Non-convex inverse problems

February 21, 2025

Correction

Exercise 1

1. Problem 1. is a direct problem, not an inverse one : the main object is the house, whose description is known, and we want to simulate an observation of it.
   Problem 2. is an inverse problem. The unknown quantity of interest is the time at which the system is turned on. The observation is a temperature measurement at 7pm.
   Problem 3. is an inverse problem. The unknown quantity of interest is the position of the object. The observations are the distances of the object to the sensors.
2. Yes, it is unique. Indeed, let $(x, y), (x', y') \in \mathbb{R}^2$ be such that $M(x, y) = M(x', y')$.
   — First case : $x^2 = x'^2 = 0$. Then $x = x' = 0$ and $y = x + y = x' + y' = y'$.
   — Second case : $x^2 = x'^2 \neq 0$. Then $y = \frac{x^2 y}{x^2} = \frac{x'^2 y'}{x'^2} = y'$ and $x = (x+y)-y = (x' + y') - y = (x' + y') - y' = x'$.

   In any case, $(x, y) = (x', y')$.

Exercise 2

1. First, let $X \in \mathbb{R}^{d_1 \times d_2}$ be a unit-normed matrix with exactly one non-zero row. We show that it is an extremal point of B. Let i_0 be the index of the non-zero row. It holds

$$\begin{aligned} ||X||_{1,2} &= ||(X_{i_0,1}, \dots, X_{i_0,d_2})||_2 \\ &= ||X||_F \text{ since } X_{i,j} = 0 \text{ for all } i \neq i_0 \\ &= 1, \end{aligned}$$

   so $X \in B$. We now show that it is extremal. Let $t \in [0; 1], X_1, X_2 \in B$ be such that

$$X = (1 - t)X_1 + tX_2.$$

   We show that either $X = X_1$ or $X = X_2$. If $t = 0$ or $t = 1$, this is true, so we can assume $0 < t < 1$. We have

$$\begin{aligned} 1 &= ||X_{i_0,:}||_2^2 \\ &= ||(1 - t)(X_1)_{i_0,:} + t(X_2)_{i_0,:}||_2^2 \\ &= (1 - t)^2||(X_1)_{i_0,:}||_2^2 + 2t(1 - t)\left\langle (X_1)_{i_0,:}, (X_2)_{i_0,:} \right\rangle + t^2||(X_2)_{i_0,:}||_2^2 \\ &\leq (1 - t)^2||(X_1)_{i_0,:}||_2^2 + 2t(1 - t)||(X_1)_{i_0,:}||_2||(X_2)_{i_0,:}||_2 + t^2||(X_2)_{i_0,:}||_2^2 \\ &= ((1 - t)||(X_1)_{i_0,:}||_2 + t||(X_2)_{i_0,:}||_2)^2 \\ &\leq 1. \end{aligned}$$

For the last line, we have used that $||(X_1)_{i_0,:}||_2 \leq ||X_1||_{1,2} \leq 1$ (as $X_1 \in B$), and same for X_2. Since the left and right-hand side of the inequality are equal, all inequalities must be equalities :

$$\begin{aligned}\langle (X_1)_{i_0,:}, (X_2)_{i_0,:} \rangle &= ||(X_1)_{i_0,:}||_2 ||(X_2)_{i_0,:}||_2, \\ ||(X_1)_{i_0,:}||_2 &= ||X_1||_{1,2} = 1, \\ ||(X_2)_{i_0,:}||_2 &= ||X_2||_{1,2} = 1.\end{aligned}$$

The first equality implies that $(X_1)_{i_0,:}$ and $(X_2)_{i_0,:}$ are colinear, with a nonnegative proportionality coefficient. From the second and third equalities, these vectors have the same norm, 1, so they are equal :

$$(X_1)_{i_0,:} = (X_2)_{i_0,:}.$$

The equality $||(X_1)_{i_0,:}||_2 = ||X_1||_{1,2}$ implies that $||(X_1)_{i,:}||_2 = 0$ for all $i \neq i_0$, so $(X_1)_{i_0,:}$ is the only non-zero row of X_1. The same holds for X_2, so that $X_1 = X_2$, and

$$X = (1-t)X_1 + tX_2 = X_1 = X_2.$$

This concludes the proof that X is extremal.
Then, let $X \in B$ be an extremal point. Let us show that it has unit Frobenius norm, and exactly one non-zero row. First, we show that X has exactly one non-zero row. We proceed by contradiction, and fix $i_0, i_1 \leq d_1$ the indices of two non-zero rows. We define $X_1, X_2 \in \mathbb{R}^{d_1 \times d_2}$ such that, for all i, j,

$$\begin{aligned}(X_1)_{i,j} &= 0 \text{ if } i \neq i_0 \\ &= \frac{X_{i,j}}{||X_{i_0,:}||_2} \text{ if } i = i_0. \\ (X_2)_{i,j} &= \frac{X_{i,j}}{1 - ||X_{i_0,:}||_2} \text{ if } i \neq i_0, \\ &= 0 \text{ if } i = i_0.\end{aligned}$$

Both matrices are different from X : the i_1-th of X_1 and the i_0-th row of X_2 are zero, while the corresponding rows of X are not. Both matrices belong to B :

$$\begin{aligned}||X_1||_{1,2} &= ||(X_1)_{i_0,:}||_2 = \frac{||X_{i_0,:}||_2}{||X_{i_0,:}||_2} = 1, \\ ||X_2||_{1,2} &= \frac{\sum_{i \neq i_0} ||X_{i,:}||_2}{1 - ||X_{i_0,:}||_2} \\ &= \frac{||X||_{1,2} - ||X_{i_0,:}||_2}{1 - ||X_{i_0,:}||_2} \\ &\leq \frac{1 - ||X_{i_0,:}||_2}{1 - ||X_{i_0,:}||_2} \\ &= 1.\end{aligned}$$

In addition,

$$X = ||X_{i_0,:}||_2 X_1 + (1 - ||X_{i_0,:}||_2) X_2,$$

so X is not extremal. We have reached a contradiction. This shows that X cannot have more than one non-zero row.
It remains to show that X has unit Frobenius norm. Since X has no more than one non-zero row, $||X||_{1,2} = ||X||_F$, so that we only have to show that $||X||_{1,2} = 1$. We observe that

$$X = ||X||_{1,2} \frac{X}{||X||_{1,2}} + (1 - ||X||_{1,2}) \times 0.$$

The matrices $\frac{X}{||X||_{1,2}}$ and 0 both belong to B. Since X is extremal and $X \neq 0$ (because 0 is not extremal), we must have $X = \frac{X}{||X||_{1,2}}$, which implies that $||X||_{1,2} = 1$.

2. In the case of compressed sensing, we have approximated the ℓ^0-norm with the ℓ^1-norm based on the argument that the unit-normed vectors with minimal ℓ^0-norm were the extremal points of the unit ℓ^1-ball. Here, in the previous question, we have shown that the unit-normed vectors with minimal ℓ^0-row norm were the extremal points of the unit mixed ℓ^1/ℓ^2-norm ball. This suggests that $||.||_{1,2}$ is a reasonable convex approximation for $||.||_{0,row}$, which leads to the following minimization problem :

$$\begin{aligned} &\text{minimize } ||X||_{1,2} \\ &\quad \text{over all } X \in \mathbb{R}^{d_1 \times d_2} \\ &\text{such that } \mathcal{L}(X) = b. \end{aligned}$$

3. In the case of compressed sensing, we say that $\mathcal{L}$ satisfies a (k, δ)-restricted isometry property if, for any vector x such that $||x||_0 \leq k$,

$$(1 - \delta)||x||_2 \leq ||\mathcal{L}(x)||_2 \leq (1 + \delta)||x||_2.$$

In the context of the exercise, we can modify the definition as follows : $\mathcal{L}$ satisfies a (k, δ)-restricted isometry property if, for any matrix $X \in \mathbb{R}^{d_1 \times d_2}$ with at most k non-zero rows,

$$(1 - \delta)||X||_F \leq ||\mathcal{L}(X)||_2 \leq (1 + \delta)||X||_F.$$

Exercise 3

1. The ℓ^0 norm is non-convex. Indeed, if x is any vector with exactly one non-zero coordinate,

$$\left|\left|\frac{0}{2} + \frac{x}{2}\right|\right|_0 = 1 > \frac{1}{2} = \frac{||0||_0}{2} + \frac{||x||_0}{2}.$$

2.

$$\begin{aligned} &\text{minimize } ||x||_1, \\ &\quad \text{over all } x \in \mathbb{R}^d, \\ &\text{such that } Ax = y. \end{aligned} \tag{ConvRel}$$

3. First, we assume that $|z_i| \leq 1$ for all $i \leq d$. Then, for any $x \in \mathbb{R}^d$,

$$\begin{aligned} ||x||_1 - \langle x, z \rangle &= \sum_{i=1}^{d} |x_i| - x_i z_i \\ &\geq \sum_{i=1}^{d} |x_i| - |x_i||z_i| \\ &\geq \sum_{i=1}^{d} |x_i| - |x_i| \\ &= 0. \end{aligned} \tag{1}$$

The minimum is therefore nonnegative. In addition, $||0||_1 - \langle 0, z \rangle = 0$, so the minimum is exactly 0.
The minimizers are the vectors x for which all inequalities in the equation block (1) are equalities. This is equivalent to

$$\begin{aligned} &\forall i \leq d, x_i z_i = |x_i||z_i| = |x_i| \\ \iff &\forall i \leq d, (x_i = 0) \text{ or } (\operatorname{sgn}(x_i) z_i = |z_i| = 1) \\ \iff &\forall i \leq d, (x_i = 0) \text{ or } (z_i = \operatorname{sgn}(x_i)) \\ \iff &\forall i \leq d \text{ s.t. } x_i \neq 0, z_i = \operatorname{sgn}(x_i). \end{aligned}$$

Second, we consider the case where there exists at least one index i such that $|z_i| > 1$. Let such an index i be fixed. Let $e_i \in \mathbb{R}^d$ denote the i-th element of the canonical basis. Then, for all $t \geq 0$,

$$\begin{aligned} \min_{x \in \mathbb{R}^d} ||x||_1 - \langle x, z \rangle &\leq ||\operatorname{sgn}(z_i) t e_i||_1 - \langle \operatorname{sgn}(z_i) t e_i, z \rangle \\ &= t(1 - |z_i|). \end{aligned}$$

Since $t(1 - |z_i|) \to -\infty$ when $t \to +\infty$, the minimum can only be $-\infty$.

4. We write (ConvRel) in min-max form :

$$\begin{aligned} \min_{x \in \mathbb{R}^d \text{ s.t. } Ax = y} ||x||_1 &= \min_{x \in \mathbb{R}^d} ||x||_1 + 1_{y = Ax} \\ &= \min_{x \in \mathbb{R}^d} \max_{b \in \mathbb{R}^m} ||x||_1 + \langle y - Ax, b \rangle \\ &= \min_{x \in \mathbb{R}^d} \max_{b \in \mathbb{R}^m} ||x||_1 - \left\langle x, A^T b \right\rangle + \langle y, b \rangle \overset{def}{=} F(x, b). \end{aligned}$$

We get the dual by switching the minimum and maximum :

$$\begin{aligned} \max_{b \in \mathbb{R}^m} \min_{x \in \mathbb{R}^d} F(x, b) &= \max_{b \in \mathbb{R}^m} \min_{x \in \mathbb{R}^d} ||x||_1 - \left\langle x, A^T b \right\rangle + \langle y, b \rangle \\ &= \max_{b \in \mathbb{R}^m} -1_{||A^T b||_\infty \leq 1} + \langle y, b \rangle \\ &= \max_{b \in \mathbb{R}^m \text{ s.t. } ||A^T b||_\infty \leq 1} \langle y, b \rangle , \end{aligned}$$

which is what was expected.

5. A pair (x, b) is primal-dual optimal if x is optimal for the primal problem, b is optimal for the dual problem, and the optimal primal and dual values are the same. Since weak duality holds, this is equivalent to

$$\text{Dual at } b = \text{Primal at } x.$$

Since

$$\begin{aligned}\text{Dual at } b &= \min_{z\in\mathbb{R}^d} F(z, b)\\ &\leq F(x, b)\\ &\leq \max_{c\in\mathbb{R}^m} F(x, c)\\ &= \text{Primal at } x,\end{aligned}$$

this holds true if and only if x is a minimizer of

$$\min_{z\in\mathbb{R}^d} F(z, b)$$

and b is a maximizer of

$$\max_{c\in\mathbb{R}^m} F(x, c).$$

The first of these two conditions is equivalent, from Question 3., to

$$||A^T b||_\infty \leq 1 \text{ and } (A^T b)_i = \text{sgn}(x_i), \forall i \leq d \text{ s.t. } x_i \neq 0,$$

which is exactly equivalent to Properties 2. and 3.. The second condition is equivalent to b being a maximizer of

$$\max_{c\in\mathbb{R}^d} \langle y - Ax, c\rangle .$$

This problem has a maximizer if and only if $y - Ax = 0$ and, in this case, any vector is a maximizer. Therefore, the second condition is equivalent to $y - Ax = 0$, which is Property 1..

6. a) Let us define $b = \text{sgn}((x_{sol})_s)A_{:s}$, for s the index of the only non-zero coordinate of x_{sol}, and show that (x_{sol}, b) is primal-dual optimal. It suffices to check the three properties from Question 5..
The first property, $Ax_{sol} = y$, is equivalent to x_{sol} being feasible for (ConvRel), which is true because x_{sol} is feasible for (CS).
For the second property, observe that, for any $i \leq d$, $|(A^T b)_i| = |\langle A_{:i}, b\rangle| = |\langle A_{:i}, A_{:s}\rangle| \leq ||A_{:i}||_2 ||A_{:s}||_2 = 1$.
For the third property, observe that, for any $i \leq d$, $(x_{sol})_i \neq 0$ if and only if $i = s$, in which case

$$(A^T b)_i = \text{sgn}((x_{sol})_s)||A_{:s}||_2^2 = \text{sgn}((x_{sol})_i).$$

We have thus shown that (x_{sol}, b) is primal-dual optimal. In particular, x_{sol} is a minimizer of (ConvRel).

b) Let x be any minimizer of (ConvRel). Let us show that $x = x_{sol}$.
Let b be defined as in the previous question. Since b is dual optimal and strong duality holds (because we have seen that (x_{sol}, b) is primal-dual optimal), and since x is primal optimal, (x, b) is a primal-dual optimal pair. In particular, it satisfies the properties from Question 5., in particular

$$\text{sgn}(x_i) = (A^T b)_i = \text{sgn}((x_{sol})_s) \langle A_{:i}, A_{:s} \rangle, \forall i \leq d \text{ s.t. } x_i \neq 0.$$

For any $i \neq s$, $|\langle A_{:i}, A_{:s} \rangle| < ||A_{:,i}||_2 ||A_{:s}||_2 = 1$, because no distinct columns of A are colinear ; therefore, $\text{sgn}(x_i) \neq \text{sgn}((x_{sol})_s) \langle A_{:i}, A_{:s} \rangle$. This shows that $x_i = 0$ for all $i \neq s$.
Since x and x_{sol} are both 1-sparse, with the same non-zero coordinate, they are colinear. Let $\lambda \in \mathbb{R}$ be such that $x = \lambda x_{sol}$. We use the fact that x is feasible for (ConvRel) :

$$y = Ax = \lambda A x_{sol} = \lambda y.$$

As $y = (x_{sol})_s A_{:s} \neq 0$, we must have $\lambda = 1$, meaning that $x = x_{sol}$.

c) The convex relaxation (ConvRel) of Problem (CS) is tight.

d) In this case, since the convex problem (ConvRel) has the same solution as (CS), it suffices to solve this convex problem (using any linear programming solver) ; the solution obtained solves (CS).

Exercise 4

1. We observe that $f(a, b) \geq 0$ for any $(a, b) \in \mathbb{R}^d \times \mathbb{R}^d$. In addition, if we denote $\mathbb{1}$ the all-one vector, it holds $f(y, \mathbb{1}) = 0$. Therefore, the minimum of f is 0, and the minimizers are

$$\begin{aligned} &\{(a, b) \in \mathbb{R}^d \times \mathbb{R}^d, f(a, b) = 0\} \\ =&\{(a, b) \in \mathbb{R}^d \times \mathbb{R}^d, a \odot b = y\}. \end{aligned}$$

2. Let $(a, b) \in \mathbb{R}^d \times \mathbb{R}^d$ be fixed. For any (h, l),

$$\begin{aligned} f(a + h, b + l) &= \frac{1}{2} ||a \odot b - y + h \odot b + a \odot l||_2^2 + o(||h|| + ||l||) \\ &= f(a, b) + \langle h \odot b + a \odot l, a \odot b - y \rangle + o(||h|| + ||l||) \\ &= f(a, b) + \langle h, (a \odot b - y) \odot b \rangle + \langle l, (a \odot b - y) \odot a \rangle + o(||h|| + ||l||). \end{aligned}$$

Therefore,

$$\nabla f(a, b) = ((a \odot b - y) \odot b, (a \odot b - y) \odot a).$$

Let us now compute the Hessian. For any (h, l),

$$\begin{aligned} &\nabla f(a + h, b + l) \\ =&\nabla f(a, b) + (b \odot b \odot h + (2a \odot b - y) \odot l, a \odot a \odot l + (2a \odot b - y) \odot h) \\ &\quad + o(||h|| + ||l||). \end{aligned}$$

Therefore, for all (h, l),

$$\nabla^2 f(a, b)(h, l) = (b \odot b \odot h + (2a \odot b - y) \odot l, a \odot a \odot l + (2a \odot b - y) \odot h).$$

3. a) For any $i \leq d$,

$$\begin{aligned}(a_t)_i^2 + (b_t)_i^2 &\geq 2|(a_t)_i||(b_t)_i| \\ &= 2\left|((a_t)_i(b_t)_i - y_i) + y_i\right| \\ &\geq 2\left(|y_i| - |(a_t)_i(b_t)_i - y_i|\right) \\ &\geq 2\left(|y_i| - ||a_t \odot b_t - y||_2\right) \\ &\geq 2\left(m - \frac{m}{2}\right) \\ &= m,\end{aligned}$$

and

$$\begin{aligned}|((a_t)_i(b_t)_i - y_i)\,(a_t)_i(b_t)_i| &\leq ||a_t \odot b_t - y||_2\,|(a_t)_i(b_t)_i| \\ &\leq ||a_t \odot b_t - y||_2\,(|y_i| + |(a_t)_i(b_t)_i - y_i|) \\ &\leq \frac{m}{2}\left(M + \frac{m}{2}\right) \\ &\leq \frac{m}{2}\left(M + \frac{M}{2}\right) \\ &= \frac{3}{4}mM.\end{aligned}$$

b) It holds $(a_{t+1}, b_{t+1}) = (a_t, b_t) - \tau\nabla f(a_t, b_t)$, hence

$$\begin{aligned}a_{t+1} &= a_t - \tau(a_t \odot b_t - y) \odot b_t, \\ b_{t+1} &= b_t - \tau(a_t \odot b_t - y) \odot a_t,\end{aligned}$$

and

$$\begin{aligned}a_{t+1} \odot b_{t+1} - y = a_t \odot b_t - y &- \tau(a_t \odot b_t - y) \odot (a_t \odot a_t + b_t \odot b_t) \\ &+ \tau^2(a_t \odot b_t - y) \odot (a_t \odot b_t - y) \odot a_t \odot b_t \\ = (a_t \odot b_t - y) \odot (\mathbb{1} &- \tau(a_t \odot a_t + b_t \odot b_t) \\ &+ \tau^2(a_t \odot b_t - y) \odot a_t \odot b_t),\end{aligned}$$

which implies

$$\begin{aligned}||a_{t+1} \odot b_{t+1} - y||_2^2 &= \sum_{i=1}^{d}((a_t)_i(b_t)_i - y_i)^2\big(1 - \tau((a_t)_i^2 + (b_t)_i^2) \\ &\qquad + \tau^2((a_t)_i(b_t)_i - y_i)(a_t)_i(b_t)_i\big)^2 \\ &\leq \sum_{i=1}^{d}((a_t)_i(b_t)_i - y_i)^2\left(1 - \tau m + \frac{3}{4}\tau^2 mM\right)^2 \\ &\leq \sum_{i=1}^{d}((a_t)_i(b_t)_i - y_i)^2\left(1 - \tau m + \frac{\tau m}{2}\right)^2 \\ &= \left(1 - \frac{\tau m}{2}\right)^2 \sum_{i=1}^{d}((a_t)_i(b_t)_i - y_i)^2.\end{aligned}$$

Taking the square root, we obtain the result.

c) The results from the previous two questions, applied iteratively, show that, if $||a_0 \odot b_0 - y||_2 \leq \frac{m}{2}$, then, for any $t \geq 0$,

$$||a_t \odot b_t - y||_2 \leq \left(1 - \frac{\tau m}{2}\right)^t ||a_0 \odot b_0 - y||_2.$$

In particular, $||a_t \odot b_t - y||_2 \to 0$ when $t \to +\infty$, and thus $f(a_t, b_t) \to 0$.

d) For any $(a, b) \in \mathbb{R}^d \times \mathbb{R}^d$,

$$\begin{aligned} ||\nabla f(a,b)||_2 &= \left(\sum_{i=1}^{d} (a_i b_i - y_i)^2 (a_i^2 + b_i^2)\right)^{1/2} \\ &\leq ||a \odot b - y||_2 ||(a,b)||_2. \end{aligned}$$

From this, we can first deduce that $(a_t, b_t)_{t\in\mathbb{N}}$ is bounded. Indeed, for any $t \in \mathbb{N}$,

$$\begin{aligned} ||(a_{t+1}, b_{t+1})||_2 &\leq ||(a_t, b_t)||_2 + \tau ||\nabla f(a_t, b_t)||_2 \\ &\leq ||(a_t, b_t)||_2 \left(1 + \tau ||a_t \odot b_t - y||_2\right) \\ &\leq ||(a_t, b_t)||_2 \left(1 + \tau ||a_0 \odot b_0 - y||_2 \left(1 - \frac{\tau m}{2}\right)^t\right) \\ &\leq ||(a_t, b_t)||_2 \left(1 + \frac{\tau m}{2} \left(1 - \frac{\tau m}{2}\right)^t\right). \end{aligned}$$

Consequently, for any t,

$$\begin{aligned} ||(a_t, b_t)||_2 &\leq \left(\prod_{s=0}^{t-1} \left(1 + \frac{\tau m}{2} \left(1 - \frac{\tau m}{2}\right)^s\right)\right) ||(a_0, b_0)||_2 \\ &= \exp\left(\sum_{s=0}^{t-1} \ln\left(1 + \frac{\tau m}{2} \left(1 - \frac{\tau m}{2}\right)^s\right)\right) ||(a_0, b_0)||_2 \\ &\leq \exp\left(\sum_{s=0}^{t-1} \frac{\tau m}{2} \left(1 - \frac{\tau m}{2}\right)^s\right) ||(a_0, b_0)||_2 \\ &\leq \exp\left(\sum_{s=0}^{+\infty} \frac{\tau m}{2} \left(1 - \frac{\tau m}{2}\right)^s\right) ||(a_0, b_0)||_2 \\ &= \exp(1) ||(a_0, b_0)||_2, \end{aligned}$$

so that the sequence is indeed bounded. Let us denote $R = \sup_t ||(a_t, b_t)||_2$. For any t,

$$\begin{aligned} ||(a_{t+1}, b_{t+1}) - (a_t, b_t)||_2 &= \tau ||\nabla f(a_t, b_t)||_2 \\ &\leq \tau R ||a_t \odot b_t - y||_2 \\ &\leq \frac{\tau m R}{2} \left(1 - \frac{\tau m}{2}\right)^t. \end{aligned}$$

In particular, for any $t, t' \in \mathbb{N}$ such that $t \leq t'$,

$$||(a_{t'}, b_{t'}) - (a_t, b_t)||_2 \leq \frac{\tau m R}{2} \sum_{s=t}^{t'-1} \left(1 - \frac{\tau m}{2}\right)^s$$

$$\leq \frac{\tau m R}{2} \sum_{s=t}^{+\infty} \left(1 - \frac{\tau m}{2}\right)^s$$
$$= R \left(1 - \frac{\tau m}{2}\right)^t.$$

The sequence of iterates is therefore a Cauchy sequence. As a consequence, it converges. Since f is continuous and $f(a_t, b_t) \to 0$ when $t \to +\infty$, the limit point of $(a_t, b_t)_{t\in\mathbb{N}}$ is a zero of f, meaning that it is a global minimizer.

e) It is a local convergence result.

4. a) Let $(a, b) \in \mathbb{R}^d \times \mathbb{R}^d$ be arbitrary. It is a first-order critical point of f if and only if $\nabla f(a, b) = 0$, i.e. for any $i \leq d$,

$$\begin{aligned} & (a_i b_i - y_i) b_i = 0 \text{ and } (a_i b_i - y_i) a_i = 0 \\ \iff & (a_i b_i - y_i = 0 \text{ or } a_i = b_i = 0). \end{aligned}$$

The set of first-order critical points is

$$\left\{(a, b) \in \mathbb{R}^d \times \mathbb{R}^d, \forall i, a_i b_i - y_i = 0 \text{ or } a_i = b_i = 0\right\}.$$

b) Let $(a, b) \in \mathbb{R}^d \times \mathbb{R}^d$ be a first-order critical point of f. Let us determine under which condition it is a second-order critical point. From the expression we have found for the Hessian, it holds for any $(h, l) \in \mathbb{R}^d \times \mathbb{R}^d$ that

$$\left\langle \nabla^2 f(a, b)(h, l), (h, l) \right\rangle = \sum_{i=1}^{d} \left(h_i^2 b_i^2 + l_i^2 a_i^2 + 2(2a_i b_i - y_i) h_i l_i\right).$$

If $a_i b_i - y_i = 0$ for all $i \leq d$, then for any $(h, l) \in \mathbb{R}^d \times \mathbb{R}^d$,

$$\begin{aligned} \left\langle \nabla^2 f(a, b)(h, l), h, l \right\rangle &= \sum_{i=1}^{d} \left(h_i^2 b_i^2 + l_i^2 a_i^2 + 2 a_i b_i h_i l_i\right) \\ &= \sum_{i=1}^{d} \left(h_i b_i - l_i a_i\right)^2 \\ &\geq 0, \end{aligned}$$

so that (a, b) is a second-order critical point.
On the contrary, if there exists $i \leq d$ such that $a_i b_i - y_i \neq 0$ then, for this index i, it holds $a_i = b_i = 0$ (and $y_i \neq 0$). Therefore, if we choose $h = e_i$ (the i-th vector of the canonical basis), and $l = \operatorname{sgn}(y_i) h$, we have

$$\begin{aligned} \left\langle \nabla^2 f(a, b)(h, l), (h, l) \right\rangle &= -2 y_i h_i l_i \\ &= -2|y_i| \\ &< 0. \end{aligned}$$

This means that (a, b) is not a second-order critical point.
Second-order critical points are exactly the minimizers of f.

c) The results seen in class on gradient descent guarantee convergence of the iterates towards a second-order critical point, provided that f is analytic (which is true) and coercive (which is false). Therefore, they do not directly apply here. However, we can expect that

   — either the iterates diverge (due to the non-coercivity of f) ;

   — or they stay in a bounded region and, then, they converge to a second-order critical point of f, i.e. a global minimizer, for almost any initial point.

   (Actually, if the stepsize is small enough, the iterates do not diverge, and convergence to a global minimizer occurs for almost any initial point, but this is not easy to prove.)