ANALYSIS OF MAX-MIN EIGENVALUE OF CONSTRAINED LINEAR COMBINATIONS OF SYMMETRIC MATRICES

Qu-Tang Cai[†], *Chun-Yi Peng*[‡], *Chang-Shui Zhang*[†]

[†] Department of Automation, Tsinghua University, Beijing 100084, P.R.China
[‡] Microsoft Research Asia, 49 Zhichun Road, Haidian District, Beijing 100084, P.R.China

ABSTRACT

This paper studies the problem whether the smallest eigenvalue of constrained linear combinations of symmetric matrices can reach a desirable value, which actually extends the mathematical problem of finding a Positive Definite Linear Combination of symmetric matrices(PDLC), and provides a universal framework to maximize the minimal eigenvalue of linear combined symmetric matrices. For solving this problem, we cast an equivalent optimization task, and propose one general algorithm framework that is proved to be globally optimal and convergent. Both theoretical analysis and experiments under a typical constraint verify our algorithm's validity and efficiency.

***Index Terms*—** Matrix multiplication, Optimization methods, Spectral analysis, Eigenvalues and eigenfunctions

1. INTRODUCTION

The smallest or largest eigenvalue of a matrix plays an important role in matrix analysis and signal processing, e.g., system stability and spectrum analysis. Among those applications, finding a Positive Definite Linear Combination of symmetric matrices(PDLC)[1] plays active in areas of signal processing, e.g. moving-average processes identification [2] and blind source separation [3]. Two algorithms have been proposed to find a PLDC [1][2], but they both assume that such a positive definite combination must exist. However unfortunately, in some cases, this assumption does not hold and their algorithms may fail and waste too much computation resources. Therefore, we study the existence problem of PDLC in this paper, and moreover, we propose a more generalized PDLC problem: how to determine whether a linear combination of symmetric matrices exists under varying constraints such that the smallest eigenvalue is larger than the given value, and to obtain one solution if exists.

The rest of the paper is organized as follows: we introduce the generalized PDLC problem and derive its equivalent optimization form in Section 2. In Section 3, we propose an iterative algorithm of two sub-optimization steps and systematically investigate its convergence and stop criteria. Finally, some experiment results are presented in Section 4, followed by our conclusion in Section 5.

2. PROBLEM FORMULATION

We introduce some basic notations here for quick reference. We use uppercase boldfaced letters for matrices with $(\cdot)^t$, $\lambda_i(\cdot)$, $tr(\cdot)$, $\|\cdot\|$, respectively, for the transpose, the i-th largest eigenvalue, the trace and the Frobenius norm, and lowercase boldfaced letters for vectors with $\|\cdot\|$ and $\|\cdot\|_\infty$ for the Euclidean norm and the ∞-norm. Let $\tilde{\mathbf{A}}$ for $\mathbf{A} \in R^{n\times n}$ represent $\mathbf{A}$'s vector form where the $(k-1)n+1:kn$ entries correspond to the k-th row in $\mathbf{A}$. Define $\langle \mathbf{A}, \mathbf{B}\rangle$ for matrices $\mathbf{A}$ and $\mathbf{B}$ by the inner product $\langle \tilde{\mathbf{A}}, \tilde{\mathbf{B}}\rangle$; Define $\max(\Lambda, \sigma\mathbf{I}) = diag(\max(\sigma, \alpha_1), \cdots, \max(\sigma, \alpha_n))$ as a σ^+-form for the diagonal matrix $\Lambda = diag(\alpha_1, \cdots, \alpha_n)$; $S_1 - S_2$ for two sets S_1 and S_2 denotes the set $\{s_0 \mid \exists s_1 \in S_1, s_2 \in S_2, s.t.\ s_0 = s_1 - s_2\}$.

Let $\mathcal{P}_\sigma^+$, $\sigma \in R$, denote the set of all the $p \times p$ symmetric matrices whose smallest eigenvalue is not less than σ, and $\mathbf{A}_1, \cdots, \mathbf{A}_n$ be $p\times p$ symmetric matrices, $S \subseteq R^n$ is a compact convex set representing particular constraints, and then $\Theta \doteq \{\sum_{k=1}^n \alpha(k)\cdot \mathbf{A}_k \mid \alpha \in S\}$ denote all the linear combinations of these symmetric matrices under constraint S. The problem is to determine whether there exists one linear combined matrix whose smallest eigenvalue is not less than σ, namely, $\mathcal{P}_\sigma^+ \cap \Theta = \emptyset$ or not, and furthermore, to find out one element in $\mathcal{P}_\sigma^+ \cap \Theta$ if $\mathcal{P}_\sigma^+ \cap \Theta \neq \emptyset$. It is equivalent to seek "zero" for the following optimization problem:

$$Minimize\, F(\mathbf{B}, \mathbf{C}) \doteq \|\mathbf{B} - \mathbf{C}\|\ , \mathbf{B} \in \mathcal{P}_\sigma^+, \mathbf{C} \in \Theta. \quad (1)$$

Obviously, if $F(\mathbf{B}, \mathbf{C}) = 0$ is achievable, and then the desired combination exists and the optimal $\mathbf{B}$ is one desirable solution, otherwise no such a combination exists. Before we go to determine whether $F(\mathbf{B}, \mathbf{C}) = 0$, let us explore the property of (1) in Lemma 1: the optimization solution to (1) exists and the optimal $\mathbf{B} - \mathbf{C}$, written by $\mathbf{B}^* - \mathbf{C}^*$, is unique. The proof can be found in Appendix A.

Lemma 1: *There exists at least one* $(\mathbf{B}^*, \mathbf{C}^*) \in \mathcal{P}_\sigma^+ \times \Theta$, *s.t.* $\|\mathbf{B}^* - \mathbf{C}^*\| = d^*$; *moreover, for any optimal* $(\mathbf{B}^*, \mathbf{C}^*)$ *(such that* $\|\mathbf{B}^* - \mathbf{C}^*\| = d^*$*),* $\mathbf{B}^* - \mathbf{C}^*$ *is unique, where* $d^* \doteq \inf\{\|\mathbf{B} - \mathbf{C}\| \mid (\mathbf{B}, \mathbf{C}) \in P_\sigma^+ \times \Theta\}$.

3. ALGORITHMIC FRAMEWORK AND ANALYSIS

3.1. Optimization Procedure

We adopt an iterative two-step procedure to optimize (1)– minimizing $\|\mathbf{B} - \mathbf{C}\|$ alternatively with respect to one of the parameters $\mathbf{B}$ and $\mathbf{C}$ with the other kept fixed. Let $F_\mathbf{C}(\mathbf{X}) : \mathcal{P}_\sigma^+ \to R$ defined by $F_\mathbf{C}(\mathbf{X}) \doteq \|\mathbf{X} - \mathbf{C}\|$, and $G_\mathbf{B}(\mathbf{X}) : \Theta \to R$ defined by $G_\mathbf{B}(\mathbf{X}) \doteq \|\mathbf{X} - \mathbf{B}\|$, and the whole optimization can be divided to the below sub-optimization problems:

3.1.1. Minimization of $F_\mathbf{C}(\mathbf{X})$

The optimal solution for $F_\mathbf{C}(\mathbf{X})$ can be given by the below proposition, whose proof can be found in [1].

Proposition 1: *Let the eigen-decomposition of symmetric matrix* $\mathbf{C} \in R^{p\times p}$ *be given by* $\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^t$, *where* $\mathbf{U}\mathbf{U}^t = \mathbf{I}_{p\times p}$ *and* $\mathbf{\Lambda}$ *is a diagonal matrix; Construct* $\mathbf{\Lambda}^{(\sigma^+)} = \max(\sigma\mathbf{I}, \mathbf{\Lambda})$, *and then* $f(\mathbf{C}) \doteq \mathbf{U}\mathbf{\Lambda}^{(\sigma^+)}\mathbf{U}^t$ *minimizes* $F_\mathbf{C}(\mathbf{X})$.

Moreover, since $\mathcal{P}_\sigma^+$ is closed and convex, the optimal X^* s.t. $F_{\mathbf{C}}(\mathbf{X}^*) = \min F_{\mathbf{C}}(\mathbf{X})$, if and only if $\mathbf{X}^* = f(\mathbf{C})$.

3.1.2. *Minimization of* $G_{\mathbf{B}}(\mathbf{X})$

Let $\mathcal{H} \doteq span\{\mathbf{A}_1, \cdots, \mathbf{A}_n\}$, $\mathcal{A} \doteq [\tilde{\mathbf{A}}_1, \cdots, \tilde{\mathbf{A}}_n]$, and $P_{\mathcal{H}}$ be the orthogonal projection operator from $R^{p\times p}$ onto $\mathcal{H}$. According to Pythagorean theorem,

$$G_{\mathbf{B}}(\mathbf{X})^2 = \|\mathbf{B} - P_{\mathcal{H}}(\mathbf{B})\|^2 + \|\mathbf{X} - P_{\mathcal{H}}(\mathbf{B})\|^2. \qquad (2)$$

Since $\|\mathbf{B} - P_{\mathcal{H}}(\mathbf{B})\|$ is fixed, *Minimize* $G_{\mathbf{B}}(\mathbf{X})^2$ turns to minimize $\|\mathbf{X} - P_{\mathcal{H}}(\mathbf{B})\|^2$. Writing $\tilde{\mathbf{X}}$ as $\tilde{\mathbf{X}} = \mathcal{A}\alpha$, then

$$\|\mathbf{X} - P_{\mathcal{H}}(\mathbf{B})\|^2 = \alpha^t\mathcal{A}^t\mathcal{A}\alpha - 2\alpha^t\mathcal{A}^t\tilde{\mathbf{B}} + const. \qquad (3)$$

That is, the minimization of $G_{\mathbf{B}}(\mathbf{X})$ can be reduced to:

$$\text{minimize } \alpha^t\mathcal{A}^t\mathcal{A}\alpha - 2\alpha^t\mathcal{A}^t\tilde{\mathbf{B}}, \text{ subject to } \alpha \in S. \qquad (4)$$

Note that (4) is a typical quadratic optimization problem, which is globally optimized [4]. Like $f(\cdot)$, the optimal solution for minimization of $G_{\mathbf{B}}(\mathbf{X})$ is also unique, and denoted by $g(\mathbf{B})$ for later use.

3.2. Convergence Property

As is known, iterative techniques can not usually converge to a global optimal value in multivariate convex optimization. However, our optimization method can be shown to finally converge in the below.

Let $(\mathbf{B}^*, \mathbf{C}^*)$ be one optimal solution of (1), and $\mathbf{D}^* \doteq \mathbf{B}^* – \mathbf{C}^*$. Assume the optimization procedure starts from arbitrary $\mathbf{C}_0$, and $\mathbf{B}_i \doteq f(\mathbf{C}_i)$, $\mathbf{C}_{i+1} \doteq g(\mathbf{B}_i)$, $\mathbf{D}_i = \mathbf{B}_i - \mathbf{C}_i$, and $d_i \doteq \|\mathbf{D}_i\|$. According to the above optimization procedure,

$$d_{i+1} \le \|\mathbf{B}_i - \mathbf{C}_{i+1}\| \le d_i. \qquad (5)$$

Furthermore, the convergence property is given as follows:

Proposition 2: *1)* $\mathbf{C}_{i+1} = \mathbf{C}_i$ *if and only if* $d_i = d^*$, *namely, the sequence* d_i*'s is strictly decreasing until* $d_i = d^*$ *for some i. 2)* $\lim_{i\to\infty} d_i = d^*$ *and* $\lim_{i\to\infty}\|\mathbf{D}_i - \mathbf{D}^*\| = 0$.

Proof: 1) If $\mathbf{C}_{i+1} = \mathbf{C}_i$, then $\mathbf{C}_i$ also minimizes $G_{\mathbf{B}_i}(\mathbf{X})$. Therefore, $\forall \mathbf{C} \in \Theta$, $\langle \mathbf{B}_i - \mathbf{C}_i, \mathbf{C} - \mathbf{C}_i\rangle \le 0$. Similarly, since $\mathbf{B}_i = f(\mathbf{C}_i)$, then $\forall \mathbf{B} \in \mathcal{P}_\sigma^+$, $\langle \mathbf{C}_i - \mathbf{B}_i, \mathbf{B} - \mathbf{B}_i\rangle \le 0$. Therefore, $\|\mathbf{B}_i - \mathbf{C}_i\|^2 \le \langle \mathbf{B} - \mathbf{C}, \mathbf{B}_i - \mathbf{C}_i\rangle$. Furthermore, by Schwartz's inequality, $\|\mathbf{B}_i - \mathbf{C}_i\| \le \|\mathbf{B} - \mathbf{C}\|$, for all $\mathbf{B}, \mathbf{C}$. Thus $\|\mathbf{B}_i - \mathbf{C}_i\| = d^*$, i.e., $d_i = d^*$. Conversely, if $d_i = d^*$, by the uniqueness in Lemma 1, one can have $\mathbf{C}_{i+1} = \mathbf{C}_i$.

It is noticed that the sequence d_i's is non-increasing from the optimization procedure. If $\exists\, i_0$, s.t. $d_{i_0} = d_{i_0+1}$, then $d_{i_0} = \|\mathbf{B}_{i_0} - \mathbf{C}_{i_0}\| \ge \|\mathbf{B}_{i_0} - \mathbf{C}_{i_0+1}\| \ge \|\mathbf{B}_{i_0+1} - \mathbf{C}_{i_0+1}\| = d_{i_0+1} = d_{i_0}$. Thus, $\|\mathbf{B}_{i_0} - \mathbf{C}_{i_0}\| = \|\mathbf{B}_{i_0} - \mathbf{C}_{i_0+1}\| = \min G_{\mathbf{B}_{i_0}}(\mathbf{X})$, so $\mathbf{C}_{i_0}$ also minimizes $G_{\mathbf{B}_{i_0}}(\mathbf{X})$. By the uniqueness, $\mathbf{C}_{i_0} = g(\mathbf{B}_{i_0}) = \mathbf{C}_{i_0+1}$ and then $d_{i_0} = d^*$.

2) Firstly, we show below that $\forall \delta > 0$, $\exists \eta_\delta \in (0,1)$, s.t. if $\|f(\mathbf{C}) - \mathbf{C}\| - d^* \ge \delta$,

$$\|f(\mathbf{C}) - g\circ f(\mathbf{C})\| - d^* \le \eta_\delta(\|f(\mathbf{C}) - \mathbf{C}\| - d^*). \qquad (6)$$

Define

$$\Xi_\delta \doteq \{\mathbf{C} \in \Theta \mid \|f(\mathbf{C}) - \mathbf{C}\| - d^* \ge \delta\}.$$

Ξ_δ is compact in $R^{p\times p}$ since $f(\cdot)$ is continuous and Θ is compact. Now consider the map $h : \Xi_\delta \to R$ defined by $h(\mathbf{C}) \doteq \frac{\|f(\mathbf{C}) - g\circ f(\mathbf{C})\| - d^*}{\|f(\mathbf{C}) - \mathbf{C}\| - d^*}$. Clearly, $h(\cdot)$ is continuous and $0 \le h(\cdot) \le 1$. Since Ξ_δ is compact, $\sup\{h(\mathbf{C}) \mid \mathbf{C} \in \Xi_\delta\}$ can be attained, denoted by η_δ. Moreover, if $\eta_\delta = 1$, then $\exists \mathbf{C}^\diamond \in \Xi_\delta$, s.t. $h(\mathbf{C}^\diamond) = 1$, i.e. $\|f(\mathbf{C}^\diamond) - g\circ f(\mathbf{C}^\diamond)\| = \|f(\mathbf{C}^\diamond) - \mathbf{C}^\diamond\|$, so $\mathbf{C}^\diamond$ minimizes $G_{f(\mathbf{C}^\diamond)}(\mathbf{X})$. As the first part of this proposition, $\|f(\mathbf{C}^\diamond) - \mathbf{C}^\diamond\| = d^*$, which is contradictory to the definition of Ξ_δ. Therefore, $\eta_\delta \in (0,1)$ and (6) holds.

Assume $d_1 > d^* + \delta$. Let

$$N_\delta = \log(\delta/(d_1 - d^*))/\log\eta_\delta, \qquad (7)$$

then $\forall\, i > N_\delta$, by (6) and the first part of Proposition 2,

$$d_i - d^* \le \max(\delta, \eta_\delta^{N_\delta}(d_1 - d^*)) = \delta \qquad (8)$$

Hence $\{d_i\}_{i=1}^\infty$ is a cauchy sequence which converges to d^*. Moreover, since $(\mathbf{D}_i + \mathbf{D}^*)/2 \in \Theta$, then

$$\begin{aligned}\|\mathbf{D}_i - \mathbf{D}^*\|^2 &= 2\|\mathbf{D}_i\|^2 + 2\|\mathbf{D}^*\|^2 - \|\mathbf{D}_i + \mathbf{D}^*\|^2 \\ &\le 2d_i^2 + 2d^{*2} - 4d^{*2} = 2d_i^2 - 2d^{*2}.\end{aligned}$$

Therefore, $\lim_{i\to\infty}\|\mathbf{D}_i - \mathbf{D}^*\| = 0$. ■

Proposition 2 shows that our iterative algorithm is monotonically convergent to a globally optimal solution, but it is difficult to determine its convergence rate, that is, to obtain one η_δ in (6). Proposition 3 presents one explicit η_δ in special case of $d^* = 0$, and the proof can be found in Appendix B.

Proposition 3: *If* $d^* = 0$, η_δ *can be given by* $\sqrt{1 - \frac{\delta^2}{diam(\Theta)^2}}$, *where* $diam(\Theta) \doteq \max\{\|\mathbf{X}_1 - \mathbf{X}_2\| \mid \mathbf{X}_1, \mathbf{X}_2 \in \Theta\}$.

3.3. Stop Condition for the Algorithm

We propose a simple criterion to judge whether $d^* = 0$, and add stop conditions for the above optimization to avoid needless computation and thus to accelerate the algorithm. Firstly, let us introduce the equivalence criterion of $d^* \ne 0$ as the below lemma.

Lemma 2: $d^* \ne 0$ *if and only if there exists a matrix* $\mathbf{V} \in R^{p\times p}$ *s.t. for* $\forall \mathbf{B} \in \mathcal{P}_\sigma^+$, $\mathbf{C} \in \Theta$, $\langle \mathbf{V}, \mathbf{B}\rangle - \langle \mathbf{V}, \mathbf{C}\rangle > 0$.

The proof can be easily derived from the separation theorem of convex sets, so it is omitted here due to space limitation. Though Lemma 2 provides an equivalent statement for $d^* \ne 0$, it is still not easy to judge whether such a matrix $\mathbf{V}$ exists. Proposition 4 find a way using $\mathbf{D}^*$ and $\mathbf{D}_i$ sequences to determine if $d^* = 0$.

Proposition 4:

1) *For* $\forall(\mathbf{B}, \mathbf{C}) \in \mathcal{P}_\sigma^+ \times \Theta$, $\langle \mathbf{D}^*, \mathbf{B}\rangle - \langle \mathbf{D}^*, \mathbf{C}\rangle \ge d^{*2}$.

2) *For* $\forall \mathbf{B} \in \mathcal{P}_\sigma^+$ *and some symmetric positive semi-definite matrix* $\mathbf{D} \in R^{p\times p}$, $\inf\langle \mathbf{D}, \mathbf{B}\rangle = \sigma\cdot tr(\mathbf{D})$, $\inf\langle \mathbf{D}_i, \mathbf{B}\rangle = \sigma\cdot tr(\mathbf{D}_i)$ *and* $\inf\langle \mathbf{D}^*, \mathbf{B}\rangle = \sigma\cdot tr(\mathbf{D}^*)$. *Furthermore,*

$$\lim_{i\to\infty}\inf\langle \mathbf{D}_i, \mathbf{B}\rangle = \inf\langle \mathbf{D}^*, \mathbf{B}\rangle. \qquad (9)$$

3) *If* $\lambda_p(\mathbf{C}_i) < \sigma$, *then*

$$\sup_{\mathbf{C}\in\Theta}\lambda_p(\mathbf{C}) \le \sigma + \frac{(\sigma - \lambda_p(\mathbf{C}_i))}{\|\mathbf{D}_i\|}\left(\sup_{\mathbf{C}\in\Theta}\langle \mathbf{D}_i, \mathbf{C}\rangle - \inf_{\mathbf{B}\in\mathcal{P}_\sigma^+}\langle \mathbf{D}_i, \mathbf{B}\rangle\right). \qquad (10)$$

4) *If* $d^* \ne 0$, *then* $\exists\, N, \forall\, i > N$ *s.t.*

$$\inf_{\mathbf{B}\in\mathcal{P}_\sigma^+}\langle \mathbf{D}_i, \mathbf{B}\rangle > \sup_{\mathbf{C}\in\Theta}\langle \mathbf{D}_i, \mathbf{C}\rangle. \qquad (11)$$

5) *If* $d^* = 0$*, then* $\forall \epsilon > 0$, $\exists N_\epsilon$*, s.t.* $\forall i > N_\epsilon$, $\lambda_p(\mathbf{C}_i) \geq \sigma - \epsilon$.

The detailed proof can be found in Appendix C. $\sup_{\mathbf{C}\in\Theta} \langle \mathbf{D}_i, \mathbf{C} \rangle$ in Proposition 4 can be computed by solving the below optimization problem:

$$\text{maximize } \sum\nolimits_{k=1}^{n} \alpha(k) \cdot \beta_i(k), \text{ subject to } \alpha \in S, \tag{12}$$

where $\beta_i(k) \doteq \langle \mathbf{D}_i, \mathbf{A}_k \rangle$. The optimal value for (12) can be generally calculated by linear programming, and can be explicitly given for certain S, e.g. the optimal value is $\sum_{k=1}^{n} |\beta_i(k)|$ for $S = \left\{\alpha \in R^n \mid \|\alpha\|_\infty \leq 1\right\}$. Until now, the algorithmic framework is summarized in Table 1, with rough upper bounds for the iteration number of the framework given in (21) and (23).

Table 1. Algorithmic Framework

Input: Symmetric matrices $\mathbf{A}_1, \cdots, \mathbf{A}_n$, a desirable value σ_0, and a relaxation factor $\epsilon \in (0, +\infty)$.

Init: Let $\sigma_1 = \sigma_0 - \epsilon$, $\sigma_2 = \sigma_0 - 2\epsilon$, $\mathcal{A} = [\tilde{\mathbf{A}}_1, \cdots, \tilde{\mathbf{A}}_n]$, and $P_{\mathcal{H}} = \mathcal{A}(\mathcal{A}^t\mathcal{A})^{-1}\mathcal{A}$. Set iteration counter $i = 0$, choose an arbitrary start point $\mathbf{C}_0 \in \Theta$ and compute $\mathbf{B}_0$.

Step I: $i = i+1$; compute $P_{\mathcal{H}}(\mathbf{B}_{i-1})$ corresponding to $P_{\mathcal{H}} \cdot \tilde{\mathbf{B}}_{i-1}$; compute $\mathbf{C}_i$ by solving (4);

Step II: Compute the eigen-decomposition $[\mathbf{U}_i, \mathbf{\Lambda}_i]$ of $\mathbf{C}_i$; if $\lambda_p(\mathbf{C}_i) \geq \sigma_2$, then $\mathbf{C}_i$ is one matrix whose smallest eigenvalue larger than σ_2 and exit; otherwise let $\mathbf{B}_i = \mathbf{U}_i \mathbf{\Lambda}^{(\sigma_1^+)} \mathbf{U}_i^t$ where $\mathbf{\Lambda}^{(\sigma_1^+)} = \max(\sigma_1 \mathbf{I}, \mathbf{\Lambda})$.

Step III: Compute $\mathbf{D}_i = \mathbf{B}_i - \mathbf{C}_i$ and $\omega_i \doteq \sup\{tr(\mathbf{D}_i\mathbf{C}) \mid \mathbf{C} \in \Theta\}$ by solving (12); if $(\omega_i - tr(\mathbf{D}_i)) < \epsilon \|\mathbf{D}_i\| / (\sigma_1 - \lambda_p(\mathbf{C}_i))$, then by (10), there is no linear combination with $\lambda_p \geq \sigma_0$ and exit; otherwise, loop back to step I.

4. EXPERIMENTAL RESULTS

We will build our experiment scenario under one typical constraint, the ellipsoid constraint $S = \left\{\alpha \in R^n \mid \|\alpha\|_2 \leq 1\right\}$ which can be extended, via scaling, translation and rotation, to the more general form $\left\{\alpha \in R^n \mid (\alpha - \alpha_0)^t \mathbf{M} (\alpha - \alpha_0) \leq r\right\}$, where $\mathbf{M}$ is symmetric positive definite and $r > 0$.

Let us derive our algorithm from the algorithm framework in Table 1. For $S = \left\{\alpha \in R^n \mid \|\alpha\|_2 \leq 1\right\}$, the optimal value for (12) in the i-th step can be explicitly given by $\|\beta\|$, and (4) can be solved using Kuhn-Tucker theorem (see [4]). Let the Lagrange multiplier be given by

$$L(\alpha, \lambda) = \alpha^t \mathcal{A}^t \mathcal{A} \alpha - 2\alpha^t \mathcal{A}^t \tilde{\mathbf{B}} + \lambda(\|\alpha\|^2 - 1),$$

and by Kuhn-Tucker theorem, the solution α^* is optimal if and only if $\exists \lambda^* \geq 0$, s.t.

$$\begin{cases} \partial L(\alpha, \lambda)/\partial \alpha = 2\mathcal{A}^t\mathcal{A}\alpha - 2\mathcal{A}^t\tilde{\mathbf{B}} + 2\lambda^*\alpha = 0, \\ \lambda^*(\|\alpha\|^2 - 1) = 0. \end{cases} \tag{13}$$

From (13), (α^*, λ^*) can be given in the below two cases:

1. When $\|P_{\mathcal{H}}(\mathbf{B})\| \leq 1$, $\lambda^* = 0$, and $\mathcal{A}\alpha^* = P_{\mathcal{H}} \cdot \tilde{\mathbf{B}}$.
2. When $\|P_{\mathcal{H}}(\mathbf{B})\| > 1$, $\alpha^* = (\mathcal{A}^t\mathcal{A} + \lambda^*)^{-1}\mathcal{A}^t\vec{B}$. Moreover, to minimize $\|P_{\mathcal{H}}(\mathbf{B}) - \alpha^*\|$, α^* should lie on the boundary of S and thus $\tilde{\mathbf{B}}^t\mathcal{A}((\mathcal{A}^t\mathcal{A}+\lambda^*)(\mathcal{A}^t\mathcal{A}+\lambda^*))^{-1}\mathcal{A}^t\tilde{\mathbf{B}} = 1$. Let $\mathcal{A}^t\mathcal{A}$ be decomposed as $\mathcal{A}^t\mathcal{A} = \mathbf{U}^t \cdot diag(u_1, \cdots, u_n) \cdot \mathbf{U}$, and then

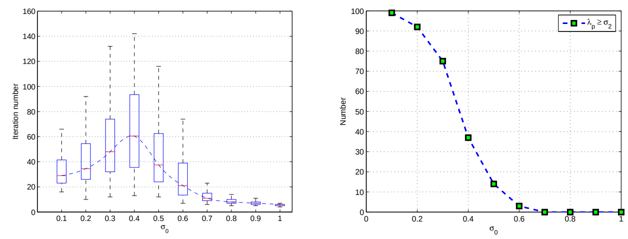
   $$\tilde{\mathbf{B}}^t\mathcal{A}\mathbf{U} \cdot diag(1/(u_1+\lambda^*)^2, \cdots, 1/(u_n+\lambda^*)^2) \cdot \mathbf{U}^t\mathcal{A}^t\tilde{\mathbf{B}} = 1.$$

   Let the k-th entry in $\mathbf{U}^t\mathcal{A}^t\tilde{\mathbf{B}}$ be denoted by υ_k. Thus,

   $$\frac{\upsilon_1^2}{(u_1+\lambda^*)^2} + \frac{\upsilon_2^2}{(u_2+\lambda^*)^2} + \cdots + \frac{\upsilon_n^2}{(u_n+\lambda^*)^2} = 1. \tag{14}$$

   The left-hand side of (14) is an increasing function w.r.t. λ^* in $[0, +\infty)$, and λ^* should belong to $\left[0, \sqrt{\upsilon_1^2 + \cdots + \upsilon_n^2}\right]$, and then (14) can be solved efficiently via bisection method, which performs significantly faster than the quadratic programming method.

Experimental results are presented for the algorithm: let $p = 10$, $n = 50$, $\epsilon = 0.01\sigma_0$, and $\sigma_0 = 0.1, 0.2, \cdots, 2$; the upper triangular entries of $\mathbf{A}_i, i = 1, \cdots, n$, are randomly generated from $[-1, 1]$ with uniform distribution. For each σ_0, 500 trials runs with the maximal iteration number set as 200. In all experiments with varying σ, these trials converge in finite steps and their mean(variant) of iteration number is plotted in Fig. 1(a). Notice that, compared to the maximal number of iteration, the actual required iteration number is relative small (about 20∼40), which verifies our algorithm works effectively and efficiently. Fig.1(b) plots that the ratio of the trials where one matrix with $\lambda_p \geq \sigma_2$ exists decreases as σ_0 increases, that is, the linear combination is getting harder to find with σ_0 increasing, which also matches with our common sense. Besides, it should be noticed that our algorithm still works efficiently to detect whether such combination exists, even when σ_0 is large and the existing assumption in [1][2] does not hold.



(a) (b)

Fig. 1. Experimental statistics versus varying the desirable the smallest eigenvalue threshold σ: (a) Number of iterations; (b) Number of trials where one matrix with $\lambda_p \geq \sigma_2$ exists.

5. CONCLUSION

This paper has proposed an extensive framework to analyze if the maximum of the smallest eigenvalue linear combinations of finite symmetric matrices under constraints can exceed the value preset. We have proved that the optimization framework is global optimal, and have given upper bounds for the number of iterations required in the framework. With little modification, the framework can be used to determine the interval to which the maximum of the smallest eigenvalues should belong, and if no truncation error takes effect, the estimation for the maximum can be achieved to whatever accuracy needed, with a bisection technique for narrowing the interval. Since the smallest eigenvalues of matrices is so important in signal processing and matrix analysis, the proposed framework promises to be useful.

6. REFERENCES

[1] Abdelhamid Zaidi, "Positive definite combination of symmetric matrices," *IEEE Trans. Signal Processing*, vol. 53, no. 11, pp. 4412–4416, 2005.

[2] L. Tong, Y. Iujiro, and R.-W. Liu, "A finite-step global convergence algorithm for the parameter estimation of multichannel ma processes," *IEEE Trans. Signal Processing*, vol. 40, no. 10, pp. 2547–2558, 1992.

[3] S. Choi, A. Cichocki, and A. Belouchrani, "Second order nonstationary source separation," *Journal of VLSI Signal Processing*, vol. 32, no. 1-2, pp. 93–104, 2002.

[4] Rangarajan K. Sundaram, *A first course in optimization theory*, Cambridge University Press, New York, 1999.

[5] Walter Rudin, *Real and complex analysis*, McGraw-Hill, New York, 3rd edition, 1987.

[6] Beresford N. Parlett, *The Symmetric eigenvalue problem*, SIAM, Philadelphia, 1998.

Appendix

A. Proof of Lemma 1

Consider $R^{p\times p}$ is a Hilbert Space. In $R^{p\times p}$, $\mathcal{P}_\sigma^+$ is closed and convex, and Θ is compact and convex, so $\mathcal{P}_\sigma^+ - \Theta$ is closed and convex and $F(\mathbf{B},\mathbf{C})$ is continuous and convex. By the definition of d^*, there exists a sequence $\{(\mathbf{B}_i,\mathbf{C}_i)\}_{i=1}^{\infty}$ s.t. $\lim_{i\to\infty}\|\mathbf{B}_i-\mathbf{C}_i\|=d^*$.

Moreover, $\|(\mathbf{B}_i+\mathbf{B}_j)/2-(\mathbf{C}_i+\mathbf{C}_j)/2\| \geq d^*$. Therefore, with Pythagoras's Theorem,

$$\|(\mathbf{B}_i-\mathbf{C}_i)-(\mathbf{B}_j-\mathbf{C}_j)\|^2 \leq \|\mathbf{B}_i-\mathbf{C}_i\|^2+\|\mathbf{B}_j-\mathbf{C}_j\|^2-4d^{*2}.$$

Since $\lim_{i\to\infty}\|\mathbf{B}_i-\mathbf{C}_i\| = d^*$, $\{\mathbf{B}_i-\mathbf{C}_i\}_{i=1}^{\infty}$ is a Cauchy sequence. By the closeness of $\mathcal{P}_\sigma^+-\Theta$, $\exists\,\mathbf{D}^*\in\mathcal{P}_\sigma^+-\Theta$, s.t. $\mathbf{D}^*=\lim_{i\to\infty}(\mathbf{B}_i-\mathbf{C}_i)$. Let $\mathbf{D}^*$ be given by $\mathbf{D}^*=\mathbf{B}^*-\mathbf{C}^*$, thus $(\mathbf{B}^*,\mathbf{C}^*)$ is an optimal solution for (1). Moreover, $\mathbf{D}^*$ is unique because, in $R^{p\times p}$, $\mathcal{P}_\sigma^+-\Theta$ is closed and convex, and $\mathbf{D}^*$ minimizes $\|\mathbf{D}\|$ w.r.t. $\mathbf{D}\in\mathcal{P}_\sigma^+-\Theta$ (see pp.79 in [5]). ■

B. Proof of Proposition 3

Let $a \doteq \langle f(\mathbf{C})-\mathbf{C},\mathbf{B}^*-\mathbf{C}\rangle/\|\mathbf{B}^*-\mathbf{C}\|^2$. Using $\mathbf{B}^*$ is an optimal solution, $\|f(\mathbf{C})-\mathbf{C}\|\leq\|\mathbf{B}^*-\mathbf{C}\|$, we have $a\leq 1$ and

$$a \geq \langle f(\mathbf{C})-\mathbf{C}, f(\mathbf{C})-\mathbf{C}\rangle/\|\mathbf{B}^*-\mathbf{C}\|^2. \tag{15}$$

Since $\mathbf{B}^*=\mathbf{C}^*$, hence $a\in\left[\|f(\mathbf{C})-\mathbf{C}\|^2/\|\mathbf{C}^*-\mathbf{C}\|^2,1\right]$ and $\mathbf{C}+a(\mathbf{C}^*-\mathbf{C})\in\Theta$. By definition of a, $\langle f(\mathbf{C})-\mathbf{C}-a*(\mathbf{C}^*-\mathbf{C}),\mathbf{C}^*-\mathbf{C}\rangle=0$. Then

$$\begin{aligned}\|f(\mathbf{C})-g\circ f(\mathbf{C})\|^2 &\leq \|f(\mathbf{C})-\{\mathbf{C}+a(\mathbf{C}^*-\mathbf{C})\}\|^2\\ &= \|f(\mathbf{C})-\mathbf{C}\|^2-a^2\|\mathbf{C}^*-\mathbf{C}\|^2\\ &\leq \|f(\mathbf{C})-\mathbf{C}\|^2-\|f(\mathbf{C})-\mathbf{C}\|^4/\|\mathbf{C}^*-\mathbf{C}\|^2\\ &\leq \|f(\mathbf{C})-\mathbf{C}\|^2\left(1-\frac{\delta^2}{diam(\Theta)^2}\right).\blacksquare\end{aligned}\tag{16}$$

C. Proof of Proposition 4

1) Since $(\mathbf{B}^*,\mathbf{C}^*)$ minimize (1), then we have

$$\langle\mathbf{C}^*-\mathbf{B}^*,\mathbf{B}-\mathbf{B}^*\rangle\leq 0 \text{ and } \langle\mathbf{B}^*-\mathbf{C}^*,\mathbf{C}-\mathbf{C}^*\rangle\leq 0.$$

Therefore, $\langle\mathbf{B}^*-\mathbf{C}^*,\mathbf{B}-\mathbf{C}\rangle \geq \langle\mathbf{B}^*-\mathbf{C}^*,\mathbf{B}^*-\mathbf{C}^*\rangle$, that is, $\langle\mathbf{D}^*,\mathbf{B}\rangle-\langle\mathbf{D}^*,\mathbf{C}\rangle\geq d^{*2}$.

2) As a symmetric positive semi-definite matrix, $\mathbf{D}$ can be decomposed into $\mathbf{D}=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^t$ and all diagonal entries in $\mathbf{\Lambda}$ are nonnegative. For $\forall\mathbf{B}\in\mathcal{P}_\sigma^+$, we have

$$\langle\mathbf{D},\mathbf{B}\rangle=tr(\mathbf{DB})=tr(\mathbf{U\Lambda U}^t\mathbf{B})=tr(\mathbf{\Lambda U}^t\mathbf{BU}). \tag{17}$$

Since $\mathbf{U}^t\mathbf{BU}\in\mathcal{P}_\sigma^+$, then the diagonal entries in $\mathbf{U}^t\mathbf{BU}$ are all larger than σ, that is, $tr(\mathbf{\Lambda U}^t\mathbf{BU})\geq\sigma\cdot tr(\mathbf{\Lambda})=\sigma\cdot tr(\mathbf{D})$; On the other side, the equality can be achieved when $\mathbf{B}=\mathbf{I}_{p\times p}$; Hence,

$$\inf_{\mathbf{B}\in\mathcal{P}_\sigma^+}\langle\mathbf{D},\mathbf{B}\rangle=\sigma\cdot tr(\mathbf{D}) \tag{18}$$

In the i-th iteration, where $\mathbf{C_i}=\mathbf{U_i\Lambda_i U_i}^t$, by Proposition 1, we can derive $\mathbf{B}_i=\mathbf{U_i\Lambda_i^{(\sigma_1^+)}U_i^t}$, where $\Lambda_i^{(\sigma_1^+)}=\max(\sigma_1\mathbf{I},\Lambda_i)$, and then

$$\mathbf{D}_i=\mathbf{U_i\Lambda_i^\ddagger U_i}^t,\text{ where } \Lambda_i^\ddagger=\max(\mathbf{0},\sigma_1\mathbf{I}-\Lambda_i). \tag{19}$$

Obviously, $\mathbf{D}^*=\lim_{i\to\infty}\mathbf{D}_i$ is also positive semi-definite. Therefore, $\inf\langle\mathbf{D}^*,\mathbf{B}\rangle=\sigma\cdot tr(\mathbf{D}^*)$, $\inf\langle\mathbf{D}_i,\mathbf{B}\rangle=\sigma\cdot tr(\mathbf{D}_i)$, and $\lim_{i\to\infty}\inf\langle\mathbf{D}_i,\mathbf{B}\rangle=\inf\langle\mathbf{D}^*,\mathbf{B}\rangle$.

3) For $\forall\,\zeta>\left(\sup_{\mathbf{C}\in\Theta}\langle\mathbf{D}_i,\mathbf{C}\rangle-\inf_{\mathbf{B}\in\mathcal{P}_\sigma^+}\langle\mathbf{D}_i,\mathbf{B}\rangle\right)/\|\mathbf{D}_i\|$, then

$$\inf_{\mathbf{B}\in\mathcal{P}_\sigma^+}\langle\mathbf{D}_i,\mathbf{B}\rangle>\sup_{\mathbf{C}\in\Theta}\langle\mathbf{D}_i,\mathbf{C}\rangle-\zeta\langle\mathbf{D}_i,\mathbf{D}_i\rangle/\|\mathbf{D}_i\|.$$

According to Lemma 2, $\mathcal{P}_\sigma^+\cap(\Theta-\zeta\mathbf{D}_i/\|\mathbf{D}_i\|)=\emptyset$, i.e. $\forall\mathbf{C}\in\Theta$, $\lambda_p(C-\zeta\mathbf{D}_i/\|\mathbf{D}_i\|)<\sigma$. By Weyl's inequalities (see [6]),

$$\begin{aligned}\lambda_p(C) &\leq \lambda_p(C-\zeta\mathbf{D}_i/\|\mathbf{D}_i\|)+\lambda_1(\zeta\mathbf{D}_i/\|\mathbf{D}_i\|)\\ &\leq \sigma+\zeta\lambda_1(\mathbf{D}_i)/\|\mathbf{D}_i\|.\end{aligned}\tag{20}$$

Combine (20) and (19), $\lambda_1(\mathbf{D}_i)=\sigma-\lambda_p(\mathbf{C}_i)$; with ζ tending to $\left(\sup_{\mathbf{C}\in\Theta}\langle\mathbf{D}_i,\mathbf{C}\rangle-\inf_{\mathbf{B}\in\mathcal{P}_\sigma^+}\langle\mathbf{D}_i,\mathbf{B}\rangle\right)/\|\mathbf{D}_i\|$, 3) holds.

4) Let $\mu\doteq diag(\Theta)$, $\delta\doteq d^{*2}\cdot\min((2\mu)^{-1},(2\sqrt{n})^{-1})$, and

$$N\doteq\log(\delta/(d_1-d^*))/\log\eta_\delta+1. \tag{21}$$

Then we have $\forall i>N$, $\|\mathbf{D}_i-\mathbf{D}^*\|<d^{*2}\min((2\mu)^{-1},(2\sqrt{n})^{-1})$ and $|tr(\mathbf{D}_i)-tr(\mathbf{D}^*)|\leq\sqrt{n}\|\mathbf{D}_i-\mathbf{D}^*\|<d^{*2}/2$. By Schwartz inequality,

$$\begin{aligned}\sup_{\mathbf{C}\in\Theta}\langle\mathbf{D}_i,\mathbf{C}\rangle &\leq \sup_{\mathbf{C}\in\Theta}\langle\mathbf{D}^*,\mathbf{C}\rangle+\sup_{\mathbf{C}\in\Theta}\langle\mathbf{D}-\mathbf{D}^*,\mathbf{C}\rangle\\ &< \sup_{\mathbf{C}\in\Theta}\langle\mathbf{D}^*,\mathbf{C}\rangle+d^{*2}/2\\ &\leq \inf_{\mathbf{B}\in\mathcal{P}_\sigma^+}\langle\mathbf{D}^*,\mathbf{B}\rangle-d^{*2}/2\\ &\leq \inf_{\mathbf{B}\in\mathcal{P}_\sigma^+}\langle\mathbf{D}_i,\mathbf{B}\rangle.\end{aligned}\tag{22}$$

5). By Weyl's inequalities, $\lambda_p(\mathbf{C}_i)=\lambda_p(\mathbf{B}_i-\mathbf{D}_i)\geq\lambda_p(\mathbf{B}_i)+\lambda_p(-\mathbf{D}_i)\geq\sigma-\rho(\mathbf{D}_i)\geq\sigma-\|\mathbf{D}_i\|$. Let

$$N_\epsilon=\log(\epsilon/d_1)/\log\eta_\epsilon. \tag{23}$$

If $d^*=0$, by (7)(16), then for $\forall i>N_\epsilon$, $\|\mathbf{D}_i\|\leq\epsilon$, so $\lambda_p(\mathbf{C}_i)\geq\sigma-\epsilon$. ■