Exercises 3

Exercise 1. 1. Let $A \in \mathbb{R}^{m\times n}$, let $I \in \mathbb{R}^{n\times n}$, $\lambda > 0$ and

$$K = \begin{bmatrix} A \\ \sqrt{\lambda} I \end{bmatrix} \in \mathbb{R}^{(m+n)\times n}.$$

Show that the singular values of K satisfy $\lambda_j \geqslant \sqrt{\lambda}$, $j = 1, \ldots, n$.

2. Let $A \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$ and $x_\lambda \in \mathbb{R}^n$ be minimizer of

$$\|Ax - y\|^2 + \lambda \|x\|^2, \quad \lambda > 0$$

Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be defined as

$$f(\lambda) = \|Ax_\lambda - y\|^2.$$

Show that

$$f'(\lambda) = 2\lambda \langle x_\lambda,\, (A^\top A + \lambda I)^{-1} x_\lambda \rangle.$$

Solution.

1. Since $K^*K = A^*A + \lambda I$, whose eigenvalues are $\sigma_i^2 + \lambda$, we have that $\lambda_i = \sqrt{\sigma_i^2 + \lambda} \geqslant \sqrt{\lambda}$
2. x_λ can be expanded in terms of the SVD as

$$x_\lambda = \sum_{i=1}^{n} \frac{\sigma_i}{\sigma_i^2 + \lambda} \langle y,\, u_i \rangle v_i$$

Hence, using that $Av_i = \sigma_i u_i$, and $y = \sum \langle y,\, u_i \rangle u_i$, we have that

$$\|Ax_\lambda - y\|^2 = \sum_{i=1}^{n} \left(\frac{\sigma_i^2}{\sigma_i^2 + \lambda} - 1 \right)^2 \langle y,\, u_i \rangle^2$$

Derivating with respect to λ yields

$$\begin{aligned} f'(\lambda) &= 2 \sum_{i=1}^{n} \left(1 - \frac{\sigma_i^2}{\sigma_i^2 + \lambda} \right) \frac{\sigma_i^2}{(\sigma_i^2 + \lambda)^2} \langle y,\, u_i \rangle^2 \\ &= 2\lambda \sum_{i=1}^{n} \frac{\sigma_i^2}{(\sigma_i^2 + \lambda)^3} \langle y,\, u_i \rangle^2 \\ &= 2\lambda \sum_{i=1}^{n} \frac{\sigma_i \langle y,\, u_i \rangle}{\sigma_i^2 + \lambda} \cdot \frac{1}{\sigma_i^2 + \lambda} \frac{\sigma_i \langle y,\, u_i \rangle}{\sigma_i^2 + \lambda} \\ &= 2\lambda \langle V^\top x_\lambda,\, (\Sigma^2 + \lambda I)^{-1} V^\top x_\lambda \rangle \\ &= 2\lambda \langle x_\lambda,\, (A^\top A + \lambda I)^{-1} x_\lambda \rangle \end{aligned}$$

■

Exercise 2. Let $A \in \mathbb{R}^{m\times n}$ such that $\operatorname{Ker} A = \{0\}$ ($m > n$, and A full column rank). Let x be the solution of $Ax = y$ for some $y \in \operatorname{Ran} A$, and let $y^\delta \in \mathbb{R}^m$ such that $\|y^\delta - y\| \leqslant \delta$.

We define

$$q(\lambda, \sigma) = \begin{cases} 1 & \text{if} \quad \sigma^2 \geqslant \lambda \\ 0 & \text{if} \quad \sigma^2 < \lambda \end{cases}$$

and the operator R_λ such that

$$R_\lambda y = \sum_{i=1}^{n} \frac{q(\lambda, \sigma_i)}{\sigma_i} \langle y,\, u_i\rangle v_i = \sum_{\sigma_i^2 \geqslant \lambda} \frac{1}{\sigma_i} \langle y,\, u_i\rangle v_i,$$

where $\{\sigma_i, u_i, v_i\}$ are the singular values and vectors of A. Let $x_\lambda^\delta = R_\lambda y^\delta$.

1. Let $\lambda = \lambda(\delta) = c\delta^\theta$ where $c > 0$ and $0 < \theta < 2$. Show that

$$\|x_\lambda^\delta - x\| \to 0 \quad \text{where} \quad \delta \to 0$$

2. Assume that $x = A^\top z$ for some $z \in \mathbb{R}^m$ (in fact this is true because $\operatorname{Ran} A^\top = \operatorname{Ran} A^\dagger = \operatorname{Span}(v_1, \ldots, v_r)$.) Deduce the θ for which the convergence of x_λ^δ towards x when $\delta \to 0$ is optimal. What is the corresponding rate?

3. Assume that $x = A^\top A w$ for some $w \in \mathbb{R}^n$ (in fact this is true because $\operatorname{Ran} A^\top A = \operatorname{Ran} A^\top = \operatorname{Span}(v_1, \ldots, v_r)$.) Deduce the θ for which the convergence of x_λ^δ is optimal. What is the corresponding rate?

Solution.

1. Let $x_\lambda = R_\lambda y$. We have

$$\|x_\lambda^\delta - x\|^2 \leqslant \|x_\lambda^\delta - x_\lambda\|^2 + \|x_\lambda - x\|^2$$

First,

$$\|x_\lambda^\delta - x_\lambda\|^2 = \|\sum_{\sigma_i^2 \geqslant c\delta^\theta} \frac{1}{\sigma_i}\langle y^d - y,\, u_i\rangle v_i\|^2 \leqslant \frac{1}{c\delta^\theta}\|y^\delta - y\|^2 \leqslant \frac{\delta^{2-\theta}}{c} \to_{\delta\to 0} 0$$

Second,

$$\|x_\lambda - x\|^2 = \|\sum_i \frac{q(c\delta^\theta, \sigma_i) - 1}{\sigma_i}\langle y,\, u_i\rangle v_i\| \to_{\delta\to 0} 0$$

2. Writing $x = A^\top z$, we have $x = \sum_{i=1}^n \sigma_i \langle z,\, u_i\rangle v_i$, and therefore

$$\begin{aligned} x_\lambda - x &= \sum_{i=1}^{n} \left(\frac{q(c\delta^\theta, \sigma_i)}{\sigma_i}\langle y,\, u_i\rangle - \sigma_i\langle z,\, u_i\rangle \right) v_i \\ &= \sum_{\sigma_i^2 \geqslant c\delta^\theta} \left(\frac{1}{\sigma_i}\langle y,\, u_i\rangle - \sigma_i\langle z,\, u_i\rangle \right) v_i - \sum_{\sigma_i^2 < c\delta^\theta} \sigma_i\langle z,\, u_i\rangle v_i \\ &= 0 - \sum_{\sigma_i^2 < c\delta^\theta} \sigma_i\langle z,\, u_i\rangle v_i \end{aligned}$$

where the last equation can be found by writing $y = AA^\top z$ and using $A^\top u_i = \sigma_i v_i$ for $i = 1, \ldots, n$. Hence

$$\|x_\lambda - x\|^2 \leqslant c\delta^\theta \|z\|^2$$

Therefore, combining everything,

$$\|x_\lambda^\delta - x\| \leqslant \frac{\delta^{1-\theta/2}}{\sqrt{c}} + \sqrt{c}\delta^{\theta/2}\|z\|$$

The rate is optimal when $\theta = 1$.

3. The reasoning is strictly identical, and we obtain

$$\|x_\lambda^\delta - x\| \leqslant \frac{\delta^{1-\theta/2}}{\sqrt{c}} + c\delta^{\theta}\|z\|$$

The optimal rate satisfy $\theta = 1 - \theta/2$: it is attained for $\theta = 2/3$.

■

Exercise 3 (Duality).

1. Compute the dual problem of the least-squares problem with Tikhonov regularization.
2. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. We define its dual norm as

$$\forall z \in \mathbb{R}^n, \quad \|z\|_* = \sup\left\{z^\top x \;;\; \|x\| \leqslant 1\right\}$$

- Show that the dual norm of the Euclidean norm $\|\cdot\|_2$ is the Euclidean norm
- Show that the dual norm of $\|\cdot\|_\infty$ is $\|\cdot\|_1$.
- Show that the dual norm of $\|\cdot\|_1$ is $\|\cdot\|_\infty$.

Solution.

1. The problem can be reformulated as

$$\min \frac{1}{2}\|z - y\|^2 + \frac{1}{2}\|x\|^2 \quad \text{s.t.} \quad z = Ax$$

The Lagrangian for this problem is

$$\mathcal{L}(x, z, \nu) = \frac{1}{2}\|z - y\|^2 + \frac{\lambda}{2}\|x\|^2 + \nu^\top(Ax - z)$$

Minimizing over x yields $\tilde{x} = -\frac{1}{\lambda}A^\top \nu$

Minimizing over z yields $\tilde{z} = \nu$.

Hence,

$$\begin{aligned} g(\nu) = \min_{x,z} \mathcal{L}(x, z, \nu) = \mathcal{L}(\tilde{x}, \tilde{z}, \nu) &= -\nu^\top y - \frac{1}{2}\|\nu\|^2 - \frac{1}{2\lambda}\|A^\top \nu\|^2 + \frac{1}{2}\|y\|^2 \\ &= -\nu^\top y - \frac{1}{2}\nu^\top(\lambda^{-1}AA^\top + I)\nu + \frac{1}{2}\|y\|^2 \end{aligned}$$

and the dual problem is simply $\max g(\nu)$

2. - By Cauchy-Schwarz, we have $|z^\top x| \leqslant \|z\|\|x\|$ with equality if $x = z/\|z\|_2$. Hence

$$\sup\left\{z^\top x\ ;\ \|x\|_2 \leqslant 1\right\} = \|z\|_2.$$

   - Let x such that $\|x\|_\infty \leqslant 1$. Then $|z^\top x| \leqslant \sum |z_i||x_i| \leqslant \sum |z_i| = \|z\|_1$, with equality if $x_i = \mathrm{sign}(z_i)$ for all i. Hence

$$\sup\left\{z^\top x\ ;\ \|x\|_\infty \leqslant 1\right\} = \|z\|_1.$$

   - Let x such that $\|x\|_1 \leqslant 1$. Then $|z^\top x| \leqslant \sum |z_i||x_i| \leqslant |z_{i_{\max}}|\|x\|_1 \leqslant \|z\|_\infty$, with equality if $x = e_{i_{\max}}$

$$\sup\left\{z^\top x\ ;\ \|x\|_1 \leqslant 1\right\} = \|z\|_\infty.$$

■