Neurocomputing 133 (2014) 280–294

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Non-orthogonal joint diagonalization algorithm based on hybrid trust region method and its application to blind source separation

Tiao Jun Zeng, Quan Yuan Feng *

The Institute of Microelectronics of Southwest Jiaotong University, Chengdu, Sichuan 610031, PR China

ARTICLE INFO

Article history:
Received 18 September 2012
Received in revised form
23 September 2013
Accepted 2 November 2013
Communicated by S. Choi
Available online 10 January 2014

Keywords:
Non-orthogonal joint diagonalization
Hybrid trust region method
Blind source separate
Superlinear convergence

ABSTRACT

We proposed an algorithm for the efficient non-orthogonal joint diagonalization of a given set of matrices. The algorithm is based on the hybrid trust region method (HTRM) and its optimization approach, on which the efficiency of the method depends. Unlike traditional trust region methods that resolve sub-problems, HTRM efficiently searches a region via a quasi-Newton approach, by which it identifies new iteration points when a trial step is rejected. Thus, the proposed algorithm improves computational efficiency. Under mild conditions, we prove that the HTRM-based algorithm has global convergence properties together with local superlinear and quadratic convergence rates. Finally, we apply the combinative algorithm to blind source separation (BSS). Numerical results show that this method is highly robust, and computer simulations indicate that the algorithms excellently performs BSS.

1. Introduction

Since the joint diagonalization (JD) of fourth-order cumulant eigenmatrices [1,2] was proposed, the JD technique has been an essential tool in signal processing, especially in blind source separation (BSS) and array processing applications; JD has been extensively studied from both theoretical and algorithmic perspectives [2]. Available JD algorithms can be classified into two core categories: orthogonal [3–5] and non-orthogonal forms [6–9]. In the application of orthogonal diagonalization to BSS, observed signals are pre-whitened, so that they are uncorrelated and have unity variance. However, preprocessing during the 'whitening' operation can adversely affect the performance of separated signals because the statistical error introduced in this stage cannot be corrected in the "effective separation" stage [8]. The limitations of orthogonal JD prompted the development of non-orthogonal JD algorithms [6–9]. In particular, the algorithm developed by Li and Zhang [9] can avoid zero solutions and any degenerate solution (nonzero but singular or ill-conditioned solutions). The aforementioned features of the algorithm enhance the robustness and efficiency of non-orthogonal JD.

In [6], two algorithms based on gradient descent approaches are presented: the first is based on a classical gradient approach, and the second is grounded on a relative gradient approach. Two algorithms adopt respectively

$$\Delta B = -\mu_\alpha \nabla C_{BD}(B) \ \text{ and } \ \Delta B = -\mu_r \nabla C_{BD}(B) B$$
$$\left(B \in C^{M\times N}, \ \nabla C_{BD}(B) = 2\frac{\partial C_{BD}(B)}{\partial B^*}\right)$$

where $C_{BD}(B)$ denotes the cost function; B^* stands for the complex conjugate of complex matrix B; ∂ is the partial derivative operator; μ_α or μ_r represents a small positive number called the step-size or adaptation coefficient, which serves as the degressive direction of the cost function. Although $C_{BD}(B)$ will decrease under condition that step-size is a small enough positive number, the condition debases the computational efficiency as well. The intrinsic drawback of the gradient descent based algorithms is that they often produce the 'zigzag' phenomenon when the iterative point is close to the optimal solution; under such a situation, the algorithms will very slowly converge to or fail to generate an optimal solution. Conversely, trust region methods have been proven to be highly efficient in addressing optimization problems [10]. Trust-region methods define a region around the current iterate within which they trust the model to be an adequate representation of the objective function, and then choose the step to be the approximate minimizer of the model in this trust region [11]. These methods are robust, can be applied to ill-conditioned problems, and have strong convergence properties under mild conditions. Many researchers have recently studied nonmonotone adaptive trust region methods (NATRM) for unconstrained optimization problems [12,13]. NATRM can automatically produce an adaptive trust region radius whenever a trial step is rejected, and will decrease

* Corresponding author. Tel: +86 28 87603147, mobile: +86 13699049169; fax: +86 28 87603212.
E-mail addresses: ztj2003atgxnu@163.com (T.J. Zeng), fengquanyua@163.com (Q.Y. Feng).

0925-2312/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.11.008

the functional values after finite iterations. The main disadvantage of NATRM lies in identifying new trial iterations; it requires considerable computational time to repeatedly solve sub-problems. Motivated by rectifying this shortcoming and inspired by Nocedal and Yuan [14], we propose a hybrid trust region method (HTRM) that adopts a different search approach at each iteration. The search direction $d_k = x_{k+1} - x_k$ is generated by solving the sub-problem of the cost function (discussed in Section 3). If d_k is rejected, the sub-problem does not need to be resolved. Using a quasi-Newton method (QNM), we compute the search direction $d_k = (-B_k^{-1}\nabla f(x_k))$. ($B_k \approx \nabla^2 f(x_k)$) is updated using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) formula to maintain the positive definitiveness of B_k [15].

The rest of the paper is organized as follows. In Section 2, we describe the non-orthogonal JD problem, then derive the formula of the gradient and Hessian matrix of the cost function. In Section 3, we introduce the HTRM-based algorithm, prove that it is well defined, and identify its convergence properties. The numerical results are also presented in this section. In Section 4, we apply the general algorithm to a BSS problem and the results of the correlation simulation are discussed in Section 5. The conclusions are summarized in Section 6.

2. Non-orthogonal joint diagonalization problem

The non-orthogonal JD problem is stated as follows [6,8,9,16]: we consider a set $\Psi = \{M_i | i = 1, 2, \cdots K,\ M_i \in R^{M\times M}\}$ (or $M_i \in C^{M\times M}$, depending on the applications). Every M_i admits the following decomposition:

$$M_i = \tilde{A}\Lambda_i\tilde{A}^T, \tag{1}$$

where $(\cdot)^T$ represents the transpose operator. The approximate JD problem seeks a ‘diagonalizing matrix’ $\tilde{A} \in R^{M\times N}$ $(N \le M)$ and K associated diagonal matrices $\Lambda_1, \Lambda_2, \cdots \Lambda_K \in R^{N\times N}$, such that

$$\Gamma(\tilde{A}, \{\Lambda_i\}) = \sum_{i=1}^{K} \| M_i - \tilde{A}\Lambda_i\tilde{A}^T \|^2, \ \forall M_i \in \Psi \tag{2}$$

is minimized, in the application of BSS, without loss of generality, let $M{=}N$ (the proof can be seen in Appendix A), $\|\bullet\|$ denotes the Euclidean norm.

Premultiplying M_i in Eq. (1) by $\tilde{A}$'s inverse matrix A and postmultiplying by A^T yields

$$AM_iA^T = \Lambda_i, \quad i = 1, 2, \cdots, K \tag{3}$$

In estimating matrix A instead of $\tilde{A}$ itself, the following quadratic cost function should be considered:

$$F(A, \{\Lambda_i\}) = \sum_{i=1}^{K} \| AM_iA^T - \Lambda_i \|^2, \ \forall M_i \in \Psi \tag{4}$$

Comparing with the cost function $\widehat{F}(A) = \sum_{i=1}^{K} \| Off(AM_iA^T) \|^2$, where the matrix operators $Diag\{\bullet\}$ and $Off\{\bullet\}$ are respectively defined as

$$DiagC = \begin{pmatrix} c_{11} & 0 & \cdots & 0 \\ 0 & c_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_{NN} \end{pmatrix} \text{ and}$$

$$OffC = C - \begin{pmatrix} c_{11} & 0 & \cdots & 0 \\ 0 & c_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_{NN} \end{pmatrix},$$

the former cost function is more general, $\widehat{F}(A)$ is the special case when the conditions $\Lambda_i = Diag\{AM_iA^T\} (i = 1, 2, \cdots, K)$ hold, but the conditions are very strict to the problem of minimizing cost function and can hardly be satisfied.

The following theorem illuminates, $F(A, \{\Lambda_i\})$ converging zero is the sufficient condition for A convergence to be a separable matrix.

Lemma 1. *if $F(A, \{\Lambda_i\}) \to 0$, then $A\tilde{A} = P_jD$, where $\tilde{A}$ is the real mixing matrix of BSS problem, D is a nonsingular diagonal matrix and P_j is the N-order permutation matrix (it contains exactly one element 1 in each row and each column).*

Proof. $\because \| AM_iA^T - \Lambda_i \| \to 0, \quad i = 1, 2, \cdots, K,$ as $F(A, \{\Lambda_i\}) \to 0$, thereof, $AM_iA^T = \Lambda_i$, on the other hand, $M_i = \tilde{A}\Lambda_i'\tilde{A}^T (\Lambda_i' = E(S(t)S^T(t - \tau_i))$, $S(t)$ is the vector of source signals), so $AM_iA^T = A(\tilde{A}\Lambda_i'\tilde{A}^T)A^T = (A\tilde{A})\Lambda_i'(A\tilde{A})^T = \Lambda_i$. Let $C = A\tilde{A}$, then $C\Lambda_i'C^T = \Lambda_i$. Moreover, Λ_i and Λ_i' are nonsingular diagonal matrices, supposing $C \ne P_jD$, it divide two cases:

(1) there only exist $N+1$ nonzero elements in the N order nonsingular square matrix C, thereof, without loss of generality, supposing its diagonal elements are nonzero, thereof, there are two nonzero elements in the m-th row and the k-th column, i.e. c_{mm}, c_{mk} and c_{kk} are nonzero elements, so, $(\Lambda_i)_{mk} = c_{mm}(\Lambda_i')_{mm}0 + c_{mk}(\Lambda_i')_{kk}(C^T)_{kk} = c_{mk}(\Lambda_i')_{kk}c_{kk} \ne 0$; this contradicts the fact that Λ_i is a diagonal matrix.
(2) There exist more than $N+1$ nonzero elements in the N order nonsingular square matrix C,

$$\because (\Lambda_i)_{mn} = (C\Lambda_i'C^T)_{mn} = \sum_{k=1}^{N} (C)_{mk}(\Lambda_i')_{kk}(C^T)_{kn}$$
$$= \sum_{k=1}^{N} (C)_{mk}(\Lambda_i')_{kk}(C)_{nk} = \sum_{k=1}^{N} c_{mk}(\Lambda_i')_{kk}c_{nk} 1 \le m, n \le N,$$
$$m \ne n, \quad i = 1, 2, \cdots, K.$$

The elements of matrix C are independent of each other, thereof, it is impossible to make the $K(N^2 - N)$ off-diagonal elements $(\Lambda_i)_{mn}$ equal to zero, so it also contradicts the fact that Λ_i is a diagonal matrix.

Base on above analysis, we conclude that $A\tilde{A} = P_jD$.

Lemma 1. *indicates that A converges in a separable matrix when $F \to 0$ (in practice, such theorem indicates that $F \to \varepsilon$, ε is a very small positive number). In other words, a BBS problem can be transformed into an unconstrained optimization problem based on the Lemma.*

Some algorithms are dedicated to a unitary matrix A and have led to Jacobi-like algorithms. When these algorithms are applied in the BSS context, the unitary constraint can be fulfilled by a classical whitening of the observations. However, this preliminary stage has been proven to limit the attainable performance. In this paper, we are proposing a solution to the non-unitary JD problem by relying on the optimization of the cost function [Eq. (4)] to estimate matrix $A \in R^{N\times N}$. To this aim, we use the hybrid trust region algorithm. First, several notations are introduced for the next derivation. Let $tr\{\cdot\}, d\{\cdot\}, vec\{\cdot\}, \otimes$ denote the trace, differential, and the row vector operators, as well as the Kronecker product, respectively. The derivation of cost function's gradient and Hessian matrices can be seen in Appendix B.

3. Algorithm based on the hybrid trust region method

We consider the following unconstrained optimization problem:

$$\min f(x), x \in R^n, \tag{5}$$

where $f : R^n \to R$ is a continuously differentiable function. The optimization problem has become an important research focus given its wide range of potential applications. Throughout the paper, we introduce some notations for convenience. $\|\cdot\|$ denotes the Euclidean norm on R^n; $\{x_k\}$ is a sequence of points generated by