumens. As the solvent (e.g. CO_2) dissolves in the bitumen reducing its viscosity, decreasing the steam requirement, and consequently decreasing the cost of bitumen and heavy oil recovery, the development of an accurate and reliable thermodynamic model to predict the solvent solubility in bitumen over wide ranges of temperatures and pressures is the key to proper design of solvent injection processes.

In hydrocarbon reservoir simulators, the most common models are regular cubic equations of state, such as the Peng-Robinson (PR) and Soave-Redlich-Kwong (SRK), which are used to model the phase behaviour of simple hydrocarbon systems where van der Waals and physical forces are the dominant interaction forces between molecules. Bitumen consists of various molecules with different hydrocarbon chains and polarities, especially asphaltene molecules which are the most polar, and the most complicated fraction of the crude oil. Asphaltenes give rise to molecular association and increase the polarity and complexity of the system. Therefore, the complex system of bitumen and solvent should be modeled using thermodynamic models, which take into account association forces between molecules.

In [1], a model for investigation of phase-stability and computation of multi-phase equilibrium at constant pressure, temperature and chemical composition was used together with two equations of state (Peng-Robinson (PR) [2] and Cubic-Plus-Association (CPA) [3] equations of state) to predict the phase behavior and solubility of CO_2, and normal alkanes from C_1 to nC_{10} in several bitumens over wide ranges of temperature and pressure. The predicted results were compared with available experimental data and modeling results available in the literature. The results show that the PR-EOS describes mixures of bitumens with CO_2, and alkanes when there is no second liquid phase or when the asphaltene content in the second liquid phase is not high. The CPA-EOS describes the phase behavior of mixtures of bitumens and CO_2, and alkanes in liquid-liquid states even when the asphaltene content of one of the phases is high. High asphatene content results in significant association and cross-association where the CPA-EOS is a natural choice.

Keywords: petroleum engineering, bitumen, asphaltenes, phase equilibrium, constant-pressure flash, Peng-Robinson equation of state, Cubic-Plus-Association eqution of state.

*Práce vznikla v rámci projektů Výpočetní metody v termodynamice vícesložkových směsí LH12064 MŠMT ČR.

Abstrakt. Vzhledem ke ztenčujícím se konvenčním zásobám ropy a rostoucí světové poptávce po energiích se v posledních letech stávají dříve alternativní způsoby těžby ropy čím dál zajímavější. Mezi tyto způsoby těžby patří např. výroba z těžkých olejů a ropných písků (bitumenů). Znalost fázového chování směsí těžkých olejů a bitumenů s různými lehkými normálními alkany a CO_2 je důležitá při efektivní výrobě ropy z těžkých ropných kapalin, zejména bitumenů. Při rozpouštění solventu (např. CO_2) v bitumenu dochází ke snížení jeho viskozity a současně k nižší spotřebě páry, v důsledku čehož dochází ke snížení nákladů na zvyšování výtěznosti bitumenu a těžkých ropných kapalin. Klíčovou roli ve správném návrhu procesů založených na vstřikování solventů hraje vývoj přesných a spolehlivých termodynamických modelů, které umožňují predikovat rozpustnost solventu v bitumenu v širokém rozsahu teplot a tlaků.

V kompozičních simulátorech ropných rezervoárů patří k nejčastěji používaným modelům obyčejné kubické stavové rovnice, jako je například Pengova-Robinsonova (PR) a Soaveho-Redlichova-Kwongova (SRK) stavová rovnice, které se používají k modelování fázového chování jednoduchých uhlovodíkových systémů, kde mezi dominantní interakční síly mezi molekulami patří van der Waalsovy síly. Bitumen se skládá z řady molekul s různými uhlovodíkovými řetězci a polaritami, zejména molekuly asfaltenu jsou silně polární a nejsložitější frakcí ropy. Asfalteny vyvolávají molekulární asociace a zvýšují polaritu a složitost systému, proto by komplexní systémy bitumenu a solventu měly být modelovány pomocí termodynamických modelů, které berou v úvahu asociační síly mezi molekulami.

V [1] byl použit model pro vyšetřování fázové stability a výpočet vícefázové rovnováhy vícesložkových směsí při konstantním tlaku, teplotě a chemickém složení spolu se dvěmi stavovými rovnicemi (Pengovou-Robinsonovou (PR) stavovou rovnicí [2] a kubickou stavovou rovnicí s asociačním členem [3]) k predikci fázového chování a rozpustnosti CO_2, a normálních alkanů C_1 až nC_{10} v několika bitumenech v širokém rozsahu teplot a tlaku. Predikované výsledky byly porovnány s dostupnými experimentálními daty a výsledky modelování dostupnými v literatuře. Výsledky ukazují, že PR-EOS popisuje dobře směsi bitumenu s CO_2 a alkany v případě nepřítomnosti druhé kapalné fáze nebo není-li obsah asfaltenu v druhé kapalné fázi vysoký. CPA-EOS popisuje fázové chování směsí bitumenu a CO_2 nebo alkanů ve stavech kapalina-kapalina, i když je obsah asfaltenu v jedné z fází vysoký. Vysoký obsah asfaltenu vede k významné asociaci molekul, kde je CPA-EOS přirozenou volbou.

Klíčová slova: ropné inženýrství, bitumen (živice), asfalteny, fázová rovnováha při konstantním tlaku, Pengova-Robinsonova stavová rovnice, kubická stavová rovnice s asociačním členem.

Full paper: This work is an abstract of article [1].

Literatura

[1] T. Jindrová, J. Mikyška, A. Firoozabadi. *Phase Behavior Modeling of Bitumen and Light Normal Alkanes and* CO_2 *by PR-EOS and CPA-EOS.* Energy&Fuels **30** (2016), 515–525

[2] D.Y. Peng, D.B. Robinson. *A New Two-Constant Equation of State.* Industrial & Engineering Chemistry Fundamentals **15(1)** (1976), 59–64.

[3] G.M. Kontogeorgis, E.C. Voutsas, I.V. Yakoumis, D.P. Tassios. *An Equation of State for Associating Fluids.* Ind. Eng. Chem. Res. **35** (1996), 4310–4318.

Doubly Trained Evolution Control for the Surrogate CMA-ES[*]

Zbyněk Pitra[†]

3rd year of PGS, email: z.pitra@gmail.com
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Martin Holeňa, Department of Nonlinear Dynamics & Complex Systems
Institute of Computer Science, CAS

Abstract. In many research and engineering tasks, optimization of real-world black-box functions that are costly to evaluate is a challenging problem of great importance. A single evaluation of the expensive function may require a great amount of resources in terms of time and performed experiments, measurements or simulations. In order to decrease the number of evaluations of the costly black-box function and still produce reasonably good solutions, a suitable regression model, also called surrogate model, of the black-box function can be employed [3]. In [4] we present a new variant of surrogate-model utilization in expensive continuous evolutionary black-box optimization. This algorithm is based on the surrogate version of the state-of-the-art evolutionary algorithm CMA-ES [2], the Surrogate Covariance Matrix Adaptation Evolution Strategy (S-CMA-ES) [1]. Similarly to the original S-CMA-ES, expensive function evaluations are saved through a surrogate model. However, the model is retrained after the points in which its prediction was most uncertain have been evaluated by the true fitness in each generation. We demonstrate that within small budget of evaluations, the new variant of S-CMA-ES using Gaussian processes [5] as a surrogate model improves the original algorithm and outperforms two state-of-the-art surrogate optimizers, except a few evaluations at the beginning of the optimization process.

Keywords: benchmarking, black-box optimization, surrogate model, Gaussian process

Abstrakt. Optimalizace tzv. black-box funkcí, tedy funkcí pro něž nelze najít správné matematické vyjádření, je významnou úlohou při řešení problémů ve výzkumu i v praxi. Pouhé jedno vyhodnocení této funkce může vyžadovat značné množství zdrojů potřebných k provedení příslušných experimentů, měření a simulací. Abychom dosáhli snížení počtu vyhodnocení drahé black-box funkce a zároveň zachovali vysokou kvalitu řešení, nabízí se použití vhodného regresního modelu, který je někdy také nazýván náhradní model [3]. V článku [4] představujeme nový způsob využití náhradního modelu v oblasti optimalizace drahých spojitých black-box funkcí. Tento nový algoritmus je založen na verzi v současnosti nejlepšího evolučního algoritmu CMA-ES [2], která používá náhradní modely, Surrogate Covariance Matrix Adaptation Evolution Strategy (S-CMA-ES) [1]. Stejně jako u původního algoritmu S-CMA-ES, jsou vyhodnocení drahé funkce šetřena pomocí náhradního modelu. Rozdílem oproti předchozímu algoritmu je

[*] This work was supported by the Grant Agency of the Czech Technical University in Prague with its grant No. SGS14/205/OHK4/3T/14 by the Czech Health Research Council project NV15-33250A, by the project "National Institute of Mental Health (NIMH-CZ)", grant number ED2.1.00/03.0078 and the European Regional Development Fund, and by the project Nr.LO1611 with a financial support from the MEYS under the NPU I program.

[†] This study has been provided in cooperation with Lukáš Bajer.

přetrénování modelu v každé generaci po přehodnocení bodů, jejichž předpovězené hodnoty měly největší míru nejistoty, pomocí skutečné black-box funkce. Ukazujeme, že při využití pouze malého počtu vyhodnocení se nová verze algoritmu S-CMA-ES, která používá gaussovské procesy [5] jako náhradní model, blíží k optimu nejen rychleji než algoritmus původní, ale i rychleji než dva ze v současnosti nejlepších algoritmů používajících náhradní modely s vyjímkou několika vyhodnocení na počátku optimalizačního procesu.

Klíčová slova: benchmarking, black-box optimalizace, náhradní modelování, gaussovské procesy

References

[1] L. Bajer, Z. Pitra, and M. Holeňa. *Investigation of Gaussian processes and random forests as surrogate models for evolutionary black-box optimization.* In 'Proceedings of the 17th GECCO Conference Companion', Madrid, (July 2015). ACM, New York.

[2] N. Hansen. *The CMA evolution strategy: A comparing review.* In 'Towards a New Evolutionary Computation', J. A. Lozano, P. Larrañaga, I. Inza, and E. Bengoetxea, (eds.), number 192 in Studies in Fuzziness and Soft Computing, Springer Berlin Heidelberg (January 2006), 75–102.

[3] Y. S. Ong, P. B. Nair, and A. J. Keane. *Evolutionary optimization of computationally expensive problems via surrogate modeling.* AIAA Journal **41** (2003), 687–696.

[4] Z. Pitra, L. Bajer, and M. Holeňa. *Doubly Trained Evolution Control for the Surrogate CMA-ES.* In 'Parallel Problem Solving from Nature - PPSN XIV - 14th International Conference, Edinburgh, UK, September 17-21, 2016, Proceedings', 59–68, (2016).

[5] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning.* Adaptative computation and machine learning series. MIT Press, (2006).

Student Skill Models in Adaptive Testing[*]

Martin Plajner

4th year of PGS, email: martin.plajner@fjfi.cvut.cz
Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jiří Vomlel, Department of Decision-Making Theory
Institute of Information Theory and Automation, CAS

Abstract. This paper provides a common framework, a generic model, for Computerized Adaptive Testing (CAT) for different model types. In CAT students are modeled by many different model types which are usually used separately. Although these models are separate and often very different, they show similarities and similar approaches when used in the CAT domain. We focus on joining these models under the common framework to allow them to share methods and to simplify the process of adaptive testing with different models. We present CAT procedure and question selection methods for the generic model. Any specific model fitting to the CAT generic framework can use these methods. We use three different types of specific models, Item Response Theory, Bayesian Networks, and Neural Networks, that instantiate the generic model. We describe these models and show how they fit into the generic structure. Moreover, we discuss an additional condition – the monotonicity – in these individual models. We illustrate the usefulness of its inclusion for CAT. With Bayesian Networks we use specific type of learning using generalized linear models to ensure the monotonicity property. We conducted simulated CAT tests on empirical data and we present methods used and results. The behavior and performance of individual models was assessed based on these tests. The best performing model was the BN model constructed by a domain expert; its parameters were learned from data under the monotonicity condition. Source codes used for our experiments are available online, with one of our data sources, and we are working to create R-language package for CAT tests.

Keywords: Bayesian Networks, Computerized Adaptive Testing, Generalized Linear Models, Item Response Theory

Abstrakt. Tento článek vytváří společný rámec, generický model, pro adaptivní testování znalostí (CAT) pro různé typy modelů. Studenti jsou v CAT modelováni různými druhy modelů, které jsou dnes obvykle všechny používány samostatně. Přestože jsou tyto modely samostatné a často velmi odlišné, sdílejí i společné přístupy a metody v kontextu CAT. V tomto článku se zaměřujeme na sjednocení těchto modelů pod společný rámec, abychom umožnili sdílení těchto metod a zjednodušili proces adaptivního testování za použití různých modelů. Pro generický model prezentujeme proces adaptivního testování a metody výběru otázky. Libovolný model spadající do společného rámce tak může využívat tyto společné metody. Prezentujeme 3 různé typy specifických modelů (teorii odpovědi na položku, bayesovské sítě a neuronové sítě) k inicializaci generického modelu. Popisujeme je a ukazujeme, jak spadají do generického rámce. Navíc se zabýváme další vlastností – monotonicitou – v těchto modelech. Ilustrujeme výhodnost využití této vlastnosti v CAT. Pro bayesovské sítě využíváme speciální způsob učení s

[*] The work on this paper has been supported by the Czech Science Foundation, GACR project No. 16-12010S and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS16/175/OHK3/2T/14.

pomocí zobecněných lineárních modelů k dodržení této vlastnosti. Na empirických datech jsme provedli simulované CAT testy a prezentujeme použité metody a výsledky. Na základě těchto testů jsme vyhodnotili chování a úspěšnost jednotlivých modelů. Jako nejlepší model nám vychází bayesovská síť, která je naučena za dodržení podmínky monotonicity pomocí zobecněných lineárních modelů. K dispozici, online na stránce jednoho z autorů, dáváme zdrojové kódy, které byly pro experiment použity (spolu s jednou datovou sadou). Pracujeme na publikaci těchto kódů v podobě balíčku pro adaptivní testování do jazyka R.

Klíčová slova: bayesovské sítě, počítačové adaptivní testování, zobecněný lineární model, teorie odpovědi na položku

Full paper: This article was presented at the 8th International Conference on Probabilistic Graphical Models (PGM2016) held in Lugano, Switzerland, 6.9.2016 – 9.9.2016. The full version is available in the conference proceedings: M. Plajner and J. Vomlel. *Student Skill Models in Adaptive Testing.* In the Proceedings of the 8th International Conference on Probabilistic Graphical Models (PGM2016), Lugano, Switzerland, 2016. Editors: A. Antonucci, G. Corani, C. de Campos.
Available at: `http://jmlr.org/proceedings/papers/v52/`

Surface Tension and Nucleation Rates of n-Heptane and n-Octane

Barbora Planková

5th year of PGS, email: barbora.plankova@gmail.com
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jan Hrubý, Department of Thermodynamics
Institute of Thermomechanics, CAS

Abstract. Homogeneous droplet nucleation is a phenomena occuring in atmosphere or industrial processes such as crude gas cleaning. In this work, we investigated the density gradient theory (DGT) and the influence of the capillary waves (CW) on the computed nucleation rates and surface tension of two n-alkanes. We found out that there are three major contributions to the surface tension of a droplet: linear increase by a Tolman's effect, quadratic decrease by the thickness of the interface and another increase by extraction of the CW. The extraction of the CW is important, because in the range of experimental data, the CW cancel out and does not play any role.

This text is a short version of the one that will be submited to **Journal of Chemical Physics**.

Keywords: nucleation, nucleation rate, surface tension, density gradient theory, capillary waves

Abstrakt. Homogenní nukleace kapek je fenomén, který se objevuje v atmosferických nebo průmyslových procesech, jako je čištění zemního plynu. V této práci jsme zkoumali povrchová napětí a nukleační rychlosti dvou alkanů, spočtené s pomocí teorie gradientu hustoty a kapilárních vln. Našli jsme tři základní příspěvky povrchovému napětí kapky: lineární přírůstek způsobený Tolmanovým efektem, kvadratický pokles způsobený tloušťkou fázového rozhraní a další přírůstek způsobený odstraněním kapilárních vln. Toto odstranění je důležité, protože v oblasti experimentálních dat se kapilární vlny vyruší a nehrají žádnou roli.

Tento text bude ve zkrácené formě poslán do žurnálu **Journal of Chemical Physics**.

Klíčová slova: nukleace, nukleační rychlosti, povrchové napětí, toerie gradientu hustoty, kapilární vlny

1 Introduction

Description of homogeneous droplet nucleation is important in many natural and industrial processes such as formation of secondary aerosols in the atmosphere, formation of water droplets in the steam turbines, or nucleation during the gas-cleaning procedures, e.g. in processing of natural gas.

Despite many attempts that resolved partial subproblems of the nucleation, there is no complete theory which would give quantitatively correct predictions.

The simplest method to describe the nucleation is the classical nucleation theory (CNT) developed by Becker and Döring[1] and extended by Zeldovich[18, 19].

However, a smooth change of the density between the two phases describes reality better. This approach is considered in the density gradient theory (DGT). The DGT was first developed in pioneering work of van der Waals[15, 13], then further elaborated by Cahn and Hilliard[3, 4].

Besides the non-zero thickness, the interface is also disturbed by a thermal motion of molecules. These undulations are called the capillary waves (CW), developed by Buff, Lovett and Stillinger[2].

2 Nucleation rates

We say that a thermodynamic system that consists of the liquid, its vapor and a phase interface between them is in a saturated state if it is in thermodynamic equilibrium stable to all fluctuations. Both phases in this system have the same temperature T, pressure p and chemical potential μ. The degree of its supersaturation is defined as

$$S = \exp\left(\frac{\mu_\mathrm{V} - \mu_{\mathrm{V}\infty}}{k_\mathrm{B}T}\right), \tag{1}$$

where μ_V is the chemical potential of the vapor, $\mu_{\mathrm{V}\infty}$ is the chemical potential (of the vapor) of the saturated state and k_B is the Boltzmann constant.

These droplets can be also viewed as clusters of n-mers. Microscopically, these clusters grow or shrink if a monomer join or leave the cluster. Of all the n-mers, important is the so-called critical cluster with n_c numbers of molecules which has the same probability to grow as to shrink. The number of droplets formed in unit of volume per unit of time is called nucleation rate, J, and is given by

$$J_\mathrm{ic} \quad = \quad \frac{\rho_\mathrm{V}\rho_{V\infty}}{\rho_{\mathrm{L}\infty}}\sqrt{\frac{2\sigma_\infty}{\pi M}}\exp\left(-\frac{\Delta\Omega}{k_\mathrm{B}T}\right), \tag{2}$$

where ρ_V is the density of the vapor phase, $\rho_{\mathrm{V},\infty}$ density of the vapor phase of the saturated state, $\rho_{\mathrm{L},\infty}$ is the liquid density and σ_∞ surface tension of the saturated state. The work of formation $\Delta\Omega$ of the critical cluster according to the CNT is given by

$$\Delta\Omega = \frac{16\pi}{3}\frac{\sigma_\infty^3}{\Delta p^2}. \tag{3}$$

3 Density gradient theory

The work of formation of a droplet according to the DGT is a volume integral consisting of homogeneous and gradient part,

$$\Delta\Omega(\rho) = \int_0^\infty \left[\Delta\omega^0(\rho) + \frac{1}{2}c\left(\frac{\mathrm{d}\rho}{\mathrm{d}r}\right)^2\right] 4\pi r^2 \mathrm{d}r, \tag{4}$$

where r is the radial coordinate; it is the distance from the center of the droplet. In Eq. (4), the grand-potential density difference is given by

$$\Delta\omega^0 = f^0 - \rho\mu_\mathrm{V} + p_\mathrm{V}, \tag{5}$$

where μ_V is the chemical potential of the vapor phase and p_V is the pressure of the vapor phase.

The work of formation (4) has its saddle point for the density profile corresponding to the so-called critical cluster; the critical cluster has its minimum with respect to all the properties in the functional space except for its size where it is maximal. This point is described by variating Eq. (4) which leads to an Euler–Lagrange equation,

$$\frac{\mathrm{d}^2\rho}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}\rho}{\mathrm{d}r} = \frac{1}{c}\Delta\mu(\rho), \tag{6}$$

where $\Delta\mu = \mu^0(\rho) - \mu_\mathrm{V}$ is the difference of local chemical potential and chemical potential of the homogeneous (vapor) phase.

To solve the problem, two boundary conditions are needed,

$$\rho(r \to \infty) = \rho_\mathrm{V}, \quad \frac{\mathrm{d}\rho}{\mathrm{d}r}(0) = 0. \tag{7}$$

4 PC–SAFT

In this work, DGT was combined with the PC–SAFT[7, 8] equation of state, which belongs to the family of modern SAFT-type EoSs being developed since 1990's[5].

The SAFT-type EoSs are defined in the form of the Helmholtz free energy, which is given as a sum of the ideal gas part F_id evaluated from the isobaric heat capacity of the ideal gas and the residual part F_res defined by the SAFT terms. An important advantage of the SAFT-type equations is that the residual part of the Helmholtz energy is defined as a sum of individual contributions accounting for various types of intermolecular interactions, e.g., the van der Waals attractions, Coulombic forces or hydrogen bonds. In PC–SAFT, the residual part consists of the hard chain contribution F_hc representing the reference fluid and the perturbation contribution F_disp.

5 Capillary waves

Thermal motion of molecules causes that the interface is not a plain surface but it is rather disturbed by the CW[2]. Meunier[11] developed so-called mode-coupling theory that adds the CW-broadening effect to the surface tension of the planar phase interface.

We define a "bare" surface tension σ_bare, as surface tension cleared of the CW. The bare surface tension still accounts for some thermal motion of molecules, but only those that are caused by non-zero thickness of the interface. The experimental surface tension σ_exp is then given by a difference of this bare surface tension and the CW contribution[11],

$$\sigma_\mathrm{exp} = \sigma_\mathrm{bare} - \frac{3}{8\pi}k_\mathrm{B}Tq_\mathrm{max}^2, \tag{8}$$

where q_max is the upper cutoff of the wave-numbers considered. q_max describes CW with the smallest wave-length that are not already accounted in the DGT[11]. The CW cause decrease of the surface tension; they are a result of an entropic effect which, in a product with temperature subtracted from internal energy, decreases energy.

Using the mode-coupling theory, $q_{\max}$ can be derived as

$$q_{\max} = \frac{1}{5.15\xi^-}. \tag{9}$$

Both correlation lengths obey a scaling law

$$\xi^\pm = \xi_0^\pm \left|1 - \frac{T}{T_\mathrm{c}}\right|^{-\nu}, \tag{10}$$

where $\nu = 0.63$ is a critical exponent.

Using Eq. (9), σ_bare in (8) can be expressed using σ_exp as[6],

$$\sigma_\mathrm{bare} = \sigma_\mathrm{exp}\left(1 + \frac{3}{8\pi}\frac{T}{T_\mathrm{c}}\frac{1}{2.70}\right). \tag{11}$$

Aside from the upper cutoff $q_{\max}$, we can also define a lower cutoff of the wave-number $q_{\min}$, the largest wave-lengths that can fit into a nanoscopic droplet. We will take a simple idea that the longest wave-length should twice fit into the circumference of the droplet,

$$2\lambda_{\max} = 2\pi r_\mathrm{s}, \tag{12}$$

where radius of the droplet is represented by the radius of the surface of tension, r_s. The lower cutoff of the wave-vector would be then

$$q_{\min} = \frac{2\pi}{\lambda_{\max}} = \frac{2}{r_\mathrm{s}}. \tag{13}$$

6 Method

Equation (6) with boundary conditions (7) forms a boundary value problem. This simply looking problem has several difficulties: the density profile near the gaseous phase has sharp shape; its slope changes abruptly from the very steep decline to an almost constant profile. Second problem is that for large droplets the density profile in the interior of the droplet changes only negligibly and is almost constant. This causes significant cumulation of numerical errors.

We developed an original algorithm based on the simple but robust shooting method combined with the halving method. The boundary value problem was modified into an initial value problem by estimating (shooting) the density in the center of the droplet ρ_0, so the initial conditions were

$$\rho(0) = \rho_0, \quad \frac{\mathrm{d}\rho}{\mathrm{d}r}(0) = 0. \tag{14}$$

The initial value problem was then solved using the Runge–Kutta method (MATLAB function ODE45). The other boundary condition, $\rho(R) = \rho_\mathrm{V}$ for R sufficiently high, had to be matched. However, we encountered several numerical problems.

Density profiles solving iterations of Eq. (6) with conditions (14) showed an oscillatory behavior around a density in the physically unstable region. For first 3 nm, the supposed

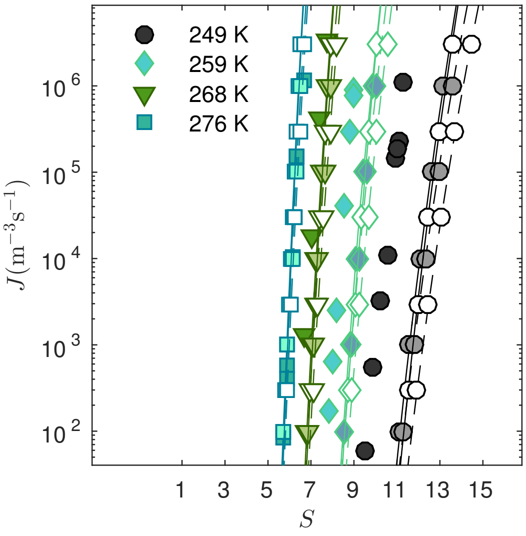


Figure 1: Nucleation rates J of n-heptane as functions of supersaturation S for various temperatures in logarithmic scale. One color always corresponds to one temperature. Comparison of theoretical nucleation rates (2) with experiments by Rudek et al. [14]. Solid lines correspond to the DGT, dashed lines to the CNT. Lines with light-color symbols corresponds to the PC–SAFT Eos, with white symbols to the PR EoS. Experimental data are depicted by dark-color symbols.

decline from liquid density $\rho_\mathrm{L} = 713\,\mathrm{kg/m^3}$ towards the vapor density $\rho_\mathrm{V} = 0.3534\,\mathrm{kg/m^3}$ can be observed. Then the density rises and oscillates around density $\rho \simeq 250\,\mathrm{kg/m^3}$ which lies in numerically stable but physically unstable region.

Therefore, the second condition of (14) had to be changed to

$$\min_{r<R} \rho(r) = \rho_\mathrm{V}, \tag{15}$$

which made it impossible to use Newton method for the initial guess iteration.

Moreover, the secant method failed to converge with the combination of ODE45, because this method uses adaptive step-size. Then, when the solution is approached, the error does not monotonically decrease and the error analysis is rather chaotic. However, ODE45 is very fast solver and we wanted to take advantage of it, so we changed the secant method to the slower, but more robust, halving method.

7 Results and discussions

We performed DGT and CNT computations for two alkanes, n-heptane and n-octane for temperatures that correspond to experimental data for the nucleation rates by Rudek et al. [14], Hung et al. [9], Luijten [10], Viisaanen et al. [16] and Wagner and Strey [17],

- n-heptane (C7): 249 K, 259 K, 268 K, 276 K,
- n-octane (C8): 241 K, 248 K, 258 K, 267 K, 287 K, 298 K, 302 K,

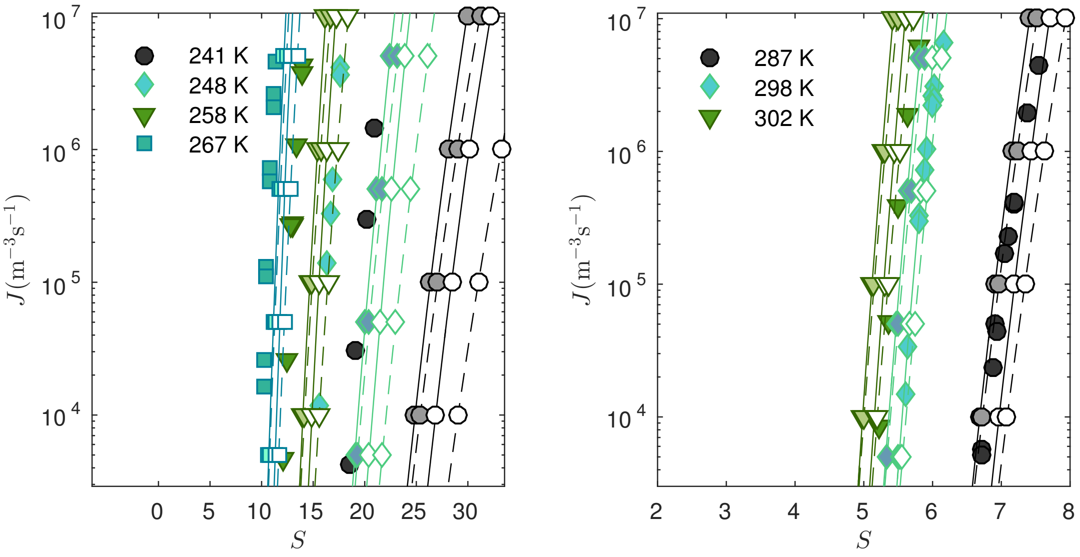


Figure 2: Nucleation rates J of n-octane as functions of supersaturation S for various temperatures in logarithmic scale. Experimental data are given by Rudek et al. [14]. Markers and lines are same as in Fig. 1.

Calculations were done using the PC–SAFT EoS and the Peng–Robinson[12] (PR) EoS. Influence parameters c were computed using the "bare" surface tension of the planar phase interface that was obtained from the experimental surface tension using Eq. (11). Using Eq. (11), we removed the effect of the capillary waves from the input of the DGT. The reason is that the DGT does not contain capillary waves and it is more consistent to exclude them.

Computations were performed using the algorithm described in Sec. 6. Program was implemented in MATLAB. The nucleation rates were evaluated using Eq. (2), DGT works of formation were computed using Eq. (4).

Figures 1 – 2 show nucleation rates depending on the supersaturations in logarithmic scales computed using the combinations of the DGT and the CNT and PC–SAFT and PR EoSs compared with the experimental data. One color and one symbol always correspond to one temperature. Lines depict the theoretical data: solid lines with light-colored symbols are the DGT–PC–SAFT computations, dashed lines with light-colored symbols correspond to CNT–PC–SAFT, solid lines with white symbols to DGT–PR, dashed lines with white symbols to CNT–PR.

Slope of experimental data quite nicely corresponds to the slope of the results, however, all the lines are somewhat shifted.

Another way how to evaluate nucleation rates with respect to experiments is via their ratios. Figure 3 shows ratios of theoretical and experimental nucleation rates corresponding to the experimental temperatures and supersaturations as functions of inversed reduced temperatures. Numerical data were therefore twice interpolated using cubic spline; first, $\ln J$ were interpolated so that their (logarithms of) supersaturations match $\ln S_{\mathrm{exp}}$ on two isotherms neighboring T_{exp}. Then these two nucleation rates were again interpolated with respect to temperature to match T_{exp}.

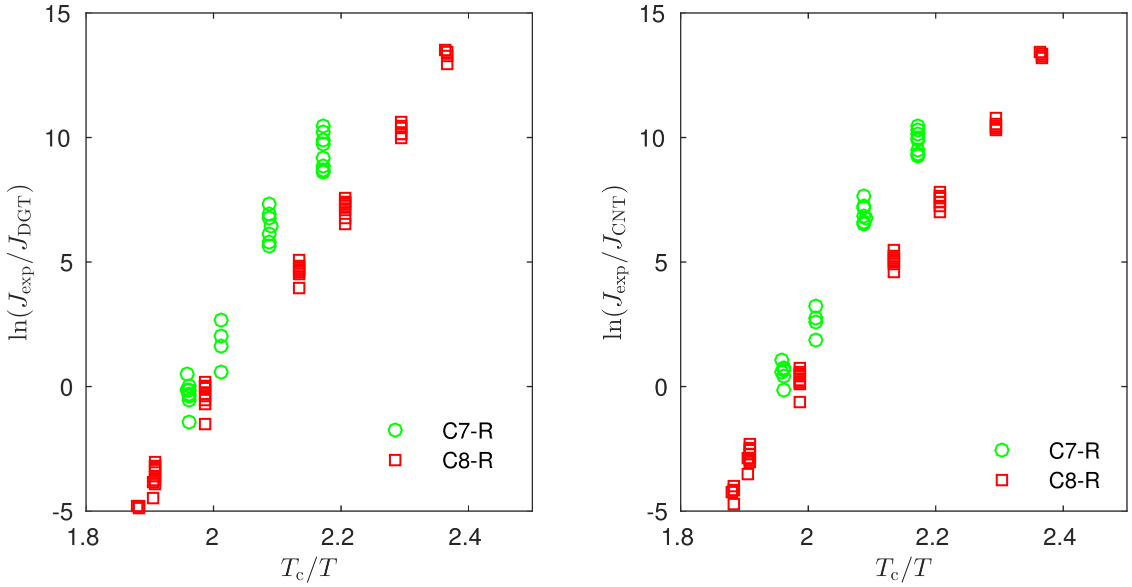


Figure 3: Ratios of nucleation rates of both n-alkanes computed using the DGT (left) and the CNT (right) and experimental nucleation rates as functions of inverse reduced temperature. Experimental data are given by Rudek et al. [14] (R).

In Fig. 3 ratios of nucleation rates computed using the DGT and the CNT and experimental nucleation rates are depicted. In both plots, one color and one symbol correspond to one group of experimental data.

The x axis is chosen so that the data from different substances are more compact and inhibit differences between the substances. Ideally, the symbols should lie on the constant line 1 (10^0), which is clearly not the case. Therefore there is a temperature trend which is not described by neither of the theories.

Our last two results discuss the effect of the capillary waves. Figure 4 shows the surface tension depending on the Laplace pressure Δp computed using the DGT and combination of the two EoSs, PR and PC–SAFT, and different types of influence parameters. We computed n-heptane at the temperature $T = 276\mathrm{K}$. Laplace pressure is connected with the radius of the droplet r via Young–Laplace equation. Therefore, $\Delta p = 0$ correspond to infinite radius, i.e. the planar phase interface with value of surface tension σ_∞. In the other end is the spinodal where surface tension vanishes $\sigma = 0$. Spinodal is not possible to reach with the DGT, therefore we used a cubic interpolation to have a continuous line.

From Fig. 4 can be seen that $\sigma_{\mathrm{bare}} > \sigma_{\mathrm{exp}}$. Surface tension of both EoSs start at either σ_{bare} or at σ_{exp} for Laplace pressure $\Delta p = 0$, but then tends to the spinodal value given by the particular EoS. The inset figure shows a detail of the trend. It can be seen that capillary waves have an effect on the surface tension.

Figure 5 shows nucleation rates depending on the supersaturation of n-heptane at the temperature $T = 276\mathrm{K}$. Again, combination of the two EoSs and the two influence parameters was used, same as in Fig. 4. The stars depict the experimental data. Erasing the capillary waves lowers the nucleation rates. The effect is quite large, which proves importance of such a treatment. Of course, adding the effect of capillary waves to the

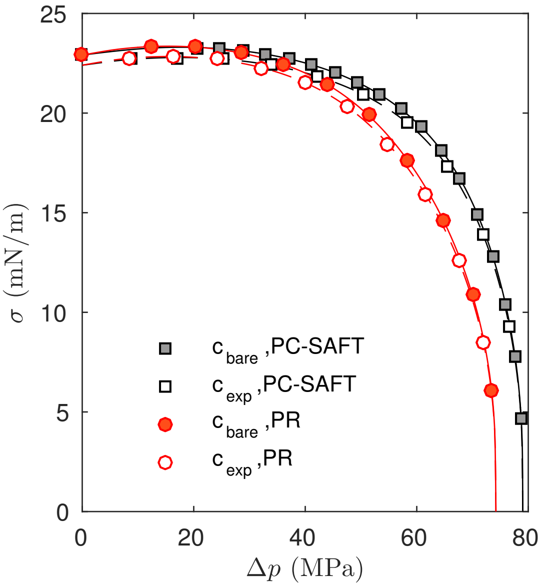


Figure 4: Surface tensions σ as functions of Laplace pressure Δp of n-heptane at temperature $T = 276\,\mathrm{K}$. Inset is the detail of the beginning where the lines cross. Lines are results of the DGT computations with PR and PC–SAFT EoS using c obtained from experimental surface tension σ_{exp} and bare surface tension σ_{bare}, Eq. (11).

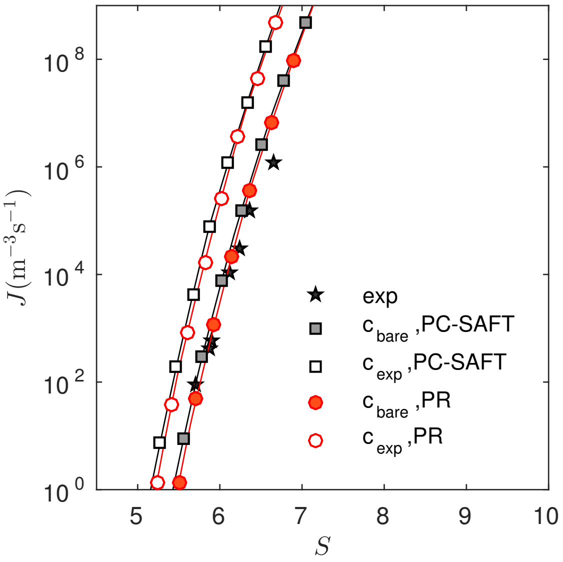


Figure 5: Nucleation rates J as functions of supersaturation S of n-heptane at temperature $T = 276\,\mathrm{K}$. Lines are the same as in Fig. 4. Nucleation rates are also compared to the experimental data by Rudek et al. [14] (black stars).

nucleation rates via surface tension and work of formation is also important. This will be a topic of the next article.

8 Conclusions

We computed nucleation rates of two alkanes: n-heptane and n-octane using the density gradient theory and the classical nucleation theory. For the calculations an original algorithm was developed in MATLAB to solve an Euler–Lagrange equation which with boundary conditions forms a boundary value problem. During the process of solution, we had to develop several numerical enhancements because the problem was numerically unstable. We used a traditional cubic Peng–Robinson EoS and a modern PC–SAFT EoS.

We also studied the effect of the CW on the surface tension and nucleation rates. We found out that in the range of the experimental data, the droplets are so small that the longest wave-lengths of the CW that could fit into a nanoscopic droplet, are already too small for the CW description. This means that the only thermal motion of the molecules is already accounted for in the DGT description and the CW play no role. Therefore, for a consistent description of droplets by the DGT it is necessary to remove CW from the input.

There are three contributions accounted for the shape of the surface tension of a droplet which is depicted in Fig. 4. First is the Tolman's linear effect which accounts for the mild increase of surface tension (for low Δp). Second, quadratic, effect accounts for the curvature bending down. This one is caused by a non-zero thickness of the interface. Third is the effect of the CW. Extraction of the CW increased the surface tension since $\sigma_{\mathrm{bare}} > \sigma_{\mathrm{exp}}$. The effect was rather significant on the surface tension and nucleation rates, so it is important to consider CW removal from the influence parameter in the DGT description.

References

[1] R. Becker and W. Döring. *Kinetische behandlung der keimbildung in übersättigten dämpfen.* Ann. Phys. **416** (1935), 719–752.

[2] F. P. Buff, R. A. Lovett, and J. F. H. Stillinger. *Interfacial density profile for fluids in the critical region.* Phys. Rev. Lett. **15** (1965), 621–623.

[3] J. W. Cahn and J. E. Hilliard. *Free energy of a nonuniform system. i. interfacial free energy.* J. Chem. Phys. **28** (1958), 258–267.

[4] J. W. Cahn and J. E. Hilliard. *Free energy of a nonuniform system. iii. nucleation in a two-component incopressible fluid.* J. Chem. Phys. **31** (1959), 688–699.

[5] W. Chapman, K. Gubbins, G. Jackson, and M. Radosz. *Saft: Equation-of-state solution model for associating fluids.* Fluid Phase Equilibr. **52** (1989), 31–38.

[6] J. Gross. *A density functional theory for vapor-liquid interfaces using the pcp-saft equation of state.* J. Chem. Phys. **131** (2009).

[7] J. Gross and G. Sadowski. *Application of perturbation theory to a hard-chain reference fluid: an equation of state for square-well chains.* Fluid Phase Equilibr. **168** (2000), 183–199.

[8] J. Gross and G. Sadowski. *Perturbed-chain saft: An equation of state based on a perturbation theory for chain molecules.* Ind. Eng. Chem. Res. **40** (2001), 1244–1260.

[9] C.-H. Hung, M. J. Krasnopoler, and J. L. Katz. *Condensation of a supersaturated vapor. viii. the homogeneous nucleation of n-nonane.* J. Chem. Phys. **90** (1989).

[10] C. Luijten. *Nucleation and Droplet Growth at High Pressure.* PhD thesis, Technische Universiteit Eindhoven, (1998).

[11] J. Meunier. *Liquid interfaces : role of the fluctuations and analysis of ellipsometry and reflectivity measurements.* J. Phys. **48** (1987), 1819–1831.

[12] D.-Y. Peng and D. B. Robinson. *A new two-constant equation of state.* Ind. Eng. Chem., Fundam., **15** (1976), 59–64.

[13] J. Rowlinson. *Translation of jd van der waals"“the thermodynamik theory of capillarity under the hypothesis of a continuous variation of density”.* J. Stat. Phys. **20** (1979), 200–244.

[14] M. M. Rudek, J. A. Fisk, V. M. Chakarov, and J. L. Katz. *Condensation of a supersaturated vapor. xii. the homogeneous nucleation of the n-alkanes.* J. Chem. Phys. **105** (1996).

[15] J. D. van der Waals. *The thermodynamic theory of capillarity under the hypothesis of a continuous variation of density.* Verhandel. Konink. Akad. Weten. **1** (1893).

[16] Y. Viisanen, P. E. Wagner, and R. Strey. *Measurement of the molecular content of binary nuclei. iv. use of the nucleation rate surfaces for the n-nonane-n-alcohol series.* J. Chem. Phys. **108** (1998), 4257.

[17] P. E. Wagner and R. Strey. *Measurements of homogeneous nucleation rates for n-nonane vapor using a two-piston expansion chamber.* J. Chem. Phys. **80** (1984), 5266.

[18] Y. B. Zeldovich. *Contribution to the theory of the formation of a new phase.* Zh. Teor. Eksp. Fiz **12** (1942), 525–538.

[19] Y. B. Zeldovich. *On the theory of new phase formation: cavitation.* Acta physicochim. URSS **18** (1943), 1–22.

Statistical Methods and Data Analysis for Mortality Comparison between Patients with Myocardial Infarction in Syria and Czech

Issam Salman

2nd year of PGS, email: issam.salman@fjfi.cvut.cz
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jiří Vomlel, Department of Decision-Making Theory
Institute of Information Theory and Automation, CAS

Abstract. Acute Myocardial Infarction (AMI) is the leading cause of death in most countries. Our research reported in this paper is twofold. In the first part of the paper we use standard statistical methods to analyze medical records of patients suffering myocardial infarction from the third world Syria and a developed country - the Czech Republic. One of our goals is to find whether there are statistically significant differences between the two countries. In the second part of the paper we compare different predictive models of hospital mortality for patients with AMI. All results presented in this paper are based on a real data about 603 patients from a hospital in the Czech Republic and about 184 patients from two hospitals in Syria. Although the learned models may be specific for the data we also draw more general conclusions that we believe are generally valid.

Keywords: Machine Learning, Data mining, Data analysis, Classification, Bayesian networks, Acute Myocardial Infarction

1 Introduction

Acute myocardial infarction (AMI) is commonly known as a heart attack. A heart attack occurs when an artery leading to the heart becomes completely blocked and the heart doesn't get enough blood or oxygen. Without oxygen, cells in that area of the heart die. AMI is responsible for more than a half of deaths in most countries worldwide. Its treatment has a significant socioeconomic impact.

One of the main objectives of our research is to design, analyze, and verify a predictive model of hospital mortality based on clinical data about patients. A model that predicts well the mortality can be used, for example, for the evaluation of the medical care in different hospitals. The evaluation based on mere mortality would not be fair to hospitals that treat often complicated cases. It seems better to measure the quality of the health care using the difference between predicted and observed mortality.

A related work was published by krumholz [1]. The authors analyze the mortality data in U.S. hospitals using the logistic regression model.

2 Data

Our data-set contains data about 787 patients characterized by 24 variables. 603 patients of them are from Czech Republic and 184 are from Syria. The attributes are listed in the Table 1. Most of the attributes are real valued, four attributes are nominal. Only a subset of attributes was measured for the Syrian patients. Most records contain missing values, i.e., for most patients only some attribute values are available. The thirty days mortality is recorded for all patients.

In the Czech Republic the results of blood tests are reported in millimoles per liter of blood. In Syria some of the measurements are reported in milligrams per liter and some in millimoles per liter. We standartize all measurements to the millimoles per liter scale.

Attribute	**Code**	**type**	**value range in data**	**Country**
Age	AGE	real	[23, 94]	SYR, CZ
Height	HT	real	[145, 205]	CZ
Weight	WT	real	[35, 150]	CZ
Body Mass Index	BMI	real	[16.65, 48.98]	CZ
Gender	SEX	nominal	{male, female}	SYR, CZ
Nationality	NAT	nominal	{Czech, Syrian}	SYR, CZ
STEMI Location	STEMI	nominal	{inferior, anterior, lateral}	SYR, CZ
Hospital	Hospital	nominal	{CZ, SYR1, SYR2}	SYR, CZ
Kalium	K	real	[2.25, 7.07]	CZ
Urea	UR	real	[1.6, 61]	SYR, CZ
Kreatinin	KREA	real	[17, 525]	SYR, CZ
Uric acid	KM	real	[97, 935]	SYR, CZ
Albumin	ALB	real	[16, 60]	SYR, CZ
HDL Cholesterol	HDLC	real	[0.38, 2.92]	SYR, CZ
Cholesterol	CH	real	[1.8, 9.9]	SYR, CZ
Triacylglycerol	TAG	real	[0.31, 11.9]	SYR, CZ
LDL Cholesterol	LDLC	real	[0.261, 7.79]	SYR, CZ
Glucose	GLU	real	[2.77, 25.7]	SYR, CZ
C-reactive protein	CRP	real	[0.3, 359]	SYR, CZ
Cystatin C	CYSC	real	[0.2, 5.22]	SYR, CZ
N-terminal prohormone of brain natriuretic peptide	NTBNP	real	[22.2, 35000]	CZ
Troponin	TRPT	real	[0, 25]	CZ
Glomerular filtration rate (based on MDRD)	GFMD	real	[0.13, 7.31]	CZ
Glomerular filtration rate (based on Cystatin C)	GFCD	real	[0.09, 7.17]	CZ

Table 1: Attributes

3 Preliminary Statistical Analysis

For a preliminary statistical analysis [2] we randomly choose 150 Czech patients and 150 Syrian patients from our dataset so that we had two groups of equal size. We selected a subset of attributes presented in both groups, namely, we considered these variables: age, nationality, gender, STEMI location, and mortality.

Since STEMI location is nominal and takes three states for most experiments we transform it to three binary variables STEMI.inf, STEMI.ant, and STEMI.lat. The nationality is encoded by a binary variable, where 0 means Czech and 1 means Syrian. The Gender is encoded by a binary variable where 0 denotes a man, while 1 stands for a female. The mortality is also encoded as a binary variable, where 0 means that the patient survived 30 days, while 1 means that he/she did not.

Already from Figure 1, where the histogram of the age values is presented, we can see that from patients that didn't survive a high percentage are young patients from Syria.

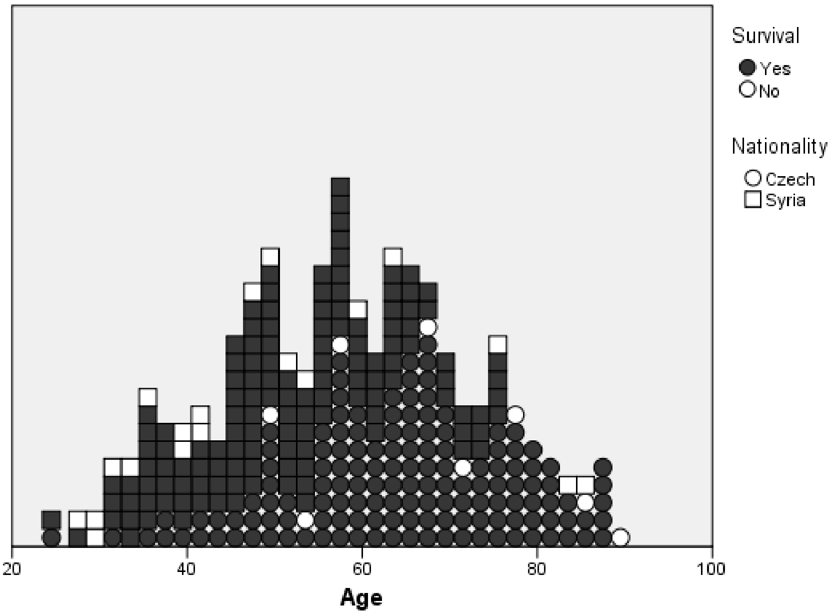


Figure 1: Histogram of the age values

In Table 2 we present the correlation matrix (since it is symmetric we present only the upper triangular part without the diagonal). The correlations that were statistically significant (at the level 0.05) are highlighted. We can see that there is a negative correlation between the age of the patients and the nationality (the Czech Republic is encoded using 0 and Syria 1). Hence, the average age of the Czech patients is greater than the age of Syrian patients. There is also a significant difference between the percentage of male and female patients in each country – the percentage of female patients is 28% in the Czech Republic and 40.6% in Syria. We also observe a significant difference between mortality of the Syrian (12%) and Czech (5.4%) patients.

Table 2: The correlations and their statistical significance

		gender	STEMI loc.	mortality	nationality
age	corr.	0.092	0.001	-0.074	**-0.460**
	sign.	0.111	0.982	0.199	**0.0001**
gender	corr.		0.034	0.018	**0.133**
	sign.		0.557	0.757	**0.021**
STEMI loc.	corr.			0.104	0.106
	sign.			0.071	0.066
mortality	corr.				**0.128**
	sign.				**0.026**

Table 3: The Chi-Square Test of conditional independence

		gender	STEMI loc.	mortality	nationality
age	value	52.63	136.7	**102.57**	**104.78**
	sign.	0.821	0.242	**0.001**	**0.001**
gender	value		1.605	0.096	**5.337**
	sign.		0.448	0.756	**0.021**
STEMI loc.	value			**10.678**	**17.173**
	sign.			**0.005**	**0.0001**
mortality	value				**4.925**
	sign.				**0.026**

The standard chi-square test of conditional independence between two variables reveals (see Table 3) that there is a significant dependence between the mortality and nationality. There is also a significant dependence between the mortality and STEMI location – the patients from Syria with a lateral infarction have the lowest probability to survive.

Table 4: The Mann–Whitney U test

		age	gender	STEMI.lat	STEMI.ant	STEMI.inf	nationality
mortality	value	3100	10036	**2833**	**2952**	3567	**2860**
	sign.	0.173	0.757	**0.002**	**0.045**	0.748	**0.027**

We also performed the Mann–Whitney U test (see Table 4) to see whether there is a significant difference between mean values of mortality if we classify patients into groups according to their age, gender, STEMI location, and nationality. From the test we can conclude the patients from the Czech Republic have lower mortality than Syrians and the patients with lateral infarction have a lower probability to survive.

Finally, we learned the logistic regression model, that describes the relationship between the considered independent variables and the mortality as the dependent variable.

We have got:

$$
\begin{aligned}
&\text{logit } P(Y = 1|X = x)\\
&\quad= \beta_0 + \beta_1 x_1 + \ldots + \beta_5 x_5\\
&\quad= -2.375 - 0.006 \cdot x_1 - 0.026 \cdot x_2 + 0.613 \cdot x_3 + 0.916 \cdot x_4 - 0.489 \cdot x_5
\end{aligned}
$$

where x_1 is the age, x_2 is the gender (0 for a male and 1 for a female), x_3 is the nationality (0 for Czech and 1 for Syrian), x_4 is the STEMI.lat (0 for no, 1 for yes), and x_5 is the STEMI.ant (0 for no, 1 for yes). But only the intercept and the variable STEMI.lat appeared to be statistically significant for mortality prediction.

From the preliminary statistical analysis we can conclude that although the diverse test we used do not suggest exactly the same relations between the studied variables they mostly agree on few significant relations:

- In Syria the mortality from AIM is significantly higher than in the Czech Republic – 87.3% Syrian patients survive, while 94.7% patients from the Czech Republic survive.
- The age of patients in Syria is lower in average (the average difference is 13 years) and there is a higher prevalence of women among the patients with AIM in Syria than in the Czech Republic.
- The STEMI location is related to the mortality.

4 Machine Learning Methods

The preliminary statistical analysis studied mostly the pairwise relations only. Since the explanatory variables may combine their influence and the influence of a variable may be mediated by another variable it is worth of studying the relations of variables all together. We will do it in two steps: (1) since the mortality prediction is of our prime interest we will compare how different classifier are able to predict the mortality, (2) to get an overall picture of the relations between all variables we will learn a Bayesian network model from the collected data.

We will work with different versions of data. They depend on how we treat variables that have more than two states: (1) real valued ordinal variables, (2) discrete valued variables (with at most five states), and (3) binary variables. We will discuss the values' transformation in more detail in the next sections.

4.1 Ordinal attributes

In our data, we have several categorical variables (sometimes also called nominal variables). These are variables that have two or more categories. For example, gender is a categorical variable having two categories (male and female). But for some machine learning methods we need ordinal attributes which are attributes whose values have an ordering of values that is natural for the quantification of their impact on the class. This is satisfied by all attributes that can take only two values – even if they are nominal, e.g.

by gender (0 for male, 1 for female), mortality (0 for survived, 1 for died). In our data it seems that the ordinality can be assumed for most real valued attributes, but note that there might also exist laboratory tests whose values deviating from a normal range in both directions (i.e. both lower and higher values) may both increase the mortality. We will refer to the ordinal data as D.ORD.

4.2 Discrete attributes

Discrete variable is a variable that can take values from a finite set. Some classification methods require discrete variables. To get a statistically reliable estimates of model parameters it is advisable to keep the number of values as low as possible while still being able to express the significant relations. We performed discretization of all real-valued attributes. It is not easy to find the optimum number and the values of split points in discretization. Fortunately, there exists the Czech National Code Book that classifies numeric laboratory results, with respect to age and gender, into nine groups $1, 2, \ldots, 9$. The group 5 corresponds to standard values in the standard population. We further reduced the number of states to 5 by joining some groups together. We will refer to data in this form as D.DISCR.

4.3 Binary attributes

Binary data are data whose variables can take on only two possible states, traditionally termed 0 and 1 in accordance with the binary numeral system and Boolean algebra. In our case all laboratory tests are encoded using two binary attributes. The first attribute takes value 0 for the standard values of the test and value 1 if the values are decreased. The second attribute takes value 0 for the standard values of the test and value 1 if the values are increased. The attributes Age, Height, and Weight are removed. From the demographic group of attributes only Gender and the Body Mass Index (BMI) were kept with BMI being encoded using two binary attributes BMI high and BMI low where the BMI is greater than the mean takes value 1, otherwise it takes value 0. We will refer to data in this form as D.BIN.

4.4 Attribute Selection

Before learning a model, we preprocess the data. Ussually, one of the most useful parts of preprocessing is the attribute selection, where irrelevant attributes are removed. Attribute selection is a process by which we automatically search for the best subset of attributes in our dataset. The notion of âĂIJbestâĂİ is relative to the problem we are trying to solve, but typically means the highest accuracy. Three key benefits of performing attributes selection on our data are:

- Reduces Overfitting: Less redundant data means less opportunity to make decisions based on a noise.
- Improves Accuracy: Less misleading data means that modeling accuracy improves.
- Reduces Training Time: Less data means that algorithms train faster.

The CfsSubsetEval method of Weka [3] selects the subsets of attributes that are highly correlated with the class while having low inter-correlation. We searched the space of all subsets by a greedy best first search with backtracking. Data D after the application of this attribute selection method will be suffixed as D.AS.

4.5 Tested classifiers

For tests we used a large subset of classifiers implemented in Weka. Classifiers that performed best in the preliminary tests qualified for the final tests. In the final tests we compared following classifiers:

- **Decision tree C4.5** [4].
- **Logistic regression** [5].
- **Naive Bayes (NB) classifier** [6] assume that the value of a particular explanatory variable (attribute) is independent of the value of any other attribute given the class variable.
- **Bayesian network (BN) classifiers** (1) learned by K2 algorithm [7] - referred as BN.K and (2) Tree Augmented Naive Bayes classifier refer as BN.TAN [8].

We use the leave-one-out cross validation as the model evaluation method. It means that N separate times, the classifier is trained on all the data except for one point and a prediction is made for that point. After that the average error is computed and used to evaluate the model.

4.6 Results of experiments

For each data record classified by a classifier there are possible classification results. Either the classifier got a positive example labeled as positive (in our data the positive example is the patient survived) or it made a mistake and marked it as negative. Conversely, a negative example may have been mislabeled as a positive one, or correctly marked as negative. This defines the following metrics:

- **True Positives (TP):** number of positive examples, labeled as such.
- **False Positives (FP):** number of negative examples, labeled as positive.
- **True Negatives (TN):** number of negative examples, labeled as such.
- **False Negatives (FN):** number of positive examples, labeled as negative.

Our results are summarized in Table 5 using the following measures of the prediction quality:

- **Accuracy** measures how often the classifier makes the correct prediction. It is the ratio between the number of correct predictions and the total number of predictions.

$$ACC \;=\; \frac{TP+TN}{TP+TN+FP+FN}$$

- **Recall** is also known as sensitivity. It is the fraction of positive instances that are correctly classified as positive (rate of true positives).

$$REC = \frac{TP}{TP + FN}$$

- **Precision** is the fraction of true positives over the number of all reported positives.

$$PRE = \frac{TP}{TP + FP}$$

- **F-measure** is the harmonic mean of the precision and the recall

$$F = 2 \cdot \frac{PRE \cdot REC}{PRE + REC}$$

- **Specificity** is the fraction of true negatives over the number of all negatives.

$$SPE = \frac{TN}{FP + TN}$$

- **Area under the ROC curve (AUC)**. The ROC curve shows how the classifier can sacrifice the true positive rate (recall or sensitivity) for the false positive rate (1-specificity) by plotting the TP rate to the FP rate. In other words, it shows you how many correct positive classifications can be gained as you allow for more and more false positives. As an example, in Figure 7 we report the ROC curve for the Naive Bayes classifier with the ordinal attributes. Its area under the curve is 0.782.

In Table 5 we compare the results of different classifiers on different versions of data. The C4.5 classifier with D.DISCR has the highest accuracy of 0.942 its recall and precision are also among the best achieved. But its area under the ROC curve is very low, only 0.371, which suggests that this classifier can not be satisfactory tuned if we want to sacrifice precision to recall or vice versa.

In Figure 2 we present the tree structure of the C4.5 learned from the discrete data. It has achieved the highest accuracy from all tested classifiers. Its structure is surprisingly simple. If the patient is Czech then it is predicted to survive if the patient is Syrian then the LDL cholesterol value should be checked. If it is below 4.78 then the patient is predicted to survive, otherwise, if LDL cholesterol value is between 4.78 and 6.28 then it depends on the Syrian hospital he/she is treated. If he/she is treated in the public hospital (SYR1) then he/she dies if it is the private one (SYR2) then he/she survives. If his/her LDL cholesterol values are higher than 6.28 then he/she dies (no matter what Syrian hospital he/she is treated in). The simplicity of the C4.5 classifier is in line with the general recommendation that in order to avoid over-fitting of training data the models should be as simple as possible.

The highest area under the ROC curve (AUC) was achieved by Naive Bayes classifier with the ordinal attributes. The highest value of F-measure was achieved by BN.K2 with discrete attributes selected by the method CfsSubsetEval method of Weka [3]. The learned BN model is actually also a Naive Bayes model – see Figure 3. We can conclude that there is a single winner – a classifier that would be the best in all considered criteria. Also, the classifiers differ in what variables they consider to be important for AMI mortality prediction. We believe that it is worth learning diverse classifiers since it may help medical specialists to get a deeper insight into the modeled problem.

Table 5: Results of experiments

Classifier	Criteria	D.ORD	D.ORD.AS	D.DISCR	D.DISCR.AS	D.BIN	D.BIN.AS
Naive Bayes	ACC	0.855	0.925	0.860	0.914	0.875	0.911
	AUC	**0.782**	0.722	0.744	0.781	0.695	0.717
	Recall	0.439	0.158	0.351	0.368	0.246	0.140
	Prec.	0.234	0.450	0.215	0.396	0.203	0.276
	F-measure	0.305	0.234	0.267	0.382	0.222	0.186
C4.5	ACC	0.935	0.933	**0.942**	0.921	0.926	0.927
	AUC	0.527	0.621	0.371	0.627	0.528	0.273
	Recall	0.263	0.105	0.246	0.123	0.070	0.035
	Prec.	0.625	0.750	0.875	0.368	0.444	0.333
	F-measure	0.370	0.185	0.384	0.184	0.121	0.063
LOG.REG	ACC	0.930	0.925	0.907	0.919	0.926	0.919
	AUC	0.746	0.755	0.622	0.746	0.675	0.746
	Recall	0.140	0.018	0.193	0.140	0.070	0.140
	Prec.	0.571	0.250	0.289	0.364	0.364	0.364
	F-measure	0.225	0.033	0.232	0.203	0.118	0.203
NB-Tree	ACC	0.932	0.936	0.914	0.920	0.913	0.920
	AUC	0.658	0.480	0.701	0.726	0.701	0.726
	Recall	0.211	0.228	0.228	0.088	0.070	0.088
	Prec.	0.600	0.684	0.310	0.313	0.211	0.313
	F-measure	0.312	0.342	0.263	0.137	0.105	0.137
BN.K2	ACC	NA	NA	0.886	0.918	0.900	0.926
	AUC	NA	NA	0.750	0.775	0.687	0.671
	Recall	NA	NA	0.316	0.368	0.193	0.105
	Prec.	NA	NA	0.265	0.429	0.256	0.462
	F-measure	NA	NA	0.288	**0.396**	0.220	0.171
BN.TAN	ACC	NA	NA	0.908	0.925	0.904	0.927
	AUC	NA	NA	0.721	0.768	0.653	0.642
	Recall	NA	NA	0.193	0.228	0.088	0.053
	Prec.	NA	NA	0.297	0,464	0.179	0.333
	F-measure	NA	NA	0.234	0.306	0.118	0.091

4.7 Bayesian Networks

A Bayesian network [9] is a probabilistic graphical model whose structure is defined by an acyclic directed graph and specifies conditional independence relations among model variables corresponding to the nodes of the graph. We learned Bayesian networks by the PC algorithm [10] implemented in Hugin [11].

We learned a joint BN model for all data (see Figure 4) but also one BN model for the Czech data (see Figure 5) and one BN model for the Syrian data (see Figure 6). We can see that latter two models are very different.

In the BN model of the Syrian data there is far less edges and many variables appear independent. We do not know the true reason, a conjecture might be it is because we have less data from Syria. Another reason might be that there are two groups of Syrian patients each getting different blood tests. However, few things can be observed: in the CZ model the mortality depends on TRPT attribute but it is not measured in Syria. The relation of Mortality to ALB is present in both models (despite in the SYR model its influence is mediated through Hospital variable). LDLC seems to play important role in both models (in the CZ model its influence is mediated through other variables).

```
Hospital <= 0: 0 (603.0/36.0)
Hospital > 0
|   LDLC <= 4.78: 0 (157.86/6.0)
|   4.78 < LDLC <= 6.28
|   |   Hospital <= 1: 1 (12.95/2.95)
|   |   Hospital > 1: 0 (9.0/1.0)
|   LDLC > 6.28: 1 (4.18/0.18)
```

Figure 2: Decision tree C4.5 learned from D.DISCR has the highest accuracy 0.943 of all tested models.

It is surprising that in all BN models the STEMI location is independent of mortality. This is in a contradiction to the basic statistical analysis where its influence on mortality was found significant by the The Mann–Whitney U test and by the Chi-Square Test of conditional independence.

The BN structure is reflected in some of the classifiers. For example, consider the decision tree presented in Figure 2. We can see that for the patients from Syria two considered variables are Hospital and LDLC. If we look into BN for Syrian patients (Figure 6) we can see that these two variables are exactly the variables that separate Mortality from the rest of the BN model.

Although these models are not learned to optimize the mortality prediction quality we compared the prediction quality of the joint model and of the combination of separate models. Again we used the leave-one-out method for learning and testing. From Figure 8 we can see that the combined model has a better performance, its AUC is 0.679 while the joint model has AUC = 0.5707282. These values are worst than values of the best performing classifiers but we again believe it is worth considering the BNs since they describe the whole modeled domain and the relations among variables even if they do not have a direct impact on better mortality prediction.

5 Conclusions

We used medical data on patients with AIM to compare results of (a) basic statistical analysis with (b) classification models and (c) Bayesian networks modeling the relations found in data. Although the conclusions might seem to be specific only for the used data here we report also a general observation.

We could see that the results of basic statistical analysis do not provide a complete picture of the modeled problem. Some variables may appear to be significantly correlated to the class variable but they are actually not directly correlated. Their correlation observed in data can be due to other variables that are correlated to both the analyzed variable and to the class variable. For the basic statistical analysis it is hard to reveal it. In principle the BN learning algorithms are able to discover the mediated correlation since they test not only pairwise independence but also the conditional independence given values of other variables. This is exactly what the PC algorithm does.

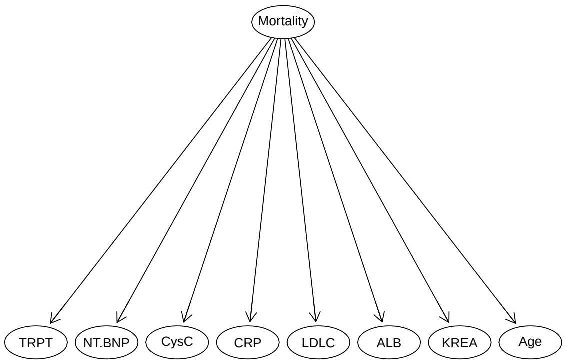


Figure 3: BN learned by BN.K2

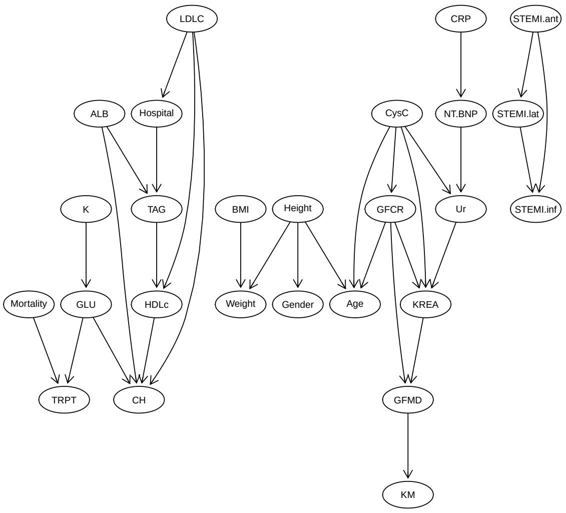
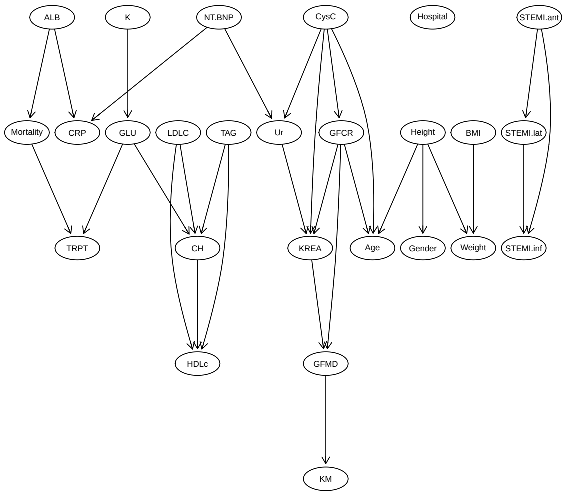


Figure 4: BN for all data



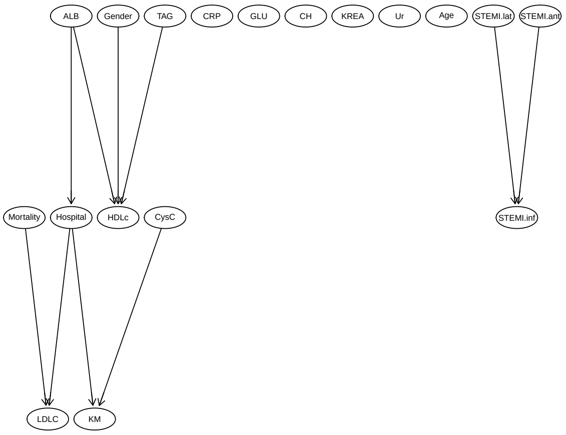


Figure 5: BN for the Czech data

Figure 6: BN for the Syrian data

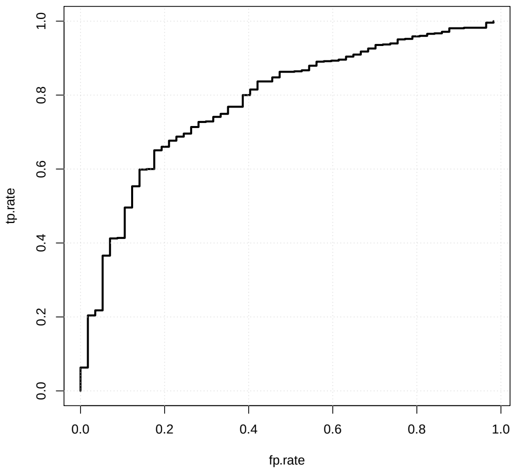


Figure 7: ROC for the Naive Bayes classifier with ordinal attributes

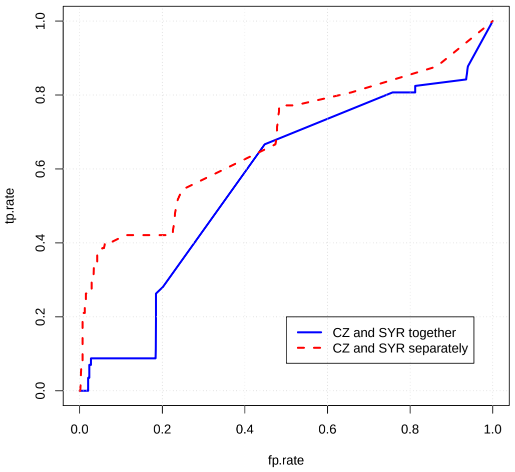


Figure 8: ROC for BNs learned by the PC algorithm

References

[1] H. M. Krumholz, S.-L. T. Normand, D. H. Galusha, J. A. Mattera, A. S. Rich, Y. Wang and Y. Wang *Risk-Adjustment Models for AMI and HF 30-Day Mortality, Methodology*. In 'Harvard Medical School, Department of Health Care Policy', (2007).

[2] L. Wasserman. *All of Statistics*. In 'Springer-Verlag New York', (2004).

[3] M. Hall and E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and H. Witten. *The WEKA Data Mining Software: An Update*. In 'ACM SIGKDD ExplorationACM SIGKDD Explorations', volume 11, Issue 1. (2009), 10–18.

[4] R. Quinlan and M. Kaufmann. *C4.5: Programs for Machine Learning*. In 'Machine Learning Journal ', volume 29, Issue 2. (1993). 131–163

[5] S. le Cessie and J.C. van Houwelingen. *Ridge estimators in logistic regression*. In 'University of Leiden, The Netherlands', volume 29, Issue 2. (1992). 131–163

[6] R. Kohavi. *Pattern Classification and Scene Analysis*. In 'Wiley-Interscience, Oxford', volume 30, Issue 1. (1973). 106–110

[7] G. F. Cooper. *A Bayesian Method for the Induction of Probabilistic Networks from Data*. In 'Machine Learning Journal', volume 9, Issue 4. (1992). 309 – 347

[8] N. Friedman, D. Geiger, and M. Goldszmidt. *Bayesian Network Classifiers*. In 'Machine Learning Journal', volume 29, Issue 2. (1997). 131–163

[9] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. In 'Morgan Kaufmann Publishers Inc. San Francisco, CA, USA', (1988).

[10] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search.* 'MIT Press, Second Edition,', (2000).

[11] A. L. Madsen, F. Jensen, U. B. Kjaerulff, and M. Lang. *The Hugin tool for probabilistic graphical models.* International Journal on Artificial Intelligence Tools, volume 14,number 3 (2005), 507–543.

Paralelní řešení úloh dvoufázového proudění v porézním prostředí pomocí MPI[*]

Jakub Solovský

1. ročník PGS, email: `jakubsolovsky@gmail.com`
Katedra matematiky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

školitel: Radek Fučík, Katedra matematiky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

Abstract. This paper deals with numerical solution of the problems of two phase flow in porous media. For solving this type of problems we propose a numerical method based on mixed hybrid finite element method. We implement several variations of this method using different approaches to solving resulting system of linear algebraic equations. We use direct and iterative solvers and describe a parallel implementation of this method based on domain decomposition using `MPI`. The method is verified on problem with known exact solution, on which we compare accuracy and computational time. Numerical experiments show that the errors are similiar for all variations of the method. The method is convergent and the experimental order of convergence is slightly less than one. There are differences in the computational time. Iterative solvers are faster and the paralelism is avantageous while using fine meshes.

Keywords: domain decomposition, two phase flow, mixed hybrid finite element method, MPI, porous media, upwind

Abstrakt. Článek se zabývá numerickým řešením úloh dvoufázového proudění v porézním prostředí. Pro řešení tohoto typu úloh navrhneme numerickou metodu vycházející z hybridní metody smíšených konečných prvků. Implementujeme několik variant této metody s použitím různých přístupů pro řešení vzniklé soustavy lineárních algebraických rovnic. Vyzkoušíme přímé i iterační řešiče a podrobně se budeme věnovat paralelizaci metody, vycházející z domain decomposition, s využitím `MPI`. Numerickou metodu testujeme na úloze, pro kterou je známé přesné řešení. Jednotlivé varianty metody porovnáváme s ohledem na chyby řešení a výpočetní čas. Ukazuje se, že chyby jsou ve všech případech téměř totožné, řešení konverguje a experimentální řád konvergence je o něco menší než jedna. Výrazné rozdíly jsou ve výpočetní náročnosti. Použití iteračních řešičů je rychlejší a přínos paralelizace se výrazně projeví na jemnějších sítích.

Klíčová slova: domain decomposition, dvoufázové proudění, hybridní metoda smíšených konečných prvků, MPI, porézní prostředí, upwind

1 Úvod

Matematické modelování dvoufázového proudění v porézním prostředí lze využít při řešení mnoha problémů, které jsou v současné době aktuální. Modelování šíření kontaminantů a látek rozpuštěných ve vodě lze využít při ochraně zdrojů pitné vody nebo při likvidaci následků nehod, při kterých došlo k úniku nebezpečných látek. U těchto úloh až na několik

[*] Tato práce byla podpořena grantem SGS11/161/OHK4/3T/14

speciálních případů není známé přesné řešení, ale s pomocí numerických metod lze nalézt poměrně dobré aproximace řešení.

V této práci se zaměříme na testování představené numerické metody na úloze dvoufázového proudění, pro kterou je známé přesné řešení. Při výpočtech na jemných sítích je důležitá nejen přesnost výsledku, ale také doba výpočtu. Proto se budeme věnovat také paralelizaci numerického řešení, která umožní rychle řešit i úlohy na velmi jemných sítích.

2 Numerická metoda

Metodu lze použít pro řešení soustavy n parciálních diferenciálních rovnic ve tvaru:

$$\sum_{j=1}^{n} N_{i,j}\frac{\partial Z_j}{\partial t} + \sum_{j=1}^{n} \boldsymbol{u}_{i,j}\cdot\nabla Z_j + \nabla\cdot\left[m_i\left(-\sum_{j=1}^{n}\boldsymbol{D}_{i,j}\nabla Z_j + \boldsymbol{w}_i\right)\right] = f_i, \tag{1}$$

kde $Z_j = Z_j(\boldsymbol{x},t)$ označují neznámé funkce $\forall t > 0$, $\forall \boldsymbol{x} \in \Omega$, kde $\Omega \subset \mathbb{R}^d$ je oblast a d je dimenze prostoru ($d \in \{1,2,3\}$). $N_{i,j}, m_i$ jsou skalární koeficienty, $\boldsymbol{u}_{i,j}, \boldsymbol{w}_i$ vektorové koeficienty a $\boldsymbol{D}_{i,j}$ symetrické tenzory druhého řádu. Dále uvedeme jen shrnutí nejdůležitějších kroků při odvození numerické metody, podrobnosti lze nalézt v [9].

Jako rychlostní člen $\boldsymbol{v}_i$ označíme výraz:

$$\boldsymbol{v}_i = -\sum_{j=1}^{n}\boldsymbol{D}_{i,j}\nabla Z_j + \boldsymbol{w}_i. \tag{2}$$

Uvažovanou oblast $\Omega \subset \mathbb{R}^d$ rozdělíme na elementy. Tyto elementy budou úsečky v jedné dimenzi, trojúhelníky ve dvou dimenzích a čtyřstěny ve třech dimenzích.

2.1 Aproximace rychlostních členů

Budeme předpokládat, že funkce $\boldsymbol{v}_i$ a $\boldsymbol{u}_{i,j}$ patří do funkcionálního prostoru $H(\mathrm{div},\Omega)$ a zároveň, že je můžeme na každém elementu aproximovat funkcemi z Raviartova-Thomasova-Nédélecova prostoru $RTN_0(K)$ [1]. Aproximace funkcí $\boldsymbol{v}_i$ a $\boldsymbol{u}_{i,j}$ budou mít na každém elementu K tvar:

$$\boldsymbol{v}_i = \sum_{E\in\partial K} v_{i,K,E}\boldsymbol{\omega}_{K,E}, \qquad \boldsymbol{u}_{i,j} = \sum_{E\in\partial K} u_{i,j,K,E}\boldsymbol{\omega}_{K,E},$$

kde $\boldsymbol{\omega}_{K,E}$ jsou bazické funkce prostoru $RTN_0(K)$. S využitím vyjádření $\boldsymbol{v}_i$ v bázi prostoru $RTN_0(K)$ a definice rychlostního členu (2) můžeme jednotlivé koeficienty $v_{i,K,E}$ určit pomocí $Z_{j,K}$ a $Z_{j,F}$:

$$v_{i,K,E} = \sum_{j=1}^{n}\left(b_{i,j,K,E}Z_{j,K} - \sum_{F\in\partial K} b_{i,j,K,E,F}Z_{j,F}\right) + w_{i,K,E}, \tag{3}$$

kde $\{B_{i,j,K}\}_{E,F} = \int_K \boldsymbol{\omega}_{K,F}^T\boldsymbol{D}_{i,j}^{-1}\boldsymbol{\omega}_{K,E}$, $\boldsymbol{b}_{i,j,K}$ je inverze $B_{i,j,K}$ a $w_{i,K,E}$ je koeficient projekce vektoru $\boldsymbol{w}_i$ do prostoru $RTN_0(K)$ podle bazického vektoru $\boldsymbol{\omega}_{K,E}$.

2.2 Aproximace zákonů zachování a časová diskretizace

Každou z rovnic (1) nejprve integrujeme přes element K a dále upravujeme. Metoda je semiimplicitní v čase. Ve všech výrazech, které jsou v Z_j lineární použijeme hodnoty z aktuální časové vrstvy, u nelineárních výrazů použijeme hodnoty z předchozí časové vrstvy (metoda zamrzlých koeficientů). Po provedení všech úprav získáme:

$$\sum_{j=1}^{n} N^k_{i,j,K} \frac{|K|_d}{\Delta t_k} \left(Z^{k+1}_{j,K} - Z^k_{j,K}\right) + \sum_{j=1}^{n} \sum_{E\in\partial K} u^k_{i,j,K,E}(Z^{k+1}_{j,E} - Z^{k+1}_{j,K}) +$$
$$+ \sum_{E\in\partial K} m^{k,upw}_{i,E} \left[\sum_{j=1}^{n} \left(b^k_{i,j,K,E} Z^{k+1}_{j,K} - \sum_{F\in\partial K} b^k_{i,j,K,E,F} Z^{k+1}_{j,F}\right) + w^k_{i,K,E}\right] = |K|_d f^k_{i,K}, \quad (4)$$

kde $|K|_d$ je d - rozměrná Lebesgueova míra elementu K a Δt_k je časový krok. Hodnota $m^{k,upw}_{i,E}$ je společná hodnota na hraně E zvolená podle směru $\boldsymbol{v}_i$ [9].

2.3 Bilance na vnitřních hranách a okrajové podmínky

K soustavě rovnic (4) popisující chování na jednotlivých elementech musíme dodat další, které mezi sebou elementy provází. Ty vyjadřují bilanci hmoty na hranicích elementů:

$$\sum_{\ell=1}^{2} \left(\sum_{j=1}^{n} \left(b^k_{i,j,K_\ell,E} Z^{k+1}_{j,K_\ell} - \sum_{F\in\partial K_\ell} b^k_{i,j,K_\ell,E,F} Z^{k+1}_{j,F}\right) + w^k_{i,K_\ell,E}\right) = 0. \quad (5)$$

Rovnice (5) popisují pouze vnitřní hrany. Pro Dirichletovu podmínku dodáme rovnice, které předepisují hodnoty stopy veličiny na hranici. Zadání Neumannovy podmínky z vlastností bazických funkcí prostoru odpovídá určení koeficientu $v_{i,K,E}$ definovaného vztahem (3).

2.4 Konstrukce matice soustavy

Z rovnice (4) lze na každém elementu vyjádřit hodnoty $Z^{k+1}_{j,K}$ pomocí $Z^{k+1}_{j,F}$ ve tvaru:

$$\boldsymbol{Z}^{k+1}_K = \sum_{F\in\partial K} \boldsymbol{Q}^{-1}_K \boldsymbol{R}_{K,F} \boldsymbol{Z}^{k+1}_F + \boldsymbol{Q}^{-1}_K \boldsymbol{R}_K, \quad (6)$$

kde $\boldsymbol{Z}^{k+1}_K$ je vektor, který obsahuje jednotlivé neznámé v čase t_{k+1} a matice $\boldsymbol{Q}_K$, $\boldsymbol{R}_{K,F}$ a vektor $\boldsymbol{R}_K$ mají následující význam:

$$(Q_K)_{i,j} = |K|_d N^k_{i,j,K} + \Delta t_k \left(-\sum_{E\in\partial K} u^k_{i,j,K,E} + \sum_{E\in\partial K} m^{k,upw}_{i,E} b^k_{i,j,K,E}\right), \quad (7)$$

$$(R_{K,F})_{i,j} = \Delta t_k \left(\sum_{E\in\partial K} m^{k,upw}_{i,E} b^k_{i,j,K,E,F} - u^k_{i,j,K,F}\right), \quad (8)$$

$$(R_K)_i = |K|_d \sum_{j=1}^{n} N^k_{i,j,K} Z^k_{j,K} + \Delta t_k \left(|K|_d f^k_{i,K} - \sum_{E\in\partial K} m^{k,upw}_{i,E} w^k_{i,K,E}\right). \quad (9)$$

Takto vyjádřené střední hodnoty $Z_{j,K}^{k+1}$ dosadíme do bilance toku na hranách (5), do zadání Neumannových okrajových podmínek a přidáme rovnice pro Dirichletovy okrajové podmínky. Jako výsledek dostaneme celkovou soustavu lineárních rovnic pro $Z_{j,F}^{k+1}$. Tuto soustavu lze symbolicky zapsat:

$$\boldsymbol{M}\boldsymbol{Z}^{k+1} = \boldsymbol{b}, \tag{10}$$

kde $\boldsymbol{Z}^{k+1} = \{\{Z_{j,F}^{k+1}\}_{j=1}^{n}\}_{F\in E_h}$ je vektor neznámých.

2.5 Algoritmus

Na závěr ještě pro přehlednost uvedeme shrnutí algoritmu.

2.5.1 Inicializace:

1a. Nastav $t_0 = 0$, $k = 0$ a nastav časový krok Δt na zvolenou hodnotu.
1b. Z počátečních podmínek urči hodnoty $Z_{j,K}^0$.
1c. Pro všechny vnitřní hrany E urči hodnoty $m_{i,E}^{upw}$ ze vztahu $m_{i,E}^{upw} = 0,5\cdot(m_{i,K_1,E} + m_{i,K_2,E})$ $(E = \partial K_1 \cup \partial K_2)$, pro vnější hrany použij informace z okrajových podmínek.

2.5.2 Hlavní výpočetní smyčka algoritmu:

2a. Spočítej mřížkové koeficienty $b_{i,j,K,E,F}$ a $b_{i,j,K,E}$.
2b. Spočítej prvky matic $\boldsymbol{Q}_K$, $\boldsymbol{R}_{K,F}$, inverzní matici $\boldsymbol{Q}_K^{-1}$ a složky vektoru $\boldsymbol{R}_K$.
2c. Sestav matici soustavy $\boldsymbol{M}$.
2d. Řešením soustavy (10) získej hodnoty $Z_{j,E}^{k+1}$.
2e. S využitím (6) spočítej hodnoty $Z_{j,K}^{k+1}$.
2f. S využitím (3) spočítej koeficienty $v_{i,K,E}$.
2g. Nastav $k = k + 1$, $t_{k+1} = t_k + \Delta t$.

3 Implementace

Implementujeme dvojrozměrnou (2D) verzi numerické metody. Numerické schéma bylo implementováno v jazyce `C++`, pro paralelizaci bylo využito standardu `MPI`[1] [7]. Pro generování trojúhelníkových sítí byl využit program `Gmsh` [4]. Výpočty probíhaly na výpočetním klastru katedry matematiky, osazeném procesory AMD Opteron 6272 (16×2, 1GHz).

3.1 Sériová implementace numerického schématu

Nejprve se budeme věnovat sériové implementaci, která je jednodušší a navíc se v porovnání s ní ukáže přínos paralelizace. Jako sériovou budeme označovat implementaci běžící jednovláknově na jednom jádru procesoru.

V základní verzi je použita knihovna UMFPACK [2]. Pro jemnější sítě, a tedy větší rozměry matice soustavy (10), výrazně narůstá doba výpočtu. Výpočet se pokusíme urychlit použitím iteračních metod pro řešení soustav lineárních algebraických rovnic.

Při řešení uvažovaných úloh nebude matice soustavy (10) symetrická, což zúží výběr možných iteračních metod. Vyzkoušíme dvě metody vycházející z metod Krylovovských

[1] Message Passing Interface

podprostorů: restartovanou metodu zobecněných minimálních reziduí (GMRES) a stabilizovanou metodu bikonjugovaných gradientů (BICGSTAB). Podrobný popis metod lze najít například v [8]. Pro obě metody budeme uvažovat předpodmínění neúplným LU rozkladem. Použijeme variantu, kdy neuvažujeme vznik zaplnění a matice ILU[2] rozkladu má stejnou řídkou strukturu jako původní matice soustavy – ILU(0) [8].

Pro iterační řešiče je velice důležité kritérium pro zastavení iterací. Budeme používat normu rezidua dělenou normou pravé strany a k zastavení iterací dojde, pokud bude splněna podmínka:

$$\frac{\|\boldsymbol{M} \cdot \boldsymbol{Z}_q^{k+1} - \boldsymbol{b}\|}{\|\boldsymbol{b}\|} \leq \epsilon, \tag{11}$$

kde $\boldsymbol{Z}_q^{k+1}$ je řešení získané v q-té iteraci a ϵ je hodnota zastavovacího kritéria.

Z důvodu jednodušší implementace a kratších výpočetních časů na testovacích úlohách v sériové implementaci se při paralelizaci zaměříme výhradně na metodu BICGSTAB.

3.2 Paralelní implementace numerického schématu

Jako předpodmínění používáme ILU(0), které se samo o sobě špatně paralelizuje. Základní idea bude rozdělit matici na bloky a provádět ILU(0) pouze na diagonálních blocích. Ty už budou nezávislé a bude je možné zpracovat současně (tj. paralelně).

Pro rozdělení matice na bloky, a tedy i distribuci dat mezi procesory, zvolíme přístup vycházející z metody Domain Decomposition [10]. Nabízí se dělení podle elementů, ale protože primární neznámé v soustavě (10) jsou stopy na hranách, je výhodnější oblast dělit podle hran. Podle počtu procesorů rozdělíme hrany do patřičného počtu skupin a tyto skupiny pak mapujeme na jednotlivé procesory. Snažíme se, aby byly všechny skupiny přibližně stejně velké, aby byla zátěž na procesory rozdělena rovnoměrně. Také budeme chtít, aby byl počet elementů, které mají hrany v různých skupinách, co nejmenší. Hodnoty z těchto hran bude využívat více procesorů a budeme proto chtít omezit množství přenášených dat. Celou oblast rozdělíme na daný počet čtverců (případně obdélníků) a hrany přiřazujeme do skupin podle toho, ve kterém čtverci (obdélníku) leží jejich střed.

Nepracujeme pouze se stopami na hranách, ale také se středními hodnotami na elementech. Pro omezení komunikace a jednodušší implementaci budeme na každý procesor mapovat všechny elementy, které mají alespoň jednu hranu mapovanou na daný procesor. Elementy, které mají hrany v různých oblastech, budeme mít uložené duplicitně (elementy jsou v tomto případě trojúhelníky – maximálně na třech různých procesorech).

3.2.1 Komunikace

Projdeme jednotlivé kroky algoritmu popsaného v sekci 2.5 a určíme, kdy bude nutné provádět komunikaci a jaká data bude nutné přenášet. Při inicializaci z počátečních podmínek (kroky **1a** - **1c**) komunikace není nutná.

Při kroku **2a** – výpočtu mřížkových koeficientů $b_{i,j,K,E,F}$ a $b_{i,j,K,E}$ – se obejdeme bez komunikace. Komunikace bude potřeba až u kroku **2b** – výpočtu matic $\boldsymbol{Q}_K$, $\boldsymbol{R}_{K,F}$ a vektorů $\boldsymbol{R}_K$. Matice přísluší danému elementu a pro jejich výpočet je nutné znát hodnoty

[2] z anglického Incomplete LU

na všech hranách tohoto elementu. Ty jsou všechny na stejném procesoru pouze v případě, že má element všechny hrany ve stejné skupině.

Při kroku **2c** – sestavování matice soustavy $\boldsymbol{M}$ – je důležité pořadí rovnic a neznámých. Pokud očíslujeme skupiny hran čísly $0, 1, 2...$, bude pořadí řádků následující: nejprve bilance na hranách skupiny 0, pak skupiny 1, apod. Uvnitř každé skupiny pak uvedeme bilanci pro jednotlivé neznámé ve vzestupném pořadí. Pořadí sloupců bude voleno tak, aby měla matice strukturu co nejvíce podobnou blokově diagonální matici. To odpovídá pořadí: nejprve neznámé odpovídající hranám skupiny 0, potom skupiny 1, apod. Pořadí uvnitř skupin bude stejné, v jakém byly procházeny bilanční rovnice při řazení řádků. Rozdělení do skupin je voleno s ohledem na to, že pro vyjádření bilance na hraně potřebujeme hodnoty ze všech hran elementů, pro které je tato hrana společná. Matice bude mezi procesory rozdělená podle řádků tak, aby měl každý procesor blok řádků, které odpovídají bilanci na hranách, které jsou mapovány na daný procesor a příslušnou část pravé strany.

Při kroku **2d** – řešení soustavy – pak bude komunikace nejvíce. Při použití metody BICGSTAB je potřeba provádět skalární součin vektorů, součin matice a vektoru a aplikovat předpodmínění [8]. Nejjednodušší operací je aplikace předpodmínění. Na každém procesoru to odpovídá konstrukci rozkladu diagonálního bloku a dvěma zpětným chodům. Protože jsou diagonální bloky na procesoru uložené celé, není potřeba žádná komunikace. Při skalárním součinu dvou vektorů, které jsou rozdělené mezi více procesorů, každý procesor spočítá skalární součin na své části vektoru. Pomocí *paralelní redukce* [7] s operací + se pak dílčí skalární součiny sečtou na jeden procesor a výsledek se pomocí operace *scatter* [7] zase rozešle všem. Při násobení matice a vektoru je na každém procesoru uložen celý blok řádků a odpovídající část vektoru. Pro získání celého součinu využijeme operaci *cyklický posun* [7]. Na začátku je na každém procesoru možné provést součin diagonálního bloku a příslušné části vektoru. Po provedení této operace pošle procesor svoji část vektoru procesoru s číslem o jedna menším a nultý procesor posílá data poslednímu. Na každém procesoru pak opět můžeme vynásobit jeden blok matice s příslušnou části vektoru a výsledek přičíst k tomu získanému z prvního kroku. Tento postup opakujeme, dokud nezískáme celý součin matice s vektorem. Tento postup je možné vylepšit, pokud využijeme znalosti o struktuře matice, kterou násobíme. Pokud předem určíme nulové bloky, můžeme místo cyklického posunu o jedna použít cyklický posun o více kroků a omezit tak komunikaci. Takové omezení komunikace vede obzvláště při použití více procesorů k výraznému zrychlení [9].

V kroku **2e** – výpočtu středních hodnot na elementech $Z_{j,K}^{k+1}$ ze stop na hranách elementů $Z_{j,E}^{k+1}$ – je opět nutná komunikace a situace je velice podobná jako v kroku **2b**. U každého elementu opět potřebujeme hodnoty na všech jeho hranách a ty si v případě, že má element hrany v různých skupinách, musíme přeposílat. V krocích **2f** a **2g** už se pak zase obejdeme bez komunikace.

Typ komunikace v krocích **2b** a **2e** je úplně stejný, liší se jen data, která je nutné odeslat. Bylo by extrémně neefektivní posílat hodnoty z jednotlivých hrany jako samostatné zprávy. Na začátku každého provedení kroků **2b** nebo **2e** můžeme sestavit jednu dlouhou zprávu, ve které budou všechny hodnoty, které bude potřeba přeposlat z jedné skupiny do druhé a odeslat je všechny najednou.

4 Výsledky

Řešená úloha bude rozšířením čistě difuzní jednorozměrné (1D) McWhorterovy–Sunadovy úlohy v homogenním prostředí. Tato úloha je definována na polopřímce $\langle 0; +\infty)$, na které je dána počáteční hodnota saturace S_i a v bodě nula je dána okrajová podmínka pro saturaci S_0. Pro nestlačitelné fáze bez vlivu gravitace lze nalézt přesné řešení [6, 3]. Pro dvojrozměrnou (2D) verzi numerické metody, budeme místo polopřímky uvažovat nekonečný pás a přesné řešení bude konstantní ve směru kolmém na polopřímku, na které je řešena původní 1D úloha.

4.1 Volba koeficientů

Pro řešení této úlohy zvolíme v obecné formulaci úlohy (1) koeficienty:

$$\boldsymbol{N} = \begin{pmatrix} -\Phi\rho_w \frac{dS_w}{dp_c} & 0 \\ -\Phi\rho_n \frac{dS_w}{dp_c} & \Phi S_n \frac{d\rho_n}{dp_n} \end{pmatrix}, \qquad \boldsymbol{u} = \boldsymbol{0}, \qquad \boldsymbol{m} = \begin{pmatrix} \rho_w \frac{\lambda_w}{\lambda_t} \\ \rho_n \frac{\lambda_n}{\lambda_t} \end{pmatrix},$$

$$\boldsymbol{D} = \begin{pmatrix} \lambda_t \boldsymbol{K} & -\lambda_t \boldsymbol{K} \\ 0 & \lambda_t \boldsymbol{K} \end{pmatrix}, \qquad \boldsymbol{w} = \begin{pmatrix} -\lambda_t \rho_w \boldsymbol{K}\boldsymbol{g} \\ \lambda_t \rho_n \boldsymbol{K}\boldsymbol{g} \end{pmatrix}, \qquad \boldsymbol{f} = \begin{pmatrix} -f_w \\ f_n \end{pmatrix},$$

kde $\lambda_t = \lambda_w + \lambda_n$. Podrobnosti k odvození koeficientů lze nalézt v [9].

Pro numerické řešení se omezíme na oblast konečné délky. Rozměry použité oblasti, značení hranic a základní trojúhelníková síť jsou znázorněné na Obrázku 1. Přesné a numerické řešení budeme porovnávat v čase $t = 60\ 000$ s, kdy čelo řešení ještě nedorazilo k hranici reprezentující nekonečno (Γ_1).

Celá oblast bude vyplněna pískem C, jehož vlastnosti jsou převzaty z [9]. Jako smáčivou fázi budeme uvažovat vodu a jako nesmáčivou fázi vzduch. Parametry použitých tekutin jsou uvedeny v [9]. Pro numerické řešení použijeme hodnotu počáteční saturace $S_i = 0,1$ a okrajové podmínky:

$$\boldsymbol{u}_n \cdot \boldsymbol{n} = 0, \qquad \boldsymbol{u}_w \cdot \boldsymbol{n} = 0, \qquad \text{na } \Gamma_1 \cup \Gamma_2 \cup \Gamma_4, \tag{12}$$

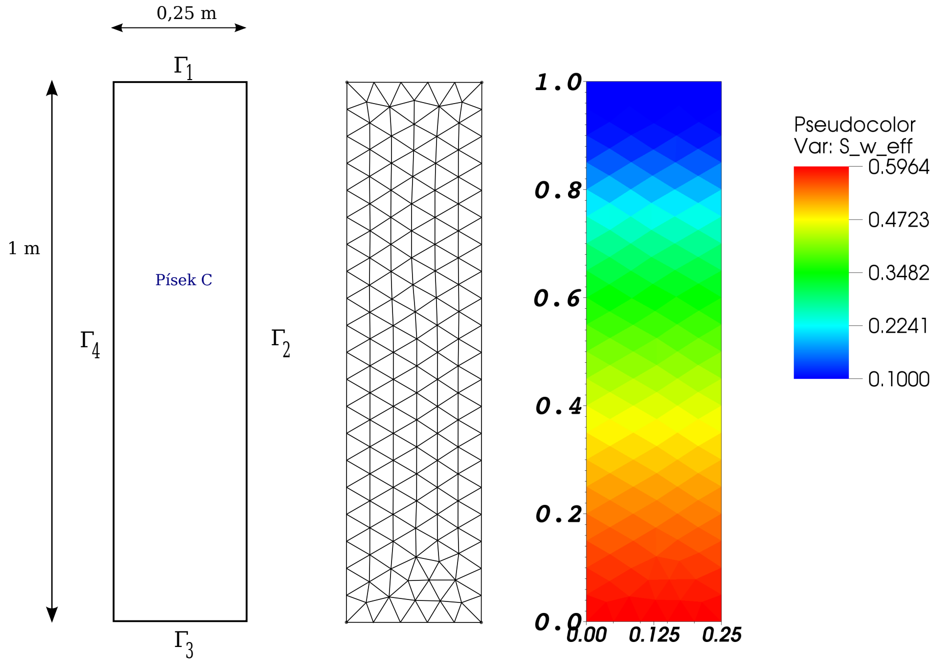
$$p_n = 10^5 \text{ Pa}, \qquad S_w = 0,6, \qquad \text{na } \Gamma_3. \tag{13}$$

4.2 Chyby řešení

Nejprve porovnáme jednotlivé variant numerické metody s ohledem na přesnost řešení. Pro iterační řešiče zvolíme zastavovací kritéria $\epsilon = 10^{-10}$ pro metodu GMRES a $\epsilon = 10^{-15}$ pro metodu BICGSTAB v sériové i paralelní verzi. Pro vyšší hodnoty ϵ chyby na jemnějších sítích výrazně narůstají a numerické řešení nekonverguje, jak je ukázáno v [9].

V Tabulce 1 jsou uvedeny chyby numerického řešení pro jednotlivé varianty metody na různých sítích, v Tabulce 2 pak hodnoty experimentálního řádu konvergence.

Z těchto výsledků je patrné, že řešení konverguje a řád konvergence je o něco menší než jedna, což jsou hodnoty typické pro metody využívající techniku upwind prvního řádu [5]. Také je vidět, že pro dané volby zastavovacího kritéria pro iterační metody, jsou hodnoty získané iteračními metodami velice podobné hodnotám, které dostaneme s použitím knihovny UMFPACK. Také vidíme, že paralelní verze metody dává téměř stejné výsledky jako sériová.



Obrázek 1: Výpočetní oblast, základní nestrukturovaná síť generovaná programem `Gmsh` (246 elementů, 394 hrany) a numerické řešení v čase $t = 60\,000$ s.

Počet elementů	UMFPACK		BICGSTAB		GMRES		Paralelně - 6 jader	
	L_1	L_2	L_1	L_2	L_1	L_2	L_1	L_2
246	$62,33$	$234,93$	$54,55$	$208,04$	$54,55$	$208,23$	$54,54$	$208,22$
984	$29,06$	$116,93$	$29,05$	$116,92$	$29,06$	$116,93$	$29,04$	$116,88$
3 396	$15,17$	$62,20$	$15,14$	$62,12$	$15,17$	$62,19$	$15,17$	$62,23$
15 744	$7,86$	$32,36$	$7,84$	$32,31$	$7,86$	$32,34$	$7,85$	$32,33$
62 976	$4,10$	$16,96$	$4,11$	$17,04$	$4,09$	$16,96$	$4,08$	$17,07$
251 904	$2,20$	$9,51$	$2,21$	$9,52$	$2,20$	$9,53$	$2,21$	$9,51$

Tabulka 1: Chyby numerických řešení pro jednotlivé varianty metody.

4.3 Výpočetní náročnost

V předchozí části jsme ukázali, že pro použité hodnoty ϵ jsou chyby při použití jednotlivých variant metody skoro stejné. Nyní nás bude zajímat srovnání výpočetních časů. Výpočetní časy pro sériové implementace jsou uvedeny v Tabulce 3, výpočetní časy pro paralelní implementaci s různým počtech výpočetních jader jsou uvedeny v Tabulce 4. Urychlení při výpočtu na p procesorech U_p a efektivita paralelizace E_p jsou uvedeny v Tabulce 5.

Z Tabulky 3 je vidět, že ze sériových implementací je nejrychlejší metoda BICGSTAB. Také GMRES je na jemnějších sítích rychlejší než knihovna UMFPACK. Z Tabulek 4 a 5 je vidět přínos paralelizace. Výhoda využití více výpočetních jader se výrazněji projeví při výpočtech na jemnějších sítích.

	UMFPACK		BICGSTAB		GMRES		Paralelně	
Zjemnění sítě	eoc_1	eoc_2	eoc_1	eoc_2	eoc_1	eoc_2	eoc_1	eoc_2
246 → 984	1,10	1,01	0,91	0,83	0,91	0,84	0,91	0,83
984 → 3 936	0,94	0,91	0,94	0,91	0,94	0,91	0,94	0,91
3 936 → 15 744	0,95	0,94	0,95	0,94	0,95	0,94	0,95	0,94
15 744 → 62 976	0,94	0,93	0,93	0,92	0,94	0,93	0,93	0,92
62 976 → 251 904	0,90	0,83	0,89	0,83	0,89	0,83	0,88	0,83

Tabulka 2: Experimentální řády konvergence pro jednotlivé varianty metody.

Počet elementů	Rozměr matice soustavy $\boldsymbol{M}$	UMFPACK	BICGSTAB	GMRES
246	788	1,3	1,4	1,7
984	3 052	7,7	5,2	36,4
3 936	12 008	67	45	800
15 744	47 632	756	373	999
62 976	189 728	9 160	3 968	7 272
251 094	757 312	186 435	34 401	84 900

Tabulka 3: Výpočetní časy pro jednotlivé sériové varianty metody v sekundách.

Počet	Rozměr matice	Počet výpočetních jader						
elementů	soustavy $\boldsymbol{M}$	1	2	4	6	8	16	32
246	788	1,4	0,5	0,4	0,4	0,4	-	-
984	3 052	5,2	2,9	1,8	1,3	1,2	-	-
3 936	12 008	45	28	11	7	7	6	-
15 744	47 632	373	294	107	79	63	37	30
62 976	189 728	3 968	2 947	1 393	921	820	387	305
251 094	757 312	34 401	21 147	12 718	8 093	7 578	6 222	3 203

Tabulka 4: Výpočetní časy pro paralelní implementaci metody v sekundách.

Počet	Počet výpočetních jader											
elementů	2		4		6		8		16		32	
	U	E	U	E	U	E	U	E	U	E	U	E
246	2,8	140 %	3,5	88 %	3,5	58 %	3,5	44 %	-	-	-	-
984	1,8	90%	2,9	72 %	4,0	67 %	4,3	54 %	-	-	-	-
3 936	1,6	80 %	4,1	102 %	6,4	107 %	6,4	80 %	7,5	47 %	-	-
15 744	1,3	63 %	3,5	87 %	4,7	78 %	6,0	74 %	10,1	63 %	12,4	39 %
62 976	1,4	67 %	2,9	71 %	4,3	71 %	4,8	60 %	10,3	64 %	13,0	41 %
251 904	1,6	81 %	2,7	67 %	4,3	70 %	4,6	57 %	5,5	35 %	10,7	34%

Tabulka 5: Urychlení (U) a efektivita (E) paralelizace při použití různého počtu výpočetních jader.

5 Závěr

V této práci jsme se zabývali numerickým řešením úloh dvoufázového proudění v porézním prostředí. Navrhli jsme numerickou metodu pro řešení tohoto typu úloh založenou na hybridní metodě smíšených konečných prvků. Metoda je implementována pro dvojrozměrné úlohy. Jsou popsány sériové implementace používající přímé řešení s použitím knihovny UMFPACK i iterační metody BICGSTAB a GMRES s předpodmíněním ILU(0). Věnujeme se paralelní variantě metody založené ma metodě Domain Decomposition, využívající standardu `MPI`.

Tuto metodu poté testujeme na úloze, pro kterou je známé přesné řešení. Je ukázáno, že numerické schéma konverguje a hodnoty experimentálního řádu konvergence se pohybují kolem jedničky. Ukazuje se, že chyby při použití různých variant metody jsou velice podobné, ale liší se jejich výpočetní náročnost. Ze sériových implementací je nejrychlejší metoda BICGSTAB a na jemnějších sítích je výhodné použití paralelizace, která přináší další urychlení.

Literatura

[1] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods.* Springer-Verlag, (1991).

[2] T. A. Davis. *Algorithm 832: UMFPACK, an unsymmetric-pattern multifrontal method.* ACM Transactions on Mathematical Software **30** (2004), 196–199.

[3] R. Fučík, T. H. Illangasekare, and M. Beneš. *Multidimensional self-similar analytical solutions of two-phase flow in porous media.* Advances in Water Resources **90** (2016), 51–56.

[4] C. Geuzaine and J. F. Remacle. *Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities.* International Journal for Numerical Methods in Engineering **79** (2009), 1309–1331.

[5] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems.* Cambridge University Press, (2002).

[6] D. B. McWhorter and D. K. Sunada. *Exact integral solutions for two-phase flow.* Watter Resources Research **26** (1990), 399–413.

[7] T. O. M. Project. Open MPI: Open Source High Performance Computing. `https://www.open-mpi.org/`. Cit. 10. 3. 2016.

[8] Y. Saad. *Iterative Methods for Sparse Linear Systems.* Society for Industrial and Applied Mathematics, (2003).

[9] J. Solovský. Matematické modelování dvoufázového vícesložkového proudění v porézním prostředí v problematice ochrany životního prostředí, (2016). Diplomová práce, ČVUT v Praze.

[10] A. Toselli and O. Windlund. *Domain Decomposition Methods – Algorithms and Theory.* Springer-Verlag, (2005).

Blind Separation of Underdetermined Linear Mixtures Based on Source Nonstationarity and AR(1) Modeling*

Ondřej Šembera

2nd year of PGS, email: sembera@utia.cas.cz
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:

Petr Tichavský, Department of Stochastic Informatics
Institute of Information Theory and Automation, CAS

Zbyněk Koldovský, The Institute of Information Technology and Electronics
Technical University Liberec

Abstract. The problem of blind separation is to estimate unknown source signals from observed mixtures of the signals, i.e. linear combinations of the original signals, without detailed knowledge of the mixing process. In this paper, we consider underdetermined mixtures, that is the number of observed mixtures is lower than the number of the original signals. We propose a method relying on the nonstationarity of the source signals. The signals are assumed to be piece-wise stationary Gaussian processes with zero means and different variances in each epoch. In comparison with the previous works [3], the sources are not assumed to be i.i.d. in each epoch. Instead, we adopt the autoregressive process of order one as a model for the source signals with different autoregressive coefficient in each epoch. This model was shown to be more appropriate for the blind separation of the natural speech signals especially for the speech records with a high sampling frequency. The proposed separation method is derived by applying the maximum likelihood estimation method to the approximate probability density of the sample covariances, which are computed in each epoch. For artificial data following the assumed model, the accuracy of the method approaches the corresponding Cramér-Rao lower bound. In the case of natural speech signals, the proposed method achieves better separation than the competing algorithms [3], [1], [2].

Keywords: Autoregressive Processes, Cramér-Rao Bound, Blind Source Separation

Abstrakt. Úlohou slepé separace je odhad neznámých původních signálů z jejich pozorovaných směsí, tj. lineárních kombinací původních signálů. Nepředpokládáme žádnou apriorní znalost samotného procesu mísení. V této práci uvažujeme nedourčené směsi, tj. počet pozorovaných směsí je menší než počet separovaných signálů. Navrhujeme metodu založenou na nestacionaritě separovaných signálů. Předpokládáme, že separované signály jsou po částech stacionární Gaussovské procesy s nulovou střední hodnotou a s různým rozptylem v každém časovém úseku. Na rozdíl od přechozích prací [3] nepředpokládáme nezávislost jednotlivých vzorků pozorovaných směsí. Namísto toho modelujeme separované signály v jednotlivých časových úsecích Gausovským autoregresním procesem prvního řádu. Tento model je vhodnější pro

*This work was supported by the Czech Science Foundation through Project No. 14-13713S.

separaci řečových promluv a to zvláště v případě vysoké vzorkovací frekvence separovaných promluv. Navrhovaná metoda byla odvozena použitím metody maximální věrohodnosti na přibližnou pravděpodobnostní hustotu empirických kovariancí, které jsou napočítány z každého úseku pozorovaných směsí. V aplikaci na umělá data odpovídající uvažovanému modelu se přesnost separace blíží Cramérově-Raově dolní mezi. V případě separace řečových promluv dosahuje navrhovaná metoda lepší separace než konkurenčí algoritmy [3], [1], [2].

Klíčová slova: Autoregresivní procesy, Cramérova-Raova mez, slepá separace signálů.

Full paper: This work was presented at 2016 IEEE International Conference on Acoustic, Speech and Signal Processing held in Shanghai on March 20-25, 2016.
O. Šembera, P. Tichavský, Z. Koldovský. *Blind separation of underdetermined linear mixtures based on source nonstationarity and AR (1) modeling.* In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016), 4323–4327.
`http://ieeexplore.ieee.org/document/7472493/?arnumber=7472493`

References

[1] De Lathauwer, Lieven, and Joséphine Castaing. *Blind identification of underdetermined mixtures by simultaneous matrix diagonalization.* Signal Processing, IEEE Transactions **56** (2008), 1096–1105.

[2] K.K. Lee, W.-K. Ma, X. Fu, T.-H. CHan, C.-Y. Chi. *A Khatri-Rao subspace approach to blind identification of mixtures of quasi-stationary sources.* Signal Processing **93** (2013), 3515–3527.

[3] P. Tichavský, Z. Koldovský. *Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources.* Signal Processing, IEEE Transactions **59** (2011), 1037–1047.

The Algorithmizable Modeling of the Object-Oriented Data Model in Craft.CASE

Ondřej Šubrt

2nd year of PGS, email: subrton2@fjfi.cvut.cz
Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Tomáš Liška, Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. The object-oriented approach usually does not follow any formal design process and is mostly ad hoc in real software development. This makes it more of an art than a science. The quality of the resultant design therefore depends to a large extent on the skills of the individual designer and cannot be evaluated easily. In this paper we present an approach to normalization of the object-oriented conceptual model based on UML class diagrams. The normalization of the object-oriented data model is performed in algorithmic way based on model transformation rules. The algorithm is able to transform the object-oriented data model from one into the other normal form following the transformation rules. The algorithm application rids the design process from the above-mentioned problems and yields a better object model by bringing formalism and taking a scientific approach. Recently, development of the CASE tool based on this approach has been started.

Keywords: Data normalization, object-oriented data model (ODM), first object-oriented normal form (1ONF), second object-oriented normal form (2ONF), third object-oriented normal form (3ONF), Craft.CASE

Abstrakt. Objektově orientovaný přístup obvykle nedodržuje žádný formální proces návrhu a je většinou ad hoc v reálném vývoji softwaru. Díky tomu je dnes vývoj softwaru více uměním než vědou. Kvalita výsledného návrhu proto závisí do značné míry na schopnosti jednotlivých designérů a nemůže být snadno ohodnocena. V tomto článku prezentujeme přístup k normalizaci objektově orientovaného konceptuálního modelu založeného na UML diagramech tříd. Normalizace objektově orientovaného datového modelu se provádí algoritmickým způsobem na základě transformačních pravidel. Algoritmus je schopen transformovat objektově orientovaný datový model z jedné normální formy do druhé podle transformačních pravidel. Použití tohoto algoritmu zbavuje proces návrhu výše uvedených problémů. Výsledkem je poté lepší objektový model v důsledku zavedení formalismu a vědeckého přístup. V poslední době byl zahájen vývoj nástroje na základě tohoto přístupu.

Klíčová slova: Datová normalizace, objektově orientovaný datový model (ODM), první objektově orientovaná normální forma (1ONF), druhá objektově orientovaná normální forma (2ONF), třetí objektově orientovaná normální forma (3ONF), Craft.CASE

1 Introduction

The object-oriented programming (OOP) has its origins in the researching of operating systems, graphic user interfaces, and particularly in programming languages, that took place in the 1970s [11]. It differs from other software engineering approaches by incorporating non-traditional ways of thinking into the field of informatics. We look at systems by abstracting the real world in the same way as in ontological, philosophical streams. The basic element is an object that describes data structures and their behavior. OOP has been and still is explained in many books [6, 9, 3, 12]. The [6], written by OOP pioneers, belongs to the best.

In real software development, the object-oriented approach usually does not follow any formal design. In this paper, we propose the transformation of object-oriented design to correct one following the transformations rules. Moreover, to make the process of transformation automatic and self-sustaining, we introduce the algorithms handling these transformations. The goal of the paper is to obtain a cohesive framework providing the resultant design with high quality. The final framework could be used in software development for design improvements.

This paper is organized as follows. Section 2 presents three normalization rules for model transformation from one into the other normal form. In Section 3, the introduction to Craft.CASE scripting is stated and the description of algorithm for algorithmizable modeling is given. In the last section, the algorithmizable modeling is investigated and evaluated on one more complicated example.

2 Three Object Normal Forms and Transformation Rules

In the data world, there is a common process called data normalization by which the data are organized in such a way as to reduce and even eliminate data redundancy, effectively increasing the cohesiveness of data entities. Data normalization only deals data and not behavior. We need to consider both when normalizing the object schema. Class normalization is a process of reorganizing the structure of object schema in such a way as to increase the cohesion of classes while minimizing the coupling between them.

In this section, three object normal forms are introduced [11, 10]. Moreover, the transformation rules from one into the other normal form are discussed in a detailed way [15, 14, 13, 2, 8].

2.1 First Normal Form Rule

Definition 1. *A class is in the first object normal form (1ONF) when its objects do not contain group of repetitive attributes. Repetitive attributes must be extracted into objects of a new class. The group of repetitive attributes is then replaced by the link at the collection of the new objects. An object schema is in 1ONF when all of its classes are in 1ONF.*

More formally; Let us have an object a in the object system Ω as $a \in \Omega$, where for $k > 1$ (length of collections of similar attributes) and $n > 1$ (number of repetition of these collections) is $data(a) = [\ldots, x_1^1, \ldots, x_k^1, \ldots, x_1^n, \ldots, x_k^n, \ldots]$ having $\forall i \in (1, \ldots, k) : class(x_i^1) = class(x_i^2) = \ldots = class(x_i^n)$.

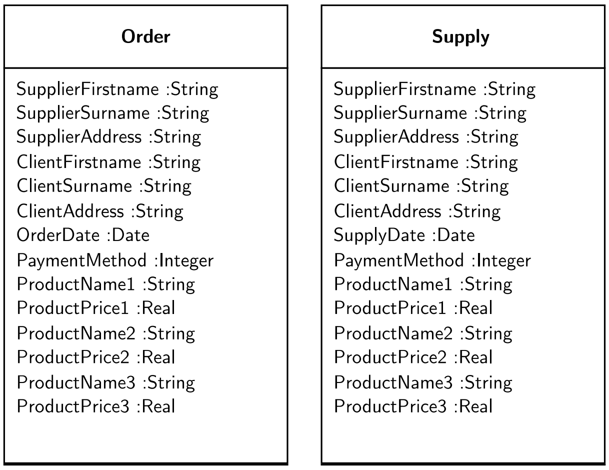


Figure 1: Model in 0ONF

Then it is required to modify object a and create new objects $b_j \in \Omega$ for $j \in (1, \ldots, n)$ as $data(a) = [\ldots, \{b_j\}, \ldots]$ and $data(b_j) = [x_1^j, \ldots, x_k^j]$.

In Figure 1, there is the example of data structure in non-normalized form and in the Figure 2, there is the same example in 1ONF.

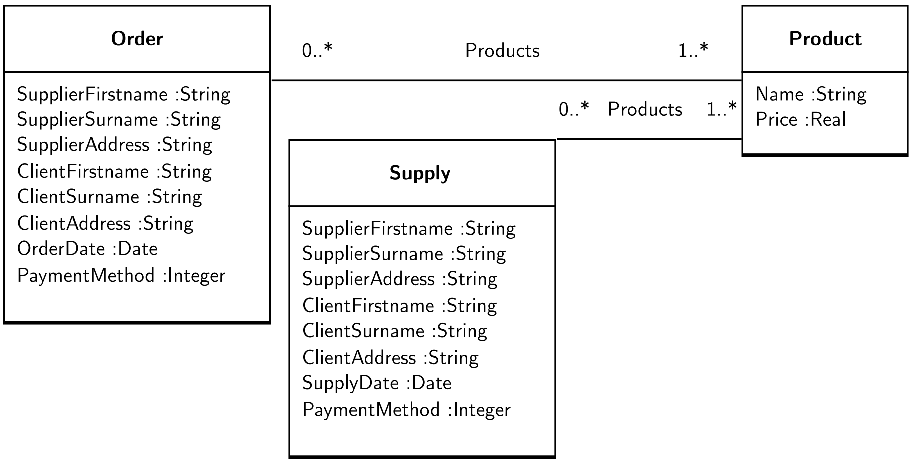


Figure 2: Model in 1ONF

2.2 Second Normal Form Rule

Definition 2. *A class is in the second object normal form (2ONF) when it is in 1ONF and when its objects do not contain attribute or group of attributes, which are shared with another object. Shared attributes must be extracted into new objects of a new class, and in all objects, where they appeared, must be replaced by the link to the object of the new class. An object schema is in 2ONF when all of its classes are in 2ONF.*

More formally; Let us have two objects $a, b \in \Omega$ for $k > 1$ (length of a collection of shared attributes) as $data(a) = [\ldots, x_1, \ldots, x_k, \ldots]$ and $data(b) = [\ldots, y_1, \ldots, y_k, \ldots]$ having $\forall i \in (1, \ldots, k) : x_i = y_i$.

Then it is required to modify objects a and b and create new object $c \in \Omega$ as $data(a) = [\ldots, c, \ldots]$ and $data(b) = [\ldots, c, \ldots]$ and $data(c) = [x_1, \ldots, x_k] = [y_1, \ldots, y_k]$.

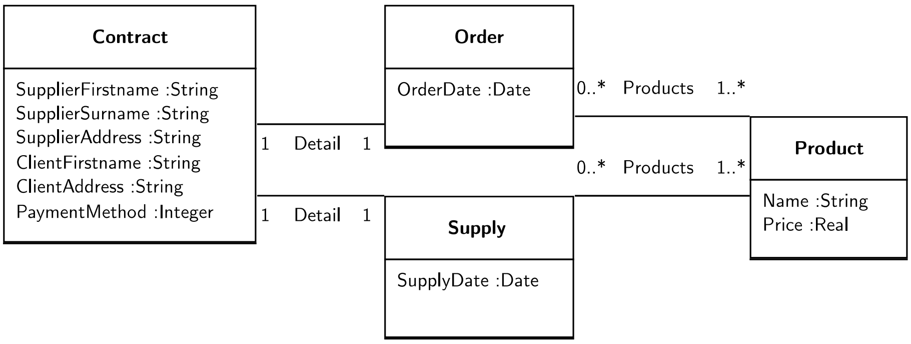


Figure 3: Model in 2ONF

In Figure 3, it concerns the attributes *SupplierFirstname*, *SupplierSurname* and *SupplierAddress* for *Supplier* and *ClientFirstname*, *ClientSurname* and *ClientAddress* for *Client* and method of payment in our example. Because these attributes are common for both concrete order and supply, it was necessary to create the new object class *Contract*.

2.3 Third Normal Form Rule

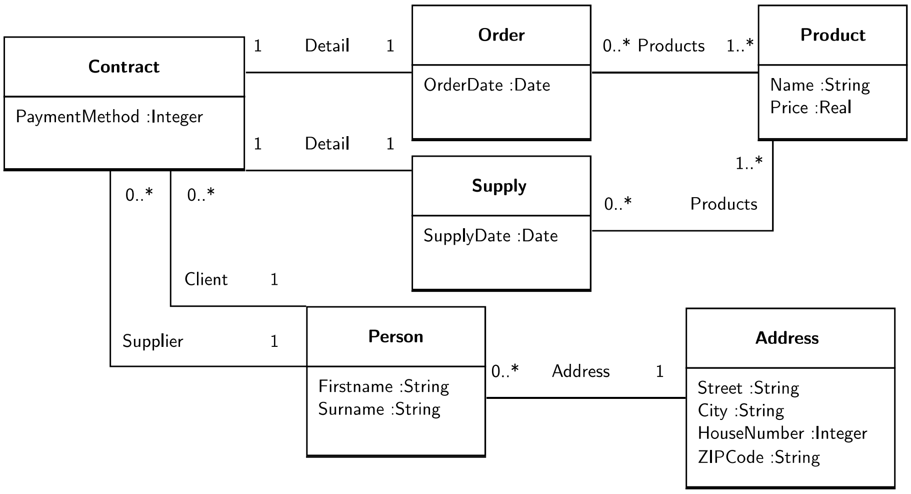


Figure 4: Model in 3ONF

Definition 3. *A class is in the third object normal form (3ONF) when it is in 2ONF and when its objects do not contain attribute or group of attributes, which have the independent interpretation in the modeled system. These attributes must be extracted into objects of a new class and in objects, where they appeared, must be replaced by the link to this new object. An object schema is in 3ONF when all of its classes are in 3ONF.*

More formally; Let us have an object $a \in \Omega$ for $k > 1$ (length of a collection of independent attributes) having $data(a) = [\ldots, x_1, \ldots, x_k, \ldots]$, where $[x_1, \ldots, x_k]$ is collection of independent attributes.

Then it is required to create new object $b \in \Omega$ and modify object a as $data(a) = [\ldots, b, \ldots]$ and $data(b) = [x_1, \ldots, x_k]$.

In Figure 4, it concerns the data about *suppliers* and *clients* in the objects of the class *Contract*. These attributes represent some *persons* having independent interpretation on contracts. The same applies to *addresses*.

3 Algorithmizable Modeling in Craft.CASE

Craft.CASE is a business process analysis (BPA) tool based on a C.C method [4]. The core of this method explains how to progress in a BPA project without forgetting anything. The C.C method consists of small steps, sequences of which are tested and validated as soon - and as often - as possible. Following this method allows processes to remain consistent even if the problem is complex. Craft.CASE leads its users step-by-step according to the C.C method. This means that the notation of the tool is rigid enough to discover, understand, and analyse processes in a consistent way.

Last but not least, another very important feature is the ability to simulate the process. Whether man use it for model validation, verification, or just to present and visualize process progress, depends on his current needs. The list of the most common features is given below:

- guidelines set according to the C.C. method
- process analysis categorization into interview, business and conceptual phases
- ability to set the user-defined properties to individual objects
- graphically visualize any object according to its values and properties
- performing process animations and simulations
- generating reports and specifying their content
- team collaboration
- advanced functions or functions not common in the field of BPA software are defined by users who can define and run their own scripts using our built-in programming language

3.1 Introduction to Craft.CASE scripting

This section explains some details of C.C programming language and the environment used for programming in the language [5, 10]. The Craft.CASE tool contains integrated development environment with a source code editor (Module Browser), a workspace (place for instant testing of pieces of code) and a debugger. C.C language is a simple programming language that user can use for:

- querying the process model developed in the Craft.CASE tool

- automatic transformation and modification of the process model in Craft.CASE
- specification of user-defined reports and exports
- extension to the Craft.CASE tool functionality (normalization of models, design patterns application, refactoring, . . .)

Table 1: Important objects for normalization of models

conceptual::class	represents a specific class
conceptual::composition	represents a specific attribute of some class
conceptual::method	represents a specific method of some class
conceptual::association	represents a relation between two classes

We have prepared a simple ODM representing relationship between a car and the owner of this car. In Figure 5, the ODM of our example is given. We use it for demonstrating of scripting in Craft.CASE.

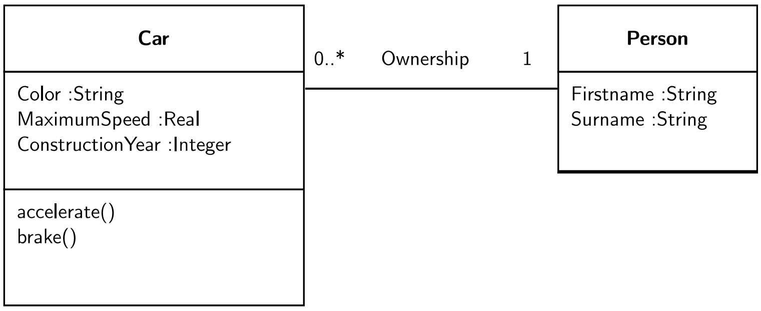


Figure 5: The ODM representing relationship between a car and the owner of this car

In the following example, the code is printing the list of all attributes, methods and associations to other classes for class *Car* in Craft.CASE.

```
# it gets all classes from project
Classes := project:elements(conceptual::class).
# initialization of ClassCar variable
ClassCar := nil.
# searching for class Car
from 1 to size(Classes) do { :I |
    if ClassCar = nil then
    {
        if Classes[I]["name"] = "Car" then
        {
            ClassCar := Classes[I].
        }.
    }.
}.
# it gets all attributes and relations to other classes
```

```
ClassLinks := element:links(ClassCar).
from 1 to size(ClassLinks) do { :I |
    if element:type(ClassLinks[I]) = "Composition" then
    {
        stream:print-nl("Attribute: " + ClassLinks[I]["name"]).
    }.
    if element:type(ClassLinks[I]) = "Association" then
    {
        stream:print-nl("Relation to class " +
        element:target(ClassLinks[I])["name"] + ": " +
        ClassLinks[I]["name"]).
    }.
}.
# it gets all methods
ClassMethods := conceptual:methods(ClassCar).
from 1 to size(ClassMethods) do { :I |
    stream:print-nl("Method: " + ClassMethods[I]["name"]).
}.
```

We run the code in Workspace of Craft.CASE. To organize code in well-arranged way, the creation of own packages is also possible in Module Browser. The output of the code follows:

```
Relation to class Person: Owner
Attribute: Color
Attribute: MaximumSpeed
Attribute: ConstructionYear
Method: accelerate
Method: brake
```

3.2 Normalization Algorithms for ODM

In this section, we introduce algorithms enabling transformation from one into the other normal form. We call this process as the normalization of ODM. Firstly, the algorithm for transformation from 0ONF to 1ONF is given, see Algorithm 1. To remind, a class is in 1ONF when specific behavior required by an attribute that is actually a collection of similar attributes is encapsulated within its own class. An object schema is in 1ONF when all of its classes are in 1ONF.

Then, the algorithm for transformation from 1ONF to 2ONF is given, see Algorithm 2. To remind, a class is in second object normal form (2ONF) when it is in 1ONF and when "share" behavior that is needed by more than one instance of the class is encapsulated within its own class(es). An object schema is in 2ONF when all of its classes are in 2ONF.

All attributes are identified only by their names and data types. It means the mentioned algorithms are dependent on well-named attributes and their uniqueness in order to transform model correctly. Of course, the violation of this restriction might cause unsuccessful and incorrect transformations.

Algorithm 1 Transformation from 0ONF into 1ONF algorithm

```
$Classes$ ← get all classes in current project
for all $Class \in Classes$ do
    $DuplicatedAttributes$ ← get all duplicated attributes of class $Class$
    for all $DuplicatedAttribute \in DuplicatedAttributes$ do
        remove attribute $DuplicatedAttribute$ from class $Class$
    end for
    for all $DuplicatedAttribute \in DuplicatedAttributes$ do
        $NewClassName$ ← get name of new class from name of $DuplicatedAttribute$
        $NewAttributeName$ ← get name of new attribute from name of $DuplicatedAttribute$
        if class with name $NewClassName$ already exists then
            $NewClass$ ← get class with name $NewClassName$
        else
            $NewClass$ ← create class with name $NewClassName$
        end if
        if attribute with name $NewAttributeName$ in class $NewClass$ does not exist yet then
            create a new attribute with name $NewClassName$ in class $NewClass$
        end if
        if association between $Class$ and $NewClass$ does not exist yet then
            create a new association between $Class$ and $NewClass$
        end if
    end for
end for
```

Finally, the algorithm for transformation from 2ONF to 3ONF should be also given. To remind, a class is in third object normal form (3ONF) when it is in 2ONF and when it encapsulates only one set of cohesive behaviors. An object schema is in 3ONF when all of its classes are in 3ONF. Unfortunately, the algorithm would be more complex than the previous ones and its implementation is going beyond the scope of this article.

To identify attribute or group of attributes having the independent interpretation in the modeled system is not a straightforward process. It must be considered what each physical attribute represents. It is not a simple task to identify these representations without any information handling them from other models (participants, function and scenarios, participant relations, business interactions, business diagrams, etc.) incorporated in a whole process of analysis.

For the identification of attributes representation, it could be also used clustering, pattern recognition, reinforcement learning, neural networks, etc. In sum, any technique based on the machine learning. We can see repeating groups of data from a data entity.

4 Test Case

In this section, we test, investigate and evaluate proposed algorithms on the example. The example is quite well-known. We can find it also in several other publications [11, 1].

Consider the class *Student* in Figure 6. This design is clearly not very cohesive. This single class is implementing functionality that is appropriate to several concepts. To transform this example from 0ONF into 1ONF, we use Algorithm 1.

With 1ONF we remove repeating groups of data from a data entity and create a new class *Seminar*. All these repeating attributes have been moved to this class. In Figure 7, we can see the resultant design in 1ONF.

Consider *Seminar* in Figure 7. It implements the behavior of maintaining both information about the course that is being taught in the seminar and about the professor teaching that course. Although this approach would work, it unfortunately does not work

Algorithm 2 Transformation from 1ONF into 2ONF algorithm

```
$Classes \leftarrow$ get all classes in current project
$AgreementNumber \leftarrow 0$
$ClassXGlobal \leftarrow nil$
$ClassYGlobal \leftarrow nil$
$TransformationTo2ONFDone \leftarrow false$
while $\neg TransformationTo2ONFDone$ do
    for all $ClassX, ClassY \in Classes$ do
        if $ClassX \neq ClassY$ then
            $CurrentAgreementNumber \leftarrow 0$
            $AttributesX \leftarrow$ get all attributes of class $ClassX$
            $AttributesY \leftarrow$ get all attributes of class $ClassY$
            for all $AttributeX \in AttributesX, AttributeY \in AttributesY$ do
                if $AttributeX$ and $AttributeY$ represent same attribute then
                    $CurrentAgreementNumber \leftarrow CurrentAgreementNumber + 1$
                end if
            end for
            if $AgreementNumber < CurrentAgreementNumber$ then    ▷ searching for two classes with the highest agreement
                $AgreementNumber \leftarrow CurrentAgreementNumber$
                $ClassXGlobal \leftarrow ClassX$
                $ClassYGlobal \leftarrow ClassY$
            end if
        else
            $Attributes \leftarrow$ get all attributes of class $ClassX$ with the same $Prefix$
            if $Attributes \neq \emptyset$ then
                remove attributes $Attributes$ from class $ClassX$
                $NewClass \leftarrow$ create class with name $Prefix$
                create all attributes $Attributes$ in class $NewClass$
                create a new association between $ClassX$ and $NewClass$
            end if
        end if
    end for
    if $AgreementNumber > 0$ then
        $AttributesX \leftarrow$ get all attributes of class $ClassXGlobal$
        $AttributesY \leftarrow$ get all attributes of class $ClassYGlobal$
        $NewClassName \leftarrow$ get name of new class from name of $ClassXGlobal$ and $ClassYGlobal$
        $NewClass \leftarrow$ create class with name $NewClassName$
        for all $AttributeX \in AttributesX, AttributeY \in AttributesY$ do
            if $AttributeX$ and $AttributeY$ represent the same attribute then
                remove attribute $AttributeX$ from class $ClassXGlobal$
                remove attribute $AttributeY$ from class $ClassYGlobal$
                create a new attribute represent the same attribute in class $NewClass$
            end if
        end for
        create a new association between $ClassXGlobal$ and $NewClass$
        create a new association between $ClassYGlobal$ and $NewClass$
        $AgreementNumber \leftarrow 0$
        $ClassXGlobal \leftarrow nil$
        $ClassYGlobal \leftarrow nil$
    else
        $TransformationTo2ONFDone \leftarrow true$
    end if
end while
```

very well. When the name of a course changes we would have to change the course name for every seminar of that course.

To transform our example from 1ONF into 2ONF, we use Algorithm 2. Figure 8 depicts the object schema in 2ONF. To improve the design of *Seminar* we have introduced two new classes, *Course* and *Professor* which encapsulate the appropriate behavior needed to implement course objects and professor objects.

Unfortunately, we do not have any algorithm for transformation from 2ONF into 3ONF, since it is going beyond the scope of this article. The exact reasons have already

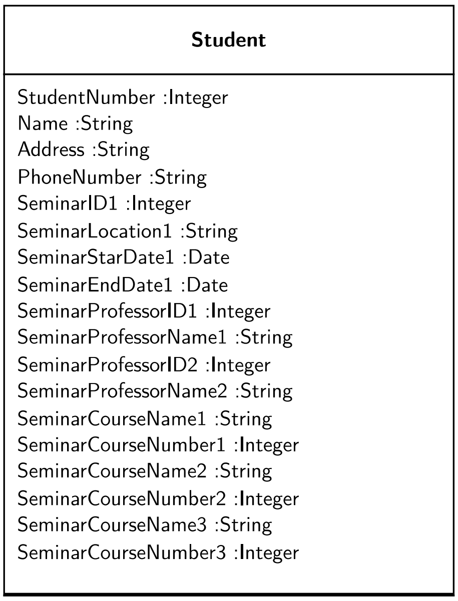

Student
StudentNumber :Integer
Name :String
Address :String
PhoneNumber :String
SeminarID1 :Integer
SeminarLocation1 :String
SeminarStarDate1 :Date
SeminarEndDate1 :Date
SeminarProfessorID1 :Integer
SeminarProfessorName1 :String
SeminarProfessorID2 :Integer
SeminarProfessorName2 :String
SeminarCourseName1 :String
SeminarCourseNumber1 :Integer
SeminarCourseName2 :String
SeminarCourseNumber2 :Integer
SeminarCourseName3 :String
SeminarCourseNumber3 :Integer

Figure 6: Test Case in 0ONF

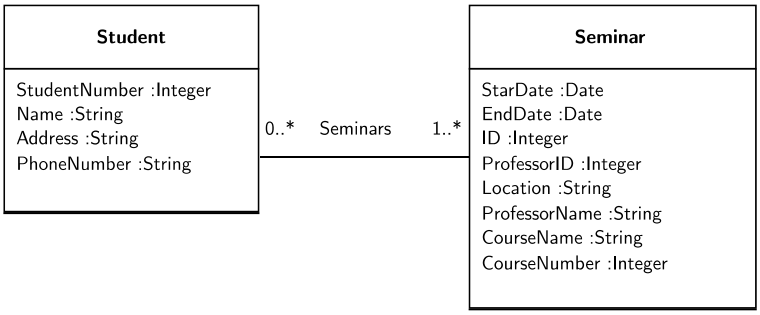


Figure 7: Test Case in 1ONF

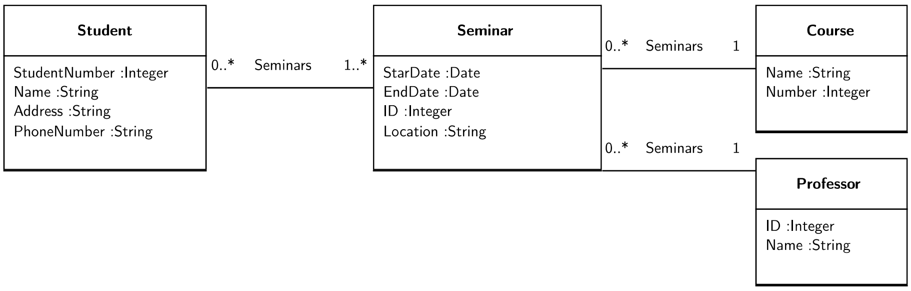


Figure 8: Test Case in 2ONF

been mentioned in previous section.

To have the whole transformation process complete, we introduce also the model design in 3ONF, however, the final transformation from 2ONF into 3ONF is done manually.

In Figure 8, the *Student* class encapsulates the behavior for both students and addresses. The first step would be to refactor *Student* into two classes, *Student* and *Address*. This would make our design more cohesive and more flexible because there is a very good

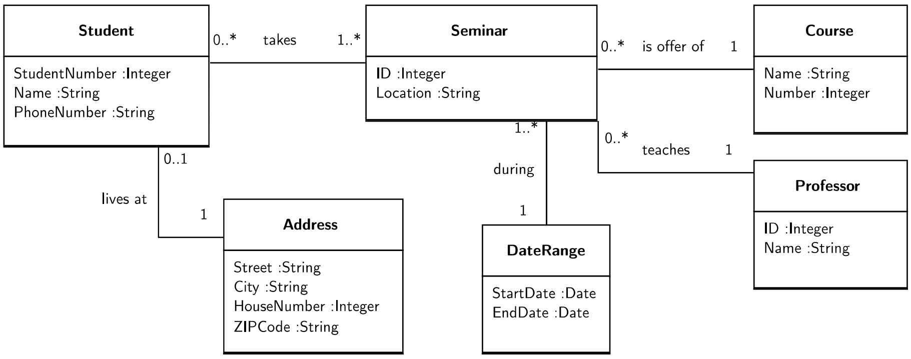


Figure 9: Test Case in 3ONF

chance that students are not the only things that have addresses. This realization leads to the class diagram presented in Figure 9.

We are still not done, because the *Seminar* class of Figure 8 implements “date range” behavior. It has a start date and an end date. Because this sort of behavior forms a cohesive whole, and because it is more than likely needed in other places, it makes sense to introduce the class *DateRange* of Figure 9.

5 Conclusion

In this paper we have shown that the principle of normalization can be applied to object-oriented design. This framework thus provides us with a formal mechanism of methodically analyzing an object-oriented design and improving its overall quality by applying these normalizations in a systematic and scientific manner.

Moreover, all normalization processes have been automated using our transformation algorithms. All classes containing well-named, unique attributes are transformable, firstly, from 0ONF into 1ONF, secondly, from 1ONF into 2ONF. The requirement on well-named and unique attributes in classes is the constraint of our research and it is absolutely essential, since we have used them for identification of particular transformation rules.

Our future research can focus on several directions. One direction could be describing the rules of our object-oriented normal forms as a sequence of refactoring steps. The algorithm for transformation from 2ONF into 3ONF is also still waiting for a deeper investigation. Moreover, there is also possibility to make our algorithms less dependent on well-named attributes and their uniqueness in our model.

References

[1] S. W. Ambler. Introduction to class normalization. `http://www.agiledata.org/essays/classNormalization.html`, (2016). [Online; accessed 1-March-2016].

[2] Byung S. Lee. *Normalization in OODB Design*, volume 24, chapter ACM SIGMOD Record, 23–27. ACM New York, NY, USA, (1995).

[3] R. G. Catell. *The Object Data Standard: ODMG 3.0.* Morgan Kaufmann, (2000).

[4] Craft.CASE Ltd. Official web sites of Craft.CASE tool. `http://www.craftcase.com/`, (2016). [Online; accessed 1-March-2016].

[5] e-FRACTAL. *Craft.CASE scripting manual.* `http://www.e-fractal.cz/`, (November 2015).

[6] A. Goldberg and K. S. Rubin. *Succeeding with objects: decision frameworks for project management.* Addison-Weysley, (1995).

[7] F. Lodhi and H. Mehdi. *Multi Topic Conference, 2003. INMIC 2003. 7th International*, chapter Normalization of object-oriented design, 446–450. IEEE, (2003).

[8] G. S. A. Mala and G. V. Uma. *PRICAI 2006: Trends in Artificial Intelligence: 9th Pacific Rim International Conference on Artificial Intelligence Guilin, China, August 7-11, 2006 Proceedings*, chapter Automatic Construction of Object Oriented Design Models [UML Diagrams] from Natural Language Requirements Specification, 1155–1159. Springer Berlin Heidelberg, Berlin, Heidelberg, (2006).

[9] V. Merunka. *Objektové modelování.* Alfa Nakladatelství, s.r.o., (2008).

[10] V. Merunka, O. Nouza, and J. Brožek. *Lecture Notes in Business Information Processing*, chapter Automated Model Transformations Using the C.C Language, 137–151. Berlin: Springer, (2008).

[11] V. Merunka and J. Tůma. *Innovations and Advances in Computer, Information, Systems Sciences, and Engineering*, chapter Normalization Rules of the Object-Oriented Data Model, 1077–1089. Springer New York, New York, NY, (2013).

[12] S. Montgomery. *Object-Oriented Information Engineering: Analysis, Design, and Implementation.* Academic Press, (2012).

[13] Z. Tari, J. Stokes, and S. Spaccapietra. *Object Normal Forms and Dependency Constraints for Object-Oriented Schemata*, volume 22, chapter ACM Transactions on Database Systems, 513–569. ACM New York, NY, USA, (1997).

[14] Wai Yin Mok, Yiu-Kai Ng and D. W. Embley. *Data and Knowledge Engineering: Theory and Applications*, chapter An Improved Nested Normal Form for Use in Object-Oriented Software Systems, 446–452. (1992). Proceedings of the 2nd International Computer Science Conference, Hong Kong.

[15] Yonghui Wu, Zhou Aoying. *Info-Tech and Info-Net*, volume 5, chapter Research on Normalization Design for Complex Object Schemes, 101–106. IEEE, (2001). Proceedings of ICII 2001, Beijing.

Permutation Entropy in Signal Analysis*

Lucie Tylová

5th year of PGS, email: lucie.tylova@fjfi.cvut.cz
Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jaromír Kukal, Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. There are many measures of chaotic behaviour: Hurst and Lyapunov exponents, various dimensions of attractor, various entropy measures, etc. Permutation entropy of equidistantly sampled data is preferred for analysing EEG signal which is a chaotic system. The novelty of the approach is in bias reduction of permutation entropy estimates, memory decrease, and time complexities of permutation analysis. Therefore, EEG signal and permutation sample lengths are not limitation. This general method was used for channel by channel analysis of Alzheimer's diseased (AD) and healthy (CN) patients to point out the differences between AD and CN groups. The technique also enables to study the influence of EEG sampling frequency in wide range.

Keywords: permutation entropy, unbiased estimation, EEG, Alzheimer's disease, hash table, resampling

Abstrakt. Existuje mnoho měr chaotického chování: Hurstovy a Lyapunovy exponenty, různé dimenze atraktorů, různé míry entropie, atd. K analýze chaotického EEG signálu je zde použita permutační entropie konstantně vzorkovaných dat. Vylepšením daného postupu je redukce odchylky odhadu permutační entropie, nižší paměťová a časová náročnost permutační analýzy. Z tohoto důvodu délka EEG signálu a permutačních vzorků není limitující. Tato metoda byla použita pro analýzu po kanál po kanále, aby určila rozdíl mezi skupinami zdravých pacientů a pacientů s Alzheimerovou chorobou. Tento postup také umožňuje studovat vliv vzorkovací frekvence.

Klíčová slova: permutační entropie, nestranný odhad, EEG, Alzheimerova choroba, hešovací tabulka, převzorkování

1 Introduction

Alzheimer's disease (AD) is the most common form of dementia, which gradually destroys the host's brain cells. Recent findings estimate that 35 million people worldwide currently suffer from AD. Clinically, AD manifests itself as a slowly progressing impairment of mental functions whose course lasts several years prior to the death of the patient. Structural changes in AD are related to the accumulation of amyloid plaques between nerve cells in the brain and with the appearance of neurofibrillary tangles inside nerve cells, particularly in the hippocampus and the cerebral cortex. Although a definite diagnosis is possible only by necropsy, a differential diagnosis with other types of dementia and with

*The paper was created with the support of CTU in Prague, Grant SGS14/208/OHK4/3T/14.

major depression should be attempted. Magnetic resonance imaging and computerized tomography can be normal in the early stages of AD, but a diffuse cortical atrophy is the main sign in brain scans. Mental status tests are also useful.

It has been shown that AD patients have lower correlation dimension (D_2) values as a measure of the underlying system dimensional complexity than control subjects [12]. Furthermore, AD patients also have significantly lower values of the largest Lyapunov (λ_1) exponent than controls in almost all EEG channels. However, estimating the non-linear dynamic complexity of physiological data using measures such as D_2 and λ_1 is problematic, as the amount of data required for meaningful results in their computation is beyond the experimental possibilities for physiological data [8]. One alternative solution lies in computing the entropy of the EEG [1]. The concept of entropy has achieved a large consensus as an indicator of complexity of nonlinear signals [10], [9]. Dauwels et al. [5] and many other authors have shown that Alzheimer's disease increases power in the delta and theta-bands in the case of EEG analysis in frequency domain but the power spectrum is a global characteristics of EEG signal which disables to study local events in the signal. A number of variants of this notion have been proposed in the literature which show different degrees of flexibility, relevance to different problems, efficiency in their computation, as well as theoretical foundations. This work investigates the potential of complexity analysis of multidimensional EEG as indicator of AD onset through permutation entropic modelling.

2 Permutation entropy

2.1 Shanon entropy and its estimation

Definition. Shannon entropy [11] H_S of a discrete random variable X with possible values $x_1, ..., x_m$ and probability mass function $p(X)$ is defined as

$$H_\mathrm{S} = -\sum_{i=1}^{m} p_i \ln p_i, \tag{1}$$

where $p_i = p(x_i)$.

If the probability function is unknown for an experimental data set, and the number of possible values is finite for random variable X, I estimate probability function p_i by relative frequency $p_{j,\mathrm{N}}$ and number of events k_N as

$$p_{j,\mathrm{N}} = \frac{n_j}{n}, \tag{2}$$

$$k_\mathrm{N} = \sum_{n_j>0} 1 \leq k, \tag{3}$$

where n_j is the number of occurrences x_i of random variable X, and n the total number of measurement results. Then I get *naive estimate* of Shannon entropy as

$$H_\mathrm{N} = -\sum_{j=1}^{k_\mathrm{N}} p_{j,\mathrm{N}} \ln p_{j,\mathrm{N}}. \tag{4}$$

This estimate is biased, and therefore it has a systematic error.

Miller [7] modified *naive estimate* H_{N} using first order Taylor expansion, which produces better estimation

$$H_{\mathrm{M}} = H_{\mathrm{N}} + \frac{k_{\mathrm{N}} - 1}{2n}. \tag{5}$$

2.2 Application to permutation analysis

Entropy estimates can be easily applied to permutation event analysis [2], [4]. Methodology from [7] estimates a smaller bias. Let time series be $\{a_k\}_{k=1}^{T}$ and sliding window $\{b_k\}_{k=1}^{w}$ of length w, then I can substitute signal values b_k in the window with their orders and then obtain permutation pattern $\{\pi_k\}_{k=1}^{w}$.

The universe of random variable X is a set of all permutation of length w. Therefore, the number of possible permutations is

$$m = w!, \tag{6}$$

but the number of various permutations in given signal cannot exceed the number of sliding samples as

$$k_n \leq n = T - w + 1. \tag{7}$$

The number of occurrences of j^{th} permutation pattern corresponds with n_j, and n is the total number of samples. The difference between typical AD and CN patients are illustrated in Fig. 1. Supposing ordering $n_{(j)} \geq n_{(j+1)}$ for j = 1,...,m-1, ten most frequent permutation patterns ($n_{(1)}$,...,$n_{(10)}$) were plotted to union diagram for f_{s} = 200 Hz, w = 14, ch = 8. In the case of AD, I observed systematic increasing or decreasing of EEG signal with small fraction of exceptions. But in the case of CN, the patterns rarely increased, no systematic decreasing was observed and EEG signal exhibited higher diversity. Therefore, this primary observation is in agreement with hypothesis of diminished EEG signal entropy in the case of AD.

Now, (4) can be used directly and the biased naive estimation H_{N} calculated as in [11]. Our methodology is based on Miller's approach [7] and direct application of (5) to permutation patterns. The difference between estimates (4) and (5) varies according to number of distinct patterns and time series length.

3 Permutation analysis for large samples

The main disadvantage of the original procedure of permutation analysis [2] is in its memory and time complexities. They realized permutation memory as a matrix of w columns and $w!$ rows together with counter vector of length $w!$. It enables permutation analysis only for $w < 13$ on a typical computer. Traditional applications [2] of permutation entropy apply window of length $w < 8$. The time complexity of single permutation counting is also $w!$, in the worst case. Therefore, I decided to use more sophisticated data structure for permutation analysis. There are many data structures and algorithms for realizing of *look-up table* as a kind of memory with fast access. Our memory has to

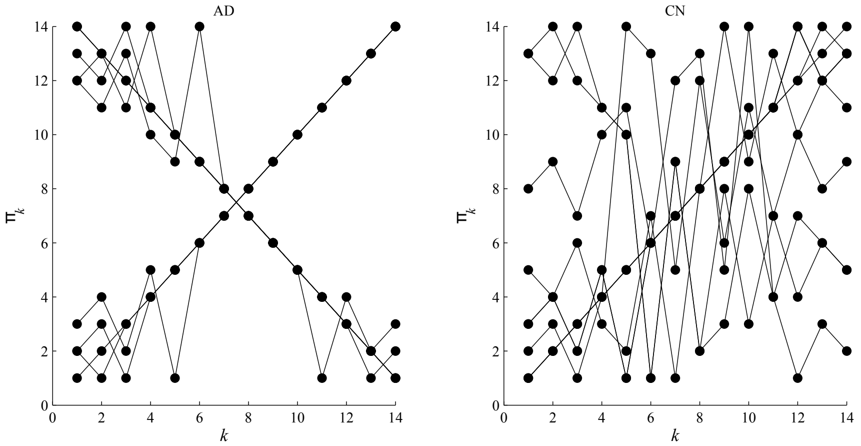


Figure 1: Ten most frequent permutation patterns as union plot for 8th EEG channel

be optimized only for two operations: FIND and INSERT. I used *hash table* with open addressing and linear probe strategy [6] as a model, which is easy to realize. Let $P > n$ be the optional prime number. Then the *loading factor* is defined as a ratio $\alpha = n/P < 1$. The mean number of permutation vector comparisons during successful FIND operation was determined [6] as

$$\mathrm{E}T_{\mathrm{OPT}} = \frac{1}{2}\left(1 + \frac{1}{1-\alpha}\right). \tag{8}$$

In the case of unsuccessful FIND operation and INSERT operation, the mean number of permutation vector comparisons is higher [6] than in the previous optimistic case

$$\mathrm{E}T_{\mathrm{PES}} = \frac{1}{2}\left(1 + \frac{1}{(1-\alpha)^2}\right). \tag{9}$$

Our tiny and fast implementation of permutation memory is a matrix of occurred permutations with w columns and $P > n$ rows together with counter vector of length P. The time complexity of single permutation counting is constant and dependent only on loading factor in the best (8) and worst (9) cases. It enables very fast permutation analysis for higher sample length w and long EEG sequences. The last implementation detail is how to realize hash function $\mathit{index} = \mathrm{h}(\boldsymbol{\pi})$ for given permutation pattern $\boldsymbol{\pi}$. By subtracting vector of units from vector $\boldsymbol{\pi}$, I obtain digital form $\boldsymbol{y} = \boldsymbol{\pi} - 1$ in the first step. Let $R = w$ be the base of digital system. In the second step, I calculate the value v of $\boldsymbol{y}$ according to base R. The resulting index into hash table has a value $\mathit{index} = v \bmod P$. In the case when $P > 3n$, we have $\alpha < 1/3$ and then the mean number of trials is less than 1.25 in the optimistic case (8) and less than 1.625 in the pessimistic (9).

The main argumentation against large window size w is in sparse sampling effect, which increases the variance of permutation entropy estimates. That is why many authors [2] prefer short windows with $w < 8$. Despite of window size constrain recommendation, large window can be efficient in classification tasks where between close differences in mean permutation entropy can also increase. There are two main aims related to classification power of permutation entropy in given classification task:

- Find optimum window length for given sampling frequency f_s which causes the best class separation measured as p_value of statistical testing.
- Find optimum sampling period f_s which also minimize p_value. It is also useful for the decreasing of length w using smaller sampling frequency f_s.

Resulting optimum values w, f_s, p_value will help to decide whether the constrain $w < 8$, undersampling and oversampling are useful in given task.

4 Application to EEG

Permutation entropy was applied to EEG signals obtained from two groups of patients. EEG data were obtained during examinations of 10 patients with moderate dementia (MMSE score 10-19). All subjects underwent brain CT, neurological and neuropsychological examinations. The other group is a control set consisting of 10 age-matched, healthy subjects who had no memory or other cognitive impairments. The average MMSE of the AD group is 16.2 (SD of 2.1). The ages of the two groups are 69.4 ± 9.2 in Alzheimer's group and 68.7 ± 7.7 in normal group, respectively. Informed consent was obtained from all included subjects and the study was approved by the local ethics committee. All recordings were performed under similar standard conditions. The subjects were in a comfortable position, on a bed, with their eyes closed. Electrodes were positioned according to the 10-20 system of electrode placement; the recording was conducted on a 21-channel digital EEG setup (TruScan 32, Alien Technik Ltd., Czech Republic) with a 22-bit AD conversion and a sampling frequency of 200 Hz. The linked ears were used as references. Stored digitized data were zero-phase digitally filtered using a bandpass FIR filter (100 coefficients, Hamming window) of 0.5-60 Hz and a bandstop filter of 49-51 Hz [6]. The analysis started by manual artefact removal. Time series length T varies between 70000 and 120000. I tried to separate these two groups of patients by two-sample t-test with null hypotheses and alternative hypothesis as

$$\mathrm{H}_0 : \mathrm{E}\hat{H}(\mathrm{AD}) = \mathrm{E}\hat{H}(\mathrm{CN}), \tag{10}$$

$$\mathrm{H}_\mathrm{A} : \mathrm{E}\hat{H}(\mathrm{AD}) \neq \mathrm{E}\hat{H}(\mathrm{CN}). \tag{11}$$

Using linear interpolation I performed resampling of original 200 Hz EEG signal to frequencies from 50 Hz to 500 Hz. For every EEG channel and sampling frequency, I found optimum window length which minimizes p_value of two-sided two-sampled t-test in the case of naive permutation entropy estimation. Using critical value 0.05, I recommend to use sampling frequency f_s up to 200 Hz for all channels. Frequency 300 Hz is suitable only for 2-12 channels and higher rate of sampling is not recommended. Minimum $p_\mathrm{value} = 0.0009$ was obtained for 8th channel and $f_\mathrm{s} = 100$ Hz using window length $w = 10$ which is in contradiction to generally accepted condition $w \leq 7$ [1], [8]. Previous constrain enforces to use $f_\mathrm{s} = 20$ Hz only (signal decimation) and $w = 4$ which brings significant AD/CN differences for 12, 14-17, 19 channels.

All the calculations were also performed for Miller correct permutation entropy with optimum window lengths in Tab. 1 and p-values in Tab. 2. Using the same critical level 0.05 in the case of Miller correction, I obtained similar results as without it. Sampling

Table 1: Optimum window length w in the case of Miller permutation entropy

Channel	f_s [Hz]								
	10	20	50	70	100	150	200	300	500
1	13	10	10	10	10	13	14	15	15
2	9	11	10	10	10	12	14	15	15
3	10	10	13	12	11	13	14	15	15
4	9	14	14	9	10	12	14	15	15
5	15	15	10	9	10	12	14	15	15
6	13	11	11	9	10	12	14	15	15
7	11	10	11	10	10	12	14	15	15
8	12	11	9	9	10	12	13	15	15
9	10	15	13	9	10	12	14	15	15
10	14	11	13	10	10	12	14	15	15
11	10	11	15	15	10	12	14	15	15
12	13	4	15	15	10	12	14	15	15
13	15	15	13	14	14	13	14	15	4
14	10	4	15	14	10	12	14	15	7
15	14	4	13	14	10	12	14	15	15
16	13	4	15	15	14	13	15	15	7
17	12	4	15	15	11	13	15	15	7
18	15	14	15	15	15	15	15	5	7
19	13	4	13	14	15	15	15	5	7

frequencies up to 200 Hz are suitable for all channels again. Frequency 300 Hz can be used only for channels 4, 8-10, 12. Optimum sampling frequency $f_\mathrm{s} = 100$ Hz with window length $w = 10$ brought $p_\mathrm{value} = 0.0006$ in the case of 8th channel, which is slightly better value then without Miller correction. Therefore, I suppose Miller correction of permutation entropy has positive effect mainly in the case of well separated group data. Traditional window length $w \leq 7$ was observed only for $f_\mathrm{s} = 20$ Hz on channels 12, 14-17, 19 again.

The p-values in Tab. 2 are results of multiple testing and Bonferroni correction is necessary. Using False Discovery Rate (FDR) methodology [3] I obtained corrected critical value $\alpha_\mathrm{FDR} = 0.0391 < 0.05$. This correction eliminated only several channels for $f_\mathrm{s} = 300$ Hz in both approaches to entropy estimations. Therefore, only $f_\mathrm{s} \leq 200$ Hz is recommended for all channels to produce significant AD/CN entropy differences under stringent FDR conditions. The difference between naive and Miller estimates is not constant because both EEG signal length and the number of occurring patterns vary within patient groups. Therefore, Miller estimate of permutation entropy causes results which differ from naive approach. Fortunately, novel estimate generates results with more clear biomedical interpretation.

5 Conclusion

Using Miller's approach instead of direct calculation of Shannon's entropy from permutation frequencies, I have developed a novel method of EEG analysis via permutation entropy. The second advantage of our method is in its very fast permutation analysis and low consumption of computer memory which enables analysis of large time series with greater length of permutation patterns and also parametric study of sampling frequency role. The method was applied to diagnose Alzheimer's disease from 19 channel EEG. I focused on the influence of sampling frequency, channel choice, and window length on separation power of permutation entropy, which were evaluated as p_{value} of standard two-sided t-test.

Significant results were obtained for raw EEG and its undersampling (10 Hz $\leq f_{\text{s}} \leq$ 200 Hz) meanwhile oversampling is not recommended strategy. The best result for naive permutation entropy was obtained for $f_{\text{s}} = 100$ Hz, $w = 10$, and 8th channel ($p_{\text{value}} = 0.0009$). Permutation entropy with Miller correction offered similar results with optimum settings $f_{\text{s}} = 100$ Hz, $w = 10$, and 8th channel ($p_{\text{value}} = 0.0006$). Therefore, Miller correction slightly improve the separation power but from the statistical point of view, all procedures with significant results are equivalent.

Table 2: Optimum p_{value} for $\mathrm{H}_0 : \mathrm{E}\hat{H}(\mathrm{AD}) = \mathrm{E}\hat{H}(\mathrm{CN})$ in the case of Miller permutation entropy

Channel	f_{s} [Hz]								
	10	20	50	70	100	150	200	300	500
1	0.0156	0.0162	0.0158	0.0132	0.0099	0.0088	0.0084	0.0817	0.6310
2	0.0156	0.0161	0.0161	0.0114	0.0054	0.0060	0.0058	0.0540	0.5578
3	0.0170	0.0159	0.0161	0.0168	0.0112	0.0104	0.0091	0.0579	0.5156
4	0.0144	0.0162	0.0163	0.0082	0.0045	0.0049	0.0048	0.0391	0.3607
5	0.0163	0.0150	0.0170	0.0079	0.0046	0.0045	0.0042	0.0465	0.4374
6	0.0159	0.0163	0.0159	0.0065	0.0032	0.0038	0.0042	0.0463	0.5431
7	0.0155	0.0155	0.0158	0.0116	0.0031	0.0039	0.0056	0.0456	0.6880
8	0.0158	0.0155	0.0111	0.0018	0.0006	0.0007	0.0013	0.0074	0.2396
9	0.0163	0.0163	0.0160	0.0100	0.0031	0.0031	0.0039	0.0220	0.2575
10	0.0153	0.0158	0.0163	0.0121	0.0042	0.0040	0.0042	0.0373	0.3840
11	0.0152	0.0165	0.0170	0.0159	0.0058	0.0049	0.0053	0.0550	0.5570
12	0.0150	0.0118	0.0164	0.0162	0.0085	0.0058	0.0067	0.0427	0.6866
13	0.0157	0.0148	0.0166	0.0168	0.0169	0.0173	0.0155	0.1109	0.4494
14	0.0158	0.0035	0.0166	0.0168	0.0131	0.0098	0.0098	0.1026	0.6807
15	0.0149	0.0061	0.0162	0.0169	0.0120	0.0091	0.0086	0.0871	0.6892
16	0.0159	0.0010	0.0162	0.0172	0.0165	0.0132	0.0135	0.1960	0.5549
17	0.0162	0.0036	0.0166	0.0169	0.0162	0.0145	0.0148	0.1410	0.4241
18	0.0157	0.0161	0.0169	0.0165	0.0169	0.0201	0.0251	0.1655	0.1653
19	0.0160	0.0040	0.0170	0.0167	0.0168	0.0178	0.0170	0.2493	0.2573

References

[1] D. E. Abasolo, R. Hornero, and P. Espino. *Approximate entropy of eeg background activity in alzheimer's disease patients.* Intelligent Automation and Soft Computing **15** (2009), 591–603.

[2] C. Bandt and B. Pompe. *Permutation entropy: A natural complexity measure for time series.* Physical Review Letters **88** (2002).

[3] Y. Benjamini and Y. Hochberg. *Multiple hypotheses testing with weights.* Scandinavian Journal of Statistics **24** (1997), 407–418.

[4] Y. Cao, W.-w. Tung, J. B. Gao, V. A. Protopopescu, and L. M. Hively. *Detecting dynamical changes in time series using the permutation entropy.* Physical Review E **70** (2004), 046217.

[5] J. Dauwels, F. Vialatte, and A. Cichocki. *Diagnosis of alzheimer's disease from eeg signals: Where are we standing?* Current Alzheimer Research **7** (2010), 487–505.

[6] D. E. Knuth. *Art of Programming. Volume 3: Sorting and Searching.* Addison-Wesley, Reading, (1998).

[7] G. A. Miller. *Note on the bias of information estimates.* Information Theory in Psychology: Problems and Methods **2** (1955), 95–100.

[8] F. C. Morabito, D. Labate, F. La Foresta, A. Bramanti, G. Morabito, and I. Palamara. *Multivariate multi-scale permutation entropy for complexity analysis of alzheimer's.* Entropy **14** (2012), 1186–1202.

[9] J.-H. Park, S. Kim, C.-H. Kim, A. Cichocki, and K. Kim. *Multiscale entropy analysis of eeg from patients under different pathological conditions.* Fractals **15** (2007), 399–404.

[10] S. M. Pincus. *Approximate entropy as a measure of system complexity.* Proceedings of the National Academy of Sciences of the United States of America **88** (1991), 2297–2301.

[11] C. E. Shannon. *A mathematical theory of communication.* The Bell System Technical Journal **27** (1948), 623–656.

[12] S. Tang, X. Jiang, Z. Liu, L. Ma, Z. Zhang, and Z. Zheng. *Entropy analysis in interacting diffusion systems on complex networks.* International Journal of Mathematics and Computers in Simulation **5** (2011), 118–125.

Proximal Algorithms in MRI Data Reconstruction

Hynek Walner

2nd year of PGS, email: walner@utia.cas.cz
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:
Jiří Boldyš, Department of Image Processing
Institute of Information Theory and Automation, CAS
Michal Šorel, Department of Image Processing
Institute of Information Theory and Automation, CAS
Jiří Dvořák, Department of Image Processing
Institute of Information Theory and Automation, CAS

Abstract. Algorithms based on proximal operators find their use in many optimization problems, such as matrix completion in computer vision or reconstruction in image processing. Brief introduction to proximal algorithms will be presented together with connection to standard methods like gradient descent or dual formulation. Furthermore, medical image reconstruction will be formulated as a variational problem using total variation regularization ready to be solved using presented methods. Finally, we will demonstrate and compare selected methods on real data acquired from MRI scanner at BTU in Brno and propose further extension of current model.

Keywords: image reconstruction, TV regularization, proximal algorithms

Abstrakt. Algoritmy založené na proximálních operátorech jsou často využívány v mnoha optimalizačních úlohách, např. doplnění dat ve strojovém učení nebo rekonstrukci obrazu. Provedeme krátké shrnutí proximálních operátorů a porovnáme je se standardními metodami, jako metodou gradientního sestupu nebo duální formulací optimalizačních úloh. Dále formulujeme rekonstrukci zdravotnických dat jakožto úlohu variačního počtu s regularizací ve tvaru totální variace v takovém tvaru, aby byla řešitelná uvedenými metodami. Nakonec vybrané algoritmy předvedeme a srovnáme na datech ze skeneru využívající magnetickou rezonanci umístěného na VUT v Brně a navrhneme další rozšíření modelu.

Klíčová slova: rekonstrukce obrazu, TV regularizace, proximální algorithmy

1 Introduction

Many inverse imaging problems such as image denoising, image deconvolution or image signal reconstrucion can be conveniently formulated as a variational problem

$$\min_{x\in\mathbb{R}^2}\left\{\lambda\int_{\Omega}|K(x)|+\frac{1}{2}\|y-Ax\|_2^2\right\},\tag{1}$$

where $\Omega \subset \mathbb{R}^2$ is image domain, $x \in L^1(\Omega)$ is the desired solution and $y \in L^1(\Omega)$ is the original data to be reconstructed. Parameter $\lambda \in \mathbb{R}_0^+$ scales the trade-off between "data" term and regularization term. Data term ensures closeness of solution to input, whereas regularization represents effort to improve visual features of image. Operator A represent transformation of output to comparable domain in which y is acquired. In medical imaging, A typically represents Fourier transformation. If K is assumed to be gradient of input image, proposed model (1) becames so-called Total Variation (TV) regularization model (or ROF model) introduced in [1]. Major advantage of incorporating TV regularization is allowing appearance of sharp discontinuities in the solution. This fact is often sought after in image processing, since edges represent important features such as boundaries of objects. However this formulation of cost functional (1) leads to difficult minimization, given the non-smooth property of the total variation. We will introduce several algorithms based on proximal operators, which can be successfully used to tackle such problems with application to MRI data reconstruction.

2 Proximal algorithms

Proximal algorithms can be percieved as a generalization of standard gradient descent. Let us suppose, that we want to solve

$$\min_{x \in \mathbb{R}^n} f(x) = \min_{x \in \mathbb{R}^n} g(x) + h(x) \tag{2}$$

where $g : \mathbb{R}^n \mapsto \mathbb{R}^n$ is convex and differentiable while $h : \mathbb{R}^n \mapsto \mathbb{R}^n$ is only convex but not necessarily differentiable. Instead of making quadratic approximation of f around x with step size $t \in \mathbb{R}^+$ to get gradient descent update for case f both convex and differentiable, it is possible to approximate only g while h stays in its original form to obtain following

$$\begin{aligned} x^+ =& \operatorname*{argmin}_z \left\{ g(x) + \nabla g(x)^T (z - x) + \frac{1}{2t} \|z - x\|_2^2 + h(z) \right\} \\ =& \operatorname*{argmin}_z \left\{ \frac{1}{2t} \left(\|z - x\|_2^2 + 2t \nabla g(x)^T (z - x) + t^2 \|\nabla g(x)\|_2^2 \right) + g(x) - \frac{2}{t} \|\nabla g(x)\|_2^2 + h(z) \right\} \\ =& \operatorname*{argmin}_z \left\{ \frac{1}{2t} \|z - (x - t\nabla g(x))\|_2^2 + h(z) \right\} \\ =& \operatorname{prox}_{t,h}((x - t\nabla g(x))), \end{aligned}$$

where we denoted minimizing term by symbol prox. Components in prox forces update to be as close to gradient step of g as possible and keeps values of h small. Using this intuitive derivation, we can formally define *proximal operator* $\operatorname{prox}_{t,h} : \mathbb{R}^n \mapsto \mathbb{R}^n$ by

$$\operatorname{prox}_{t,h}(x) = \operatorname*{argmin}_z \left\{ \frac{1}{2t} \|z - x\|_2^2 + h(z) \right\}.$$

Combining proximal operator with gradient descent, leads to writing minimizing algorithm of (2) as

Algorithm 1 General proximal operator minimization

```
1. Initialize $x^0 \in \mathbb{R}^n$.
2. Let $x^+ = (x^{k-1} - t\nabla g(x^{k-1}))$.
3. Define $x^k = \mathrm{prox}_{t_k,h}(x^+)$.
```

Last step can we also written in gradient descent manner as

$$x^k = x^{k-1} - t_k G_{t_k}(x^{k-1}), \quad G_t(x) = \frac{x - \mathrm{prox}_{t,h}(x - t\nabla g(x))}{t},$$

where $G_t(x)$ is so-called *generalized gradient.* Notice, that evaluation of proximal operator depends only on gradient of g and h itself, thus it can be conveniently used when proximal operator of h is known. Algorithms using proximal operators are especially useful, when evaluating of $\mathrm{prox}_{t,h}(x)$ is sufficiently quick. Generally speaking, one can achieve rate of convergence of order $O(1/k^2)$, or even $O(1/e^k)$ for special cases, whereas subgradient descent methods converge at $O(1/\sqrt{k})$. This is the case of h in form of Total Variation as presented by (1). In following sections, we will present several algorithms that employ different approach for evaluating proximal operator.

2.1 FISTA

Standard minimizing procedure based on proximal operator is Iterative Shrinkage Thresholding Algorithm (or ISTA) and its accelerated version Fast ISTA (or FISTA). Let us consider minimization problem (2) with Total Variation regularization, i.e.

$$\min_x \frac{1}{2}\|x\|_2^2 + \lambda\|x\|_1.$$

Respective proximal operator is of form

$$\mathrm{prox}_{\mathrm{t},\|\cdot\|_1}(x) = \operatorname*{argmin}_z \left\{ \frac{1}{2t}\|z - x\|_2^2 + \lambda\|z\|_1 \right\} \tag{3}$$

where solution to this equation can be written as a *soft thresholding operator* $S_{\lambda t}(x)$ where

$$S_\lambda(x) = \mathrm{sgn}(x)(|x| - \lambda)_+.$$

It can be easily shown, that $S_{\lambda t}(x)$ minimizes term in (3) and is easily numerically computed.

Let us now develop FISTA algorithm with regards to minimizing LASSO (Least Absolute Shrinkage and Selection Operator) problem

$$\min_x \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1$$

which is simpler version of general model (1). Gradient of g is computed as

$$\nabla g(x) = -A^T(y - Ax).$$

and when plugged into Algorithm 1 together with soft thresholding operator, we get easily computable steps which minimize LASSO problem. Such algorithm is known as ISTA. For the sake of completeness, we will present its accelerated version, which was actually implemented. Faster version of ISTA interpolates results from two consecutive steps in addition to original algorithm.

Algorithm 2 FISTA algorithm

```
1. Initialize $x^0, y^1 \in \mathbb{R}^n$, $\alpha_0 = 0$ and let $\alpha_k = \frac{1+\sqrt{1+4\alpha_{s-1}^2}}{2}$, $\gamma_0 = \frac{1-\alpha_s}{\alpha_{s+1}}$.
2. Let $x^+ = (x^{k-1} + tA^T(y - Ax))$.
3. Set $y^{k+1} = S_{\lambda t}(x^+)$ and $x^k = (1-\gamma_k)y^{k+1} + \gamma_k y^k$.
```

2.2 ADMM

Following algorithm is based on slightly different approach on handling minimization of variational problems with non-smooth regularization. Such method is called Alternating Direction Method of Mutlipliers (ADMM) and is built on minimizing each function from

$$\min_{x\in\mathbb{R}^n} g(x) + h(x)$$

separately. This approach is called dual minimization or Douglas-Racheford splitting and its main advantage is when evaluating proximal operator of $f + g$ is more numerically demanding, than computing each proximal operator separately.

Derivation of ADMM originates from minimizing augmented Lagrangian. Firstly, assume LASSO problem once more and rewrite it as a constrain optimization problem

$$\min_x \frac{1}{2}\|Ax - y\|_2^2 + \lambda\|z\|_1 \quad \text{s.t.} \quad x - z = 0.$$

Furthermore, we write augmented Lagrangian of such problem as

$$L_\rho(x, z, u) = \frac{1}{2}\|Ax - y\|_2^2 + \lambda\|z\|_1 + \rho u^T(x - z) + \frac{\rho}{2}\|x - z\|_2^2. \tag{4}$$

Notice, that additional terms equal to zero at optimal point by definiton of constraint $x - z = 0$. Minimizing of augmented Lagrangian (4) is treated separately over its primal variables x and z.

Finding optimal value $x^\star$ minimizing (4) in its first variable can be easily attained in closed-form solution

$$x^\star = (A^TA + \rho)^{-1}(A^Ty + \rho(z - u)),$$

where attaining optimal $z^\star$ can be percieved via proximal algorithms, namely soft thresholding operator defined in previous section. We define

$$z^\star = S_{\lambda/\rho}(x + u).$$

Finally, dual variable u is updated by gained value of constrain and we can readily write ADMM algorithm. Notable feature of ADMM is, that it converges fast at early stage, but requires fair number of iterations for high precision results.

Algorithm 3 ADMM algorithm for LASSO problem

```
1. Initialize $x^0, y^0, z^0 \in \mathbb{R}^n$, $\rho \in \mathbb{R}^+$.
2. Let $x^k = (A^T A + \rho)^{-1}(A^T y^{k-1} + \rho(z^{k-1} - u^{k-1}))$.
3. Let $z^k = S_{\lambda/\rho}(x^k + u^{k-1})$.
4. Update $u^k = u^k + x^k - z^k$.
```

2.3 Chambolle–Pock

Finally, following algorithm proposed by [3] can be seen as a generalization of ADMM or Arrow-Hurwicz model. It is derivated from primal-dual formulation of original problem (2), which assumes form

$$\min_x \max_p \langle Kx, p\rangle + g(x) - h^*(p), \tag{5}$$

where h^* is convex conjugate of h. By taking consecutive saddle point of primal-dual problem, we obtain iterative steps for finding optimal minimizer of (2). Note, that symbol ∂ refers to subgradient of a function.

Algorithm 4 General Chambolle–Pock algorithm

```
1. Initialize $x^0, p^0 \in \mathbb{R}^n$, $\tau, \sigma \in \mathbb{R}^+$, $\theta \in [0, 1]$ and $z^0 = x^0$.
2. Let $p^k = (I + \sigma\partial h^*)^{-1}(p^{k-1} + \sigma K z^{k-1})$.
3. Let $x^k = (I + \tau\partial g)^{-1}(x^{k-1} - \tau K^* p^{k-1})$.
4. Update $z^k = x^k + \theta(x^k - x^{k-1})$.
```

For detailed derivation of proposed algorithm see [3], we will present its implementation on ROF model (1). According to (5), primal-dual formulation of ROF is

$$\min_x \max_p - \langle x, \mathrm{div} p\rangle + \frac{\lambda}{2}\|x - y\|_2^2 - \delta_P(p),$$

where δ_P is indicator function of convex set P given by $P = \{p : \|p\|_\infty \leq 1\}$. Given this particular form of $h^*(p)$, it is possible to express solution to dual maximization as

$$p = (I + \sigma\partial h^*)^{-1}(\tilde{p}) \Longleftrightarrow p = \frac{\tilde{p}}{\max(1, |\tilde{p}|)},$$

i.e. as a pointwise projection onto L^2 balls. Evaluation of proximal operator with respect to g is again quadratic problem and is given by

$$x = (I + \tau\partial g)^{-1}(\tilde{u}) \Longleftrightarrow x = \frac{\tilde{u} + \tau\lambda y}{1 + \tau\lambda}.$$

Main distinction of Chambolle–Pock algorithm is evaluating convex conjugate of regularization function, rather than its proximal operator. This feature may turn out to be useful, when evaluating proximal operator is analytically or numerically unfeasible.

3 MRI application

Presented algorithms was applied in reconstruction of MRI images. Data originates from Agilent 9.4T MRI Small Animal Scanner which collects signal in k-space (i.e. Fourier domain). Although available data was sampled at full rate (resulting in 128×128 complex matrix), implementation of reconstruction algorithms incorporated "degradation" matrix, which filters out coefficients in signal. This was motivated by preparation to employ reconstruction when using *golden angle* (see [4]) subsampling, which allows acquiring MRI data at faster rate. This leads to ROF model written in following form

$$\min_{x \in \mathbb{R}^2} \left\{ \frac{1}{2} \|y - MFx\|_2^2 + \lambda \|Kx\|_1 \right\},$$

where F corresponds to 2D Fourier transform, M selects used coefficients and K computes image gradient.

To complete overall information regarding implementation, FISTA includes small number of ADMM steps when evaluating proximal operator of total variation. ADMM itselft was computed in Fourier domain, due to simplified calculations. Chambolle–Pock was implemented in its stated form.

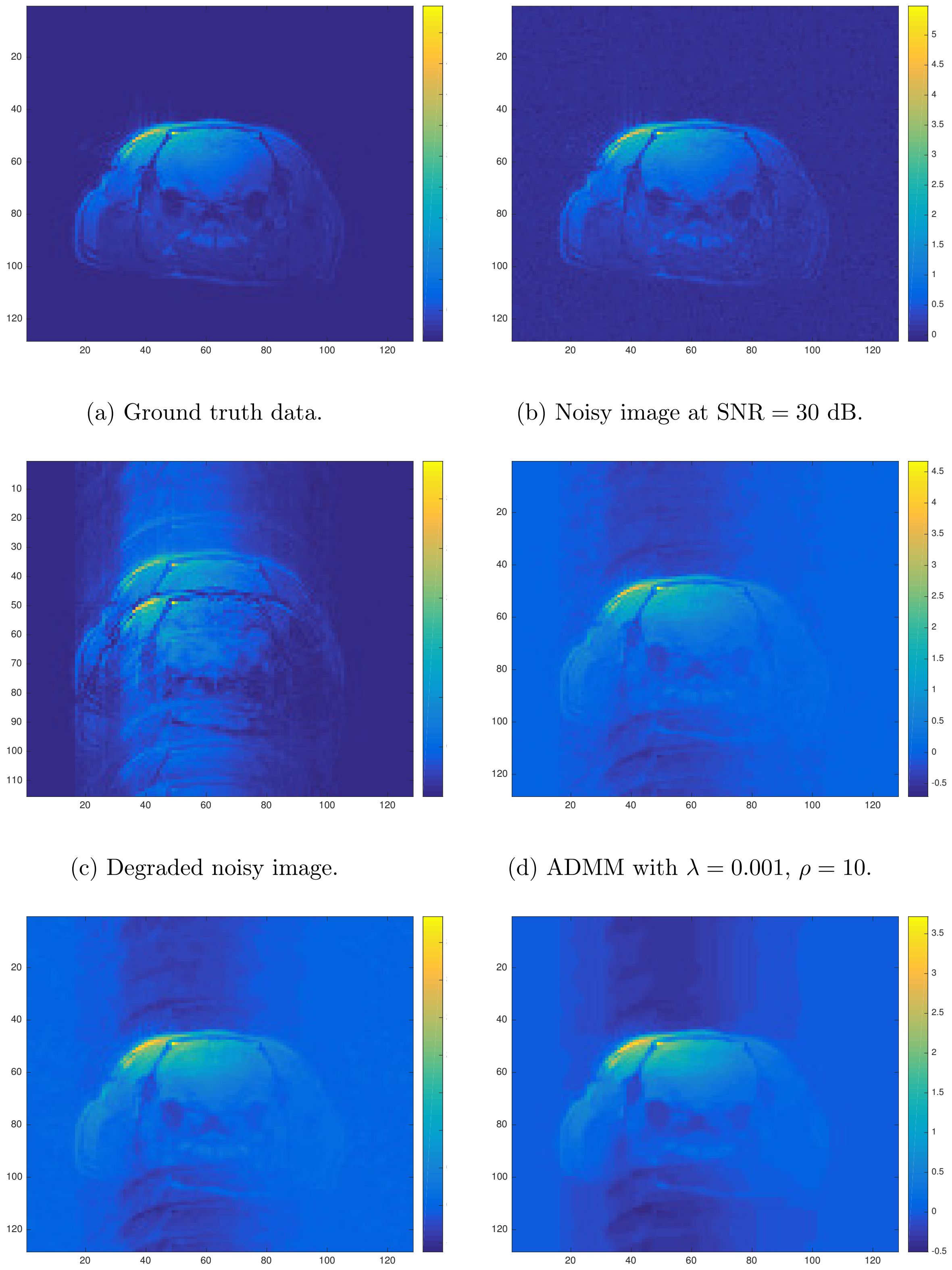
Evaluation of algorithms proceeded as follows. Firstly, ground truth data was corrupted by Additive Gaussian White Noise, with Signal-to-Noise ratio equals to 30dB. Secondly, only 90% of Fourier coeffcients were selected by simple rule of ignoring each 10th component. Reconstruction results for various values of regularization parameter λ can be found on Figure 1. As can be seen on resulting images, proximal algorithms achieve visually very appealing outcomes, given knowledge of degraded process.

In order to compare implemented algorithms by speed of convergence, we firstly performed 10000 iterations of each method (i.e. FISTA, ADMM and Chambolle–Pock) to get as ideal output as possible. Consequently, distance (in terms of Frobenius matrix norm) of generated outcomes and this "ground truth" result was measured and can be seen on Figure 2 and Figure 3.

In general, ADMM converges at fast rate at first, but requires aditional iterations to achieve more precise results, whereas FISTA slowly improves result through every iteration. Chambolle–Pock indicates similar behaviour as ADMM but it is even faster at a early stage. To conclude convergence analysis, we present Table (1) with required iterations to achieve given precision $\varepsilon =$1e-4 for distance between current outcome and "optimal" result. FISTA did not achieve desired stability, as was improving by order of 1e-2 at final iteration.

FISTA	ADMM	Chambolle–Pock
·	6621	9989

Table 1: Number of iterations to reach given stability.



(a) Ground truth data.

(b) Noisy image at SNR = 30 dB.

(c) Degraded noisy image.

(d) ADMM with $\lambda = 0.001$, $\rho = 10$.

(e) ADMM with $\lambda = 0.001$, $\rho = 100$.

(f) ADMM with $\lambda = 0.005$, $\rho = 100$

Figure 1: Reconstruction of degraded image using ADMM method for various weights on regularization.

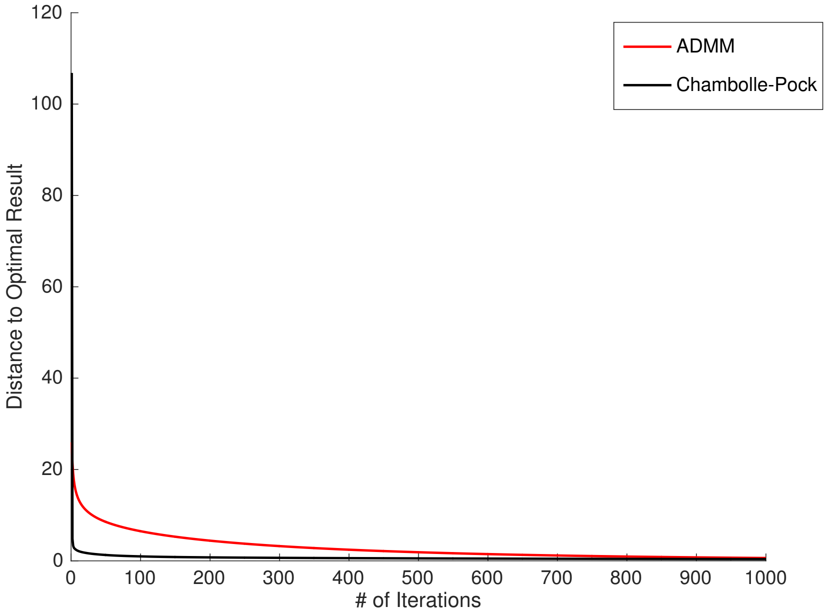


Figure 2: ADMM and Chambolle–Pock convergence speed comparison.

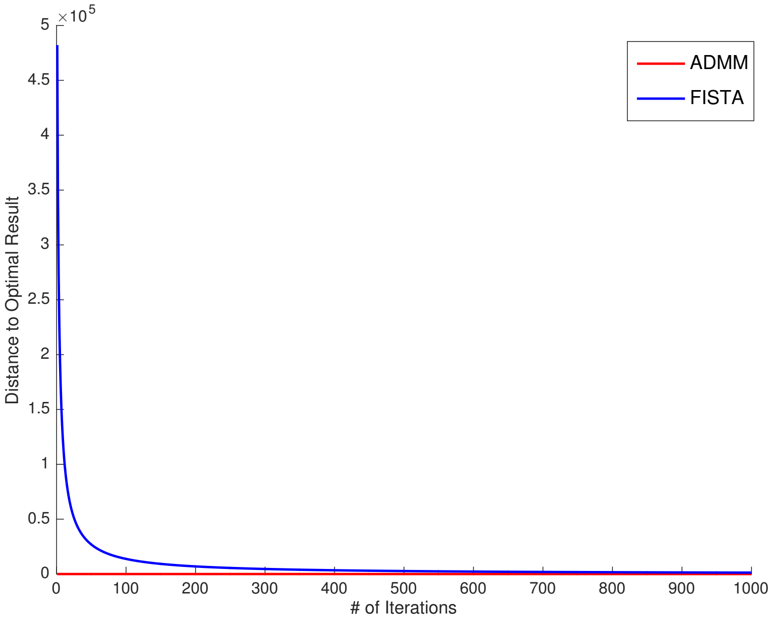


Figure 3: FISTA and ADMM convergence speed comparison.

4 Conclusion

We have introduced several algorithms based on proximal operators and demonstrated its application on recontruction MRI data. Proximal algorithms can be seen as generalization of methods using gradient (or subgradient) descent or dual formulation constraint optimization and are conveniently used when it is feasible to evaluate proximal operator of used regularization term in original problem. Total variation has both simple proximal evaluation as well as strong use in image processing. Variational denoising model was extended to reconstruct subsampled data and was solved numerically using several presented algorithms. Further extension to this model is to incorporate *golden angle* subsampling, assume dynamical data and make use of prior knowledge given for used input data.

References

[1] L. Rudin, S. J. Osher, E. Fatemi. *Nonlinear total variation based noise removal algorithms.* Physica D., 60. (1992), 259–268.

[2] N. Parikh, S. Boyd *Proximal Algorithms.* Foundations and Trends in Optimization, Vol. 1, No. 3 (2013), 123–231.

[3] A. Chambolle, T. Pock. *A first-order primal-dual algorithm for convex problems with applications to imaging.* T. J Math Imaging Vis, 40:120, (2011).

[4] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers and O. Doessel. *An Optimal Radial Profile Order Based on the Golden Ratio for Time-Resolved MRI.* IEEE Transactions on medical imaging, Vol. 26, No. 1. (2007).

[5] S. Sykora. *K-space formulation of MRI.* Stan's Library, Vol.I. (2005).

[6] M. V. Afonso, J.M. Bioucas-Dias. *Fast Image Recovery Using Variable Splitting and Constrained Optimization.* IEEE Transactions on image processing, Vol. 19., No. 9. (2010).