A Unified Theorem on SDP Rank Reduction

Anthony Man–Cho So*, Yinyu Ye†, Jiawei Zhang‡

November 19, 2006

Abstract

We consider the problem of finding a low–rank approximate solution to a system of linear equations in symmetric, positive semidefinite matrices. Specifically, let $A_1, \ldots, A_m \in \mathbb{R}^{n \times n}$ symmetric, positive semidefinite matrices, and let $b_1, \ldots, b_m \geq 0$. We show that if there exists a symmetric, positive semidefinite matrix X to the following system of equations:

$$A_i \bullet X = b_i \qquad \text{for } i = 1, \ldots, m$$

then for any *fixed* $d = 1, \ldots, O(\log m)$, there exists an $X_0 \succeq 0$ of rank at most d such that:

$$\beta \cdot b_i \leq A_i \bullet X_0 \leq \alpha \cdot b_i \qquad \text{for } i = 1, \ldots, m$$

where:

$$\alpha = 1 + O\left(\frac{\log m}{d}\right), \quad \beta = \begin{cases} \Omega\left(m^{-2/d}\right) & \text{for } d = O\left(\dfrac{\log m}{\log\log m}\right) \\ \Omega\left((\log m)^{-3\log m/(d \log\log m)}\right) & \text{otherwise} \end{cases}$$

Moreover, such an X_0 can be found in randomized polynomial time. This complements a result of Barvinok [2] and provides a unified treatment of and generalizes several results in the literature [3, 6, 7, 8].

1 Introduction

In this note we consider the problem of finding a low–rank approximate solution to a system of linear equations in symmetric, positive semidefinite (psd) matrices. Specifically, let $A_1, \ldots, A_m \in \mathbb{R}^{n \times n}$ be symmetric psd matrices, and let $b_1, \ldots, b_m \geq 0$. Consider the following system of linear equations:

$$A_i \bullet X = b_i \quad \text{for } i = 1, \ldots, m; \;\; X \succeq 0, \text{ symmetric} \tag{1}$$

It is well–known [1] (see also [2, 9]) that if (1) is feasible, then there exists a solution $X \succeq 0$ of rank no more than $\sqrt{2m}$. However, in many applications, such as graph realization [10] and

*Department of Computer Science, Stanford University, Stanford, CA 94305. E–mail: `manchoso@cs.stanford.edu`

†Department of Management Science and Engineering, Stanford University, Stanford, CA 94305. E–mail: `yinyu-ye@stanford.edu`

‡IOMS–OM, Stern School of Business, New York University, New York, NY 10012. E–mail: `jzhang@stern.nyu.edu`

dimension reduction [7], it is desirable to have a low–rank solution, say, a solution of rank at most d, where $d \geq 1$ is fixed. Of course, such a low–rank solution may not exist, and even if it does exist, one may not be able to find it efficiently. Thus, it is natural to ask whether one can *efficiently* find an $X_0 \succeq 0$ of rank at most d (where $d \geq 1$ is fixed) such that X_0 satisfies (1) *approximately*, i.e.:

$$\beta(m,n,d) \cdot b_i \leq A_i \bullet X_0 \leq \alpha(m,n,d) \cdot b_i \qquad \text{for } i = 1, \ldots, m \tag{2}$$

for some functions $\alpha \geq 1$ and $\beta \in (0,1]$. The quality of the approximation will be determined by how close α and β are to 1. Our main result is the following:

Theorem 1 *Let $A_1, \ldots, A_m \in \mathbb{R}^{n\times n}$ be symmetric psd matrices, and let $b_1, \ldots, b_m \geq 0$. Suppose that there exists an $X \succeq 0$ such that $A_i \bullet X = b_i$ for $i = 1, 2, \ldots, m$. Let $r = \min\{\sqrt{2m}, n\}$. Then, for any $d \geq 1$, there exists an $X_0 \succeq 0$ with rank$(X_0) \leq d$ such that:*

$$\beta(m,n,d) \cdot b_i \leq A_i \bullet X_0 \leq \alpha(m,n,d) \cdot b_i \qquad \textit{for } i = 1, \ldots, m$$

where:

$$\alpha(m,n,d) = \begin{cases} 1 + \dfrac{12\log(4mr)}{d} & \textit{for } 1 \leq d \leq 12\log(4mr) \\[2ex] 1 + \sqrt{\dfrac{12\log(4mr)}{d}} & \textit{otherwise} \end{cases} \tag{3}$$

and

$$\beta(m,n,d) = \begin{cases} \dfrac{1}{5e} \cdot \dfrac{1}{m^{2/d}} & \textit{for } 1 \leq d \leq \dfrac{2\log m}{\log\log(2m)} \\[2ex] \dfrac{1}{4e} \cdot \dfrac{1}{\log^{f(m)/d}(2m)} & \textit{for } \dfrac{2\log m}{\log\log(2m)} < d \leq 4\log(4mr) \\[2ex] 1 - \sqrt{\dfrac{4\log(4mr)}{d}} & \textit{for } d > 4\log(4mr) \end{cases} \tag{4}$$

and $f(m) = \frac{3\log m}{\log\log(2m)}$. Moreover, such an X_0 can be found in randomized polynomial time.

Remarks:

(a) From the definition of r, we see that the bounds above can be made independent of n and the ranks of $A_1, \ldots, A_m$.

(b) Note that $f(m)/d \leq 3/2$ in the region $d > \frac{2\log m}{\log\log(2m)}$.

(c) If $\max_{1\leq i\leq m} \text{rank}(A_i) = O(1)$, then the lower bound can be sharpened to $\Omega\left(m^{-2/d}\right)$ for *all* $d \in \{1, \ldots, 4\log(4mr)\}$; see the proof of Proposition 2.

(d) The constants can be improved if we only consider one–sided inequalities.

It turns out that Theorem 1 provides a unified treatment of and generalizes several results in the literature:

(a) *(Metric Embedding)* Let ℓ_2^m be the space $\mathbb{R}^m$ equipped with the Euclidean norm, and let ℓ_2 be the space of infinite sequences $x = (x_1, x_2, \ldots)$ of real numbers such that $\|x\|_2 \equiv \left(\sum_{j\geq 1} |x_j|^2\right)^{1/2} < \infty$. Given an n–point set $V = \{v_1, \ldots, v_n\}$ in ℓ_2^m, we would like to embed it into a low–dimensional Euclidean space as faithfully as possible. Specifically, we say that a map $f : V \to \ell_2$ is an D–embedding (where $D \geq 1$) if there exists a number $r > 0$ such that for all $u, v \in V$, we have:

$$r \cdot \|u - v\|_2 \leq \|f(u) - f(v)\|_2 \leq D \cdot r \cdot \|u - v\|_2$$

The goal is to find an f such that D is as small as possible. It is known [3, 7] that for any *fixed* $d \geq 1$, an $O\left(n^{2/d}\left(d^{-1}\log n\right)^{1/2}\right)$–embedding into ℓ_2^d exists. We now show how to derive this result from Theorem 1. Let e_i be the i–th standard basis vector in ℓ_2^d, and define $E_{ij} = (e_i - e_j)(e_i - e_j)^T$ for $1 \leq i < j \leq n$. Let U be the $m \times n$ matrix whose i–th column is the vector v_i, where $i = 1, \ldots, n$. Then, it is clear that the matrix $X = U^T U$ satisfies the following system of equations:

$$E_{ij} \bullet X = \|v_i - v_j\|_2^2 \qquad \text{for } 1 \leq i < j \leq n$$

Now, Theorem 1 implies that we can find an $X_0 \succeq 0$ of rank at most d such that:

$$\Omega\left(n^{-4/d}\right) \cdot \|v_i - v_j\|_2^2 \leq E_{ij} \bullet X_0 \leq O\left(\frac{\log n}{d}\right) \cdot \|v_i - v_j\|_2^2 \qquad \text{for } 1 \leq i < j \leq n$$

Upon taking the Cholesky factorization $X_0 = U_0^T U_0$, we recover a set of points $u_1, \ldots, u_n \in \ell_2^d$ such that:

$$\Omega\left(n^{-2/d}\right) \cdot \|v_i - v_j\|_2 \leq \|u_i - u_j\|_2 \leq O\left(\sqrt{\frac{\log n}{d}}\right) \cdot \|v_i - v_j\|_2 \qquad \text{for } 1 \leq i < j \leq n$$

as desired. We should point out that by using different techniques, Matoušek [7] was able to show that in fact an $\Theta(n)$–embedding into ℓ_2^d exists for the cases where $d = 1, 2$.

We remark that if we do not restrict the dimension of the range of f, then by the Johnson–Lindenstrauss lemma [3, 4], for any $\epsilon > 0$, there exists an $(1+\epsilon)$–embedding of V into ℓ_2^d, where $d = O(\epsilon^{-2}\log n)$. In [2, Chapter V, Proposition 6.1], Barvinok generalizes this result and shows that if the assumptions of Theorem 1 are satisfied, then for any $\epsilon \in (0,1)$ and $d \geq 8\epsilon^{-2}\log(4m)$, there exists an $X_0 \succeq 0$ of rank at most d such that:

$$(1-\epsilon)b_i \leq A_i \bullet X_0 \leq (1+\epsilon)b_i \qquad \text{for } i = 1, \ldots, m$$

Thus, Theorem 1 complements Barvinok's result and generalizes the corresponding results in the study of bi–Lipschitz embeddings into low–dimensional Euclidean space [3, 7].

(b) *(Quadratic Optimization with Homogeneous Quadratic Constraints)* Consider the following optimization problems:

$$\begin{aligned} v^*_{maxqp} = \text{maximize} \quad & x^T A x \\ \text{subject to} \quad & x^T A_i x \leq 1 \qquad i = 1, \ldots, m \end{aligned} \tag{5}$$

$$\begin{aligned} v^*_{minqp} = \text{minimize} \quad & x^T A x \\ \text{subject to} \quad & x^T A_i x \geq 1 \qquad i = 1, \ldots, m \end{aligned} \tag{6}$$

where $A_1, \ldots, A_m$ are symmetric positive semidefinite matrices. Both of these problems arise from various applications (see [6, 8]) and are NP–hard. Their natural SDP relaxations are given by:

$$\begin{array}{rll} v^*_{maxsdp} = \text{maximize} & A \bullet X & \\ \text{subject to} & A_i \bullet X \leq 1 & i = 1, \ldots, m \\ & X \succeq 0 & \end{array} \tag{7}$$

$$\begin{array}{rll} v^*_{minsdp} = \text{minimize} & A \bullet X & \\ \text{subject to} & A_i \bullet X \geq 1 & i = 1, \ldots, m \\ & X \succeq 0 & \end{array} \tag{8}$$

It is clear that if $X = xx^T$ is a rank–1 feasible solution to (7) (resp. (8)), then x is a feasible solution to (5) (resp. (6)). Now, let X^*_{maxsdp} be an optimal solution to (7). It has been shown in [8] that one can extract a rank–1 matrix X_0 from X^*_{maxsdp} such that (i) X_0 is feasible to (7) and (ii) $A \bullet X_0 \geq \Omega\left(\frac{1}{\log m}\right) \cdot v^*_{maxqp}$. We now derive a similar result using Theorem 1. By definition, the matrix X^*_{maxsdp} satisfies the following system:

$$A \bullet X^*_{maxsdp} = v^*_{maxsdp}, \quad A_i \bullet X^*_{maxsdp} = b_i \leq 1 \qquad \text{for } i = 1, \ldots, m$$

As we shall see from the proof of Theorem 1, one can find a rank–1 matrix $X'_0 \succeq 0$ such that:

$$\mathbb{E}\left[A \bullet X'_0\right] = v^*_{maxsdp}, \quad A_i \bullet X'_0 \leq O(\log m) \cdot b_i \qquad \text{for } i = 1, \ldots, m$$

It follows that the matrix $X_0 = \Omega\left(\frac{1}{\log m}\right) \cdot X'_0 \succeq 0$ is feasible to (7), and that $\mathbb{E}\left[A \bullet X_0\right] = \Omega\left(\frac{1}{\log m}\right) \cdot v^*_{maxsdp} \geq \Omega\left(\frac{1}{\log m}\right) \cdot v^*_{maxqp}$.

In a similar fashion, if X^*_{minsdp} is an optimal solution to (8), then one can extract a rank–1 matrix $X'_0 \succeq 0$ from X^*_{minsdp} such that $X_0 = O(m^2) \cdot X'_0$ is feasible for (8) and $\mathbb{E}\left[A \bullet X_0\right] = O(m^2) \cdot v^*_{minqp}$, thus recovering a result of Luo et al. [6].

In [6] the authors also consider a complex version of (5) and (6), in which the matrices A and A_i are complex Hermitian and the components of the decision vector x can take on complex values. They show that if X^*_{maxsdp} (resp. X^*_{minsdp}) is an optimal solution to the corresponding SDP relaxation (7) (resp. (8)), then one can extract a complex rank–1 solution that achieves $\Omega\left(\frac{1}{\log m}\right)$ (resp. $O(m)$) times the optimum value. Our result shows that these bounds are also achievable for the real version of (7) and (8) if we allow the solution matrix to have rank at most 2.

2 Proof of the Main Result

We first make some standard preparatory moves (see, e.g., [2, 6, 8]). Let $X \succeq 0$ be a solution to the system (1). By a result of Barvinok [1] and Pataki [9], we may assume that $r_0 \equiv \text{rank}(X) <$

$\sqrt{2m}$. Let $X = UU^T$ for some $U \in \mathbb{R}^{n \times r_0}$, and set $A_i' = U^T A_i U \in \mathbb{R}^{r_0 \times r_0}$, where $i = 1, \ldots, m$. Then, we have $A_i' \succeq 0$, $\mathrm{rank}(A_i') \leq \min\{\mathrm{rank}(A_i), r_0\}$, and

$$b_i = A_i \bullet X = \left(U^T A_i U\right) \bullet I = A_i' \bullet I = \mathrm{Tr}(A_i')$$

Moreover, if $X_0' \succeq 0$ satisfies the inequalities:

$$\beta(m,n,d) \cdot b_i \leq A_i' \bullet X_0' \leq \alpha(m,n,d) \cdot b_i \qquad \text{for } i = 1, \ldots, m$$

then upon setting $X_0 = U X_0' U^T \succeq 0$, we see that $\mathrm{rank}(X_0) \leq \mathrm{rank}(X_0')$, and

$$A_i \bullet X_0 = \left(U^T A_i U\right) \bullet X_0' = A_i' \bullet X_0'$$

i.e. X_0 satisfies the inequalities in (2). Thus, in order to establish Theorem 1, it suffices to establish the following:

Theorem 1' *Let $A_1, \ldots, A_m \in \mathbb{R}^{n \times n}$ be symmetric psd matrices, where $n < \sqrt{2m}$. Then, for any $d \geq 1$, there exists an $X_0 \succeq 0$ with rank$(X_0) \leq d$ such that:*

$$\beta(m,n,d) \cdot Tr(A_i) \leq A_i \bullet X_0 \leq \alpha(m,n,d) \cdot Tr(A_i) \qquad \textit{for } i = 1, \ldots, m$$

where $\alpha(m,n,d)$ and $\beta(m,n,d)$ are given by (3) and (4), respectively.

The proof of Theorem 1' relies on the following estimates of a chi–square random variable.

Proposition 1 *Let $\xi_1, \ldots, \xi_n$ be i.i.d. standard Gaussian random variables. Let $\alpha \in (1, \infty)$ and $\beta \in (0,1)$ be constants, and set $U_n = \sum_{i=1}^n \xi_i^2$. Note that $U_n \sim \chi_n^2$. Then, the following hold:*

$$\Pr\left(U_n \geq \alpha^2 n\right) \leq \left[\alpha \exp\left(\frac{1-\alpha^2}{2}\right)\right]^n = \exp\left[\frac{n}{2}\left(1 - \alpha^2 + 2\log\alpha\right)\right] \tag{9}$$

$$\Pr\left(U_n \leq \beta^2 n\right) \leq \left[\beta \exp\left(\frac{1-\beta^2}{2}\right)\right]^n = \exp\left[\frac{n}{2}\left(1 - \beta^2 + 2\log\beta\right)\right] \tag{10}$$

Proof To establish (9), we let $t \in [0, 1/2)$ and compute:

$$\begin{aligned}
\Pr\left(U_n \geq \alpha^2 n\right) &= \Pr\left\{\exp\left[t\left(U_n - \alpha^2 n\right)\right] \geq 1\right\} \\
&\leq \mathbb{E}\left[\exp\left[t\left(U_n - \alpha^2 n\right)\right]\right] && \text{(by Markov's inequality)} \\
&= \exp\left(-t\alpha^2 n\right) \cdot \left(\mathbb{E}\left[\exp\left(t\xi_1^2\right)\right]\right)^n && \text{(by independence)} \\
&= \exp\left(-t\alpha^2 n\right) \cdot (1-2t)^{-n/2}
\end{aligned}$$

Let $f : [0, 1/2) \to \mathbb{R}$ be given by $f(t) = \exp\left(-t\alpha^2 n\right) \cdot (1-2t)^{-n/2}$. Then, we have:

$$f'(t) = -\exp\left(-t\alpha^2 n\right) \alpha^2 n (1-2t)^{-n/2} + \exp\left(-t\alpha^2 n\right) n (1-2t)^{-(n/2+1)}$$

and hence f is minimized at $t^* = \left(1 - \alpha^{-1}\right)/2$. Note that $t^* \in (0, 1/2)$ whenever $\alpha \in (1, \infty)$. Thus, we conclude that:

$$\Pr\left(U_n \geq \alpha^2 n\right) \leq f\left(t^*\right) = \left[\alpha \exp\left(\frac{1-\alpha^2}{2}\right)\right]^n$$

To establish (10), we proceed in a similar fashion. For $t \geq 0$, we have:

$$\begin{aligned}
\Pr\left(U_n \leq \beta^2 n\right) &= \Pr\left\{\exp\left[t\left(\beta^2 n - U_n\right)\right] \geq 1\right\} \\
&\leq \mathbb{E}\left[\exp\left[t\left(\beta^2 n - U_n\right)\right]\right] && \text{(by Markov's inequality)} \\
&= \exp\left(t\beta^2 n\right) \cdot \left(\mathbb{E}\left[\exp\left(-t\xi_1^2\right)\right]\right)^n && \text{(by independence)} \\
&= \exp\left(t\beta^2 n\right) \cdot (1+2t)^{-n/2}
\end{aligned}$$

Now, let $f : [0,\infty) \to \mathbb{R}$ be given by $f(t) = \exp\left(t\beta^2 n\right) \cdot (1+2t)^{-n/2}$. Then, we have:

$$f'(t) = \exp\left(t\beta^2 n\right)\beta^2 n(1+2t)^{-n/2} - \exp\left(t\beta^2 n\right) n(1+2t)^{-(n/2+1)}$$

and hence f is minimized at $t^* = \left(\beta^{-2} - 1\right)/2$. Moreover, we have $t^* > 0$ whenever $\beta < 1$. It follows that:

$$\Pr\left(U_n \leq \beta^2 n\right) \leq f\left(t^*\right) = \left[\beta \exp\left(\frac{1-\beta^2}{2}\right)\right]^n$$

as desired. □

In the sequel, let $d \geq 1$ be a given integer. Consider the following randomized procedure for generating an $X_0 \succeq 0$ of rank at most d:

Algorithm 1 Procedure GenSoln

Input: An integer $d \geq 1$.

Output: An psd matrix X_0 of rank at most d.

1: generate i.i.d. Gaussian random variables ξ_i^j with mean 0 and variance $1/d$, and define $\xi^j = (\xi_1^j, \ldots, \xi_n^j)$, where $i = 1, \ldots, n$; $j = 1, \ldots, d$

2: return $X_0 = \sum_{j=1}^d \xi^j \left(\xi^j\right)^T$

We remark that the above procedure is different from those in [6, 8]. Let $X_0 \succeq 0$ be the output of GenSoln. The following propositions form the heart of our analysis.

Proposition 2 *Let $H \in \mathbb{R}^{n\times n}$ be a symmetric positive semidefinite matrix. Consider the spectral decomposition $H = \sum_{k=1}^r \lambda_k v_k v_k^T$, where $r = \mathit{rank}(H)$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$. Set $\bar{\lambda}_k = \lambda_k/(\lambda_1 + \cdots + \lambda_r)$. Then, for any $\beta \in (0,1)$, we have:*

$$\Pr\left(H \bullet X_0 \leq \beta\, Tr(H)\right) \leq r \cdot \exp\left[\frac{d}{2}(1 - \beta + \log\beta)\right] \leq r \cdot \exp\left[\frac{d}{2}(1 + \log\beta)\right] \tag{11}$$

On the other hand, if β satisfies $e\beta\log r \leq 1/5$, then (11) can be sharpened to:

$$\Pr\left(H \bullet X_0 \leq \beta\, Tr(H)\right) \leq \left(\sqrt{\frac{5e\beta}{2}}\right)^d \tag{12}$$

Proof We first establish (11). Let $q_k = \sqrt{\lambda_k} \cdot v_k$. Then, we have $H = \sum_{k=1}^r q_k q_k^T$. Observe that $q_k^T \xi^j$ is a Gaussian random variable with mean 0 and variance $\sigma_k^2 \equiv d^{-1} \sum_l \left(q_k^T e_l\right)^2$, where e_l is the l–th coordinate vector. Moreover, we have:

$$\sum_{k=1}^r \sigma_k^2 = \frac{1}{d} \sum_{k=1}^r \sum_l \left(q_k^T e_l\right)^2 = \frac{1}{d} \cdot \mathrm{Tr}(H) \quad \text{and} \quad \mathbb{E}\left[\sum_{j=1}^d \left(q_k^T \xi^j\right)^2\right] = d \cdot \sigma_k^2$$

Hence, we conclude that:

$$\Pr\left(\sum_{j=1}^d \left(q_k^T \xi^j\right)^2 \le \beta d \sigma_k^2\right) = \Pr\left(U_d \le \beta d\right) \le \exp\left[\frac{d}{2}\left(1 - \beta + \log \beta\right)\right] \qquad \text{for } k = 1, \ldots, r$$

Now, observe that $H \bullet X_0 = \sum_{k=1}^r \sum_{j=1}^d \left(q_k^T \xi^j\right)^2$. Hence, we conclude that:

$$\Pr\left(H \bullet X_0 \le \beta \mathrm{Tr}(H)\right) \le \sum_{k=1}^r \Pr\left(\sum_{j=1}^d \left(q_k^T \xi^j\right)^2 \le \beta d \sigma_k^2\right) \le r \cdot \exp\left[\frac{d}{2}\left(1 - \beta + \log \beta\right)\right]$$

To establish (12), we proceed as follows. Clearly, we have $H \bullet X_0 = \sum_{k=1}^r \sum_{j=1}^d \lambda_k \left(v_k^T \xi^j\right)^2$. Now, observe that $u = \left(v_k^T \xi^j\right)_{k,j} \sim \mathcal{N}(0, d^{-1} I_{rd})$. Indeed, $v_k^T \xi^j$ is a Gaussian random variable, as it is the sum of Gaussian random variables. Moreover, we have:

$$\mathbb{E}\left[v_k^T \xi^j\right] = 0 \quad \text{and} \quad \mathbb{E}\left[\left(v_k^T \xi^j\right)\left(v_l^T \xi^{j'}\right)\right] = \frac{1}{d} \cdot v_k^T v_l = \frac{1}{d} \cdot \mathbf{1}_{\{k=l\}}$$

It follows that $H \bullet X_0$ has the same distribution as $\sum_{k=1}^r \sum_{j=1}^d \lambda_k \tilde{\xi}_{kj}^2$, where $\tilde{\xi}_{kj}$ are i.i.d. Gaussian random variables with mean 0 and variance $1/d$. Now, we compute:

$$\begin{aligned}
\Pr\left(H \bullet X_0 \le \beta \mathrm{Tr}(H)\right) &= \Pr\left(\sum_{k=1}^r \sum_{j=1}^d \lambda_k \tilde{\xi}_{kj}^2 \le \beta \sum_{k=1}^r \lambda_k\right) \\
&= \Pr\left(\sum_{k=1}^r \sum_{j=1}^d \bar{\lambda}_k \tilde{\xi}_{kj}^2 \le \beta\right) \qquad \left(\text{where } \bar{\lambda}_k = \lambda_k \Big/ \sum_{k=1}^r \lambda_k\right)
\end{aligned}$$

Define:

$$p\left(r, \bar{\lambda}, \beta\right) \equiv \Pr\left(\sum_{k=1}^r \sum_{j=1}^d \bar{\lambda}_k \tilde{\xi}_{kj}^2 \le \beta\right)$$

Then, by Proposition 1, we have:

$$p\left(r, \bar{\lambda}, \beta\right) \le \Pr\left(\sum_{k=1}^r \sum_{j=1}^d \bar{\lambda}_r \tilde{\xi}_{kj}^2 \le \beta\right) = \Pr\left(d \sum_{k=1}^r \sum_{j=1}^d \tilde{\xi}_{kj}^2 \le \frac{\beta}{r \bar{\lambda}_r} \cdot rd\right) \le \left(\frac{e\beta}{r \bar{\lambda}_r}\right)^{rd/2}$$

On the other hand, we have:

$$p\left(r,\bar{\lambda},\beta\right) \le \Pr\left(\sum_{k=1}^{r-1}\sum_{j=1}^{d}\bar{\lambda}_k\tilde{\xi}_{kj}^2 \le \beta\right) \le \Pr\left(\sum_{k=1}^{r-1}\sum_{j=1}^{d}\frac{\bar{\lambda}_k}{1-\bar{\lambda}_r}\tilde{\xi}_{kj}^2 \le \frac{\beta}{1-\bar{\lambda}_r}\right)$$

Now, observe that:

$$\frac{1}{1-\bar{\lambda}_r}\sum_{k=1}^{r-1}\bar{\lambda}_k = 1$$

whence:

$$p\left(r,\bar{\lambda},\beta\right) \le p\left(r-1,\frac{\bar{\lambda}_{1:r-1}}{1-\bar{\lambda}_r},\frac{\beta}{1-\bar{\lambda}_r}\right)$$

It then follows from an easy inductive argument that:

$$p\left(r,\bar{\lambda},\beta\right) \le \min_{1\le k\le r}\left\{\left(\frac{e\beta}{k\bar{\lambda}_k}\right)^{kd/2}\right\} \tag{13}$$

Let $\alpha = p\left(r,\bar{\lambda},\beta\right)^{2/d}$. Note that $\alpha \in (0,1)$. By (13), we have $\bar{\lambda}_k \le \left(k\alpha^{1/k}\right)^{-1}e\beta$ for $k=1,\ldots,r$. Upon summing over k and using the fact that $\sum_{k=1}^{r}\bar{\lambda}_k = 1$, we obtain:

$$\sum_{k=1}^{r}\frac{1}{k\alpha^{1/k}} \ge \frac{1}{e\beta} \tag{14}$$

If $r=1$, then we have $\alpha \le e\beta$. Henceforth, we shall assume that $r \ge 2$. Note that for any $\alpha \in (0,1)$, the function $t \mapsto (t\alpha^{1/t})^{-1}$ is decreasing for all $t \ge 1$, since we have:

$$\frac{d}{dt}\left(\frac{1}{t\alpha^{1/t}}\right) = \frac{\log\alpha - t}{t^3\alpha^{1/t}} < 0$$

Hence, it follows that:

$$\sum_{k=1}^{r}\frac{1}{k\alpha^{1/k}} \le \frac{1}{\alpha} + \int_1^r \frac{1}{t\alpha^{1/t}}\,dt = \frac{1}{\alpha} + \int_{\frac{\log(1/\alpha)}{r}}^{\log(1/\alpha)}\frac{e^t}{t}\,dt \tag{15}$$

where we use the change of variable $z = -t^{-1}\log(1/\alpha)$ in the last step. Using the expansion:

$$\frac{e^t}{t} = \frac{1}{t}\sum_{j\ge 0}\frac{t^j}{j!} = \frac{1}{t} + \sum_{j\ge 0}\frac{t^j}{(j+1)!}$$

we compute:

$$\begin{aligned}
\int_{\frac{\log(1/\alpha)}{r}}^{\log(1/\alpha)} \frac{e^t}{t}\,dt &= \log r + \sum_{j\ge 0} \frac{t^{j+1}}{(j+1)(j+1)!}\Big|_{\frac{\log(1/\alpha)}{r}}^{\log(1/\alpha)} \\
&= \log r + \sum_{j\ge 0} \frac{\log^{j+1}(1/\alpha)}{(j+1)(j+1)!}\left(1-\frac{1}{r^{j+1}}\right) \\
&\le \log r + \sum_{j\ge 0} \frac{\log^{j+1}(1/\alpha)}{(j+1)!} \\
&= \log r + \frac{1}{\alpha} - 1 \\
&\le \log r + \frac{1}{\alpha} \qquad (16)
\end{aligned}$$

Upon combining (14), (15) and (16), we conclude that:

$$\frac{1}{e\beta} \le \frac{2}{\alpha} + \log r$$

which, together with the assumption that $e\beta \log r \le 1/5$, implies that $\alpha \le 5e\beta/2$. □

Proposition 3 *Let $H \in \mathbb{R}^{n\times n}$ be a symmetric positive semidefinite matrix. Consider the spectral decomposition $H = \sum_{k=1}^r \lambda_k v_k v_k^T$, where $r = rank(H)$ and $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_r > 0$. Then, for any $\alpha > 1$, we have:*

$$\Pr\left(H \bullet X_0 \ge \alpha\, Tr(H)\right) \le r \cdot \exp\left[\frac{d}{2}\left(1-\alpha+\log\alpha\right)\right] \qquad (17)$$

Proof As before, let $q_k = \sqrt{\lambda_k} \cdot v_k$. Then, using the arguments in the proof of Proposition 2, we conclude that:

$$\Pr\left(\sum_{j=1}^{d} \left(q_k^T \xi^j\right)^2 \ge \alpha d \sigma_k^2\right) = \Pr(U_d \ge \alpha d) \le \exp\left[\frac{d}{2}\left(1-\alpha+\log\alpha\right)\right] \qquad \text{for } k = 1,\ldots,r$$

Now, observe that $H \bullet X_0 = \sum_{k=1}^r \sum_{j=1}^d \left(q_k^T \xi^j\right)^2$. Hence, we have:

$$\Pr\left(H \bullet X_0 \ge \alpha \mathrm{Tr}(H)\right) \le r \cdot \exp\left[\frac{d}{2}\left(1-\alpha+\log\alpha\right)\right] \qquad (18)$$

as desired. □

Proof of Theorem 1' We first establish the upper bound. We write $\alpha = 1+\alpha'$ for some $\alpha' > 0$. Using the inequality $\log(1+x) \le x - x^2/2 + x^3/3$, which is valid for all $x > 0$, it is easy to show that:

$$1-\alpha+\log\alpha = -\alpha' + \log(1+\alpha') \le \begin{cases} -\dfrac{\alpha'}{6} & \text{for } \alpha' \ge 1 \\[2ex] -\dfrac{\alpha'^2}{6} & \text{for } 0 < \alpha' < 1 \end{cases} \qquad (19)$$

Let $T = \frac{12\log(4mn)}{d}$. If $T \geq 1$, then set $\alpha' = T$; otherwise, set $\alpha' = \sqrt{T}$. In the former case, we have $\alpha' \geq 1$, and hence by Proposition 3 and the bound in (19), for each $i = 1, \ldots, m$, we have:

$$\Pr\left(A_i \bullet X_0 \geq \alpha \mathrm{Tr}(A_i)\right) \leq \mathrm{rank}(A_i) \cdot \exp\left(-\frac{d\alpha'}{12}\right) \leq \frac{1}{4m}$$

where the last inequality follows from the fact that $\mathrm{rank}(A_i) \leq n$. In the latter case, we have $\alpha' \in (0,1)$, and a similar calculation shows that:

$$\Pr\left(A_i \bullet X_0 \geq \alpha \mathrm{Tr}(A_i)\right) \leq \mathrm{rank}(A_i) \cdot \exp\left(-\frac{d\alpha'^2}{12}\right) \leq \frac{1}{4m}$$

for each $i = 1, \ldots, m$. Hence, we conclude that:

$$\Pr\left(A_i \bullet X_0 \leq \alpha(m,n,d) \cdot \mathrm{Tr}(A_i) \;\text{ for all } i = 1, \ldots, m\right) \geq 1 - \frac{1}{4} = \frac{3}{4} \tag{20}$$

where $\alpha(m,n,d)$ is given by (3). Next, we establish the lower bound. We consider the following cases:

<u>Case 1</u>: $1 \leq d \leq \frac{2\log m}{\log\log(2m)}$

Let $\beta = \left(5em^{2/d}\right)^{-1}$ in Proposition 2. Since $r < \sqrt{2m}$, we have:

$$e\beta \log r < \frac{1}{10m^{2/d}} \log 2m \leq \frac{1}{10} < \frac{1}{5}$$

by our choice of d. It follows that (12) of Proposition 2 applies, and we conclude that:

$$\Pr\left(A_i \bullet X_0 \leq \beta \mathrm{Tr}(A_i)\right) \leq \left(\frac{1}{2}\right)^{d/2} \cdot \frac{1}{m} \qquad \text{for } i = 1, \ldots, m$$

Together with (20), we have:

$$\Pr\left(\beta \mathrm{Tr}(A_i) \leq A_i \bullet X_0 \leq \alpha(m,n,d) \cdot \mathrm{Tr}(A_i) \;\text{ for all } i = 1, \ldots, m\right) \geq \frac{3}{4} - \left(\frac{1}{2}\right)^{d/2} > 0$$

for all $d \geq 1$.

<u>Case 2</u>: $\frac{2\log m}{\log\log(2m)} < d \leq 4\log(4mn)$

Suppose that $d = \frac{k\log m}{\log\log(2m)}$ for some $k > 2$. Let $\beta = \left(4e\log^{3/k}(2m)\right)^{-1}$ in Proposition 2. Upon noting that $m^{3/d} = \log^{3/k}(2m)$ and using (11) of Proposition 2, we have:

$$\Pr\left(A_i \bullet X_0 \leq \beta \mathrm{Tr}(A_i)\right) \leq \mathrm{rank}(A_i) \cdot \left(\sqrt{e\beta}\right)^d \leq \mathrm{rank}(A_i) \cdot \left(\frac{1}{2}\right)^d \cdot \frac{1}{m^{3/2}} < \sqrt{2} \cdot \left(\frac{1}{2}\right)^d \cdot \frac{1}{m}$$

Together with (20), we have:

$$\Pr\left(\beta \mathrm{Tr}(A_i) \leq A_i \bullet X_0 \leq \alpha(m,n,d) \cdot \mathrm{Tr}(A_i) \;\text{ for all } i = 1, \ldots, m\right) \geq \frac{3}{4} - \sqrt{2}\left(\frac{1}{2}\right)^d > 0$$

for all $d \geq 2$.

<u>**Case 3**</u>: $d > 4\log(4mn)$
We write $\beta = 1 - \beta'$ for some $\beta' \in (0,1)$. Using the inequality $\log(1-x) \leq -x - x^2/2$, which is valid for all $x \in [0,1]$, we have:

$$1 - \beta + \log\beta = \beta' + \log(1-\beta') \leq -\frac{\beta'^2}{2}$$

Let $\beta' = \left(\frac{4\log(4mn)}{d}\right)^{1/2}$. By assumption, we have $\beta' \in (0,1)$. By (11) of Proposition 2, for each $i = 1,\ldots,m$, we have:

$$\Pr\left(A_i \bullet X_0 \leq \beta \mathrm{Tr}(A_i)\right) \leq \mathrm{rank}(A_i) \cdot \exp\left(-\frac{d\beta'^2}{4}\right) \leq \frac{1}{4m}$$

It follows that:

$$\Pr\left(\beta\mathrm{Tr}(A_i) \leq A_i \bullet X_0 \leq \alpha(m,n,d)\cdot\mathrm{Tr}(A_i) \text{ for all } i = 1,\ldots,m\right) \geq \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$$

This completes the proof of Theorem 1'. □

3 A Refinement

In this section we show how Theorem 1' can be refined using the following set of estimates for a chi–square random variable:

Fact 1 *(Laurent, Massart [5]) Let $\xi_1,\ldots,\xi_n$ be i.i.d. standard Gaussian random variables. Let $a_1,\ldots,a_n \geq 0$, and set:*

$$|a|_\infty = \max_{1\leq i\leq n}|a_i|, \quad |a|_2^2 = \sum_{i=1}^n a_i^2$$

Define $V_n = \sum_{i=1}^n a_i(\xi_i^2 - 1)$. Then, for any $t > 0$, we have:

$$\Pr\left(V_n \geq \sqrt{2}|a|_2 t + |a|_\infty t^2\right) \leq e^{-t^2/2} \tag{21}$$

$$\Pr\left(V_n \leq -\sqrt{2}|a|_2 t\right) \leq e^{-t^2/2} \tag{22}$$

Fact 1 allows us to use the condition number of the given matrix H to compute the deviation probabilities in Propositions 2 and 3. To carry out this program, let us first recall some notations. Let H be a symmetric positive semidefinite matrix. Define $r = \mathrm{rank}(H)$, and let $K = \lambda_1/\lambda_r$ be the condition number of H, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ are the eigenvalues of H. Set $\bar{\lambda}_k = \lambda_k/(\lambda_1 + \cdots + \lambda_r)$. We then have the following proposition:

Proposition 4 *The following inequalities hold:*

(a) $\frac{1}{r} \leq |\bar{\lambda}|_\infty \leq \frac{K}{r}$;

(b) $|\bar{\lambda}|_2^2 \leq \frac{1}{r-1+K} + \frac{K(K-1)}{(r-1+K)^2}$;

(c) $\sqrt{1+\frac{r-1}{K^2}} \cdot |\bar{\lambda}|_\infty \le |\bar{\lambda}|_2$*;*

(d) $|\bar{\lambda}|_2^2 \le K|\bar{\lambda}|_\infty$.

Proof

(a) The first inequality follows from the fact that $\sum_{j=1}^r \bar{\lambda}_j = 1$. To establish the second inequality, suppose to the contrary that $|\bar{\lambda}|_\infty > K/r_i$. Then, we have $\bar{\lambda}_r > 1/r$, whence $\sum_{j=1}^r \bar{\lambda}_j > (r-1)/r + K/r > 1$, which is a contradiction.

(b) Let $\bar{\lambda}_r = x$. Then, we have $\bar{\lambda}_1 = Kx$. To bound $|\bar{\lambda}|_2^2$, we first observe that for $x < u \le v < Kx$ and $\epsilon \ge \min\{u-x, Kx-v\} > 0$, we have $(u-\epsilon)^2 + (v+\epsilon)^2 > u^2 + v^2$. This implies that the vector $\bar{\lambda}^*$ that maximizes $|\bar{\lambda}|_2^2$ satisfies $(r-1)\bar{\lambda}_r^* + K\bar{\lambda}_r^* = 1$, or equivalently, $\bar{\lambda}_r^* = \frac{1}{r-1+K}$. This in turn yields:

$$|\bar{\lambda}|_2^2 \le \frac{r-1}{(r-1+K)^2} + \frac{K^2}{(r-1+K)^2} = \frac{1}{r-1+K} + \frac{K(K-1)}{(r-1+K)^2}$$

as desired.

(c) We have:

$$\frac{|\bar{\lambda}|_2^2}{|\bar{\lambda}|_\infty^2} = 1 + \sum_{j=2}^r \frac{\bar{\lambda}_j^2}{\bar{\lambda}_1^2} \ge 1 + \frac{r-1}{K^2}$$

as desired.

(d) We compute:

$$\frac{|\bar{\lambda}|_2^2}{|\bar{\lambda}|_\infty} = \bar{\lambda}_1 + \sum_{j=2}^r \frac{\bar{\lambda}_j^2}{\bar{\lambda}_1} \le \bar{\lambda}_1 + (r-2)\bar{\lambda}_1 + \frac{\bar{\lambda}_1}{K^2} \le \frac{K}{r}\left(r-1+\frac{1}{K^2}\right) \le K$$

where we use the fact that $|\bar{\lambda}|_\infty = \bar{\lambda}_1 \le K/r$ in the second inequality.

$\square$

Using Fact 1 and Proposition 4, we obtain the following refinements to Theorem 1':

Theorem 2 *Under the setting of Theorem 1' and the additional assumptions that* $\min_i r_i = \Omega(\log m)$ *and* $K \equiv \max_{1\le i\le m} K_i = O(\sqrt{r})$*, the event:*

$$\{A_i \bullet X_0 \ge \Theta(1) \cdot Tr(A_i) \ \textit{for } i = 1, \ldots, m\}$$

occurs with constant probability.

Proof Using the fact that $\sum_{k=1}^{r}\sum_{j=1}^{d}\bar{\lambda}_k = d$ and setting $t = \frac{1-\beta}{\sqrt{2}} \cdot \frac{\sqrt{d}}{|\bar{\lambda}|_2}$, we conclude by (22) that:

$$
\begin{aligned}
\Pr\left(\sum_{k=1}^{r}\sum_{j=1}^{d}\bar{\lambda}_k\tilde{\xi}_{kj}^2 \le \beta\right) &= \Pr\left\{\sum_{k=1}^{r}\sum_{j=1}^{d}\bar{\lambda}_k\left(\tilde{\xi}_{kj}^2 - \frac{1}{d}\right) \le \beta - 1\right\} \\
&= \Pr\left\{\sum_{k=1}^{r}\sum_{j=1}^{d}\bar{\lambda}_k\left(\tilde{\xi}_{kj}^2 - \frac{1}{d}\right) \le -\sqrt{2}|\bar{\lambda}|_2 t\right\} \\
&\le \exp\left(-\frac{(1-\beta)^2}{4} \cdot \frac{d}{|\bar{\lambda}|_2^2}\right) \\
&\le \exp\left[-\Omega(\beta r_i)\right]
\end{aligned}
$$

Hence, by taking $\beta = \Theta(1)$, we conclude that:

$$
\Pr\left(A_i \bullet X_0 \le \Theta(1) \cdot \mathrm{Tr}(A_i)\right) = O(1/m) \qquad \text{for } i = 1, \ldots, m
$$

which in turn implies the theorem. □

Theorem 3 *Under the setting of Theorem 1' and the additional assumptions that* $\min_i r_i = \Omega(\log m)$ *and* $K \equiv \max_{1\le i\le m} K_i = O(1)$*, the event:*

$$
\{A_i \bullet X_0 \le \Theta(1) \cdot \mathit{Tr}(A_i) \ \textit{ for } i = 1, \ldots, m\}
$$

occurs with constant probability.

Proof Using the arguments in the proof of Theorem 2, we see that:

$$
\Pr\left(H \bullet X_0 \ge \alpha \mathrm{Tr}(H)\right) = \Pr\left\{\sum_{k=1}^{r}\sum_{j=1}^{d}\bar{\lambda}_k\left(\tilde{\xi}_{kj}^2 - \frac{1}{d}\right) \ge \alpha - 1\right\}
$$

Let

$$
t = \frac{\sqrt{|\bar{\lambda}|_2^2 + 2|\bar{\lambda}|_\infty(\alpha-1)} - |\bar{\lambda}|_2}{\sqrt{2}|\bar{\lambda}|_\infty} \tag{23}
$$

It then follows from (21) and the definition of t in (23) that:

$$
\Pr\left(H \bullet X_0 \ge \alpha \mathrm{Tr}(H)\right) \le \exp\left(-t^2/2\right) \tag{24}
$$

Note that:

$$
\begin{aligned}
t &= \frac{1}{\sqrt{2}} \cdot \frac{|\bar{\lambda}_i|_2}{|\bar{\lambda}_i|_\infty} \cdot \left(\sqrt{1 + 2\frac{|\bar{\lambda}_i|_\infty}{|\bar{\lambda}_i|_2^2}(\alpha-1)} - 1\right) && \text{(by equation (23))} \\
&\ge \frac{1}{\sqrt{2}} \cdot \sqrt{1 + \frac{r_i - 1}{K_i^2}} \cdot \left(\sqrt{1 + \frac{2(\alpha-1)}{K_i}} - 1\right) && \text{(by Proposition 4(c),(d))} \\
&= \Omega(\sqrt{\alpha r_i})
\end{aligned}
$$

Hence, by taking $\alpha = \Theta(1)$, we conclude that:

$$\Pr\left(A_i \bullet X_0 \geq \Theta(1) \cdot \mathrm{Tr}(A_i)\right) = O(1/m) \qquad \text{for } i = 1, \ldots, m$$

which in turn implies the theorem. $\square$

4 Conclusion

In this note we have considered the problem of finding a low–rank approximate solution to a system of linear equations in symmetric, positive semidefinite matrices. As we have demonstrated, our result provides a unified treatment of and generalizes several results in the literature. A main ingredient in our analysis is a set of tail estimates of a chi–squared random variable. We believe that these estimates could be of independent interest.

As a further illustration of our techniques, suppose that we are given positive semidefinite matrices A_k of rank r_k, where $k = 1, \ldots, K$. Consider a knapsack semidefinite matrix equality:

$$\sum_{k=1}^{K} A_k \bullet X_k = b, \quad X_k \succeq 0 \qquad \text{for } k = 1, \ldots, K$$

Our goal is to find a rank–one matrix $X_k^0 \succeq 0$ for each X_k such that:

$$\beta \cdot b \leq \sum_{k=1}^{K} A_k \bullet X_k^0 \leq \alpha \cdot b$$

Then, our result implies that the distortion rates would be on the order of $\log(K(\sum_k r_k))$ as opposed to $K(\sum_k r_k)$ obtained from the standard analysis where the terms are treated as $K(\sum_k r_k)$ independent equalities.

References

[1] Alexander I. Barvinok, *Problems of Distance Geometry and Convex Properties of Quadratic Maps*, Discrete and Computational Geometry 13:189–202, 1995.

[2] Alexander Barvinok, *A Course in Convexity*, Graduate Studies in Mathematics Volume 54, American Mathematical Society, 2002.

[3] Sanjoy Dasgupta, Anupam Gupta, *An Elementary Proof of the Johnson–Lindenstrauss Lemma*, Technical Report TR–99–06, International Computer Science Institute, Berkeley, CA, 1999.

[4] W. B. Johnson, J. Lindenstrauss, *Extensions of Lipschitz Mapping into Hilbert Space*, Contemporary Mathematics 26:189–206, 1984.

[5] B. Laurent, P. Massart, *Adaptive Estimation of a Quadratic Functional by Model Selection*, The Annals of Statistics 28(5):1302–1338, 2000.

[6] Zhi–Quan Luo, Nicholas D. Sidiropoulos, Paul Tseng, Shuzhong Zhang, *Approximation Bounds for Quadratic Optimization with Homogeneous Quadratic Constraints*, to appear in SIAM Journal on Optimization, 2006.

[7] Jiří Matoušek, *Bi–Lipschitz Embeddings into Low–Dimensional Euclidean Spaces*, Commentationes Mathematicae Universitatis Carolinae 31(3):589–600, 1990.

[8] A. Nemirovski, C. Roos, T. Terlaky, *On Maximization of Quadratic Form over Intersection of Ellipsoids with Common Center*, Mathematical Programming, Series A 86:465–473, 1999.

[9] Gábor Pataki, *On the Rank of Extreme Matrices in Semidefinite Programs and the Multiplicity of Optimal Eigenvalues*, Mathematics of Operations Research 23(2):339–358, 1998.

[10] Anthony Man–Cho So, Yinyu Ye, *A Semidefinite Programming Approach to Tensegrity Theory and Realizability of Graphs*, Proceedings of the 17th Annual ACM–SIAM Symposium on Discrete Algorithms, pp. 766–775, 2006.