Non-convex inverse problems

March 16, 2026, 2 hours

Solution

Exercise 1

1. The map f is C^∞. Its gradient is

$$\nabla f : x \in \mathbb{R}^d \to 4x_1(x_1^2-2)\left(1+\frac{1}{4}||x||_2^2\right)e_1 + \frac{1}{2}(9+(x_1^2-2)^2)x.$$

Let us compute all vectors $x \in \mathbb{R}^d$ such that $\nabla f(x) = 0$.
For any x such that $\nabla f(x) = 0$, it holds

$$x = -2\frac{4x_1(x_1^2-2)\left(1+\frac{1}{4}||x||_2^2\right)}{9+(x_1^2-2)^2}e_1.$$

(Note that the denominator is strictly positive, and in particular cannot be zero.) Therefore, x is colinear to e_1. It suffices to find all scalar numbers $\lambda \in \mathbb{R}$ such that $\nabla f(\lambda e_1) = 0$.
Let $\lambda \in \mathbb{R}$ be fixed.

$$\begin{aligned}\nabla f(\lambda e_1) &= \left(4\lambda(\lambda^2-2)\left(1+\frac{1}{4}\lambda^2\right)+\frac{1}{2}(9+(\lambda^2-2)^2)\lambda\right)e_1\\ &= \frac{3}{2}\lambda\left(\lambda^4-1\right)e_1.\end{aligned}$$

This is zero if and only if $\lambda \in \{-1,0,1\}$. Therefore, the first-order critical points of f are $-e_1, 0$ and e_1.
Now, we compute the second-order critical points. We have to determine, for each first-order critical point, whether the Hessian of f at this point is semidefinite positive. For any $x, h \in \mathbb{R}^d$,

$$\begin{aligned}\text{Hess}\, f(x)(h,h) = 4(3x_1^2-2)&\left(1+\frac{1}{4}||x||_2^2\right)h_1^2 + 4x_1(x_1^2-2)\langle x,h\rangle h_1\\ &+\frac{1}{2}\left(9+(x_1^2-2)^2\right)||h||_2^2.\end{aligned}$$

Therefore, for any $h \in \mathbb{R}^d$,

$$\begin{aligned}\text{Hess}\, f(e_1)(h,h) &= h_1^2 + 5||h||_2^2\\ &\geq 0.\end{aligned}$$

This Hessian is positive semidefinite, so that e_1 is second-order critical. The Hessian is identical at $-e_1$, so $-e_1$ is also second-order critical.
Now, for any $h \in \mathbb{R}^d$,

$$\begin{aligned}\text{Hess}\, f(0)(h,h) &= -8h_1^2 + \frac{13}{2}||h||_2^2\\ &= -\frac{3}{2}h_1^2 + \frac{13}{2}(h_2^2+\cdots+h_d^2).\end{aligned}$$

In particular, $\text{Hess}\, f(0)(e_1,e_1) = -\frac{3}{2} < 0$, so the Hessian is not positive semidefinite.
The only second-order critical points are e_1 and $-e_1$.
The map f is coercive : for any x, $f(x) \geq 9\left(1+\frac{1}{4}||x||_2^2\right) \geq \frac{9}{4}||x||_2^2$, so that $f(x) \overset{||x||_2\to+\infty}{\longrightarrow} +\infty$. In addition, f is continuous. Therefore, it has a minimum over $\mathbb{R}^d$.

A minimizer of f is necessarily a second-order critical point, so the set of minimizers is non-empty and included in $\{-e_1, e_1\}$. We observe that

$$f(-e_1) = \frac{25}{2} = f(e_1).$$

As f has the same value at e_1 and $-e_1$, so it is not possible that one of the points is a minimizer, and the other is not. Therefore, both e_1 and $-e_1$ are global minimizers of f, and they are the only ones.

2. We apply the theorem seen in class about convergence of gradient descent to a second-order critical point starting from almost any initial point.
   We have already seen that f is coercive. In addition, the first-order critical points are isolated. Therefore, from the remark after Theorem 2.13, we can say that, for almost any $x_0 \in \mathbb{R}^d$, gradient descent with constant stepsize starting at x_0 converges towards a second-order critical point of f, hence a global minimizer (all second-order critical points are minimizers, from the previous question), provided that the stepsize is small enough.

Exercise 2

In the whole exercice, for any $k \in \{1, \ldots, d\}$, we denote $A_{:k}$ the k-th column of A.

1. Let (P) be the property that no two distinct columns of A are colinear.
   Note that, since $d \geq 2$, (P) implies that no column of A is zero.
   Let us assume that A satisfies (P) and show that uniqueness holds. Let x_1, x_2 be 1-sparse, such that $Ax_1 = Ax_2$. Let i_1, i_2 be the indices of the unique non-zero coordinate of x_1, x_2. Then

   $$(x_1)_{i_1} A_{:i_1} = Ax_1 = Ax_2 = (x_2)_{i_2} A_{:i_2}.$$

   So either $(x_1)_{i_1} = (x_2)_{i_2} = 0$, which means that $x_1 = x_2 = 0$, or the columns $A_{:i_1}$ and $A_{:i_2}$ are colinear. In this latter case, from Property (P), it must hold that $i_1 = i_2$. Also, since the columns of A are non-zero, we must have $(x_1)_{i_1} = (x_2)_{i_2}$. Therefore, x_1, x_2 are 1-sparse vectors, with the same non-zero coefficient, at the same index : they are equal. This shows uniqueness.
   Now, let us assume that (P) does not hold. Let $i_1 \neq i_2$ be two indices such that $A_{:i_1}$ and $A_{:i_2}$ are colinear. Let $\alpha_1, \alpha_2 \in \mathbb{R}$ be two numbers, not both zero, such that

   $$\alpha_1 A_{:i_1} = \alpha_2 A_{:i_2}.$$

   We define $x_1 \in \mathbb{R}^d$ as the 1-sparse vector such that

   $$(x_1)_{i_1} = \alpha_1,$$
   $$(x_1)_i = 0, \forall i \neq i_1,$$

   and $x_2 \in \mathbb{R}^d$ as

   $$(x_2)_{i_2} = \alpha_2,$$
   $$(x_2)_i = 0, \forall i \neq i_2.$$

   Then, it holds that

   $$Ax_1 = \alpha_1 A_{:i_1} = \alpha_2 A_{:i_2} = Ax_2,$$

   but $x_1 \neq x_2$. Therefore, uniqueness does not hold.
2.

for $k = 1, \dots, d$ **do**
  if *y_* is colinear to the k-th column of A* **then**
    Denote $\alpha \in \mathbb{R}$ such that $y_* = \alpha A_{:k}$.
    Return $x_* = \alpha e_k$, where e_k is the k-th vector of the canonical basis of $\mathbb{R}^d$.
  end
end

3. It is an instance of a plug-and-play algorithm, based on the forward-backward formulation described in class. Here, H is the denoiser (it is the projection onto the set of 1-sparse signals), while $x \to x - \tau A^T(Ax - y_*)$ is a gradient step (with stepsize τ) for the data fidelity term

$$x \in \mathbb{R}^d \to \frac{1}{2}||Ax - y_*||_2^2.$$

4. a) Let $i \in \{1, \dots, d\}$ be fixed.

$$\begin{aligned}
\left|\left((\mathrm{Id} - \tau A^T A)(x_t - x_*)\right)_i\right| &\leq |(x_t - x_*)_i| + \tau \left|\left(A^T A(x_t - x_*)\right)_i\right| \\
&= |(x_t - x_*)_i| + \tau |\langle A_{:i}, A(x_t - x_*)\rangle| \\
&\leq |(x_t - x_*)_i| + \tau ||A_{:i}||_2 ||A(x_t - x_*)||_2 \\
&\leq |(x_t - x_*)_i| + \tau m_2 M ||x_t - x_*||_2 \\
&\leq (1 + \tau m_2 M)||x_t - x_*||_2.
\end{aligned}$$

b) For any i,

$$\begin{aligned}
\left|\left(x_t - \tau A^T(Ax_t - y_*)\right)_i\right| &= \left|\left(x_t - \tau A^T(Ax_t - Ax_*)\right)_i\right| \\
&= \left|\left(x_* + x_t - x_* - \tau A^T A(x_t - x_*)\right)_i\right| \\
&= \left|(x_*)_i + \left((\mathrm{Id} - \tau A^T A)(x_t - x_*)\right)_i\right|.
\end{aligned}$$

Therefore, for $i = i_{x_*}$,

$$\begin{aligned}
\left|\left(x_t - \tau A^T(Ax_t - y_*)\right)_{i_{x_*}}\right| &\geq |(x_*)_{i_{x_*}}| - \left|\left((\mathrm{Id} - \tau A^T A)(x_t - x_*)\right)_{i_{x_*}}\right| \\
&\geq |(x_*)_{i_{x_*}}| - (1 + \tau m_2 M)||x_t - x_*||_2 \\
&= ||x_*||_2 - (1 + \tau m_2 M)||x_t - x_*||_2 \\
&\geq \left(1 - \frac{1 + \tau m_2 M}{4}\right)||x_*||_2 \\
&> \frac{||x_*||_2}{2}.
\end{aligned}$$

At the third line, we have used the equality $|(x_*)_{i_{x_*}}| = ||x_*||_2$, which is valid because x_* is 1-sparse. At the last line, we have used the assumption that $\tau < \frac{1}{m_2 M}$.
On the other hand, for $i \in \{1, \dots, d\} \setminus \{i_{x_*}\}$, using the fact that $(x_*)_i = 0$,

$$\begin{aligned}
\left|\left(x_t - \tau A^T(Ax_t - y_*)\right)_i\right| &= \left|(x_*)_i + \left((\mathrm{Id} - \tau A^T A)(x_t - x_*)\right)_i\right| \\
&= \left|\left((\mathrm{Id} - \tau A^T A)(x_t - x_*)\right)_i\right| \\
&\leq (1 + \tau m_2 M)||x_t - x_*||_2 \\
&\leq (1 + \tau m_2 M)\frac{||x_*||_2}{4} \\
&< \frac{||x_*||_2}{2}.
\end{aligned}$$

c) From the previous subquestion, for any $i \in \{1, \dots, d\} \setminus \{i_{x_*}\}$,

$$\left|\left(x_t - \tau A^T(Ax_t - y_*)\right)_i\right| < \frac{||x_*||_2}{2} < \left|\left(x_t - \tau A^T(Ax_t - y_*)\right)_{i_{x_*}}\right|.$$

Therefore, the single largest coordinate of $x_t - \tau A^T(Ax_t - y_*)$ (in absolute value) is i_{x_*}, meaning that $x_{t+1} = H(x_t - \tau A^T(Ax_t - y_*))$ is a 1-sparse vector, and its unique non-zero coefficient is at index i_{x_*}.

d) We assume that $x_t \in B\left(x_*, \frac{||x_*||_2}{8}\right)$. As x_* and x_{t+1} are 1-sparse, and their only non-zero coordinate is the one at index i_{x_*},

$$\begin{aligned}
||x_{t+1} - x_*||_2 &= \left|(x_{t+1} - x_*)_{i_{x_*}}\right| \\
&= \left|\left((\mathrm{Id} - \tau A^T A)(x_t - x_*)\right)_{i_{x_*}}\right| \\
&\leq (1 + \tau m_2 M)||x_t - x_*||_2 \\
&\leq (1 + \tau m_2 M)\frac{||x_*||_2}{8} \\
&< \frac{||x_*||_2}{4}.
\end{aligned}$$

5. a)

$$\begin{aligned}
\left|\left(x_t - \tau A^T(Ax_t - y_*)\right)_{i_{x_*}} - (x_*)_{i_{x_*}}\right| &= \left|-(x_*)_{i_{x_*}} + \left(x_t - \tau A^T A(x_t - x_*)\right)_{i_{x_*}}\right| \\
&= \left|(x_t - x_*)_{i_{x_*}} - \tau\left(A^T A(x_t - x_*)\right)_{i_{x_*}}\right| \\
&= \left|(x_t - x_*)_{i_{x_*}} - \tau(x_t - x_*)_{i_{x_*}}\left(A^T A_{:i_{x_*}}\right)_{i_{x_*}}\right| \\
&= \left|(x_t - x_*)_{i_{x_*}} - \tau(x_t - x_*)_{i_{x_*}}||A_{:i_{x_*}}||_2^2\right| \\
&= \left|1 - \tau||A_{:i_{x_*}}||_2^2\right| |(x_t - x_*)_{i_{x_*}}|.
\end{aligned}$$

From the assumptions, we know that

$$0 < 1 - \tau m_2 M \leq 1 - \tau||A_{:i_{x_*}}||_2^2 \leq 1 - \tau m_1^2.$$

The left-handside inequality is due to the fact that $||A_{:i_{x_*}}||_2^2 \leq ||A_{:i_{x_*}}||_2|||A||| \leq m2M$. Therefore,

$$\begin{aligned}
\left|\left(x_t - \tau A^T(Ax_t - y_*)\right)_{i_{x_*}} - (x_*)_{i_{x_*}}\right| &= \left(1 - \tau||A_{:i_{x_*}}||_2^2\right)|(x_t - x_*)_{i_{x_*}}| \\
&\geq \left(1 - \tau m_1^2\right)|(x_t - x_*)_{i_{x_*}}| \\
&= \left(1 - \tau m_1^2\right)||x_t - x_*||_2.
\end{aligned}$$

The last inequality is due to the fact that x_t and x_* are 1-sparse, with i_{x_*} as common non-zero coordinate.

b) From Question 4., we know that x_{t+1} is 1-sparse, and its non-zero coordinate is at index i_{x_*}. Therefore,

$$\begin{aligned}
||x_{t+1} - x_*||_2 &= \left|(x_{t+1} - x_*)_{i_{x_*}}\right| \\
&= \left|\left(x_t - A^T(Ax_t - y_*)\right)_{i_{x_*}} - (x_*)_{i_{x_*}}\right| \\
&\leq (1 - \tau m_1^2)||x_t - x_*||_2.
\end{aligned}$$

6. a) Let us assume that $x_0 \in B\left(x_*, \frac{||x_*||}{8}\right)$. Then, from 4.c) and d), we know that x_1 is 1-sparse, with the same non-zero coordinate as x_*, and $x_1 \in B\left(x_*, \frac{||x_*||_2}{4}\right)$.
Now, we prove by iteration over t that, for any $t \in \mathbb{N}^*$, x_t is 1-sparse, with the same non-zero coordinate as x_*, and

$$||x_t - x_*||_2 \leq (1 - \tau m_1^2)^{t-1} ||x_1 - x_*||_2. \tag{1}$$

For $t = 1$, we have already shown that x_t is 1-sparse, with $i_{x_1} = i_{x_*}$ and the inequality is an equality.
Now, if the property holds true for some $t \in \mathbb{N}^*$, it holds that

$$||x_t - x_*||_2 \leq (1 - \tau m_1^2)^{t-1} ||x_1 - x_*||_2 \leq ||x_1 - x_*||_2 \leq \frac{||x_*||_2}{4}.$$

Therefore, from 4.c), we have that x_{t+1} is 1-sparse, with the same non-zero coordinate as x_*. From Question 5.b) and using Equation (1),

$$||x_{t+1} - x_*||_2 \leq (1 - \tau m_1^2)||x_t - x_*||_2 \leq (1 - \tau m_1^2)^t ||x_1 - x_*||_2.$$

This proves Equation (1) for $t + 1$, hence establishes it for any $t \geq 1$.
From Equation (1), $||x_t - x_*||_2 \overset{t \to +\infty}{\longrightarrow} 0$, hence $(x_t)_{t \in \mathbb{N}}$ goes to x_*.

b) It is a local convergence result for the considered algorithm.

Exercise 3

1. The relaxation is tight, in the sense used in class, if (Min TV) has a unique solution, and this solution is x_*.
2. a) Let $\bar{B}_{TV}$ be this set. The conjecture is that extremal points of $\bar{B}_{TV}$ are exactly the elements in $\mathcal{E}_\perp$ with one jump and jump gap equal to ± 1. More formally, the extremal points are exactly

$$\{x_{jump,k}\}_{k=1,\dots,d-1} \cup \{-x_{jump,k}\}_{k=1,\dots,d-1},$$

where, for any k, $x_{jump,k}$ is the vector such that

$$\begin{aligned} (x_{jump,k})_s &= \tfrac{k}{d} - 1 \text{ for all } s = 1, \dots, k, \\ &= \tfrac{k}{d} \text{ forall } s = k+1, \dots, d. \end{aligned}$$

(Observe that these vectors belong to $\mathcal{E}_\perp$ and are the only elements of $\mathcal{E}_\perp$ with exactly one jump and TV norm equal to 1.)
If the conjecture holds true, then the vector x_* can be written as a convex combination of a few extremal points of its TV ball. Therefore, it intuitively belongs to a kind of corner, or sharp edge, of this set. This particular geometry makes it possible for $\{x_*\} + \operatorname{Ker} A$ to go through x_* without intersecting the rest of the TV ball $\{x \in \mathcal{E}_\perp, ||x||_{TV} \leq ||x_*||_{TV}\}$. This is the same intuition we described in class to justify the choice of the ℓ^1-norm as a convex relaxation of the ℓ^0-norm.

b) Let us define

$$\begin{aligned} D \ : \ \mathcal{E}_\perp &\to \mathbb{R}^{d-1} \\ x &\to (x_{i+1} - x_i)_{i=1,\dots,d-1}. \end{aligned}$$

This map is injective : for any vector $x \in \mathcal{E}_\perp$, if $D(x) = 0$, then $x_1 = x_2 = \dots = x_d$, and, as $0 = \langle x, \mathbb{1}_d \rangle = dx_1$, we must have $x_1 = 0$ and $x = 0$. Since it is a linear map between $(d-1)$-dimensional vector spaces, it is a bijection.
For any $x \in \mathcal{E}_\perp$, it holds

$$||x||_{TV} = ||D(x)||_1.$$

Therefore, a vector $x \in \mathcal{E}_\perp$ belongs to $\bar{B}_{TV}$ if and only if $D(x) \in \bar{B}_{\ell^1}$, where $\bar{B}_{\ell^1}$ is the unit ball of $\mathbb{R}^{d-1}$ for the ℓ^1-norm. This shows that

$$\bar{B}_{\ell^1} = D(\bar{B}_{TV}).$$

A bijective linear map sends the extremal points of a convex set to the extremal points of the image of the set through the map. Therefore, a vector $x \in \mathcal{E}_\perp$ is an extremal point of $\bar{B}_{TV}$ if and only if $D(x)$ is an extremal point of $\bar{B}_{\ell^1}$, i.e. there exists $k \in \{1, \dots, d-1\}$ such that

$$D(x) = e_k \text{ or } D(x) = -e_k.$$

Here, e_k is the k-th vector of the canonical basis of $\mathbb{R}^{d-1}$.
For any vector $x \in \mathcal{E}_\perp$ and $k \in \{1, \dots, d-1\}$,

$$D(x) = e_k \iff x = x_{jump,k}$$

(because $D(x_{jump,k}) = e_k$ and D is injective). Therefore, a vector $x \in \mathcal{E}_\perp$ is an extremal point of $\bar{B}_{TV}$ if and only if there exists $k \in \{1, \dots, d-1\}$ such that

$$x = x_{jump,k} \text{ or } x = -x_{jump,k}.$$

3. We write the primal problem in min-max form :

$$\min_{x\in\mathcal{E}_\perp} \max_{z\in\mathbb{R}^m} ||x||_{TV} + \langle z, y_* - Ax\rangle$$
$$= \min_{x\in\mathcal{E}_\perp} \max_{z\in\mathbb{R}^m} \langle z, y_*\rangle + ||x||_{TV} - \left\langle x, A^T z\right\rangle .$$

We get the dual problem by exchanging the minimum and maximum :

$$\max_{z\in\mathbb{R}^m} \min_{x\in\mathcal{E}_\perp} \langle z, y_*\rangle + ||x||_{TV} - \left\langle x, A^T z\right\rangle$$
$$\max_{z\in\mathbb{R}^m} \langle z, y_*\rangle + \min_{x\in\mathcal{E}_\perp} \left(||x||_{TV} - \left\langle x, A^T z\right\rangle\right) .$$

Let us consider any $z \in \mathbb{R}^m$ and compute $\min_{x\in\mathcal{E}_\perp} \left(||x||_{TV} - \left\langle x, A^T z\right\rangle\right)$.
For any $k \in \{1, \dots, d-1\}$ and $\lambda \in \mathbb{R}$, we have

$$\begin{aligned}
||\lambda x_{jump,k}||_{TV} - \left\langle \lambda x_{jump,k}, A^T z\right\rangle &= |\lambda| - \lambda\left(\frac{k}{d} - 1\right)\sum_{i=1}^{k}(A^T z)_i - \lambda\frac{k}{d}\sum_{i=k+1}^{d}(A^T z)_i \\
&= |\lambda| + \lambda\sum_{i=1}^{k}(A^T z)_i - \lambda\frac{k}{d}\sum_{i=1}^{d}(A^T z)_i \\
&= |\lambda| + \lambda\sum_{i=1}^{k}(A^T z)_i - \lambda\frac{k}{d}\left\langle \mathbb{1}_d, A^T z\right\rangle \\
&= |\lambda| + \lambda\sum_{i=1}^{k}(A^T z)_i - \lambda\frac{k}{d}\left\langle A\mathbb{1}_d, z\right\rangle \\
&= |\lambda| + \lambda\sum_{i=1}^{k}(A^T z)_i.
\end{aligned}$$

Therefore, if $\sum_{i=1}^k (A^T z)_i < -1$, then

$$||\lambda x_{jump,k}||_{TV} - \left\langle \lambda x_{jump,k}, A^T z\right\rangle \overset{\lambda\to+\infty}{\longrightarrow} -\infty.$$

On the other hand, if $\sum_{i=1}^{k}(A^T z)_i > 1$, then

$$||\lambda x_{jump,k}||_{TV} - \langle \lambda x_{jump,k}, A^T z \rangle \overset{\lambda \to -\infty}{\longrightarrow} -\infty.$$

This shows that, for any $k \in \{1, \dots, d-1\}$, if $\left|\sum_{i=1}^{k}(A^T z)_i\right| > 1$, then

$$\min_{x \in \mathcal{E}_\perp} \left(||x||_{TV} - \langle x, A^T z \rangle \right) = -\infty.$$

Now, let us consider the situation where, for any k, $\left|\sum_{i=1}^{k}(A^T z)_i\right| \leq 1$. Then, for any $x \in \mathcal{E}_\perp$,

$$\begin{aligned}
&||x||_{TV} - \langle x, A^T z \rangle \\
&= \sum_{i=1}^{d-1} |x_{i+1} - x_i| - \sum_{s=1}^{d} x_s (A^T z)_s \\
&= \sum_{i=1}^{d-1} |x_{i+1} - x_i| - \sum_{s=1}^{d} \left(x_d + \sum_{i=s}^{d-1} (x_i - x_{i+1}) \right) (A^T z)_s \\
&= \sum_{i=1}^{d-1} |x_{i+1} - x_i| - \sum_{s=1}^{d} \sum_{i=s}^{d-1} (x_i - x_{i+1})(A^T z)_s - x_d \sum_{s=1}^{d} (A^T z)_s \\
&= \sum_{i=1}^{d-1} |x_{i+1} - x_i| - \sum_{i=1}^{d-1} (x_i - x_{i+1}) \left(\sum_{s=1}^{i} (A^T z)_s \right) - x_d \sum_{s=1}^{d} (A^T z)_s \\
&= \sum_{i=1}^{d-1} |x_{i+1} - x_i| - \sum_{i=1}^{d-1} (x_i - x_{i+1}) \left(\sum_{s=1}^{i} (A^T z)_s \right) \\
&\geq \sum_{i=1}^{d-1} |x_{i+1} - x_i| - \sum_{i=1}^{d-1} |x_i - x_{i+1}| \left| \sum_{s=1}^{i} (A^T z)_s \right| \\
&\geq \sum_{i=1}^{d-1} |x_{i+1} - x_i| - \sum_{i=1}^{d-1} |x_i - x_{i+1}| \\
&= 0.
\end{aligned}$$

Therefore,

$$\min_{x \in \mathcal{E}_\perp} \left(||x||_{TV} - \langle x, A^T z \rangle \right) \geq 0.$$

Actually, since the value 0 is attained at $x = 0$, we have

$$\min_{x \in \mathcal{E}_\perp} \left(||x||_{TV} - \langle x, A^T z \rangle \right) = 0.$$

We plug this into the max-min expression above and get as dual problem

$$\max_{z \in \mathbb{R}^m} \langle z, y_* \rangle + 1_{\forall k=1,\dots,d-1, \left|\sum_{i=1}^{k}(A^T z)_i\right| \leq 1},$$

which is exactly what we had to prove.

4. For any pair $(x, z) \in \mathcal{E}_\perp \times \mathbb{R}^m$, we define

$$F(x, z) = \langle z, y_* \rangle + ||x||_{TV} - \langle x, A^T z \rangle$$

We have seen that the objective of the primal problem can be written as

$$x \in \mathcal{E}_\perp \to \max_{z \in \mathbb{R}^m} F(x, z),$$

while the objective of the dual problem is

$$z \in \mathbb{R}^m \to \min_{x \in \mathcal{E}_\perp} F(x, z).$$

Let $(x_o, z_o) \in \mathcal{E}_\perp \times \mathbb{R}^m$ be fixed. Because of weak duality, it is primal-dual optimal if and only if the value of the primal at x_o is the same as the value of the dual at z_o, i.e.

$$\max_{z \in \mathbb{R}^m} F(x_o, z) = \min_{x \in \mathcal{E}_\perp} F(x, z_o).$$

Since $\min_{x \in \mathcal{E}_\perp} F(x, z_o) \leq F(x_o, z_o) \leq \max_{z \in \mathbb{R}^m} F(x_o, z)$, this is also equivalent to

$$\max_{z \in \mathbb{R}^m} F(x_o, z) = F(x_o, z_o)$$
$$\text{and } \min_{x \in \mathcal{E}_\perp} F(x, z_o) = F(x_o, z_o).$$

The first equality holds true if and only if $Ax_o = y_*$, i.e. x_o is primal-feasible. As to the second inequality, it is equivalent to

$$||x_o||_{TV} - \left\langle x_o, A^T z_o \right\rangle = \min_{x \in \mathcal{E}_\perp} ||x||_{TV} - \left\langle x, A^T z \right\rangle. \tag{2}$$

Let us find a simple property which is equivalent to Equation (2). First, if the equation is verified, we must have

$$\min_{x \in \mathcal{E}_\perp} ||x||_{TV} - \left\langle x, A^T z_o \right\rangle \neq -\infty,$$

so, from the reasoning in the previous question, it must hold

$$\left| \sum_{i=1}^{k} (A^T z_o)_i \right| \leq 1, \forall k = 1, \ldots, d-1 \tag{3}$$

(which is equivalent to dual-feasibility of z_o) and

$$||x_o||_{TV} - \left\langle x_o, A^T z_o \right\rangle = \min_{x \in \mathcal{E}_\perp} ||x||_{TV} - \left\langle x, A^T z_o \right\rangle = 0.$$

Now, from the chain of equalities and inequalities in the answer to the previous question, we have that, in order for $||x_o||_{TV} - \left\langle x_o, A^T z_o \right\rangle$ to be 0, we must have, for any $i = 1, \ldots, d-1$,

$$\begin{aligned} &(x_o)_i - (x_o)_{i+1} = 0 \\ \text{or } &\left((x_o)_i - (x_o)_{i+1} > 0 \text{ and } \sum_{s=1}^{i} (A^T z_o)_s = 1 \right) \\ \text{or } &\left((x_o)_i - (x_o)_{i+1} < 0 \text{ and } \sum_{s=1}^{i} (A^T z_o)_s = -1 \right). \end{aligned} \tag{4}$$

We have proved that Equation (2) implies Equations (3) and (4).
Conversely, if Equations (3) and (4) hold, we have (from Equation (3))

$$\min_{x \in \mathcal{E}_\perp} ||x||_{TV} - \left\langle x, A^T z_o \right\rangle = 0$$

and, from Equation (4) and the chain of inequalities (which become equalities) in the answer to the previous question,

$$||x_o||_{TV} - \left\langle x_o, A^T z_o \right\rangle = 0.$$

Therefore, Equation (2) is also true.
In summary, (x_o, z_o) is primal-dual optimal if and only if x_o is primal-feasible, z_o is dual-feasible and Equation (4) is verified for any $i = 1, \ldots, d-1$.

Exercise 4

Basic solution : we can exploit the absolute value of the linear measurements only, namely the fact that

$$|z_k| = |\langle x_{sol}, u_k\rangle|, \forall k = 1, \dots, m',$$

and treat these new measurements as the y_k. This leads to defining

$$\hat{M}_{new} = \frac{1}{m + d/2}\left(\sum_{k=1}^{m} y_k^2 v_k v_k^* + \sum_{k=1}^{d/2} |z_k|^2 u_k u_k^*\right),$$

and x_0 as the main eigenvector of this matrix.

More elaborate solution : let us denote $\mathcal{U} = \text{Vect}\{u_1, \dots, u_{m'}\}$. For any vector $x \in \mathbb{R}^d$, we write

$$x = x_{\mathcal{U}} + x_{\perp},$$

where $x_{\mathcal{U}} \in \mathcal{U}$ and $x_{\perp} \in \mathcal{U}^{\perp}$.

From $z_1, \dots, z_{m'}$, we can compute $x_{sol\,\mathcal{U}}$ by solving the linear system

$$z_k = \langle x_{sol\,\mathcal{U}}, u_k\rangle, \forall k = 1, \dots, m',$$
$$\text{and } x_{sol\,\mathcal{U}} \in \mathcal{U}.$$

What we have to estimate is $x_{sol\,\perp}$. Let $x_{0\,sp}$ be the result of the classical spectral estimation procedure. If the procedure is successful, there exists $\epsilon \in \{-1; 1\}$ such that

$$x_{0\,sp} \approx \epsilon x_{sol} = \epsilon x_{sol\,\mathcal{U}} + \epsilon x_{sol\,\perp}.$$

First, we try to determine ϵ. We compute

$$\langle x_{0\,sp}, x_{sol\,\mathcal{U}}\rangle \approx \epsilon ||x_{sol\,\mathcal{U}}||_2^2$$

and set $\epsilon_0 = \text{sign}(\langle x_{0\,sp}, x_{sol\,\mathcal{U}}\rangle)$. We can expect that, with high probability, $\epsilon_0 = \epsilon$. Under this event,

$$\epsilon_0 x_{0\,sp} \approx x_{sol\,\mathcal{U}} + x_{sol\,\perp}$$
$$\Rightarrow (I_d - P_{\mathcal{U}})(\epsilon_0 x_{0\,sp}) \approx x_{sol\,\perp},$$

where $P_{\mathcal{U}}$ is the orthogonal projector onto $\mathcal{U}$. This leads to setting

$$x_0 = x_{sol\,\mathcal{U}} + (I_d - P_{\mathcal{U}})(\epsilon_0 x_{0\,sp}).$$

Different solution : we keep the notation from the previous solution. For any k, we have

$$\begin{aligned} y_k^2 &= |\langle x_{sol\,\mathcal{U}}, v_{k\,\mathcal{U}}\rangle + \langle x_{sol\,\perp}, v_{k\,\perp}\rangle|^2 \\ &= |\langle x_{sol\,\mathcal{U}}, v_{k\,\mathcal{U}}\rangle|^2 + 2\langle x_{sol\,\mathcal{U}}, v_{k\,\mathcal{U}}\rangle\langle x_{sol\,\perp}, v_{k\,\perp}\rangle + |\langle x_{sol\,\perp}, v_{k\,\perp}\rangle|^2. \end{aligned}$$

From there, we can imagine, for instance, to look at

$$\frac{1}{m}\sum_{k=1}^{m} y_k^2 v_{k\,\perp}.$$

As $\mathcal{U}$ is a random m'-dimensional subspace of $\mathbb{R}^d$, with uniform distribution, and independent from $v_1, \dots, v_m$, we can check that, conditionally on $\mathcal{U}$, $v_{1\,\mathcal{U}}, \dots, v_{m\,\mathcal{U}}, v_{1\,\perp}, \dots, v_{m\,\perp}$ are independent Gaussian vectors, with laws

$$\mathcal{N}(0, P_{\mathcal{U}}) \text{ for } v_{1\,\mathcal{U}}, \dots, v_{m\,\mathcal{U}},$$

$$\mathcal{N}(0, I_d - P_{\mathcal{U}}) \text{ for } v_{1\,\perp}, \dots, v_{m\,\perp}.$$

This yields that, for any k, conditionally on $\mathcal{U}$,

$$\begin{aligned}
\mathbb{E}(y_k^2 v_{k\,\perp}) &= \mathbb{E}\left(|\langle x_{sol\,\mathcal{U}}, v_{k\,\mathcal{U}}\rangle|^2 v_{k\,\perp}\right) \\
&\qquad + 2\mathbb{E}\left(\langle x_{sol\,\mathcal{U}}, v_{k\,\mathcal{U}}\rangle \langle x_{sol\,\perp}, v_{k\,\perp}\rangle v_{k\,\perp}\right) \\
&\qquad + \mathbb{E}\left(|\langle x_{sol\,\perp}, v_{k\,\perp}\rangle|^2 v_{k\,\perp}\right) \\
&= 2\mathbb{E}\left(\langle x_{sol\,\mathcal{U}}, v_{k\,\mathcal{U}}\rangle \langle x_{sol\,\perp}, v_{k\,\perp}\rangle v_{k\,\perp}\right) \\
&= 2\langle x_{sol\,\mathcal{U}}, v_{k\,\mathcal{U}}\rangle x_{sol\,\perp}.
\end{aligned}$$

Therefore, we can, for instance, define

$$x_{0\,\perp} = \frac{\sum_{k=1}^m \operatorname{sign}(\langle x_{sol\,\mathcal{U}}, v_{k\,\mathcal{U}}\rangle) y_k^2 v_{k\,\perp}}{2\sum_{k=1}^m |\langle x_{sol\mathcal{U}}, v_{k\,\mathcal{U}}\rangle|}.$$

In expectation, this sum is equal to $x_{sol\,\perp}$. We may therefore expect that, for m large enough, it approximates $x_{sol\,\perp}$ with high probability, and set

$$x_0 = x_{sol\,\mathcal{U}} + x_{0\,\perp}.$$