Robust and Efficient Computation of Eigenvectors in a Generalized Spectral Method for Constrained Clustering

Chengming Jiang
University of California, Davis
cmjiang@ucdavis.edu

Huiqing Xie
East China University of Science and Technology
hqxie@ecust.edu.cn

Zhaojun Bai
University of California, Davis
zbai@ucdavis.edu

Abstract

FAST-GE is a generalized spectral method for constrained clustering [Cucuringu et al., AISTATS 2016]. It incorporates the must-link and cannot-link constraints into two Laplacian matrices and then minimizes a Rayleigh quotient via solving a generalized eigenproblem, and is considered to be simple and scalable. However, there are two unsolved issues. Theoretically, since both Laplacian matrices are positive semi-definite and the corresponding pencil is singular, it is not proven whether the minimum of the Rayleigh quotient exists and is equivalent to an eigenproblem. Computationally, the locally optimal block preconditioned conjugate gradient (LOBPCG) method is not designed for solving the eigenproblem of a singular pencil. In fact, to the best of our knowledge, there is no existing eigensolver that is immediately applicable. In this paper, we provide solutions to these two critical issues. We prove a generalization of Courant-Fischer variational principle for the Laplacian singular pencil. We propose a regularization for the pencil so that LOBPCG is applicable. We demonstrate the robustness and efficiency of proposed solutions for constrained image segmentation. The proposed theoretical and computational solutions can be applied to eigenproblems of positive semi-definite pencils arising in other machine learning algorithms, such as generalized linear discriminant analysis in dimension reduction and multisurface classification via eigenvectors.

Proceedings of the 20^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2017, Fort Lauderdale, Florida, USA. JMLR: W&CP volume 54.

1 INTRODUCTION

Clustering is one of the most important techniques for statistical data analysis, with applications ranging from machine learning, pattern recognition, image analysis, bioinformatics to computer graphics. It attempts to categorize or group date into clusters on the basis of their similarity. Normalized Cut [Shi and Malik, 2000] and Spectral Clustering [Ng et al., 2002] are two popular algorithms.

Constrained clustering refers to the clustering with a *prior* domain knowledge of grouping information. Here relatively few must-link (ML) or cannot-link (CL) constraints are available to specify regions that must be grouped in the same partition or be separated into different ones [Wagstaff et al., 2001]. With constraints, the quality of clustering could be improved dramatically. In the past few years, constrained clustering has attracted a lot of attentions in many applications such as transductive learning [Chapelle et al., 2006, Joachims, 2003], community detection [Eaton and Mansbach, 2012, Ma et al., 2010] and image segmentation [Chew and Cahill, 2015, Cour et al., 2007, Cucuringu et al., 2016, Eriksson et al., 2011, Wang et al., 2014, Xu et al., 2009, Yu and Shi, 2004].

Existing constrained clustering methods based on spectral graph theory can be organized in two classes. One class is to impose the constraints on the indicator vectors (eigenspace) *explicitly*, namely, the constraints are either encoded in a linear form [Cour et al., 2007, Eriksson et al., 2011, Xu et al., 2009, Yu and Shi, 2004] or in a bilinear form [Wang et al., 2014]. The other class of the methods *implicitly* incorporates the constraints into Laplacians. The semi-supervised normalized cut [Chew and Cahill, 2015] casts the constraints as a low-rank matrix to the Laplacian of the data graph. In a recent proposed generalized spectral method (FAST-GE) [Cucuringu et al., 2016], the ML and CL con-

straints are incorporated by two Laplacians L_G and L_H. The clustering is realized by solving the optimization problem

$$\inf_{\substack{x\in\mathbb{R}^n \\ x^T L_H x>0}} \frac{x^T L_G x}{x^T L_H x}, \tag{1}$$

which subsequently is converted to the problem of computing a few eigenvectors of the generalized eigenproblem

$$L_G x = \lambda L_H x. \tag{2}$$

It is shown that the FAST-GE algorithm works in nearly-linear time and provides some theoretical guarantees for the quality of the clusters [Cucuringu et al., 2016]. FAST-GE demonstrates its superior quality to the other spectral approaches, namely CSP [Wang et al., 2014] and COSf [Rangapuram and Hein, 2012].

However, there are two critical unsolved issues associated with the FAST-GE algorithm. The first one is theoretical. Since both Laplacians L_G and L_H are symmetric positive semi-definite and the pencil $L_G - \lambda L_H$ is singular, i.e., $\det(L_G - \lambda L_H) \equiv 0$ for all λ. The Courant-Fischer variational principle [Golub and Van Loan, 2012, sec.8.1.1] is not applicable. Consequently, it is not proven whether the infimum of the Rayleigh quotient (1) is equivalent to the smallest eigenvalue of problem (2). The second issue is computational. LOBPCG is not designed for the generalized eigenproblem of a singular pencil. In fact, to the best of our knowledge, there is no existing eigensolver that is applicable to the large sparse eigenproblem (2).

In this paper, we address these two critical issues. Theoretically, we first derive a canonical form of the singular pencil $L_G - \lambda L_H$ and show the existence of the finite real eigenvalues. Then we generalize the Courant-Fischer variational principle to the singular pencil $L_G - \lambda L_H$ and prove that the infimum of the Rayleigh quotient (1) can be replaced by the minimum, namely, the existence of minima is guaranteed. Based on these theoretical results, we can claim that the optimization problem (1) is indeed equivalent to the problem of finding the smallest finite eigenvalue and associated eigenvectors of the generalized eigenproblem (2). Computationally, we propose a regularization to transform the singular pencil $L_G - \lambda L_H$ to a positive definite pencil $K - \sigma M$, where K and M are symmetric and M is positive definite. Consequently, we can directly apply LOBPCG and other eigensolvers, such as Lanczos method in ARPACK [Lehoucq et al., 1998]. We demonstrate the robustness and efficiency of the proposed approach for constrained segmentations of a set of large size images.

We note that the essence of the proposed theoretical and computational solutions in this paper is about mathematically rigorous and computationally effective treatments of large sparse symmetric positive semi-definite pencils. The proposed results in this paper could be applied to eigenproblems arising in other machine learning techniques, such as generalized linear discriminant analysis in dimension reduction [He et al., 2005, Park and Park, 2008, Zhu and Huang, 2014] and multisurface proximal support vector machine classification via generalized eigenvectors [Mangasarian and Wild, 2006].

2 PRELIMINARIES

A weighted undirected graph G is represented by a pair (V, W), where $V = \{v_1, v_2, \ldots, v_n\}$ denotes the set of vertices, and $W = (w_{ij})$ is a symmetric weight matrix, such that $w_{ij} \geqslant 0$ and $w_{ii} = 0$ for all $1 \leqslant i, j \leqslant n$. The pair (v_i, v_j) is an edge of G *iff* $w_{ij} > 0$. The degree d_i of a vertex v_i is the sum of the weights of the edges adjacent to v_i:

$$d_i = \sum_{i=j}^{n} w_{ij}.$$

The degree matrix is $D = \mathrm{diag}(d_1, d_2, ...d_n)$. The Laplacian L of G is defined by $L = D - W$ and has the following well-known properties:
(a) $x^T L x = \frac{1}{2}\sum_{i,j} w_{ij}(x_i - x_j)^2$;
(b) $L \succeq 0$ if $w_{ij} \geqslant 0$ for all i, j;
(c) $L \cdot \mathbf{1} = 0$;
(d) Let λ_i denote the ith smallest eigenvalue of L. If the underlying graph of G is connected, then $0 = \lambda_1 < \lambda_2 \leqslant \lambda_3 \leqslant \ldots \leqslant \lambda_n$, and $\dim(\mathcal{N}(L)) = 1$, where $\mathcal{N}(L)$ denotes the nullspace of L.

Let A be a subset of vertices V and $\bar{A} = V/A$, the quantity

$$\mathrm{cut}_G(A) = \sum_{v_i\in A, v_j\in\bar{A}} w_{ij} \tag{3}$$

is called the *cut* of A on graph G. The *volume* of A is the sum of the weights of all edges adjacent to vertiecs in A:

$$\mathrm{vol}(A) = \sum_{v_i\in A}\sum_{j=1}^{n} w_{ij}.$$

It can be shown, see for example [Gallier, 2013], that

$$\mathrm{cut}_G(A) = \frac{x^T L x}{(a-b)^2}, \tag{4}$$

where x is the indicator vector such that its i element $x_i = a$ if $v_i \in A$, and $x_i = b$ if $v_i \notin A$, and a, b are two distinct real numbers.

For a given weighted graph $G = (V, W)$, the k-way partitioning is to find a partition $(A_1, \ldots, A_k)$ of V,

such that the edges between different subsets have very low weight and the edges within a subset have very high weight. The Normalized Cut (Ncut) method [Shi and Malik, 2000] is a popular method for unconstrained partitioning. For the given weighted graph G, the objective of Ncut for a 2-way partition $(A, \bar{A})$ is to minimize the quantity

$$\mathrm{Ncut}(A) = \frac{\mathrm{cut}(A)}{\mathrm{vol}(A)} + \frac{\mathrm{cut}(\bar{A})}{\mathrm{vol}(\bar{A})}. \tag{5}$$

It is shown that

$$\min_{A} \mathrm{Ncut}(A) = \min_{x} \frac{x^T L x}{x^T D x} \tag{6}$$

subject to $x^T D\mathbf{1} = 0$, where x is a binary indicator vector and the i-th element x_i of x satisfying $x_i \in \{1, -b\}$ and $b = \mathrm{vol}(A)/\mathrm{vol}(\bar{A})$. $\mathbf{1}$ is a vector of all ones. Note that $D \succ 0$ under the connectivity assumption of G.

The binary optimization problem (6) is known to be NP-complete. After relaxing x to be a real-valued vector $x \in \mathbb{R}^n$, it becomes solving the Rayleigh quotient optimization problem

$$\inf_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{x^T L x}{x^T D x} \quad \text{subject to } x^T D\mathbf{1} = 0. \tag{7}$$

By the Courant-Fischer variational principle, see for example [Golub and Van Loan, 2012, sec.8.1.1], under the assumption $D \succ 0$, there is a minimum of (7) and the minimum is reached by the eigenvector corresponding to the second smallest eigenvalue λ_2 of the generalized symmetric definite eigenproblem

$$Lx = \lambda D x. \tag{8}$$

For a k-way partitioning, the spectral clustering method [Ng et al., 2002] is widely used. There the normalized Laplacian $L_{\mathrm{n}} = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}$ is firstly constructed, and then eigenvectors $X = [x_1, \ldots, x_k]$ corresponding to the k smallest eigenvalues of L_{n} are computed. Subsequently, a matrix $Y \in \mathbb{R}^{n \times k}$ is formed by normalizing each row of X. Treated as points in high dimensional space, rows of Y are clustered into k clusters via k-means.

3 FAST-GE MODEL

In this section, we present the FAST-GE model [Cucuringu et al., 2016] for a k-way constrained spectral clustering. For a given weighted graph $G_D = (V, W_D)$ and k disjoint vertex sets $\{V_1, V_2, \ldots, V_k\}$ of constraints, where $V_i \subset V$ and vertices in the same V_i form the ML constraints and any two vertices in different V_i form the CL constraints, the objective of FAST-GE is to find a partition $(A_1, A_2, \ldots, A_k)$ of V such that $V_i \subseteq A_i$, and the edges between different subsets have very low weight, the edges within a subset have very high weight.

Let $\mathrm{cut}_{G_D}(A_p)$ be the cut of the subset A_p on G_D defined in (3), and we introduce a graph

$$G_M = (V, W_M), \tag{9}$$

where the weight matrix $W_M = \sum_{\ell=1}^{k} W_{M_\ell}$. The entries of W_{M_ℓ} are defined as follows: if v_i and v_j are in the same V_ℓ, then $W_{M_\ell}(i,j) = d_i d_j/(d_{\min} d_{\max})$, where d_i and d_j are the degrees of v_i and v_j in G_D, $d_{\min} = \min_i d_i$ and $d_{\max} = \max_i d_i$. Otherwise $W_{M_\ell}(i,j) = 0$. Then the quantity measuring the violation of ML constraints of the cut of A_p is defined by

$$\mathrm{cut}_{G_M}(A_p) = \sum_{v_i \in A_p, v_j \in \bar{A}_p} W_M(i,j).$$

To measure the cut of A_p that satisfies the CL constraints, we consider a graph

$$G_H = (V, W_H), \tag{10}$$

where $W_H = W_C + W_C^T + K^{(c)}/n$, and the entries of W_C are defined as follows: if $v_i \in V_{\ell_1}$ and $v_j \in V_{\ell_2}$ ($\ell_1 \neq \ell_2$), then $W_C(i,j) = d_i d_j/(d_{\min} d_{\max})$. Otherwise, $W_C(i,j) = 0$. $K^{(c)}$ is called a demanding matrix defined by $K^{(c)}(i,j) = (d_i^{(c)} \cdot d_j^{(c)})/\sum_i d_i^{(c)}$, where $d_i^{(c)}$ is the degree of v_i in $(V, W_C + W_C^T)$. Then the quantity to measure the cut of A_p that satisfies the CL constraints is defined by

$$\mathrm{cut}_{G_H}(A_p) = \sum_{v_i \in A_p, v_j \in \bar{A}_p} W_H(i,j).$$

Based on the quantities $\mathrm{cut}_{G_D}(A_p)$, $\mathrm{cut}_{G_M}(A_p)$ and $\mathrm{cut}_{G_H}(A_p)$, the measure of "badness" for the partitioning $(A_p, \bar{A}_p)$ is defined as

$$\phi_p = \frac{\mathrm{cut}_{G_D}(A_p) + \mathrm{cut}_{G_M}(A_p)}{\mathrm{cut}_{G_H}(A_p)}. \tag{11}$$

By denoting the graphs $G = (V, W_D + W_M)$ and $H = G_H$, it can be shown that the quantity (11) for measuring the "badness" can be rewritten as

$$\phi_p = \frac{\mathrm{cut}_G(A_p)}{\mathrm{cut}_H(A_p)}. \tag{12}$$

The objective of FAST-GE for a k-way constrained partitioning is then given by

$$\min_{A_1, \ldots, A_k} \max_p \phi_p. \tag{13}$$

When $k = 2$, we have $\phi_1 = \phi_2$. Denote $A = A_1$, the minimax problem (13) is simplified to

$$\min_A \phi_2 = \min_A \frac{\text{cut}_G(A)}{\text{cut}_H(A)}. \tag{14}$$

By the definition of the quantity "cut" in (4), the optimization problem (14) is equivalent to the following binary optimization of the Rayleigh quotient

$$\min_{x^T L_H x > 0} \frac{x^T L_G x}{x^T L_H x}, \tag{15}$$

where $x = (x_i)$ is a binary indicator vector with $x_i \in \{a, b\}$. L_G and L_H are the Laplacians of G and H, namely, $L_G = (D_D + D_M) - (W_D + W_M)$ and $L_H = D_H - W_H = D_H - (W_C + W_C^T + K^{(c)})/n$ and D_D, D_M and D_H are the degree matrices of G_D, G_M and G_H, respectively. Note that the constraint vertex sets $\{V_1, V_2, \ldots, V_k\}$ are used to define W_M and W_H, so the ML and CL constraints are incorporated into L_G and L_H. One can readily verify that both L_G and L_H are positive semi-definite. Furthermore, the pencil $L_G - \lambda L_H$ is singular, i.e., $\det(L_G - \lambda L_H) \equiv 0$ for all scalar λ. The common nullspace of L_G and L_H is $\mathcal{N}(L_G) \cap \mathcal{N}(L_H) = \text{span}\{\mathbf{1}\}$. This is due to the fact that when the underlying graph of G_D is connected, the dimension of the nullspace of L_G is 1.

Similar to the argument for the Ncut (6), the optimization problem (15) is NP-complete. In practice, the binary indicator vector x in (15) is relaxed to a real-valued vector $x \in \mathbb{R}^n$, then the optimization (15) is relaxed to

$$\inf_{\substack{x \in \mathbb{R}^n \\ x^T L_H x > 0}} \frac{x^T L_G x}{x^T L_H x}. \tag{16}$$

We notice that the minimum "min" in (15) must be replaced by the infimum "inf" in (16) since the existence of the minimum in the real domain $\mathbb{R}^n$ is not proven, when both L_G and L_H are positive semi-definite and share a non-empty common nullspace.

4 VARIATIONAL PRINCIPLE AND EIGENVALUE PROBLEM

Since L_H is positive semi-definite in (16), Courant-Fischer variational principle is not applicable. The optimization problem (16) cannot be immediately transformed to an equivalent eigenproblem. Furthermore, since the pencil $L_G - \lambda L_H$ is singular, it is unknown about the existence and number of the finite eigenvalues of the pencil. To address these theoretical issues, we have proven the following theorem for the existence of the finite eigenvalues and a generalization of the Courant-Fischer variational principle.

Theorem 1. *Let $n \times n$ matrices L_G and L_H be positive semi-definite, then (a) the pencil $L_G - \lambda L_H$ has r finite non-negative eigenvalues $0 \leqslant \lambda_1 \leqslant \cdots \leqslant \lambda_r$, where $r = rank(L_H)$. (b) The i-th finite eigenvalue λ_i has the following variational characterization*

$$\lambda_i = \max_{\substack{\mathcal{X} \subseteq \mathbb{R}^n \\ \dim(\mathcal{X}) = n+1-i}} \min_{\substack{x \in \mathcal{X} \\ x^T L_H x > 0}} \frac{x^T L_G x}{x^T L_H x}. \tag{17}$$

In particular,

$$\lambda_1 = \min_{\substack{x \in \mathbb{R}^n \\ x^T L_H x > 0}} \frac{x^T L_G x}{x^T L_H x}. \tag{18}$$

Proof. See Appendix A. □

By Theorem 1(a), we know that if L_H is a nonzero matrix, then the pencil $L_G - \lambda L_H$ has finite non-negative eigenvalues. By Theorem 1(b), we know that the infimum "inf" in (16) is obtainable and can be replaced by "min". In addition, the minimum is reached by the eigenvector corresponding to the *smallest finite* eigenvalue of the generalized eigenproblem

$$L_G x = \lambda L_H x. \tag{19}$$

In general, for a k-way constrained spectral partitioning, the computational kernel is to compute k eigenvectors corresponding to the k smallest finite eigenvalues of (19).

We note that the pencil $L_G - \lambda L_H$ is a special case of a general class of positive semi-definite pencils. The variational principles of general positive semi-definite pencils [Liang et al., 2013][Liang and Li, 2014] have similar forms (17) and (18), where the "min" and "max" are replaced by "inf" (infimum) and "sup" (supremum), respectively. With a detailed analysis, here we show that when both L_G and L_H are positive semi-definite, there have minimum and maximum, and the optimization problem (16) is equivalent to the problem of computing the eigenvectors corresponding to the k smallest finite eigenvalues of (19).

Now let us turn to the problem of solving the eigenproblem (19). If L_H is positive definite, then the eigenproblem (19) is a well-studied generalized symmetric definite eigenproblem. There exist a number of highly efficient solvers and software packages, such as Lanczos method in ARPACK [Lehoucq et al., 1998], and LOBPCG [Knyazev, 2001]. However, the eigenproblem (19) is defined by two positive semi-definite matrices L_G and L_H, and the pencil $L_G - \lambda L_H$ is singular. These solvers are not applicable. To address this computational issue, we propose a regularization scheme to transform (19) to a generalized symmetric definite eigenproblem while maintaining the sparsity of L_G and L_H.

Theorem 2. *Suppose $L_G - \lambda L_H$ has the finite eigenvalues $\lambda_1 \leqslant \cdots \leqslant \lambda_r$, where $r = rank(L_H)$. Let*

$$K = -L_H \quad and \quad M = L_G + \mu L_H + ZSZ^T, \quad (20)$$

where $Z \in \mathbb{R}^{n\times s}$ is an orthonormal basis of the common nullspace of L_G and L_H. $S \in \mathbb{R}^{s\times s}$ is an arbitrary positive definite matrix, and μ is a positive scalar. Then (a) M is positive definite. (b) The eigenvalues of $K - \lambda M$ are $\sigma_1 \leqslant \cdots \leqslant \sigma_r < \sigma_{r+1} = \cdots = \sigma_n = 0$, where $\sigma_i = -1/(\lambda_i + \mu)$ for $i = 1, 2, \ldots, r$.

Proof. See Appendix B. □

The matrix transformation (20) plays the role of *regularization* of the singular pencil $L_G - \lambda L_H$. By Theorem 2(a), the first k smallest finite eigenvalues of (19) can be obtained by computing the k smallest eigenvalues of the generalized symmetric definite eigenproblem

$$Kx = \sigma Mx. \quad (21)$$

By Theorem 2(b), the symmetric definite pencil $K - \lambda M$ also plays the role of the *spectral enhancement* as shown in Fig. 1. The desired smallest eigenvalues, say λ_1 and λ_2 of $L_G - \lambda L_H$ are mapped to the largest eigenvalues σ_1 and σ_2 in absolute values of $K - \sigma M$. The gap between λ_1 and λ_2 could be significantly amplified. As a result, the new pencil $K - \sigma M$ has an eigenvalue distribution with much better separation properties than the original pencil $L_G - \lambda L_H$ and requires far less iterations to convergence in an eigensolver. Numerical examples are provided in Sec. 6.3. This *spectral enhancement* is similar to the shift-and-invert spectral transformation widely used in large-scale eigensolvers [Ericsson and Ruhe, 1980, Parlett, 1998]. However, here, the proposed spectral enhancement is *inverse-free*! Finally, we note that since $M \succ 0$, we can use an approximation of M^{-1} as a preconditioner in a preconditioned eigensolver. We will discuss this in Sec. 5.

5 FAST-GE-2.0

An algorithm, referred to as FAST-GE-2.0, for a k-way partition of a given data graph $G_D = (V, W_D)$ with constraint subsets $\{V_1, V_2, \ldots, V_k\}$ is summarized in Alg. 1. FAST-GE-2.0 is a modified version of FAST-GE and is supported by rigorous mathematical theory (Theorem 1) and an effective regularization and spectral enhancement (Theorem 2).

A few remarks are in order:

1. At line 3, since the pencil (K, M) is symmetric definite, we can use any available eigensolver for the generalized symmetric definite eigenproblem. In this paper, we use LOBPCG [Knyazev, 2001].

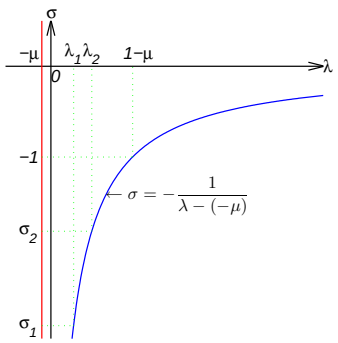

Figure 1: Spectral Transformation

Algorithm 1 FAST-GE-2.0

Input:
Data graph $G_D = (V, W_D)$, constraint sets $\{V_1, \ldots, V_k\}$, and μ, Z and S for the regularlization (20)

Output:
Indicator $c \in \mathbb{R}^n$, with $c_j \in \{1, \ldots, k\}$.

1: construct $G_M = (V, W_M)$ and $G_H = (V, W_H)$ by (9) and (10);
2: compute Laplacians L_G and L_H of $G = (V, W_D + W_M)$ and $H = G_H$ respectively;
3: compute eigenvectors $X = [x_1, \ldots, x_k]$ of the regularized pencil $K - \sigma M$ in (20);
4: renormalize X to Y;
5: $c = \texttt{k-means}(Y, k)$.

For the matrix-vector products Kv and Mv, since both K and M are sparse plus low-rank, we can apply an efficient communication-avoiding algorithm [Knight et al., 2013] for the products. The proper basis selection and maintenance of orthogonality are crucial and costly in LOBPCG [Hetmaniuk and Lehoucq, 2006]. One may exploit some kind of approximations, such as Nyström aprximation [Gisbrecht et al., 2010]. The efficiency of LOBPCG strongly depends on the choice of a preconditioner. A natural preconditioner here is $T \approx M^{-1}$. Applying the preconditioner T is equivalent to solve the block linear system $MW = R$ for W approximately. We can use the preconditioned conjugate gradient (PCG) with the Jacobi precoditioner, namely, the diagonal matrix of M. PCG will be referred to as "inner iterations" in contrast to the outer LOBPCG iterations. Numerical results in Sec. 6.3 show that a small number (2 to 4) of inner iterations is sufficient.

2. The renormalization of X before k-means clustering at line 4 consists of the following steps:
(a) $D_c = \text{diag}(\|x_1\|, \ldots, \|x_k\|)$;

(b) $\widehat{X} = XD_c^{-1}$;
(c) $\widehat{Y} = \widehat{X}^T$;
(d) $D_r = \text{diag}(\|\widehat{y}_1\|, \ldots, \|\widehat{y}_n\|)$;
(e) $Y = D_r^{-1}\widehat{X}$.
Step (b) normalizes the columns of X. Since x_i is M-orthogonal, the values of $\|x_i\|_2$ for different i are not the same. This normalization balances the absolute values of the elements of eigenvectors. Step (e) normalizes the rows of the eigenvectors. It balances the absolute values of the elements of rows, as points in high-dimensional space. This is also a main step in the Spectral Clustering [Ng et al., 2002]. The effect of the renormalization will be discussed in Sec. 6.2.

3. After the renormalization step, any geometric partitioning algorithm can be applied in line 5. Here we use the k-means method to cluster the rows of Y into k disjointed sets. If $c_i = j$, then the vertex v_i is in the cluster A_j.

6 EXPERIMENTS

In this section, we present the experimental results of FAST-GE-2.0 and provide empirical study on the performance tuning. Alg. 1 is implemented in MATLAB. The eigensolver LOBPCG is obtained from MathWorks File Exchange[1]. In the spirit of reproducible research, the scripts of the implementation of FAST-GE-2.0 and the data that used to generate experimental results presented in this paper can be obtained from the URL `https://github.com/aistats2017239/fastge2`.

For all experimental results, unless otherwise stated, we use $\mu = 10^{-3}$ and $S = I$ for defining the pencil $K - \sigma M$. By the definitions of L_G and L_H, $Z = \mathbf{1}$ is a basis of the common nullspace of L_G and L_H. The blocksize of LOBPCG is chosen to be the number of clusters k and the stopping criterion $\varepsilon = 10^{-4}$. For the inner iteration to apply the preconditioner $T \approx M^{-1}$, the residual tolerance $tol\_in = 10^{-4}$ and the maximum number of inner iterations $maxit\_in = 4$ by default. The justification for the choices of μ and $maxit\_in$ is in Sec. 6.3. All experiements were run on a machine with Intel(R) Core(TM) i7-3612QM CPU@2.10 GHz and 6GB RAM.

6.1 Synthetic Data

This simple synthetic example is to show that if LOBPCG is applied to compute smallest eigenvalue of $L_G - \lambda L_H$ for the indicator vector x as suggested in FAST-GE, it could terminate prematurely. As in [Chew and Cahill, 2015], we generate sets S_1, S_2 and S_3 with 100 points each, see Fig. 2(a). For constrained clustering, we add a pair of ML constraints in S_1 and S_3, shown as red dots in Fig. 2(a). In addition, we add one pair of ML constraints in S_2, shown as the green dots in Fig. 2(a). We observed that if LOBPCG is used to compute the smallest eigenpair of $\{L_G, L_H\}$, it terminates at the first iteration, with the error "*the residual is not full rank or/and operatorB is not positive definite*". Clearly, this is due to the singularity of L_H. In contrast, when LOBPCG is applied for the regularized pencil $K - \sigma M$, it converges successfully with the output indicator vector x_1 shown in Fig. 2(b). Consequently, the vector x_1 leads to a successful desired 2-way constrained partition $(S_1 \cup S_3, S_2)$ shown in Fig. 2(c).

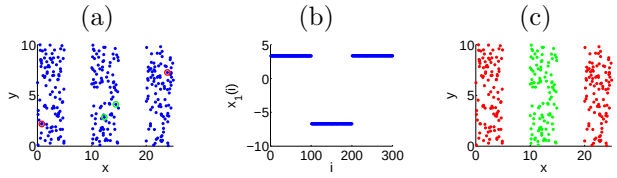

Figure 2: Synthetic data (a) Data and Constraints, (b) Indicator Vector x_1, (c) Constrianed Clustering by FAST-GE-2.0.

6.2 Constrained Image Segmentations

A comparison of the quality of the constrained image segmentation of FAST-GE with COSf and CSP is reported [Cucuringu et al., 2016]. Here we focus on the comparison of FAST-GE and proposed FAST-GE-2.0. Columns in Fig. 3 are in the order of original images, images with constraints, the heatmap of eigenvalue x_1, the heatmap of the renormalized vector y_1, and the image segmentation by FAST-GE-2.0. The related data are in Table 1.

Few comments are in order. (i) for these large size images, the number of constraint points m, is relatively small, see m-column in Table 1. (ii) After the renormalization, y_1 has much sharper difference than the indicator vector x_1, see in the heatmaps of x_1 and y_1. (iii) The results of constrained image segmentation successfully satisfy ML and CL constraints. They clearly separate the object and background in the 2-way clustering (rows 1 to 4) and different regions of images in the k-way clustering (rows 5 to 7). (iv) The total CPU running time is shown in the last t_t-column of Table 1. As we can see that nearly 80% to 90% of the running time is on solving the eigenproblem, t_e-column. *The eigenproblem is the computational bottleneck!* We note that the reported CPU running time for the 5-way segmentation of "Patras" image by FAST-GE is under 3 seconds. Unfortunately, the computer

[1] https://www.mathworks.com/matlabcentral/fileexchange/48-lobpcg-m

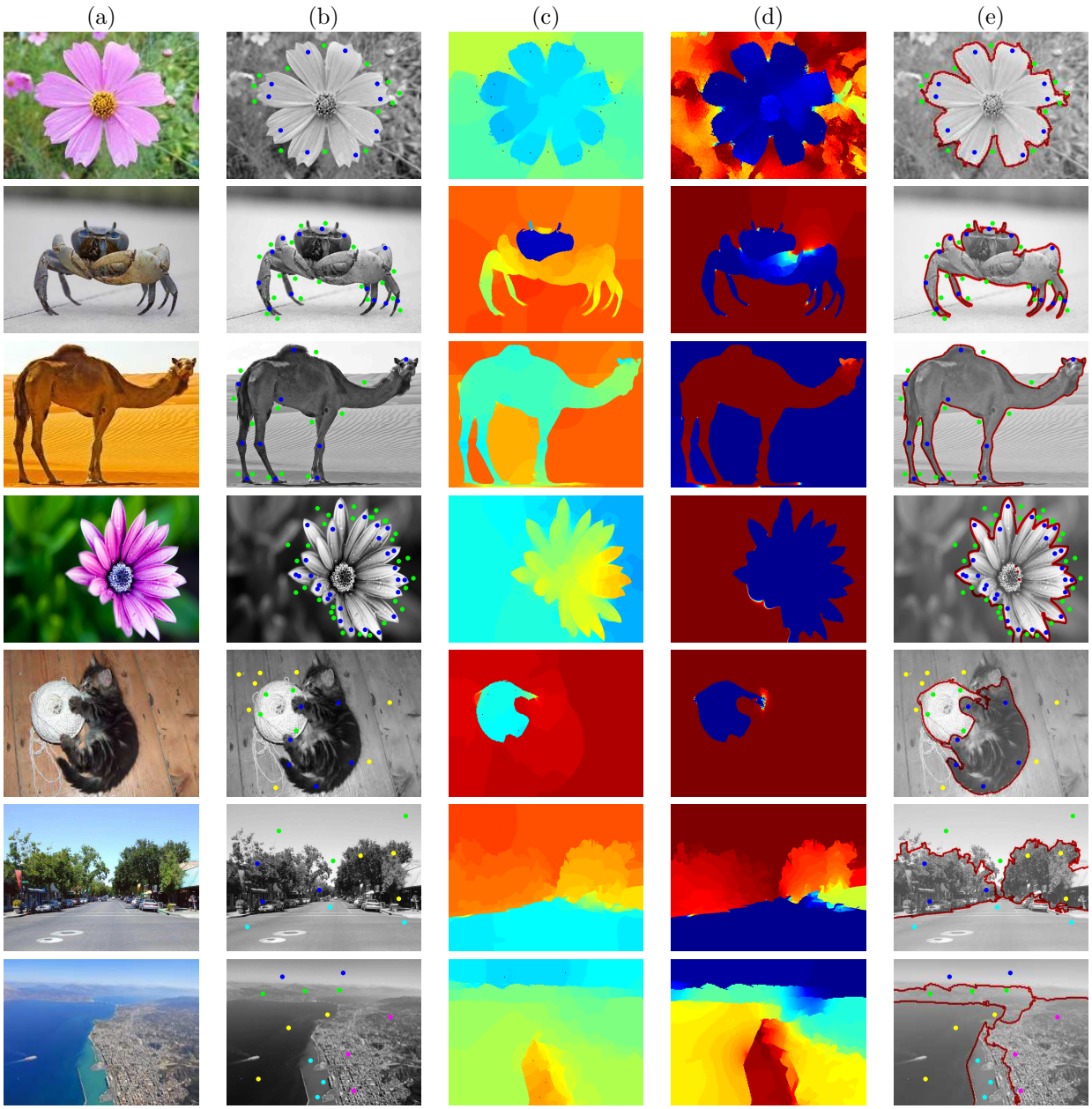

Figure 3: (a) Original Images, (b) Constraints, (c) Indicator Vector x_1, (d) Renormalized Indicator Vector y_1, (e) Constrained Segmentation by FAST-GE-2.0.

platform and stopping threshold ε of the eigensolver were not reported. We observed that a larger stopping threshold ε will significantly decrease the running time t_e of FAST-GE-2.0. However, for accuracy, we have used a relatively small threshold $\varepsilon = 10^{-4}$ throughout our experiments.

The effectiveness of the preconditioner $T = M^{-1}$ reflects in the reduction of the running time. For example, in the 2-way partitioning of "Flower" image, it took 3227 LOBPCG iterations and 52.04 seconds if the preconditioner is not used. With the precondi-

Table 1: Problem Size and Running Time for Fig. 3

Image	Pixels n	k	m	t_e(sec)	t_t (sec)
Flower	30,000	2	23	5.88	7.14
Crab	143,000	2	32	55.71	62.14
Camel	240,057	2	23	144.94	159.67
Daisy	1,024,000	2	58	1518.88	1677.51
Cat	50,325	3	18	12.45	15.16
Davis	235,200	4	12	77.90	89.75
Patras	44,589	5	14	11.81	13.72

Table 2: Effect of Shift μ in Spectral Transformation

μ	σ_1	σ_2	σ_3	*iter*	t (sec)
10	-0.099	-0.099	-0.099	223	11.78
1	-0.993	-0.990	-0.907	148	7.61
10^{-1}	-9.411	-9.080	-4.949	121	6.85
10^{-2}	-61.538	-49.678	-8.924	72	5.90
10^{-3}	-137.928	-89.850	-9.703	71	5.89
10^{-4}	-157.477	-97.755	-9.789	69	5.68
10^{-5}	-159.741	-98.622	-9.797	69	5.88

tioiner T, it reduced to 88 LOBPCG iterations and 5.88 seconds. This is about a factor of 9 speedup. Similarly, for the 5-way partitioning of "Patras" image, LOBPCG without preconditioning took 2032 iterations and 101.28 seconds. In constrast, with the preconditioner T, it reduced to 61 iterations and 11.81 seconds. This is again about a factor of 9 speedup.

6.3 Performance Tuning

Parameter Tuning of μ. The shift μ is a key parameter in the matrix transformation (20) that directly affects the efficiency of an eigensolver. To numerically study the effect of μ, we use "Flower" image (200×150) as an example. We compute the three largest eigenvalues in absolute values $\{\sigma_i\}$ of $K-\sigma M$ to obtain the three smallest finite eigenvalues $\{\lambda_i\}$ of $L_G-\lambda L_H$ by $\lambda_i=-1/\sigma_i-\mu$ for $i=1,2,3$. Table 2 shows the transformed eigenvalues σ_i, the number of LOBPCG iterations (*iter*) and the running time of LOBPCG (t), with respect to different values of μ, where the same random initial vector X_0 of LOBPCG is used for all μ's. By Table 2, we see that initially decreasing the value of μ reduces the number of LOBPCG iterations (*iter*) and runtime (t). This is due to the widen gaps between the desired eigenvalues. However, after a certain point, the benefit saturates. Therefore, we suggest $\mu=10^{-2}\sim10^{-4}$ as a default value.

Inner-Outer Iterations. We have used the maximum number of inner PCG iteration $maxit\_in=4$ for applying the preconditioner $T\approx M^{-1}$ in Secs. 6.1 and 6.2. One naturally asks whether by increasing $maxit\_in$, it would take less number of outer LOBPCG iterations and reduce the running time. We tested the trade-off of inner-outer iterations. The threshold for the inner PCG iterations is $tol\_in=10^{-4}$. We note that in our experiments, the inner PCG never reaches tol_in even with $maxit\_in=20$. Fig. 4 shows the number of outer iterations, the running time of inner and outer iterations with respect to the different number of $maxit\_in$ on an image of size $n=38,400$. We see that as the number of inner iterations $maxit\_in$ increases, the number of outer LOBPCG iterations decreases (blue line). But the total CPU time (green line) increases due to the growing cost of applying the preconditioning (red line). The tradeoff suggests that a small $maxit\_in=2\sim4$ achieves the best of overall performance.

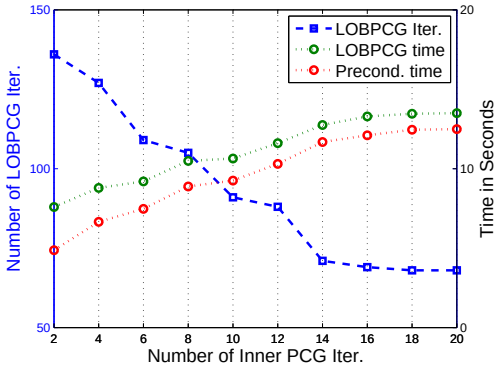

Figure 4: The Tradeoff between the Numbers of Inner PCG Iterations for Applying Preconditioner and the Numbers of Outer LOBPCG Iterations for Finding Eigenvalues.

7 CONCLUDING REMARKS

The FAST-GE-2.0 algorithm is a modified version of FAST-GE [Cucuringu et al., 2016] and is established on a solid mathematical foundation of an extended Courant-Fischer variational principle. The matrix transformation from the positive semi-definite pencil $L_G-\lambda L_H$ to the positive definite pencil $K-\sigma M$ provides computational robustness and efficiency of FAST-GE-2.0. The eigensolver is the computational bottleneck. Our future study includes further exploiting structures of the underlying sparse plus low rank structure of Laplacians for high performance computing. In addition, we plan to investigate applying and developing extended Courant-Fischer variational principle and the regularization scheme to large scale eigenvalue problems arising from other applications in machine learning, such as generalized linear discriminant analysis in dimension reduction [He et al., 2005, Park and Park, 2008, Zhu and Huang, 2014] and multisurface classification via generalized eigenvectors [Mangasarian and Wild, 2006].

Acknowledgements

The research of Jiang and Bai was supported in part by the NSF grants DMS-1522697 and CCF-1527091. Part of this work was done while Xie was visiting the University of California, Davis, supported by Shanghai Natural Science Fund (No.15ZR1408400), China.

References

[Chapelle et al., 2006] Chapelle, O., Scholkopf, B., and Zien, A. (2006). *Semi-Supervised Learning*. MIT Press, London, England.

[Chew and Cahill, 2015] Chew, S. E. and Cahill, N. D. (2015). Semi-Supervised Normalized Cuts for image segmentation. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1716–1723.

[Cour et al., 2007] Cour, T., Srinivasan, P., and Shi, J. (2007). Balanced graph matching. *Advances in Neural Information Processing Systems*, 19:313–320.

[Cucuringu et al., 2016] Cucuringu, M., Koutis, I., Chawla, S., Miller, G., and Peng, R. (2016). Simple and scalable constrained clustering: A generalized spectral method. In *Proceedings of International Conference on Artificial Intelligence and Statistics*, pages 445–454.

[Eaton and Mansbach, 2012] Eaton, E. and Mansbach, R. (2012). A spin-glass model for semi-supervised community detection. In *In Proceedings of AAAI Conference on Artificial Intelligence*, pages 900–906.

[Ericsson and Ruhe, 1980] Ericsson, T. and Ruhe, A. (1980). The spectral transformation Lanczos method for the numerical solution of large sparse generalized symmetric eigenvalue problems. *Mathematics of Computation*, 35:1251–1268.

[Eriksson et al., 2011] Eriksson, A., Olsson, C., and Kahl, F. (2011). Normalized cuts revisited: A reformulation for segmentation with linear grouping constraints. *Journal of Mathematical Imaging and Vision*, 39:45–61.

[Gallier, 2013] Gallier, J. (2013). Notes on elementary spectral graph theory. Applications to graph clustering using normalized cuts. *arXiv preprint arXiv:1311.2492*.

[Gisbrecht et al., 2010] Gisbrecht, A., Mokbel, B., and Hammer, B. (2010). The Nyström approximation for relational generative topographic mappings. In *NIPS Workshop on Challenges of Data Visualization*.

[Golub and Van Loan, 2012] Golub, G. H. and Van Loan, C. F. (2012). *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, 4 edition.

[He et al., 2005] He, X., Cai, D., Yan, S., and Zhang, H.-J. (2005). Neighborhood preserving embedding. In *Proceedings of IEEE International Conference on Computer Vision*, volume 2, pages 1208–1213.

[Hetmaniuk and Lehoucq, 2006] Hetmaniuk, U. and Lehoucq, R. (2006). Basis selection in lobpcg. *Journal of Computational Physics*, 218:324–332.

[Joachims, 2003] Joachims, T. (2003). Transductive learning via spectral graph partitioning. In *Proceedings of Internaitonl Conference on Machine Learning*, volume 3, pages 290–297.

[Knight et al., 2013] Knight, N., Carson, E., and Demmel, J. (2013). Exploiting data sparsity in parallel matrix powers computations. In *Proceedings of International Conference on Parallel Processing and Applied Mathematics*, pages 15–25.

[Knyazev, 2001] Knyazev, A. V. (2001). Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method. *SIAM Journal on Scientific Computing*, 23:517–541.

[Lehoucq et al., 1998] Lehoucq, R. B., Sorensen, D. C., and Yang, C. (1998). *ARPACK Users' Guide: Solution of Large-scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. SIAM, Philadelphia,PA.

[Liang and Li, 2014] Liang, X. and Li, R.-C. (2014). Extensions of Wielandts min–max principles for positive semi-definite pencils. *Linear and Multilinear Algebra*, 62:1032–1048.

[Liang et al., 2013] Liang, X., Li, R.-C., and Bai, Z. (2013). Trace minimization principles for positive semi-definite pencils. *Linear Algebra and its Applications*, 438:3085–3106.

[Ma et al., 2010] Ma, X., Gao, L., Yong, X., and Fu, L. (2010). Semi-supervised clustering algorithm for community structure detection in complex networks. *Physica A: Statistical Mechanics and its Applications*, 389:187–197.

[Mangasarian and Wild, 2006] Mangasarian, O. L. and Wild, E. W. (2006). Multisurface proximal support vector machine classification via generalized eigenvalues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):69–74.

[Ng et al., 2002] Ng, A. Y., Jordan, M. I., and Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, 2:849–856.

[Park and Park, 2008] Park, C. H. and Park, H. (2008). A comparison of generalized linear discriminant analysis algorithms. *Pattern Recognition*, 41:1083–1097.

[Parlett, 1998] Parlett, B. N. (1998). *The symmetric eigenvalue problem*. SIAM.

[Rangapuram and Hein, 2012] Rangapuram, S. S. and Hein, M. (2012). Constrained 1-spectral clustering. In *Proceedings of International Conference on Artificial Intelligence and Statistics*, volume 30, pages 90–102.

[Shi and Malik, 2000] Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:888–905.

[Wagstaff et al., 2001] Wagstaff, K., Cardie, C., Rogers, S., and Schrödl, S. (2001). Constrained k-means clustering with background knowledge. In *Proceedings of Internaitonl Conference on Machine Learning*, volume 1, pages 577–584.

[Wang et al., 2014] Wang, X., Qian, B., and Davidson, I. (2014). On constrained spectral clustering and its applications. *Data Mining and Knowledge Discovery*, 28:1–30.

[Xu et al., 2009] Xu, L., Li, W., and Schuurmans, D. (2009). Fast normalized cut with linear constraints. In *Proceedings of Computer Vision and Pattern Recognition*, pages 2866–2873.

[Yu and Shi, 2004] Yu, S. X. and Shi, J. (2004). Segmentation given partial grouping constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:173–183.

[Zhu and Huang, 2014] Zhu, L. and Huang, D.-S. (2014). A Rayleigh–Ritz style method for large-scale discriminant analysis. *Pattern Recognition*, 47:1698–1708.

Appendix A. Proof of Theorem 1

In this appendix, we first derive a canonical form of the pencil $L_G - \lambda L_H$, and then prove the variational principle in Theorem 1. For the simplicity of notation, in this appendix, we denote $A = L_G$ and $B = L_H$. We begin with the following lemma.

Lemma 1. *If $A - \lambda B$ is a symmetric matrix pencil of order n with $A \succeq 0$ and $B \succeq 0$, then there exists an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ such that*

$$Q^T A Q = \begin{array}{c} \\ r \\ n_1 \\ m \end{array} \begin{array}{c} \begin{array}{ccc} r & n_1 & m \end{array} \\ \begin{bmatrix} \widehat{A}_{11} & \widehat{A}_{12} & \\ \widehat{A}_{12}^T & \widehat{A}_{22} & \\ & & 0 \end{bmatrix} \end{array} \equiv \begin{array}{c} \\ r+n_1 \\ m \end{array} \begin{array}{c} \begin{array}{cc} r+n_1 & m \end{array} \\ \begin{bmatrix} \widehat{A} & \\ & 0 \end{bmatrix} \end{array}, \tag{1}$$

$$Q^T B Q = \begin{array}{c} \\ r \\ n_1 \\ m \end{array} \begin{array}{c} \begin{array}{ccc} r & n_1 & m \end{array} \\ \begin{bmatrix} \widehat{B}_{11} & & \\ & 0 & \\ & & 0 \end{bmatrix} \end{array} \equiv \begin{array}{c} \\ r+n_1 \\ m \end{array} \begin{array}{c} \begin{array}{cc} r+n_1 & m \end{array} \\ \begin{bmatrix} \widehat{B} & \\ & 0 \end{bmatrix} \end{array}, \tag{2}$$

where $\widehat{A}_{22} \succ 0$ and $\widehat{B}_{11} \succ 0$. Furthermore, the sub-pencil $\widehat{A} - \lambda\widehat{B}$ is regular and $\widehat{A} \succeq 0$ and $\widehat{B} \succeq 0$.

Proof. Since $B \succeq 0$, there exists an orthogonal matrix $Q_1 \in \mathbb{R}^{n\times n}$ such that

$$B^{(0)} \equiv Q_1^T B Q_1 = \begin{array}{c} \\ r \\ d \end{array} \begin{array}{c} \begin{array}{cc} r & d \end{array} \\ \begin{bmatrix} \widehat{B}_{11} & \\ & 0 \end{bmatrix} \end{array}, \tag{3}$$

where $\widehat{B}_{11} \succ 0$. Applying transformation Q_1 to matrix A, we have

$$A^{(0)} \equiv Q_1^T A Q_1 = \begin{array}{c} \\ r \\ d \end{array} \begin{array}{c} \begin{array}{cc} r & d \end{array} \\ \begin{bmatrix} \widehat{A}_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix} \end{array}.$$

Note that $A_{22} \succeq 0$ due to the fact that $A \succeq 0$.

For the $d \times d$ block matrix A_{22}, there exists an orthogonal matrix $Q_{22} \in \mathbb{R}^{d\times d}$ such that

$$Q_{22}^T A_{22} Q_{22} = \begin{array}{c} \\ n_1 \\ m \end{array} \begin{array}{c} \begin{array}{cc} n_1 & m \end{array} \\ \begin{bmatrix} \widehat{A}_{22} & \\ & 0 \end{bmatrix} \end{array},$$

where $\widehat{A}_{22} \succ 0$.

Let $Q_2 = \mathrm{diag}(I_r, Q_{22})$. Then we have

$$A^{(1)} \equiv Q_2^T A^{(0)} Q_2 = \begin{array}{c} \\ r \\ n_1 \\ m \end{array} \begin{array}{c} \begin{array}{ccc} r & n_1 & m \end{array} \\ \begin{bmatrix} \widehat{A}_{11} & \widehat{A}_{12} & \widehat{A}_{13} \\ \widehat{A}_{12}^T & \widehat{A}_{22} & \\ \widehat{A}_{13}^T & & 0 \end{bmatrix} \end{array},$$

$$B^{(1)} \equiv Q_2^T B^{(0)} Q_2 = \begin{array}{c} \\ r \\ n_1 \\ m \end{array} \begin{array}{c} \begin{array}{ccc} r & n_1 & m \end{array} \\ \begin{bmatrix} \widehat{B}_{11} & & \\ & 0 & \\ & & 0 \end{bmatrix} \end{array},$$

where $\begin{bmatrix} \widehat{A}_{12}, & \widehat{A}_{13} \end{bmatrix} = A_{12}Q_{22}$. Note that since $A^{(1)} \succeq 0$, we must have $\widehat{A}_{13} = 0$. Otherwise, if there exists an element $a_{ij} \neq 0$ in $\widehat{A}_{13}$, then the 2 by 2 sub-matrix $\begin{bmatrix} \widehat{a}_{ii} & a_{ij} \\ a_{ij} & 0 \end{bmatrix}$ of $A^{(1)}$ is indefinite, where $\widehat{a}_{ii}$ is the i-th diagonal element of $\widehat{A}_{11}$. This contradicts to the positive semi-definiteness of $A^{(1)} \succeq 0$.

Denote $Q = Q_1 Q_2$. Then Q is orthogonal, and $Q^T A Q,\ Q^T B Q$ have the form (1).

Finally, we show the pencil $\widehat{A} - \lambda \widehat{B}$ is regular. For any $\lambda \in \mathbb{C}$, straightforward calculation gives that

$$\begin{aligned} \det(\widehat{A} - \lambda \widehat{B}) &= \det \begin{pmatrix} \widehat{A}_{11} - \lambda \widehat{B}_{11} & \widehat{A}_{12} \\ \widehat{A}_{12}^T & \widehat{A}_{22} \end{pmatrix} \\ &= \det \begin{pmatrix} \widehat{A}_{11} - \widehat{A}_{12}\widehat{A}_{22}^{-1}\widehat{A}_{12}^T - \lambda \widehat{B}_{11} & \\ & \widehat{A}_{22} \end{pmatrix} \\ &= \det(\widehat{A}_{22}) \det(\widehat{A}_{11} - \widehat{A}_{12}\widehat{A}_{22}^{-1}\widehat{A}_{12}^T - \lambda \widehat{B}_{11}). \end{aligned}$$

Recall that $\widehat{A}_{22} \succ 0$. Furthermore, since $\widehat{B}_{11} \succ 0$, $\det(\widehat{A}_{11} - \widehat{A}_{12}\widehat{A}_{22}^{-1}\widehat{A}_{12}^T - \lambda \widehat{B}_{11}) \not\equiv 0$. Hence, $\det(\widehat{A} - \lambda \widehat{B}) \not\equiv 0$. This means the pencil $\widehat{A} - \lambda \widehat{B}$ is regular. □

By Lemma 1, we have the following canonical form of the matrix pair $\{A, B\}$ to show that the matrices A and B are simultaneously diagonalizable with a congruence transformation.

Lemma 2. *If $A - \lambda B$ is a symmetric matrix pencil of order n with $A \succeq 0$ and $B \succeq 0$, then there exists a nonsingular matrix $X \in \mathbb{R}^{n\times n}$ such that*

$$X^T A X = \begin{array}{c} \\ r \\ n_1 \\ m \end{array} \begin{array}{c} \begin{array}{ccc} r & n_1 & m \end{array} \\ \begin{bmatrix} \Lambda_r & & \\ & I & \\ & & 0 \end{bmatrix} \end{array}, \quad X^T B X = \begin{array}{c} \\ r \\ n_1 \\ m \end{array} \begin{array}{c} \begin{array}{ccc} r & n_1 & m \end{array} \\ \begin{bmatrix} I & & \\ & 0 & \\ & & 0 \end{bmatrix} \end{array}, \tag{4}$$

where Λ_r is a diagonal matrix of non-negative diagonal elements $\lambda_1, \ldots, \lambda_r$, $r = \operatorname{rank}(B)$, $m = \dim(\mathcal{N}(A) \cap \mathcal{N}(B))$ and $n_1 = \dim(\mathcal{N}(B)) - m$.

Proof. By Lemma 1, there exists an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ such that

$$A^{(1)} \equiv Q^T A Q = \begin{array}{c} \\ r \\ n_1 \\ m \end{array} \begin{array}{c} \begin{array}{ccc} r & n_1 & m \end{array} \\ \begin{bmatrix} \widehat{A}_{11} & \widehat{A}_{12} & \\ \widehat{A}_{12}^T & \widehat{A}_{22} & \\ & & 0 \end{bmatrix} \end{array} \quad \text{and} \quad B^{(1)} \equiv Q^T B Q = \begin{array}{c} \\ r \\ n_1 \\ m \end{array} \begin{array}{c} \begin{array}{ccc} r & n_1 & m \end{array} \\ \begin{bmatrix} \widehat{B}_{11} & & \\ & 0 & \\ & & 0 \end{bmatrix} \end{array}.$$

Let

$$X_1 = \begin{bmatrix} I_r & & \\ -\widehat{A}_{22}^{-1}\widehat{A}_{12}^T & \widehat{A}_{22}^{-1/2} & \\ & & I_s \end{bmatrix}.$$

Then

$$A^{(2)} \equiv X_1^T A^{(1)} X_1 = \begin{bmatrix} \widehat{A}_{11} - \widehat{A}_{12}^{(1)}\widehat{A}_{22}^{-1}\widehat{A}_{12}^T & & \\ & I_{n_1} & \\ & & 0_s \end{bmatrix} \quad \text{and} \quad B^{(2)} \equiv X_1^T B^{(1)} X_1 = \begin{bmatrix} \widehat{B}_{11} & & \\ & 0_{n_1} & \\ & & 0_s \end{bmatrix}.$$

Since $\widehat{B}_{11} \succ 0$, there exists a nonsingular matrix $\widehat{X}_2$ such that

$$\widehat{X}_2^T [\widehat{A}_{11} - \widehat{A}_{12}^{(1)}\widehat{A}_{22}^{-1}\widehat{A}_{12}^T]\widehat{X}_2 = \Lambda,\ \widehat{X}_2^T \widehat{B}_{11} \widehat{X}_2 = I_r.$$

Let $X_2 = \operatorname{diag}(\widehat{X}_2, I_{n_1}, I_s)$. Then we have

$$X_2^T A^{(2)} X_2 = \operatorname{diag}(\Lambda, I_{n_1}, 0_s),\ X_2^T B^{(2)} X_2 = \operatorname{diag}(I_r,\ 0_{n_1},\ 0_s).$$

Denote $X = Q X_1 X_2$. Then we obtain (4). The remaining results are eaily obtained from the canonical form (2). □

The following remarks are in order:

1. By Lemma 2, we know (i) there are $r = \text{rank}(B)$ finite eigenvalues of the pencil $A - \lambda B$ and all finite eigenvalues are real, nonnegative and non-defective. and (ii) there are $n_1 = \dim(\mathcal{N}(B)) - \dim(\mathcal{N}(A) \cap \mathcal{N}(B))$ non-defective infinite eigenvalues.

2. The canonical form (4) has been derived in [Newcomb, 1961]. Here we give the values of indices r, n_1, m in (4) and our proof seems more compact.

3. Lemma 3.8 in [Liang et al., 2013] deals with the canonical form of a general positive semi-definite pencil. Obviously, the pencil $A-\lambda B$ considered here is a special case of positive semi-definite pencil. So Lemma 3.8 is applicable here. Our proof is constructive based on Fix-Heiberger's reduction [Fix and Heiberger, 1972].

We now provide a proof of the variational principle in Theorem 1. Without loss of generality, we assume that pencil $A - \lambda B$ is in the canonical form (4), i.e.,

$$A = \begin{array}{c} \\ r \\ n_1 \\ m \end{array} \begin{array}{c} \begin{array}{ccc} r & n_1 & m \end{array} \\ \begin{bmatrix} \Lambda_r & & \\ & I & \\ & & 0 \end{bmatrix} \end{array}, \quad B = \begin{array}{c} \\ r \\ n_1 \\ m \end{array} \begin{array}{c} \begin{array}{ccc} r & n_1 & m \end{array} \\ \begin{bmatrix} I & & \\ & 0 & \\ & & 0 \end{bmatrix} \end{array}. \tag{5}$$

Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a subspace of dimension $n+1-i$, where $1 \leqslant i \leqslant r$ and $x \in \mathcal{X}$ be partitioned into $x = [x_1^T, x_2^T, x_3^T]^T$ conformally with the form (5), then

$$\inf_{\substack{x \in \mathcal{X} \\ x^T B x > 0}} \frac{x^T A x}{x^T B x} = \inf_{\substack{x \in \mathcal{X} \\ x_1^T x_1 > 0}} \frac{x_1^T \Lambda_r x_1 + x_2^T x_2}{x_1^T x_1} = \inf_{\substack{x \in \mathcal{X} \\ x_1^T x_1 > 0}} \frac{x_1^T \Lambda_r x_1}{x_1^T x_1}. \tag{6}$$

Let $\mathcal{X}^{(1)} = \{[I_r, 0_{n-r}]x \mid x \in \mathcal{X}\}$. Evidently, $\mathcal{X}^{(1)}$ is a subspace of $\mathbb{R}^r$. Moreover,

$$n + 1 - i \geqslant \dim(\mathcal{X}^{(1)}) \geqslant n + 1 - i - n_1 - s = r + 1 - i.$$

Then there exists a subspace $\widetilde{\mathcal{X}} \subseteq \mathbb{R}^r$ of dimension $r + 1 - i$ such that $\widetilde{\mathcal{X}} \subseteq \mathcal{X}^{(1)}$. For the matrix Λ_r, by Courant-Fischer min-max principle, we have

$$\inf_{\substack{x \in \mathcal{X} \\ x_1^T x_1 > 0}} \frac{x_1^T \Lambda_r x_1}{x_1^T x_1} = \min_{\substack{x_1 \in \mathcal{X}^{(1)} \\ x_1^T x_1 > 0}} \frac{x_1^T \Lambda_r x_1}{x_1^T x_1} \leqslant \min_{\substack{x_1 \in \widetilde{\mathcal{X}} \\ x_1^T x_1 > 0}} \frac{x_1^T \Lambda_r x_1}{x_1^T x_1} \leqslant \max_{\substack{\dim(\mathcal{S}) = r+1-i \\ \mathcal{S} \subseteq \mathbb{R}^r}} \min_{\substack{x_1 \in \mathcal{S} \\ x_1^T x_1 > 0}} \frac{x_1^T \Lambda_r x_1}{x_1^T x_1} = \lambda_i.$$

Combining above equation with (6), we know that for any subspace $\mathcal{X} \subseteq \mathbb{R}^n$ with dimension $n + 1 - i$,

$$\min_{\substack{x \in \mathcal{X} \\ x^T B x > 0}} \frac{x^T A x}{x^T B x} \leqslant \lambda_i. \tag{7}$$

On the other hand, let us consider a special choice of the subspace $\mathcal{X}$:

$$\mathcal{S}_i = \mathcal{R}(S_i),$$

where

$$S_i = \begin{array}{c} \\ i-1 \\ r+1-i \\ n-r \end{array} \begin{array}{c} \begin{array}{cc} r+1-i & n-r \end{array} \\ \begin{bmatrix} 0 & \\ I & 0 \\ & I \end{bmatrix} \end{array}.$$

Then $\dim(\mathcal{S}_i) = n + 1 - i$, and

$$S_i^T A S_i = \text{diag}(\widetilde{\Lambda}_i,\ I_{n_1},\ 0_s), \quad S_i^T B S_i = \text{diag}(I_{r+1-i},\ 0_{n_1},\ 0_s),$$

where $\widetilde{\Lambda}_i = \text{diag}(\lambda_i, \cdots, \lambda_r)$. Let $x_* = S_i e_1 \in \mathcal{S}_i$, where e_1 is a unit vector of dimension $n + r - i$, then

$$\frac{x_*^T A x_*}{x_*^T B x_*} = \lambda_i.$$

Consequently, Eq.17 (Sec.4) follows from above equation and (7). Taking $i = 1$ in (7), we get Eq.18 (Sec.4).

Appendix B. Proof of Theorem 2

Similar to Appendix A, for the simplicity of notation, we denote $A = L_G$ and $B = L_H$. By the definitions of K and M in Theorem 2, we have

$$K = -B, \quad M = A + \mu B + ZSZ^T.$$

By Lemma 1, there exists an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ such that

$$Q^T A Q = \begin{array}{c} \\ n-m \\ m \end{array} \begin{array}{c} \begin{array}{cc} n-m & m \end{array} \\ \begin{bmatrix} \widehat{A} & \\ & 0 \end{bmatrix} \end{array}, \quad Q^T B Q = \begin{array}{c} \\ n-m \\ m \end{array} \begin{array}{c} \begin{array}{cc} n-m & m \end{array} \\ \begin{bmatrix} \widehat{B} & \\ & 0 \end{bmatrix} \end{array}, \tag{8}$$

where the $(n-m)\times(n-m)$ sub-pencil $\widehat{A} - \lambda\widehat{B}$ is regular and $\widehat{A} \succeq 0$ and $\widehat{B} \succeq 0$.

Let Q in (8) be conformally partitioned in the form $Q = [Q_1, Q_2]$, where $Q_2 \in \mathbb{R}^{n\times m}$. Then Q_2 is also an orthonormal basis of $\mathcal{N}(A) \cap \mathcal{N}(B)$, i.e.,

$$Z = Q_2 G \tag{9}$$

for some orthogonal matrix G.

For the regular pair $\{\widehat{A}, \widehat{B}\}$, by Lemma 2, there exists a nonsingular matrix $\widetilde{X} \in \mathbb{R}^{(n-m)\times(n-m)}$ such that

$$\widetilde{X}^T \widehat{A} \widetilde{X} = \mathrm{diag}(\Lambda_r, I_{n_1}), \ \widetilde{X}^T \widehat{B} \widetilde{X} = \mathrm{diag}(I_r, 0_{n_1}), \tag{10}$$

where $\Lambda_r = \mathrm{diag}(\lambda_1, \cdots, \lambda_r) \succeq 0$.

Let $X = Q\,\mathrm{diag}(\widetilde{X}, I_m)$. Then

$$\begin{aligned} X^T K X &= \mathrm{diag}(\widetilde{X}^T, I_m) Q^T (-B) Q\, \mathrm{diag}(\widetilde{X}, I_m) \\ &= \mathrm{diag}(\widetilde{X}^T, I_m)\, \mathrm{diag}(-\widehat{B}, 0_m)\, \mathrm{diag}(\widetilde{X}^T, I_m) \qquad \text{by (8)} \\ &= \mathrm{diag}(-I_r,\ 0_{n_1}, 0_m) \qquad \text{by (10)}, \end{aligned}$$

and

$$\begin{aligned} X^T M X &= \mathrm{diag}(\widetilde{X}^T,\ I_m) Q^T (A + \mu B + ZSZ^T) Q\, \mathrm{diag}(\widetilde{X},\ I_m) \\ &= \mathrm{diag}(\widetilde{X}^T,\ I_m)\, \mathrm{diag}(\widehat{A} + \mu\widehat{B}, GSG^T)\, \mathrm{diag}(\widetilde{X},\ I_m) \qquad \text{by (8) and (9)} \\ &= \mathrm{diag}(\Lambda_r + \mu I_r,\ I_{n_1}, GSG^T) \qquad \text{by (10)}. \end{aligned}$$

Since $\Lambda_r \succeq 0$, $S > 0$ and $\mu > 0$, $M \succ 0$. The nonzero eigenvalues of the pencil $K - \sigma M$ are $\sigma_i = -1/(\lambda_i + \mu)$ for $i = 1, \ldots, r$.

References

[Fix and Heiberger, 1972] Fix, G. and Heiberger, R. (1972). An algorithm for the ill-conditioned generalized eigenvalue problem. *SIAM Journal on Numerical Analysis*, 9:78–88.

[Liang et al., 2013] Liang, X., Li, R.-C., and Bai, Z. (2013). Trace minimization principles for positive semi-definite pencils. *Linear Algebra and its Applications*, 438:3085–3106.

[Newcomb, 1961] Newcomb, R. W. (1961). On the simultaneous diagonalization of two semi-definite matrices. *Quarterly of Applied Mathematics*, 19:144–146.