# MATH-329 Nonlinear optimization
# Exercise session 7: Intuitive constrained optimization

Instructor: Nicolas Boumal
TAs: Andrew McRae, Andreea Musat

Document compiled on October 29, 2024

**1. Warm-up: drawing tangent and normal cones.** Consider the following sets:

1. The half disk $\{x \in \mathbb{R}^2 : \|x\| \leq 1 \text{ and } x_1 \geq 0\}$.
2. The "opposite" of the half disk, $\mathbb{R}^2 \setminus \{x \in \mathbb{R}^2 : \|x\| < 1 \text{ and } x_1 > 0\}$.
3. The triangle limited by the vertices $(0, 0)$, $(1, 0)$ and $(0, 1)$.

For each of these, do the following. (Since this is a warm-up exercise, no proofs needed: just draw.)

1. Draw the set.
2. Identify interesting points, and draw the tangent cones there.
3. At the same points, draw also the normal cones.

Consider the definition of stationary points for minimization problems constrained to those sets, specifically the formulation that relates gradients and normal cones. Make sure this concept makes sense to you.

**Answer.** We highlight in green the tangent cones and in red the normal cones (we draw only vectors of the cones with norm bounded by a certain constant)

1. See Figure **??**.
2. See Figure **??**. Note that at the corner, we drew a single red point indicating that the normal cone contains only the zero vector.
3. See Figure **??**.

■

**2. Necessary optimality conditions in the interior of the constraint set.** Consider a set $S$ and a point $x$ in the interior of $S$ (that is, there exists a neighborhood of $x$ which is entirely contained in $S$). What are the tangent cones and normal cones at $x$? What are the necessary optimality conditions there? Does it make sense?

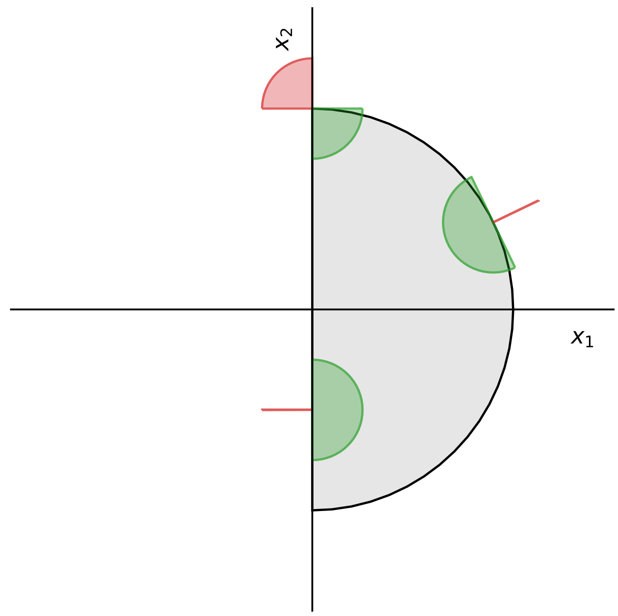


Figure 1: Half disk.

**Answer.** In this case the tangent cone is the whole embedding space $\mathcal{E}$. Let $v \in \mathcal{E}$ and $c(t) = x + tv$ be a curve. Then for all sufficiently small $t$ we have $c(t) \in S$. This is because $x$ is in the interior of $S$. So we find that $c'(0) = v$ is in $\mathrm{T}_x S$. We deduce immediately that the normal cone contains only the zero vector. This means that the first-order stationarity condition at $x$ is $\nabla f(x) = 0$. We recover exactly the same condition as when the optimization is unconstrained. This makes sense since none of the constraints are active at $x$. ■

**3. Optimizing on the unit circle.** Consider the following optimization problem:

$$\min_{(x,y)\in\mathbb{R}^2} x + y \qquad \text{subject to} \qquad x^2 + y^2 = 1.$$

1. Draw the feasible set and the gradient vector field of the cost function. Based on this drawing, can you guess what the solutions are? Can you guess what the stationary points are (see lecture notes)?

2. We may be tempted to solve the problem by eliminating $y$ with the change of variable $y = \pm\sqrt{1 - x^2}$. Do this, and show that the two possible signs lead to two different answers. Use the necessary optimality conditions (see lecture notes) to identify the right one.

**Answer.**

1. The search space is the unit circle; the gradient vector field is constant, pointing to the upper-right direction at a 45 degrees angle. Intuitively the solution is the point on the circle that is as much as possible in the negative gradient direction.

2. We use the changes of variable

$$y = \sqrt{1 - x^2} \qquad \text{and} \qquad y = -\sqrt{1 - x^2}$$

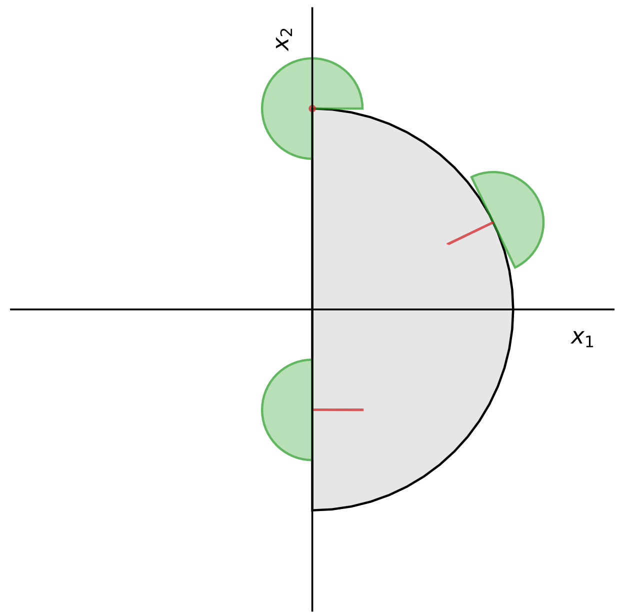


Figure 2: Opposite of half disk.

with $x \in [-1, 1]$. If we take the positive root, we need to solve

$$\min\; g_+(x) := x + \sqrt{1-x^2} \qquad \text{subject to} \qquad -1 \leq x \leq 1. \tag{1}$$

We seek for the critical points of $g_+$ in $]-1,1[$ by solving $g'_+(x) = 0$. We find:

$$g'_+(x) = 0 \iff \sqrt{1-x^2} = x \iff \begin{cases} x \geq 0 \\ 1 - x^2 = x^2 \end{cases} \iff x = \frac{\sqrt{2}}{2}.$$

This is the unique critical point of $g_+$ in $]-1,1[$ and we have $g_+(\frac{\sqrt{2}}{2}) = \sqrt{2}$, $g_+(-1) = -1$, $g_+(1) = 1$. So the solution to (**??**) is $x = -1$. As a consequence, a first candidate solution to the original problem is $(x^*_+, y^*_+) = (-1, 0)$.

Now consider the negative root. The problem becomes

$$\min\; g_-(x) := x - \sqrt{1-x^2} \qquad \text{subject to} \qquad -1 \leq x \leq 1. \tag{2}$$

We seek for the critical points of $g_-$ in $]-1,1[$ by solving $g'_-(x) = 0$. We find:

$$g'_-(x) = 0 \iff \sqrt{1-x^2} = -x \iff \begin{cases} x \leq 0 \\ 1 - x^2 = x^2 \end{cases} \iff x = -\frac{\sqrt{2}}{2}.$$

This is once again the unique critical point of $g_-$ in $]-1,1[$ and we have $g_-(\frac{\sqrt{2}}{2}) = -\sqrt{2}$, $g_-(-1) = -1$, $g_-(1) = 1$. So the solution to (**??**) is $x = -\frac{\sqrt{2}}{2}$. As a consequence $(x^*_-, y^*_-) = (-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$ is the other candidate solution for the original problem.

Since we have $x^*_+ + y^*_+ = -1 < -\sqrt{2} = x^*_- + y^*_-$, we conclude that taking the positive root leads us to a suboptimal candidate. This highlights a need for a more general and robust

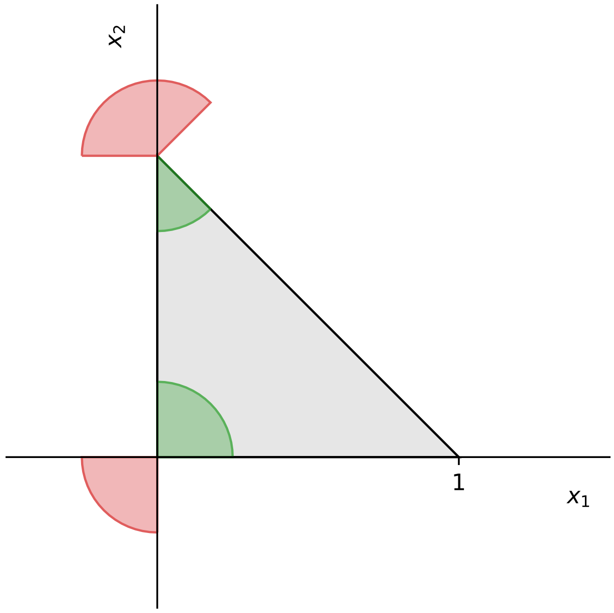


Figure 3: Triangle.

strategy to determine and assess constrained optimizers, and this is be the topic of the second part of the course.

Let $S = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$ be the unit circle. Let $p = (x, y) \in S$. The tangent cone of $S$ at $p$ is

$$\mathrm{T}_p S = \{v \in \mathbb{R}^2 \mid \langle p, v \rangle = 0\}.$$

The gradient of the cost function is $\begin{bmatrix} 1 & 1 \end{bmatrix}^\top$ everywhere. So if the point $p$ is an optimum then it satisfies that for all $v \in \mathrm{T}_p S$ we have

$$v_1 + v_2 \geq 0.$$

This is the first-order necessary condition. We can check that the only point of $S$ that satisfies this condition is $(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$. We conclude that the suboptimal candidate that we found is actually not even first-order stationary.

■

**4. Optimizing on the unit disk.** Find the *maxima* of $f(x, y) = xy$ on the closed unit disk, defined by the inequality $x^2 + y^2 \leq 1$. To do this, consider the tangent cones to the unit disk both in the interior and on the boundary; use this to determine the stationary points. As usual, we recommend you draw the situation. In particular sketch the gradient field.

**Answer.** See Figure **??**. The extreme value theorem (often named after Karl Weierstrass) states that a continuous function on a compact set achieves its maximum and minimum on the set. We know that extrema are attained either in the interior or on the boundary of the constraint set. If the function is differentiable, an extremum in the interior must be a point where the gradient is

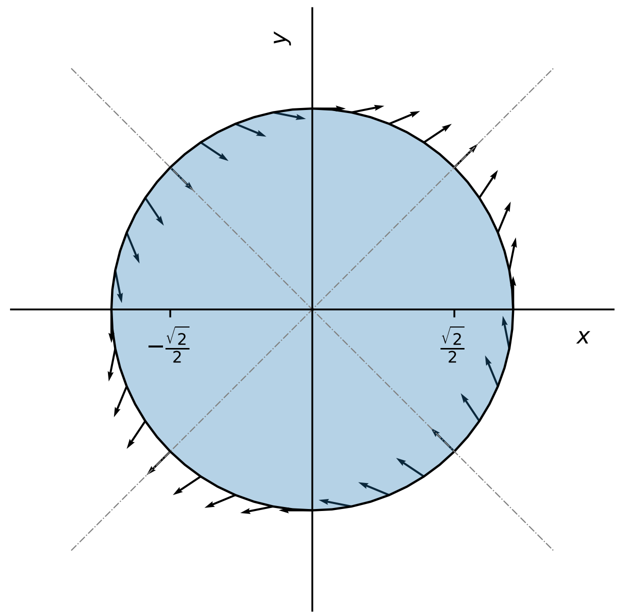


Figure 4: Unit disk and gradient of $f$ on the boundary.

zero. Thus we can decompose the problem into finding the maxima of $f$ in the interior and on the boundary of the unit disk.

In the interior of the disk, maxima can only be attained where the gradient is zero. They must satisfy

$$\nabla f(x, y) = (y, x) = 0.$$

So the unique candidate in the interior is $(0, 0)$. We can check that the Hessian of $f$ at that point is

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

which has one positive eigenvalue and one negative eigenvalue. Hence $(0, 0)$ is a saddle point and cannot be an optimum. We conclude that there is no optimum in the interior.

Now let's consider the boundary, which is the unit circle $S = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$. In the previous exercise we found that the tangent cone of $S$ at a point $p = (x, y) \in S$ is

$$\mathrm{T}_p S = \{v \in \mathbb{R}^2 \mid \langle p, v \rangle = 0\}.$$

Let $p = (x, y) \in S$. We have $\nabla f(x, y) = (y, x)$ so $p$ is first-order stationary if for all $v \in \mathrm{T}_p S$ we have

$$\langle \nabla f(x, y), v \rangle = y v_1 + x v_2 \geq 0.$$

Notice however that $\mathrm{T}_p S$ is a linear space. This implies that $p$ is stationary if and only if $\nabla f(p)$ is orthogonal to $\mathrm{T}_p S$. (Indeed if it is orthogonal then clearly it satisfies the inequality above. Conversely if it is not orthogonal then there exists $v \in \mathrm{T}_p S$ such that $\langle \nabla f(x, y), v \rangle \neq 0$ and either $v$ or $-v$ will not satisfy the condition.) The linear space $\mathrm{T}_p S$ is 1-dimensional and a basis for it is $(y, -x)$. So $p$ is first-order stationary if and only if $\nabla f(p)$ is orthogonal to $v = (y, -x)$, or equivalently, if and only if $y^2 = x^2$. We conclude that only the four points of the circle $S$ that intersect the cross $y^2 = x^2$ are stationary. If we compute the value of $f$ at these points we find that the two optimal solutions are

$$(x_1, y_1) = \Big(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\Big) \qquad \text{and} \qquad (x_2, y_2) = \Big(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\Big)$$

The objective function at these points is $\frac{1}{2}$. ■

**5. Tangent cone of the infinity norm ball.** Consider the closed infinity norm ball of unit radius given by

$$B_\infty = \left\{ x \in \mathbb{R}^n : \|x\|_\infty = \max_{i=1,\dots,n} |x_i| \leq 1 \right\}.$$

1. Draw this set for $n = 1$ and for $n = 2$.
2. What is the tangent cone for a point in the interior of $B_\infty$?
3. What is the tangent cone for a point on the boundary of $B_\infty$? Start with $n = 1$ and $n = 2$.

**Answer.**

1. For $n = 1$, it's the interval $[-1, 1]$ of the real line. For $n = 2$, it's a square with side of length 2.
2. The interior of $B_\infty$ is non-empty. In this case the tangent cone of a point $x$ in the interior is be the whole ambient space, i.e. $\mathrm{T}_x B_\infty = \mathbb{R}^n$. (Proven in exercise 2.)
3. We decompose the boundary as
$$\partial B_\infty = \cup_{i=1}^n B_\infty^i$$
where $B_\infty^i = \{x \in \mathbb{R}^n \mid x_i = 1, |x_j| \leq 1 \text{ if } j \neq i\}$. Each of the $B_\infty^i$ is a portion of the hyperplane $\langle e_i, x \rangle = 1$, where $e_i$ is the $i$th canonical vector. Given $x \in B_\infty^i$ and $v \in \mathbb{R}^n$, the curve $\gamma_{x,v}(t) = x + tv$ remains in $B_\infty$ for all sufficiently small $t \geq 0$ if and only if
$$\langle e_j, v \rangle = v_j \leq 0$$
for all $j$ such that $x_j = 1$. We let $L_j = \{v \in \mathbb{R}^n \mid v_j \leq 0\}$ and $I(x) = \{j \in [n] \mid x_j = 1\}$. The set $I(x)$ contains all the indices $j$ for which $x \in \partial B_\infty^j$. Using the considerations above we conclude that
$$\mathrm{T}_x B_\infty = \bigcap_{j \in I(x)} L_j.$$

■

# Supplementary exercises

**1. Optimizing on the unit circle with change of variable.** In an exercise above we saw that a change of variables can introduce spurious solutions. However some variable changes are better than others. In this one you will parametrize the circle using $\theta \mapsto (\cos\theta, \sin\theta)$. Consider the constrained optimization problem.

$$\min_{(x,y)\in\mathbb{R}^2} xy \quad \text{subject to} \quad x^2 + y^2 = 1$$

1. Draw the gradient of $f(x, y) = xy$ at various points of the unit circle. Based on this drawing, guess where the stationary points are, and guess which points are the minimizers.
2. Introduce the change of variable described above to obtain an optimization problem of one variable.
3. Solve this single-variable optimization problem. Explain how you use your findings to determine the minimizers of $f$ on the circle.

**Answer.**

1. We suspect stationary points will occur when the gradient is orthogonal to the unit circle, that is at the intersection of the circle and the diagonals of the four quadrants (see Figure **??**).

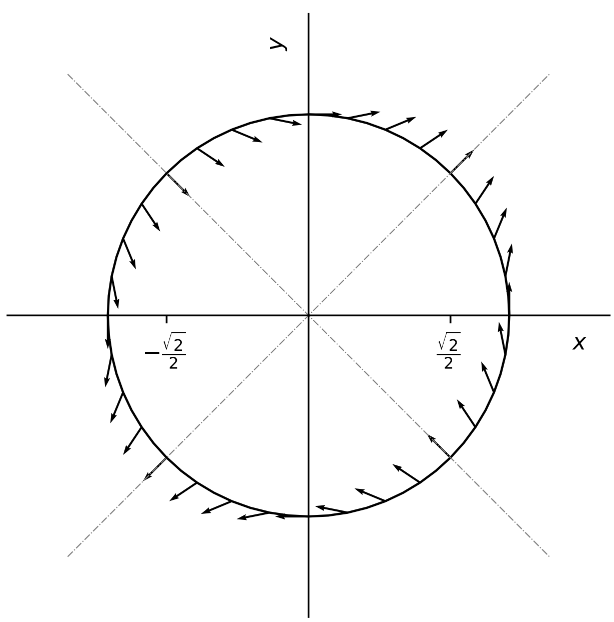


Figure 5: Unit circle and gradient of $f$.

2. In polar coordinates, the feasible set can be parametrized as

$$x = \cos(\theta),\ y = \sin(\theta) \qquad \text{with } \theta \in [0, 2\pi[.$$

(Note: we could also omit the constraint on $\theta$ because of the periodicity of the parametrization: in this case the arguments must be slightly adapted.) The optimization problem becomes

$$\min_{\theta} g(\theta) := \cos(\theta)\sin(\theta) \quad \text{subject to} \quad 0 \leq \theta < 2\pi.$$

3. We seek critical points of $g$ in $]0, 2\pi[$ by solving

$$g'(\theta) = 0 \quad \Leftrightarrow \quad -\sin(\theta)^2 + \cos(\theta)^2 = 0 \quad \Leftrightarrow \quad 2\cos(\theta)^2 = 1$$

which has four solutions

$$\theta_1 = \frac{\pi}{4}, \quad \theta_2 = \frac{3\pi}{4}, \quad \theta_3 = \frac{5\pi}{4}, \quad \theta_4 = \frac{7\pi}{4}.$$

Evaluating the objective we find

$$g(\theta_1) = \frac{1}{2}, \quad g(\theta_2) = -\frac{1}{2}, \quad g(\theta_3) = \frac{1}{2}, \quad g(\theta_4) = -\frac{1}{2}$$

From this, we deduce the original optimization problem has 2 different solutions $(x_1^*, y_1^*) = \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ and $(x_2^*, y_2^*) = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$. (Details omitted.)

■

**2. Infinity norm minimization.** Let $F\colon \mathbb{R}^n \to \mathbb{R}^m$ be a smooth vector function and consider the unconstrained optimization problem of minimizing $f(x)$ where

$$f(x) = \|F(x)\|_\infty = \max_{i=1,\ldots,m} |F_i(x)|.$$

Notice that this cost function is (generally) *not* smooth. Reformulate the problem as a *smooth* constrained optimization problem, that is: describe the search space $S$ with smooth inequalities, and arrange for the cost function to be smooth as well. *Hint: the essential trick is to add a new, "fake" variable. You can look in the Nocedal & Wright for inspiration.*

**Answer.** A trivial yet effective translation of this problem to a constrained setting involves defining an extra variable $t \in \mathbb{R}$ and solve

$$\min_{x,t} t \quad \text{subject to} \quad \begin{cases} t - F_i(x) \geq 0 \\ t + F_i(x) \geq 0 \end{cases} \quad \forall i = 1, \ldots, m.$$

This is indeed a smooth constrained optimization problem as both the objective function and the functions defining the feasible set are smooth. Note however that there is no magic happening: the problem is now smooth but the additional constraints may be difficult to handle. ■