

A regularized nonnegative third order tensor decomposition using a Primal-Dual Projected Gradient Algorithm: application to 3D fluorescence spectroscopy

Karima EL Qate¹, Mohammed El Rhabi², Abdelilah Hakim¹, Eric Moreau³,
and Nadège Thirion-Moreau³

¹ LAMAI, FSTG Marrakech, University of Cady Ayyad, Morocco
karima.elqate@gmail.com abdelilah.hakim@gmail.com

² Applied Mathematics and Computer Science Department, Ecole des Ponts
ParisTech (ENPC), Paris, France
mohammed.el-rhabi@enpc.fr

³ Aix Marseille Université, Université de Toulon, CNRS UMR 7020, LIS, Marseille,
France
thirion@univ-tln.fr moreau@univ-tln.fr

Abstract. This paper investigates the use of Primal-Dual optimization algorithms on multidimensional signal processing problems. The data blocks interpreted in a tensor way can be modeled by means of multi-linear decomposition. Here we will focus on the Canonical Polyadic Decomposition (CPD), and we will present an application to fluorescence spectroscopy using this decomposition. In order to estimate the factors or latent variables involved in these decompositions, it is usual to use criteria optimization algorithms. A classical cost function consists of a measure of the modeling error (fidelity term) to which a regularization term can be added if necessary. Here, we consider one of the most efficient optimization methods, Primal-Dual Projected Gradient. The effectiveness and the robustness of the proposed approach are shown through numerical examples.

Keywords: Constrained optimization · Nonnegative tensor decomposition · Primal-Dual · Regularization · Projected Gradient.

1 Introduction

This work deals with the Canonical Polyadic Decomposition (CPD) problem which has received much attention in the last ten years in various fields, ranging from telecommunications to chemometrics, spectral unmixing, neuroimaging, machine learning and Signal Processing for Biomedical Engineering. The CPD is a compact and flexible model which consists of decomposing a tensor into a minimal sum of rank-1 tensors. Initially developed by Harshman in psychometry [1], it was later referred to as Canonical Decomposition (Candecomp) [2], Parallel Factor Model (Parafac) [3] [4], and Topographic Components Model [5].

Many researchers have addressed the problem of computing the CPD of multi-way arrays, rewriting it as an optimization problem and more precisely as a minimization problem involving a sum of a (not necessarily convex) differentiable function and a (not necessarily differentiable) convex function:

$$\underset{z \in \mathbb{R}^N}{\text{minimize}} \underbrace{F(z)}_{\text{Fidelity}} + \underbrace{R(z)}_{\text{Regularization}} \quad (1)$$

The most popular approach is resorted to an iterative Alternating Least Squares (ALS) procedure [2]. Other iterative algorithms based on first and second order optimization methods such as gradient or conjugate gradient have also been proposed ([6] [7] for a full comparison of computation cost). Recently, a set of iterative algorithms based on a reduced functional has been introduced in [8]. In this article, we consider the minimization of a function which is the sum of a convex differentiable function F and a convex function R , which is not differentiable. A standard approach in this context consists of using the Primal-Dual Projected Gradient algorithm, which alternates a subgradient step. This proposed scheme differs from the other classical methods because it allows to obtain efficient combinatorial algorithms, in terms of approximation factor and calculation time. The main idea is to work simultaneously on the primal and the dual by finding an adequate solution for the dual, then to improve it at each step by optimizing an associated restricted primal problem.

The rest of the paper is organized as follows: Section 2 introduces a reminder of the principles of 3D fluorescence spectroscopy, and its links with the CPD problems. Section 3 describes the general principles of Primal-Dual Projected Gradient Algorithm. In section 4, we will explain how it can be used to solve the CPD problem and we will provide the resulting algorithm. Finally, section 5 provides some numerical results and a discussion on the algorithm performance.

2 Problem statement: CPD of fluorescent data

2.1 3D fluorescence spectroscopy

Whether it appears disturbing or playful, the phenomenon of fluorescence stirs our curiosity as much as it captures our gaze. This transitory manifestation results from the interactions between light and matter. When these are illuminated by incident light, some elements emit some of the energy at different wavelengths. The intensity of this fluorescence light varies as a function of the wavelengths of the incident light and the light emitted. The shape of these variations forms the fluorescence spectrum of the illuminated element and can be measured using a spectrofluorometer. We can distinguish two types of fluorescence spectrum: excitation and emission spectra. Therefore, spectroscopic fluorescence analysis is based on the processing of these signals. By successively, using the two monochromator of the Spectrofluorometer (in excitation and in emission), it is possible to measure the emission spectra for different excitation wavelengths. The Fluorescence Excitation-Emission Matrices (FEEM) are thus obtained [9]. When a

first order approximation of the Beer-Lambert law [10] is considered (for weak or low absorbance of the fluorophores), the intensity measured at (λ_e, λ_f) can be written as:

$$I(\lambda_e, \lambda_f, k) = C\phi I_0(\lambda_e)\gamma(\lambda_f)\varepsilon(\lambda_e)c_k \quad (2)$$

where C is a constant depending on the device, $I_0(\lambda_e)$ is the intensity of the light source, ϕ is the fluorescence quantum yield, ε denotes the relative absorbance spectrum (sometimes called the excitation spectrum), λ_f is the fluorescence emission wavelength, λ_e stands for the excitation wavelength, γ is the fluorescence relative emission spectrum and c_k is the concentration of the fluorophore in the sample number k . In the case of a mixture of N fluorophores, we obtain a generalized version of equation 2

$$I(\lambda_e, \lambda_f, k) = CI_0(\lambda_e) \sum_{n=1}^N \phi_n \gamma_n(\lambda_f) \varepsilon(\lambda_e) c_{k,n} \quad (3)$$

where $c_{k,n}$ stands for the concentration of n -th fluorescent solute in the k -th sample. The goal is to estimate the individual spectra of each fluorophore using the Canonical Polyadic Decomposition.

2.2 CP decomposition of 3-way arrays

We thus have a data set, denoted by X , containing the measurements of a physical quantity x_{ijk} function of three parameters, i, j, k . A trilinear model [11] [12] of X consists of a linear combination of three variables a, b and c depending respectively on i, j, k and a common parameter n such that:

$$x_{ijk} = \sum_{n=1}^N a_{in} b_{jn} c_{kn}, \quad \forall(i, j, k) \quad (4)$$

where the three involved matrices $A = (a_{in}) \in \mathbb{R}^{I \times N}$, $B = (b_{jn}) \in \mathbb{R}^{J \times N}$ and $C = (c_{kn}) \in \mathbb{R}^{K \times N}$ are the so-called loading matrices, whose N columns are the loading factors.

Matrix writing The data can be grouped in a single matrix by unfolding the tensor X in a preferred direction [13]. We denote by $X_{(1)}^{I,KJ}$ the matrix (I, KJ) representing the tensor unfolded in the direction i . We then look for the matrix relation existing between $X_{(1)}^{I,KJ}$ and the three matrices to be determined. For this, we introduce a special tensor product called Khatri-Rao product denoted by (\odot) , such that for two matrices A and C having the same number of columns N , we have:

$$A \odot C = [a_{:1} \otimes c_{:1} \quad a_{:2} \otimes c_{:2} \dots a_{:N} \otimes c_{:N}]$$

where \otimes refers to the Kronecker product. $X_{(1)}^{I,KJ}$ can then be written as:

$$X_{(1)}^{I,KJ} = A(C \odot B)^T \quad (5)$$

It is possible to unfold X according to j into a matrix $X_{(2)}^{J,KI}$, or again according to k into a matrix $X_{(3)}^{K,IJ}$. Thanks to the same reasoning, we obtain the two following formulas:

$$X_{(2)}^{J,KI} = B(C \odot A)^T, \quad X_{(3)}^{K,IJ} = C(B \odot A)^T \quad (6)$$

These three relations will allow us to estimate A , B and C .

In the particular case of Fluorescence Spectroscopy analysis, X contains different FEEMs corresponding to mixtures of N fluorophores in various proportions, then a_{in} represents the fluorescence factor (the product of the concentration and the quantum yield) of the fluorophore n , b_{jn} represents the value of the emission spectrum of the fluorophore n at the wavelength j and c_{kn} represents the value of the excitation spectrum of the fluorophore n at the wavelength k .

3 Primal-Dual Projected Gradient algorithm

An efficient approach for solving the aforementioned general minimization problem (1) consists of using the Primal-Dual Projected Gradient Algorithm. We refer the interested reader to [14],[15] and [16] for further details.

Using the Fenchel-Legendre transform, the previous minimization problem can be formulated as the search for a saddle point, more precisely, we have the following proposition:

Proposition 1. *Each of the following two problems admit a solution*

1. $\hat{z}_1 \in \arg \min_{z_1 \in Z_1} f_0(z_1) + f_1(Kz_1)$
2. $\hat{z}_2 \in \arg \max_{z_2 \in Z_2} -(f_0^*(-K^*z_2) - f_1^*(z_2))$

The first is called "primal problem" and the second is called "dual problem". In addition, we have:

$$\min_{z_1 \in Z_1} f_0(z_1) + f_1(Kz_1) = \max_{z_2 \in Z_2} -(f_0^*(-K^*z_2) - f_1^*(z_2)) \quad (7)$$

Let us now give one of the essential theorems in this context

Theorem 1. *Let $f_0 : U \rightarrow \mathbb{R}$ be a closed and convex functional on the set U , f_1 a closed and convex functional on the set V and let $K : U \rightarrow V$ be a continuous linear operator. Then we have the following equivalence:*

$$\underbrace{\min_{z \in U} \{f_0(z) + f_1(Kz)\}}_{\text{Primal}} = \underbrace{\min_{z \in U} \max_{\varphi \in V^*} \{ \langle Kz, \varphi \rangle - f_1^*(\varphi) + f_0(z) \}}_{\text{Primal-Dual}} \quad (8)$$

where z and φ are the primal and dual variables, respectively, f_1^* is the convex conjugate of f_1 , V^* is the dual space of V , and $\langle \cdot, \cdot \rangle$ is the inner product.

Variational problem We will now see how the primal-dual problem of theorem (1) can be written as a variational problem.

Before considering the specific problems, we first introduce a general saddle point problem notation

$$\min_{u \in U} \max_{v \in V} \{\mathbf{L}(u, v)\} \tag{9}$$

where U and V are closed convex, and \mathbf{L} is a convex-concave function defined over $U \times V$. In particular, $\mathbf{L}(\cdot, v)$ is convex for every $v \in V$, and $\mathbf{L}(u, \cdot)$ is concave for every $u \in U$.

Now we reduce the problem of approximating a saddle point (9) of \mathbf{L} on $U \times V$, the resolution of the associated variational inequality find $z^* \in X := U \times V$ s.t

$$\langle z - z^*, H(z^*) \rangle \geq 0 \quad \forall z \in Z \tag{10}$$

where

$$z = \begin{pmatrix} u \\ v \end{pmatrix} \quad \text{and} \quad H(z) = \begin{pmatrix} \partial_u \mathbf{L}(u, v) \\ -\partial_v \mathbf{L}(u, v) \end{pmatrix}$$

Projected gradient method Now, we will see how the preceding variational inequality (10) can be solved using the gradient projection method, for this we give the following proposition

Proposition 2. *Let r be a positive parameter and Z a convex set. An element z^* is solution of (10) if and only if*

$$z^* = P_Z(z^* - rH(z^*)) \tag{11}$$

where $P_Z(z)$ denote the orthogonal projection of the point z onto the nonempty, close, convex set Z .

Now we will use the fixed-point method to solve the last equation defined in the previous proposition:

Given $z \in Z$ compute the solution at step $n + 1$ by iterating the scheme

$$z_{n+1}^* = P_Z(z_n^* - rH(z_n^*)) \tag{12}$$

4 Application to the penalized nonnegative third order tensor factorization problem

4.1 Proposed Algorithm

Estimate \widehat{A} , \widehat{B} and \widehat{C} of A , B and C results from

$$\underset{A \in \mathbb{R}^{I \times N}}{\text{minimize}} \quad \mathbf{F}(A, B, C) + \mathbf{R}_1(A) \quad \text{s.t.} \quad A \geq 0. \tag{13}$$

$$\underset{B \in \mathbb{R}^{J \times N}}{\text{minimize}} \mathbf{F}(A, B, C) + \mathbf{R}_2(B) \quad \text{s.t.} \quad B \geq 0. \quad (14)$$

$$\underset{C \in \mathbb{R}^{K \times N}}{\text{minimize}} \mathbf{F}(A, B, C) + \mathbf{R}_3(C) \quad \text{s.t.} \quad C \geq 0. \quad (15)$$

where

$$\begin{aligned} \mathbf{F}(A, B, C) &= \frac{1}{2} \| X_{(1)}^{I,KJ} - A(C \odot B)^T \|_F^2 \\ &= \frac{1}{2} \| X_{(2)}^{J,KI} - B(C \odot A)^T \|_F^2 \\ &= \frac{1}{2} \| X_{(3)}^{K,JI} - C(B \odot A)^T \|_F^2 \end{aligned} \quad (16)$$

and $\| \cdot \|_F$ denotes the Frobenius norm. We opt for the following regularization terms

$$\mathbf{R}_1(A) = \alpha_A \| A \|_1, \quad \mathbf{R}_2(B) = \alpha_B \| B \|_1 \quad \text{and} \quad \mathbf{R}_3(C) = \alpha_C \| C \|_1 \quad (17)$$

where α_A, α_B and α_C are non negative regularization parameters, and $\| \cdot \|_1$ is the l_1 -norm.

Now we will follow the steps of the previous paragraph to solve our optimization problem. It becomes obvious, by comparing our following optimization problem with the primal-dual model in theorem (1), we get the following notation:

$$f_0(A) = \frac{1}{2} \| X_{(1)}^{I,KJ} - A(C \odot B)^T \|_F^2 \quad (18)$$

$$f_1(KA) = \alpha_A \sum_{i,n} |A(i, n)| \quad (19)$$

Now, we define the functionals:

$$\begin{aligned} K : U &\rightarrow V & \text{and} & & f_1 : V &\rightarrow \mathbb{R} \\ A &\mapsto KA & & & Y &\mapsto \sum_i \sum_n f(Y(i, n)) \end{aligned} \quad (20)$$

with $U = \mathbb{R}_+^{I \times N}$ and $V = \mathbb{R}_+^{I \times N}$ two reflexive spaces and $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = |x|$. The conjugate f^* of the function f is found as:

$$\begin{aligned} f^* : \mathbb{R} &\rightarrow \mathbb{R} \\ s &\mapsto \begin{cases} 0 & \text{if } |s| \leq 1 \\ +\infty & \text{if } |s| > 1 \end{cases} \end{aligned} \quad (21)$$

and therefore the convex conjugate f_1^* of f is

$$\begin{aligned} f_1^* : V &\rightarrow \mathbb{R} \\ Y &\mapsto \sum_i \sum_n f^*(Y(i, n)) \end{aligned} \quad (22)$$

Thus, using theorem (1), we now get the equivalent primal-dual problem:

$$\min_{A \in U} \max_{Y \in V^*} \{ \alpha_A \langle A, Y \rangle + \frac{1}{2} \| X_{(1)}^{I,KJ} - A(C \odot B)^T \|_F^2 \} \quad (23)$$

where $V^* = \{Y \in \mathbb{R}_+^{I \times N}; \sum_{i,n} |Y(i, n)| \leq 1\}$

Using the notation in (10) we get for each (i, n)

$$H(A, Y) = \begin{pmatrix} \alpha_A Y - \left(X_{(1)}^{I,KJ} - A(C \odot B)^T \right) (C \odot B) \\ -\alpha_A A \end{pmatrix} \quad (24)$$

Now by proceeding the last step which is used to apply the projection algorithm, and to provide a current estimate $(A^{(k+1)}, Y^{(k+1)})$ at iteration step $k + 1$, we obtain the following algorithm:

Algorithm

Initialize $A^{(0)}, Y^{(0)}$.

Repeat

1. $Y^{(k+1)} = P_{V^*} \{Y^{(k)} + r_1 \alpha_A A^{(k)}\}$
2. $A^{(k+1)} = P_U \{A^{(k)} - r_2 [\alpha_A Y^{(k+1)} (X_{(1)}^{I,KJ} - A(C \odot B)^T) (C \odot B)]\}$

until convergence

Where r_1, r_2 small positive constants. We do exactly the same for the estimation of matrices B and C .

Note, that the nonnegativity of the factors is ensured by the projection onto the subspaces U and V^* , introduced in the algorithm above.

5 Numerical simulations

The purpose of this subsection is to evaluate the performance of the Primal Dual Gradient Projected algorithm on synthetic data tensor built as follows: emission and excitation spectra and the three resulting Fluorescence Excitation Emission Matrices of three fluorophores, in our case tyrosine, phenylalanine and tryptophan have been downloaded at the following address: <http://omlc.orgi.edu/spectra/PhotochemCAD/index.html>. In the case of a perfect trilinear model and a known rank, our algorithm is able to recover the true solution. Moreover, by comparing our algorithm with Conjugate Gradient algorithm (with regularization terms l_1) and Gradient algorithm (with regularization terms l_1), we observe that our algorithm is the less computer time consuming ($\alpha_A = \alpha_B = \alpha_C = 0.01$ and $r1 = 0.01, r2 = 1/700$)

Primal Dual	Conjugate Gradient	Gradient
9.1984	28.6114	28.1782

Table1: Computer elapsed time (in second) for the different methods after 10000 iterations

On figure 1, we have given the 3 FEEM of reference, and the 3 FEEM obtained by application of primal-dual algorithm with regularization terms.

On figure 2, we have given the reference emission spectrum, and the estimated emission spectra using our algorithm starting from a random initialization ($\alpha_A = \alpha_B = \alpha_C = 0.01$) and ($r_1 = 0.01, r_2 = 1/700$).

On figure 3, we have given the reference excitation spectrum, and the estimated excitation spectra using our algorithm starting from a random initialization ($\alpha_A = \alpha_B = \alpha_C = 0.01$) and ($r_1 = 0.01, r_2 = 1/700$).

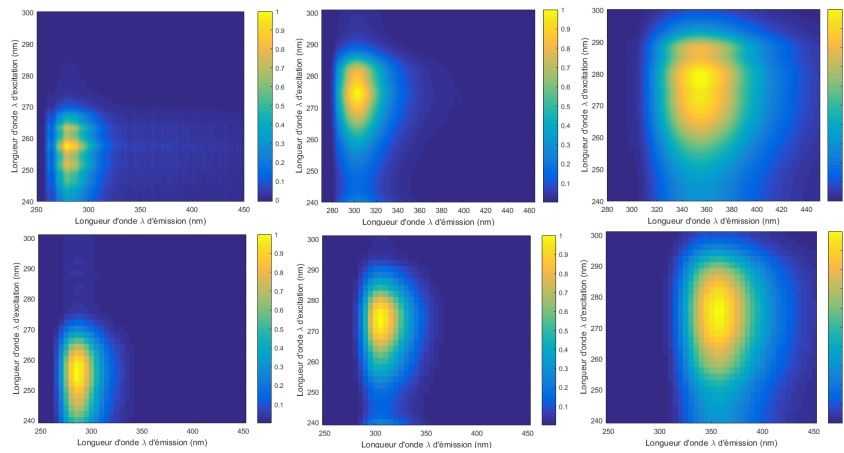


Fig. 1. Reference FEEM (Top), the estimated FEEM (bottom) using Primal-Dual Projected Gradient

6 Conclusion

Our work has investigated the problem of the nonnegative CPD of three-way array (third order tensors) arising in a variety of disciplines in the sciences and engineering. In particular, we have shown its interest in the field of 3D fluorescence spectroscopy. Efficiently, to solve that problem we have suggested Primal-Dual Projected Gradient Algorithm. Numerical results have proven the interest of the proposed approach.

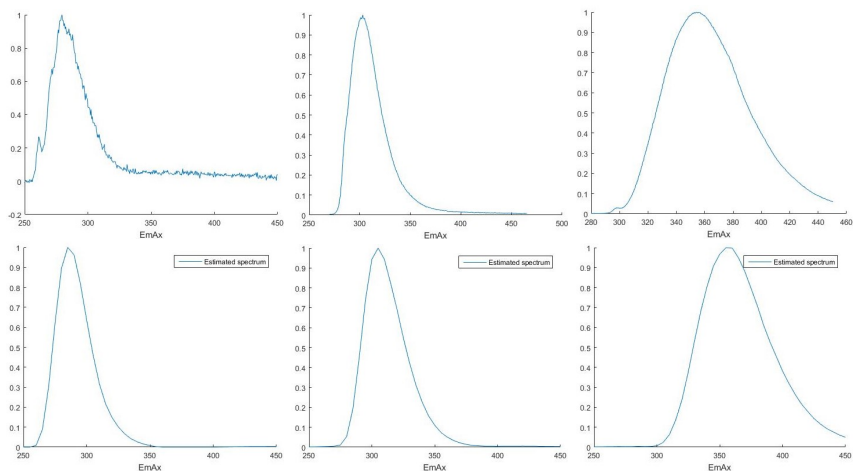


Fig. 2. Reference spectrum (Top), the estimated emission spectra (bottom): phenylalanine, tyrosine and tryptophan

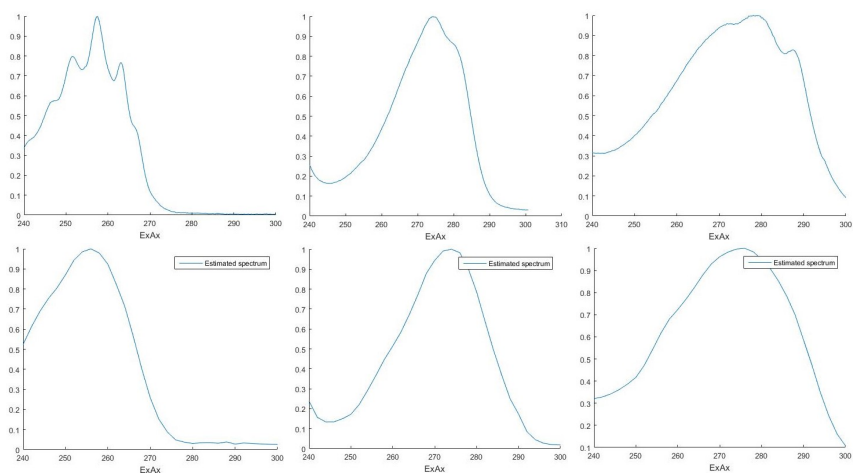


Fig. 3. Reference spectrum (Top), the estimated excitation spectra (bottom): phenylalanine, tyrosine and tryptophan

References

1. R. A. Harshman.: Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16:1-84, 1970.
2. Carroll, J. Douglas et Chang, Jih-Jie.: Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, vol. 35, no 3, p. 283-319, 1970.
3. Harshman, Richard A. and Lundy, Margaret E.: PARAFAC: Parallel factor analysis. *Computational Statistics and Data Analysis*, vol. 18, no 1, p. 39-72, 1994.
4. Harshman, Richard A. and Lundy, Margaret E.: The PARAFAC model for three-way factor analysis and multidimensional scaling. *Research methods for multimode data analysis*, vol. 46, p. 122-215, 1984.
5. Mocks, J.: Topographic components model for event-related potentials and some biophysical considerations. *IEEE transactions on biomedical engineering*, vol. 35, no 6, p. 482-484, 1988.
6. Tendeiro, Jorge, Dosse, M. Bennani, and Ten Berge, Jos MF.: First and second-order derivatives for CP and INDSCAL. *Chemometrics and Intelligent Laboratory Systems*, vol. 106, no 1, p. 27-36, 2011.
7. Acar, Evrim, Dunlavy, Daniel M., Kolda, Tamara G., et al.: Scalable tensor factorizations with missing data. In : *Proceedings of the 2010 SIAM international conference on data mining*. Society for Industrial and Applied Mathematics, p. 701-712, 2010.
8. Kindermann, Stefan et Navasca, Carmeliza.: News algorithms for tensor decomposition based on a reduced functional. *Numerical Linear Algebra with Applications*, vol. 21, no 3, p. 340-374, 2014.
9. P. G. Coble.: Characterization of marine and terrestrial dom in seawater using excitation-emission matrix spectroscopy, *Marine Chemistry*, vol. 52, pp. 325-346, 1996.
10. Smilde, Age, Bro, Rasmus, et Geladi, Paul. *Multi-way analysis: applications in the chemical sciences*. John Wiley and Sons, 2005.
11. J.-P. Royer, N. Thirion-Moreau, and P. Comon.: Computing the polyadic decomposition of nonnegative third order tensors, *Eurasip Signal Processing*, vol. 91, no. 9, pp. 2159-2171, 2011.
12. F. L. Hitchcock.: The expression of a tensor or a polyadic as a sum of products, *J. Math. and Phys.*, vol. 6, pp. 165-189, 1927.
13. A. Cichocki, R. Zdunek, A. H. Phan, and S. I. Amari.: *Non negative matrix and tensor factorizations: Application to exploratory multi-way data analysis and blind separation*. Wiley, 2009.
14. A. Chambolle, T. Pock.: *A first-order primal-dual algorithm for convex problems with applications to imaging*, 2010
15. M. Zhu, T. Chan.: *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, Tech. rep., UCLA CAM Report 08-34 2008.
16. Ekeland, Ivar, and Roger T.: *Convex analysis and variational problems*, Vol. 28. Siam, 1999.