NMSA403 Optimization Theory – Exercises

Martin Branda

Charles University, Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

Version: November 28, 2018

Contents

1 Introduction and motivation

1.1 Operations Research/Management Science and Mathematical Programming

Goal: improve/stabilize/set of a system. You can reach the goal in the following steps:

- Problem understanding
- Problem description – probabilistic, statistical and econometric models
- Optimization – mathematical programming (formulation and solution)
- Verification – backtesting, stresstesting
- Implementation (Decision Support System)
- Decisions

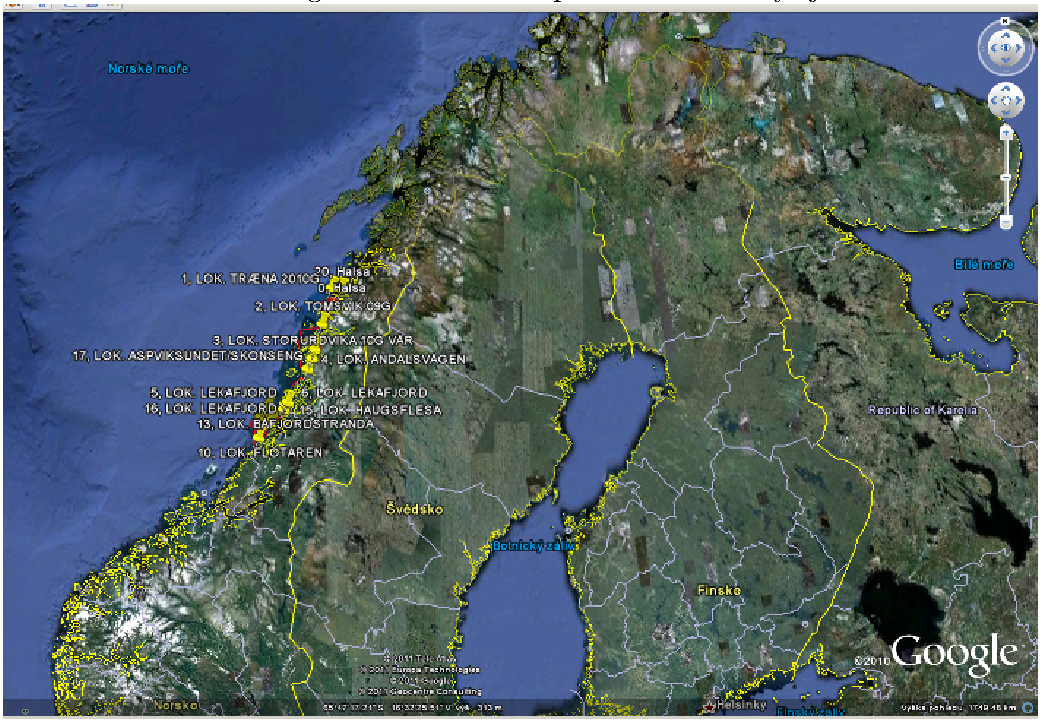
1.2 Marketing – Optimization of advertising campaigns

- **Goal** – maximization of the effectiveness of a advertising campaign given its costs or vice versa
- **Data** – "peoplemeters", public opinion poll, historical advertising campaigns
- **Target group** – (potential) customers (age, region, education level ...)
- **Effectiveness criteria**
  - GRP (TRP) – rating(s)
  - Effective frequency – relative number of persons in the target group hit k-times by the campaign
- Nonlinear (nonconvex) or integer programming

1.3 Logistic – Vehicle routing problems

- **Goal** – maximize *filling rate* of the ships (operation planning), fleet composition, i.e. capacity and number of ships (strategic planning)
- **Rich Vehicle Routing Problem**
  - time windows
  - heterogeneous fleet
  - several depots and inter-depot trips
  - several trips during the planning horizon
  - *non-Euclidean distances* (fjords)
- Integer programming :-(, constructive heuristics and tabu search

Figure 1: A boat trip around Norway fjords



Literature: M.B., K. Haugen, J. Novotný, A. Olstad (2017).

Related problems with increasing complexity:

- Traveling Salesman Problem
- Uncapacitated Vehicle Routing Problem (VRP)
- Capacitated VRP
- VRP with Time Windows
- ...

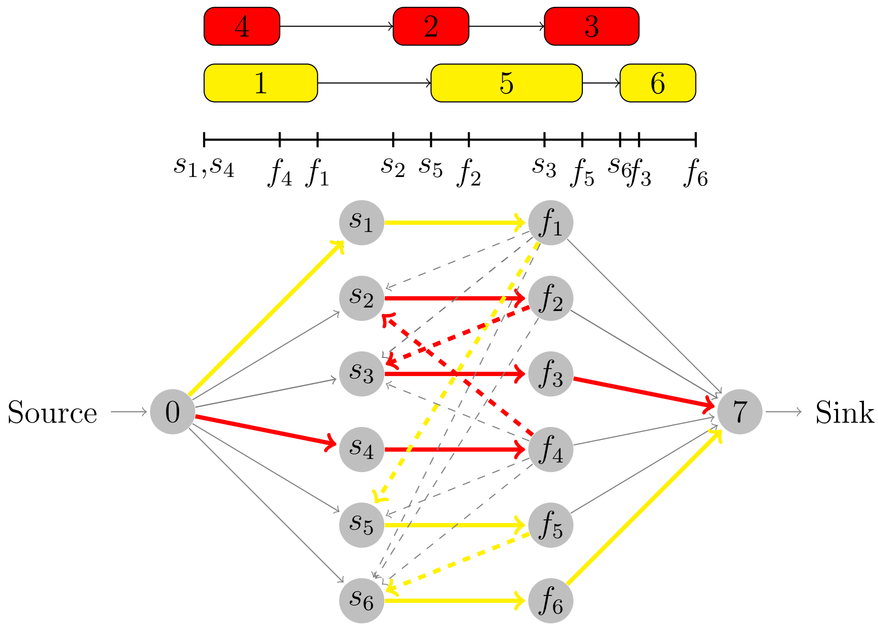
Approach to problem solution:

1. Mathematical formulation
2. Solving using GAMS based on historical data
3. Heuristic(s) implementation
4. Implementation to a Decision Support System

1.4 Scheduling – Reparations of oil platforms

- **Goal** – send the right workers to the oil platforms taking into account uncertainty (bad weather – helicopter(s) cannot fly – jobs are delayed)
- **Scheduling** – jobs = reparations, machines = workers (highly educated, skilled and costly)

Figure 2: Fixed interval schedule (assignment of 6 jobs to 2 machines) and corresponding network flow



- Integer and stochastic programming

Literature: M. Branda, J. Novotný, A. Olstad (2016), M. Branda, S. Hájek (2017)

1.5 Insurance – Pricing in nonlife insurance

- **Goal** – optimization of prices in MTPL/CASCO insurance taking into account riskiness of contracts and competitiveness of the prices on the market
- **Risk** – compound distribution of random losses over 1y (Data-mining & GLM)
- Nonlinear stochastic optimization (probabilistic or expectation constraints)
- See Table 1

Literature and detailed information: M.B. (2012, 2014)

1.6 Power industry – Bidding, power plant operations

Energy markets

- **Goal** – profit maximization and risk minimization
- **Day-ahead bidding** from wind (power) farm
- Nonlinear stochastic programming

Power plant operations

Table 1: Multiplicative tariff of rates: final price can be obtained by multiplying the coefficients

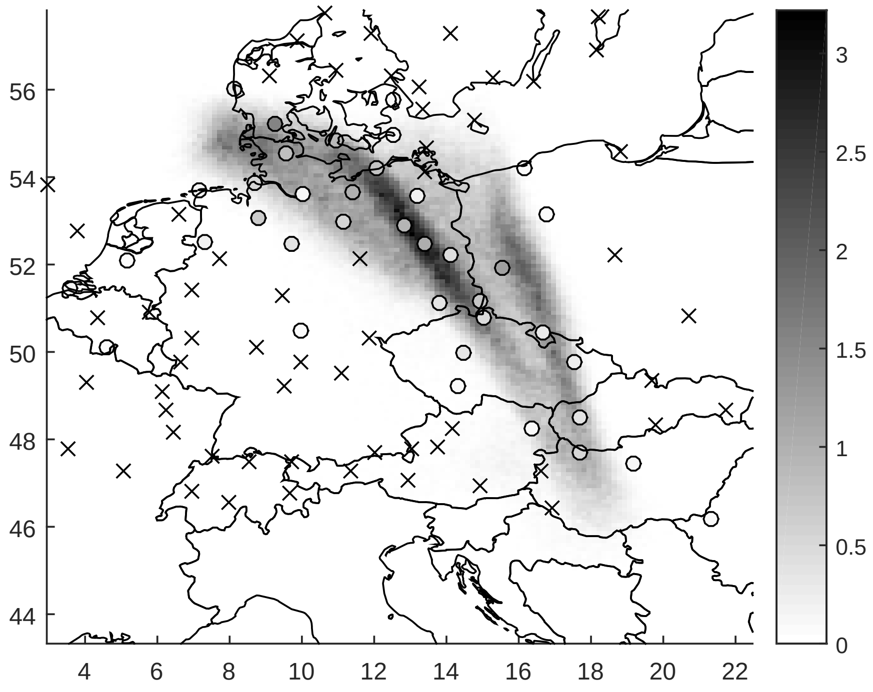
		GLM	SP model (ind.)	SP model (col.)
TG	up to 1000 ccm	3 805	9 318	5 305
TG	1000–1349 ccm	4 104	9 979	5 563
TG	1350–1849 ccm	4 918	11 704	6 296
TG	1850–2499 ccm	5 748	13 380	7 125
TG	over 2500 ccm	7 792	17 453	9 169
Region	Capital city	1.61	1.41	1.41
Region	Large towns	1.16	1.18	1.19
Region	Small towns	1.00	1.00	1.00
Region	Others	1.00	1.00	1.00
Age	18–30y	1.28	1.26	1.27
Age	31–65y	1.06	1.11	1.11
Age	over 66y	1.00	1.00	1.00
DL less that 5y	NO	1.00	1.00	1.00
DL more that 5y	YES	1.19	1.13	1.12

- **Goal** – profit maximization and risk minimization
- **Coal power plants** – demand seasonality, ...
- Stochastic linear programming (multistage/multiperiod)

1.7 Environment – Inverse modelling in atmosphere

- **Goal** – identification of the source and the amount released into the atmosphere
- **Standard approach** – dynamic Bayesian models
- **New approach** – Sparse optimization – Nonlinear/quadratic integer programming (weighted least squares with nonnegativity and sparsity constraints)
- **Applications**: nuclear power plants accidents, volcano accidents, nuclear tests, emission of pollutants ...
- **Project homepage:** http://stradi.utia.cas.cz/

Literature and detailed information: L. Adam, M.B. (2016).



2 Introduction to optimization

Repeat

- Cones
- Farkas theorem
- Convexity of sets and functions
- Symmetric Local Optimality Conditions (SLPO)

3 Convex sets and functions

Repeat the rules for estimating convexity of functions and sets: intersection of convex sets, function composition, level sets of convex functions, nonnegative combinations of convex function, first end second order derivatives, Hessian matrix, epigraph ...

For a function $f : \mathbb{R}^n \to \mathbb{R}^*$, we define its epigraph

$$\mathrm{epi}(f) = \left\{(x, \nu) \in \mathbb{R}^{n+1} : \; f(x) \leq \nu\right\}$$

Example 3.1. *Prove the equivalence between the possible definitions of convex functions $f : \mathbb{R}^n \to \mathbb{R}^*$:*

1. $\mathrm{epi}(f)$ *is a convex set*
2. $\mathrm{Dom}(f)$ *is a convex set and for all $x, y \in \mathrm{Dom}(f)$ and $\lambda \in (0, 1)$ we have*

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Example 3.2. *Decide if the following sets are convex:*

$$\begin{aligned}
M_1 &= \left\{(x,y)\in\mathbb{R}^2_+ :\ ye^{-x}-x\ge 1\right\}, && (1)\\
M_2 &= \left\{(x,y)\in\mathbb{R}^2 :\ x\ge 2+y^2\right\}, && (2)\\
M_3 &= \left\{(x,y)\in\mathbb{R}^2 :\ x^2+y\log y^4\le 139,\ y\ge 2\right\}, && (3)\\
M_4 &= \left\{(x,y)\in\mathbb{R}^2 :\ \log x+y^2\ge 1,\ x\ge 1,\ y\ge 0\right\}, && (4)\\
M_5 &= \left\{(x,y)\in\mathbb{R}^2 :\ (x^3+e^y)\log(x^3+e^y)\le 49,\ x\ge 0,\ y\ge 0\right\}, && (5)\\
M_6 &= \left\{(x,y)\in\mathbb{R}^2 :\ x\log x+xy\ge 0,\ x\ge 1\right\}, && (6)\\
M_7 &= \left\{(x,y)\in\mathbb{R}^2 :\ 1-xy\le 0,\ x\ge 0\right\}, && (7)\\
M_8 &= \left\{(x,y,z)\in\mathbb{R}^3 :\ \frac{1}{2}(x^2+y^2+z^2)+yz\le 1,\ x\ge 0,\ y\ge 0\right\}, && (8)\\
M_9 &= \left\{(x,y,z)\in\mathbb{R}^2 :\ 3x-2y+z=1\right\}. && (9)
\end{aligned}$$

Solution: M_2 is an epigraph of function $f(y)=2+y^2$, which is obviously convex.

Example 3.3. *Establish conditions under which the following sets are convex:*

$$\begin{aligned}
M_{10} &= \left\{x\in\mathbb{R}^n :\ \alpha\le a^Tx\le\beta\right\},\ \text{for some } a\in\mathbb{R}^n,\alpha,\beta\in\mathbb{R}, && (10)\\
M_{11} &= \left\{x\in\mathbb{R}^n :\ \left\|x-x^0\right\|\le\|x-y\|,\ \forall y\in S\right\},\ \text{for some } S\subseteq\mathbb{R}^n, && (11)\\
M_{12} &= \left\{x\in\mathbb{R}^n :\ x^Ty\le 1,\ \forall y\in S\right\},\ \text{for some } S\subseteq\mathbb{R}^n. && (12)
\end{aligned}$$

Example 3.4. *Verify if the following functions are convex:*

$$\begin{aligned}
f_1(x,y) &= x^2y^2+\frac{x}{y},\ x>0,y>0, && (13)\\
f_2(x,y) &= xy, && (14)\\
f_3(x,y) &= \log(e^x+e^y)-\log x,\ x>0, && (15)\\
f_4(x,y) &= \exp\{x^2+e^{-y}\},\ x>0,\ y>0, && (16)\\
f_5(x,y) &= -\log(x+y),\ x>0,\ y>0, && (17)\\
f_6(x,y) &= \sqrt{e^x+e^{-y}}, && (18)\\
f_7(x,y) &= x^3+2y^2+3x, && (19)\\
f_8(x,y) &= -\log(cx+dy),\ c,d\in\mathbb{R}, && (20)\\
f_9(x,y) &= \frac{x^2}{y},\ y>0, && (21)\\
f_{10}(x,y) &= xy\log xy,\ x>0,y>0, && (22)\\
f_{11}(x,y) &= |x+y|, && (23)\\
f_{12}(x) &= \sup_{y\in\mathrm{Dom}(f)}\{x^Ty-f(y)\}=f^*(x),\ f:\mathbb{R}^n\to\mathbb{R}, && (24)\\
f_{13}(x) &= \|Ax-b\|_2^2, && (25)
\end{aligned}$$

Solution: Consider $f_6(x,y)=\sqrt{e^x+e^{-y}}$. The first order partial derivatives are equal to

$$\begin{aligned}
\frac{\partial f_6}{\partial x}(x,y) &= \frac{e^x}{2\sqrt{e^x+e^{-y}}},\\
\frac{\partial f_6}{\partial y}(x,y) &= \frac{-e^{-y}}{2\sqrt{e^x+e^{-y}}},
\end{aligned}$$

and the second order derivatives are

$$\begin{aligned}
\frac{\partial^2 f_6}{\partial x^2}(x,y) &= \frac{e^{2x}+2e^{x-y}}{4(e^x+e^{-y})^{\frac{3}{2}}}, \\
\frac{\partial^2 f_6}{\partial y^2}(x,y) &= \frac{e^{-2y}+2e^{x-y}}{4(e^x+e^{-y})^{\frac{3}{2}}}, \\
\frac{\partial^2 f_6}{\partial y \partial x}(x,y) = \frac{\partial^2 f_6}{\partial x \partial y}(x,y) &= \frac{e^{x-y}}{4(e^x+e^{-y})^{\frac{3}{2}}}
\end{aligned}$$

To verify that the Hessian matrix is positive definite, it is sufficient to look on the numerators, because the common denominator $4(e^x+e^{-y})^{\frac{3}{2}}$ is always positive. Obviously $e^{2x}+2e^{x-y}$ is positive, thus it remains to verify that

$$(e^{2x}+2e^{x-y})(e^{-2y}+2e^{x-y})-(e^{x-y})^2>0.$$

Example 3.5. *Let $f(x,y)$ be a convex function and C is a convex set. Then*

$$g(x) \quad = \quad \inf_{y\in C} f(x,y). \tag{26}$$

is convex.

Example 3.6. *(Vector composition) Let $g_i:\mathbb{R}^n\to\mathbb{R}$, $i=1,\dots,k$ and $h:\mathbb{R}^k\to\mathbb{R}$ be convex functions. Moreover let h be nondecreasing in each argument. Then*

$$f(x)=h\big(g_1(x),\dots,g_k(x)\big). \tag{27}$$

is convex.
Apply to

$$f(x)=\log\left(\sum_{i=1}^k e^{g_i(x)}\right),$$

where g_i are convex.
Hint: the first part can be verified using the definition of convexity, in the second part compute the Hessian matrix $\mathcal{H}(x)$ and use the Cauchy-Schwarz inequality $(a^Ta)(b^Tb)\geq(a^Tb)^2$ to verify that $v^T\mathcal{H}(x)v\geq 0$ for all $v\in\mathbb{R}^k$.

Solution: Consider the function

$$h(z)=\log\left(\sum_{i=1}^k e^{z_i}\right).$$

Obviously it is nondecreasing in each argument. We can show that it is also convex. We compute its second order partial derivatives

$$\begin{aligned}
\frac{\partial^2 h}{\partial z_j^2}(z) &= \frac{e^{z_j}(\sum_{i=1}^k e^{z_i})-e^{z_j}e^{z_j}}{(\sum_{i=1}^k e^{z_i})^2},\ j=1,\dots,k, \\
\frac{\partial^2 h}{\partial z_j \partial z_l}(z) &= \frac{-e^{z_j}e^{z_l}}{(\sum_{i=1}^k e^{z_i})^2},\ j\neq l.
\end{aligned}$$

If we use the notation $y = (e^{z_1}, \dots, e^{z_k})^T$ and $\mathbb{I} = (1, \dots, 1)^T \in \mathbb{R}^k$, we can write the Hessian matrix in the form

$$\mathcal{H}_h(y) = \frac{1}{(\mathbb{I}^T y)^2} \left(\mathrm{diag}(y)(\mathbb{I}^T y) - yy^T \right),$$

where $\mathbb{I}^T y = \sum_{i=1}^k y_i = \sum_{i=1}^k e^{z_i}$ and $\mathrm{diag}(y)$ denotes the diagonal matrix with elements y. We would like to verify that $v^T \mathcal{H}_h(y) v \geq 0$ for arbitrary $v \in \mathbb{R}^k$. We can compute

$$v^T \mathcal{H}_h(y) v = \frac{(\sum_{i=1}^k y_i v_i^2)(\sum_{i=1}^k y_i) - (\sum_{i=1}^k y_i v_i)^2}{(\sum_{i=1}^k y_i)^2}.$$

By setting $a_i = \sqrt{y_i} v_i$ and $b_i = \sqrt{y_i}$ and using the Cauchy-Schwarz inequality in the form $(a^T a)(b^T b) - (a^T b)^2 \geq 0$, we can obtain that the numerator is nonnegative, i.e. $v^T \mathcal{H}_h(y) v \geq 0$.

Example 3.7. *Verify that the geometric mean is concave:*

$$f(x) \quad = \quad \left(\prod_{i=1}^n x_i \right)^{1/n}, \ x \in (0, \infty)^n. \tag{28}$$

Hint: compute the Hessian matrix and use the Cauchy-Schwarz inequality $(a^T a)(b^T b) \geq (a^T b)^2$.

4 Separating hyperplane theorems

Remind: theorem about projection of a point to a convex set (obtuse angle), separation of a point and a convex set, proper and strict separability.

Using the theorem about the separation of a point and a convex set, prove the following lemma about the existence of a supporting hyperplane.

Lemma 4.1. *Let $\emptyset \neq K \subset \mathbb{R}^n$ be a convex set and $x \in \partial K$. Then, there is $\gamma \in \mathbb{R}^n$, $\gamma \neq 0$ such that*

$$\inf\{\langle \gamma, y \rangle : \ y \in K\} \geq \langle \gamma, x \rangle .$$

Hint: separate a sequence $x_n \notin K$ which converge to the point x on the boundary, show the convergence of separating hyperplanes characterized by $\gamma_n \neq 0$.

Example 4.2. *Find a separating or supporting hyperplane for the following sets and points:*

$$\begin{array}{ll} x_1 = (-1,-1) & K_1 = \{(x,y);\ x \geq 0, y \geq 0\}, \\ x_2 = (3,1) & K_2 = \{(x,y);\ x^2 + y^2 < 10\}, \\ x_3 = (3,0,0) & K_3 = \{(x,y,z);\ x^2 + y^2 + z^2 \leq 9\}, \\ x_4 = (0,2,0) & K_4 = \{(x,y,z);\ x + y + z \leq 1\}. \end{array}$$

Hint: Use pictures and realize that γ is the normal vector of the separating/supporting hyperplane. For x_1, K_1 we can use $\gamma = (1,1)$, then

$$\min_{(x,y) \in K} x + y = 0 > -1 - 1 = -2.$$

Note that other choices are also possible, in particular $(\gamma_1, \gamma_2) \neq 0$ with $\gamma_1, \gamma_2 \geq 0$.

Example 4.3. *Let $K \subseteq \mathbb{R}^n$, $K \neq \emptyset$. Show that K is a closed convex set if and only if it is an intersection of all closed half-spaces which contain K.*

Hint: Show that if $y \notin K$, then it is not contained in the intersection using the theorem about separation of a point and a convex set.

Example 4.4. *Provide a description of the circle in $\mathbb{R}^2$ and ball in $\mathbb{R}^3$ as a intersection of supporting halfspaces.*

Prove the following theorem which gives a sufficient condition for proper separability of two convex sets.

Theorem 4.5. *Let $A, B \subset \mathbb{R}^n$ be non-empty convex sets. If $\mathrm{rint}(A) \cap \mathrm{rint}(B) = \emptyset$, then A and B can be properly separated.*

Hint: Separate set $K = A - B$ and point 0. First, show that $0 \notin \mathrm{rint}K$.

Example 4.6. *Verify whether the following pairs of (convex ?) sets are properly or strictly separable or not. If they are separable, suggest a possible value of γ.*

$$\begin{array}{ll} A_1 = \{(x,y);\ y \geq |x|\}, & B_1 = \{(x,y);\ 2y + x \leq 0\}, \\ A_2 = \{(x,y);\ xy \geq 1, x > 0\}, & B_2 = \{(x,y;\ x \leq 0, y \leq 0\}, \\ A_3 = \{(x,z);\ x + y + z \leq 1\}, & B_3 = \{(x,y,z);\ (x-2)^2 + (y-2)^2 + (z-2)^2 \leq 3\}, \\ A_4 = \{(x,y,z);\ 0 \leq x,y,z \leq 1\}, & B_4 = \{(x,y,z);\ (x-2)^2 + (y-2)^2 + (z-2)^2 \leq 3\}. \end{array}$$

Hint: Use pictures. Sets A_1, B_2 are properly separable using $\gamma_1 = (1, 2)$:

$$\min_{(x,y) \in A_1} x + 2y = 0 \geq 0 = \max_{(x,y) \in B_1} x + 2y.$$

Example 4.7. *Discuss the proof of the Farkas theorem.*

Hint: Use an alternative formulation of the FT.

5 Subdifferentiability and subgradient

From Introduction to optimization (or similar course), you should remember the following property which holds for any differentiable convex function $f : X \to \mathbb{R}$:

$$\forall x, y \in X\ f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle .$$

This property can be generalized by the notation of subdifferentiability. Any subgradient $a \in \mathbb{R}^n$ of function f at $x \in X$ fulfills

$$f(y) - f(x) \geq \langle a, y - x \rangle\ \ \forall y \in X.$$

Set of all subgradients at x is called subdifferential of f at x and denoted by $\partial f(x)$.

Optimality condition

$$0 \in \partial f(x^*)$$

is necessary and sufficient for $x^* \in X$ being a global minimum.

Example 5.1. *Consider (do not necessiraly prove, rather think about) the following properties of subgradient:*

1. *a is subgradient of f at x if and only if $(a, -1)$ supports $\mathrm{epi}(f)$ at $(x, f(x))$.*
2. *if f is convex, then $\partial f(x) \neq \emptyset$ for all $x \in \mathrm{rint\ dom} f$.*
3. *if f is convex and differentiable, then $\partial f(x) = \{\nabla f(x)\}$.*
4. *if $\partial f(x) = \{g\}$ (is singleton), then $g = \nabla f(x)$.*
5. *$\partial f(x)$ is a closed convex set.*

Hint: 1. Apply the definition of the supporting hyperplane to an epigraph, i.e. use $\gamma = (a, -1)$:

$$\max_{(y,z)\in \mathrm{epi}(f)} a^T y - z \leq a^T x - f(x).$$

Now, realize that $(y, f(y)) \in \mathrm{epi}(f)$ and $f(y)$ is the smallest value of z leading to

$$\forall y \in \mathrm{dom}(f)\ a^T y - f(y) \leq a^T x - f(x).$$

Finally, it is sufficient to reorganize the formula to get the definition of subgradient

$$\forall y \in \mathrm{dom}(f)\ f(y) - f(x) \geq a^T y - a^T x.$$

Example 5.2. *Derive the subdifferential for the following functions:*

$$\begin{array}{rcll} f_1(x) & = & |x| & \\ f_2(x) & = & x^2 & \text{if } x \leq -1, \\ & & -x & \text{if } x \in [-1, 0], \\ & & x^2 & \text{if } x \geq 0, \\ f_3(x, y) & = & |x + y| & \end{array}$$

Hint: Use pictures and the definition.

Lemma 5.3. *Let $f_1, \ldots, f_k$ be convex functions and let*

$$f(x) = f_1(x) + \cdots + f_k(x).$$

Then

$$\partial f_1(x) + \cdots + \partial f_k(x) \subseteq \partial f(x).$$

Hint: Use the definition.

6 Generalizations of convex functions

6.1 Quasiconvex functions

Definition 6.1. *We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is quasiconvex, if all its level sets are convex.*

Example 6.2. *Find several examples of functions which are quasiconvex, but they are not convex. Try to find an example of function which is not continous on the interior of its domain (thus it cannot be convex).*

Example 6.3. *Show that the following property is equivalent to the definition of quasiconvexity*

$$f(\lambda x + (1-\lambda)y) \le \max\{f(x), f(y)\}$$

for all x, y and $\lambda \in [0,1]$.

Example 6.4. *Verify that the following functions are quasiconvex on given sets:*

$$f(x,y) = xy \text{ for } (x,y) \in \mathbb{R}_+ \times \mathbb{R}_-,$$

$$f(x) = \frac{a^T x + b}{c^T x + d} \text{ for } c^T x + d > 0.$$

Hint: Use the definition.

Lemma 6.5. *Continuous function $f : \mathbb{R} \to \mathbb{R}$ is quasiconvex if and only if one of the following conditions holds*

- *f is nondecreasing,*
- *f is nonincreasing,*
- *there is a $c \in \mathbb{R}$ such that f is nonincreasing on $(-\infty, c]$ and nondecreasing on $[c, \infty)$.*

Hint: Realize that the level sets are intervals.

Example 6.6. *Let f be differentiable. Show that f is quasiconvex if and only if it holds*

$$f(y) \le f(x) \implies \nabla f(x)^T (y - x) \le 0.$$

Example 6.7. *Let f be a differentiable quasiconvex function. Show that the condition*

$$\nabla f(\overline{x}) = 0$$

implies that $\overline{x}$ is a local minimum of f.

Hint: Consider the previous lemma.

Example 6.8. *Let f_1, f_2 be quasiconvex functions, g be a nondecreasing function and $t \geq 0$ be a scalar. Prove that the following operations preserve quasiconvexity*

- tf_1,
- $\max\{f_1, f_2\}$,
- $g \circ f_1$.

Example 6.9. *Let f_1, f_2 be quasiconvex functions. Find counterexamples that the following operations DO NOT preserve quasiconvexity:*

- $f_1 + f_2$,
- $f_1 f_2$.

Example 6.10. *Verify that the following functions are quasiconvex on given sets:*

$$\begin{aligned}
f_1(x,y) &= \frac{1}{xy} \text{ on } \mathbb{R}^2_{++},\\
f_2(x,y) &= \frac{x}{y} \text{ on } \mathbb{R}^2_{++},\\
f_3(x,y) &= \frac{x^2}{y} \text{ on } \mathbb{R} \times \mathbb{R}_{++},\\
f_4(x,y) &= \sqrt{|x+y|} \text{ on } \mathbb{R}^2.
\end{aligned}$$

Hint: 1-3. Use the definition. 4. Use the above rules.

Example 6.11. *Let S be a nonempty convex subset of $\mathbb{R}^n$, $g : S \to \mathbb{R}_+$ be convex and $h : S \to (0, \infty)$ be concave. Show that the function defined by*

$$f(x) = \frac{g(x)}{h(x)}$$

is quasiconvex on S.

Hint: Use the definition based on the level sets.

Definition 6.12. *We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly quasiconvex if*

$$f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\}$$

for all x, y with $f(x) \neq f(y)$ and $\lambda \in (0,1)$.

Lemma 6.13. *Let f be strictly quasiconvex and S be a convex set. Then any local minimum $\overline{x}$ of $\min_{x \in S} f(x)$ is also a global minimum.*

6.2 Pseudoconvex functions

Definition 6.14. *Consider $S \subset \mathbb{R}^n$ a nonempty open set. We say that differentiable function $f : S \to \mathbb{R}$ is pseudoconvex if it holds*

$$\nabla f(x)^T(y-x) \geq 0 \implies f(y) \geq f(x)$$

for all $x, y \in S$.

Example 6.15. *Find a pseudoconvex function which is not convex.*

Hint: Consider increasing functions.

Example 6.16. *Use the definition to show that the following fractional linear function is pseudoconvex:*

$$f(x) = \frac{a^T x + b}{c^T x + d} \text{ for } c^T x + d > 0.$$

Hint: Use the definition.

Example 6.17. *Consider function f as defined in Example 6.11. Moreover, let S be open and g, h be differentiable on S. Show that f is pseudoconvex.*

Example 6.18. *Let f be a differentiable function. Show that if f is convex, then it is also pseudoconvex.*

Hint: Use the first order characterization of the differentiable convex functions.

Example 6.19. *Let f be a differentiable function. Show that if f is pseudoconvex, than it is also quasiconvex.*

Hint: Use the alternative definition of quasiconvex functions based on the maximum.

Example 6.20. *Show that if $\nabla f(\overline{x}) = 0$ for a pseudoconvex f, then $\overline{x}$ is a global minimum of f.*

Hint: Use the definition.

Example 6.21. *The following table summarizes relations between the stationary points and minima of a differentiable function f:*

f general:	*$\overline{x}$ global min.*	$\Longrightarrow$	*$\overline{x}$ local min.*	$\Longrightarrow$	$\nabla f(\overline{x}) = 0$
f quasiconvex:	*$\overline{x}$ global min.*	$\Longrightarrow$	*$\overline{x}$ local min.*	$\Longrightarrow$	$\nabla f(\overline{x}) = 0$
f strictly quasiconvex:	*$\overline{x}$ global min.*	$\Longleftrightarrow$	*$\overline{x}$ local min.*	$\Longrightarrow$	$\nabla f(\overline{x}) = 0$
f pseudoconvex:	*$\overline{x}$ global min.*	$\Longleftrightarrow$	*$\overline{x}$ local min.*	$\Longleftrightarrow$	$\nabla f(\overline{x}) = 0$
f convex:	*$\overline{x}$ global min.*	$\Longleftrightarrow$	*$\overline{x}$ local min.*	$\Longleftrightarrow$	$\nabla f(\overline{x}) = 0.$

7 Optimality conditions

7.1 Optimality conditions based on directions

Example 7.1. *Consider the global optimization problem*

$$\min 2x_1^2 - x_1x_2 + x_2^2 - 3x_1 + e^{2x_1+x_2}.$$

Find a descent direction at point (0,0).

Hint: Compute the gradient.

Example 7.2. *Verify the optimality conditions at point* $(2, 4)$ *for problem*

$$\begin{aligned} \min\ & (x_1 - 4)^2 + (x_2 - 6)^2 \\ \text{s.t. } & x_1^2 \le x_2, \\ & x_2 \le 4. \end{aligned}$$

Consider the same point for the problem with the second inequality constraint in the form

$$x_2 \le 5.$$

Hint: Use the basic optimality conditions derived for a convex objective function and a convex set of feasible solutions.

Example 7.3. *Consider open* $\emptyset \neq S \subseteq \mathbb{R}^n$, $f : S \to \mathbb{R}$, *and define set of improving directions of* f *at* $x \in S$

$$F_f(x) = \{s \in \mathbb{R}^n :\ s \neq 0,\ \exists \delta > 0\ \forall 0 < \lambda < \delta :\ f(x + \lambda s) < f(x)\}.$$

For differentiable f, *define its approximation*

$$F_{f,0}(x) = \{s \in \mathbb{R}^n :\ \langle \nabla f(x), s\rangle < 0\}.$$

Show that it holds

$$F_{f,0}(x) \subseteq F_f(x).$$

Moreover, if f *is pseudoconvex at* x *with respect to a neighborhood of* x, *then*

$$F_{f,0}(x) = F_f(x).$$

If f *is convex, then*

$$F_f(x) = \{\alpha(y - x) :\ \alpha > 0,\ f(y) < f(x),\ y \in S\}.$$

Hint: Use the scalarization function.

Example 7.4. *Consider the global optimization problem*

$$\min 2x_1^2 - 3x_1x_2^2 + x_2^4.$$

Derive the set of improving directions at (0,0).

Example 7.5. *Consider open* $\emptyset \neq S \subseteq \mathbb{R}^n$*, functions* $g_i : S \to \mathbb{R}$*, and the set of feasible solutions*

$$M = \{x \in S : \ g_i(x) \leq 0, \ i = 1, \dots, m\}.$$

Define the set of feasible directions of M *at* x

$$D_M(x) = \{s \in \mathbb{R}^n : \ s \neq 0, \ \exists \delta > 0 \ \forall 0 < \lambda < \delta : \ x + \lambda s \in M\}.$$

If M *is a convex set, then*

$$D_M(x) = \{\alpha(y - x) : \ \alpha > 0, \ y \in M, \ y \neq x\}.$$

For differentiable g_i *define*

$$\begin{aligned} G_{g,0}(x) &= \{s \in \mathbb{R}^n : \ \langle \nabla g_i(x), s \rangle < 0, \ i \in I_g(x)\}, \\ G'_{g,0}(x) &= \{s \in \mathbb{R}^n : \ s \neq 0, \ \langle \nabla g_i(x), s \rangle \leq 0, \ i \in I_g(x)\}. \end{aligned}$$

In general, it holds

$$G_{g,0}(x) \subseteq D_M(x) \subseteq G'_{g,0}(x).$$

Hint: Use the scalarization function.

Example 7.6. *Discuss the above defined sets of directions for the sets*

$$\begin{aligned} M_1 &= \{(x, y) : \ -(x-2)^2 \geq y - 2, \ -(y-2)^2 \geq x - 2\}, \\ M_2 &= \{(x, y) : \ (x-2)^2 \geq y - 2, \ (y-2)^2 \geq x - 2\}, \end{aligned}$$

at point (2,2).

Hint: Use pictures to decide which of the approximations are tight.

Example 7.7. *Discuss the above defined sets of directions for a polyhedral set*

$$M = \{x \in \mathbb{R}^n : \ Ax \leq b\}.$$

Example 7.8. *Discuss the above defined sets of directions for the problem*

$$\begin{aligned} \min \ & (x_1 - 3)^2 + (x_2 - 2)^2 \\ \text{s.t.} \ & x_1^2 + x_2^2 \leq 5, \\ & x_1 + x_2 \leq 3, \\ & x_1 \geq 0, \ x_2 \geq 0, \end{aligned}$$

at point (2,1). Apply the Farkas theorem to the conditions on directions.

Example 7.9. *Discuss the above defined sets of directions for the problem*

$$\begin{aligned} \min \ & (x_1 - 3)^2 + (x_2 - 3)^2 \\ \text{s.t.} \ & x_1^2 + x_2^2 = 4, \end{aligned}$$

at point $(\sqrt{2}, \sqrt{2})$*. Then consider the set of improving directions for equality constraints* $h_j(x) = 0$*, where* $h_j : S \to \mathbb{R}$

$$H_{h,0}(x) = \{s \in \mathbb{R}^n : \ \langle \nabla h_j(x), s \rangle = 0\}.$$

7.2 Karush–Kuhn–Tucker optimality conditions

7.2.1 A few pieces of the theory

Consider a **nonlinear programming problem** with inequality and equality constraints:

$$\begin{aligned} \min\ & f(x) \\ \text{s.t. } & g_i(x) \le 0,\ i = 1, \dots, m, \\ & h_j(x) = 0,\ j = 1, \dots, l, \end{aligned} \tag{29}$$

where $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ are differentiable functions. We denote by M the set of feasible solutions.

Define the **Lagrange function** by

$$L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{j=1}^{l} v_j h_j(x),\ u_i \ge 0. \tag{30}$$

The **Karush–Kuhn–Tucker optimality conditions** are then

$$\begin{aligned} & \nabla_x L(x, u, v) = 0, \\ & u_i g_i(x) = 0,\ u_i \ge 0,\ i = 1, \dots, m. \end{aligned} \tag{31}$$

Any point (x, u, v) which fulfills the above conditions is called a KKT point. The KKT point is feasible if $x \in M$.

If a Constraint Qualification (CQ) condition is fulfilled, then the KKT conditions are necessary for local optimality of a point. Basic CQ conditions are:

- **Slater CQ:** $\exists \tilde{x} \in M$ such that $g_i(\tilde{x}) < 0$ for all i and the gradients $\nabla_x h_j(\tilde{x})$, $j = 1, \dots, l$ are linearly independent.
- **Linear independence CQ** at $\hat{x} \in M$: all gradients

  $$\nabla_x g_i(\hat{x}),\ i \in I_g(\hat{x}),\ \nabla_x h_j(\hat{x}),\ j = 1, \dots, l$$

  are linearly independent.

These conditions are quite strong and are sufficient for weaker CQ conditions, e.g. the Kuhn–Tucker condition (Mangasarian–Fromovitz CQ, Abadie CQ, ...).

Consider the set of active (inequality) constraints and its partitioning

$$\begin{aligned} I_g(x) &= \{i :\ g_i(x) = 0\}, \\ I_g^0(x) &= \{i :\ g_i(x) = 0,\ u_i = 0\}, \\ I_g^+(x) &= \{i :\ g_i(x) = 0,\ u_i > 0\}, \end{aligned} \tag{32}$$

i.e.

$$I_g(x) = I_g^0(x) \cup I_g^+(x).$$

We say that the **second-order sufficient condition** (SOSC) is fulfilled at a feasible KKT point (x, u, v) if for all $0 \neq z \in \mathbb{R}^n$ such that

$$\begin{aligned} z^T \nabla_x g_i(x) &= 0,\ i \in I_g^+(x), \\ z^T \nabla_x g_i(x) &\leq 0,\ i \in I_g^0(x), \\ z^T \nabla_x h_j(x) &= 0,\ j = 1, \dots, l, \end{aligned} \tag{33}$$

it holds

$$z^T\, \nabla^2_{xx} L(x, u, v)\, z > 0. \tag{34}$$

Then x is a strict local minimum of the nonlinear programming problem (29).

To summarize, we are going to practice the following relations:

1. Feasible KKT point and convex problem $\rightarrow$ global optimality at x.
2. Feasible KKT point and SOSC $\rightarrow$ (strict) local optimality at x.
3. Local optimality at x and a constraint qualification (CQ) condition $\rightarrow \exists (u, v)$ such that (x, u, v) is a KKT point.

7.2.2 Karush–Kuhn–Tucker optimality conditions

Example 7.10. *Consider the nonlinear programming problems from examples 7.2, 7.8. Compute the Lagrange multipliers at given points.*

Example 7.11. *Consider the problem*

$$\begin{aligned} \min\ & 2e^{x_1 - 1} + (x_2 - x_1)^2 + x_3^2 \\ \text{s.t. }\ & x_1 x_2 x_3 \leq 1, \\ & x_1 + x_3 \geq c, \\ & x \geq 0. \end{aligned}$$

For which values of c does $\bar{x} = (1, 1, 1)$ fulfill the KKT conditions? Is it a global solution?

Example 7.12. *Consider the problem*

$$\begin{aligned} \min\ & \frac{x_1 + 3x_2 + 3}{2x_1 + x_2 + 6} \\ \text{s.t. }\ & 2x_1 + x_2 \leq 12, \\ & -x_1 + 2x_2 \leq 4, \\ & x_1, x_2 \geq 0. \end{aligned}$$

Verify that the KKT conditions are fulfilled for all points on the line between (0,0) and (6,0). Are the KKT conditions sufficient for global optimality?

Example 7.13. *Consider the problem*

$$\begin{aligned} \min \ & -\sum_{i=1}^{n} \log(\alpha_i + x_i) \\ \text{s.t.} \ & \sum_{i=1}^{n} x_i = 1 \\ & x_i \geq 0, \end{aligned}$$

where $\alpha_i > 0$ are parameters. Using the KKT conditions find the solutions.

Example 7.14. *Let $n \geq 2$. Consider the problem*

$$\begin{aligned} \min \ & x_1 \\ \text{s.t.} \ & \sum_{i=1}^{n} \left(x_i - \frac{1}{n} \right)^2 \leq \frac{1}{n(n-1)}, \\ & \sum_{i=1}^{n} x_i = 1. \end{aligned}$$

Show that

$$\left(0, \frac{1}{n-1}, \ldots, \frac{1}{n-1} \right)$$

is an optimal solution.

Solution: First, realize that the considered point is feasible. Write the Lagrange function

$$L(x_1, \ldots, x_n, u, v) = x_1 + u \left(\sum_{i=1}^{n} \left(x_i - \frac{1}{n} \right)^2 - \frac{1}{n(n-1)} \right) + v \left(\sum_{i=1}^{n} x_i - 1 \right),$$

where $u \geq 0$ and $v \in \mathbb{R}$. The KKT conditions (optimality and complementarity) are

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= 1 + 2u \left(x_1 - \frac{1}{n} \right) + v = 0, \\ \frac{\partial L}{\partial x_i} &= 2u \left(x_1 - \frac{1}{n} \right) + v = 0, \ i \neq 1, \\ u & \left(\sum_{i=1}^{n} \left(x_i - \frac{1}{n} \right)^2 - \frac{1}{n(n-1)} \right) = 0. \end{aligned} \tag{35}$$

Realize that the inequality constraint is active at the considered point, i.e.

$$\left(0 - \frac{1}{n} \right)^2 + \sum_{i=2}^{n} \left(\frac{1}{n-1} - \frac{1}{n} \right)^2 = \frac{1}{n(n-1)}.$$

To obtain the values of Lagrange multipliers, we solve the optimality conditions

$$\begin{aligned} & 1 - \frac{2u}{n} + v = 0, \\ & 2u \left(\frac{1}{n-1} - \frac{1}{n} \right) + v = 0, \ (\forall i \neq 1). \end{aligned} \tag{36}$$

By solving this linear system for u and v, we obtain the values

$$\begin{aligned} u &= \frac{n-1}{2} \geq 0, \\ v &= \frac{-1}{n} \in \mathbb{R}. \end{aligned} \tag{37}$$

Thus, we have obtained a KKT point

$$(x, u, v) = \left(0, \frac{1}{n-1}, \ldots, \frac{1}{n-1}, \frac{n-1}{2}, \frac{-1}{n}\right),$$

Since the objective function is convex (linear), the inequality constraint is convex and the equality constraint is linear, the considered point is a global solution (minimum) of the problem.

Example 7.15. *Using the KKT conditions find the closest point to (0,0) in the set defined by*

$$M = \{x \in \mathbb{R}^2 : \; x_1 + x_2 \geq 4, \; 2x_1 + x_2 \geq 5\}.$$

Can several points (solutions) exist?

Hint: Formulate a nonlinear programming problem.

Example 7.16. *Consider the problem*

$$\begin{aligned} \min \; & \sum_{j=1}^{n} \frac{c_j}{x_j} \\ \text{s.t.} \; & \sum_{j=1}^{n} a_j x_j = b, \\ & x_j \geq \varepsilon, \end{aligned}$$

where $a_j, b, c_j, \varepsilon > 0$ *are parameters. Using the KKT conditions find an optimal solution.*

Example 7.17. *Consider the problem*

$$\begin{aligned} \min \; & x \\ s.t. \; & (x-1)^2 + (y-1)^2 \leq 1 \\ & (x-1)^2 + (y+1)^2 \leq 1. \end{aligned}$$

The optimal solution is obviously the only feasible point $(1, 0)$. *Why are not the KKT conditions fulfilled?*

Hint: Discuss the Constraint Qualification conditions.

Example 7.18. *Consider the problem*

$$\begin{aligned} \min\ & -x_1 \\ s.t.\ & -(1-x_1)^3 + x_2 \le 0, \\ & x_2 \ge 0. \end{aligned}$$

Use the picture to show that $(1,0)$ *is the global optimal solution. Why are not the KKT conditions fulfilled?*

Hint: Discuss the Constraint Qualification conditions.

Example 7.19. *Write the KKT conditions for a linear programming problem.*

Example 7.20. *Verify that the point* $(x, y) = (\frac{4}{5}, \frac{8}{5})$ *is a local/global solution of the problem*

$$\begin{aligned} \min\ & x^2 + y^2, \\ s.t.\ & x^2 + y^2 \le 5, \\ & x + 2y = 4, \\ & x, y \ge 0. \end{aligned}$$

Example 7.21. *Derive the least square estimate for coefficients in the linear regression model under linear constraints, i.e. solve the problem*

$$\begin{aligned} \min_{\beta}\ & \|Y - X\beta\|^2, \\ \text{s.t. } & A\beta = b. \end{aligned}$$

7.2.3 Second Order Sufficient Condition (SOSC)

When the problem is not convex, then the solutions of the KKT conditions need not to correspond to global optima. The Second Order Sufficient Condition (SOSC) can be used to verify if the KKT point (its x part) is at least a local minimum.

Example 7.22. *Consider the problem*

$$\begin{aligned} \min\ & x^2 - y^2 \\ s.t.\ & x - y = 1 \\ & x, y \ge 0. \end{aligned}$$

Using the KKT optimality conditions find all stationary points. Using the SOSC verify if some of the points corresponds to a (strict) local minimum.

Solution: Write the Lagrange function

$$L(x, y, u_1, u_2, v) = x^2 - y^2 - u_1 x - u_2 y + v(x - y - 1),\ u_1, u_2 \ge 0.$$

Derive the KKT conditions

$$
\begin{aligned}
&\frac{\partial L}{\partial x} = 2x - u_1 + v = 0,\\
&\frac{\partial L}{\partial y} = -2y - u_2 - v = 0,\\
&-u_1 x = 0,\\
&-u_2 y = 0.
\end{aligned} \tag{38}
$$

Solving this conditions together with feasibility leads to one feasible KKT point

$$(x, y, u_1, u_2, v) = (1, 0, 0, 2, -2).$$

Since the problem is non-convex, we can apply SOSC (33), (34). We have $I_g(1,0) = I_g^+(1,0) = \{2\}$ and $I_g^0(1,0) = \emptyset$, so the conditions on $0 \neq z \in \mathbb{R}^2$ are:

$$
\begin{aligned}
z_1 - z_2 &= 0,\\
-z_2 &= 0.
\end{aligned}
$$

Since no $z \neq 0$ exists, the SOSC is fulfilled. (It is not necessary to compute $\nabla^2_{xx} L$.)

Example 7.23. *Consider the problem*

$$
\begin{aligned}
\min\ & -x\\
s.t.\ & x^2 + y^2 \leq 1\\
& (x-1)^3 - y \leq 0.
\end{aligned}
$$

Using the KKT optimality conditions find all stationary points. Using the SOSC verify if some of the points corresponds to a (strict) local minimum.

Example 7.24. *Consider the problem*

$$
\begin{aligned}
\min\ & -(x-2)^2 - (y-3)^2\\
s.t.\ & 3x + 2y \geq 6,\\
& -x + y \leq 3,\\
& x \leq 2.
\end{aligned}
$$

Using the KKT optimality conditions find all stationary points. Using the SOSC verify if some of the points corresponds to a (strict) local minimum.

Literature

- L. Adam, M. Branda (2016). **Sparse optimization for inverse problems in atmospheric modelling.** Environmental Modelling & Software 79, 256–266. (free Matlab codes available)

- Bazaraa, M.S., Sherali, H.D., and Shetty, C.M. (2006). **Nonlinear programming: theory and algorithms**, Wiley, Singapore, 3rd edition.
- Boyd, S., Vandenberghe, L. (2004). **Convex optimization**, Cambridge University Press, Cambridge.
- M. Branda (2012). **Underwriting risk control in non-life insurance via generalized linear models and stochastic programming**. Proceedings of the 30th International Conference on MME 2012, 61–66.
- M. Branda (2014). **Optimization approaches to multiplicative tariff of rates estimation in non-life insurance**. Asia-Pacific Journal of Operational Research 31(5), 1450032, 17 pages, 2014.
- M. Branda, S. Hájek (2017). **Flow-based formulations for operational fixed interval scheduling problems with random delays**. Computational Management Science 14 (1), 161–177.
- M. Branda, J. Novotný, A. Olstad (2016). **Fixed interval scheduling under uncertainty - a tabu search algorithm for an extended robust coloring formulation**. Computers & Industrial Engineering 93, 45–54.
- M. Branda, K. Haugen, J. Novotný, A. Olstad (2017). **Downstream logistics optimization at EWOS Norway**. Mathematics for Applications 6 (2), 127-141.
- R.T. Rockafellar (1972). **Convex analysis**, Princeton University Press, New Jersey.