Exercises 5

Exercise 1 (Maximum likelihood estimation).

1. Let $x \in \mathbb{R}$, and $y_i = x + w_i$ where $w_i \sim \mathcal{N}(0, 1)$. Give the Maximum Likelihood Estimator of x, *i.e.*

$$\hat{x} = \mathrm{argmax}_x \, p(y|x)$$

2. Same question assuming now a multiplicative Gaussian noise, *i.e.* $y_i \sim x w_i$ with $w_i \sim \mathcal{N}(0, 1)$.

Solution.

1. If $y = x + w$ with $w \sim_{i.i.d.} \mathcal{N}(0, 1)$, then $y \sim_{i.i.d.} \mathcal{N}(x, 1)$, and hence

$$p(y|x) = C \prod_{i=1}^{n} \exp\left(-\frac{(y_i - x)^2}{2}\right)$$

Taking the log of this expression yields

$$\ln(C) - \frac{1}{2} \sum_{i=1}^{n} (y_i - x)^2$$

Maximizing with respect to x yields

$$\hat{x} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

2. Similarly, $y \sim_{i.i.d.} \mathcal{N}(0, x^2)$, and hence

$$p(y|x) = \frac{C}{x^n} \prod_{i=1}^{n} \exp\left(-\frac{y_i^2}{2x^2}\right).$$

Take the log

$$\ln(C) - n \ln(x) - \frac{1}{2x^2} \sum_{i=1}^{n} y_i^2$$

Differentiating with respect to x and canceling yields

$$-\frac{n}{x} + \frac{1}{x^3} \sum y_i^2 = 0$$

and hence $x = \frac{1}{\sqrt{n}} \|y\|_2$

$\blacksquare$

Exercise 2. Let $A \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$, $\eta > 0$ and let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^m$. Show that the solution of the optimization problem

$$x_* = \operatorname{argmin} \|z\|_1 \quad \text{s.t.} \quad \|Az - y\| \leqslant \eta$$

is m-sparse in the case of the uniqueness of the solution. *Hint:* show that the system of columns $\{a_j \ ; \ j \in \operatorname{Supp} x_*\}$ is linearly independent.

Solution. Suppose that there exists $v \in \operatorname{Ker} A$ such that $v \neq 0$ and $\operatorname{Supp} v = \operatorname{Supp} x_* = I$. Then for t sufficiently small

$$\begin{aligned}\|x_*\|_1 < \|x_* + tv\|_1 &= \sum_I \operatorname{sign}(x_i + tv_i)(x_i + tv_i) \\ &= \sum_I \operatorname{sign}(x_i)(x_i + tv_i) \\ &= \|x_*\|_1 + t \sum_I \operatorname{sign}(x_i) v_i\end{aligned}$$

and we can choose k such that the second term is negative, which leads to a contradiction. Hence A_I is injective, and since necessarily $\operatorname{rank} A \leqslant m$, the solution is at most m-sparse (for any larger subset I, A_I can't be injective since $\operatorname{rank} A \leqslant m$). ∎

Exercise 3 (Null Space Property).

1. Prove the uniform recovery theorem: every k-sparse x_0 is the unique solution of

$$\min \|x\|_1 \quad \text{s.t.} \quad Ax = Ax_0. \tag{BP}$$

if and only if A satisfies the Null Space Property of order k, *i.e.* :

$$\forall I : |I| \leqslant k, \quad \forall h \in \operatorname{Ker} A \setminus 0, \quad \|h_I\|_1 < \|h_{I^c}\|_1$$

2. Show that if NSP(k) holds, then the solution of (BP) is also a solution of

$$\min \|x\|_0 \quad \text{s.t.} \quad Ax = Ax_0$$

3. Let $x_0 \in \mathbb{R}^n$ (not necessarily k-sparse), let $y = Ax_0 + w$, $\|w\| \leqslant \varepsilon$, and let x be a solution of

$$\min \|x\|_1 \quad \text{s.t.} \quad \|Ax - y\| \leqslant \varepsilon \tag{BP-ε}$$

The goal of this question is to prove the uniform robust recovery theorem: if A obeys the robust NSP of order k, *i.e.*

$$\exists 0 < \rho < 1, \quad \exists \tau > 0, \quad \forall I : |I| = k, \quad \forall h \in \operatorname{Ker} A \setminus 0, \quad \|h_I\|_1 \leqslant \rho \|h_{I^c}\|_1 + \tau \|Ah\|_2$$

then for all $x_0 \in \mathbb{R}^n$, any solution of (BP-ε) satisfies

$$\|x - x_0\|_1 \leqslant 2\frac{1+\rho}{1-\rho}\sigma_k(x_0)_1 + 4\frac{\tau}{1-\rho}\varepsilon$$

where $\sigma_k(x_0)_1 := \inf\{\|x_0 - z\|_1 \ ; \ \|z\|_0 \leqslant k\}$ is the best k-sparse approximation with respect to the ℓ^1-norm. We assume that A satisfies the robust NSP of order k.

(a) Let $h = x - x_0$. Show that, for any subset I,

$$\|x_0\|_1 + \|h_{I^c}\| \leqslant 2\|(x_0)_{I^c}\|_1 + \|h_I\|_1 + \|x\|_1$$

(b) Deduce that for a well chosen subset I,

$$\|h_{I^c}\|_1 \leqslant \frac{1}{1-\rho}(2\sigma_k(x_0)_1 + 2\tau\varepsilon)$$

(c) Conclude.

Solution.

1. Assume that every k-sparse vector x_0 is the unique solution of

$$\min \|x\|_1 \quad \text{s.t.} \quad Ax = Ax_0.$$

Let $k \in \mathbb{N}$, and let I such that $|I_0| = k$. Let $h \in \operatorname{Ker} A \setminus 0$, then h_I is the unique solution of the ℓ^1-minimization with $Ax = Ah_I$. But $Ah = A(h_I + h_{I^c}) = 0$, hence $A(-h_{I^c}) = Ah_I$ and thus $\|h_I\|_1 < \|h_{I^c}\|$. Reciprocally, if A satisfies NSP(k), then for x minimizer, let $h = x - x_0$, and we have, for $I = \operatorname{Supp} x_0$, (hence $|I| \leqslant k$),

$$\|x\|_1 = \|x_0 + h\|_1 = \|x_0 + h_I\|_1 + \|h_{I^c}\| \geqslant \|x_0\|_1 - \|h_I\|_1 + \|h_{I^c}\|_1 > \|x_0\|_1$$

unless $h = 0$, *i.e.* unless $x = x_0$.

2. Let z be a minimezer of

$$\min \|x\|_0 \quad \text{s.t.} \quad Ax = Ax_0$$

Then

$$\|z\|_0 \leqslant \|x\|_0 = k$$

so z is k-sparse and obeys $Az = Ax_0$. Therefore from NSP(k), z is the unique minimizer of (BP), hence $z = x_0$.

3. (a) We have

$$\begin{aligned}
\|x_0\|_1 + \|h_{I^c}\| &= \|(x_0)_{I^c}\| + \|(x_0)_I\| + \|h_{I^c}\|_1 \\
&\leqslant \|(x_0)_{I^c}\| + \|h_I\| + \|x_I\|_1 + \|h_{I^c}\|_1 \\
&\leqslant \|(x_0)_{I^c}\| + \|h_I\| + \|x_I + x_{I^c}\|_1 + \|(x_0)_{I^c}\|_1
\end{aligned}$$

(b) Let I be the support of the k largest entries of x_0. Then $|I| \leqslant k$, and $\|(x_0)_{I^c}\|_1 = \sigma_k(x_0)_1$. Additionally, from the robust NSP,

$$\|h_I\|_1 \leqslant \rho\|h_{I^c}\|_1 + \tau\|Ah\| \leqslant \rho\|h_{I^c}\|_1 + \tau(\|Ax - y\| + \|y - Ax_0\|) \leqslant \rho\|h_{I^c}\|_1 + 2\tau\varepsilon.$$

Hence, using the inequality from the previous question we obtain

$$\|x_0\|_1 + (1-\rho)\|h_{I^c}\|_1 \leqslant 2\sigma_k(x_0)_1 + 2\tau\varepsilon + \|x\|_1$$

Noting that $\|x\|_1 - \|x_0\|_1 \leqslant 0$ (since x is a minimizer), we obtain the desired inequality.

4. From the robust NSP we obtain

$$\|h_I\|_1 \leqslant \rho \|h_{I^c}\|_1 + \tau \|Ah\| \leqslant \frac{\rho}{1-\rho}(2\sigma_k(x_0)_1 + 2\tau\varepsilon) + 2\tau\varepsilon \leqslant \frac{\rho}{1-\rho} 2\sigma_k(x_0)_1 + \frac{1}{1-\rho} 2\tau\varepsilon$$

which finally yields

$$\|h\|_1 = \|h_I\|_1 + \|h_{I^c}\|_1 \leqslant 2\frac{1+\rho}{1-\rho}\sigma_k(x_0)_1 + 4\frac{\tau}{1-\rho}\varepsilon$$

■