University of Vienna
Faculty of Mathematics

Nonlinear Optimization Problems

Milutin Popovic

Contents

1 Sheet 1

1.1 Exercise 1

Let $X \subseteq \mathbb{R}^n$ be an nonempty convex set and $f : \mathbb{R}^n \to \mathbb{R}$ a convex function. Show that every local minimum of f w.r.t. X is a global minimum of f w.r.t. X.

Let $x^* \in X$ be the local minimum of f w.r.t. X. Then there is an ε-ball $B_\varepsilon(x^*) = \{x \in X : \|x^* - x\| \leq \varepsilon\}$ such that $f(x^*) \leq f(x)$ for all $x \in B_\varepsilon(x^*) \cap X$. Now suppose x^* is not a global minimum then there is a $x_0 \in X$ such that $f(x_0) < f(x^*)$. Since X is convex there is a line connecting x_0 and x^* in X. This line has elements of the form

$$L = \{(1-\lambda)\, x_0 - \lambda x^* : \ \lambda \in [0,1]\} \subseteq X \tag{1}$$

Now for all $z \in L$ with $z = (1-\lambda)x_0 - \lambda x^*$ $(\forall \lambda \in [0,1])$ it holds that

$$\begin{align} \|x^* - (1-\lambda)x_0 - \lambda x^*\| \leq \varepsilon \tag{2} \\ (1-\lambda)\|x^* - x_0\| \leq \varepsilon \ \Rightarrow \ \lambda \simeq 1 \tag{3} \end{align}$$

and

$$\begin{align} f\left((1-\lambda)\, x_0 + \lambda x\right) &\leq \lambda f(x^*) + (1-\lambda) f(x_0) \tag{4} \\ &< \lambda f(x_0) + (1-\lambda) f(x_0) \tag{5} \\ &= f(x_0) \tag{6} \end{align}$$

This means that we found a $\tilde{x} = \lambda x^* - (1-\lambda)x_0 \in B_\varepsilon(x^*)$ such that $f(\tilde{x}) < f(x^*)$ which is a contradiction since x^* is a local minimum in $B_\varepsilon(x^*)$. This means that x^* is a global minimum of f w.r.t X.

1.2 Exercise 2

Let $X \subseteq \mathbb{R}^n$ be a nonempty, $x_0 \in X$. Show that

1. $T_X(x_0)$ is a nonempty closed cone
2. If X is convex, then $T_X(x_0) = \mathrm{cl}\left(\bigcup_{\lambda \geq 0} \lambda\,(X - x_0)\right)$
3. If X is convex then $(T_X(x_0))^* = -N_X(x_0))$

The statements will be proven in chronological order. Starting with 1. the Bouligand tangent cone to X is defined as

$$T_X(x_0) = \{d \in \mathbb{R}^n : \exists (x^k)_{k\geq 0} \subset X, \exists (t_k)_{k \geq 0} \searrow 0 : \frac{x^k - x_0}{t_k} \to d\} \tag{7}$$

Now $T_X(x_0)$ is nonempty since for $x^k = x_0$ for all k and a sequence $t_k = \frac{1}{k}$ we have

$$\frac{x^k - x_0}{t_k} \to 0 \in T_X(x_0) \tag{8}$$

To show that $T_X(x_0)$ is closed, consider a sequence $(d_k)_{k\geq 0} \subset T_X(x_0)$ with convergence point $d \in \mathbb{R}^n$. To show that it is closed we need to show that $d \in T_X(x_0)$. So for all $d_n \in (d_k)_{k\geq 0} \subset T_X(x_0)$ there exists a sequence $(x^{n,k})_{k\geq 0} \subset X$ and a sequence $(t_{n,k})_{k\geq 0}$ with $t_{n,k} \searrow 0$ as $k \to \infty$ such that

$$\frac{x^{n,k} - x_0}{t_{n,k}} \to d_n \quad \forall n \geq 0, \tag{9}$$

and

$$\lim_{n\to\infty} \lim_{k\to\infty} \frac{x^{n,k} - x_0}{t_{n,k}} = d. \tag{10}$$

Then there exist K_ε and N_δ for $\varepsilon, \delta > 0$ such that

$$\|d_k - d\| \leq \varepsilon \;\; \forall k \geq K_\varepsilon, \tag{11}$$

$$\|\frac{x^{n,K_\varepsilon} - x_0}{t_{n,K_\varepsilon}} - d_{K_\varepsilon}\| \leq \delta \;\; \forall n \geq N_\delta. \tag{12}$$

To conclude the proof consider

$$\|\frac{x^{N_\delta,K_\varepsilon} - x_0}{t_{N_\delta,K_\varepsilon}} - d\| \tag{13}$$

$$= \|\left(\frac{x^{N_\delta,K_\varepsilon} - x_0}{t_{N_\delta,K_\varepsilon}} - d_{K_\varepsilon}\right) + (d_{K_\varepsilon} - d)\,\| \tag{14}$$

$$\leq \varepsilon + \delta \tag{15}$$

Meaning that $d \in T_X(x_0)$. Then $T_X(x_0)$ is really a cone because $\forall d \in T_X(x_0)$ and $\lambda > 0$ we have that $\lambda d \in T_X(x_0)$ by the choice $t_{k,\lambda} = \frac{1}{\lambda} t_k$

$$\frac{x^k - x}{t_{k,\lambda}} = \lambda \frac{x^k - x_0}{t_k} \to \lambda d \in T_X(x_0) \tag{16}$$

For number 2 additionally X is a convex set. And by definition of a cone $\bigcup_{\lambda\geq 0}(X - x_0)$ is a cone. And the union of convex sets is also convex the set $\mathrm{cl}\left(\bigcup_{\lambda\geq_0} \lambda(X - x_0)\right)$ is also. convex. For number 3 we simply calculate

$$-\left(T_X(x_0)\right)^* = -\{s \in R^n : s^T d \geq 0 \forall d \in T_X(x_0)\} \tag{17}$$

$$= \{s \in R^n : s^T d \leq 0 \forall d \in T_X(x_0)\} \tag{18}$$

$$= \{s \in R^n : s^T (x - x_0) \leq 0 \forall x \in X\} \tag{19}$$

Since $forall x \in X$ there exists an appropriate sequence converging to x subjected to the tangent cone elements.

1.3 Exercise 3

Let $X \subseteq \mathbb{R}^n$ be nonempty and $\operatorname{dist}_X : \mathbb{R}^n \to \mathbb{R}$, with $\operatorname{dist}_X(y) = \inf\{\|y - x\| : x \in X\}$. Then consider the directional derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ at direction d at $x_0 \in \mathbb{R}^n$

$$f'(x_0; d) = \lim_{t \downarrow 0} \frac{f(x_0 + td) - f(x_0)}{t} \tag{20}$$

Show that if $X \subseteq \mathbb{R}^n$ is nonempty and convex then the tangent cone can be written as

$$T_X(x_0) = \{d \in \mathbb{R}^n : \left(\operatorname{dist}_X\right)'(x_0; d) = 0\}. \tag{21}$$

First we note

$$(\operatorname{dist}_X)'(x_0; d) = \lim_{t \downarrow 0} \frac{\operatorname{dist}_X(x_0 + td)}{t} = 0, \tag{22}$$

is true for all $x_0 + td \in X$. Since X is convex we have that

$$\begin{aligned} T_X(x_0) &= \operatorname{cl}\left(\bigcup_{\lambda} (X - x_0)\right) && (23)\\ &= \operatorname{cl}\left(\{\lambda(x - x_0) : x \in X, \lambda \geq 0\}\right) && (24) \end{aligned}$$

Then $(\operatorname{dist}_X)'(x_0; d) = 0$ holds only for vectors of the form $d = \lambda(x - x_0)$, $\lambda \geq 0$ and $x \in X$.

1.4 Exercise 4

We consider the general constrained optimization problem

$$\begin{aligned} \min \quad & f(x), && (25)\\ \text{s.t.} \quad & g_i(x) \leq 0, i = 0, \ldots, m \\ & h_j(x) = 0, j = 0, \ldots, p \\ & x \in \mathbb{R}^n \end{aligned}$$

for $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}$, $i = 0, \ldots, m$ and $j = 0, \ldots, p$ continuously differentiable. Let x_0 be a feasible element of the problem and

$$\begin{aligned} X_{\text{lin}} := \{x \in \mathbb{R}^n : \quad & g_i(x_0) + \nabla g_i(x_0)^T (x - x_0), i = 0, \ldots, m && (26)\\ & h_j(x_0) + \nabla h_j(x_0)^T (x - x_0), j = 0, \ldots, p\}. && (27) \end{aligned}$$

Show that $T_{\text{lin}}(x_0) = T_{X_{\text{lin}}}(x_0)$.
For the reminder

$$\begin{aligned} T_{\text{lin}}(x_0) := \{d \in \mathbb{R}^n : \quad & \nabla g_i(x_0)^T d, \forall i \in \mathcal{A}(x_0) && (28)\\ & \nabla h_j(x_0)^T d, j = 0, \ldots, p\}, && (29) \end{aligned}$$

where $\mathcal{A}(x_0) = \{i = i, \ldots, m : g_i(x_0) = 0\}$. First we show that X_{lin} is convex then we can use the second part of exercise 2. So for all $x, y \in X_{\text{lin}}$ we need to show that $z = (1 - \lambda)x\lambda y \in X_{\text{lin}}$, this is done by checking the conditions

$$\begin{aligned} & \nabla g_i(x_0) + \nabla g_i(x_0)^T (z - x_0) && (30)\\ &= \nabla g_i(x_0) + \nabla g_i(x_0)^T ((1 - \lambda)x + \lambda y - x_0) && (31)\\ &= \nabla g_i(x_0) + \nabla g_i(x_0)^T (x - \lambda x + \lambda y - x_0) && (32)\\ &= \nabla g_i(x_0) + \nabla g_i(x_0)^T (x - x_0) + \nabla g_i(x_0)^T (-\lambda x + \lambda y) && (33)\\ &\leq \lambda \nabla g_i(x_0)^T (-x + y) \qquad (\lambda \geq 0) && (34)\\ &\leq \nabla g_i(x_0)^T (-x + y + x_0 - x_0) + g_i(x_0) - g_i(x_0) && (35)\\ &= -\nabla g_i(x_0) - \nabla g_i(x_0)^T (x - x_0) + \nabla g_i(x_0) + \nabla g_i(x_0)^T (y - x_0) && (36)\\ &\leq 0 && (37) \end{aligned}$$

and similarly for h_j

$$\begin{aligned}
&\nabla h_j(x_0) + \nabla h_j(x_0)^T(z - x_0) && (38)\\
&= \nabla h_j(x_0) + \nabla h_j(x_0)^T((1-\lambda)x + \lambda y - x_0) && (39)\\
&= \nabla h_j(x_0) + \nabla h_j(x_0)^T(x - \lambda x + \lambda y - x_0) && (40)\\
&= \nabla h_j(x_0) + \nabla h_j(x_0)^T(x - x_0) + \nabla h_j(x_0)^T(-\lambda x + \lambda y) && (41)\\
&= \lambda \nabla h_j(x_0)^T(-x + y) && (42)\\
&= \lambda \nabla h_j(x_0)^T(-x + y + x_0 - x_0) + \lambda h_j(x_0) - \lambda h_j(x_0) && (43)\\
&= \lambda \left(-\nabla h_j(x_0) - \nabla h_j(x_0)^T(x - x_0) + \nabla h_j(x_0) + \nabla h_j(x_0)^T(y - x_0)\right) && (44)\\
&= 0 && (45)
\end{aligned}$$

Thereby X_{lin} is convex and

$$T_{X_{\text{lin}}}(x_0) = \bigcup_{\lambda \geq 0} \lambda\left(X_{\text{lin}} - x_0\right). \tag{46}$$

Additionally $T_{lin}(x_0)$ is a polyhedral so per definition

$$T_{\text{lin}}(x_0) = \bigcup_{\lambda \geq 0} \lambda\left(X_{\text{lin}} - x_0\right) \tag{47}$$

1.5 Exercise 5

Let $g : \mathbb{R}^2 \to \mathbb{R}^4$ with

$$g(x, y) = \begin{pmatrix} \pi - 2x \\ -y - 1 \\ 2x - 3\pi \\ y - \sin(x) \end{pmatrix}. \tag{48}$$

and

$$X = \{(x, y) \in \mathbb{R}^2 : g(x, y) \leqq 0\}. \tag{49}$$

The set X has the following constrains on x, y

$$\begin{aligned}
\pi - 2x \leq 0 \;\Rightarrow x \geq \frac{\pi}{2} && (50)\\
-y - 1 \leq 0 \Rightarrow y \geq -1 && (51)\\
2x - 3\pi \leq\Rightarrow x \leq \frac{3\pi}{2} && (52)\\
y - \sin(x) \leq 0 \Rightarrow y \leq 1, && (53)
\end{aligned}$$

meaning $x \in [\frac{\pi}{2}, \frac{3\pi}{2}]$ and $y \in [-1, 1]$. The tangent cones of X w.r.t $x_1 = (\frac{\pi}{2}, 1)^T, x_2 = (\pi, 0)^T$ and $x_3 = (\frac{3\pi}{2}, -1)^T$ are

$$\begin{aligned}
T_X(x_1) = \{(0, -\lambda)^T, (\lambda, 1)^T : \lambda > 0\} && (54)\\
T_X(x_2) = \{\lambda(\cos(x), \sin(x))^T, (\lambda, 1)^T : \lambda > 0\} && (55)\\
T_X(x_3) = \{(-\lambda, 0)^T, (0, \lambda)^T : \lambda > 0\} && (56)
\end{aligned}$$

Graphically it represents the following

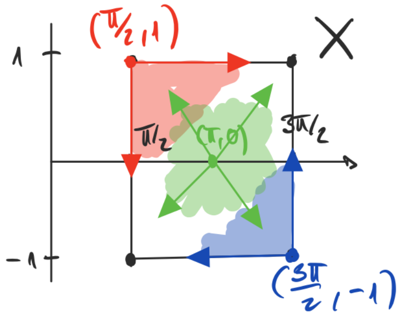

Figure 1: Graphical representation of X and tangent cones $T_X(x_i)$, $i = 1, 2, 3$

Then we look at the linearized tangent cones $T_{\text{lin}}(x_i) = \{d \in \mathbb{R}^2 : \nabla g_i(x_i)^T d \leq 0 \forall i \in \mathcal{A}(x_i)\}$. First we calculate the gradients of entries of g

$$\nabla g_1(x, y) = (-2, 0)^T \qquad \nabla g_2(x, y) = (0, -1)^T \tag{57}$$
$$\nabla g_3(x, y) = (2, 0)^T \qquad \nabla g_4(x, y) = (-\cos(x), 1)^T \tag{58}$$

then for all $j \in \{1, 2, 3, 4\}$ and $i \in \{1, 2, 3\}$ we check if $g_j(x_i) \leq 0$ and construct $\mathcal{A}(x_i)$ and then find d subjected to the condition. For x_1 we have

$$g_1(x_1) = 0, \quad g_2(x_1) \neq 0, \quad g_3(x_1) \neq 0, \quad g_4(x_1) = 0 \tag{59}$$
$$\Rightarrow \mathcal{A}(x_1) = \{1, 2\} \tag{60}$$

Then

$$T_{\text{lin}} = \{d \in \mathbb{R}^2 : (0, 1)d \leq 0, (-2, 0)d \leq 0\} \tag{61}$$
$$= \{(\lambda, 0)^T, (0, -\lambda)^T : \lambda > 0\} \tag{62}$$

For x_2 we have

$$g_1(x_1) \neq 0, \quad g_2(x_1) \neq 0, \quad g_3(x_1) \neq 0, \quad g_4(x_1) \neq 0 \tag{63}$$
$$\Rightarrow \mathcal{A}(x_1) = \{4\} \tag{64}$$

Then

$$T_{\text{lin}} = \{d \in \mathbb{R}^2 : (1, 1)d \leq 0\} \tag{65}$$
$$= \{(-\lambda, \lambda)^T, (\lambda, -\lambda)^T, (-\lambda, 0)^T, (0, -\lambda)^T : \lambda > 0\} \tag{66}$$

For x_3 we have

$$g_1(x_1) \neq 0, \quad g_2(x_1) = 0, \quad g_3(x_1) = 0, \quad g_4(x_1) = 0 \tag{67}$$
$$\Rightarrow \mathcal{A}(x_1) = \{2, 3, 4\} \tag{68}$$

Then

$$T_{\text{lin}} = \{d \in \mathbb{R}^2 : (0, -1)d \leq 0, (2, 0)d \leq 0, (0, 1)d \leq 0\} \tag{69}$$
$$= \{(-\lambda, 0)^T : \lambda > 0\} \tag{70}$$

We conclude that $T_{X_{\text{lin}}}(x_i) = T_{\text{lin}}(x_i)$ only for $i = 1$.

1.6 Exercise 6

Let $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^m$. Prove using the strong duality theorem of linear optimization that the following statements are equivalent

1. The system $Ax = b$ has a solution $x \leqq 0$
2. It holds $b^T d \geq 0$ for all $d \in \mathbb{R}^m$ with $A^T d \leqq 0$

we show that 1 is equivalent to $\neg\exists d \in R^m : Ad \leq 0$ and $b^T d > 0$. Consider the following primal dual

$$\begin{aligned} \max \quad & 0^T x \\ \text{s.t.:} \quad & Ax = b \\ & x \leqq 0 \end{aligned} \tag{71}$$

and

$$\begin{aligned} \min \quad & b^T d \\ \text{s.t.:} \quad & Ad \leq 0 \\ & d \in R^m \end{aligned} \tag{72}$$

$$\tag{73}$$

Consider a solution of $Ax = b$ such that $x \geq 0$. This means that the primal is true, so there exists an optimal solution since $0^T x = 0$ then 0 must be this primal optimal. By the duality 0 must be the optimal of the dual problem, which means $b^T d = 0$ so there is no $d \in \mathbb{R}^m : b^T d > 0$. On the other hand $\neg\exists d \in \mathbb{R}^m : Ad \leq 0$ and $b^T d > 0$ so $\forall d' \in \mathbb{R}^m$ we have that $Ad' > 0$ or $b^T d \leq 0$ such that $b^T d \leq 0$ for all $d \in \mathbb{R}^m$. So we can conclude that there exists a solution because the primal has at least one feasible $x \in \mathbb{R}^m : Ax = 0, x \geqq 0$. Thereby the two statements above are equivalent.