A Unified Theorem on SDP Rank Reduction

Anthony Man–Cho So
Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong
email: manchoso@se.cuhk.edu.hk

Yinyu Ye
Department of Management Science and Engineering and, by courtesy, Electrical Engineering, Stanford University, Stanford, CA 94305, USA
email: yinyu-ye@stanford.edu

Jiawei Zhang
Department of Information, Operations, and Management Sciences, Stern School of Business, New York University, New York, NY 10012, USA
email: jzhang@stern.nyu.edu

We consider the problem of finding a low–rank approximate solution to a system of linear equations in symmetric, positive semidefinite matrices, where the approximation quality of a solution is measured by its maximum relative deviation, both above and below, from the prescribed quantities. We show that a simple randomized polynomial–time procedure produces a low–rank solution that has provably good approximation qualities. Our result provides a unified treatment of and generalizes several well–known results in the literature. In particular, it contains as special cases the Johnson–Lindenstrauss lemma on dimensionality reduction, results on low–distortion embeddings into low–dimensional Euclidean space, and approximation results on certain quadratic optimization problems.

Key words: Semidefinite Programming; Low–rank Matrices; Randomized Algorithm; Metric Embedding; Quadratic Optimization

MSC2000 Subject Classification: Primary: 90C22, 90C90; Secondary: 68W20, 68W25

OR/MS subject classification: Primary: Programming; Secondary: Mathematics

---

1. Introduction In this paper 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:

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

where $P \bullet Q = \mathrm{Tr}(P^T Q)$ is the Frobenius inner product of the two matrices P and Q. It is well–known (Barvinok [2]; see also Barvinok [3], Pataki [13]) 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 (So and Ye [14]) and dimensionality reduction (Matoušek [10], Weinberger and Saul [15]), 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 1.1 *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:*

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

where:

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

and

$$\beta(m,n,d) = \begin{cases} \dfrac{1}{e(2m)^{2/d}} & \text{for } 1 \le d \le 4\ln(2m) \\ \max\left\{ \dfrac{1}{e(2m)^{2/d}},\, 1-\sqrt{\dfrac{4\ln(2m)}{d}} \right\} & \text{for } d > 4\ln(2m) \end{cases} \tag{4}$$

Moreover, there exists an efficient randomized algorithm for finding such an X_0.

Before we discuss the proof and the applications of Theorem 1.1, several remarks are in order.

Remarks.

(i) While the upper bound (3) depends on the parameter r (which can be viewed as a worst–case bound on $\max_{1\le i\le m}\operatorname{rank}(A_i)$), the lower bound (4) does not have such a dependence.

(ii) From the definition of r, we see that the upper bound (3) can be made independent of n and the ranks of $A_1,\dots,A_m$.

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

2. Proof of the Main Result We first make some standard preparatory moves (see, e.g., Barvinok [3], Luo et al. [9], Nemirovski et al. [12]). Let $X \succeq 0$ be a solution to the system (1). By a result of Barvinok [2] and Pataki [13], we may assume that $r_0 \equiv \operatorname{rank}(X) < \sqrt{2m}$. Let $X = UU^T$ for some $U \in \mathbb{R}^{n\times r_0}$, and set $A_i' = U^TA_iU \in \mathbb{R}^{r_0\times r_0}$, where $i=1,\dots,m$. Then, we have $A_i' \succeq 0$, $\operatorname{rank}(A_i') \le \min\{\operatorname{rank}(A_i), r_0\}$, and

$$b_i = A_i \bullet X = \left(U^TA_iU\right) \bullet I = A_i' \bullet I = \operatorname{Tr}(A_i')$$

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

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

then upon setting $X_0 = UX_0'U^T \succeq 0$, we see that $\operatorname{rank}(X_0) \le \operatorname{rank}(X_0')$, and

$$A_i \bullet X_0 = \left(U^TA_iU\right) \bullet X_0' = A_i' \bullet X_0'$$

i.e. X_0 satisfies the inequalities in (2). Thus, in order to establish Theorem 1.1, it suffices to establish the following:

Theorem 1.1’ *Let $A_1,\dots,A_m \in \mathbb{R}^{n\times n}$ be symmetric psd matrices, where $n < \sqrt{2m}$. Then, for any $d \ge 1$, there exists an $X_0 \succeq 0$ with rank$(X_0) \le d$ such that:*

$$\beta(m,n,d)\cdot \mathit{Tr}(A_i) \le A_i \bullet X_0 \le \alpha(m,n,d)\cdot \mathit{Tr}(A_i) \qquad \text{for } i=1,\dots,m \tag{5}$$

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

In the sequel, let $d \ge 1$ be a given integer. The proof of Theorem 1.1’ involves analyzing the following simple randomized procedure:

Algorithm 1 Procedure GenSoln

Input: An integer $d \ge 1$.
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 $\xi^j = (\xi_1^j,\dots,\xi_n^j)$, where $i=1,\dots,n$; $j=1,\dots,d$
2: return $X_0 = \sum_{j=1}^d \xi^j\left(\xi^j\right)^T$

Let X_0 be the output of Procedure GenSoln. Clearly, we have $X_0 \succeq 0$ and $\operatorname{rank}(X_0) \le d$. Moreover, for any $H \in \mathbb{R}^{n\times n}$, we have $\mathbb{E}\left[H \bullet X_0\right] = \operatorname{Tr}(H)$. We now claim that X_0 satisfies (5) with constant probability. This is established via the following propositions, which form the heart of our analysis.

Proposition 2.1 *Let $H \in \mathbb{R}^{n\times n}$ be a symmetric psd matrix of rank at least* 1. *Then, for any $\beta \in (0,1)$, we have:*

$$\Pr\left(H \bullet X_0 \le \beta\, \mathit{Tr}(H)\right) \le \exp\left[\frac{d}{2}\left(1-\beta+\ln\beta\right)\right] \le \exp\left[\frac{d}{2}\left(1+\ln\beta\right)\right] \tag{6}$$

PROOF. Consider the spectral decomposition $H = \sum_{k=1}^r \lambda_k v_k v_k^T$, where $r \equiv \text{rank}(H) \geq 1$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$. Then, 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:

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

Since uncorrelated Gaussian random variables are independent, 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 $1/d$. In particular, we have:

$$\Pr\left(H \bullet X_0 \leq \beta \text{Tr}(H)\right) = \Pr\left(\sum_{k=1}^r \sum_{j=1}^d \lambda_k \tilde{\xi}_{kj}^2 \leq \beta \sum_{k=1}^r \lambda_k\right) = \Pr\left(\sum_{k=1}^r \sum_{j=1}^d \bar{\lambda}_k \tilde{\xi}_{kj}^2 \leq \beta\right)$$

where $\bar{\lambda}_k = \lambda_k / (\lambda_1 + \cdots + \lambda_r)$ for $k = 1, \ldots, r$. Now, we compute:

$$\begin{aligned}
\Pr\left(\sum_{k=1}^r \sum_{j=1}^d \bar{\lambda}_k \tilde{\xi}_{kj}^2 \leq \beta\right) &= \Pr\left(\exp\left(-t \sum_{k=1}^r \sum_{j=1}^d \bar{\lambda}_k \tilde{\xi}_{kj}^2\right) \geq \exp(-t\beta)\right) && \text{(for all } t \geq 0) \\
&\leq \exp(t\beta) \cdot \mathbb{E}\left[\exp\left(-t \sum_{k=1}^r \sum_{j=1}^d \bar{\lambda}_k \tilde{\xi}_{kj}^2\right)\right] && \text{(by Markov's inequality)} \\
&= \exp(t\beta) \cdot \prod_{k=1}^r \left(\mathbb{E}\left[\exp\left(-t\bar{\lambda}_k \tilde{\xi}_{11}^2\right)\right]\right)^d && \text{(by independence)}
\end{aligned}$$

Recall that for a standard Gaussian random variable ξ, we have $\mathbb{E}\left[\exp\left(-t\xi^2\right)\right] = (1+2t)^{-1/2}$ for all $t \geq 0$. Thus, it follows that:

$$\Pr\left(H \bullet X_0 \leq \beta \text{Tr}(H)\right) \leq \exp(t\beta) \cdot \prod_{k=1}^r \left(1 + \frac{2t\bar{\lambda}_k}{d}\right)^{-d/2} = \exp(t\beta) \cdot \exp\left[-\frac{d}{2} \sum_{k=1}^r \ln\left(1 + \frac{2t\bar{\lambda}_k}{d}\right)\right]$$

Now, note that for any fixed $t \geq 0$, the function $g_t : \mathbb{R}^r \to \mathbb{R}$ defined by $g_t(x) = -(d/2) \sum_{k=1}^r \ln(1 + 2tx_k/d)$ is convex. Hence, its maximum over the simplex $\{x \in \mathbb{R}^r : \sum_{k=1}^r x_i = 1, \, x \geq 0\}$ is attained at a vertex of the simplex. This implies that:

$$\Pr\left(H \bullet X_0 \leq \beta \text{Tr}(H)\right) \leq \exp\left[t\beta - \frac{d}{2} \ln\left(1 + \frac{2t}{d}\right)\right]$$

It is easy to show that the function $t \mapsto \exp(t\beta - (d/2)\ln(1 + 2t/d))$ is minimized at $t^* = d(1-\beta)/2\beta$. Moreover, we have $t^* > 0$ whenever $\beta \in (0,1)$. It follows that:

$$\Pr\left(H \bullet X_0 \leq \beta \text{Tr}(H)\right) \leq \exp\left[\frac{d}{2}(1 - \beta + \ln\beta)\right]$$

as desired. □

PROPOSITION 2.2 *Let $H \in \mathbb{R}^{n \times n}$ be a symmetric psd matrix with $r \equiv$ rank$(H) \geq 1$. Then, for any $\alpha > 1$, we have:*

$$\Pr\left(H \bullet X_0 \geq \alpha \, Tr(H)\right) \leq r \cdot \exp\left[\frac{d}{2}(1 - \alpha + \ln\alpha)\right] \tag{7}$$

PROOF. Consider the rank–1 decomposition $H = \sum_{k=1}^r q_k q_k^T$. Then, we have $H \bullet X_0 = \sum_{k=1}^r \sum_{j=1}^d \left(q_k^T \xi^j\right)^2$. 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 $\sum_{k=1}^r \sigma_k^2 = \frac{1}{d} \sum_{k=1}^r \sum_l \left(q_k^T e_l\right)^2 = \frac{1}{d} \cdot \text{Tr}(H)$. It follows that:

$$\Pr\left(H \bullet X_0 \geq \alpha \text{Tr}(H)\right) = \Pr\left(\sum_{k=1}^r \sum_{j=1}^d \left(q_k^T \xi^j\right)^2 \geq \alpha d \sum_{k=1}^r \sigma_k^2\right) \leq \sum_{k=1}^r \Pr\left(\sum_{j=1}^d \left(q_k^T \xi^j\right)^2 \geq \alpha d \sigma_k^2\right) \tag{8}$$

To bound the last quantity, we first note that $\mathbb{E}\left[\sum_{j=1}^d \left(q_k^T \xi^j\right)^2\right] = d \cdot \sigma_k^2$. Hence, for any $t \in [0,1/2)$ and $k = 1,\ldots,r$, we have:

$$\begin{aligned}\Pr\left(\sum_{j=1}^d \left(q_k^T \xi^j\right)^2 \geq \alpha d \sigma_k^2\right) &= \Pr\left[\exp\left(t\sum_{j=1}^d \sigma_k^{-2}\left(q_k^T \xi^j\right)^2\right) \geq \exp(t\alpha d)\right] \\ &\leq \exp(-t\alpha d) \cdot (1-2t)^{-d/2}\end{aligned}$$

Now, the function $t \mapsto \exp\left(-t\alpha d\right) \cdot (1-2t)^{-d/2}$ is minimized at $t^* = (\alpha - 1)/2\alpha$. Moreover, we have $t^* \in (0,1/2)$ whenever $\alpha \in (1,\infty)$. It follows that:

$$\Pr\left(\sum_{j=1}^d \left(q_k^T \xi^j\right)^2 \geq \alpha d \sigma_k^2\right) \leq \alpha^{d/2} \cdot \exp\left(-\frac{d(\alpha-1)}{2}\right) = \exp\left[\frac{d}{2}(1-\alpha+\ln\alpha)\right] \tag{9}$$

Upon combining (8) and (9), we obtain:

$$\Pr\left(H \bullet X_0 \geq \alpha \mathrm{Tr}(H)\right) \leq r \cdot \exp\left[\frac{d}{2}\left(1-\alpha+\ln\alpha\right)\right]$$

as desired. □

Remarks. The reader may wonder why we do not follow the proof of Proposition 2.1 and get rid of the extra factor of r in (7). Indeed, following the argument in Proposition 2.1, we have:

$$\Pr\left(H \bullet X_0 \geq \alpha \mathrm{Tr}(H)\right) \leq \exp\left[-t\alpha - \frac{d}{2}\ln\left(1-\frac{2t}{d}\right)\right] \tag{10}$$

for all $t \in [0,1/2)$. Now, the quantity on the right–hand side is minimized at $t^* = d(\alpha-1)/2\alpha$. If $d = 1$, then we have $t^* \in (0,1/2)$, whence we obtain the following improvement over (7):

$$\Pr\left(H \bullet X_0 \geq \alpha \mathrm{Tr}(H)\right) \leq \exp\left[\frac{1}{2}(1-\alpha+\ln\alpha)\right]$$

However, if $d \geq 2$, then we have $t^* \geq 1/2$ whenever $\alpha \geq d/(d-1)$. In particular, if $d \geq 2$ and $\alpha \geq d/(d-1)$, then the minimum of the function $t \mapsto -t\alpha - (d/2)\ln(1-2t/d)$ over $[0,1/2]$ occurs at $t^* = 1/2$. Upon substituting this into (10), we have:

$$\Pr\left(H \bullet X_0 \geq \alpha \mathrm{Tr}(H)\right) \leq \exp\left[-\frac{1}{2}\left(\alpha + d\ln\left(1-\frac{1}{d}\right)\right)\right] \leq 2\exp(-\alpha/2)$$

which can be inferior to (7) in the applications that we are interested in.

Proof of Theorem 1.1'. We first establish the lower bound. Let $\beta = \left(e(2m)^{2/d}\right)^{-1}$. Note that $\beta \in (0,1)$ for all $d \geq 1$. Hence, by Proposition 2.1, we have:

$$\Pr\left(A_i \bullet X_0 \leq \beta \mathrm{Tr}(A_i)\right) \leq \exp\left[\frac{d\ln(e\beta)}{2}\right] = \frac{1}{2m} \qquad \text{for } i = 1,\ldots,m$$

which implies that:

$$\Pr\left(A_i \bullet X_0 \geq \frac{1}{e(2m)^{2/d}} \cdot \mathrm{Tr}(A_i) \ \text{ for all } i = 1,\ldots,m\right) \geq \frac{1}{2} \tag{11}$$

On the other hand, if $d > 4\ln(2m)$, then we can obtain an alternative bound as follows. Write $\beta = 1 - \beta'$ for some $\beta' \in (0,1)$. Using the inequality $\ln(1-x) \leq -x - x^2/2$, which is valid for all $x \in [0,1]$, we have:

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

Now, let $\beta' = \sqrt{\frac{4\ln(2m)}{d}}$. Since $d > 4\ln(2m)$, we have $\beta' \in (0,1)$. It then follows from Proposition 2.1 that:

$$\Pr\left(A_i \bullet X_0 \leq \beta \mathrm{Tr}(A_i)\right) \leq \exp\left(-\frac{d\beta'^2}{4}\right) = \frac{1}{2m} \qquad \text{for } i = 1,\ldots,m$$

which in turn implies that:

$$\Pr\left(A_i \bullet X_0 \geq \left(1 - \sqrt{\frac{4\ln(2m)}{d}}\right) \cdot \mathrm{Tr}(A_i) \text{ for all } i = 1, \ldots, m\right) \geq \frac{1}{2} \tag{12}$$

Upon combining (11) and (12), we obtain:

$$\Pr\left(A_i \bullet X_0 \geq \beta(m,n,d) \cdot \mathrm{Tr}(A_i) \text{ for all } i = 1, \ldots, m\right) \geq \frac{1}{2} \tag{13}$$

where $\beta(m,n,d)$ is given by (4).

Next, we establish the upper bound. We write $\alpha = 1 + \alpha'$ for some $\alpha' > 0$. Using the inequality $\ln(1+x) \leq x - x^2/2 + x^3/3$, which is valid for all $x > 0$, it is easy to show that:

$$1 - \alpha + \ln\alpha = -\alpha' + \ln(1+\alpha') \leq \begin{cases} -\dfrac{\alpha'}{6} & \text{for } \alpha' \geq 1 \\ -\dfrac{\alpha'^2}{6} & \text{for } 0 < \alpha' < 1 \end{cases} \tag{14}$$

Let $T = \frac{12\ln(4mn)}{d}$. 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 2.2 and the bound in (14), for $i = 1, \ldots, m$, 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 $i = 1, \ldots, m$. Hence, we conclude that:

$$\Pr\left(A_i \bullet X_0 \leq \alpha(m,n,d) \cdot \mathrm{Tr}(A_i) \text{ for all } i = 1, \ldots, m\right) \geq 1 - \frac{1}{4} = \frac{3}{4} \tag{15}$$

where $\alpha(m,n,d)$ is given by (3).

Finally, upon combining (13) and (15), we conclude that:

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

This completes the proof of Theorem 1.1'. □

3. Some Applications of the Main Result It turns out that Theorem 1.1 provides a unified treatment of and generalizes several results in the literature. These results in turn give some indication on the sharpness of the bounds derived in Theorem 1.1:

(i) *(Metric Embedding)* Let ℓ_2^p be the space $\mathbb{R}^p$ 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^p, 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 (Dasgupta and Gupta [4], Matoušek [10]) that for any *fixed* $d \geq 1$, an $O\left(n^{2/d}\left(d^{-1}\log n\right)^{1/2}\right)$–embedding[1] into ℓ_2^d exists. We now show how to derive this result from Theorem 1.1. 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 $m \times n$

[1]Given two functions $f, g : \mathbb{R}_+ \to \mathbb{R}_+$, we say that (i) $f(x) = O(g(x))$ if there exist constants $c > 0$ and $x_0 > 0$ such that $f(x) \leq c \cdot g(x)$ for all $x \geq x_0$, (ii) $f(x) = \Omega(g(x))$ if there exist constants $c > 0$ and $x_0 > 0$ such that $f(x) \geq c \cdot g(x)$ for all $x \geq x_0$; (iii) $f(x) = \Theta(g(x))$ if we have both $f(x) = O(g(x))$ and $f(x) = \Omega(g(x))$.

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:

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

Now, Theorem 1.1 implies that we can find an $X_0 \succeq 0$ of rank at most d such that:

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

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:

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

as desired. Clearly, any improvements on either (3) or (4) will immediately yield an improved bound on D for embeddings into ℓ_2^d. On the other hand, for any $d \ge 1$, there exists an n–point set V in ℓ_2^{d+1} such that *any* embedding of V into ℓ_2^d requires $D = \Omega\left(n^{1/\lfloor (d+1)/2 \rfloor}\right)$ (Matoušek [11]). We should also point out that by using different techniques, Matoušek [10] was able to show that in fact an $\Theta(n)$–embedding into ℓ_2^d exists for the cases where $d = 1, 2$.

If we do not restrict the dimension of the range of f, then by the Johnson–Lindenstrauss lemma (Johnson and Lindenstrauss [7], Dasgupta and Gupta [4]), for any $\epsilon > 0$ and any n–point set V in ℓ_2, there exists an $(1+\epsilon)$–embedding of V into ℓ_2^d, where $d = O(\epsilon^{-2} \log n)$. In Barvinok [3, Chapter V, Proposition 6.1], the author generalizes this result and shows that if the assumptions of Theorem 1.1 are satisfied, then for any $\epsilon \in (0,1)$ and $d \ge 8\epsilon^{-2} \log(4m)$, there exists an $X_0 \succeq 0$ of rank at most d such that:

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

Thus, Theorem 1.1 complements Barvinok's result and generalizes the corresponding results in the study of bi–Lipschitz embeddings into low–dimensional Euclidean space (Dasgupta and Gupta [4], Matoušek [10]). We remark that Alon [1] has shown that the dependence of d on ϵ in the Johnson–Lindenstrauss lemma is almost tight. Specifically, there exists an n–point set V in ℓ_2 such that for any $\epsilon \in (n^{-1/2}, 1/2)$, say, an $(1+\epsilon)$–embedding of V into ℓ_2^d will require $d = \Omega((\epsilon^2 \log(1/\epsilon))^{-1} \log n)$.

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

$$\begin{array}{rll} v^*_{maxqp} = \text{maximize} & x^T A x & \\ \text{subject to} & x^T A_i x \le 1 & i = 1, \ldots, m \end{array} \tag{16}$$

$$\begin{array}{rll} v^*_{minqp} = \text{minimize} & x^T A x & \\ \text{subject to} & x^T A_i x \ge 1 & i = 1, \ldots, m \end{array} \tag{17}$$

where $A_1, \ldots, A_m$ are symmetric psd matrices. Both of these problems arise from various applications (see Luo et al. [9], Nemirovski et al. [12]) and are NP–hard. Their natural SDP relaxations are given by:

$$\begin{array}{rll} v^*_{maxsdp} = \text{maximize} & A \bullet X & \\ \text{subject to} & A_i \bullet X \le 1 & i = 1, \ldots, m \\ & X \succeq 0 & \end{array} \tag{18}$$

$$\begin{array}{rll} v^*_{minsdp} = \text{minimize} & A \bullet X & \\ \text{subject to} & A_i \bullet X \ge 1 & i = 1, \ldots, m \\ & X \succeq 0 & \end{array} \tag{19}$$

It is clear that if $X = xx^T$ is a rank–1 feasible solution to (18) (resp. (19)), then x is a feasible solution to (16) (resp. (17)). Now, consider the maximization problem (16) and its SDP relaxation (18), where the objective matrix A can be indefinite. Suppose that X^*_{maxsdp} is an optimal solution to (18). It has been shown in Nemirovski et al. [12] that one can extract a rank–1 matrix X_0 from X^*_{maxsdp} such that (i) X_0 is always feasible for (18) and (ii) $A \bullet X_0 \ge \Omega\left(\frac{1}{\log m}\right) \cdot v^*_{maxqp}$

with high probability. We now derive a similar result using Theorem 1.1 and some tools from probability theory. Let $r = \mathrm{rank}(X^*_{maxsdp})$, and let $U \in \mathbb{R}^{n\times r}$ be such that $X^*_{maxsdp} = UU^T$. By definition, the matrix X^*_{maxsdp} satisfies the following system:

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

Thus, by Theorem 1.1, the rank–1 matrix $X'_0 = U\xi\xi^T U^T \succeq 0$, where $\xi \in \mathbb{R}^r$ is a standard Gaussian random vector, satisfies:

$$A_i \bullet X'_0 \le O(\log m) \cdot b_i \qquad \text{for } i = 1, \ldots, m \tag{20}$$

with high probability, say, at least 49/50 (cf. (15)). Moreover, we have:

$$\mathbb{E}\left[A \bullet X'_0\right] = A \bullet X^*_{maxsdp} = v^*_{maxsdp}$$

Note that $v^*_{maxsdp} \ge v^*_{maxqp} \ge 0$, since $x = 0$ is a feasible solution to (16). Now, in order to recover the result of Nemirovski et al. [12], it suffices to show that:

$$\Pr\left(A \bullet X'_0 \ge v^*_{maxsdp}\right) > \frac{1}{25} \tag{21}$$

Indeed, (20) and (21) together would imply that the matrix $X'_0 \succeq 0$ satisfies the following system:

$$A \bullet X'_0 \ge v^*_{maxsdp}, \quad A_i \bullet X'_0 \le O(\log m) \cdot b_i \qquad \text{for } i = 1, \ldots, m$$

with probability at least $1-(1/50+24/25) = 1/50$. It then follows that with constant probability, the matrix $X_0 = \Omega\left(\frac{1}{\log m}\right) \cdot X'_0 \succeq 0$ is feasible for (18), and that:

$$A \bullet X_0 \ge \Omega\left(\frac{1}{\log m}\right) \cdot v^*_{maxsdp} \ge \Omega\left(\frac{1}{\log m}\right) \cdot v^*_{maxqp}$$

as required. We remark that the gap between v^*_{maxsdp} and v^*_{maxqp} can be as large as $\Omega(\log m)$; see Nemirovski et al. [12].

In order to prove (21), we proceed as follows. First, observe that:

$$\begin{aligned}\Pr\left(A \bullet X'_0 \ge v^*_{maxsdp}\right) &= \Pr\left(\sum_{i=1}^r \sum_{j=1}^r (U^T A U)_{ij}\xi_i\xi_j \ge v^*_{maxsdp}\right) \\ &= \Pr\left(\sum_{1\le i<j\le r} (U^T A U)_{ij}\xi_i\xi_j \ge 0\right)\end{aligned}$$

Now, let

$$Y = \sum_{1\le i<j\le r} w_{ij}\xi_i\xi_j \qquad \text{where } w_{ij} = \left(\sum_{1\le i<j\le r}\left[(U^TAU)_{ij}\right]^2\right)^{-1/2} (U^TAU)_{ij}$$

Note that $\mathbb{E}[Y] = 0$ and $\mathbb{E}\left[Y^2\right] = 1$. The following fact shows that we can bound the probability $\Pr(Y \ge 0)$ using bounds on the moments of Y:

FACT 3.1 *(He et al. [6]) Let Z be a random variable such that $\mathbb{E}[Z] = 0$, $\mathbb{E}\left[Z^2\right] = 1$, and $\mathbb{E}\left[|Z|^p\right] \le \tau$ for some $p > 2$ and $\tau > 0$. Then, we have:*

$$\Pr(Z \ge 0) > \frac{1}{4}\tau^{-\frac{2}{p-2}}$$

For $p = 3$, the bound above can be sharpened to:

$$\Pr(Z \ge 0) \ge \frac{8(-5+\sqrt{7})^2}{27(1+\sqrt{7})}\tau^{-2}$$

It turns out that the problem of estimating the moments of Y has been extensively studied in the literature. For instance, we have the following theorem:

FACT 3.2 *(Gnedenko [5]) Let $\xi_1, \ldots, \xi_r$ be i.i.d. standard Gaussian random variables, and let w_{ij}, where $1 \le i < j \le r$, be any real numbers. Then, for any $p \ge 3$, we have:*

$$\mathbb{E}\left[\left|\sum_{1\le i<j\le r} w_{ij}\xi_i\xi_j\right|^p\right] \le 2^{-p/2}\mathbb{E}\left[|U_1 - 1|^p\right] \tag{22}$$

where U_1 is a chi–square random variable with one degree of freedom.

We remark that Fact 3.2 was originally stated for i.i.d. Bernoulli random variables (i.e. $\xi_i = \pm 1$ with equal probability for $i = 1, \ldots, r$). However, an application of the Central Limit Theorem immediately yields the version stated above.

Now, by Fact 3.2, we have:

$$\mathbb{E}\left[|Y|^3\right] \le \frac{1}{2^{3/2}\cdot(2\pi)^{1/2}}\left(\int_0^1 x^{-1/2}e^{-x/2}(1-x)^3\,dx + \int_1^\infty x^{-1/2}e^{-x/2}(x-1)^3\,dx\right) \tag{23}$$

We first bound:

$$\int_0^1 x^{-1/2}e^{-x/2}(1-x)^3\,dx \le \int_0^1 x^{-1/2}(1-x)^3\,dx = \frac{32}{35} \tag{24}$$

Next, by using integration by parts, we compute:

$$\begin{aligned}\int_1^\infty x^{-1/2}e^{-x/2}(x-1)^3\,dx &= \int_1^\infty \left(x^{5/2} - 3x^{3/2} + 3x^{1/2} - x^{-1/2}\right)\,dx \\ &= 24e^{-1/2} + 8\int_1^\infty x^{-1/2}e^{-x/2}\,dx\end{aligned}$$

Since $x^{-1/2}e^{-x/2} \le e^{-x/2}/2$ for $x \ge 4$, we bound:

$$\begin{aligned}\int_1^\infty x^{-1/2}e^{-x/2}\,dx &\le \int_1^4 x^{-1/2}e^{-x/2}\,dx + \int_4^\infty \frac{e^{-x/2}}{2}\,dx \\ &\le \int_1^2 e^{-x/2}\,dx + \int_2^4 \frac{e^{-x/2}}{\sqrt{2}}\,dx + e^{-2} \\ &= 2\left(e^{-1/2} - e^{-1}\right) + \sqrt{2}\left(e^{-1} - e^{-2}\right) + e^{-2}\end{aligned}$$

It follows that:

$$\int_1^\infty x^{-1/2}e^{-x/2}(x-1)^3\,dx \le 40e^{-1/2} - 8\sqrt{2}(\sqrt{2}-1)e^{-1} - 8(\sqrt{2}-1)e^{-2} < \frac{221}{10} \tag{25}$$

Upon substituting (24) and (25) into (23), we have:

$$\mathbb{E}\left[|Y|^3\right] < \frac{1}{2^{3/2}\cdot(2\pi)^{1/2}}\cdot\frac{1611}{70}$$

whence by Fact 3.1 we conclude that:

$$\Pr(Y \ge 0) > \frac{1}{25}$$

thus establishing (21). Incidentally, our bound in (21) is slightly stronger than the one established in He et al. [6].

Let us now turn our attention to the minimization problem (17) and its SDP relaxation (19). We assume that the objective matrix A is psd, so that $v^*_{minqp} \ge v^*_{minsdp} \ge 0$. Let $X^*_{minsdp} = UU^T$ be an optimal solution to (19), where $U \in \mathbb{R}^{n\times r}$ and $r = \operatorname{rank}(X^*_{minsdp})$. Let $\xi \in \mathbb{R}^r$ be a standard Gaussian random vector. Then, by Theorem 1.1, the rank–1 matrix $X_0' = U\xi\xi^TU^T \succeq 0$ satisfies:

$$A_i \bullet X_0' \ge \Omega\left(m^{-2}\right)\cdot\left(A_i \bullet X^*_{minsdp}\right) \qquad \text{for } i = 1, \ldots, m$$

with high probability. Moreover, we have $\mathbb{E}\left[A \bullet X_0'\right] = v^*_{minsdp}$. Now, by Markov's inequality (since $A \bullet X_0'$ is a non–negative random variable), we have:

$$\Pr\left(A \bullet X_0' \le 2v^*_{minsdp}\right) \ge \frac{1}{2}$$

It follows that with constant probability, the matrix $X_0 = O(m^2) \cdot X_0' \succeq 0$ is feasible for (19), and that $A \bullet X_0 \leq O(m^2) \cdot v^*_{minqp}$. In particular, we have recovered a result of Luo et al. [9]. We remark that the gap between v^*_{minqp} and v^*_{minsdp} can be as large as $\Omega(m^2)$; see Luo et al. [9].

In Luo et al. [9] the authors also considered a complex versions of (16) and (17), in which the matrices A and A_i are complex Hermitian psd 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 (18) (resp. (19)), 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 versions of (18) and (19) if we allow the solution matrix to have rank at most 2. In particular, the complex versions of (16) and (17) with real symmetric psd A and A_i's (i.e. only the decision vector takes on complex values) correspond precisely to the real versions of (18) and (19) with a rank–2 constraint on X.

4. A Refinement of the Main Result In this section we show how Theorem 1.1' can be refined using the following set of estimates for a chi–square random variable:

FACT 4.1 *(Laurent and Massart [8]) 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:

$$\Pr\left(V_n \geq \sqrt{2}|a|_2 t + |a|_\infty t^2\right) \leq e^{-t^2/2} \tag{26}$$

$$\Pr\left(V_n \leq -\sqrt{2}|a|_2 t\right) \leq e^{-t^2/2} \tag{27}$$

Fact 4.1 allows us to use the condition number of the given matrix H to compute the deviation probabilities in Propositions 2.1 and 2.2. To carry out this program, let us first recall some notations. Let H be a symmetric psd 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 4.1 *The following inequalities hold:*

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

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

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

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

PROOF.

(i) 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.

(ii) Let $\bar{\lambda}_r = x$. Then, we have $\bar{\lambda}_1 = Kx$. To bound $|\bar{\lambda}|_2^2$, we first observe that for $x < u \leq v < Kx$ and $\epsilon \geq \min\{u - x, Kx - v\} > 0$, 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:

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

as desired.

(iii) We have:

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

as desired.

(iv) We compute:

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

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

□

Using Fact 4.1 and Proposition 4.1, we obtain the following refinements to Theorem 1.1':

THEOREM 4.1 *Let $m \geq 2$, and let r_i be the rank of A_i, where $i = 1, \ldots, m$. Set $K \equiv \max_{1 \leq i \leq m} K_i$. Under the setting of Theorem 1.1' and the additional assumption that $\min_i r_i \geq 32K^2 \ln m$, the event:*

$$\left\{ A_i \bullet X_0 \geq \left(1 - \frac{1}{2d}\right) Tr(A_i) \text{ for all } i = 1, \ldots, m \right\}$$

occurs with constant probability.

PROOF. Let $\bar{\lambda}_1^i \geq \bar{\lambda}_2^i \geq \cdots \geq \bar{\lambda}_{r_i}^i > 0$ be the normalized eigenvalues of A_i, where $i = 1, \ldots, m$. Using the fact that $\sum_{k=1}^{r_i} \sum_{j=1}^{d} \bar{\lambda}_k^i = d$ and setting $t_i = \frac{1-\beta}{\sqrt{2}} \cdot \frac{d}{|\bar{\lambda}^i|_2}$, we have:

$$\begin{aligned}
\Pr\left(A_i \bullet X_0 \leq \beta \mathrm{Tr}(A_i)\right) &= \Pr\left(\sum_{k=1}^{r_i}\sum_{j=1}^{d} \bar{\lambda}_k^i \left(\tilde{\xi}_{kj}^2 - \frac{1}{d}\right) \leq \beta - 1\right) \\
&= \Pr\left(\sum_{k=1}^{r_i}\sum_{j=1}^{d} \bar{\lambda}_k^i \left(d\tilde{\xi}_{kj}^2 - 1\right) \leq -\sqrt{2}|\bar{\lambda}^i|_2 t_i\right) \\
&\leq \exp\left(-\frac{(1-\beta)^2}{4} \cdot \frac{d^2}{|\bar{\lambda}^i|_2^2}\right) && \text{(by (27))} \\
&\leq \exp\left(-\frac{(1-\beta)^2}{4} \cdot \frac{d^2 r_i}{K^2}\right) && \text{(by Proposition 4.1(i),(iv))}
\end{aligned}$$

Set $\beta = 1 - \frac{1}{2d} \in (0,1)$. Then, since $r_i \geq 32K^2 \ln m$, we have $\Pr\left(A_i \bullet X_0 \leq \beta \mathrm{Tr}(A_i)\right) \leq \frac{1}{m^2}$. It follows that:

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

as required. □

THEOREM 4.2 *Let $m \geq 2$, and let r_i be the rank of A_i, where $i = 1, \ldots, m$. Set $K \equiv \max_{1 \leq i \leq m} K_i$. Under the setting of Theorem 1.1' and the additional assumption that $\min_i r_i \geq 3 \ln m$, the event:*

$$\left\{ A_i \bullet X_0 \leq \left(1 + \frac{K}{2d}(2K+1)^2\right) Tr(A_i) \text{ for all } i = 1, \ldots, m \right\}$$

occurs with constant probability.

PROOF. Using the notations and arguments in the proof of Theorem 4.1, we see that:

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

for $i = 1, \ldots, m$. Now, let

$$t_i = \frac{\sqrt{|\bar{\lambda}^i|_2^2 + 2|\bar{\lambda}^i|_\infty(\alpha - 1)d} - |\bar{\lambda}^i|_2}{\sqrt{2}|\bar{\lambda}^i|_\infty} \qquad \text{for } i = 1, \ldots, m \tag{28}$$

It then follows from (26) and the definition of t_i in (28) that:

$$\Pr\left(A_i \bullet X_0 \geq \alpha \mathrm{Tr}(A_i)\right) \leq \exp\left(-t_i^2/2\right) \tag{29}$$

Upon letting $\alpha = 1 + \frac{K}{2d}(2K+1)^2$, we have:

$$\begin{aligned}
t_i &= \frac{1}{\sqrt{2}} \cdot \frac{|\bar{\lambda}^i|_2}{|\bar{\lambda}^i|_\infty} \cdot \left(\sqrt{1 + 2\frac{|\bar{\lambda}^i|_\infty}{|\bar{\lambda}^i|_2^2}(\alpha - 1)d} - 1 \right) && \text{(by equation (28))} \\
&\geq \frac{1}{\sqrt{2}} \cdot \sqrt{1 + \frac{r_i - 1}{K_i^2}} \cdot \left(\sqrt{1 + \frac{2(\alpha - 1)d}{K_i}} - 1 \right) && \text{(by Proposition 4.1(iii),(iv))} \\
&\geq \frac{1}{\sqrt{2}} \cdot \sqrt{\frac{r_i - 1}{K^2}} \cdot 2K \\
&\geq \sqrt{2(3\ln m - 1)} && \text{(since } r_i \geq 3\ln m \text{ for } i = 1, \ldots, m)
\end{aligned}$$

for $i = 1, \ldots, m$. It follows from (29) that:

$$\Pr\left(A_i \bullet X_0 \geq \alpha \mathrm{Tr}(A_i)\right) \leq \frac{e}{m^3} \qquad \text{for } i = 1, \ldots, m$$

whence:

$$\Pr\left(A_i \bullet X_0 \leq \alpha \mathrm{Tr}(A_i) \;\text{ for all } i = 1, \ldots, m\right) \geq 1 - \frac{e}{m^2} \geq \frac{1}{4}$$

as required. □

5. Conclusion In this paper we have considered the problem of finding a low–rank approximate solution to a system of linear equations in symmetric, positive semidefinite matrices. Our result provides a unified treatment of and generalizes several well–known results in the literature. As a further illustration, suppose that we are given symmetric psd 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 $\ln(K(\sum_k r_k))$, 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.

We also remark that our result can be applied to the following standard form SDP:

$$\min \; C \bullet X \quad \text{subject to } A_i \bullet X = b_i \quad \text{for } i = 1, \ldots, m; \;\; X \succeq 0 \tag{30}$$

Indeed, first recall that the dual of (30) is given by:

$$\max \sum_{i=1}^{m} b_i y_i \quad \text{subject to } \sum_{i=1}^{m} y_i A_i + S = C; \;\; S \succeq 0 \tag{31}$$

When $\bar{X}$ is optimal for (30), then under certain regularity conditions there will be a feasible solution $(\bar{S}, \bar{y})$ to the dual (31) such that $\bar{S} \bullet \bar{X} = 0$. Thus, in rounding the SDP solution $\bar{X}$ into a lower rank one, we can include the equality constraint $\bar{S} \bullet X = 0$. In particular, our rounding method will yield a low–rank X_0 such that $\bar{S} \bullet X_0 = 0$, i.e. X_0 is optimal for a “nearby” problem of the original SDP.

Acknowledgment. Part of this work was done when the first author was a PhD student in the Computer Science Department at Stanford University. Research by the first and second author was supported in part by NSF grant DMS–0604513. We would like to express our gratitude to the anonymous referees for their careful reading of the manuscript. In particular, we thank the referee who pointed out an erroneous argument in the original manuscript regarding our application of the main theorem to recover Nemirovski et al. and Luo et al.'s results on quadratic optimization. We also thank the referee who suggested the current version of Proposition 2.1, which improves upon our previous version.

References

[1] Noga Alon, *Problems and Results in Extremal Combinatorics I*, Discrete Mathematics **273** (2003), 31–53.

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

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

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

[5] B. D. Gnedenko, *The Best Constant in an Inequality of Khinchin Type for Quadratic Forms*, Russian Mathematical Surveys **55(2)** (2000), 340–341.

[6] Simai He, Zhi–Quan Luo, Jiawang Nie, Shuzhong Zhang, *Semidefinite Relaxation Bounds for Indefinite Homogeneous Quadratic Optimization*, Technical Report SEEM2007–01, Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, 2007; to appear in SIAM Journal on Optimization.

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

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

[9] Zhi–Quan Luo, Nicholas D. Sidiropoulos, Paul Tseng, Shuzhong Zhang, *Approximation Bounds for Quadratic Optimization with Homogeneous Quadratic Constraints*, SIAM Journal on Optimization **18(1)** (2007), 1–28.

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

[11] Jiří Matoušek, *Lectures on Discrete Geometry*, Springer–Verlag, 2002.

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

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

[14] 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 (2006), 766–775.

[15] Kilian Q. Weinberger, Lawrence K. Saul, *An Introduction to Nonlinear Dimensionality Reduction by Maximum Variance Unfolding*, Proceedings of the 21st National Conference on Artificial Intelligence, 2006.