Non-convex optimization: solution of the exercise

Irène Waldspurger*

December 8, 2022

Exercise

1. For any $x \in \mathbb{R}^n$, $\mathcal{L}(x) \geq 0$ (since it is a sum of squares). In addition,

$$\mathcal{L}(x_*) = \sum_{k=1}^{m} \left(\langle x_*, v_k \rangle^2 - | \langle x_*, v_k \rangle |^2 \right)^2 = 0.$$

Consequently, $\min \mathcal{L} = 0$ and, for any $x \in \mathbb{R}^n$,

$$\begin{aligned}
& x \text{ is a global minimizer of } \mathcal{L} \\
\iff & \mathcal{L}(x) = 0 \\
\iff & \sum_{k=1}^{m} (\langle x, v_k \rangle^2 - y_k^2)^2 = 0 \\
\iff & \langle x, v_k \rangle^2 - y_k^2 = 0, \quad \forall k = 1, \dots, m \\
\iff & \langle x, v_k \rangle = \pm | \langle x_*, v_k \rangle |, \quad \forall k = 1, \dots, m \\
\iff & | \langle x, v_k \rangle | = | \langle x_*, v_k \rangle |, \quad \forall k = 1, \dots, m.
\end{aligned}$$

2. The map $\mathcal{L}$ is polynomial in the coordinates of x. It is thus C^∞.

Let us compute its derivatives.

For any $x \in \mathbb{R}^n$, the gradient $\nabla \mathcal{L}(x)$ is the only vector such that

$$\mathcal{L}(x + w) = \mathcal{L}(x) + \langle \nabla \mathcal{L}(x), w \rangle + o(||w||).$$

*waldspurger@ceremade.dauphine.fr

And, for any x, w,

$$\begin{aligned}
\mathcal{L}(x+w) &= \sum_{k=1}^{m} \left(\langle x+w, v_k\rangle^2 - y_k^2\right)^2 \\
&= \sum_{k=1}^{m} \left(\langle x, v_k\rangle^2 - y_k^2 + 2\langle x, v_k\rangle \langle w, v_k\rangle + o(||w||)\right)^2 \\
&= \sum_{k=1}^{m} \Big[\left(\langle x, v_k\rangle^2 - y_k^2\right)^2 \\
&\qquad\qquad + 4\left(\langle x, v_k\rangle^2 - y_k^2\right)\langle x, v_k\rangle \langle w, v_k\rangle + o(||w||)\Big] \\
&= \mathcal{L}(x) + 4\sum_{k=1}^{m} \left(\langle x, v_k\rangle^2 - y_k^2\right)\langle x, v_k\rangle \langle w, v_k\rangle + o(||w||) \\
&= \mathcal{L}(x) + \left\langle 4\sum_{k=1}^{m} \left(\langle x, v_k\rangle^2 - y_k^2\right)\langle x, v_k\rangle v_k, w \right\rangle + o(||w||).
\end{aligned}$$

We thus have

$$\nabla\mathcal{L}(x) = 4\sum_{k=1}^{m} \left(\langle x, v_k\rangle^2 - y_k^2\right)\langle x, v_k\rangle v_k.$$

As to the Hessian at a point $x \in \mathbb{R}^n$, it is the only quadratic function such that, for any $l \in \mathbb{R}^n$,

$$\langle \nabla\mathcal{L}(x+h), l\rangle = \langle \nabla\mathcal{L}(x), l\rangle + \nabla^2\mathcal{L}(x)\cdot(h,l) + o(||h||).$$

For any x, h, l,

$$\begin{aligned}
\langle \nabla\mathcal{L}(x+h), l\rangle &= 4\sum_{k=1}^{m} \left(\langle x+h, v_k\rangle^2 - y_k^2\right)\langle x+h, v_k\rangle \langle v_k, l\rangle \\
&= 4\sum_{k=1}^{m} \left(\langle x, v_k\rangle^2 - y_k^2 + 2\langle x, v_k\rangle \langle h, v_k\rangle + o(||h||)\right) \\
&\qquad\qquad \times \left(\langle x, v_k\rangle + \langle h, v_k\rangle\right)\langle v_k, l\rangle \\
&= 4\sum_{k=1}^{m} \big[(\langle x, v_k\rangle^2 - y_k^2)\langle x, v_k\rangle \langle v_k, l\rangle
\end{aligned}$$

$$+ (3\langle x, v_k\rangle^2 - y_k^2)\langle h, v_k\rangle\langle v_k, l\rangle\Big] + o(||h||)$$

$$= \langle \nabla\mathcal{L}(x), l\rangle + 4\sum_{k=1}^m (3\langle x, v_k\rangle^2 - y_k^2)\langle v_k, h\rangle\langle v_k, l\rangle + o(||h||).$$

Consequently

$$\nabla^2\mathcal{L}(x)\cdot(h,l) = 4\sum_{k=1}^m (3\langle x, v_k\rangle^2 - y_k^2)\langle h, v_k\rangle\langle l, v_k\rangle,$$

which implies that, for any x, h,

$$\nabla^2\mathcal{L}(x)\cdot(h,h) = 4\sum_{k=1}^m (3\langle x, v_k\rangle^2 - y_k^2)\langle h, v_k\rangle^2.$$

Another possibility to solve the question would have been to compute the partial derivatives. Indeed, we know that, for any $x, h \in \mathbb{R}^n$,

$$\nabla\mathcal{L}(x) = \begin{pmatrix} \frac{\partial\mathcal{L}}{\partial x_1}(x) \\ \vdots \\ \frac{\partial\mathcal{L}}{\partial x_n}(x) \end{pmatrix} \quad \text{and} \quad \nabla^2 f(x)\cdot(h,h) = \sum_{i,j}\frac{\partial^2\mathcal{L}}{\partial x_i\partial x_j}(x)h_ih_j.$$

3. For any $x \in \mathbb{R}^n$,

$$\begin{aligned}\mathbb{E}(\nabla\mathcal{L}(x)) &= 4\sum_{k=1}^m \mathbb{E}\left(\langle x, v_k\rangle^3 v_k\right) - 4\sum_{k=1}^m \mathbb{E}\left(\langle x_*, v_k\rangle^2\langle x, v_k\rangle v_k\right)\\ &= 4\sum_{k=1}^m (3||x||^2 x) - 4\sum_{k=1}^m (2\langle x_*, x\rangle x_* + ||x_*||^2 x)\\ &= 4m\left(\left(3||x||^2 - ||x_*||^2\right)x - 2\langle x_*, x\rangle x_*\right).\end{aligned}$$

For any $x, h \in \mathbb{R}^n$,

$$\begin{aligned}\mathbb{E}\left(\nabla^2\mathcal{L}(x)\cdot(h,h)\right) &= 12\sum_{k=1}^m \mathbb{E}\left(\langle x, v_k\rangle^2\langle h, v_k\rangle^2\right) - 4\sum_{k=1}^m \mathbb{E}\left(\langle x_*, v_k\rangle^2\langle h, v_k\rangle^2\right)\\ &= 12\sum_{k=1}^m (2\langle x, h\rangle^2 + ||x||^2||h||^2)\\ &\qquad - 4\sum_{k=1}^m (2\langle x_*, h\rangle^2 + ||x_*||^2||h||^2)\\ &= 4m\left(6\langle x, h\rangle^2 - 2\langle x_*, h\rangle^2 + (3||x||^2 - ||x_*||^2)||h||^2\right).\end{aligned}$$

4. We start with the first-order stationary points. For any $x \in \mathbb{R}^n$, $\nabla(\mathbb{E}\mathcal{L})(x) = 0$ if and only if

$$4m\left(\left(3||x||^2 - ||x_*||^2\right)x - 2\left\langle x_*, x\right\rangle x_*\right) = 0.$$

This happens if and only if

$$3||x||^2 - ||x_*||^2 = \left\langle x_*, x\right\rangle = 0 \tag{1}$$

or

$$3||x||^2 - ||x_*||^2 \neq 0 \quad \text{and} \quad x = \frac{2\left\langle x_*, x\right\rangle}{3||x||^2 - ||x_*||^2}x_*. \tag{2}$$

The set of vectors x satisfying Equation (1) is

$$\left\{ \frac{||x_*||}{\sqrt{3}}u \,\middle|\, u \in \{x_*\}^\perp, ||u|| = 1 \right\}.$$

Additionally, a vector x satisfies Equation (2) if and only if it is colinear to x_* (that is, $x = \lambda x_*$ for some $\lambda \in \mathbb{R}$) and the colinearity factor λ is such that

$$0 \neq 3||\lambda x_*||^2 - ||x_*||^2 = (3\lambda^2 - 1)||x_*||^2$$

and

$$\begin{aligned}
\lambda x_* &= x \\
&= \frac{2\left\langle x_*, \lambda x_*\right\rangle}{3||\lambda x_*||^2 - ||x_*||^2} \\
&= \frac{2\lambda}{3\lambda^2 - 1}x_*.
\end{aligned}$$

These two equations are equivalent to the following conditions:

(a) $3\lambda^2 - 1 \neq 0$;

(b) $\lambda = \frac{2\lambda}{3\lambda^2-1}$, that is $\lambda = 0$ or $1 = \frac{2}{3\lambda^2-1}$, that is $\lambda \in \{-1, 0, 1\}$.

Consequently, the set of vectors x which satisfy Equation (2) is

$$\{-x_*, 0, x_*\}.$$

We have therefore shown that the set of first order critical points of $\mathbb{E}\mathcal{L}$ is

$$\left\{ \frac{||x_*||}{\sqrt{3}} u \,\middle|\, u \in \{x_*\}^\perp, ||u|| = 1 \right\} \cup \{-x_*, 0, x_*\}.$$

A second order critical point of $\mathbb{E}\mathcal{L}$ is a point x such that

(a) x is first order critical;

(b) $\nabla^2 \mathbb{E}\mathcal{L}(x) \succeq 0$.

Let us consider a first order critical point x, and determine whether $\nabla^2 \mathbb{E}\mathcal{L}(x) \succeq 0$.

- First case: $x = \frac{||x_*||}{\sqrt{3}} u$ for some unit-normed vector u orthogonal to x_*.

  For any h,

  $$\begin{aligned} &\nabla^2 \mathbb{E}\mathcal{L}(x) \cdot (h, h) \\ =&4m \left(6 \left\langle \frac{||x_*||}{\sqrt{3}} u, h \right\rangle^2 - 2 \langle x_*, h \rangle^2 + (3 \left|\left| \frac{||x_*||}{\sqrt{3}} u \right|\right|^2 - ||x_*||^2)||h||^2 \right) \\ =&4m \left(2||x_*||^2 \langle u, h \rangle^2 - 2 \langle x_*, h \rangle^2 \right). \end{aligned}$$

  In particular,

  $$\nabla^2 \mathbb{E}\mathcal{L}(x) \cdot (x_*, x_*) = -8m||x_*||^2 < 0.$$

  Therefore, $\nabla^2 \mathbb{E}\mathcal{L}(x) \not\succeq 0$.

- Second case: $x = 0$.

  For any h,

  $$\nabla^2 \mathbb{E}\mathcal{L}(x) \cdot (h, h) = -4m(2 \langle x_*, h \rangle + ||x_*||^2||h||^2).$$

  In particular,

  $$\nabla^2 \mathbb{E}\mathcal{L}(x) \cdot (x_*, x_*) = -12||x_*||^2 < 0.$$

  Therefore, $\nabla^2 \mathbb{E}\mathcal{L}(x) \not\succeq 0$.

- Third case: $x = \pm x_*$.
  For any h,

  $$\nabla^2 \mathbb{E}\mathcal{L}(x) \cdot (h, h) = 8m \left(2 \langle x_*, h \rangle^2 + ||x_*||^2 ||h||^2\right).$$

  This is a sum of squares, hence always nonnegative: $\nabla^2 \mathbb{E}\mathcal{L}(x) \succeq 0$.

The only second-order critical points are $-x_*$ and x_*.