Introduction to Optimization

Outline:

- Standard form optimization problem and terminology.
- Convex optimization problems.
- Lagrange duality.
- Optimization methods.

Optimization problem in standard form

$$
\begin{array}{lrcll}
\text{minimize} & f_0(\boldsymbol{x}) & & & \\
\text{subject to} & f_i(\boldsymbol{x}) & \leq & 0\,, & i = 1, \ldots, m \\
 & h_i(\boldsymbol{x}) & = & 0\,, & i = 1, \ldots, p
\end{array}
$$

where

$\boldsymbol{x} \in \mathbb{R}^n$ is the **optimized vector** of variables.

$f_0 \colon \mathbb{R}^n \to \mathbb{R}$ is the **objective function**.

$f_i \colon \mathbb{R}^n \to \mathbb{R}$ is the **inequality constraint** function.

$h_i \colon \mathbb{R}^n \to \mathbb{R}$ is the **equality constraint** function.

Explicit constraints are $f_i(\boldsymbol{x}) \le 0$ and $h_i(\boldsymbol{x}) = 0$; **unconstrained problem** has no explicit constrains (i.e. $m = p = 0$).

Implicit constraint is $\boldsymbol{x} \in \mathcal{D}$ where $\mathcal{D}$ is a common domain of the objective function and constraint functions

$$\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i\ .$$

Feasible set: contains points which satisfy implicit and explicit constraints

$$\mathcal{X}_{\text{feas}} = \mathcal{D} \cap \{\boldsymbol{x} \mid f_i(\boldsymbol{x}) \le 0\,, i = 1, \ldots, m\,, h_j(\boldsymbol{x}) = 0\,, j = 1\ \ldots, n\}$$

Example: (minimal entropy discrete distribution)

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} x_i \log x_i \\ \text{subject to} & \sum_{i=1}^{n} x_i = 1\,. \end{array}$$

which has explicit constraint $\sum_{i=1}^{n} x_i = 1$, *implicit constraints* $x_i > 0$ *and feasible set* $\mathcal{X}_{\text{feas}} = \{\boldsymbol{x} \mid \sum_{i=1}^{n} x_i = 1\,, x_i > 0, i = 1, \ldots .n\}$.

LP problem

$$\begin{array}{lrcl} \text{minimize} & & \boldsymbol{c}^T\boldsymbol{x} & \\ \text{subject to} & \mathbf{A}\boldsymbol{x} & = & \boldsymbol{b} \\ & \mathbf{D}\boldsymbol{x} & \leq & \boldsymbol{q} \end{array}$$

QP problem

$$\begin{array}{lrcl} \text{minimize} & \multicolumn{3}{l}{\frac{1}{2}\boldsymbol{x}^T\mathbf{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x}} \\ \text{subject to} & \mathbf{A}\boldsymbol{x} & = & \boldsymbol{b} \\ & \mathbf{D}\boldsymbol{x} & \leq & \boldsymbol{q} \end{array}$$

where

$\boldsymbol{x} \in \mathbb{R}^n$ is a vector of optimized variables

$\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^p$, $\boldsymbol{q} \in \mathbb{R}^m$ are vectors

$\mathbf{A} \in \mathbb{R}^{p\times n}$, $\mathbf{D} \in \mathbb{R}^{m\times n}$, $\mathbf{H} \in \mathbb{R}^{n\times n}$ are matrices

Note that LP and QP can be always rewritten to a simpler form using the **slack variables trick:** the inequality constraints $\mathbf{D}\boldsymbol{x} \leq \boldsymbol{q}$ are replaced by equivalent constraints $\mathbf{D}\boldsymbol{x} + \boldsymbol{\xi} = \boldsymbol{q}$ and $\boldsymbol{\xi} \geq \mathbf{0}$.

(Globally) optimal value:

$$p^* = \inf\{f_0(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathcal{X}_{\text{feas}}\}$$

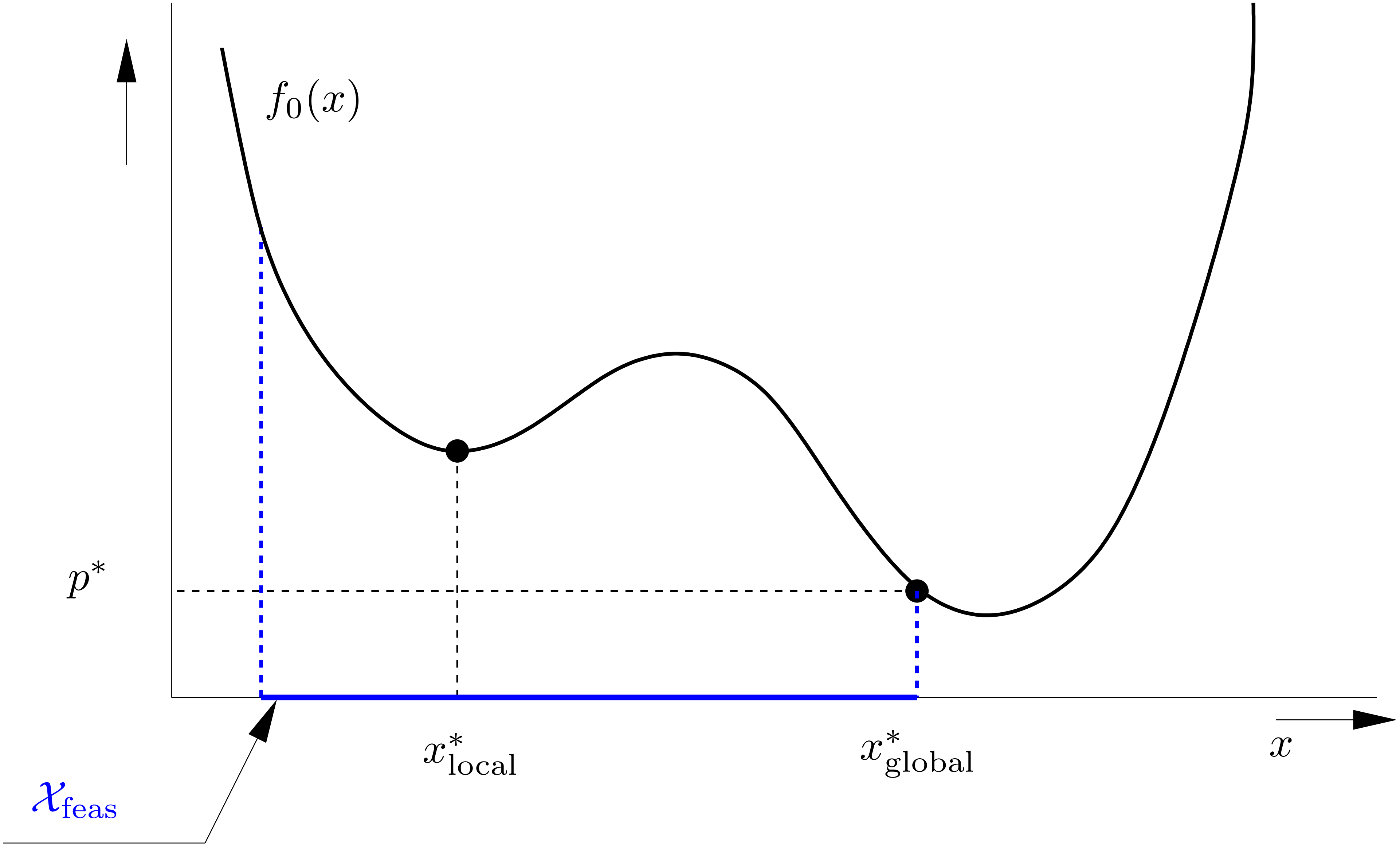
- $p^* = \infty$ if the problem is infeasible, i.e., $\mathcal{X}_{\text{feas}} = \{\emptyset\}$.
- $p^* = -\infty$ if the problem is unbounded.

Optimal solutions: $\boldsymbol{x}$ is the optimal solution if it is feasible and $f(\boldsymbol{x}) = p^*$; $\mathcal{X}_{\text{opt}} = \{\boldsymbol{x} \mid f_0(\boldsymbol{x}) = p^*, \boldsymbol{x} \in \mathcal{X}_{\text{feas}}\}$ is the set of optimal solutions.

Locally optimal: $\boldsymbol{x}$ is locally optimal if there exist $R > 0$ such that $\boldsymbol{x}$ is optimal for

$$\begin{array}{ll} \text{minimize} & f_0(\boldsymbol{y}) \\ \text{subject to} & \boldsymbol{y} \in \mathcal{X}_{\text{feas}} \cap \{\boldsymbol{y} \mid \|\boldsymbol{x} - \boldsymbol{y}\| \le R\} \end{array}$$

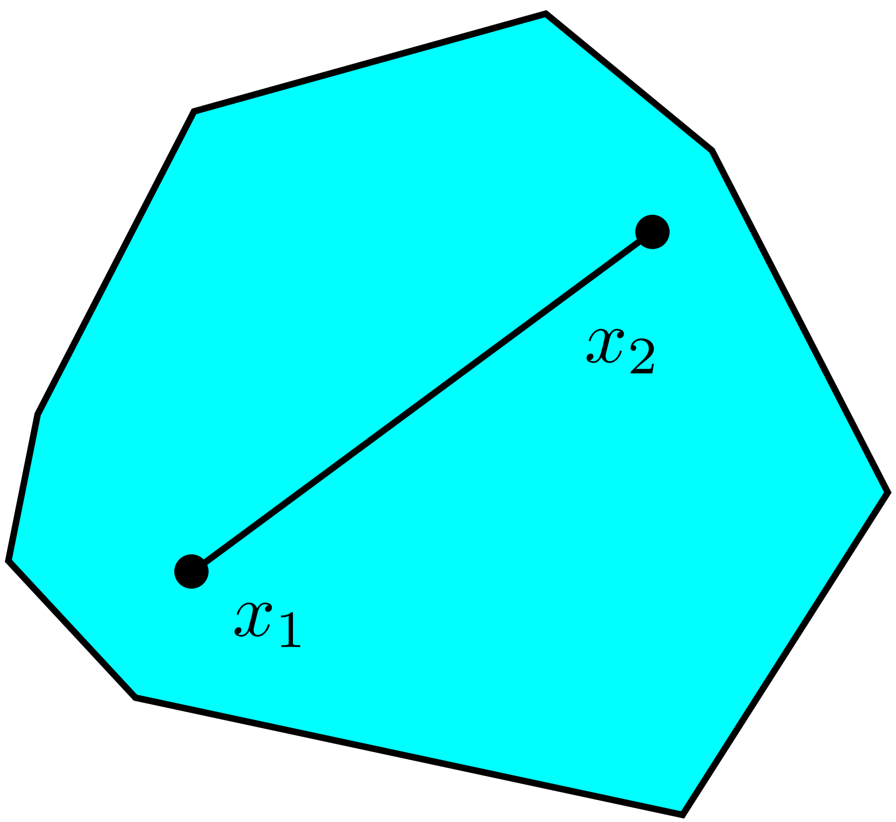
Example: Optimal and locally optimal solution



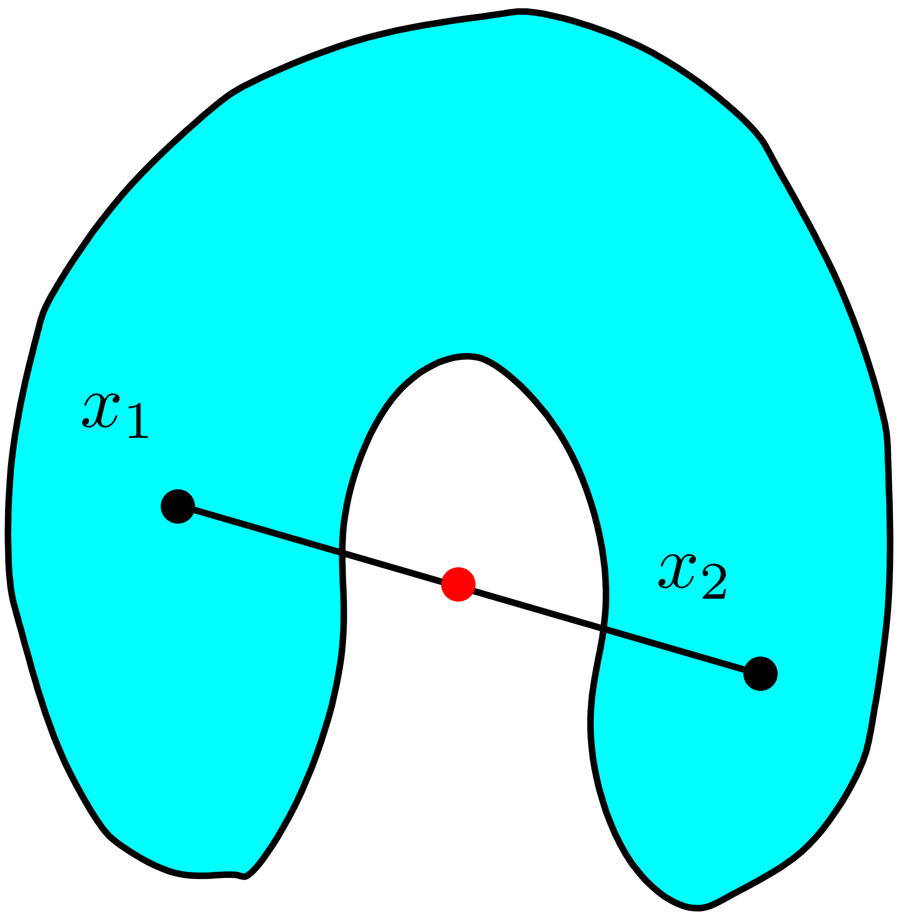
A set $\mathcal{X} \subseteq \mathbb{R}^n$ is convex if the line segment connecting any two points from $\mathcal{X}$ lies in $\mathcal{X}$, i.e., for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$ and all θ such that $0 \leq \theta \leq 1$ it holds

$$\boldsymbol{x}_1(1-\theta) + \theta \boldsymbol{x}_2 \in \mathcal{X} \ .$$

Convex set



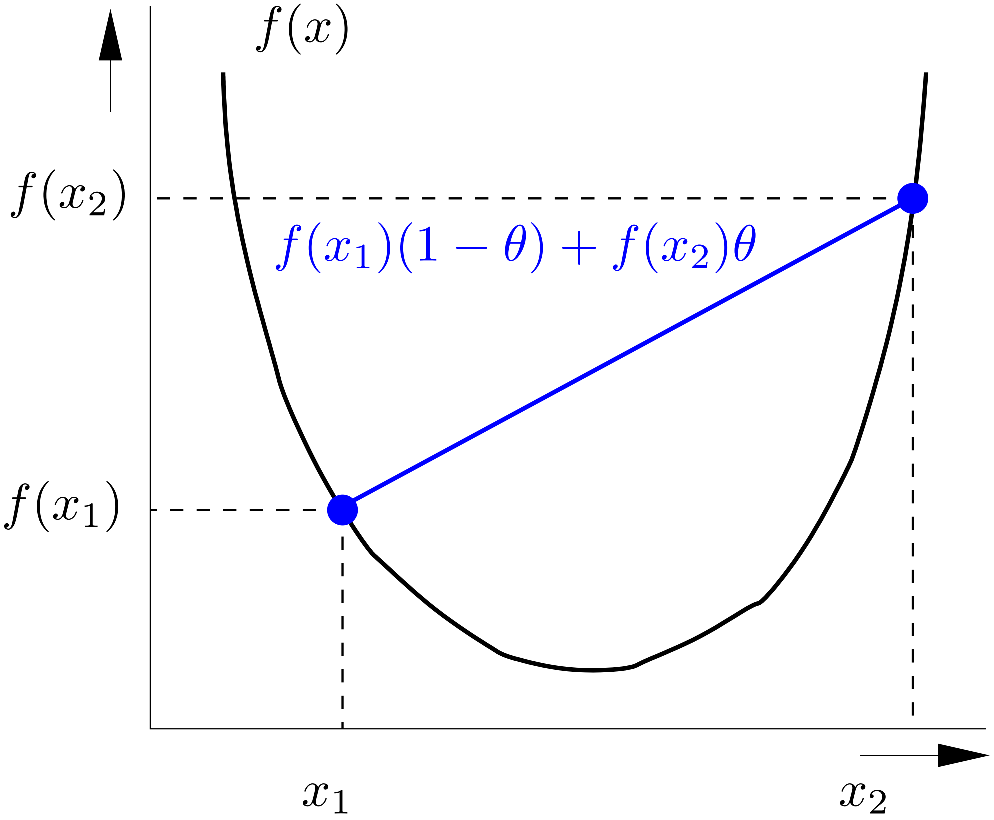
Non-convex set



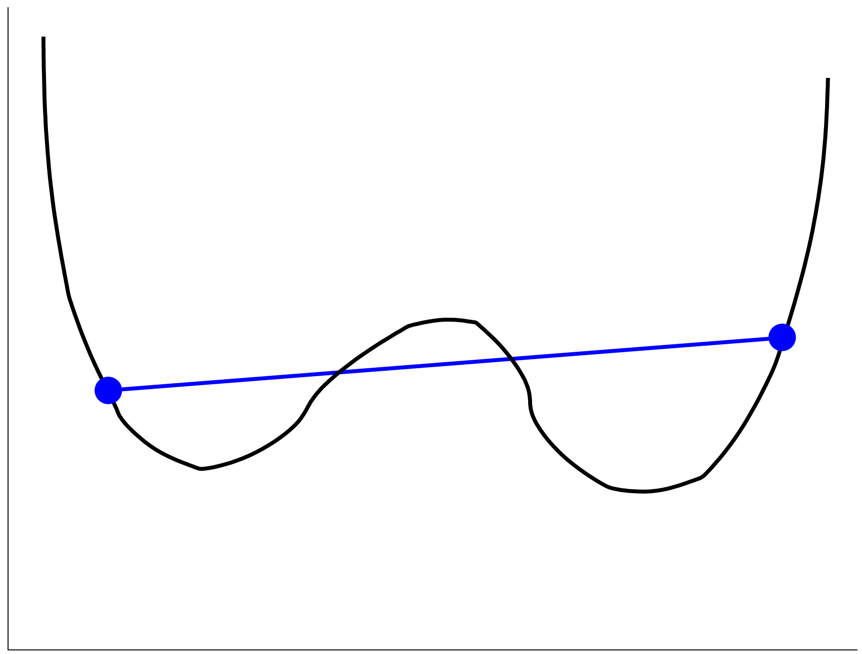
A function $f \in \mathbb{R}^n \to \mathbb{R}$ is convex if $\mathbf{dom}\, f$ is convex and for all $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in \mathbf{dom}\, f$ and all θ such that $0 \leq \theta \leq 1$ it holds

$$f(\boldsymbol{x}_1(1-\theta) + \boldsymbol{x}_2\theta) \leq f(\boldsymbol{x}_1)(1-\theta) + f(\boldsymbol{x}_2)\theta\,.$$

Convex function

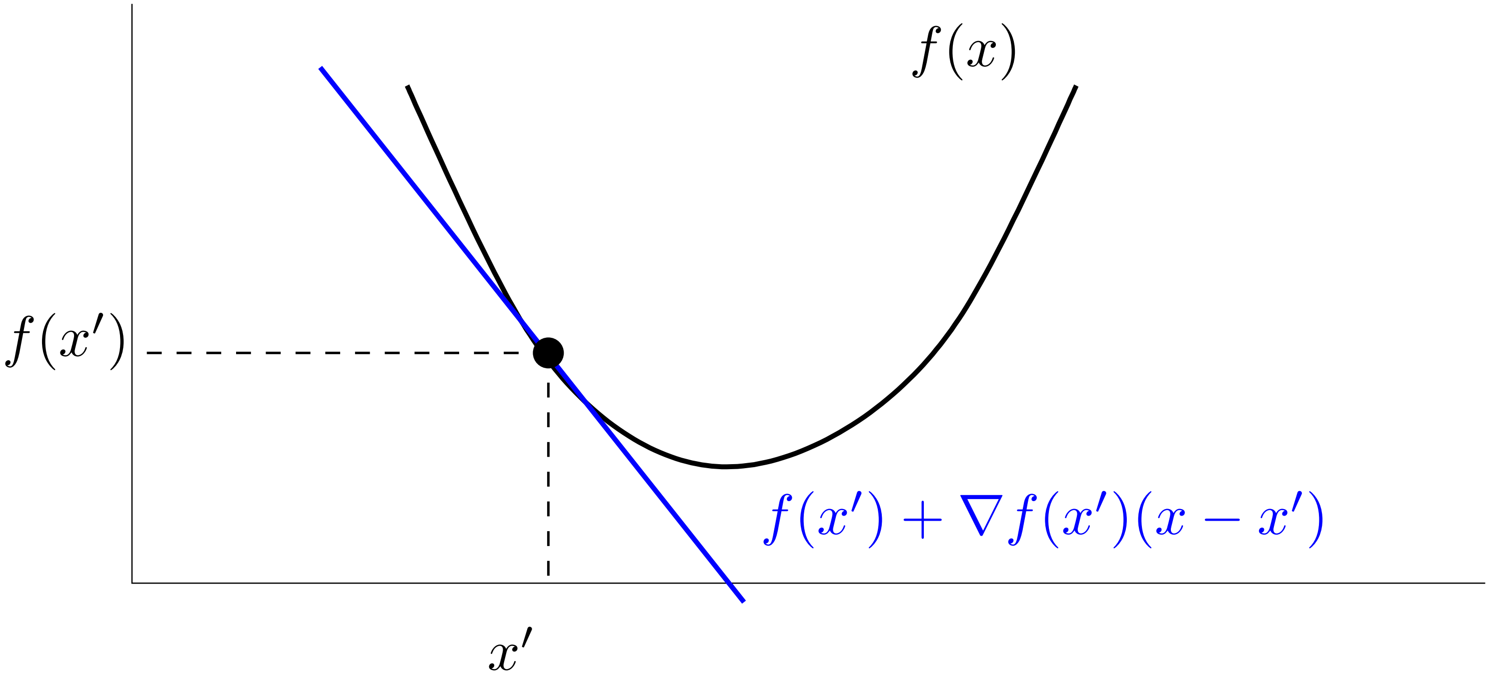


Non-convex function



First-order condition: Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}$ is differentiable, i.e., gradient $\nabla f(\boldsymbol{x}) \in \mathbb{R}^n$ exists at each point $\boldsymbol{x} \in \mathbf{dom}\, f$. Then f is convex if and only if $\mathbf{dom}\, f$ is convex and

$$f(\boldsymbol{x}) \leq f(\boldsymbol{x}') + \nabla f(\boldsymbol{x}')^T (\boldsymbol{x} - \boldsymbol{x}') \,, \qquad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X} \,.$$



Second-order condition: Suppose that f twice differentiable, i.e., the Hessian matrix of second derivatives $\nabla^2 f(\boldsymbol{x})$ exists at each point $\boldsymbol{x} \in \mathbf{dom}\, f$. Then f is convex if and only if $\mathbf{dom}\, f$ is convex and $\nabla^2 f(\boldsymbol{x})$ is positive semi-definite for all $\boldsymbol{x} \in \mathbf{dom}\, f$.

Convex optimization problem

The optimization problem is convex if the objective function $f_0(\boldsymbol{x})$ is convex and the feasible set $\mathcal{X}_{\text{feas}}$ is convex.

- In particular, the problem is convex if $f_0, f_1, \ldots, f_m$ are convex and the equality constraints h_i are affine, i.e., $h_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{x} - b_i = 0$.
- The **standard form** of the convex optimization problem

$$\begin{array}{lrcl} \text{minimize} & f_0(\boldsymbol{x}) & & \\ \text{subject to} & f_i(\boldsymbol{x}) & \leq & 0\,, \quad i = 1, \ldots, m \\ & \mathbf{A}\boldsymbol{x} & = & \boldsymbol{x} \end{array}$$

- **Basic property of the convex problems:** Any locally optimal solution is globally optimal $\Rightarrow$ greatly simplifies optimization.
  - We can use **descent methods**: iteratively move in a descent direction until we reach the optimum.
  - For **non-convex problems we can get stuck in a local optimum**; it is difficult to identify whether the attained optimum is local or global.

LP problem	**QP problem**
$$\begin{array}{lrcl} \text{minimize} & \boldsymbol{c}^T\boldsymbol{x} & & \\ \text{subject to} & \mathbf{A}\boldsymbol{x} & = & \boldsymbol{b} \\ & \mathbf{D}\boldsymbol{x} & \leq & \boldsymbol{q} \end{array}$$	$$\begin{array}{lrcl} \text{minimize} & \frac{1}{2}\boldsymbol{x}^T\mathbf{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} & & \\ \text{subject to} & \mathbf{A}\boldsymbol{x} & = & \boldsymbol{b} \\ & \mathbf{D}\boldsymbol{x} & \leq & \boldsymbol{q} \end{array}$$

Linear Programming is a convex problem since the objective is a convex function, the equality functions are affine, the inequality constraints define a convex set.

Quadratic Programming is a convex problem if and only if the matrix $\mathbf{H}$ is positively semi-definite;

Recall the Second-order condition and notice that for QP the Hessian matrix $\nabla^2 f(\boldsymbol{x}) = \mathbf{H}$.

Suppose that f_0 is differentiable. Then a vector $\boldsymbol{x}$ is the optimal solution if and only if it is feasible $\boldsymbol{x} \in \mathcal{X}_{\text{feas}}$ and

$$\nabla f_0(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) \geq 0 \quad \text{for all} \quad \boldsymbol{y} \in \mathcal{X}_{\text{feas}} \,.$$

How to see this?

- Recall the definition of the directional derivative

$$f_0'(\boldsymbol{x}; \boldsymbol{\delta}) = \lim_{h \to 0_+} \frac{f_0(\boldsymbol{x} + h\boldsymbol{\delta})}{h} = \nabla f_0(\boldsymbol{x})^T \boldsymbol{\delta} \,.$$

  The sign of $f_0'(\boldsymbol{x}; \boldsymbol{\delta})$ determines whether f_0 increases or decreases when we move from $\boldsymbol{x}$ in the direction $\boldsymbol{\delta}$.

- Moving from a feasible point $\boldsymbol{x}$ along a feasible direction $\boldsymbol{\delta} = \boldsymbol{y} - \boldsymbol{x}$, $\boldsymbol{y} \in \mathcal{X}_{\text{feas}}$ by sufficiently small step produces a feasible point.

- A vector $\boldsymbol{x}$ is optimal iff there is no feasible direction which decreases the objective function, i.e., for each $\boldsymbol{y} \in \mathcal{X}_{\text{feas}}$ moving along $\boldsymbol{\delta} = \boldsymbol{y} - \boldsymbol{x}$ increases the objective so that

$$f_0'(\boldsymbol{x}; \boldsymbol{\delta}) \geq 0 \quad \Rightarrow \quad \nabla f_0(\boldsymbol{x})^T \boldsymbol{\delta} \geq 0 \quad \Rightarrow \nabla f_0(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) \geq 0 \,.$$

What are we going to do?

- For the optimized problem (called primal in this context) we derive a dual optimization problem.

What is it good for?

- **Optimality certificate**. Primal objective function is an upper bound and the **dual objective function is a lower bound on the optimal value** $\Rightarrow$ theoretically justified stopping conditions for optimization.

- **Simplifies optimization**. The dual problem can be of lesser complexity; in some cases the primal solution can be easily obtained from the dual solution.

- **New insight**. The dual problem can bring a new insight to the problem (e.g. Max-flow/Min-cut problems from graph theory are dual, or Maximum-likelihood/Minimum-entropy density estimation problems are dual).

Primal optimization problem in standard form

$$\begin{array}{lrcll} \text{minimize} & f_0(\boldsymbol{x}) & & & \\ \text{subject to} & f_i(\boldsymbol{x}) & \leq & 0\,, & i = 1, \ldots, m \\ & h_j(\boldsymbol{x}) & = & 0\,, & j = 1, \ldots, p \end{array}$$

where $\mathcal{D}$ is the problem domain, p^* is the optimal value.

Lagrangian: $L\colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ with domain $\mathbf{dom}\, L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} h_i \nu_i$$

- sum of objective function plus weighted sum of constraint functions
- λ_i is Lagrange multiplier associated with $f_i(\boldsymbol{x}) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(\boldsymbol{x}) = 0$

Lagrange dual function $g\colon \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$
\begin{aligned}
g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\
&= \inf_{\boldsymbol{x} \in \mathcal{D}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x}) \right)
\end{aligned}
$$

- $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is a **concave function** since it is point-wise infimum of convex functions of $(\boldsymbol{\lambda}, \boldsymbol{\nu})$; note that it holds in general even for non-convex primal problems.
- For many important problem $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ has an analytical form.

Example: Lagrange dual function for LP problem

- We start form the primal LP problem

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \\ & \mathbf{D}\boldsymbol{x} \leq \boldsymbol{q} \end{array}$$

- We form the Lagrangian (using matrix notation for brevity)

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) &= f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} h_i \nu_i \\ &= \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{\lambda}^T(\mathbf{D}\boldsymbol{x} - \boldsymbol{q}) + \boldsymbol{\nu}^T(\mathbf{A}\boldsymbol{x} - \boldsymbol{b}) \\ &= (\boldsymbol{c} + \mathbf{D}^T\boldsymbol{\lambda} + \mathbf{A}^T\boldsymbol{\nu})^T\boldsymbol{x} - \boldsymbol{\lambda}^T\boldsymbol{q} - \boldsymbol{\nu}^T\boldsymbol{b} \end{aligned}$$

- We get the Lagrange dual function by minimizing w.r.t primal variables

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\boldsymbol{\lambda}^T\boldsymbol{q} - \boldsymbol{\nu}^T\boldsymbol{b} & \text{if} \quad \boldsymbol{c} + \mathbf{D}^T\boldsymbol{\lambda} + \mathbf{A}^T\boldsymbol{\nu} = \mathbf{0} \\ -\infty & \text{otherwise} \end{cases}$$

Weak duality: If $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\boldsymbol{x} \in \mathcal{X}_{\text{feas}}$ then $f_0(\boldsymbol{x}) \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$, i.e. the **Lagrange dual function is a lower bound** on the primal objective. In particular, it lower bounds the optimal value $p^* \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$, $\forall \boldsymbol{\lambda} \geq \mathbf{0}$, $\forall \boldsymbol{\nu}$.

To see this recall the Lagrangian

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} h_i \nu_i$$

and notice that for $\boldsymbol{x} \in \mathcal{X}_{\text{feas}}$ we have:

1. $f_i(\boldsymbol{x}) \leq 0$ and thus $\sum_i \lambda_i f(\boldsymbol{x}) \leq 0$ since $\lambda_i \geq 0$,
2. $h_i(\boldsymbol{x}) = 0$ and thus $\sum_i \nu_i h_i(\boldsymbol{x}) = 0$,

therefore

$$f_0(\boldsymbol{x}) \geq f_0(\boldsymbol{x}) + \underbrace{\sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x})}_{\leq 0} + \underbrace{\sum_{i=1}^{p} h_i \nu_i}_{=0} = L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \geq \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \,.$$

Note that the **weak duality holds in general** regardless the primal problem is convex or not.

Dual problem

$$\begin{array}{ll} \text{maximize} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{subject to} & \boldsymbol{\lambda} \geq 0 \end{array}$$

where we optimize w.r.t $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\nu} \in \mathbb{R}^p$; the optimal value denoted by d^*.

- Solving the dual problem $\approx$ **finding the best lower bound** d^* **on primal optimal value** p^* which can be obtained from the Lagrangian.
- **Duality gap** is the difference between the primal and the dual optimal values $p^* - d^* \geq 0$, i.e., it determines the tightness of the lower bound.
- The **dual problem is always convex** since $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is a concave function regardless the primal problem is convex or not.
- $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ are **dual feasible** if $\boldsymbol{\lambda} \geq 0$ and $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\inf$, i.e. for dual feasible points we have non-trivial lower bound.

  It usually helps if the constraint $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\inf$ is expressed explicitly in the dual problem.

The **primal LP problem**

$$\begin{array}{lrcl} \text{minimize} & \boldsymbol{c}^T\boldsymbol{x} & & \\ \text{subject to} & \mathbf{A}\boldsymbol{x} & = & \boldsymbol{b} \\ & \mathbf{D}\boldsymbol{x} & \leq & \boldsymbol{q} \end{array}$$

with the Lagrange dual function

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\boldsymbol{\lambda}^T\boldsymbol{q} - \boldsymbol{\nu}^T\boldsymbol{b} & \text{if} \quad \boldsymbol{c} + \mathbf{D}^T\boldsymbol{\lambda} + \mathbf{A}^T\boldsymbol{\nu} = \mathbf{0} \\ -\infty & \text{otherwise} \end{cases}$$

The dual problem reads

$$\begin{array}{ll} \text{maximize} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{subject to} & \boldsymbol{\lambda} \geq 0 \end{array}$$

Making the constraint $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\inf$ explicit, i.e., $\boldsymbol{c} + \mathbf{D}^T\boldsymbol{\lambda} + \mathbf{A}^T\boldsymbol{\nu} = \mathbf{0}$, we get the **dual LP problem**

$$\begin{array}{lrcl} \text{maximize} & -\boldsymbol{\lambda}^T\boldsymbol{q} - \boldsymbol{\nu}^T\boldsymbol{b} & & \\ \text{subject to} & \boldsymbol{\lambda} & \geq & 0 \\ & \boldsymbol{c} + \mathbf{D}^T\boldsymbol{\lambda} + \mathbf{A}^T\boldsymbol{\nu} & = & \mathbf{0} \end{array}$$

Strong duality holds if the duality gap is zero, i.e., $p^* = d^*$ and the Lagrangian lower bound is tight.

When does it happen?

- It **does not hold in general**.
- It **holds if the primal problem is convex and the Slater's condition** (also called constraint qualification) holds:

  Slater's condition holds if there exists a strictly feasible point, i.e., there exists $\boldsymbol{x} \in \mathcal{X}_{\text{feas}}$ such that $f_i(\boldsymbol{x}) < 0$, $i = 1, \ldots, m$; note that this condition is very mild.
- There also exist non-convex problems for which the strong duality holds.

A triplet $(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ satisfy the Karush-Kuhn-Tucker conditions if:

$\dfrac{\partial L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})}{\partial \boldsymbol{x}} = \mathbf{0}$	partial derivative of L w.r.t $\boldsymbol{x}$ vanishes
$\dfrac{\partial L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})}{\partial \boldsymbol{\lambda}} \leq \mathbf{0}$	implies $f_i(\boldsymbol{x}) \leq 0$, $i = 1, \ldots, m$
$\dfrac{\partial L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} = \mathbf{0}$	implies $h_i(\boldsymbol{x}) = 0$, $i = 1, \ldots, p$.
$\boldsymbol{\lambda} \geq \mathbf{0}$	duality constraint holds
$\lambda_i f_i(\boldsymbol{x}) = 0$, $i = 1, \ldots, m$	so called complementary slackness

- If **strong duality holds** then **KKT conditions are necessary** for $(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ to be optimal.
- If **primal problem is convex** and **Slater's condition holds** then **KKT conditions are necessary and sufficient** for $(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ to be optimal.

The primal LP problem

$$\begin{array}{lrcl} \text{minimize} & \boldsymbol{c}^T\boldsymbol{x} & & \\ \text{subject to} & \mathbf{A}\boldsymbol{x} & = & \boldsymbol{b} \\ & \mathbf{D}\boldsymbol{x} & \leq & \boldsymbol{q} \end{array}$$

with the Lagrangian

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{\lambda}^T(\mathbf{D}\boldsymbol{x} - \boldsymbol{q}) + \boldsymbol{\nu}^T(\mathbf{A}\boldsymbol{x} - \boldsymbol{b})$$

The KKT conditions read:

$$\begin{array}{lll} \dfrac{\partial L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})}{\partial \boldsymbol{x}} = \mathbf{0} & \Rightarrow & \boldsymbol{c} + \mathbf{D}^T\boldsymbol{\lambda} + \mathbf{A}^T\boldsymbol{\nu} = \mathbf{0} \\ \dfrac{\partial L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})}{\partial \boldsymbol{\lambda}} \leq \mathbf{0} & \Rightarrow & \mathbf{D}\boldsymbol{x} - \boldsymbol{q} \leq \mathbf{0} \\ \dfrac{\partial L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} = \mathbf{0} & \Rightarrow & \mathbf{A}\boldsymbol{x} - \boldsymbol{b} = \mathbf{0} \\ \boldsymbol{\lambda} \geq \mathbf{0} & \Rightarrow & \boldsymbol{\lambda} \geq \mathbf{0} \\ \lambda_i f_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m & \Rightarrow & \boldsymbol{\lambda}^T(\mathbf{D}\boldsymbol{x} - \boldsymbol{q}) = \mathbf{0} \end{array}$$

Let us consider an unconstrained convex problem

$$\text{minimize} \quad f(\boldsymbol{x})$$

General descent method:

Initialization: set $\boldsymbol{x} \in \mathbf{dom}\, f$.

repeat

1. Determine a descent dirrection $\boldsymbol{\delta}$.
2. Line-search: find a step size $t = \text{argmin}_{t'>0} f(\boldsymbol{x} + t'\boldsymbol{\delta})$.
3. Update $\boldsymbol{x} := \boldsymbol{x} + t\boldsymbol{\delta}$.

until stopping condition is satisfied.

- It generates a sequence of $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \dots$ such that $f(\boldsymbol{x}^{(k)}) > f(\boldsymbol{x}^{(k+1)})$.
- For f differentiable, a vector $\boldsymbol{\delta}$ is a descent direction if

$$f'(\boldsymbol{x}; \boldsymbol{\delta}) = \lim_{h \to 0_+} \frac{f(\boldsymbol{x} + h\boldsymbol{\delta})}{h} = \nabla f(\boldsymbol{x})^T \boldsymbol{\delta} < 0$$

e.g., gradient descent methods use $\boldsymbol{\delta} = -\nabla f(\boldsymbol{x})$.

Let us consider equality constrained convex problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

- Using the KKT optimality conditions, $\boldsymbol{x} \in \textbf{dom}\, f$ is optimal iff there exist $\boldsymbol{\nu}$ such that

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b}\,, \qquad \nabla f(\boldsymbol{x}) + \mathbf{A}^T\boldsymbol{\nu} = 0\,.$$

- For a **convex quadratic function** $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\mathbf{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x}$ the **KKT conditions** lead to an efficiently solvable **set of linear equations**:

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b}\,, \qquad \mathbf{H}\boldsymbol{x} + \boldsymbol{c} + \mathbf{A}^T\boldsymbol{\nu} = 0\,.$$

- **Newton method** is applicable for a general twice differentiable function $f(\boldsymbol{x})$: it iteratively approximates $f(\boldsymbol{x})$ by a quadratic function

$$\hat{f}(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}')\nabla^2 f(\boldsymbol{x}')(\boldsymbol{x}-\boldsymbol{x}') + \nabla f(\boldsymbol{x}')^T(\boldsymbol{x}-\boldsymbol{x}') + f(\boldsymbol{x}')$$

and solves the KKT conditions for the approximation $\hat{f}(\boldsymbol{x})$.

Let us consider equality constrained convex problem

$$\begin{array}{lrcl} \text{minimize} & f_0(\boldsymbol{x}) & & \\ \text{subject to} & f_i(\boldsymbol{x}) & \leq & 0\,, \quad i = 1, \ldots, m \\ & \mathbf{A}\boldsymbol{x} & = & \boldsymbol{b} \end{array}$$

- Constraints $f_i(\boldsymbol{x}) \leq 0$ can be made implicit using the **barrier function**

$$\phi_i(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \; f_i(\boldsymbol{x}) \leq 0 \\ \infty & \text{if } \; f_i(\boldsymbol{x}) > 0 \end{cases}$$

  i.e., we can equivalently optimized **equality constraint problem**

$$\begin{array}{lrcl} \text{minimize} & f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \phi_i(\boldsymbol{x}) & & \\ \text{subject to} & \mathbf{A}\boldsymbol{x} & = & \boldsymbol{b} \end{array}$$

- Functions $\phi_i(\boldsymbol{x})$ are approximated by a **differentiable convex functions**

$$\hat{\phi}_i(\boldsymbol{x}) = -\frac{1}{t}\log(-f_i(\boldsymbol{x}))\,,$$

  which for high t well approximates the step barrier function $\phi_i(\boldsymbol{x})$.

Materials used to prepare this lecture:

- S. Boyd, L. Vandenberghe: *Convex optimization*. Cambridge University Press. 2004.

  Available at: `http://www.stanford.edu/~boyd/cvxbook/`

- S. Boyd: *Lecture notes for EE364*, Stanford University. 2007-2008.

  Available at: `http://www.stanford.edu/class/ee364/`

- H. Hindi: A Tutorial on Convex Optimization II: Duality and Interior Point Methods. Palo Alto Research Center, California.

  Google: `hindi tutorial convex`

Further recommended literature:

- D.P. Bertsekas. *Nonlinear Programming*. (2nd edition), Athena Scientific, Belmont, Massachusetts, 1999.

- J.F. Bonnans, et. al: *Numerical Optimization*. (2nd edition), Springer, Heidelberg, 2006.