Exercises 4

Exercise 1. Show that, for $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$,

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad a^\top x = b$$

always has a one-sparse vector as minimizer. (This is perhaps the simplest illustration of the general phenomenon that solutions of undetermined ℓ^1 minimization problems tend to be sparse.)

Solution. We have

$$|b| = \left|\sum a_i x_i\right| \leqslant \sum |a_i| \leqslant \|a\|_\infty \|x\|_1$$

and equality holds if $x_i = 0$ for all $i \neq i_{\max}$ (where $|a_{i_{\max}}| = \|a\|_\infty$), and $x_{i_{\max}} = \pm\|a\|_\infty^{-1}$. Hence setting

$$\begin{cases} x_i = 0 \quad \text{for} \quad i \neq i_{\max} \\ x_{i_{\max}} = \operatorname{sign}(b)\|a\|_\infty^{-1} \end{cases}$$

yields a solution to the problem. ■

Exercise 2. Compute the dual of the TV denoising problem

$$\min \frac{1}{2}\|x - y\|^2 + \lambda\|Bx\|_1.$$

Solution. We can reformulate this problem as

$$\min \frac{1}{2}\|x - y\|^2 + \lambda\|z\|_1 \quad \text{s.t.} \quad z = Bx$$

The Lagrangian reads

$$\mathcal{L}(x, z, \nu) = \frac{1}{2}\|x - y\|^2 + \lambda\|z\|_1 + \nu^\top(Bx - z).$$

Canceling the gradient with respect to x yields

$$x = y - B^\top \nu$$

Minimizing with respect to z yields

$$\lambda\|z_*\|_1 - \nu^\top z_* = \begin{cases} 0 \quad \text{if} \quad \|\nu\|_\infty \leqslant \lambda \\ -\infty \quad \text{otherwise} \end{cases}$$

Therefore, the dual problem is given by

$$\max \nu^\top By - \frac{1}{2}\|B^\top \nu\|^2 \quad \text{s.t.} \quad \|\nu\|_\infty \leqslant \lambda$$

■

Exercise 3. 1. Compute the proximal operator of the following functions:

- $f(x) = \frac{1}{2}\|Ax - y\|^2$
- $f(x) = \mathbb{1}_{Ax=y}(x)$, where A has full row-rank
- $f(x) = -\sum \log(x_i)$ (with $x_i > 0$)

2. Solve

$$\min_x \frac{1}{2}\|Bx - z\|_2^2 \quad \text{s.t.} \quad Ax = y$$

assuming that the solution exists and is unique.

Solution.

1. - We have

$$x_* = \operatorname{argmin}_x \frac{1}{2}\|z - x\|^2 + \frac{\gamma}{2}\|Ax - y\|^2 \iff x_* - z + \gamma A^\top(Ax_* - y) = 0$$
$$\iff x_* = (I + \gamma A^\top A)^{-1}(z + \gamma A^\top y)$$

- The KKT conditions for

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

give

$$\begin{cases} x - z \in \operatorname{Ran} A^\top \\ Ax = y \end{cases}$$

Therefore

$$\begin{cases} x - z = A^\top \nu \\ Ax = y \end{cases} \implies y - Az = AA^\top \nu \implies \nu = (AA^\top)^{-1}(y - Az)$$

and thus $x = z + A^\top(AA^\top)^{-1}(y - Az)$. $AA^\top$ is invertible because A has full row-rank, and in that case $A^\top(AA^\top)^{-1} = A^\dagger$.

- We want to minimize, for each i, $(z_i - x_i)^2/2 - \gamma \log x_i$ with respect to x_i, which gives

$$x_i - z_i + \frac{\gamma}{x_i} = 0 \iff x_i^2 - z_i x_i - \gamma = 0$$
$$\iff x_i = \frac{z_i + \sqrt{z_i^2 + 4\gamma}}{2}$$

since $x_i > 0$.

2. The optimality conditions read

$$\begin{cases} B^\top(Bx - z) = A^\top \nu \\ Ax = y \end{cases} \iff \begin{cases} B^\top Bx - A^\top \nu = B^\top z \\ Ax = y \end{cases}$$

which can be written

$$\begin{bmatrix} B^\top B & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \nu \end{bmatrix} = \begin{bmatrix} B^\top z \\ y \end{bmatrix}.$$

If there is a unique solution, it then given by

$$\begin{bmatrix} x \\ \nu \end{bmatrix} = \begin{bmatrix} B^\top B & A^\top \\ A & 0 \end{bmatrix}^{-1} \begin{bmatrix} B^\top z \\ y \end{bmatrix}$$

■

Exercise 4 (Elastic Net). The "Elastic Net" regularized problem combines both ℓ^1 and ℓ^2 regularization. It si given as

$$\min_x \|y - Ax\|^2 + \lambda_2 \|x\|_2^2 + \lambda_1 \|x\|_1$$

1. What is a suitable decomposition of the objective to perform proximal splitting on this problem?
2. Describe forward-backward updates that can be used to solve it.

Solution.

1. Several decompositions of the objective are possible, but the simplest one is to consider

$$F(x) = \|y - Ax\|^2 + \lambda_2 \|x\|_2^2$$
$$G(x) = \lambda_1 \|x\|_1$$

2. The proximal operator for $\|x\|_1$ is the soft-thresholding operator S_γ (with $\gamma = \lambda_1$ here). The gradient step on the smooth part of the objective reads

$$x_{k+1} = x_k - \tau 2A^\top (Ax_k - y) - 2\tau\lambda_2 x_k = x_k - \tau \left((A^\top A + \lambda_2 I)x_k + A^\top y\right)$$

hence the forward-backward update reads

$$x_{k+1} = S_{\lambda_1}\left(x_k - \tau((A^\top A + \lambda_2 I)x_k + A^\top y)\right)$$

■