# SPARSE AND LOW-RANK MATRIX QUANTILE ESTIMATION WITH APPLICATION TO QUADRATIC REGRESSION

## Supplementary Material

## S1 Proof of Theorem 1

Let $\mathbf{\Delta} = \widehat{\mathbf{B}} - \mathbf{B}$, and define

$$Q(\mathbf{Z}_i; \mathbf{\Delta}) = \rho_\tau \left( y_i - \langle \mathbf{B} + \mathbf{\Delta}, \mathbf{Z}_i \rangle \right) - \rho_\tau \left( y_i - \langle \mathbf{B}, \mathbf{Z}_i \rangle \right).$$

By the optimality of $\widehat{\mathbf{B}}$, we have

$$\frac{1}{n} \sum_{i=1}^{n} Q(\mathbf{Z}_i; \mathbf{\Delta}) \leq \lambda\alpha\mathcal{R}_1(\mathbf{B}) + \lambda(1-\alpha)\mathcal{R}_2(\mathbf{B}) - \lambda\alpha\mathcal{R}_1(\mathbf{B}+\mathbf{\Delta}) - \lambda(1-\alpha)\mathcal{R}_2(\mathbf{B}+\mathbf{\Delta}). \tag{S1.1}$$

Since $\rho_\tau(\cdot)$ is convex, we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} Q(\mathbf{Z}_i; \mathbf{\Delta}) &\geq \langle -\frac{1}{n} \sum_{i=1}^{n} \left( \tau - I\{ y_i - \langle \mathbf{B}, \mathbf{Z}_i \rangle \leq 0 \} \right) \mathbf{Z}_i, \mathbf{\Delta} \rangle \\ &\geq - \min \left\{ \|\mathbf{E}\|_{op}/\alpha, \|\mathbf{E}\|_\infty/(1-\alpha) \right\} \left( \alpha \|\mathbf{\Delta}\|_* + (1-\alpha)\|\mathbf{\Delta}\|_1 \right), \end{aligned} \tag{S1.2}$$

where we define $\mathbf{E} = \frac{1}{n}\sum_{i=1}^{n}\left(\tau - I\{y_i - \langle \mathbf{B}, \mathbf{Z}_i\rangle \le 0\}\right)\mathbf{Z}_i$, and in the last step we used Lemma 1.

Following the proof of Lemma 2, and using Markov's inequality, we can easily obtain $\|\mathbf{E}\|_{op} \le C\sqrt{(d_1+d_2)/n}$ and $\|\mathbf{E}\|_{\infty} \le C\sqrt{\mathrm{log}p/n}$, with probability approaching one. Thus we have $\lambda \ge 2\min\left\{\|\mathbf{E}\|_{op}/\alpha, \|\mathbf{E}\|_{\infty}/(1-\alpha)\right\}$, which combined with (S1.2) yields

$$\frac{1}{n}\sum_{i=1}^{n} Q(\mathbf{Z}_i; \mathbf{\Delta}) \ge -\frac{\lambda}{2}\left(\alpha\|\mathbf{\Delta}\|_* + (1-\alpha)\|\mathbf{\Delta}\|_1\right). \qquad \text{(S1.3)}$$

Recalling that we define $\mathbf{\Delta}''$ to be the projection of $\mathbf{\Delta}$ on $\bar{\mathbb{M}}_1^{\perp}$ and $\mathbf{\Delta}' = \mathbf{\Delta} - \mathbf{\Delta}''$, we also have

$$\begin{aligned}\mathcal{R}_1(\mathbf{B}+\mathbf{\Delta}) =& \mathcal{R}_1\left(\mathbf{B}+\mathbf{\Delta}''+\mathbf{\Delta}'\right)\\ \ge& \mathcal{R}_1\left(\mathbf{B}+\mathbf{\Delta}''\right) - \mathcal{R}_1\left(\mathbf{\Delta}'\right)\\ =& \mathcal{R}_1\left(\mathbf{B}\right) + \mathcal{R}_1\left(\mathbf{\Delta}''\right) - \mathcal{R}_1\left(\mathbf{\Delta}'\right),\end{aligned}$$

where the last equality used the decomposability property since $\mathbf{B} \in \mathbb{M}_1$ and $\mathbf{\Delta}'' \in \bar{\mathbb{M}}_1^{\perp}$. Thus we have $\mathcal{R}_1(\mathbf{B}) - \mathcal{R}_1(\mathbf{B}+\mathbf{\Delta}) \le \mathcal{R}_1(\mathbf{\Delta}') - \mathcal{R}_1(\mathbf{\Delta}'')$. Similarly, we can show $\mathcal{R}_2(\mathbf{B}) - \mathcal{R}_2(\mathbf{B}+\mathbf{\Delta}) \le \mathcal{R}_2(\mathbf{\Delta}_{\mathcal{S}}) - \mathcal{R}_2(\mathbf{\Delta}_{\mathcal{S}^{\perp}})$. Combined with (S1.1), (S1.3), we proved that $\mathbf{\Delta} \in \mathbb{C}$, that is, $\mathbf{\Delta}$ satisfies

$$\alpha\mathcal{R}_1(\mathbf{\Delta}'') + (1-\alpha)\mathcal{R}_2(\mathbf{\Delta}_{\mathcal{S}^{\perp}}) \le 3(\alpha\mathcal{R}_1(\mathbf{\Delta}') + (1-\alpha)\mathcal{R}_2(\mathbf{\Delta}_{\mathcal{S}})). \qquad \text{(S1.4)}$$

By Lemma 3, assumption C3, (S1.1), and that $\mathbf{\Delta} \in \mathbb{C}$, we get

$$\begin{aligned} & c_1(\|\mathbf{\Delta}\|_F^2 \wedge \|\mathbf{\Delta}\|_F) - C\|\mathbf{\Delta}\|_F \min\left\{\frac{(\alpha\sqrt{r}+(1-\alpha)\sqrt{s})}{\alpha}\sqrt{\frac{d_1+d_2}{n}}, \frac{(\alpha\sqrt{r}+(1-\alpha)\sqrt{s})}{1-\alpha}\sqrt{\frac{\log p}{n}}\right\} \\ \leq \ & \lambda(\alpha\mathcal{R}_1(\mathbf{\Delta}) + (1-\alpha)\mathcal{R}_2(\mathbf{\Delta})) \leq 4\lambda(\alpha\mathcal{R}_1(\mathbf{\Delta}') + (1-\alpha)\mathcal{R}_2(\mathbf{\Delta}_{\mathcal{S}})) \\ \leq \ & C\lambda(\alpha\sqrt{r}+(1-\alpha)\sqrt{s})\|\mathbf{\Delta}\|_F. \end{aligned}$$

This implies $\|\mathbf{\Delta}\|_F \leq C\lambda(\alpha\sqrt{r}+(1-\alpha)\sqrt{s})$. □

**Lemma 1.** *For any* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{d_1 \times d_2}$ *and* $\alpha \in [0,1]$*, we have*

$$\langle \mathbf{A}, \mathbf{B}\rangle \leq \min\left\{\|\mathbf{B}\|_{op}/\alpha, \|\mathbf{B}\|_\infty/(1-\alpha)\right\}(\alpha\|\mathbf{A}\|_* + (1-\alpha)\|\mathbf{A}\|_1),$$

*where* $\|\mathbf{B}\|_{op}$ *is the operator norm and* $\|\mathbf{B}\|_\infty = \max_{j,k}|B_{jk}|$.

**Proof.** Using $\langle \mathbf{A}, \mathbf{B}\rangle \leq \|\mathbf{A}\|_*\|\mathbf{B}\|_{op}$ and $\langle \mathbf{A}, \mathbf{B}\rangle \leq \|\mathbf{A}\|_1\|\mathbf{B}\|_\infty$ we have

$$\begin{aligned} \langle \mathbf{A}, \mathbf{B}\rangle & \leq \min\left\{\alpha\|\mathbf{A}\|_*\frac{\|\mathbf{B}\|_{op}}{\alpha}, (1-\alpha)\|\mathbf{A}\|_1\frac{\|\mathbf{B}\|_\infty}{1-\alpha}\right\} \\ & \leq \min\left\{\frac{\|\mathbf{B}\|_{op}}{\alpha}, \frac{\|\mathbf{B}\|_\infty}{1-\alpha}\right\}(\alpha\|\mathbf{A}\|_* + (1-\alpha)\|\mathbf{A}\|_1). \end{aligned}$$

□

**Lemma 2.** *If* $\mathbf{z}_i = \mathrm{vec}(\mathbf{Z}_i)$ *is sub-Gaussian, then* $\forall \gamma > 0$,

$$E[\exp\{\gamma\|\sum_i \epsilon_i \mathbf{Z}_i\|_{op}\}] \leq 20^{d_1+d_2}e^{Cn\gamma^2}$$

*and*

$$E[\exp\{\gamma\|\sum_i \epsilon_i \mathbf{Z}_i\|_\infty\}] \leq 2pe^{Cn\gamma^2},$$

*where* $\epsilon_i \in \{-1, 1\}$ *are independent Rademacher variables.*

**Proof.** Let $\mathbf{E}$ be any matrix of size $d_1 \times d_2$. Let $\{\mathbf{u}_i, i = 1, \ldots, M_1\}$ be a 1/4-covering of the unit sphere in $\mathbb{R}^{d_1}$ and $\{\mathbf{v}_i, i = 1, \ldots, M_2\}$ be a 1/4 covering of the unit sphere in $\mathbb{R}^{d_2}$, with $M_1 \leq 20^{d_1}$ and $M_2 \leq 20^{d_2}$ (the bound for $M_1, M_2$ is due to Lemma 2.5 of van der Geer (2000)). Thus, for any $\mathbf{u}, \mathbf{v}$ with $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$ there exists $\mathbf{u}_i, \mathbf{v}_j$ in the covering such that $\|\mathbf{u} - \mathbf{u}_i\| \leq 1/4$ and $\|\mathbf{v} - \mathbf{v}_j\| \leq 1/4$ and then

$$\mathbf{u}^\top \mathbf{E}\mathbf{v} = \mathbf{u}^\top \mathbf{E}(\mathbf{v} - \mathbf{v}_j) + (\mathbf{u} - \mathbf{u}_i)^\top \mathbf{E}\mathbf{v}_j + \mathbf{u}_i^\top \mathbf{E}\mathbf{v}_j \leq \frac{1}{4}\|\mathbf{E}\|_{op} + \frac{1}{4}\|\mathbf{E}\|_{op} + \mathbf{u}_i^\top \mathbf{E}\mathbf{v}_j.$$

Thus we have

$$\|\mathbf{E}\|_{op} = \sup_{\|\mathbf{u}\|=\|\mathbf{v}\|=1} \mathbf{u}^\top \mathbf{E}\mathbf{v} \leq \frac{1}{2}\|\mathbf{E}\|_{op} + \max_{\mathbf{u}_i, \mathbf{v}_j} \mathbf{u}_i^\top \mathbf{E}\mathbf{v}_j,$$

which implies

$$\|\mathbf{E}\|_{op} \leq 2 \max_{\mathbf{u}_i, \mathbf{v}_j} \mathbf{u}_i^\top \mathbf{E}\mathbf{v}_j.$$

Then we have

$$\begin{aligned}
& E[\exp\{\gamma\|\sum_i \epsilon_i \mathbf{Z}_i\|_{op}] \\
\leq\ & 20^{d_1+d_2} \max_{\mathbf{u}_j, \mathbf{v}_k} E[\exp\{2\gamma \mathbf{u}_j^\top (\sum_i \epsilon_i \mathbf{Z}_i)\mathbf{v}_k\}] \\
=\ & 20^{d_1+d_2} \max_{\mathbf{u}_j, \mathbf{v}_k} \prod_{i=1}^n E\left[\exp\{2\gamma \mathbf{u}_j^\top (\epsilon_i \mathbf{Z}_i)\mathbf{v}_k\}\right] \\
\leq\ & 20^{d_1+d_2} e^{C\gamma^2 n},
\end{aligned}$$

where the last step used assumption C2 and note that $\epsilon_i \mathbf{Z}_i$ is also sub-Gaussian and $\mathbf{u}_j^\top (\epsilon_i \mathbf{Z}_i)\mathbf{v}_k = (\mathbf{v}_k \otimes \mathbf{u}_j)^\top \text{vec}(\epsilon_i \mathbf{Z}_i)$.

The second part is easier. We have

$$\begin{aligned}
& E[\exp\{\gamma \max_j |\sum_i z_{ij}\epsilon_i|\}] \\
= \; & E[\max_j \exp\{\gamma|\sum_i z_{ij}\epsilon_i|\}] \\
\leq \; & p \max_j E[\exp\{\gamma|\sum_i z_{ij}\epsilon_i|\}] \\
\leq \; & 2p \max_j E[\exp\{\gamma(\sum_i z_{ij}\epsilon_i)\}],
\end{aligned}$$

where the last step used the fact that for any symmetric random variable $z$, $E[e^{|z|}] \leq e[e^z + e^{-z}] = 2E[e^z]$. Using $z_{ij}$ is sub-Gaussian and thus $z_{ij}\epsilon_i$ is also sub-Gaussian, we get $E[\exp\{\gamma(\sum_i z_{ij}\epsilon_i)\}] = (e^{C\gamma^2})^n$ which proved the lemma. □

**Lemma 3.** *Under the assumptions of Theorem 1, with probability approaching one, we have*

$$\begin{aligned}
& \sup_{\substack{\boldsymbol{\Delta}\in\mathbb{C} \\ \|\boldsymbol{\Delta}\|_F\leq t}} \left| \frac{1}{n}\sum_{i=1}^n Q(\mathbf{Z}_i;\boldsymbol{\Delta}) - EQ(\mathbf{Z}_i;\boldsymbol{\Delta}) \right| \\
\leq \; & Ct \min\left\{ \frac{(\alpha\sqrt{r} + (1-\alpha)\sqrt{s})}{\alpha}\sqrt{\frac{d_1+d_2}{n}}, \frac{(\alpha\sqrt{r} + (1-\alpha)\sqrt{s})}{1-\alpha}\sqrt{\frac{logp}{n}} \right\}.
\end{aligned}$$

**Proof.** Let

$$\begin{aligned}
A(t) = \; & \sup_{\substack{\boldsymbol{\Delta}\in\mathbb{C} \\ \|\boldsymbol{\Delta}\|_F\leq t}} \left| \frac{1}{n}\sum_{i=1}^n Q(\mathbf{Z}_i;\boldsymbol{\Delta}) - EQ(\mathbf{Z}_i;\boldsymbol{\Delta}) \right| \\
= \; & \sup_{\substack{\boldsymbol{\Delta}\in\mathbb{C} \\ \|\boldsymbol{\Delta}\|_F\leq t}} \frac{1}{\sqrt{n}} \left| \mathbb{G}_n Q(\mathbf{Z}_i;\boldsymbol{\Delta}) \right|,
\end{aligned}$$

where $\mathbb{G}_n Q = \sqrt{n}\,(P_n Q - PQ)$ is the empirical process.

By the Lipschitz property of $\rho_\tau$, we have for any $\mathbf{\Delta}$ with $\|\mathbf{\Delta}\|_F \le t$,

$$\mathrm{Var}(Q(\mathbf{Z}_i;\mathbf{\Delta}) - EQ(\mathbf{Z}_i;\mathbf{\Delta})) \le E\left(\mathbf{z}_i^t\boldsymbol{\delta}\right)^2 \le \sigma_{\max}(\mathbf{J})t^2,$$

where $\sigma_{\max}(\mathbf{J})$ is the maximum singular value of $\mathbf{J}$. Let $B(t) = \frac{1}{\sqrt{n}} \sup_{\substack{\mathbf{\Delta}\in\mathbb{C} \\ \|\mathbf{\Delta}\|_F\le t}} |\mathbb{G}_n^o Q(\mathbf{Z}_i;\mathbf{\Delta})|$ with $\mathbb{G}_n^o Q(\mathbf{Z}_i;\mathbf{\Delta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i Q(\mathbf{Z}_i;\mathbf{\Delta})$ and $\epsilon_i$ are independent Rademacher variables. Then by Lemma 2.3.7 in Van der Vaart and Wellner (1996), we get

$$P\left(A(t) \ge M\right) \le \frac{2P\left(B(t) \ge M/4\right)}{1 - 4\sigma_{\max}(\mathbf{J})t^2/(nM^2)}.$$

Let $\mathbf{W} = \frac{1}{n}\sum_{i=1}^n \epsilon_i \mathbf{Z}_i$. Then we have, for any $\eta > 0$,

$$\begin{aligned}
&P\left(B(t) \ge M/4\right) \\
\le &\exp\left\{-\frac{1}{4}\eta M\right\} E\exp\left\{\eta B(t)\right\} && \text{(S1.5)} \\
\le &\exp\left\{-\frac{1}{4}\eta M\right\} E\exp\left\{C\eta \sup_{\substack{\mathbf{\Delta}\in\mathbb{C} \\ \|\mathbf{\Delta}\|_F\le t}} \left|\frac{1}{n}\sum_{i=1}^n \epsilon_i\langle \mathbf{Z}_i, \mathbf{\Delta}\rangle\right|\right\} && \text{(S1.6)} \\
\le &\exp\left\{-\frac{1}{4}\eta M\right\} E\exp\left\{C\eta \min\left\{\frac{\|\mathbf{W}\|_{op}}{\alpha}, \frac{\|\mathbf{W}\|_\infty}{1-\alpha}\right\}\left(\alpha\sqrt{r}t + (1-\alpha)\sqrt{s}t\right)\right\} \\
& && \text{(S1.7)} \\
\le &\min\left\{\exp\left\{-\frac{1}{4}\eta M\right\} E\exp\left\{C\eta\frac{\|\mathbf{W}\|_{op}}{\alpha}\left(\alpha\sqrt{r}t + (1-\alpha)\sqrt{s}t\right)\right\},\right. \\
&\qquad \left.\exp\left\{-\frac{1}{4}\eta M\right\} E\exp\left\{C\eta\frac{\|\mathbf{W}\|_\infty}{1-\alpha}\left(\alpha\sqrt{r}t + (1-\alpha)\sqrt{s}t\right)\right\}\right\}
\end{aligned}$$

$$\leq \min\left\{(20)^{d_1+d_2}\exp\left\{-\frac{1}{4}\eta M\right\}\cdot\exp\left\{\frac{C\eta^2(\alpha^2 r+(1-\alpha)^2 s)t^2}{n\alpha^2}\right\},\right.$$
$$\left.2p\exp\left\{-\frac{1}{4}\eta M\right\}\cdot\exp\left\{\frac{C\eta^2(\alpha^2 r+(1-\alpha)^2 s)t^2}{n(1-\alpha)^2}\right\}\right\} \tag{S1.8}$$
$$\leq \min\left\{(20)^{d_1+d_2}\exp\left\{-\frac{CM^2 n\alpha^2}{(\alpha^2 r+(1-\alpha)^2 s)t^2}\right\},\right.$$
$$\left.2p\exp\left\{-\frac{CM^2 n(1-\alpha)^2}{(\alpha^2 r+(1-\alpha)^2 s)t^2}\right\}\right\}, \tag{S1.9}$$

where (S1.5) uses Markov's inequality, (S1.6) uses the contraction property of the Rademacher process (see Theorem 2.3 in Koltchinskii (2011)), (S1.7) is obtained Lemma 1 and that any $\mathbf{\Delta}\in\mathbb{C}$ satisfies $\alpha\|\mathbf{\Delta}\|_{op}+(1-\alpha)\|\mathbf{\Delta}\|_\infty \leq 4(\alpha\|\mathbf{\Delta}'\|_{op}+(1-\alpha)\|\mathbf{\Delta}_{\mathcal{S}}\|_\infty) \leq C(\alpha\sqrt{r}\|\mathbf{\Delta}\|_F+(1-\alpha)\sqrt{s}\|\mathbf{\Delta}\|_F)$, (S1.8) uses Lemma 2 and (S1.9) is obtained by setting $\eta \asymp \frac{Mn\alpha^2}{(\alpha^2 r+(1-\alpha)^2 s)t^2}$ for the first term and $\eta \asymp \frac{Mn(1-\alpha)^2}{(\alpha^2 r+(1-\alpha)^2 s)t^2}$ for the second term.

Finally, taking $M \asymp \min\left\{t\frac{(\alpha\sqrt{r}+(1-\alpha)\sqrt{s})}{\alpha}\sqrt{\frac{d_1+d_2}{n}}, t\frac{(\alpha\sqrt{r}+(1-\alpha)\sqrt{s})}{1-\alpha}\sqrt{\frac{\log p}{n}}\right\}$ proves the lemma. □

# S2 Condition C3

**Lemma 4.** *Suppose the conditional density $f_{y_i|\mathbf{Z}_i}$ satisfies $f_{y_i|\mathbf{Z}_i}(\langle\mathbf{B},\mathbf{Z}_i\rangle) > \underline{f} > 0$ and $|f'_{y_i|\mathbf{Z}_i}(\cdot)| \leq \overline{f'}$, matrix $\mathbf{J} = E[\mathbf{z}_i\mathbf{z}_i^\top]$ is positive definite and its minimum eigenvalue is denoted by $\sigma_{\min}(\mathbf{J})$, and the restricted nonlinear* *impact coefficient*

$$q := \frac{3}{2}\frac{\underline{f}^{\frac{3}{2}}}{\overline{f'}} \inf_{\mathbf{\Delta}\in\mathbb{C}} \frac{(E|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle|^2)^{\frac{3}{2}}}{E|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle|^3} > 0.$$

*We have* $E[\rho_\tau(y_i-\langle\mathbf{B}+\mathbf{\Delta},\mathbf{Z}_i\rangle)]-E[\rho_\tau(y_i-\langle\mathbf{B},\mathbf{Z}_i\rangle)] \geq \frac{1}{4}\underline{f}^{\frac{1}{2}}\sigma_{\min}^{\frac{1}{2}}(\mathbf{J})\left(\underline{f}^{\frac{1}{2}}\sigma_{\min}^{\frac{1}{2}}(\mathbf{J})\|\mathbf{\Delta}\|_F^2 \wedge q\|\mathbf{\Delta}\|_F\right).$

**Proof.** By Knight's identity

$$\rho_\tau(u-v)-\rho_\tau(u) = -v\psi_\tau(u) + \int_0^v \left(I\{u\leq s\}-I\{u\leq 0\}\right)ds,$$

where $\psi_\tau(u)=\tau-I\{u<0\}$, we have

$$\begin{aligned}
&E[\rho_\tau(y_i-\langle\mathbf{B}+\mathbf{\Delta},\mathbf{Z}_i\rangle)]-E[\rho_\tau(y_i-\langle\mathbf{B},\mathbf{Z}_i\rangle)]\\
=&E\int_0^{\langle\mathbf{\Delta},\mathbf{Z}_i\rangle}\left(I\{y_i-\langle\mathbf{B},\mathbf{Z}_i\rangle\leq t\}-I\{y_i-\langle\mathbf{B},\mathbf{Z}_i\rangle\leq 0\}\right)dt\\
=&E\int_0^{\langle\mathbf{\Delta},\mathbf{Z}_i\rangle}\left[F_{y_i|\mathbf{Z}_i}(\langle\mathbf{B},\mathbf{Z}_i\rangle+t)-F_{y_i|\mathbf{Z}_i}(\langle\mathbf{B},\mathbf{Z}_i\rangle)\right]dt\\
=&E\int_0^{\langle\mathbf{\Delta},\mathbf{Z}_i\rangle}\left[f_{y_i|\mathbf{Z}_i}(\langle\mathbf{B},\mathbf{Z}_i\rangle)t+\frac{1}{2}f'_{y_i|\mathbf{Z}_i}(\langle\mathbf{B},\mathbf{Z}_i\rangle+\delta t)t^2\right]dt\\
\geq&\frac{1}{2}\underline{f}E\left|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^2-\frac{1}{6}\overline{f'}E\left|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^3\\
=&\frac{1}{4}\underline{f}E\left|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^2+\frac{1}{4}\underline{f}E\left|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^2-\frac{1}{6}\overline{f'}E\left|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^3.
\end{aligned}$$

When $\left(\underline{f}E\left|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^2\right)^{\frac{1}{2}}\leq q$, we have $\frac{1}{4}\underline{f}E\left|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^2\geq\frac{1}{6}\overline{f'}E\left|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^3$, and then

$$E[\rho_\tau(y_i-\langle\mathbf{B}+\mathbf{\Delta},\mathbf{Z}_i\rangle)]-E[\rho_\tau(y_i-\langle\mathbf{B},\mathbf{Z}_i\rangle)]\geq\frac{1}{4}\underline{f}E\left|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^2\geq\frac{1}{4}\underline{f}\sigma_{\min}(\mathbf{J})\|\mathbf{\Delta}\|_F^2. \tag{S2.10}$$

On the other hand, if $\left(\underline{f}E\left|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^2\right)^{\frac{1}{2}} > q$, let $\theta = \frac{q}{\left(\underline{f}E\left|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^2\right)^{\frac{1}{2}}}$. Thus $\left(\underline{f}E\left|\langle\theta\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^2\right)^{\frac{1}{2}} = q$. Then, we get

$$\begin{aligned}
&E[\rho_\tau(y_i - \langle\mathbf{B}+\mathbf{\Delta},\mathbf{Z}_i\rangle)] - E[\rho_\tau(y_i - \langle\mathbf{B},\mathbf{Z}_i\rangle)]\\
\geq&\frac{1}{\theta}E[\rho_\tau(y_i - \langle\mathbf{B}+\theta\mathbf{\Delta},\mathbf{Z}_i\rangle)] - E[\rho_\tau(y_i - \langle\mathbf{B},\mathbf{Z}_i\rangle)]\\
\geq&\frac{1}{\theta}\frac{1}{4}\underline{f}\theta^2E\left|\langle\mathbf{\Delta},\mathbf{Z}_i\rangle\right|^2\\
\geq&\frac{1}{4}\underline{f}^{\frac{1}{2}}q\sigma_{\min}^{\frac{1}{2}}(\mathbf{J})\|\mathbf{\Delta}\|_F,
\end{aligned}$$

where the first inequality follows from the convexity $\rho_\tau(\theta(y_i - \langle\mathbf{B}+\mathbf{\Delta},\mathbf{Z}_i\rangle) + (1-\theta)(y_i - \langle\mathbf{B},\mathbf{Z}_i\rangle)) \leq \theta\rho_\tau(y_i - \langle\mathbf{B}+\mathbf{\Delta},\mathbf{Z}_i\rangle) + (1-\theta)\rho_\tau(y_i - \langle\mathbf{B},\mathbf{Z}_i\rangle)$, and the second inequality follows from the first inequality of (S2.10), and the last one follows the definition of $\theta$. Therefore, we get

$$E[\rho_\tau(y_i - \langle\mathbf{B}+\mathbf{\Delta},\mathbf{Z}_i\rangle)] - E[\rho_\tau(y_i - \langle\mathbf{B},\mathbf{Z}_i\rangle)] \geq \frac{1}{4}\underline{f}^{\frac{1}{2}}\sigma_{\min}^{\frac{1}{2}}(\mathbf{J})\left(\underline{f}^{\frac{1}{2}}\sigma_{\min}^{\frac{1}{2}}(\mathbf{J})\|\mathbf{\Delta}\|_F^2 \wedge q\|\mathbf{\Delta}\|_F\right).$$

□

## Bibliography

Koltchinskii, V. *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*. New York: Springer (2011).

van der Geer, S. A. *Empirical Processes in M-Estimation*. Cambridge: Cambridge University Press (2000).

Van der Vaart, A. W. and Wellner, J. A. *Weak Convergence and Empirical Processes*. New York: Springer (1996).