# SUPPLEMENTARY MATERIAL FOR "QUADRO: A SUPERVISED DIMENSION REDUCTION METHOD VIA RAYLEIGH QUOTIENT OPTIMIZATION"

By Jianqing Fan*, Zheng Tracy Ke*, Han Liu† and Lucy Xia*

*Princeton University*

In this supplement we present the technical proofs and numerical tables for simulation, for the main article. Section A contains proofs of Propositions 2.1, 5.1 and 6.2. Section B contains proofs of Lemmas A.1, 9.1, 9.2 and 9.3. Section C contains numerical tables corresponding to Figure 3 and Figure 4 in the main article.

## APPENDIX A

**A.1. Proof of Proposition 2.1.** We first present a lemma which is proved in Section B.

Lemma A.1. *If $\boldsymbol{U}$ follows a uniform distribution on $\mathcal{S}^{d-1}$, for any $d \times d$ diagonal matrix $\mathbf{S}$ and any vector $\boldsymbol{\beta} \in \mathbb{R}^d$, we have*

- $\mathbb{E}(\boldsymbol{U}^\top \mathbf{S} \boldsymbol{U}) = \frac{\mathrm{tr}(\mathbf{S})}{d}$, $\mathbb{E}[(\boldsymbol{U}^\top \mathbf{S} \boldsymbol{U})^2] = \frac{2\,\mathrm{tr}(\mathbf{S}^2)+[\mathrm{tr}(\mathbf{S})]^2}{d^2+2d}$;
- $\mathbb{E}(\boldsymbol{U}^\top \boldsymbol{\beta}) = 0$, $\mathbb{E}[(\boldsymbol{U}^\top \boldsymbol{\beta})^2] = \frac{\|\boldsymbol{\beta}\|^2}{d}$;
- $\mathbb{E}(\boldsymbol{U}^\top \mathbf{S} \boldsymbol{U} \boldsymbol{U}^\top \boldsymbol{\beta}) = 0$.

Now, we show the claim of Proposition 2.1. Let $\boldsymbol{Y} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{Z} - \boldsymbol{\mu})$, then $\boldsymbol{Y} = \xi \boldsymbol{U}$ where $\boldsymbol{U}$ follows a uniform distribution on $\mathcal{S}^{d-1}$ and is independent of $\xi$. The quadratic form $Q(\boldsymbol{Z})$ can be rewritten as

$$
\begin{aligned}
Q(\boldsymbol{Z}) &= \boldsymbol{Z}^\top \boldsymbol{\Omega} \boldsymbol{Z} - 2\boldsymbol{\delta}^\top \boldsymbol{Z} \\
&= \boldsymbol{Y}^\top \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Omega} \boldsymbol{\Sigma}^{1/2} \boldsymbol{Y} + 2\boldsymbol{Y}^\top \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Omega}\boldsymbol{\mu} - \boldsymbol{\delta}) + \boldsymbol{\mu}^\top \boldsymbol{\Omega} \boldsymbol{\mu} - 2\boldsymbol{\mu}^\top \boldsymbol{\delta} \\
&= \bar{Q}(\boldsymbol{Y}) + c,
\end{aligned}
$$

where $c = \boldsymbol{\mu}^\top \boldsymbol{\Omega} \boldsymbol{\mu} - 2\boldsymbol{\mu}^\top \boldsymbol{\delta}$. Therefore, $\mathbb{E}[Q(\boldsymbol{Z})] = \mathbb{E}[\bar{Q}(\boldsymbol{Y})] + c$ and $\mathrm{var}[Q(\boldsymbol{Z})] = \mathrm{var}[\bar{Q}(\boldsymbol{Y})]$.

*Supported in part by NSF Grant DMS-1206464 and NIH Grants R01GM100474 and R01-GM072611.

†Supported in part by NSF Grant III-1116730, NSF III-1332109, an NIH sub-award and a FDA sub-award from Johns Hopkins University, and an NIH-subaward from Harvard University.

Furthermore, we let
$$\mathbf{\Sigma}^{1/2}\mathbf{\Omega}\mathbf{\Sigma}^{1/2} = \mathbf{K}\mathbf{S}\mathbf{K}^\top$$
be the eigenvalue decomposition of $\mathbf{\Sigma}^{1/2}\mathbf{\Omega}\mathbf{\Sigma}^{1/2}$, where $\mathbf{K}$ is an orthogonal matrix and $\mathbf{S}$ is a diagonal matrix. We also define
$$\boldsymbol{\beta} = \mathbf{K}^\top\mathbf{\Sigma}^{1/2}(\mathbf{\Omega}\boldsymbol{\mu} - \boldsymbol{\delta}).$$
Notice that $\mathbf{Y}^\top\mathbf{\Sigma}^{1/2}\mathbf{\Omega}\mathbf{\Sigma}^{1/2}\mathbf{Y} = \xi^2\boldsymbol{U}^\top\mathbf{K}\mathbf{S}\mathbf{K}^\top\boldsymbol{U} = \xi^2\boldsymbol{U}_1^\top\mathbf{S}\boldsymbol{U}_1$, where $\boldsymbol{U}_1 = \mathbf{K}^\top\boldsymbol{U}$. Since $\mathbf{K}$ is an orthogonal matrix, $\boldsymbol{U}_1$ follows the same distribution as $\boldsymbol{U}$ and is also independent of $\xi$. Moreover, we can write $\boldsymbol{Y}^\top\mathbf{\Sigma}^{1/2}(\mathbf{\Omega}\boldsymbol{\mu} - \boldsymbol{\delta}) = \xi\boldsymbol{U}_1^\top\boldsymbol{\beta}$. To save notation, we still use $\boldsymbol{U}$ to represent $\boldsymbol{U}_1$. It follows that
$$\bar{Q}(\boldsymbol{Y}) = \xi^2\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U} + 2\xi\boldsymbol{U}^\top\boldsymbol{\beta}.$$

Let's calculate $\mathbb{E}[\bar{Q}(\boldsymbol{Y})]$ first.
$$\begin{aligned}\mathbb{E}[\bar{Q}(\boldsymbol{Y})] &= \mathbb{E}(\xi^2)\mathbb{E}(\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U}) + 2\mathbb{E}(\xi)\mathbb{E}[\boldsymbol{U}^\top\boldsymbol{\beta}]\\ &= \frac{\mathbb{E}(\xi^2)}{d}\operatorname{tr}(\mathbf{S}) = \operatorname{tr}(\mathbf{\Omega}\mathbf{\Sigma}).\end{aligned}$$
The first equality is due to the fact that $\xi$ and $\boldsymbol{U}$ are independent; the second equality is from Lemma A.1; and the last inequality is because $\mathbb{E}(\xi^2) = d$ and $\operatorname{tr}(\mathbf{S}) = \operatorname{tr}(\mathbf{\Sigma}^{1/2}\mathbf{\Omega}\mathbf{\Sigma}^{1/2}) = \operatorname{tr}(\mathbf{\Omega}\mathbf{\Sigma})$. It follows that
$$\mathbb{E}[Q(\boldsymbol{Z})] = \mathbb{E}[\bar{Q}(\boldsymbol{Y})] + c = \operatorname{tr}(\mathbf{\Omega}\mathbf{\Sigma}) + \boldsymbol{\mu}^\top\mathbf{\Omega}\boldsymbol{\mu} - 2\boldsymbol{\mu}^\top\boldsymbol{\delta}.$$

Next, we calculate $\operatorname{var}[\bar{Q}(\mathbf{Y})]$. It follows that
$$\begin{aligned}\operatorname{var}[\bar{Q}(\mathbf{Y})] &= \operatorname{var}(\xi^2\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U} + 2\xi\boldsymbol{U}^\top\boldsymbol{\beta})\\ &= \operatorname{var}(\xi^2\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U}) + 4\operatorname{var}(\xi\boldsymbol{U}^\top\boldsymbol{\beta}) + 4\operatorname{cov}(\xi^2\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U}, \xi\boldsymbol{U}^\top\boldsymbol{\beta}).\end{aligned}$$
Let's look at them term by term. First,
$$\begin{aligned}\operatorname{var}(\xi^2\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U}) &= \mathbb{E}[\xi^4(\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U})^2] - \mathbb{E}^2(\xi^2\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U})\\ &= \mathbb{E}(\xi^4)\mathbb{E}[(\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U})^2] - \mathbb{E}^2(\xi^2)\mathbb{E}^2(\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U})\\ &= \mathbb{E}(\xi^4)\frac{2\operatorname{tr}(\mathbf{S}^2) + \operatorname{tr}^2(\mathbf{S})}{2d + d^2} - \mathbb{E}^2(\xi^2)\frac{\operatorname{tr}^2(\mathbf{S})}{d^2}\\ &= 2(\gamma+1)\operatorname{tr}(\mathbf{S}^2) + \gamma\operatorname{tr}^2(\mathbf{S})\\ &= 2(\gamma+1)\operatorname{tr}(\mathbf{\Omega}\mathbf{\Sigma}\mathbf{\Omega}\mathbf{\Sigma}) + \gamma[\operatorname{tr}(\mathbf{\Omega}\mathbf{\Sigma})]^2.\end{aligned}$$
The third equality comes from Lemma A.1; the last equality follows from the fact that $\operatorname{tr}(\mathbf{S}^2) = \operatorname{tr}(\mathbf{\Sigma}^{1/2}\mathbf{\Omega}\mathbf{\Sigma}\mathbf{\Omega}\mathbf{\Sigma}^{1/2}) = \operatorname{tr}(\mathbf{\Omega}\mathbf{\Sigma}\mathbf{\Omega}\mathbf{\Sigma})$. Second,
$$\operatorname{var}(\xi\boldsymbol{U}^\top\boldsymbol{\beta}) = \mathbb{E}(\xi^2(\boldsymbol{U}^\top\boldsymbol{\beta})^2) - \mathbb{E}^2(\xi\boldsymbol{U}^\top\boldsymbol{\beta})$$

$$= \mathbb{E}(\xi^2)\mathbb{E}[(\boldsymbol{U}^\top \boldsymbol{\beta})^2] - \mathbb{E}^2(\xi)\mathbb{E}^2(\boldsymbol{U}^\top \boldsymbol{\beta})$$
$$= \mathbb{E}(\xi^2)\frac{\|\boldsymbol{\beta}\|^2}{d} = (\boldsymbol{\Omega\mu} - \boldsymbol{\delta})^\top \boldsymbol{\Sigma}(\boldsymbol{\Omega\mu} - \boldsymbol{\delta}).$$

In the last equality, we have used $\mathbb{E}(\xi^2) = d$, $\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Omega\mu} - \boldsymbol{\delta})$ and $\mathbf{K}^\top\mathbf{K} = \mathbf{I}_d$. Last,

$$\begin{aligned}\mathrm{cov}(\xi^2\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U}, \xi\boldsymbol{U}^\top\boldsymbol{\beta}) &= \mathbb{E}(\xi^3)\mathbb{E}(\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\beta}) - \mathbb{E}(\xi^2)\mathbb{E}(\mathbf{U}^\top\mathbf{S}\mathbf{U})\mathbb{E}(\xi\mathbf{U}^\top\boldsymbol{\beta})\\ &= \mathbb{E}(\xi^3)\mathbb{E}(\boldsymbol{U}^\top\mathbf{S}\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\beta}) = 0\end{aligned}$$

Combining the above gives

$$\begin{aligned}\mathrm{var}[Q(\boldsymbol{Z})] = \mathrm{var}[\bar{Q}(\boldsymbol{Y})] &= 2(\gamma+1)\,\mathrm{tr}(\boldsymbol{\Omega\Sigma\Omega\Sigma}) + \gamma[\mathrm{tr}(\boldsymbol{\Omega\Sigma})]^2\\ &\quad + 4(\boldsymbol{\Omega\mu} - \boldsymbol{\delta})^\top\boldsymbol{\Sigma}(\boldsymbol{\Omega\mu} - \boldsymbol{\delta}).\end{aligned}$$

□

**A.2. Proof of Proposition 5.1.** For any $d \times 1$ vector $\mathbf{v}$ and $d \times d$ matrix $\mathbf{A}$, we denote by $\mathrm{Supp}(\mathbf{v})$ the support of $\mathbf{v}$, which is contained in $\{1, \cdots, d\}$, and by $\mathrm{Supp}(\mathbf{A})$ the support of $\mathbf{A}$, which is contained in $\{1, \cdots, d\} \times \{1, \cdots, d\}$. Let $\theta = \sqrt{1 + v_1(1+\gamma)/2} > 1$ and

$$c = (1+\kappa)\min\left\{(1+\gamma){v_2}^2,\ 4(1 - 1/\theta^2)v_2\right\}.$$

The claim then becomes $\Theta(S, 0) \geq c$, or in other words,

$$\mathbf{x}^\top\mathbf{Q}\mathbf{x} \geq c|\mathbf{x}|^2, \qquad \text{when } \mathrm{Supp}(\mathbf{x}) \subset S.$$

First, using (13) and Lemma 9.1, we find that for each $\mathbf{x}$, there exits unique $(\boldsymbol{\Omega}, \boldsymbol{\delta})$ such that $\mathbf{x} = \mathbf{x}(\boldsymbol{\Omega}, \boldsymbol{\delta})$ and $\mathbf{x}^\top\mathbf{Q}\mathbf{x} = L(\boldsymbol{\Omega}, \boldsymbol{\delta})$. Second, by definition of $U'$ and $V$, $\mathrm{Supp}(\mathbf{x}) \subset S$ implies that $\mathrm{Supp}(\boldsymbol{\Omega}) \subset U' \times U'$ and $\mathrm{Supp}(\boldsymbol{\delta}) \subset V$. Therefore, it suffices to show

(A.1)
$$L(\boldsymbol{\Omega}, \boldsymbol{\delta}) \geq c(|\boldsymbol{\Omega}|^2 + |\boldsymbol{\delta}|^2), \qquad \text{when } \mathrm{Supp}(\boldsymbol{\Omega}) \subset U' \times U' \text{ and } \mathrm{Supp}(\boldsymbol{\delta}) \subset V.$$

Now, we show (A.1). From (5) and that $\gamma \geq 0$,

$$L_k(\boldsymbol{\Omega}, \boldsymbol{\delta}) \geq 2(1+\gamma)\,\mathrm{tr}(\boldsymbol{\Omega\Sigma}_k\boldsymbol{\Omega\Sigma}_k) + 4(\boldsymbol{\Omega\mu}_k - \boldsymbol{\delta})^\top\boldsymbol{\Sigma}_k(\boldsymbol{\Omega\mu}_k - \boldsymbol{\delta}), \qquad k = 1, 2.$$

Let $\widetilde{\boldsymbol{\Omega}}$ be the submatrix of $\boldsymbol{\Omega}$ by restricting rows and columns to the set $U' \cup V$, and $\widetilde{\boldsymbol{\delta}}$ be the subvector of $\boldsymbol{\delta}$ by restricting the elements to the set $U' \cup V$. It is easy to see that when $\mathrm{Supp}(\boldsymbol{\Omega}) \subset U' \times U'$ and $\mathrm{Supp}(\boldsymbol{\delta}) \subset V$,

$$\mathrm{tr}(\boldsymbol{\Omega\Sigma}_k\boldsymbol{\Omega\Sigma}_k) = \mathrm{tr}(\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{\Sigma}}_k\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{\Sigma}}_k),$$

$$(\boldsymbol{\Omega}\boldsymbol{\mu}_k - \boldsymbol{\delta})^\top \boldsymbol{\Sigma}_k (\boldsymbol{\Omega}\boldsymbol{\mu}_k - \boldsymbol{\delta}) = (\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{\mu}}_k - \widetilde{\boldsymbol{\delta}})^\top \widetilde{\boldsymbol{\Sigma}}_k (\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{\mu}}_k - \widetilde{\boldsymbol{\delta}}),$$

where we recall that $\widetilde{\boldsymbol{\Sigma}}_k$ is the submatrix of $\boldsymbol{\Sigma}_k$ by restrcting rows and columns to the set $U' \cup V$, $\widetilde{\boldsymbol{\mu}}_k$ to the set $U' \cup V$. It follows that

$$\begin{aligned}
& L_k(\boldsymbol{\Omega}, \boldsymbol{\delta}) \\
\geq\ & 2(1+\gamma)\operatorname{tr}(\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{\Sigma}}_k\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{\Sigma}}_k) + 4(\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{\mu}}_k - \widetilde{\boldsymbol{\delta}})^\top \widetilde{\boldsymbol{\Sigma}}_k (\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{\mu}}_k - \widetilde{\boldsymbol{\delta}}) \\
=\ & 2(1+\gamma)\operatorname{tr}(\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{\Sigma}}_k\widetilde{\boldsymbol{\Omega}}(\widetilde{\boldsymbol{\Sigma}}_k - v_1\widetilde{\boldsymbol{\mu}}_k\widetilde{\boldsymbol{\mu}}_k^\top)) + 4(1-1/\theta^2)\widetilde{\boldsymbol{\delta}}^\top\widetilde{\boldsymbol{\Sigma}}_k\widetilde{\boldsymbol{\delta}} \\
& + 4(\theta\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{\mu}}_k - \theta^{-1}\widetilde{\boldsymbol{\delta}})^\top \widetilde{\boldsymbol{\Sigma}}_k (\theta\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{\mu}}_k - \theta^{-1}\widetilde{\boldsymbol{\delta}}) \\
\text{(A.2)} \geq\ & 2(1+\gamma)\operatorname{tr}(\widetilde{\boldsymbol{\Omega}}\widetilde{\boldsymbol{\Sigma}}_k\widetilde{\boldsymbol{\Omega}}(\widetilde{\boldsymbol{\Sigma}}_k - v_1\widetilde{\boldsymbol{\mu}}_k\widetilde{\boldsymbol{\mu}}_k^\top)) + 4(1-1/\theta^2)\lambda_{\min}(\widetilde{\boldsymbol{\Sigma}}_k)|\widetilde{\boldsymbol{\delta}}|^2.
\end{aligned}$$

Denote by $I_1$ the first term in (A.2). We aim to derive a lower bound for $I_1$. It is well known that $\operatorname{tr}(\mathbf{A}^\top\mathbf{B}\mathbf{C}\mathbf{D}^\top) = \operatorname{vec}(\mathbf{A})^\top(\mathbf{D}\otimes\mathbf{B})\operatorname{vec}(\mathbf{C})$, where $\operatorname{vec}(\mathbf{A})$ be the vectorization of $\mathbf{A}$ by stacking all the columns, $\mathbf{D}\otimes\mathbf{B}$ is the Kronecker product of $\mathbf{D}$ and $\mathbf{B}$. Using this formula and that $\boldsymbol{\Sigma}_k$ is symmetric, we find that

$$\begin{aligned}
I_1 &= 2(1+\gamma)\operatorname{vec}(\widetilde{\boldsymbol{\Omega}})^\top[(\widetilde{\boldsymbol{\Sigma}}_k - v_1\widetilde{\boldsymbol{\mu}}_k\widetilde{\boldsymbol{\mu}}_k^\top)\otimes\widetilde{\boldsymbol{\Sigma}}_k]\operatorname{vec}(\widetilde{\boldsymbol{\Omega}}) \\
&\geq 2(1+\gamma)|\widetilde{\boldsymbol{\Omega}}|^2\lambda_{\min}((\widetilde{\boldsymbol{\Sigma}}_k - v_1\widetilde{\boldsymbol{\mu}}_k\widetilde{\boldsymbol{\mu}}_k^\top)\otimes\widetilde{\boldsymbol{\Sigma}}_k) \\
&\geq (1+\gamma)\lambda_{\min}^2(\widetilde{\boldsymbol{\Sigma}}_k)|\widetilde{\boldsymbol{\Omega}}|^2.
\end{aligned}$$

The last inequality is from the property that $\lambda_{\min}(\mathbf{A}\otimes\mathbf{B}) = \lambda_{\min}(\mathbf{A})\lambda_{\min}(\mathbf{B})$ when $\mathbf{A}$ and $\mathbf{B}$ are positive semi-definite, and also the assumption that $\lambda_{\min}(\widetilde{\boldsymbol{\Sigma}}_k - v_1\widetilde{\boldsymbol{\mu}}_k\widetilde{\boldsymbol{\mu}}_k) \geq \frac{1}{2}\lambda_{\min}(\widetilde{\boldsymbol{\Sigma}}_k)$. Plugging $I_1$ into (A.2), we have

$$\begin{aligned}
& L_1(\boldsymbol{\Omega}, \boldsymbol{\delta}) + \kappa L_2(\boldsymbol{\Omega}, \boldsymbol{\delta}) \\
\geq\ & (1+\gamma)[\lambda_{\min}^2(\widetilde{\boldsymbol{\Sigma}}_1) + \kappa\lambda_{\min}^2(\widetilde{\boldsymbol{\Sigma}}_2)]|\widetilde{\boldsymbol{\Omega}}|^2 + 4(1-1/\theta^2)[\lambda_{\min}(\widetilde{\boldsymbol{\Sigma}}_1) + \kappa\lambda_{\min}(\widetilde{\boldsymbol{\Sigma}}_2)]|\widetilde{\boldsymbol{\delta}}|^2 \\
\geq\ & (1+\gamma)(1+\kappa){v_2}^2|\widetilde{\boldsymbol{\Omega}}|^2 + 4(1-1/\theta^2)(1+\kappa)v_2|\widetilde{\boldsymbol{\delta}}|^2 \\
\geq\ & c(|\widetilde{\boldsymbol{\Omega}}|^2 + |\widetilde{\boldsymbol{\delta}}|^2) = c(|\boldsymbol{\Omega}|^2 + |\boldsymbol{\delta}|^2).
\end{aligned}$$

This proves (A.1). □

**A.3. Proof of Proposition 6.2.** Given $(\boldsymbol{\Omega}, \boldsymbol{\delta}, t)$, recall that $R_k = R_k(\boldsymbol{\Omega}, \boldsymbol{\delta})$, for $k = 1, 2$. Let $x_1 = \left[(1-t)^2 R_1\right]^{-1}$, $x_2 = \left[t^2 R_2\right]^{-1}$, and $x = \pi x_1 + (1-\pi)x_2$. By direct calculation,

(A.3)

$$\overline{\mathrm{Err}}(\boldsymbol{\Omega}, \boldsymbol{\delta}, t) = \pi H(x_1) + (1-\pi)H(x_2), \qquad H\left(\frac{\pi}{(1-t)^2 R_{\kappa(t)}}\right) = H(x).$$

Since $H$ is twice continuously differentiable, from the Taylor expansion,

$$H(x_1) = H(x) + H'(x)(x_1 - x) + \frac{1}{2}H''(z_1)(x_1 - x)^2,$$

$$H(x_2) = H(x) + H'(x)(x_2 - x) + \frac{1}{2}H''(z_2)(x_2 - x)^2,$$

where $z_1$ is a number between $x_1$ and $x$, and $z_2$ is a number between $x_2$ and $x$. Noticing that $\pi(x_1 - x) + (1-\pi)(x_2 - x) = 0$, we further obtain

$$\begin{aligned}\pi H(x_1) + (1-\pi)H(x_2) &= H(x) + \frac{\pi}{2}H''(z_1)(x_1 - x)^2 + \frac{1-\pi}{2}H''(z_2)(x_2 - x)^2 \\ &= H(x) + \frac{\pi(1-\pi)}{2}[(1-\pi)H''(z_1) + \pi H''(z_2)](x_1 - x_2)^2.\end{aligned}$$

Here, the second equality is because $x_1 - x = (1-\pi)(x_1 - x_2)$ and $x_2 - x = \pi(x_2 - x_1)$. Let $A = \sup_{z\in[\min\{x_1,x_2\},\max\{x_1,x_2\}]} |H''(z)|$. It follows that

$$\big|\pi H(x_1) + (1-\pi)H(x_2) - H(x)\big| \le \frac{\pi(1-\pi)}{2} \cdot A(x_1 - x_2)^2. \tag{A.4}$$

Now, we bound $A(x_1 - x_2)^2$. Write for short $x_{\min} = \min\{x_1, x_2\}$ and $x_{\max} = \max\{x_1, x_2\}$. It is easy to see that

$$A(x_1 - x_2)^2 \le \sup_{z\in[x_{\min},x_{\max}]} |z^2 H''(z)| \cdot \left(\frac{x_1 - x_2}{x_{\min}}\right)^2. \tag{A.5}$$

By direct calculation, the function $z^2H''(z)$ has the expression

$$z^2H''(z) = \frac{1}{4\sqrt{2\pi}}\left(\frac{1}{z} - 3\right) e^{-\frac{1}{2z}} \frac{1}{z^{1/2}}.$$

On one hand, as $z \to \infty$, $|z^2H''(z)| \le Cz^{-1/2}$; on the other hand, as $z \to 0$, $|z^2H''(z)| \le Ce^{-1/(2z)}z^{-3/2} \le Cz^{1/2}$. Therefore, we have $|z^2H''(z)| \le C\min\{\sqrt{z}, 1/\sqrt{z}\}$. Plugging it into (A.5), we have

$$A(x_1 - x_2)^2 \le C\big[\max\big\{x_1 \wedge (1/x_1),\ x_2 \wedge (1/x_2)\big\}\big]^{1/2} \cdot \left(\frac{x_1 - x_2}{x_{\min}}\right)^2.$$

Note that $x_1 \wedge (1/x_1) = V_1$, $x_2 \wedge (1/x_2) = V_2$ and $|x_1 - x_2|/x_{\min} = |V - 1|$. It follows that

$$A(x_1 - x_2)^2 \le C\big[\max\{V_1, V_2\}\big]^{1/2} \cdot |V - 1|^2. \tag{A.6}$$

We combine (A.3), (A.4) and (A.6), and note that $\pi(1-\pi) \le 1/4$. The first claim then follows.

When $t = 1/2$, we find $V - 1 = \Delta R/R_0$. The second claim follows immediately. □

# APPENDIX B

**B.1. Proof of Lemma A.1.** Consider $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. Then $\boldsymbol{Y} \overset{(d)}{=} R\boldsymbol{U}$, where $R^2 \sim \chi^2_d$ and it is independent of $\boldsymbol{U}$. Since $\mathbb{E}(R^2) = d$ and $\mathrm{var}(R^2) = 2d$, it follows that

$$\begin{aligned}
\mathbb{E}(\boldsymbol{Y}^\top \mathbf{S}\boldsymbol{Y}) &= \mathbb{E}(R^2)\mathbb{E}(\boldsymbol{U}^\top \mathbf{S}\boldsymbol{U}) = d\mathbb{E}(\boldsymbol{U}^\top \mathbf{S}\boldsymbol{U}), \\
\mathbb{E}[(\boldsymbol{Y}^\top \mathbf{S}\boldsymbol{Y})^2] &= \mathbb{E}(R^4)\mathbb{E}[(\boldsymbol{U}^\top \mathbf{S}\boldsymbol{U})^2] = (d^2+2d)\mathbb{E}[(\boldsymbol{U}^\top \mathbf{S}\boldsymbol{U})^2], \\
\mathbb{E}(\boldsymbol{Y}^\top \boldsymbol{\beta}) &= \mathbb{E}(R)\mathbb{E}(\boldsymbol{U}^\top \boldsymbol{\beta}), \\
\mathbb{E}[(\boldsymbol{Y}^\top \boldsymbol{\beta})^2] &= \mathbb{E}(R^2)\mathbb{E}[(\boldsymbol{U}^\top \boldsymbol{\beta})^2] = d\mathbb{E}[(\boldsymbol{U}^\top \boldsymbol{\beta})^2], \\
\mathbb{E}(\boldsymbol{Y}^\top \mathbf{S}\boldsymbol{Y}\boldsymbol{Y}^\top \boldsymbol{\beta}) &= \mathbb{E}(R^3)\mathbb{E}(\boldsymbol{U}^\top \mathbf{S}\boldsymbol{U}\boldsymbol{U}^\top \boldsymbol{\beta}).
\end{aligned}$$

First, note that $\boldsymbol{Y}^\top \boldsymbol{\beta} \sim N(0, \|\boldsymbol{\beta}\|^2)$. So $\mathbb{E}(\boldsymbol{Y}^\top \boldsymbol{\beta}) = 0$ and $\mathbb{E}[(\boldsymbol{Y}^\top \boldsymbol{\beta})^2] = \|\boldsymbol{\beta}\|^2$. We immediately have

$$\mathbb{E}(\boldsymbol{U}^\top \boldsymbol{\beta}) = 0, \qquad \mathbb{E}[(\boldsymbol{U}^\top \boldsymbol{\beta})^2] = \|\boldsymbol{\beta}\|^2/d.$$

Second, write $\mathbf{S} = \mathrm{diag}(s_1, \cdots, s_d)$ and $\boldsymbol{Y}^\top \mathbf{S}\boldsymbol{Y} = \sum_{i=1}^d s_i Y_i^2$, where $Y_i \overset{iid}{\sim} N(0,1)$. Therefore,

$$\begin{aligned}
\mathbb{E}(\boldsymbol{Y}^\top \mathbf{S}\boldsymbol{Y}) &= \sum_{i=1}^d s_i \mathbb{E}(Y_i^2) = \sum_{i=1}^d s_i = \mathrm{tr}(\mathbf{S}), \\
\mathrm{var}(\boldsymbol{Y}^\top \mathbf{S}\boldsymbol{Y}) &= \sum_{i=1}^d s_i^2 \mathrm{var}(Y_i^2) = \sum_{i=1}^d 2s_i^2 = 2\,\mathrm{tr}(\mathbf{S}^2).
\end{aligned}$$

We immediately have

$$\mathbb{E}(\boldsymbol{U}^\top \mathbf{S}\boldsymbol{U}) = \mathrm{tr}(\mathbf{S})/d, \qquad \mathbb{E}[(\boldsymbol{U}^\top \mathbf{S}\boldsymbol{U})^2] = \frac{[\mathrm{tr}(\mathbf{S})]^2 + 2\,\mathrm{tr}(\mathbf{S}^2)}{d^2+2d}.$$

Last, note that

$$\boldsymbol{Y}^\top \mathbf{S}\boldsymbol{Y}\boldsymbol{Y}^\top \boldsymbol{\beta} = (\sum_{i=1}^d s_i Y_i^2)(\sum_{j=1}^d \beta_j Y_j) = \sum_{i=1}^d s_i \beta_i Y_i^3 + \sum_{i \neq j} s_i \beta_j Y_i^2 Y_j.$$

So $\mathbb{E}(\boldsymbol{Y}^\top \mathbf{S}\boldsymbol{Y}\boldsymbol{Y}^\top \boldsymbol{\beta}) = 0$. Since $\mathbb{E}(R^3) \neq 0$, we immediately have $\mathbb{E}(\boldsymbol{U}^\top \mathbf{S}\boldsymbol{U}\boldsymbol{U}^\top \boldsymbol{\beta}) = 0$. $\square$

**B.2. Proof of Lemma 9.1.** Recall that

$$\begin{aligned} M(\boldsymbol{\Omega}, \boldsymbol{\delta}) &= -\boldsymbol{\mu}_1^\top \boldsymbol{\Omega}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^\top \boldsymbol{\Omega}\boldsymbol{\mu}_2 + 2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \boldsymbol{\delta} - \operatorname{tr}(\boldsymbol{\Omega}(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)) \\ &= \operatorname{tr}\left[\boldsymbol{\Omega}(\boldsymbol{\Sigma}_2 + \boldsymbol{\mu}_2\boldsymbol{\mu}_2^\top - \boldsymbol{\Sigma}_1 - \boldsymbol{\mu}_1\boldsymbol{\mu}_1^\top)\right] + 2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \boldsymbol{\delta}. \end{aligned}$$

It is well known that for any matrices $\mathbf{A}$ and $\mathbf{B}$, $\operatorname{tr}(\mathbf{A}^\top \mathbf{B}) = \operatorname{vec}(\mathbf{A})^\top \operatorname{vec}(\mathbf{B})$. So we have

$$\begin{aligned} M(\boldsymbol{\Omega}, \boldsymbol{\delta}) &= \operatorname{vec}(\boldsymbol{\Omega})^\top \operatorname{vec}(\boldsymbol{\Sigma}_2 + \boldsymbol{\mu}_2\boldsymbol{\mu}_2^\top - \boldsymbol{\Sigma}_1 - \boldsymbol{\mu}_1\boldsymbol{\mu}_1^\top) + 2\boldsymbol{\delta}^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\ &= \mathbf{x}^\top \mathbf{q}. \end{aligned}$$

Moreover, for $k = 1, 2$,

$$\begin{aligned} L_k(\boldsymbol{\Omega}, \boldsymbol{\delta}) &= 2(1+\gamma)\operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{\Sigma}_k\boldsymbol{\Omega}\boldsymbol{\Sigma}_k) + \gamma[\operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{\Sigma}_k)]^2 + 4(\boldsymbol{\Omega}\boldsymbol{\mu}_k - \boldsymbol{\delta})^\top \boldsymbol{\Sigma}_k(\boldsymbol{\Omega}\boldsymbol{\mu}_k - \boldsymbol{\delta}) \\ &= 2(1+\gamma)\operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{\Sigma}_k\boldsymbol{\Omega}\boldsymbol{\Sigma}_k) + \gamma[\operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{\Sigma}_k)]^2 + 4\boldsymbol{\mu}_k^\top \boldsymbol{\Omega}\boldsymbol{\Sigma}_k\boldsymbol{\Omega}\boldsymbol{\mu}_k - 8\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_k\boldsymbol{\Omega}\boldsymbol{\mu}_k + 4\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_k\boldsymbol{\delta} \\ &= 2\operatorname{tr}\left[\boldsymbol{\Omega}\boldsymbol{\Sigma}_k\boldsymbol{\Omega}((1+\gamma)\boldsymbol{\Sigma}_k + 2\boldsymbol{\mu}_k\boldsymbol{\mu}_k^\top)\right] - 8\operatorname{tr}(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_k\boldsymbol{\Omega}\boldsymbol{\mu}_k) + 4\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_k\boldsymbol{\delta} + \gamma[\operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{\Sigma}_k)]^2. \end{aligned}$$

From linear algebra, $\operatorname{tr}(\mathbf{A}^\top \mathbf{B}\mathbf{C}\mathbf{D}^\top) = \operatorname{vec}(\mathbf{A})^\top (\mathbf{D} \otimes \mathbf{B}) \operatorname{vec}(\mathbf{C})$. It follows that

$$\begin{aligned} L_k(\boldsymbol{\Omega}, \boldsymbol{\delta}) &= 2\operatorname{vec}(\boldsymbol{\Omega})^\top \left[((1+\gamma)\boldsymbol{\Sigma}_k + 2\boldsymbol{\mu}_k\boldsymbol{\mu}_k^\top) \otimes \boldsymbol{\Sigma}_k\right] \operatorname{vec}(\boldsymbol{\Omega}) - 8\boldsymbol{\delta}^\top(\boldsymbol{\mu}_k^\top \otimes \boldsymbol{\Sigma}_k)\operatorname{vec}(\boldsymbol{\Omega}) \\ &\quad + 4\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_k\boldsymbol{\delta} + \gamma \operatorname{vec}(\boldsymbol{\Omega})^\top \operatorname{vec}(\boldsymbol{\Sigma}_k)\operatorname{vec}(\boldsymbol{\Sigma}_k)^\top \operatorname{vec}(\boldsymbol{\Omega}) \\ &= 2\begin{bmatrix}\operatorname{vec}(\boldsymbol{\Omega})^\top & \boldsymbol{\delta}^\top\end{bmatrix} \begin{bmatrix} ((1+\gamma)\boldsymbol{\Sigma}_k + 2\boldsymbol{\mu}_k\boldsymbol{\mu}_k^\top) \otimes \boldsymbol{\Sigma}_k & -2(\boldsymbol{\mu}_k \otimes \boldsymbol{\Sigma}_k) \\ -2(\boldsymbol{\mu}_k^\top \otimes \boldsymbol{\Sigma}_k) & 2\boldsymbol{\Sigma}_k \end{bmatrix} \begin{bmatrix}\operatorname{vec}(\boldsymbol{\Omega}) \\ \boldsymbol{\delta}\end{bmatrix} \\ &\quad + \gamma \operatorname{vec}(\boldsymbol{\Omega})^\top \operatorname{vec}(\boldsymbol{\Sigma}_k)\operatorname{vec}(\boldsymbol{\Sigma}_k)^\top \operatorname{vec}(\boldsymbol{\Omega}) \\ &= \mathbf{x}^\top \mathbf{Q}_k \mathbf{x}. \end{aligned}$$

Note that $\mathbf{Q} = \mathbf{Q}_1 + \kappa\mathbf{Q}_2$ and $L = L_1 + \kappa L_2$. This immediately implies $L(\boldsymbol{\Omega}, \boldsymbol{\delta}) = \mathbf{x}^\top \mathbf{Q}\mathbf{x}$. □

**B.3. Proof of Lemma 9.2.** Let $\Delta_n = \max\{|\widehat{\boldsymbol{\Sigma}}_k - \boldsymbol{\Sigma}_k|_\infty, |\widehat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k|_\infty, k = 1, 2\}$ to save notation. We recall that $|\widehat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k|_\infty \le |\boldsymbol{\mu}_k|_\infty \le 1$ and $|\widehat{\boldsymbol{\Sigma}}_k - \boldsymbol{\Sigma}_k|_\infty \le |\boldsymbol{\Sigma}_k|_\infty$, for $k = 1, 2$, as assumed in the beginning of Section 5.2.

Consider $|\widehat{\mathbf{q}} - \mathbf{q}|_\infty$ first. Note that $|\operatorname{vec}(\mathbf{A}) - \operatorname{vec}(\mathbf{B})|_\infty = |\mathbf{A} - \mathbf{B}|_\infty$ for any matrices $\mathbf{A}$ and $\mathbf{B}$. It follows that

$$\begin{aligned} |\widehat{\mathbf{q}} - \mathbf{q}|_\infty &\le \sum_{k=1}^{2} \left(|\widehat{\boldsymbol{\Sigma}}_k - \boldsymbol{\Sigma}_k|_\infty + |\widehat{\boldsymbol{\mu}}_k\widehat{\boldsymbol{\mu}}_k^\top - \boldsymbol{\mu}_k\boldsymbol{\mu}_k^\top|_\infty + 2|\widehat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k|_\infty\right) \\ &\le C\Delta_n + \sum_{k=1}^{2} |\widehat{\boldsymbol{\mu}}_k\widehat{\boldsymbol{\mu}}_k^\top - \boldsymbol{\mu}_k\boldsymbol{\mu}_k^\top|_\infty. \end{aligned}$$

Write $\boldsymbol{\mu}_k = \boldsymbol{\mu}$ for short. We have $|\widehat{\mu}(i)\widehat{\mu}(j) - \mu(i)\mu(j)| \leq |\widehat{\mu}(j)||\widehat{\mu}(i) - \mu(i)| + |\mu(i)||\widehat{\mu}(j) - \mu(j)| \leq (|\widehat{\boldsymbol{\mu}}|_\infty + |\boldsymbol{\mu}|_\infty)|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}|_\infty$. Since $|\widehat{\boldsymbol{\mu}}|_\infty \leq |\boldsymbol{\mu}|_\infty + |\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}|_\infty \leq 2|\boldsymbol{\mu}|_\infty \leq 2$, it follows that $|\widehat{\mu}(i)\widehat{\mu}(j) - \mu(i)\mu(j)| \leq 3|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}|_\infty \leq 3\Delta_n$. As a result,

$$|\widehat{\boldsymbol{\mu}}_k \widehat{\boldsymbol{\mu}}_k^\top - \boldsymbol{\mu}_k \boldsymbol{\mu}_k^\top|_\infty \leq 3\Delta_n, \qquad k = 1, 2. \tag{B.1}$$

We immediately have $|\widehat{\mathbf{q}} - \mathbf{q}|_\infty \leq C\Delta_n$.

Next, consider $|\widehat{\mathbf{Q}} - \mathbf{Q}|_\infty$. It is easy to see that

$$\begin{aligned}
|\widehat{\mathbf{Q}}_k - \mathbf{Q}_k|_\infty \leq\; & 2(1+\gamma)|\widehat{\boldsymbol{\Sigma}}_k \otimes \widehat{\boldsymbol{\Sigma}}_k - \boldsymbol{\Sigma}_k \otimes \boldsymbol{\Sigma}_k|_\infty + 4|(\widehat{\boldsymbol{\mu}}_k \widehat{\boldsymbol{\mu}}_k^\top) \otimes \widehat{\boldsymbol{\Sigma}}_k - (\boldsymbol{\mu}_k \boldsymbol{\mu}_k^\top) \otimes \boldsymbol{\Sigma}_k|_\infty \\
& + \gamma|\operatorname{vec}(\widehat{\boldsymbol{\Sigma}}_k)\operatorname{vec}(\widehat{\boldsymbol{\Sigma}}_k)^\top - \operatorname{vec}(\boldsymbol{\Sigma}_k)\operatorname{vec}(\boldsymbol{\Sigma}_k)^\top|_\infty \\
& + 4|\widehat{\boldsymbol{\mu}}_k \otimes \widehat{\boldsymbol{\Sigma}}_k - \boldsymbol{\mu}_k \otimes \boldsymbol{\Sigma}_k|_\infty + 4|\widehat{\boldsymbol{\Sigma}}_k - \boldsymbol{\Sigma}_k|_\infty.
\end{aligned}$$

Here $|\widehat{\boldsymbol{\Sigma}}_k - \boldsymbol{\Sigma}_k|_\infty \leq \Delta_n$, and using a similar argument as in (B.1), we can show that $|\operatorname{vec}(\widehat{\boldsymbol{\Sigma}}_k)\operatorname{vec}(\widehat{\boldsymbol{\Sigma}}_k)^\top - \operatorname{vec}(\boldsymbol{\Sigma}_k)\operatorname{vec}(\boldsymbol{\Sigma}_k)^\top|_\infty \leq C|\operatorname{vec}(\widehat{\boldsymbol{\Sigma}}_k) - \operatorname{vec}(\boldsymbol{\Sigma}_k)|_\infty \leq C\Delta_n$. To bound the other terms, it suffices to show that

$$|\mathbf{A} \otimes \mathbf{B} - \mathbf{A}' \otimes \mathbf{B}'|_\infty \leq C(|\mathbf{A} - \mathbf{A}'|_\infty + |\mathbf{B} - \mathbf{B}'|_\infty), \tag{B.2}$$

when $|\mathbf{A} - \mathbf{A}'|_\infty \leq |\mathbf{A}|_\infty \leq C$ and $|\mathbf{B} - \mathbf{B}'|_\infty \leq |\mathbf{B}|_\infty \leq C$. To see this, note that $|A(i,j)B(k,l) - A'(i,j)B'(k,l)| \leq |B'(k,l)||A(i,j) - A'(i,j)| + |A(i,j)||B(k,l) - B'(k,l)| \leq (|\mathbf{A}|_\infty + |\mathbf{B}'|_\infty)(|\mathbf{A} - \mathbf{A}'|_\infty + |\mathbf{B} - \mathbf{B}'|_\infty) \leq C(|\mathbf{A} - \mathbf{A}'|_\infty + |\mathbf{B} - \mathbf{B}'|_\infty)$. This proves (B.2). It follows immediately that $|\widehat{\mathbf{Q}} - \mathbf{Q}|_\infty \leq C\Delta_n$. □

**B.4. Proof of Lemma 9.3.** Write $\mathbf{x}^* = \mathbf{x}^*_{\lambda_0}$ for short. From KKT conditions, there exists a dual variable $\theta$ such that

$$\mathbf{q}^\top \mathbf{x}^* = 1, \qquad 2\mathbf{Q}\mathbf{x}^* + \lambda_0 \widetilde{\operatorname{sign}}(\mathbf{x}^*) = \theta \mathbf{q}.$$

Here $\widetilde{\operatorname{sign}}(\mathbf{x}^*)$ is the vector whose $j$-th coordinate is $-1$, $1$ or any value between $[-1, 1]$ when $x_j^* < 0$, $x_j^* > 0$ and $x_j^* = 0$, respectively. We multiply both sides of the second equation by $(\mathbf{x}^*)^\top$, and note that $(\mathbf{x}^*)^\top \widetilde{\operatorname{sign}}(\mathbf{x}^*) = |\mathbf{x}^*|_1$. It follows that $\theta = 2(\mathbf{x}^*)^\top \mathbf{Q}\mathbf{x}^* + \lambda_0 |\mathbf{x}^*|_1$ and

$$\begin{aligned}
\mathbf{Q}\mathbf{x}^* &= \Big[(\mathbf{x}^*)^\top \mathbf{Q}\mathbf{x}^* + \frac{\lambda_0}{2}|\mathbf{x}^*|_1\Big]\mathbf{q} - \frac{\lambda_0}{2}\widetilde{\operatorname{sign}}(\mathbf{x}^*) \\
&= V^* \mathbf{q} + \frac{\lambda_0}{2}\big[|\mathbf{x}^*|_1 \mathbf{q} - \widetilde{\operatorname{sign}}(\mathbf{x}^*)\big],
\end{aligned} \tag{B.3}$$

where $V^* = (\mathbf{x}^*)^\top \mathbf{Q}\mathbf{x}^* = [R(\mathbf{x}^*)]^{-1}$. Let $\bar{V}^* = (V^*)^{1/2}$. Then $\Gamma(\mathbf{x}) = \frac{1}{\bar{V}^*}|\mathbf{Q}\mathbf{x}^* - V^*\mathbf{q}|_\infty$, and (B.3) implies

$$\Gamma(\mathbf{x}^*) \le \frac{\lambda_0}{2\bar{V}^*}(|\mathbf{x}^*|_1|\mathbf{q}|_\infty + 1) = \frac{|\mathbf{q}|_\infty}{2\bar{V}^*}\lambda_0|\mathbf{x}^*|_1 + \frac{\lambda_0}{2}[R(\mathbf{x}^*)]^{1/2}. \tag{B.4}$$

From the Cauchy-Schwartz inequality,

$$|\mathbf{x}^*|_1 \le |\mathbf{x}^*|_0^{1/2}|\mathbf{x}^*| \le |\mathbf{x}^*|_0^{1/2}\frac{1}{\sqrt{c_0}}\sqrt{(\mathbf{x}^*)^\top \mathbf{Q}\mathbf{x}^*} = \frac{\bar{V}^*}{\sqrt{c_0}}|\mathbf{x}^*|_0^{1/2},$$

where we have used the fact that $(\mathbf{x}^*)^\top \mathbf{Q}\mathbf{x}^* \ge \lambda_{\min}(\mathbf{Q}_{SS})|\mathbf{x}^*|^2 \ge c_0|\mathbf{x}^*|^2$. Plugging it into (B.4) gives

$$\begin{aligned}\Gamma(\mathbf{x}^*) &\le \frac{|\mathbf{q}|_\infty}{2\sqrt{c_0}}\lambda_0|\mathbf{x}^*|_0^{1/2} + \frac{\lambda_0}{2}[R(\mathbf{x}^*)]^{1/2}\\ &\le C_1\lambda_0\big[\max\{|\mathbf{x}^*|_0, R(\mathbf{x}^*)\}\big]^{1/2},\end{aligned}$$

where $C_1$ is a positive constant that only depends on $c_0$ (noting that $|\mathbf{q}|_\infty \le C\max\{|\boldsymbol{\mu}_k|_\infty, |\boldsymbol{\Sigma}_k|_\infty, k = 1, 2\} \le C$).

## APPENDIX C

For the experiments for Gaussian distributions, the means and standard deviations (in the parenthesis) of 100 replications are reported here. The corresponding boxplots are in Figure 3.

| | QUADRO | SLR | L-SLR | ROAD | P-LDA | FAIR |
|---|---|---|---|---|---|---|
| Model 1 | 0.179<br>(0.016) | 0.235<br>(0.028) | 0.191<br>(0.017) | 0.246<br>(0.074) | 0.192<br>(0.011) | 0.185<br>(0.018) |
| Model 1L | 0.162<br>(0.016) | 0.214<br>(0.030) | 0.172<br>(0.017) | 0.217<br>(0.070) | 0.170<br>(0.011) | 0.162<br>(0.016) |
| Model 2 | 0.144<br>(0.016) | 0.224<br>(0.018) | 0.470<br>(0.008) | 0.491<br>(0.010) | 0.476<br>(0.021) | 0.481<br>(0.017) |
| Model 3 | 0.109<br>(0.013) | 0.164<br>(0.018) | 0.176<br>(0.016) | 0.235<br>(0.068) | 0.202<br>(0.015) | 0.218<br>(0.018) |

For the experiments for general elliptical distributions, the means and standard deviations (in the parenthesis) of 100 replications are reported here. The corresponding boxplots are in Figure 4.

| | QUADRO | QUADRO-0 | SLR | L-SLR | ROAD | P-LDA | FAIR |
|---|---|---|---|---|---|---|---|
| Model 4 | 0.136<br>(0.015) | 0.144<br>(0.015) | 0.167<br>(0.019) | 0.157<br>(0.017) | 0.197<br>(0.069) | 0.159<br>(0.017) | 0.196<br>(0.028) |
| Model 5 | 0.161<br>(0.012) | 0.173<br>(0.014) | 0.184<br>(0.014) | 0.184<br>(0.016) | 0.232<br>(0.060) | 0.174<br>(0.010) | 0.210<br>(0.019) |
| Model 6 | 0.130<br>(0.017) | 0.129<br>(0.016) | 0.152<br>(0.018) | 0.211<br>(0.018) | 0.268<br>(0.055) | 0.193<br>(0.012) | 0.235<br>(0.023) |

J. Fan, H. Liu and L. Xia
Department of Operations Research
and Financial Engineering
Princeton University
Princeton, New Jersey, 08544
USA
E-mail: jqfan@princeton.edu
hanliu@princeton.edu
lxia@princeton.edu

Z. Ke
Department of Statistics
University of Chicago
Chicago, Illinois, 60637
USA
E-mail: zke@princeton.edu