Exercises Week 12: Optimization on nonsmooth sets through lifts

Instructor: Nicolas Boumal
TA: Andreea Musat (andreea.musat@epfl.ch)

May 8, 2025

1 Minimization on an interval

Let $f : \mathbb{R} \to \mathbb{R}$ be twice continuously differentiable. Consider the *constrained* minimization problem

$$\min_{x \in \mathbb{R}} f(x) \quad \text{subject to} \quad -1 \leq x \leq 1. \tag{P}$$

Alternatively, consider the *unconstrained* minimization problem

$$\min_{y \in \mathbb{R}} g(y) \quad \text{where} \quad g = f \circ \sin\,. \tag{Q}$$

1. Argue that (P) has at least one global minimizer.

   Answer. A continuous function on a compact domain has at least one global minimizer. ■

2. Argue that the optimal values of (P) and (Q) are equal.

   Answer. the lift $\sin\colon \mathbb{R} \to [-1, 1]$ is surjective, so the range of values attained by g is the same as for f, since $g(\mathbb{R}) = f(\sin(\mathbb{R})) = f([-1, 1])$. ■

A second-order critical (SOC) point for the unconstrained minimization problem (Q) is a point $y \in \mathbb{R}$ such that $g'(y) = 0$ and $g''(y) \geq 0$.

3. Argue carefully that if y is second-order critical for (Q) then $x = \sin(y)$ is stationary for (P).

   Answer. A point $y \in \mathbb{R}$ is SOC for (Q) if $g'(y) = 0$ and $g''(y) \geq 0$. We have that

$$g'(y) = \frac{\mathrm{d}}{\mathrm{d}y} f(\sin(y)) = f'(\sin(y)) \cos(y) \tag{1}$$

$$g''(y) = \frac{\mathrm{d}}{\mathrm{d}y} g'(y) = f''(\sin(y)) \cos^2(y) - f'(\sin(y)) \sin(y) \tag{2}$$

Consider the case $x \in (-1, 1)$ and let $x = \sin(y)$. Observe that $y \neq k\pi/2$ for $k \in \mathbb{Z}$. Since $\cos(y) \neq 0$, the condition $g'(0) = 0$ implies that $f'(x) = 0$, which shows that x is stationary for (P).

Now, consider the case $x = 1$, which corresponds to $y = \pi/2 + 2k\pi$ for some $k \in \mathbb{Z}$. Using that $\cos(y) = 0$, we have that

$$0 \leq g''(y) = -f'(x) \tag{3}$$

However, the normal cone at x is $N_x[-1,1] = \{v \in \mathbb{R} \mid v \geq 0\}$, so we have that $-f'(x) \in N_x[-1,1]$, which shows that x is indeed stationary for (P).

The case $x = -1$ is handled similarly. ■

4. Now assume f is convex. Deduce that if y is a local minimizer of g then in fact y is a global minimizer of g.

Answer. Since y local minimizer for g, it is a SOC for problem (Q). By the previous exercise, this implies that $x = \sin(y)$ is stationary for f. Since (P) is a convex problem when f is convex, it follows that x is the global minimizer for (P). ■

2 Sphere-to-simplex lift

Let $\mathbb{S}^{d-1} = \{y \in \mathbb{R}^d \mid \|y\| = 1\}$ be the unit sphere in $\mathbb{R}^d$ and define the simplex

$$\Delta^{d-1} = \{y \in \mathbb{R}^d \mid y_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{d} y_i = 1\}.$$

Consider the sphere-to-simplex (or Hadamard) lift, given by

$$\varphi \colon \mathbb{S}^{d-1} \to \Delta^{d-1}, \quad \varphi(y) = y^{\odot 2} = (y_1^2, \ldots, y_d^2) \tag{sphere-to-simplex}$$

We say that the lift φ satisfies the $2 \Rightarrow 1$ property if for any smooth function $f \colon \mathcal{E} \to \mathbb{R}$ and $g = f \circ \varphi$, if $y \in \mathbb{S}^{d-1}$ is a second order critical point for the problem $\min_{y' \in \mathbb{S}^{d-1}} g(y')$, then $x = \varphi(y)$ is a stationary point for the problem $\min_{x' \in \Delta^{d-1}} f(x')$.

Show that the lift φ satisfies the $2 \Rightarrow 1$ property. You may find it easier to first treat the case $d = 2$.

Hint: The tangent cone of the standard simplex has the following expression:

$$T_x \Delta^{d-1} = \{v \in \mathbb{R}^d \mid \sum_{i=1}^{d} v_i = 0 \text{ and } v_i \geq 0 \text{ for } i \notin \operatorname{supp}(x)\},$$

where for a point $x \in \Delta^{d-1}$, the support is $\operatorname{supp}(x) = \{i \mid x_i > 0\}$.
(This can be shown by using that for a convex set S, it holds that $\mathrm{T}_x S = \overline{\mathrm{K}_x S}$, where the cone of feasible directions $\mathrm{K}_x S$ is defined as $\mathrm{K}_x S = \{\alpha(y - x) \mid y \in S, \alpha \geq 0\}$. See Section 9.1 in these lecture notes for details.)

Answer. We start with some reminders from constrained optimization. Let $S \subset \mathcal{E}$ be a subset of a Euclidean space. The set $\mathrm{T}_x S$ is a cone. The polar of a cone C is the set

$$C^\circ = \{w \in \mathcal{E} \mid \langle w, v \rangle \leq 0 \text{ for all } v \in C\} \tag{4}$$

We define the normal cone of of S at x as $\mathrm{N}_x S = (\mathrm{T}_x \Delta^{d-1})^\circ$. The following is a standard fact from constrained optimization: if $f \colon \mathcal{E} \to \mathbb{R}$ is differentiable, we say that x^* is stationary for the problem $\min_{x \in S} f(x)$ if $\mathrm{D}f(x^*)[v] \geq 0$ for all $v \in \mathrm{T}_{x^*} S$. This is equivalent to $-\nabla f(x^*) \in \mathrm{N}_{x^*} S$. For details, see Chapter 7 in these lecture notes.

In particular, for the standard simplex, it can be shown that the normal cone has the following expression

$$N_x\Delta^{d-1} = \left\{ v \in \mathbb{R}^d \;\middle|\; \begin{array}{ll} v_i = \lambda & \text{for all } i \in \operatorname{supp}(x), \\ v_i \leq \lambda & \text{for all } i \notin \operatorname{supp}(x) \end{array} \quad \text{for some } \lambda \in \mathbb{R} \right\}.$$

We compute the (Riemannian) gradient $\nabla g(y)$ and (Riemannian) Hessian $\nabla^2 g(y)$ of g at y in terms of the gradient and Hessian of f. Let $y \in \mathcal{M}$, $x = \varphi(y)$ and $\dot{y} \in \mathrm{T}_y\,\mathcal{M}$. We have that

$$\mathrm{D}g(y)[\dot{y}] = \mathrm{D}(f \circ \varphi)(y)[\dot{y}] = \mathrm{D}f(x)\big[\mathrm{D}\varphi(y)[v]\big] = \langle \nabla f(x), \mathrm{D}\varphi(y)[\dot{y}]\rangle = \langle \mathrm{D}\varphi(y)^*[\nabla f(x)], \dot{y}\rangle.$$

Since $\mathrm{D}\varphi(y)[\dot{y}] = 2\mathrm{diag}(y)[\dot{y}]$, we obtain that the (Riemannian) gradient of g is

$$\nabla g(y) = 2\mathrm{Proj}_y\Big(\mathrm{diag}(y)\nabla f(\varphi(y))\Big).$$

Therefore, the first order criticality condition is

$$\nabla g(y) = 0 \quad \iff \quad \text{there exists } \lambda \in \mathbb{R} \text{ such that } [\nabla f(x)]_i = \lambda \text{ for all } i \in \operatorname{supp}(x). \tag{FOC}$$

Assume that y is a first order critical point and $y_i \neq 0$ for all $i = 1, \ldots, d$. Then, by (FOC), we obtain that there exists $\lambda \in \mathbb{R}$ such that $-\nabla f(x) = \lambda 1_d$. But since $N_x\Delta^{d-1} = \{\lambda 1_d \mid \lambda \in \mathbb{R}\}$, this shows that indeed x is stationary for f. It remains to study the case when there exists some i such that $y_i = 0$. For this, we use the second order criticality (SOC) condition.

The point $y \in \mathcal{S}^{d-1}$ is SOC if $\langle \dot{y}, \nabla^2 f(y)[\dot{y}]\rangle \geq 0$ for any $\dot{y} \in \mathrm{T}_y\,\mathbb{S}^{d-1}$. Take some smooth curve c on $\mathbb{S}^{d-1}$ with $c(0) = y$, $c'(0) = \dot{y}$ and $c''(0) = 0$, where $c''(t)$ is the intrinsic acceleration of c at t. Then, y is a SOC if and only if

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}t^2}\right|_{t=0} g(c(t)) = \langle \nabla^2 g(c(0))[c'(0)], c'(0)\rangle + \langle \nabla g(c(0)), c''(0)\rangle = \langle \nabla^2 g(y)[\dot{y}], \dot{y}\rangle \geq 0.$$

Let $\gamma = \varphi \circ c \colon (-\varepsilon, \varepsilon) \to \mathbb{R}^d$. Since $g \circ c = f \circ \varphi \circ c = f \circ \gamma$, we can equivalently write the SOC condition as

$$0 \leq \left.\frac{\mathrm{d}^2}{\mathrm{d}t^2}\right|_{t=0} f(\gamma(t)) = \langle \nabla^2 f(\gamma(0))[\gamma'(0)], \gamma'(0)\rangle + \langle \nabla f(\gamma(0)), \gamma''(0)\rangle \tag{SOC}$$

Take $c(t) = \cos(t)y + \sin(t)v$ (check that this satisfies the requirements above!) and since $\gamma(t) = c(t)^{\odot 2}$, we have that

$$\gamma'(t) = 2c(t) \odot \frac{\mathrm{d}}{\mathrm{d}t}c(t) \quad \text{and} \quad \gamma''(t) = 2\left(\Big(\frac{\mathrm{d}}{\mathrm{d}t}c(t)\Big)^{\odot 2} + c(t) \odot \frac{\mathrm{d}^2}{\mathrm{d}t^2}c(t)\right).$$

Pick some index $i \neq \operatorname{supp}(y)$ and let $\dot{y} = e_i \in \mathrm{T}_y\,\mathbb{S}^{d-1}$, since $y^\top e_i = y_i = 0$. Observe also that $y \odot \dot{y} = 0$. Then

$$\gamma'(0) = 2y \odot \dot{y} = 0 \quad \text{and} \quad \gamma''(0) = 2(\dot{y}^{\odot 2} - y^{\odot 2}).$$

That is, $[\gamma''(0)]_k = -y_k^2$ if $k \neq i$ and $[\gamma''(0)]_k = 1$ if $k = i$. With this, (SOC) becomes

$$\langle \nabla f(x), \dot{y}^{\odot 2} - y^{\odot 2}\rangle = \frac{\mathrm{d}f}{\mathrm{d}x_i}(x) - \sum_{k \in \operatorname{supp}(y)} y_k^2 \frac{\mathrm{d}f}{\mathrm{d}x_k}(x) \geq 0 \tag{5}$$

From (FOC), we have that there exists $\lambda \in \mathbb{R}$ such that $\frac{\mathrm{d}f}{\mathrm{d}x_k}(x) = \lambda$ for all $k \in \operatorname{supp}(y)$. Therefore, from Eq. (5), we obtain

$$-\frac{\mathrm{d}f}{\mathrm{d}x_i}(x) \leq -\lambda = -\frac{\mathrm{d}f}{\mathrm{d}x_k}(x), \text{ for any } k \in \operatorname{supp}(x).$$

This confirms that $-\nabla f(x) \in N_x\Delta^{d-1}$, so x is stationary for f on Δ^{d-1}. ∎