Non-convex inverse problems

March 1, 2024

Correction

Exercise 1

Imagine a computer criminal, which wants to read an email containing sensitive information. The email is encrypted. If the criminal manages to intercept the email, then she faces an inverse problem : she must identify the original text (which is the unknown) from its image through the encryption procedure (which is the observation).

Exercise 2

[Caution : this is a *non-linear* inverse problem. Therefore, it cannot be analyzed with the result on *linear* inverse problems from our first lecture.]

Reconstruction is unique : for any $(x_1, x_2) \in \mathbb{R}^2$ and associated measurements (y_1, y_2), it holds $(x_1, x_2) = (y_1, (1 + y_1^2)y_2)$. Therefore, the measurements (y_1, y_2) uniquely determine (x_1, x_2).

Reconstruction is not stable. Indeed, there exist pairs (x_1, x_2) and (x'_1, x'_2), with associated measurements $(y_1, y_2), (y'_1, y'_2)$ such that

$$||(y_1, y_2) - (y'_1, y'_2)||_2 \ll ||(y_1, y_2)||_2$$

but

$$||(x_1, x_2) - (x'_1, x'_2)||_2 \not\ll ||(x_1, x_2)||_2.$$

To show it, we can consider the pair $(x_1, x_2) = (t, t)$, for t large, and define $(x'_1, x'_2) = (t, 0)$. Then

$$\frac{||(x_1, x_2) - (x'_1, x'_2)||_2}{||(x_1, x_2)||_2} = \frac{t}{\sqrt{2}t} = \frac{1}{\sqrt{2}} \not\ll 1$$

while

$$\begin{aligned} \frac{||(y_1, y_2) - (y'_1, y'_2)||_2}{||(y_1, y_2)||_2} &= \frac{\frac{t}{1+t^2}}{t\sqrt{1 + \frac{1}{(1+t^2)^2}}} \\ &= \frac{1}{(1+t^2)\sqrt{1 + \frac{1}{(1+t^2)^2}}} \\ &\sim \frac{1}{t^2} \\ &\ll 1. \end{aligned}$$

Exercise 3

1. It is a non-convex problem, because $\mathcal{E}_k$ is not-convex. Indeed, $\mathcal{E}_k$ contains $\mathbb{R}e_E$ for all $E \subset \{1, \dots, d\}$, therefore $\text{Conv}(\mathcal{E}_k) \supset \text{Vect}\left(\{e_E\}_{E\subset\{1,\dots,d\}}\right) = \mathbb{R}^d$. But $\mathcal{E}_k \neq \mathbb{R}^d$ (it is a finite union of k-dimensional vector subspaces of $\mathbb{R}^d$, and $k < d$), so $\mathcal{E}_k$ is different from its convex hull.

2. First, let x be any point of this set, which we denote $\mathcal{M}$. We assume that $x \neq e_E$ for all E and show that x is not extremal. For all i, $x_i \in [-1;1]$. Since $x \neq e_E$ for all E, there exists an index i such that $x_i \neq -1$ and $x_i \neq 1$ (i.e. $x_i \in]-1;1[$). We define $y, z \in \mathbb{R}^d$ such that

$$\begin{aligned} y_j &= z_j = x_j, \forall j \neq i, \\ y_i &= 1, \\ z_i &= -1. \end{aligned}$$

Then $x = (1-t)y + tz$ for $t = \frac{1-x_i}{2} \in [0;1]$. The vectors y, z belong to $\mathcal{M}$ and are different from x. Therefore, x is not extremal.
Now, let us fix $E \subset \{1, \dots, d\}$. The vector e_E is in $\mathcal{M}$. Let us show that it is extremal. Let $y, z \in \mathcal{M}$ and $t \in [0;1]$ be such that $e_E = (1-t)y + tz$. We must show that either y or z is equal to e_E.
If $t = 0$, then $y = e_E$. If $t = 1$, then $z = e_E$. Let us assume $0 < t < 1$ and show that $y = z = e_E$. For all $i \leq d$, if $i \in E$, then $(e_E)_i = 1$. Since $y_i \leq 1$ and $z_i \leq 1$, it holds

$$1 = (e_E)_i = (1-t)y_i + tz_i \leq (1-t) + t = 1.$$

The inequality must be an equality, meaning that $y_i = z_i = 1 = (e_E)_i$. The same reasoning applies if $i \notin E$. It shows that y, z and e_E are equal.

3. The problem we want to solve is (assuming that a solution exists)

$$\begin{aligned} &\text{minimize } ||x||_{reg} \\ &\text{over } x \in \mathbb{R}^d \text{ such that } Ax = y, \end{aligned}$$

where, for any x, $||x||_{reg} = \min\{s \in \{1, \dots, d\}, x \in \mathcal{E}_s\}$.
Following the intuition discussed in the lecture, since the vectors e_E are the extremal points of the ℓ^∞ ball, it makes sense to approximate $||.||_{reg}$ with the infinity ball, yielding the convex problem

$$\begin{aligned} &\text{minimize } ||x||_\infty \\ &\text{over } x \in \mathbb{R}^d \\ &\text{such that } Ax = y. \end{aligned} \tag{Relax-Reg}$$

4. Mimicking the definition of k-restricted isometry constant for sparse recovery, we could define the k-restricted isometry constant of A as the smallest positive number $\delta_k > 0$ (if it exists) such that, for all $x \in \mathcal{E}_k$,

$$(1-\delta_k)||x||_2 \leq ||Ax||_2 \leq (1+\delta_k)||x||_2.$$

5. a) Problem (Relax-Reg) can be rewritten as a min-max problem :

$$\begin{aligned} &\min_{x\in\mathbb{R}^d} \max_{z\in\mathbb{R}^m} ||x||_\infty + \langle y - Ax, z\rangle \\ &= \min_{x\in\mathbb{R}^d} \max_{z\in\mathbb{R}^m} \langle y, z\rangle + ||x||_\infty - \left\langle x, A^T z\right\rangle \end{aligned}$$

We compute the dual by exchanging the minimum and maximum :

$$\begin{aligned}
&\max_{z\in\mathbb{R}^m}\min_{x\in\mathbb{R}^d} \langle y,z\rangle + ||x||_\infty - \left\langle x, A^T z\right\rangle \\
&\qquad = \max_{z\in\mathbb{R}^m} \langle y,z\rangle + \min_{x\in\mathbb{R}^d}\left(||x||_\infty - \left\langle x, A^T z\right\rangle\right) \\
&\qquad = \max_{\substack{z\in\mathbb{R}^m \\ ||A^T z||_1\le 1}} \langle y,z\rangle .
\end{aligned}$$

Indeed, $\min_{x\in\mathbb{R}^d} ||x||_\infty - \left\langle x, A^T z\right\rangle = -\infty$ if $||A^T z||_1 > 1$: denoting $h \in \mathbb{R}^d$ the vector such that $h_i = 1$ if $(A^T z)_i \geq 0$ and $h_i = -1$ if $(A^T z)_i < 0$, we have, for all $t \geq 0$,

$$\begin{aligned}
||th||_\infty - \left\langle th, A^T z\right\rangle &= t - t||A^T z||_1 \\
&= -t(||A^T z||_1 - 1),
\end{aligned}$$

which goes to $-\infty$ when t goes to ∞.
On the other hand, if $||A^T z||_1 \leq 1$, then $||x||_\infty - \left\langle x, A^T z\right\rangle \geq ||x||_\infty - ||x||_\infty ||A^T z||_1 \geq 0$ for all $x \in \mathbb{R}^d$. Since $x = 0$ yields the value 0, the minimum is zero.
The dual problem is

$$\begin{aligned}
\text{maximize } & \langle y,z\rangle , \\
\text{over } & z\in\mathbb{R}^d \\
\text{such that } & ||A^T z||_1 \leq 1.
\end{aligned}$$

b) We must show that c is a feasible point for the dual problem, with the same objective value as the primal at x_0, that is

$$||x_0||_\infty = \langle y,c\rangle .$$

The vector c is feasible for the dual problem because $||A^T c||_1 = 1$. In addition,

$$\begin{aligned}
\langle y,c\rangle &= \langle Ax_0, c\rangle \\
&= \left\langle x_0, A^T c\right\rangle \\
&= \sum_{i=1}^d (x_0)_i (A^T c)_i \\
&= \sum_{i=1}^d |(x_0)_i||(A^T c)_i| \\
&= \sum_{i=1}^d ||x_0||_\infty |(A^T c)_i| \\
&= ||x_0||_\infty ||A^T c||_1 \\
&= ||x_0||_\infty .
\end{aligned}$$

The fourth equality is true because the components of x_0 and $A^T c$ have the same sign. The fifth one is true because, if $|(x_0)_i| < ||x_0||_\infty$, then $(A^T c)_i = 0$, hence in all situations, $|(x_0)_i||(A^T c)_i| = ||x_0||_\infty |(A^T c)_i|$.

Exercise 4

1. a) Let us define

$$\begin{array}{rcl} g: \mathbb{R}^d & \to & \mathbb{R} \\ x & \to & \frac{1}{2m}\sum_{k=1}^m \left(u_k^{(1)}\cos\left(\left\langle v_k^{(1)}, x\right\rangle\right) + u_k^{(2)}\cos\left(\left\langle v_k^{(2)}, x\right\rangle\right) - y_k\right)^2. \end{array}$$

If the inverse problem which consists in recovering x_{sol} from $y_1, \dots, y_m$ has a solution, then this problem is equivalent to finding a global minimizer of g over $\mathbb{R}^d$. Indeed, the minimum of g is 0 (since it is a nonnegative function, which is equal to 0 at x_{sol}). Therefore, a global minimizer is exactly a point at which g cancels, that is, a point at which, for all k,

$$u_k^{(1)}\cos\left(\left\langle v_k^{(1)}, x\right\rangle\right) + u_k^{(2)}\cos\left(\left\langle v_k^{(2)}, x\right\rangle\right) = y_k.$$

b) We can choose an arbitrary point x_0, then try to minimize g by gradient descent with backtracking linesearch.

2. a) $\text{Hess}f(x_{sol}) = \left(1 - e^{-2||x_{sol}||^2}\right) I_d + 4e^{-2||x_{sol}||^2} x_{sol}x_{sol}^T$.

b) For all x going to x_{sol},

$$\begin{aligned} &\langle \nabla f(x), x - x_{sol}\rangle \\ &= \left\langle \nabla f(x_{sol}) + \text{Hess}f(x_{sol})(x - x_{sol}) + O(||x - x_{sol}||^2), x - x_{sol}\right\rangle \\ &= \left\langle \text{Hess}f(x_{sol})(x - x_{sol}) + O(||x - x_{sol}||^2), x - x_{sol}\right\rangle \\ &= \langle \text{Hess}f(x_{sol})(x - x_{sol}), x - x_{sol}\rangle + O(||x - x_{sol}||^3) \\ &= \left\langle \left(1 - e^{-2||x_{sol}||^2}\right)(x - x_{sol}) + 4e^{-2||x_{sol}||^2} x_{sol}\langle x_{sol}, x - x_{sol}\rangle, x - x_{sol}\right\rangle \\ &\qquad + O(||x - x_{sol}||^3) \\ &= \left(1 - e^{-2||x_{sol}||^2}\right)||x - x_{sol}||^2 + 4e^{-2||x_{sol}||^2}\langle x_{sol}, x - x_{sol}\rangle^2 \\ &\qquad + O(||x - x_{sol}||^3). \end{aligned}$$

c) For all x close to x_{sol},

$$\begin{aligned} ||x - \mu\nabla f(x) - x_{sol}||^2 &= ||x - x_{sol}||^2 - 2\mu\langle \nabla f(x), x - x_{sol}\rangle + \mu^2||\nabla f(x)||^2 \\ &= ||x - x_{sol}||^2 - 2\mu\left(1 - e^{-2||x_{sol}||^2}\right)||x - x_{sol}||^2 \\ &\quad - 8\mu e^{-2||x_{sol}||^2}\langle x_{sol}, x - x_{sol}\rangle^2 + \mu^2||\nabla f(x)||^2 \\ &\qquad + O(||x - x_{sol}||^3) \\ &\leq ||x - x_{sol}||^2 - 2\mu\left(1 - e^{-2||x_{sol}||^2}\right)||x - x_{sol}||^2 \\ &\quad + \mu^2||\nabla f(x)||^2 + O(||x - x_{sol}||^3) \\ &= \left(1 - 2\mu\left(1 - e^{-2||x_{sol}||^2}\right)\right)||x - x_{sol}||^2 \\ &\quad + \mu^2||\nabla f(x)||^2 + O(||x - x_{sol}||^3). \end{aligned}$$

d) Let $\mu > 0$ be such that

$$\mu \leq \frac{1 - e^{-2||x_{sol}||^2}}{50}.$$

Then, from the previous question and the hint,

$$\begin{aligned}
||x - \mu\nabla f(x) - x_{sol}||^2 &\leq \left(1 - 2\mu\left(1 - e^{-2||x_{sol}||^2}\right)\right)||x - x_{sol}||^2 \\
&\quad + 25\mu^2||x - x_{sol}||^2 + O(||x - x_{sol}||^3) \\
&\leq \left(1 - 2\mu\left(1 - e^{-2||x_{sol}||^2}\right)\right)||x - x_{sol}||^2 \\
&\quad + \frac{\mu}{2}\left(1 - e^{-2||x_{sol}||^2}\right)||x - x_{sol}||^2 + O(||x - x_{sol}||^3) \\
&= \left(1 - \frac{3}{2}\mu\left(1 - e^{-2||x_{sol}||^2}\right)\right)||x - x_{sol}||^2 \\
&\quad + O(||x - x_{sol}||^3).
\end{aligned}$$

For x close enough to x_{sol}, the "$O(||x - x_{sol}||^3)$" term is smaller than

$$\frac{1}{2}\mu\left(1 - e^{-2||x_{sol}||^2}\right)||x - x_{sol}||^2.$$

Therefore, when x is close enough to x_{sol},

$$||x - \mu\nabla f(x) - x_{sol}||^2 \leq \left(1 - \mu\left(1 - e^{-2||x_{sol}||^2}\right)\right)||x - x_{sol}||^2.$$

e) Let $\rho > 0$ and $\mu_0 > 0$ be such that the inequality of the previous question holds true for all $x \in B(x_{sol}, \rho)$ and any $\mu \in]0; \mu_0]$.
Let us consider the gradient descent iterates $(x_t)_{t\in\mathbb{N}}$, for some $x_0 \in B(x_{sol}, \rho)$, with stepsize smaller than μ_0. Then, for each t,

$$||x_{t+1} - x_{sol}||^2 \leq \left(1 - \mu\left(1 - e^{-2||x_{sol}||^2}\right)\right)||x_t - x_{sol}||^2.$$

(The proof is by iteration over t : it is true for $t = 0$ from the inequality established in the previous question. Assuming it is true up to some $t-1$, it implies that $||x_t - x_{sol}|| \leq ||x_{t-1} - x_{sol}|| \leq \dots \leq ||x_0 - x_{sol}|| \leq \rho$. Therefore, the inequality of Question 2.d) can be applied to x_t, which shows the result for t.)
As a consequence, for all t,

$$||x_t - x_{sol}||^2 \leq \left(1 - \mu\left(1 - e^{-2||x_{sol}||^2}\right)\right)^t ||x_0 - x_{sol}||^2,$$

so that $||x_t - x_{sol}|| \to 0$ when $t \to +\infty$.

f) It is a local convergence result.

3. a) For any $\lambda \in \mathbb{R}$,

$$\begin{aligned}
&\nabla f(\lambda x_{sol}) \\
&= \left(-2e^{-2\lambda^2||x_{sol}||^2} + e^{-\frac{(\lambda+1)^2}{2}||x_{sol}||^2} + e^{-\frac{(\lambda-1)^2}{2}||x_{sol}||^2}\right)\lambda x_{sol}
\end{aligned}$$

$$+\left(e^{-\frac{(\lambda+1)^2}{2}||x_{sol}||^2} - e^{-\frac{(\lambda-1)^2}{2}||x_{sol}||^2}\right) x_{sol}$$

$$= \left(-2\lambda e^{-2\lambda^2||x_{sol}||^2} + (\lambda+1)e^{-\frac{(\lambda+1)^2}{2}||x_{sol}||^2} + (\lambda-1)e^{-\frac{(\lambda-1)^2}{2}||x_{sol}||^2}\right) x_{sol}$$

$$= \left(-(a+b)e^{-\frac{(a+b)^2}{2}} + ae^{-\frac{a^2}{2}} + be^{-\frac{b^2}{2}}\right) \frac{x_{sol}}{||x_{sol}||},$$

where $a = (\lambda+1)||x_{sol}||$ and $b = (\lambda-1)||x_{sol}$. From the hint, this is zero if and only if $a = 0$ (which is equivalent to $\lambda = -1$) or $b = 0$ (which is equivalent to $\lambda = 1$) or $a = -b$ (which is equivalent to $\lambda = 0$).

b) First, let x be any critical point. From the expression of the gradient, two cases are possible :

1. $-2e^{-2||x||^2} + e^{-\frac{||x+x_{sol}||^2}{2}} + e^{-\frac{||x-x_{sol}||^2}{2}} = 0$;
2. $-2e^{-2||x||^2} + e^{-\frac{||x+x_{sol}||^2}{2}} + e^{-\frac{||x-x_{sol}||^2}{2}} \neq 0$, in which case

$$x = \frac{e^{-\frac{||x+x_{sol}||^2}{2}} - e^{-\frac{||x-x_{sol}||^2}{2}}}{-2e^{-2||x||^2} + e^{-\frac{||x+x_{sol}||^2}{2}} + e^{-\frac{||x-x_{sol}||^2}{2}}} x_{sol},$$

which notably implies that x is colinear to x_{sol}.

We discuss the two cases separately. In the first case, in order for the gradient to be zero, since $x_{sol} \neq 0$, we must have

$$0 = e^{-\frac{||x+x_{sol}||^2}{2}} - e^{-\frac{||x-x_{sol}||^2}{2}}$$
$$= e^{-\frac{||x+x_{sol}||^2}{2}} \left(1 - e^{2\langle x, x_{sol}\rangle}\right).$$

Therefore, $1 - e^{2\langle x, x_{sol}\rangle} = 0$, which is equivalent to $\langle x, x_{sol}\rangle = 0$. Then,

$$0 = -2e^{-2||x||^2} + e^{-\frac{||x+x_{sol}||^2}{2}} + e^{-\frac{||x-x_{sol}||^2}{2}}$$
$$= 2\left(-e^{-2||x||^2} + e^{-\frac{||x||^2+||x_{sol}||^2}{2}}\right)$$
$$= 2e^{-\frac{1}{2}||x||^2}\left(-e^{-\frac{3}{2}||x||^2} + e^{-\frac{||x_{sol}||^2}{2}}\right).$$

This implies that $3||x||^2 = ||x_{sol}||^2$, therefore $||x|| = \frac{||x_{sol}||}{\sqrt{3}}$.

Let us now discuss the second case. Since, in this case, x is colinear to x_{sol}, we must have $x = -x_{sol}$ or $x = 0$ or $x = x_{sol}$, from the previous question.

To summarize, the only points which can be first-order critical are

$$\left\{x \in \mathbb{R}^d, \langle x, x_{sol}\rangle = 0, ||x|| = \frac{||x_{sol}||}{\sqrt{3}}\right\} \cup \{-x_{sol}, 0, x_{sol}\}. \quad (1)$$

We check that these points are actually first-order critical. For $-x_{sol}, 0, x_{sol}$, it is a consequence of the previous question. Now, for some $x \in \mathbb{R}^d$ such that $\langle x, x_{sol}\rangle = 0$ and $||x|| = \frac{||x_{sol}||}{\sqrt{3}}$,

$$\nabla f(x) = \left(-2e^{-\frac{2}{3}||x_{sol}||^2} + 2e^{-\frac{||x||^2+||x_{sol}||^2}{2}}\right) x$$

$$
\begin{aligned}
&+ \left(e^{-\frac{||x||^2+||x_{sol}||^2}{2}} - e^{-\frac{||x||^2+||x_{sol}||^2}{2}} \right) x_{sol} \\
&= \left(-2e^{-\frac{2}{3}||x_{sol}||^2} + 2e^{-\frac{||x||^2+||x_{sol}||^2}{2}} \right) x \\
&= \left(-2e^{-\frac{2}{3}||x_{sol}||^2} + 2e^{-\frac{\frac{4}{3}||x_{sol}||^2}{2}} \right) x \\
&= 0.
\end{aligned}
$$

As a consequence, the first-order critical points are exactly the points given in Equation (1).

c) We have seen at Question 2.a) that

$$\mathrm{Hess} f(x_{sol}) = \left(1 - e^{-2||x_{sol}||^2}\right) I_d + 4e^{-2||x_{sol}||^2} x_{sol} x_{sol}^T.$$

It is the sum of two semidefinite positive matrices. Hence, it is semidefinite positive, so x_{sol} is a second-order critical point. As $\mathrm{Hess} f(x_{sol}) = \mathrm{Hess} f(-x_{sol})$, $-x_{sol}$ is also a second-order critical point.
On the contrary,

$$\mathrm{Hess} f(0) = -2\left(1 - e^{-\frac{||x_{sol}||^2}{2}}\right) I_d - 2e^{-\frac{||x_{sol}||^2}{2}} x_{sol} x_{sol}^T.$$

It is the sum of two semidefinite negative (and non zero) matrices, hence 0 is not a second-order critical point.
Now, let x be such that $\langle x, x_{sol} \rangle = 0$ and $||x|| = \frac{||x_{sol}||}{\sqrt{3}}$. Then

$$\mathrm{Hess} f(x) = 6e^{-\frac{2}{3}||x_{sol}||^2} x x^T - 2e^{-\frac{2}{3}||x_{sol}||^2} x_{sol} x_{sol}^T.$$

In particular,

$$\langle \mathrm{Hess} f(x)(x_{sol}), x_{sol} \rangle = -2e^{-\frac{2}{3}||x_{sol}||^2} ||x_{sol}||^4 < 0.$$

Therefore, $\mathrm{Hess} f(x) \not\succeq 0$, so x is not a second-order critical point.
The only second-order critical points are $-x_{sol}$ and x_{sol}.