CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.2021.nla.03

Vol. **2**, No. 2, pp. 297-312
June 2021

A Convergence Analysis on the Iterative Trace Ratio Algorithm and its Refinements

Zhongxiao Jia* and Fuhui Lai

Department of Mathematical Sciences, Tsinghua University, Beijing 100084, P.R. China.

Received 20 May 2020; Accepted 28 September 2020

Abstract. In many applications one needs to process massive data of high dimensionality. In order to make data compact and reduce computational complexity, dimensionality reduction algorithms are commonly used. Linear discriminant analysis (LDA) is one of the most used approaches, which leads to a matrix trace ratio problem, i.e., maximization of the trace ratio $\rho(V)=\mathrm{Tr}(V^TAV)/\mathrm{Tr}(V^TBV)$, where A and B are $n\times n$ real symmetric matrices with B positive definite, and V is an $n\times p$ column orthonormal matrix. In this paper, we consider a commonly used Iterative Trace Ratio (ITR) algorithm developed by Ngo et al., *The trace ratio optimization problem, SIAM Review, 54 (3) (2012), pp. 545–569*. In implementations, it is common to use the symmetric Lanczos method to compute the p eigenvectors of a certain large matrix corresponding to its p largest eigenvalues at each iteration, and the resulting algorithm is abbreviated as ITR_L. We establish the global convergence and local quadratic convergence of the trace ratio itself. We then make two improvements over ITR_L: (i) using the refined Lanczos method to compute the desired eigenvectors at each iteration and (ii) providing a better initial guess via solving a generalized eigenvalue problem of the matrix pair (A,B). The resulting algorithms are abbreviated as ITR_RL and ITR_GeigRL, respectively. Numerical experiments demonstrate that ITR_RL and ITR_GeigRL outperform ITR_L substantially.

AMS subject classifications: 65F15, 65F30, 62H30, 15A18

Key words: Dimensionality reduction, trace ratio, global convergence, eigenvalue, eigenvector, refined projection, refined Lanczos algorithm, implicitly restarted, preprocess.

1 Introduction

In many applications such as data mining, machine learning and bioinformatics, one has to process massive data. Due to excessive storage requirement and computational cost,

*Corresponding author. *Email addresses:* `jiazx@tsinghua.edu.cn` (Z. Jia), `laifh14@163.com` (F. Lai)

in order to overcome the difficulty of high dimensionality, make an effective analysis and process the data, it is necessary to use dimensionality reduction, whose aim is to reduce the high dimensionality and meanwhile to retain the useful information of data features and structures. Some dimensionality reduction approaches are from statistics and geometry, e.g., the linear discriminant analysis (LDA) [5] and the local preserving projection (LPP) [7], etc. They ultimately lead to the matrix trace ratio problem: Find the column orthonormal solution V_* to the optimization problem

$$\max_{V^TV=I_p} \frac{\mathrm{Tr}(V^TAV)}{\mathrm{Tr}(V^TBV)}, \tag{1.1}$$

where A and B are $n\times n$ real symmetric, V is $n\times p$ column orthonormal, I_p is the identity of order p with $1<p\ll n$ being the classification number of data features, and 'Tr(·)' is the trace of a square matrix. For convenience, we suppose that B is symmetric positive definite, and define $\rho(V)=\frac{\mathrm{Tr}(V^TAV)}{\mathrm{Tr}(V^TBV)}$ to be the objective function. Therefore, the trace ratio optimization problem is to find V_* and the maximum trace ratio $\rho_*=\rho(V_*)$.

The Foley-Sammon transform (FST) method [4] is based on the Fisher discriminant criterion [3]. Exploiting this criterion, Sammon proposes an optimal set of discriminant surfaces [16], which is extended to the optimal set of discriminant vectors [4]. One can use FST method to obtain optimal sets of discriminant vectors successively. The method has received much attention in the field of pattern recognitions; see, e.g., [8]. There have been some FST based methods under different conditions, for example, Liu's method [14]. Since the discriminant vectors are successively obtained and only locally optimal, there is no guarantee that the discriminant vectors obtained ultimately are globally optimal. This is a typical shortcoming that many local optimization methods share.

There have been some methods for solving (1.1). Yan and Tang [20] propose a multiscale search algorithm, which, at each iteration, solves a symmetric matrix eigenvalue problem, but no convergence analysis is made. It appears that this method converges slowly, similar to the bisection methods proposed by Xiang et al. [19] and Guo et al. [6]. Wang et al. [18] improve these methods and propose a Newton like method for solving problem (1.1), in which they give up multiscale and bisection searches and instead solve a symmetric matrix eigenvalue problem at each iteration so as to accelerate the convergence. Ngo et al. [15] develop an Iterative Trace Ratio (ITR) method. In implementations, they suggest to use the symmetric Lanczos method at iteration k to find the p eigenvectors of $A-\rho_kB$ associated with the p largest eigenvalues where ρ_k is the current approximation to ρ_*, and we shall call the resulting ITR algorithm ITR_L. Specifically, the implicitly restarted Lanczos algorithm is used.

The convergence of ITR has been analyzed in, e.g., [1, 22, 23], where the global linear convergence of $\{\rho_k\}$ and the local quadratic convergence of $\{V_k\}$ has been proved with the columns of V_k being the unit-length eigenvectors of $A-\rho_{k-1}B$ corresponding to its p largest eigenvalues counting their multiplicities. Throughout the paper, the p largest eigenvalues are always meant to count their multiplicities. The convergence of