Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 2025, 19(3), 350–361
https://doi.org/10.12913/22998624/195644
ISSN 2299-8624, License CC-BY 4.0

Received: 2024.10.18
Accepted: 2025.01.15
Published: 2025.02.01

Fast approximation of matrix exponential and its application to independent component analysis problem

Dariusz Mika[1*], Jerzy Józwik[2], Fabio Leccese[3], Alessandro Ruggiero[4]

[1] Institute of Technical Sciences and Aviation, The University College of Applied Sciences in Chelm, 22-100 Chelm, Poland
[2] Department of Production Engineering, Faculty of Mechanical Engineering, Lublin University of Technology, Poland
[3] Dipartimento di Scienze, Universitàdegli Studi "Roma Tre", Roma, Italy
[4] Department of Industrial Engineering, University of Salerno, Fisciano, Italy
* Corresponding author's e-mail: dmika@panschelm.edu.pl

ABSTRACT

This paper introduces a new low-computational method for approximating the skew-symmetric and skew-Hermitian matrix exponential. Our method belongs to the splitting methods, which we modify and combine with new low-cost analytic formula for sparse skew-symmetric and skew-Hermitian matrix exponential, similar to the Euler-Rodrigues formula, well known for the skew-symmetric matrices in $\mathbb{R}^3$. Our new approximation procedure for skew-symmetric (skew-Hertmitian) matrix exponential that we use is computationally very cheap, which ensures high speed of algorithms using this operation in their structure. To evaluate this approximation method we used it for the optimization problem of Independent Component Analysis (ICA) type. The results are compared to other known ICA algorithms such as well known Infomax and JADE. The average increase in convergence speed in the studied range of the number of source images was approximately 7% compared to the second fastest ICA algorithm using the standard and universal matrix exponential formula. High quality of separation was also obtained, comparable to well-known ICA algorithms such as Infomax or JADE. Obtained results confirm the effectiveness of the proposed method in technical applications and indicate potential use in on-line applications.

Keywords: matrix exponential, skew-Hermitian matrix, splitting methods, independent component analysis.

INTRODUCTION

Consider the ordinary differential equation (ODE) with constant coefficient and the standard Cauchy problem

$$\dot{x} = Ax, x(0) = x_0 \in \mathfrak{u}(n) \tag{1}$$

where: $A \in \mathfrak{u}(n)$ is a given, fixed, complex skew-Hermitian matrix. A solution vector $x(t)$ is given by $x(t) = e^{tA}x_0$, where e^{tA} is an element of unitary group $\mathrm{U}(n)$ and can be expressed by the convergent power series.

$$e^{tA} = I + tA + \frac{1}{2}t^2A^2 + \frac{1}{6}t^3A^3 + \cdots \in \mathrm{U}(n) \tag{2}$$

In a real case, $y_0, A \in \mathfrak{so}(n)$ are skew-symmetric matrices and e^{tA} is a special orthogonal matrix. This type of ordinary differential equation and the related exponentiation of skew-Hermitian matrix arise in many areas of sciences, engineering or control theory. Problems related with dynamical systems and rigid body dynamics as, for example, robotics can be modeled by this type of differential equations.

In [1] authors develop the basic theory and applications of mechanics with an emphasis on the role of symmetry in a context of dynamical systems and the use of geometric methods and new applications to integrable and chaotic systems, control systems, stability and bifurcation. In [2] the general concept of Be´zier curves to curved spaces is presented, and this generalization is illustrated with an application in kinematics problem of trajectory generation or motion interpolation for a moving rigid body. In [3] authors present an algorithm for generating a twice-differentiable curve on the rotation group SO(3) that interpolated a given ordered set of rotation matrices at their specified knot times. The paper [4] addresses the problem of generating smooth trajectories between an initial and a final position and orientation in space. The authors use the notions of Riemannian metric and covariant derivative from differential geometry and formulate the problem as a variational problem on the Lie group of spatial rigid body displacements.

The motion, in which one point of the rigid body remains constant, is known as the spherical motion. This type of motion is described by the equation (1), where matrix A, in this case is the skew-symmetric part of the velocity gradient tensor

$$= \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \in \mathfrak{so} \tag{3}$$

Assuming $\vec{\omega} = (\omega_1, \omega_2, \omega_3) \in \mathbb{R}^3$ is a vector of the angular velocity of the rigid body then $x(t) = e^{tA}x_0 = Rx_0$ means the rotation of the point x_0 of the rigid body by the angle $t\omega$ around the axis defined by $\vec{\omega}$ and the matrix $R = e^{tA}$ is the rotation matrix.

Important practical field of application of skew-Hermitian matrix exponential is optimization with unitarity constraints, which occurs e.g. in Independent Component Analysis ICA. ICA has found many practical applications in such areas as signal processing, machine learning, sound and image processing, pattern recognition [5], biomedical signal processing [6; 7], financial and astronomical data analysis [8; 9; 10; 11]. The authors using standard ICA algorithms obtain interesting results of source separation in this type of data. On the other hand complex ICA ranges from biomedical signal processing such as extracting spatial maps and time courses from functional magnetic resonance imaging (fMRI) data to communication areas of applications in Multi-Input Multi-Output wireless communication systems [12]. A wide range of applications of complex ICA finds in time-frequency analysis. In this context, there is currently great interest in the use of blind source separation methods including ICA in the diagnostic of technical systems where the on-line detection of fault signals plays a crucial role [13; 14; 15]. As demonstrated in [16; 17] it is possible to separate sources using ICA on magnitude of the Fourier representation of single mixed signal, i.e. on spectrogram only.

In the Lie group methods the search direction is typically chosen as a negative skew-Hermitian gradient of the cost function. This gradient matrix belongs to the Lie algebra space of the unitary $\mathrm{U}(n)$ group. In each iteration step, the skew-Hermitian matrix is mapped by the exponentiation operation into the Lie group $\mathrm{U}(n)$ and thereby to ensure the continuous presence on the optimization surface i.e., on the Lie group $\mathrm{SO}(n)$. Lie group optimization techniques are very efficient, stable and universal [18; 19; 20]. Every Lie group methods require a number of matrix exponentials i.e., mapping a Lie algebra $\mathfrak{g}$ to a Lie group G. This operation however needs to be approximated to the order of the underlain ODE solving method. This approximation should reside in the Lie group G associated to the Lie algebra $\mathfrak{g}$. This fundamental requirement in general is not fulfilled by many standard approximations methods unless it is calculated exactly. In the well-known work [21], the authors present nineteen methods to compute the exponential of a matrix. In [22] and in [23] authors presents exact methods to compute the matrix exponential.

Exact formula for matrix exponential of any matrix is given by (2) [24]. Some simple approximation methods for a low dimensional skew-symmetric matrix are given in [25; 26; 27]. The well known formula for 3×3 skew-symmetric matrices (4)

$$e^A = I + \frac{\sin(\alpha)}{\alpha} A + \frac{1}{2}\left(\frac{\sin\left(\frac{\alpha}{2}\right)}{\frac{\alpha}{2}}\right)^2 A^2 \tag{4}$$

where: $\alpha = \frac{1}{2}\|A\|_F$ and $\|\cdot\|_F$ is a Frobenius matrix norm are known as the Euler-Rodrigues formula and calculate exponential exactly.

Cayley-Hamilton theorem [28] gives another exact method of approximation of skew-symmetric matrix exponentials. However this is well known that, first, it requires high computational cost of high powers of matrix and second it use direct the characteristic polynomial of the matrix so it may lead to some computational instabilities.

Standard approximants based on Padé and Chebyshev rational approximation generally do not reside on the Lie group G associated with Lie algebra $\mathfrak{g}$. In case of skew-Hermitian or skew-symmetric approximation only diagonal Padé approximation maps Lie algebras $\mathfrak{u}(n)$ and $\mathfrak{so}(n)$ to underlying Lie group $\mathrm{U}(n)$ and $\mathrm{SO}(n)$ [29] but this property is not satisfied in the case of the $\mathfrak{sl}(n)$. It can be shown that the only analytic function f, that maps $\mathfrak{sl}(n)$ into $\mathrm{SL}(n)$, consistently with the exponential function is the exponential itself.

Splitting the matrix A into a sum of low-rank bordered matrices A_i belonging to $\mathfrak{g}$ is another very effective method of approximation of matrix. In [29; 30; 31] authors present some methods for the approximation of the matrix exponential in a Lie-algebraic setting. Splitting methods can be applied in different fields for example in: the partial differential equations context [24], constructing volume-preserving methods [32] and constructing symplectic methods [33]. Some excellent studies of the splitting methods can be found in [34].

The Strang-type splitting of the form (5)

$$e^{tA} \approx F(tA) = e^{\frac{1}{2}tA_1} e^{\frac{1}{2}tA_2} \cdots e^{\frac{1}{2}tA_{k-1}} e^{tA_k} e^{\frac{1}{2}tA_{k-1}} \cdots e^{\frac{1}{2}tA_2} e^{\frac{1}{2}tA_1} \tag{5}$$

where: $A = \sum_{i=1}^{k} A_i$ is well know splitting methods and approximate e^{tA} to second order. As long as $A_1, A_2, \ldots, A_k \in \mathfrak{g}$ it follows at once from the definition of Lie group that the approximation resides in G [35].

Generally, using a splitting methods it is possible to express e^{tA} for $A \in \mathfrak{g}$ and small t as (6)

$$e^{tX} = e^{g_1(t)X_1} e^{g_2(t)X_2} \ldots . e^{g_m(t)X_m} \tag{6}$$

where: $g_i(t)$ is the scalar function analytic at the origin and $X = \{X_1, X_2, \ldots, X_k\}$ be a basis of $\mathfrak{g}$ and $\dim \mathfrak{g} = m$. This splitting is known as a canonical coordinate of the second kind and was pioneered by Owren and Marthinsen [36] in the context of general Lie group methods. A properly selected base X of the Lie algebra $\mathfrak{g}$ leads to robust and affordable algorithms.

In this paper we present some fast approximation of skew-symmetric and skew-Hermitian matrix exponential based on splitting methods and its application in optimization problem. Often, in some technical applications, e.g. in on-line optimization algorithms, the speed of the algorithm is more important than the quality of the approximation. Our motivation is to explore fast methods of matrix exponential in context of the balance between quality and speed of approximation.

BACKGROUND THEORY ON SPLITTING METHODS

The basic idea of splitting methods relies on decomposition of $n \times n$ matrix $A \in \mathfrak{g}$ in the form (7)

$$A = \sum_{i=1}^{k} A_i \tag{7}$$

The e^{tA} can be approximated by (8)

$$F(tA) = \mathrm{e}^{tA_1} \cdot \ldots \cdot \mathrm{e}^{tA_k} \tag{8}$$

Splitting $A \in \mathfrak{g}$ into appropriate low rank terms $A_i \in \mathfrak{g}$ so that the cost of computing every e^{tA_i} is small enough compared to e^{tA} is the main idea of splitting methods. The simplest choice of splitting of $A \in \mathfrak{u}(\mathrm{n})$ is columnwise decomposition. Taking $A_0 = A$ we let (9)

$$A_1 = a_1^{(0)} e_1^T - e_1 \left(a_1^{(0)}\right)^H \in \mathfrak{u}(n) \tag{9}$$

where: $a_1^{(0)}$ is the first nonzero column of A_0 and e_1 is canonical base of $\mathbb{R}^n$. In general taking $A_{i+1} = A_{i-1} - A_i$ for $i = 1, \ldots, n-1$ and (10)

$$A_{i+1} = a_{i+1}^{(i)} e_{i+1}^T - e_{i+1} \left(a_{i+1}^{(i)}\right)^H \in \mathfrak{u}(n) \tag{10}$$

where: $a_{i+1}^{(i)}$ is the first nonzero column of A_i and e_{i+1} is $(i+1)$-th canonical basis vector of $\mathbb{R}^n$. This type of approximation is order one [29].

Using the same kind of splitting, the function (11)

$$F(tA) = e^{\frac{1}{2}tA_1} e^{\frac{1}{2}tA_2} \cdot \ldots \cdot e^{\frac{1}{2}tA_{k-1}} e^{tA_k} e^{\frac{1}{2}tA_{k-1}} \cdot \ldots \cdot e^{\frac{1}{2}tA_2} e^{\frac{1}{2}tA_1} \tag{11}$$

known as the generalized Strang splitting is the second-order approximant of e^{tA} [30]. Moreover as long as $A_1, A_2, \ldots, A_k \in \mathfrak{g}$ the approximant $F(tA)$ resides in G. The general rule is that more than k multiplications of exponentials e^{tB_i} are needed to obtain an approximation of order higher than one. Using the technique known as canonical coordinate of the second kind [37] it can be increase the order of approximation without increasing the number of evaluation and multiplication of the term e^{tA_i}. This method approximates e^{tA} as a composition of exponentials of the form $e^{g_i(t)X_i}$, where $\{X_1, X_2, \ldots, X_k\}$ is a basis of $\mathfrak{g}$ and $g_i(t)$ for $i = 1, \ldots, n-1$ are the scalar functions analytic in the origin. Clever choice of the function $g_i(t)$ and bases $\{X_i\}$ which exploit the Lie-algebraic structure can reduce the computational cost by several orders and leads to robust and affordable algorithms. This method makes very good use of matrix sparsity. For sparse tridiagonal (12)

$$A = \begin{bmatrix} 0 & \beta_1\ 0 & \ldots & 0 \\ -\beta_1 & 0\ \beta_2 & & \vdots \\ 0 & -\beta_2 \ddots & & 0 \\ \vdots & & 0 & \beta_{n-1} \\ 0 & \cdots 0 & -\beta_{n-1} & 0 \end{bmatrix} = \sum_{k=1}^{n-1} \beta_k F_{k,k+1} \in \mathfrak{so}(n) \tag{12}$$

and $n-1$ dimensional basis $X_k = F_{k,k+1} = e_k e_{k+1}^T - e_{k+1} e_k^T$ for $k = 1, \ldots, n-1$, the first-order approximant $F(tA)$ needs only $n-1$ multiplications of exponentials $e^{t\beta_k F_{k,k+1}}$ with very simple form (13)

$$A = \begin{bmatrix} 1 & \ldots & & & \ldots & 0 \\ \vdots & \ddots & & & & \vdots \\ & & \cos(t\beta_k) & \sin(t\beta_k) & & \\ & & -\sin(t\beta_k) & \cos(t\beta_k) & & \\ \vdots & & & & \ddots & \vdots \\ 0 & \ldots & & & \ldots & 1 \end{bmatrix} \tag{13}$$

and with very small computational cost. The first-order approximation takes the form (14)

$$F(tA) = e^{t\beta_1 F_{1,2}} \cdot \ldots \cdot e^{t\beta_{n-1} F_{n-1,n}} \tag{14}$$

Second-order approximation $F(tA)$ uses matrices of the form $F_{k,k+2}$ and takes the form (15) [29; 30]

$$F(tA) = e^{t\beta_{n-1}F_{n-1,n}} \cdot e^{t\beta_{n-2}F_{n-2,n-1}} \cdot e^{\frac{1}{2}t^2\beta_{n-2}\beta_{n-1}F_{n-2,n}} \cdot \ldots \cdot e^{t\beta_2 F_{2,3}} \cdot e^{\frac{1}{2}t^2\beta_2\beta_3 F_{2,4}} \cdot\cdot\ e^{t\beta_1 F_{1,2}} \cdot e^{\frac{1}{2}t^2\beta_1\beta_2 F_{1,3}} \tag{15}$$

The Strang-type splitting in this case needs $2(n-1)$ multiplications of the form (16)

$$F(tA) = e^{\frac{1}{2}t\beta_1 F_{1,2}} \cdot \ldots \cdot e^{\frac{1}{2}t\beta_{n-2}F_{n-2,n-1}} \cdot e^{t\beta_{n-1}F_{n-1,n}} \cdot e^{\frac{1}{2}t\beta_{n-2}F_{n-2,n-1}} \cdot \ldots \cdot e^{\frac{1}{2}t\beta_1 F_{1,2}} \tag{16}$$

GEODESIC-FLOW OPTIMIZATION METHODS IN INDEPENDENT COMPONENT ANALYSIS

The basic concepts of Independent Component Analysis relies on estimating a sequence of n latent variables $s = (s_1, \ldots, s_n)^T$ (source signals) from their linear mixtures $x = (x_1, \ldots, x_n)^T$. The main assumption of this method is the statistical independence of s_i, i.e., Independent Components (ICs). The observed mixed signal x can be modeled (17)

$$x = As \quad (17)$$

where: $A \in \mathrm{Gl}(n)$ is the $n \times n$ real or complex unknown invertible mixing matrix.

A solution for the ICA problem consists in finding the demixing (or filtration) matrix $W \cong A^{-1}$. Independent components s_i are obtained from (18)

$$\mathrm{s} \cong \hat{s} = Wx \quad (18)$$

where: $\hat{s}$ is the estimator of a source signal s. Given N observations of x, the goal of ICA is to estimate the demixing matrix W and thereby recover the source signal s.

The basic ICA model assumes that the number m of ICs s_i is known and equal to the number n of observed mixed signal x_{i}. Although, in general ICA model, this assumption do not have to be met and an ICA model where $m \neq n$ is also considered. In the extreme case, when only one observed signal is given, the ICA problem is called as single channel source separation. This challenging problem is also possible to solve [16; 17; 38; 39].

In all very approach to solve ICA problem there are two main concepts: optimization of cost function and algebraic modification of demixing matrix W. In the first class a nonlinear cost function is optimized. This function represents a measure of statistical independence of estimated ICs. In Maximum Likelihood approach the negative log-likelihood cost function is used and takes the form (19)

$$f(W) = -\log p(x) = -\log|\det(W)| - \sum_{i=1}^{n} \log p_i(w_i x) \quad (19)$$

where: w_i is the i-th row of demixing matrix W and $p_i(\cdot)$ is the pdf function of i-th source signal s_i.

Another very important method in this class of ICA is the Negentropy Maximization approach. This method uses some non-Gaussianity measure as a cost function [40]. In basic concept this function is the negative entropy, commonly called as negentropy. This function is a natural measure of entropic distance of pdf of estimated source from variable with standard Gaussian distribution. Negentropy can be defined as (20)

$$J(W) \triangleq H(\nu) - H(s) \quad (20)$$

where: $H(\cdot) \triangleq -E\{log\, p(\cdot)\}$ is the differential entropy of the given probability distribution and ν is the Gaussian variable with the same variance as source s. Since $H(\nu) = \mathrm{const}$, then maximizing negentropy $J(W)$ leads to minimization of entropy $H(s)$. The main drawback in this concept relies on difficulty connected with negentropy calculation. Definition requires knowledge of the pdf function. This information is a priori unknown and in practice certain approximations of negentropy are considered, such as the Taylor, Gram-Charlier and Edgeworth expansions. However in certain cases these expansions can leads to a rather poor approximation of negentropy with high sensitivity to outliers. To overcome this problem a non-polynomial expansions are used, which leads to an approximation of the form (21)

$$J(W) \approx \frac{1}{2} \sum_{i=1}^{n} E\{G^i(Wx)\}^2 \quad (21)$$

where: $G^i, i = 1, \dots, n$ is any set of orthonormal nonquadratic function.

In general the demixing matrix W belongs to general linear group $\mathrm{Gl}(n)$, i.e., matrix W satisfies only invertibility property $\det W \neq 0$. The whitening (or sphering) of observed mixed signal reduces the problem complexity. Assuming, without losing the generality, that the source signal is zero-mean and with unit-variance, i.e., $\mathrm{E}\{ss^H\} = I$, the whitening process decorrelate signal using eigendecomposition of the correlation matrix, i.e., $C_x = \mathrm{E}\{xx^H\} = U\Lambda U^H$, where U is the unitary matrix with eigenvectors of C_x as a column of U and Λ is diagonal matrix with eigenvalues on main diagonal. Transformed signal $y = \Lambda^{-1/2}U^H\mathrm{x}$ has following property: $\mathrm{E}\{yy^H\} = \mathrm{E}\{\Lambda^{-1/2}U^H xx^H U\Lambda^{-1/2}\} = \Lambda^{-1/2}U^H\mathrm{E}\{xx^H\}U\Lambda^{-1/2} = I$, i.e., the whitened signal y is decorrelated and has unit variance. It means that new mixing matrix $\tilde{A}$, defined as $y = Vx = VAs = \tilde{A}s$, has unitarity property, i.e., $\mathrm{E}\{yy^H\} = \tilde{A}\mathrm{E}\{ss^H\}\tilde{A}^{\mathrm{H}} = \tilde{A}\tilde{A}^H = I$, so a new demixing matrix $\widetilde{W} = \tilde{A}^{-1}$ has also the unitarity property $\widetilde{W}\widetilde{W}^H = I$. In optimization context it also means that whitening process simplifies the ICA problem from operation on general linear group $\mathrm{Gl}(n)$ to operation on unitary group $\mathrm{U}(n)$.

Let $\gamma_{X,I}(t)\colon t \to \mathrm{U}(n)$ be a carve in $\mathrm{U}(n)$ generated by left invariant tangent vector X and emanating from I. This carve define the one-parameter subgroup in $\mathrm{U}(n)$, which satisfies the condition

$\gamma_{X,I}(t)^H \gamma_{X,I}(t) = I$. Differentiating this equation the tangent space $T_W \mathrm{U}(n)$ at the point $W(t) = \gamma_{X,I}(t)$ can be defined as (22)

$$T_W \mathrm{U}(n) = \{X \in \mathbb{C}^{n\times n} | X^H W + W^H X = 0\} \tag{22}$$

The Lie algebra $\mathfrak{u}(n)$ is obtained by substituting $W = I$ in (22) and is given by (23)

$$\mathfrak{u}(n) = T_I \mathrm{U}(n) = \{X \in \mathbb{C}^{n\times n} | X^H + X = 0\} \tag{23}$$

This algebra is a set of skew-Hermitian matrices. The contravariant gradient $\widehat{\nabla}_{\mathrm{W}^*} f$ of the cost function f on Riemannian manifold $\mathrm{U}(n)$ at the point W can be expressed as (24)

$$\widehat{\nabla}_{\mathrm{W}^*} f = \nabla_{\mathrm{W}^*} f - W \nabla^{\mathrm{H}}_{\mathrm{W}^*} f W \in T_W \mathrm{U}(n) \tag{24}$$

where: $\nabla_{\mathrm{W}^*} f$ is a gradient vector calculated in Euclidean space $\mathbb{C}^{n\times n}$. Using an exponential map [41] a carve $\gamma_{X,I}(t)$ can be expressed as (25)

$$\gamma_{X,I}(t) = \exp(tX) \in \mathrm{U}(n) \tag{25}$$

This carve defines the geodesic line, i.e., the carve minimizing length between two points on Riemannian manifold. Using left translation the geodesic emanating from W takes the form (26)

$$\gamma_{X,W}(t) = W\gamma_{X,I}(t) = W\exp\,(tX) \tag{26}$$

The optimization procedure that uses this expression of geodesic line is known as geodesic flow method. The search direction Ω in these methods are chosen as negative gradient $-\widehat{\nabla}_{\mathrm{W}^*} f$ at point W (defined in (24)) translated to the identity, i.e., into Lie algebra $\mathfrak{u}(n)$ by expression (27)

$$\Omega = W^{-1} \widehat{\nabla}_{W^*} f = W^H \widehat{\nabla}_{W^*} f = W^H \nabla_{W^*} f - \nabla^{\mathrm{H}}_{W^*} f W \in \mathfrak{u}(n) \tag{27}$$

Therefore the optimization procedure is conducted by geodesic line defined as (28)

$$\gamma(t) = W\exp\,(-t\Omega) \tag{28}$$

and the optimization scheme has the form (29)

$$W_{k+1} = W_k R_{\Omega_k} = W_k \exp\,(-t_{opt}\Omega_k) \tag{29}$$

where: k is the number of iteration and t_{opt} is the value of t that minimizes f on geodesic $\gamma(t)$ in every iteration step.

FORMULA FOR APPROXIMATION OF SPARSE SKEW-SYMMETRIC AND SKEW-HERMITIAN MATRIX EXPONENTIAL

In some technical areas there is no need to use the full structure of skew-Hermitian matrices. An optimization problem of the search for the extreme of the function $f(W), W \in \mathbb{C}^{n\times n}$ with the constraint $WW^H = I_n$ e.g. in the ICA problem, can be solved using a skew-symmetric matrix that approximates the gradient of the function f, built only from one row and one column [42]. The simple structure of this type of matrix makes it possible to create an analytical exponentiation formula with a low computational cost. Appropriate construction of such matrix ensures a solution to the optimization problem with a faster convergence rate compared to conventional algorithms using the full form of the skew-Hermitian gradient matrix of function f.

In [43] an analytical Rodrigues-like formula for approximation of skew-symmetric matrices is given. This formula requires however, the Schur decomposition, which introduces an additional computational cost. Therefore, in some practical applications with speed priority, the exponential procedure using skew-Hermitian matrices directly, i.e., without any preprocessing would be exceptionally useful. In [42] the author presents a simple formula for exponentiation of skew-symmetric matrices with one row and one column. This formula, using the 2-norm $\|\cdot\|_2$ comes down to the well-known Rodrigues-like formula.

Theorem 1. Consider a sparse skew-symmetric matrix C of the form (30)

$$C = \begin{bmatrix} 0 & \cdots\times & \dots & 0 \\ \vdots & \ddots\times & & \vdots \\ \times & \times\, 0 & \times & \times \\ \vdots & \times & \ddots & \vdots \\ 0 & \cdots\times & \cdots & 0 \end{bmatrix} i \in \mathfrak{so}(n) \tag{30}$$

with only $n-1$ free variables. The exact exponential of matrix C takes the Euler-Rodrigues-like form (31)

$$\exp C = I_n + \frac{\sin\alpha}{\alpha} C + \frac{1}{2}\left(\frac{\sin\frac{\alpha}{2}}{\frac{\alpha}{2}}\right)^2 C^2 \tag{31}$$

Proof. See [44]
In a complex the analogous formula comes from:

Theorem 2. Consider a sparse skew-Hermitian matrix C of the form (32)

$$C = \begin{bmatrix} 0 & \cdots c_1 & \dots & 0 \\ \vdots & \ddots\ \vdots & & \vdots \\ -c_1^* & i|c_m| & & -c_n^* \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots c_n & \cdots & 0 \end{bmatrix} \in \mathfrak{u}(n) \tag{32}$$

The 4th-order approximation of exponential of matrix C takes the form (33)

$$\exp C = I_n + \frac{\sin\alpha}{\alpha}(C + c_m E_m) + \frac{1}{2}\left(\frac{\sin\frac{\alpha}{2}}{\frac{\alpha}{2}}\right)^2 (C^2 + c_m^2 E_m) + R_4 \tag{33}$$

where the residual matrix R_4 takes the form (34)

$$R_4 = -\frac{c_m}{6} PP^H - \left(\frac{c_m^2}{6} + \frac{c_m\alpha^2 + c_m^3}{24}\right)(P - P^H) - \left(c_m + \frac{c_m^2}{2} - \frac{2c_m^2\alpha^2 + c_m^4}{24}\right) E_m \tag{34}$$

where: P is the matrix containing only i-th column of C such that $C = P - P^H - c_m E_m$, and E_m is a zero matrix with only one (m,m)-th element equal to 1.
Proof.See [44]

As you can see, this formula contains an Euler-Rodrigues-like term similar to the real case but with a modified (m,m)-th element of C and C^2. The residual matrix R_4 modifies the main Euler-Rodrigues component in different way. The first term with PP^H modifies each element of the result matrix, the second term with $P - P^H$ only m-th row and m-th column, and the third expression with E_m modifies only (m,m)-th element of the result matrix.

IMPLEMENTATION OF THE ALGORITHM AND COMPLEXITY

The proposed above formulas can be used for approximation of exponential of any skew-symmetric and skew-Hermitian matrix. The proposed strategy consists in decomposition of the matrix A in incomplete form (35)

$$\tilde{A} = \sum_{i\in\{i_p\}} C_i = A - R \tag{35}$$

where: the matrices C_i takes the form (30) or (32), index $i \in \{i_p\} \in (1, \dots, n-1)$ denotes the number of column of matrix A, taking part in the splitting procedure (35) and $R = A - \tilde{A}$ is the residual matrix. The desired approximation accuracy is achieved by appropriate selection of the number of C_i matrices involved in splitting (35). This selection involves determining the size of the index set $\{i_p\}$ as well as selecting its elements. In this second aspect, a selection strategy was adopted based on the decreasing value of the norm of columns v_i of matrix A. For example, for $p = 3$ and selecting the set of indices $\{i_p\} = \{k, l, r\}$ for which $\|v_k\| \geq \|v_l\| \geq \|v_r\|$ and $\|v_k\| = \max_j(\|v_j\|), j = 1, \dots, n-1$, the proposed splitting procedure takes the form (36)

$$\tilde{A} = \sum_{i\in\{k,l,r\}} C_i = C_k + C_l + C_r = (P_k - P_k^T) + (P_l - P_l^T) + (P_r - P_r^T) \begin{bmatrix} 0 & \times & \times & & \times\,0 \\ \times & \times & \times\times & \times & \times\times \\ & \times & \times & & \times \\ \times & \times & \times\times & \times & \times\times \\ & \times & \times & & \times \\ \times & \times & \times\times & \times & \times\times \\ 0 & \times & \times & \times & 0 \end{bmatrix} \begin{matrix} k \\ \\ l \\ \\ r \end{matrix} \quad (36)$$

$$k\,l\,r$$

where: $P_k = \begin{bmatrix} 0 & \cdots\times\cdots & 0 \\ \vdots & \times & \vdots \\ \vdots & \times & \vdots \\ 0 & \cdots\times\cdots & 0 \end{bmatrix}$ (k); $P_l = \begin{bmatrix} 0 & \cdots\times\cdots & 0 \\ \vdots & \times & \vdots \\ \vdots & \times & \vdots \\ 0 & \cdots\times\cdots & 0 \end{bmatrix}$ (l); $P_r = \begin{bmatrix} 0 & \cdots\times\cdots & 0 \\ \vdots & \times & \vdots \\ \vdots & \times & \vdots \\ 0 & \cdots\times\cdots & 0 \end{bmatrix}$ (r)

and $P_l(l,k) = 0,\ P_r(l,r) = 0,\ P_r(k,r) = 0$

We have used the notation used in [45]. The symbol '×' denotes a matrix element different from zero. The remaining elements are all equal to zero. Below we describe the proposed procedure in more detail.

Algorithm

% Purpose – Approximation of the exponential with splitting skew-symmetric or
% skew-Hermitian matrix A in incomplete form (35) with column selection.
% In – $A: n \times n$ skew-symmetric or skew-Hermitian matrix
% Out – Approximation of $\exp tA \approx F(tA) = \exp C_{i_1} \exp C_{i_2} \ldots \exp C_{i_m}$ and $\exp C_{i_p}$ is
% computed by formula (31) or (33)
for$i = 1{:}n$
    $\alpha(i) = \|v_i\|^2 = A(:,i)^T A(:,i)$
end
$[\sim, index] = sort(\alpha);$
for$i = 1{:}m$
$P_i = zeros(n);$
$k = index(i);$
$P_i\,(:,i) = tA(:,i);$
for $h = 1{:}(k-1)$
$P_i(index(i-h),k) = 0;$
end
$C_i = P_i - P_i^T;$
$exp\ C_i$; (equation (31) or (33))
$exp(tA) \approx F(tA) = exp\,C_1\,exp\,C_2 \ldots exp\,C_i\,;$
end

The main computational cost of the proposed algorithm consist of m-times evaluation of $\exp C_i$ from equation (31) or (33) and $(m-1)$-times multiplying the results $\exp C_1 \exp C_2 \ldots \exp C_i$. The cost of computing $\exp C_i$ from formula (31) or (33) is $2n^2 + 4n + 2 + 2c_{\sin}$ flops in the real case and $4(3n^2 + 6n + 14 + 2c_{\sin})$ flops in the complex case, where $c_{\sin}$ is the computational cost of the built-in Matlab sin function. In the complex case, the estimation assumes that 1 complex multiplication is equal to 4 real floating point operations [46]. The multiplication stage costs $(m-1)n^3$ flops and $4(m-1)n^3$ flops in real and complex case, respectively. In summary, the total cost of the proposed algorithm can be estimated as $(m-1)n^3 + (2m+1)n^2 + 6mn + +2(c_{\sin}+1)m$ flops in the real case and $4[(m-1)n^3 + n^2 + m(3n^2 + 7n + 14 + 2c_{\sin})]$ flops in the complex case.

It is clear from this analysis that computational cost depends significantly on the number of m columns used. When only one column $m = 1$ is used, the computational cost of the proposed method is only of the order $O(n^2)$. For $2 \leq m \leq n-1$ the prevailing term in the computational cost is $(m-1)n^3$ flops and with full splitting the cost is $(n-1)n^3 + O(n^2)$.

NUMERICAL EXPERIMENTS AND APPLICATIONS

In the experiment, the proposed algorithm (denoted as ICCS) of approximation of skew-symmetric matrix exponential was used in ICA problem with geodesic flow method of optimization. In all simulations the natural monochrome images were used as a source signal, each having pixels. Six images (pixel matrix) was transformed into a row vector, normalized and combined into a 6-row matrix. A random non-orthogonal mixing matrix was used to create a linear combination of the pictures vector, i.e. a multichannel input signal. Figures 1 (a), (b) and (c) show the source signals (images), the mixed signals and recovered images, respectively.

The negentropy function, defined in (21), was chosen as the classical measure of statistical signal independence and was used in all the algorithms studied [40]. The step size of each iteration was chosen experimentally. Geodesic-flow algorithm with proposed approximation of matrix exponential was compared with the same geodesic-flow algorithm (denoted as Exp) but with built-in Matlab function expm and other approximation methods of matrix exponential. We used General Polar Decomposition (denoted as GPD) [47], 1-order approximation of the form (14) ad 2-order approximation of Strang type defined in (16) both with Lanczos tridiagonalization in preprocessing stage. The built-in function expm uses the diagonal *(1,1)*-Padé approximation and calculates the exponential to nearly machine accuracy. We also used in simulation a classical ICA method as Infomax (with Euclidean Steepest Descent optimization scheme) and JADE. The proposed algorithm was carried out for the optimal number of column m^* obtained in [44]. As a stopping criterion we used a condition of cost function convergence in the form: if $J_k - J_{k-1} < \varepsilon = 10^{-4}$ break. The simulation was carried out in Matlab 7.9 on a PC (Intel i7 2.8 GHz CPU, 8GB RAM). The mixing matrix was chosen randomly using built-in Matlab function rand. Simulation was carried out with a different number of sources (images) from $n = 2$ to $n = 8$. The results were averaged over 500 trials. The summary of the averaged convergence times obtained in the experiment are shown in the Figure 2.

Collected results of average convergence times are the shortest for each for the proposed ICCS algorithm (red line in Figure 2). The largest speed increase was observed for larger numbers of source images, i.e. for $n \geq 6$. The average increase in convergence speed in the studied range of the number of source images was approximately 7% compared to the second fastest algorithm Exp.

The performance of separation was measured by the Amari Performance Index (API) defined as [48] (37)

$$\mathrm{API} = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{|p_{ij}|}{\max_k|p_{ik}|} - 1\right) + \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{|p_{ij}|}{\max_k|p_{kj}|} - 1\right) \quad (37)$$

Figure 1. Separation results using four natural images of 128×128 pixels in size (a) original source images, (b) mixed images, (c) recovered sources using the proposed agorithm

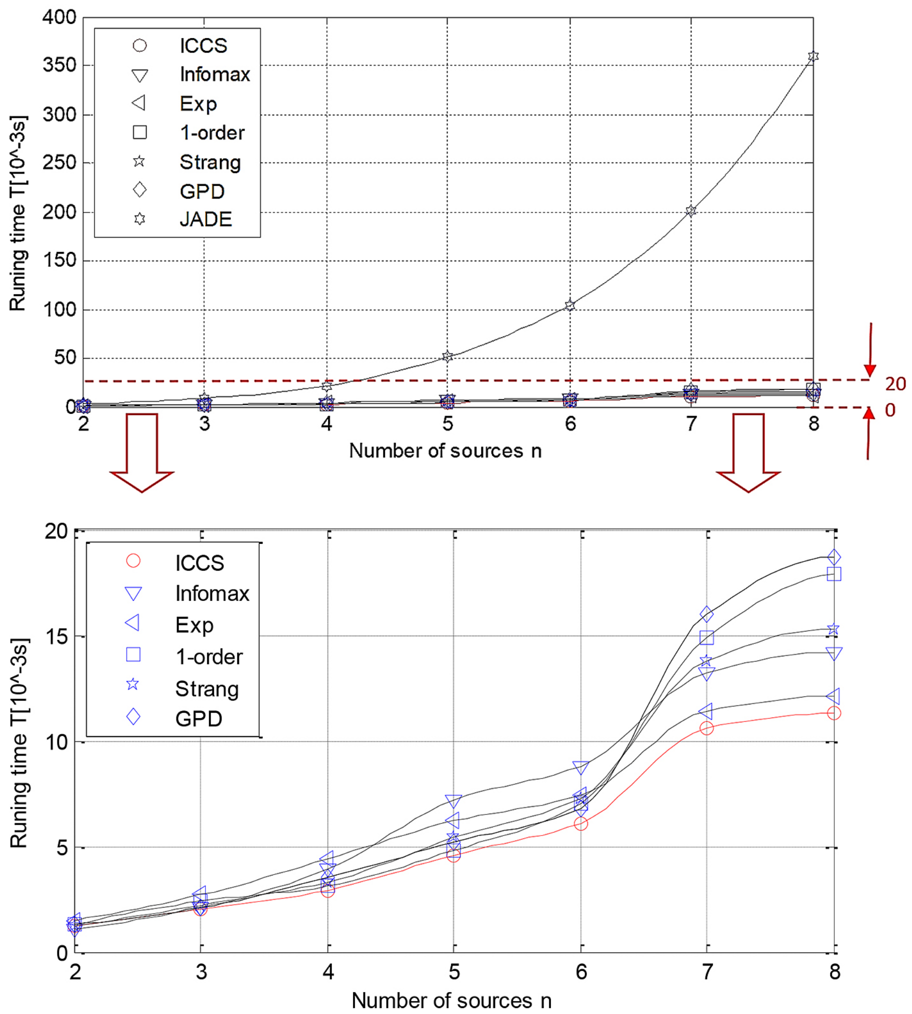

Figure 2. Average convergence time (runing time of the algorithms) of ICA algorithms used in experiment with respect to number of sources

Table 1. Averaged Amari Performace Index for algorithms used in simulation

ICA algorithm	ICCS	Infomax	Exp	1-order	Strang	GPD	JADE
API [dB]	-24.8	-25.0	-24.0	-23.7	-24.2	-23.6	-24.7

where: p_{ij} is (i, j)-th element of the global matrix $P = WA^{-1}$. The average values of API parameters for individual ICA algorithms used in the experiment are presented in Table 1.

It should be emphasized that high separation quality was obtained using only one ($m = 1$) column (with the maximum norm). This means that, from the optimization point of view, a single column (with the maximum norm) of the skew-symmetric matrix can well define the search direction. Using of only a single column results in an increase in the number of iterations by only about 3–5 iterations. A small increase in the number of iterations of the optimization process with a reduced computational cost of exponentiation approximation explains the high speed of the proposed algorithm.

CONCLUSIONS

In this paper, the new approximation algorithm for skew-symmetric and skew-Hermitian matrix exponential has been introduced and tested. The algorithm uses a new low-cost analytical formula for cross-type single-column skew-symmetric and skew-Hermitian matrix exponential. In our new approach we use the splitting method in incomplete form for reducing computational complexity of exponential stage. The proposed algorithm was used and tested in the Independent Component Analysis problem. In the simulation test, a geodetic flow type cost function optimization method was used, where the proposed algorithm and classic matrix exponential methods were used in the exponential stage. With the optimal selection of the number of columns the average increase in convergence speed in the studied range of the number of source images was approximately 7% compared to the second fastest algorithm with a high separation quality comparable to well-known ICA algorithms such as Infomax or JADE. It should also be added that high separation quality was also obtained using only a single column of the matrix (with the largest norm) in approximation with only a slight increase in the number of iterations. The obtained results of average convergence times and their comparison with classical methods confirm the high effectiveness of the proposed method and its potential application in online applications.

Acknowledgements

This work was prepared within the project PM-II/SP/0093/2024/02 entitled "Building the algorithms of blind signal processing based on geometrical structure of information", financed by the Ministry of Education and Science (Poland) as a part of the Polish Metrology II Programme and This work was prepared within the project PM-II/SP/0040/2024/02 entitled "Multi-sensor system for measuring thermo-mechanical interactions with comprehensive condition analysis", financed by the Ministry of Education and Science (Poland) as a part of the Polish Metrology II Programme.

REFERENCES

1. Marsden, J.E. and Ratiu, T.S. Introduction to mechanics and symmetry: a basic exposition of classical mechanical systems. s.l.: Springer Science & Business Media, 2013.
2. Park, F.C. and Ravani, B. Bézier curves on Riemannian manifolds and Lie groups with kinematics applications. In: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers. 1994. 15–21.
3. Park, F.C. and Ravani, B. Smooth invariant interpolation of rotations. ACM Transactions on Graphics (TOG). 1997, 16(3), 277–295.
4. Zefran, M., Kumar, V. and Croke, C.B. On the generation of smooth three-dimensional rigid body motions. IEEE Transactions on Robotics and Automation. 1998, 14(4), 576–589.
5. Lu, C.-J., Shao, Y.E. and Li, P.-H. Mixture control chart patterns recognition using independent component analysis and support vector machine. Neurocomputing. 2011, 74(11), 1908–1914.
6. Li, H., Adali, T., Correa, N., Rodriguez, P.A., Calhoun, V.D. Flexible complex ICA of fMRI data. s.l.: In: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2010, 2050–2053.
7. Yu, M.-C., Lin, Q.-H., Kuang, L.-D., et al. ICA of full complex-valued fMRI data using phase information of spatial maps. Journal of neuroscience methods. 2015, 249, 75–91.
8. Chen, K.-H., and Khaldoun, K. The Reconstruction of Financial Signals Using Fast ICA for Systemic Risk. In: 2015 IEEE Symposium Series on Computational Intelligence. IEEE. 2015, 885–889.
9. Akutsu, T., Ando, M., Arai, K., et al. Application of independent component analysis to the iKAGRA data. Progress of Theoretical and Experimental Physics. 2020, 5: 053F01.
10. Bai, L., Xu, A., Guo, P., et al. Kernel ICA feature extraction for spectral recognition of celestial objects. In: 2006 IEEE International Conference on Systems, Man and Cybernetics.. s.l.: In: 2006 IEEE International Conference on Systems, Man and Cybernetics. 2006, 3922–3926.
11. Chattopadhyay, T., Fraix-Burnet, D., Mondal, S. Unsupervised Classification of Galaxies. I. ICA feature selection. arXiv preprint arXiv:1802.02856. Saptarshi. 2018.
12. Gesbert, D., Bolcskei, H., Gore, D., Paulraj, A. MIMO wireless channels: Capacity and performance prediction. s.l.: Globecom'00-IEEE. Global Telecommunications Conference. Conference Record (Cat. No. 00CH37137). IEEE, 2000, 1083–1088.
13. Chen, Z., Zhao, F., Zhou, J., et al. Fault diagnosis of loader gearbox based on an Ica and SVM algorithm. International Journal of Environmental Research and Public Health. 2019, 16(23), 4868.
14. Hao, Y., Song, L., Wang, M., et al. Underdetermined source separation of bearing faults based on optimized intrinsic characteristic-scale decomposition and local non-negative matrix factorization. IEEE Access. 2019, 7, 11427–11435.
15. Ypma, A., Leshem, A. and Duin, R.P. Blind separation of rotating machine sources: bilinear forms and convolutive mixtures. Neurocomputing. 2002, 49(1–4), 349–368.
16. Mika, D., Budzik, G. and Józwik, J. Single Channel Source Separation with ICA-Based Time-Frequency

Decomposition. Sensors. 2020, 20(7), 2019.
17. Mika, D., Budzik, G. and Józwik, J. 1ICA-based single channel source separation with time-frequency decomposition. s.l.: In: 2020 IEEE 7th International Workshop on Metrology for AeroSpace (MetroAeroSpace). IEEE, 2020, 238–243.
18. Mika, D. Fast Gradient Algorithm with Toral Decomposition for Complex ICA. Mechanical Systems and Signal Processing. 2022, 178, 109266.
19. Mika, D. and Jozwik, J. Lie Group Methods in Blind Signal Processing. s.l.: Sensors, 2020, 20(2), 440.
20. Mika, D. Fast gradient algorithm for complex ICA and its application to the MIMO systems. Scientific Reports. 2023, 13(1), 11633.
21. Moler, C. and Van Loan, C. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM review. 2003, 45(1), 3–49.
22. Salman, M.A. and Borkar, V.C. The exact methods to compute the matrix exponential. IOSR Journal of Mathematics 2016, 12(1), Ver. IV.
23. Ciric, I.R. New series formulae for computing accurately matrix exponentials. s.l.: In: 2018 18th International Symposium on Antenna Technology and Applied Electromagnetics (ANTEM). IEEE, 2018, 1–2.
24. Schatzman, M. Higher order alternate directions methods. Technical report. s.l.: Universite de Lyonn, Laboratoire d'analyse numerique, 1997.
25. Leite, F.S. and Crouch, P. Closed forms for the exponential mapping on matrix Lie groups based on Putzer's method. Journal of Mathematical Physics. 1999, 40(7), 3561–3568.
26. Politi, T. A formula for the exponential of a real skew-symmetric matrix of order 4. BIT Numerical Mathematics. 2001, 41, 842–845.
27. Gallier, J. and Xu, D. Computing exponentials of skew-symmetric matrices and logarithms of orthogonal matrices. International Journal of Robotics and Automation. 2003, 18(1), 10–20.
28. Horn, R.A. and Johnson, C.R. Matrix analysis. s.l.: Cambridge university press, 2012.
29. Celledoni, E. and Iserles, A. Approximating the exponential from a Lie algebra to a Lie group. Mathematics of Computation. 2000, 69(232), 1457–1480.
30. Celledoni, E., Iserles, A. Methods for the approximation of the matrix exponential in a Lie-algebraic setting. IMA Journal of Numerical Analysis. 2001, 21(2), 463–488.
31. Fiedler, B., Gröger, K. and Sprekels, J. Equadiff 99: (In 2 Volumes). World Scientific Publishing Company. 2000.
32. Kang, F. and Zai-Jiu, S. Volume-preserving algorithms for source-free dynamical systems. Numerische Mathematik. 1995, 71, 451–463.
33. Sanz-Serna, J.-M. and Calvo, M.-P. Numerical hamiltonian problems. s.l.: Courier Dover Publications, 2018.
34. Mclachlan, R.I. On the numerical integration of ordinary differential equations by symmetric composition methods. SIAM Journal on Scientific Computing. 1995, 16(1), 151–168.
35. Yoshida, H. Construction of higher order symplectic integrators. Physics letters A. 1990, 150(5–7), 262–268.
36. Owren, B. and Marthinsen, A. Integration methods based on canonical coordinates of the second kind. Numerische Mathematik. 2001, 87, 763–790.
37. Varadarajan, V.S. Lie groups, Lie algebras, and their representations. s.l.: Springer Science & Business Media, 2013.
38. Mika, D. and Kleczkowski, P. ICA-based single channel audio separation: new bases and measures of distance. Archives of Acoustics. 2011, 36(2), 311–331.
39. Mika, D., Kleczkowski, P. Lisbon,Automatic clustering of components for single channel ICA-based signal demixing. Portugal: s.n., 2010, 5350–5359.
40. Hyvarinen A., Karhunen J., Oja E. Independent Component Analysis. New York: John Wiley & Sons, 2001.
41. Wojtyński, W. Grupy i algebry Liego. Warszawa PWN, 1986. Vol. Biblioteka matematyczna t.60.
42. Tang, Y. Independent component analysis employing exponentials of sparse antisymmetric matrices. Neurocomputing. 2019, 325, 172–181.
43. Higham, N.J. Functions of matrices: theory and computation. s.l.: Society for Industrial and Applied Mathematics, 2008.
44. Mika, D. Incomplete cross splitting in the approximation of skew-symmetric matrix exponential. International Journal for Numerical Methods in Engineering, PREPRINT.
45. Iserles, A. and Zanna, A. Efficient computation of the matrix exponential by generalized polar decompositions. SIAM Journal on Numerical analysis. 2005, 42(5), 2218–2256.
46. Albera, L., Kachenoura, A, Comon, P., et al. ICA-based EEG denoising: a comparative analysis of fifteen methods. Bulletin of the Polish Academy of Sciences: Technical Sciences. Special issue on Data Mining in Bioengineering: 2012, 60(3), 407–418.
47. Zanna, A. and Munthe-Kaas, H.Z. Generalized polar decompositions for the approximation of the matrix exponential. SIAM journal on matrix analysis and applications. 2002, 23(3), 840–862.
48. Amari, S., Cichocki, A., Yanget, H.H., et. al. A new learning algorithm for blind signal separation. In: Advances in neural information processing systems. Morgan Kaufmann Publishers. 1996, 757–763.