# Optimisation I

Markus Grasmair

Department of Mathematics,
Norwegian University of Science and Technology,
Trondheim, Norway

Trondheim,
April 24, 2023

# Questionnaire

Please fill in the questionnaire at `https://nettskjema.no/a/333926`, if you have not done so yet!

# Outline

1. Theory of optimisation
   - Basics
   - Unconstrained problems
   - Constrained optimisation
   - Convex optimisation
   - Lagrangian duality
   - Multi-criteria optimisation

2. Numerical optimisation
   - Line search methods
   - CG and quasi-Newton
   - Trust region methods
   - Non-linear least squares methods
   - Penalty and barrier methods
   - Linear and quadratic programming
   - Sequential quadratic programming

# Formulation of optimisation problems

Want to solve a problem of the form

$$\min_x f(x) \qquad \text{s.t. } x \in \Omega \tag{P}$$

for some $\Omega \subset \mathbb{R}^d$.

- Unconstrained optimisation: $\Omega = \mathbb{R}^d$.

Typically in the case of constrained optimisation:

$$x \in \Omega \iff \begin{cases} c_i(x) = 0, & i \in \mathcal{E}, \\ c_i(x) \geq 0, & i \in \mathcal{I}. \end{cases}$$

# Local and global solutions

A point $x^* \in \Omega$ is:

- a *global solution* of $(P)$, if

$$f(x^*) \leq f(x) \qquad \text{for all } x \in \Omega,$$

- a *local solution* of $(P)$, if there exists $\varepsilon > 0$ such that

$$f(x^*) \leq f(x) \qquad \text{for all } x \in \Omega \text{ with } \|x - x^*\| \leq \varepsilon.$$

# Existence of solutions

The problem admits a global solution if (e.g.):

- $\Omega$ is compact (closed and bounded) and $f$ is (lower semi-)continuous.
- $\Omega$ is closed, and the function $f$ is coercive ($f(x) \to +\infty$ as $\|x\| \to \infty$) and (lower semi-)continuous.
- $f$ is (lower semi-)continuous, and there exists a closed and bounded subset $\Omega_0 \subset \Omega$ and $x_0 \in \Omega_0$ such that

$$f(x) \geq f(x_0) \qquad \text{for all } x \in \Omega \setminus \Omega_0.$$

Note:

- $\Omega$ is closed if the functions $c_i$ are all continuous.

# Outline

# Optimality conditions for unconstrained problems

- If $x^*$ is a local solution of $(P)$, then
$$\nabla f(x^*) = 0$$
and
$$H_f(x^*) \qquad \text{is positive semi-definite.}$$
- If
$$\nabla f(x^*) = 0$$
and
$$H_f(x^*) \qquad \text{is positive definite,}$$
then $x^*$ is a (strict) local solution of $(P)$.

# Outline

# Tangent cones

Let $\Omega \subset \mathbb{R}^d$ be non-empty and closed.

- A vector $p \in \mathbb{R}^d$ is a tangent to $\Omega$ at $x$, if there exist sequences

$$
\begin{aligned}
&(z_k)_{k\in\mathbb{N}} \subset \Omega, && z_k \to x,\\
&(t_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{>0}, && t_k \to 0,
\end{aligned}
$$

  such that

$$p = \lim_{k\to\infty} \frac{z_k - x}{t_k}.$$

- The tangent cone $T_\Omega(x)$ is the set of all tangents to $\Omega$ at $x$.
- If $x^*$ is a local solution of $(P)$, then

$$\langle \nabla f(x^*), p\rangle \geq 0 \qquad \text{for all } p \in T_\Omega(x^*).$$

# Linearised feasible directions and LICQ

- The active set $\mathcal{A}(x)$ consists of all indices $i \in \mathcal{E} \cup \mathcal{I}$ with $c_i(x) = 0$.
- The set of all linearised feasible directions $\mathcal{F}(x)$ consists of all vectors $p \in \mathbb{R}^d$ such that

$$\begin{aligned} \langle \nabla c_i(x), p \rangle &= 0 && \text{for all } i \in \mathcal{E}, \\ \langle \nabla c_i(x), p \rangle &\geq 0 && \text{for all } i \in \mathcal{I} \cap \mathcal{A}(x). \end{aligned}$$

- We say that the Linear Independence Constraint Qualification (LICQ) is satisfied at $x \in \Omega$, if the family of vectors

$$\nabla c_i(x), \qquad i \in \mathcal{A}(x),$$

  is linearly independent.
- We always have that $T_\Omega(x) \subset \mathcal{F}(x)$.
- If LICQ holds at $x$, then $T_\Omega(x) = \mathcal{F}(x)$.

# KKT conditions

Define the Lagrangian $\mathcal{L} \colon \mathbb{R}^d \times \mathbb{R}^{\mathcal{E} \cup \mathcal{I}} \to \mathbb{R}$,

$$\mathcal{L}(x, \lambda) = f(x) - \sum_i \lambda_i c_i(x).$$

Assume that $x^*$ is a local solution of $(P)$ and that LICQ holds at $x^*$. Then there exists $\lambda^*$ such that (KKT conditions):

$$\begin{aligned}
\nabla f(x^*) &= \sum_i \lambda_i^* \nabla c_i(x^*), & \\
c_i(x^*) &= 0, & i \in \mathcal{E}, \\
c_i(x^*) &\geq 0, & i \in \mathcal{I}, \\
\lambda_i^* &\geq 0, & i \in \mathcal{I}, \\
\lambda_i^* c_i(x^*) &= 0, & i \in \mathcal{I}.
\end{aligned}$$

# Second order conditions

Assume that $x^*$ is a KKT point with Lagrange multiplier $\lambda^*$. Denote

$$\mathcal{C}(x^*) := \big\{p \in \mathcal{F}(x^*) : \langle \nabla f(x^*), p\rangle = 0\big\}.$$

- If $x^*$ is a local solution of $(P)$ and LICQ holds at $x^*$, then

  $$\big\langle p, H_{\mathcal{L}}(x^*, \lambda^*)p\big\rangle \geq 0$$

  for all $p \in \mathcal{C}(x^*)$.
- If

  $$\big\langle p, H_{\mathcal{L}}(x^*, \lambda^*)p\big\rangle > 0$$

  for all $p \in \mathcal{C}(x^*) \setminus \{0\}$, then $x^*$ is a strict local solution of $(P)$.

# Outline

# Convex sets and functions

- The set $U \subset \mathbb{R}^d$ is convex if

$$\lambda x + (1-\lambda)y \in U \qquad \text{whenever } x,\ y \in U \text{ and } 0 < \lambda < 1.$$

- If $U \subset \mathbb{R}^d$ is convex, then the function $f\colon U \to \mathbb{R}$ is convex if

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$$

  whenever $x$, $y \in U$, and $0 < \lambda < 1$.

- The function $f$ is convex if and only if

$$f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle$$

  for all $x$, $y \in U$.

- The function $f$ is convex if and only if

$$H_f(x) \qquad \text{is positive semi-definite for all } x.$$

# Optimality conditions

Assume that the function $f$ is convex, that the functions $c_i$, $i \in \mathcal{E}$, are linear, and the functions $c_i$, $i \in \mathcal{I}$, are concave.

- If the KKT conditions hold at $x^* \in \Omega$, then $x^*$ is a *global* solution of $(P)$.
- We say that *Slater's constraint qualification* holds, if there exists $x$ with $c_i(x) = 0$, $i \in \mathcal{E}$, and $c_i(x) > 0$, $i \in \mathcal{I}$.[1]
- If Slater's constraint qualification holds, then the KKT conditions are necessary and sufficient.

Note: In the unconstrained case $\Omega = \mathbb{R}^d$, the point $x^*$ is a global solution of $(P)$, if and only if $\nabla f(x^*) = 0$.

[1] A small, but potentially significant, generalisation is possible and useful: In the case of linear inequality constraints $a_i^T x \geq b_i$, we do not need to require that these are satisfied strictly.

# Feasible directions and projections

Assume that $\Omega \subset \mathbb{R}^d$ is closed and convex and that $f\colon \Omega \to \mathbb{R}$ is convex.

- A direction $p \in \mathbb{R}^d$ is feasible at $x$ if $x + tp \in \Omega$ for some $t > 0$.
- The projection $\pi_\Omega(z)$ of a point $z$ onto $\Omega$ is the unique solution of the problem
$$\min_x \frac{1}{2}\|x - z\|_2^2 \qquad \text{s.t. } x \in \Omega.$$
- The projection $\pi_\Omega(z)$ is characterised by the variational inequality
$$\big\langle z - \pi_\Omega(z), x - \pi_\Omega(z)\big\rangle \leq 0 \qquad \text{for all } x \in \Omega.$$
- The point $x^* \in \Omega$ is a global solution of $(P)$ if and only if
$$\langle \nabla f(x^*), p\rangle \geq 0 \qquad \text{for all feasible directions } p \text{ at } x^*.$$
- The point $x^* \in \Omega$ is a global solution of $(P)$ if and only if
$$x^* = \pi_\Omega\big(x^* - \alpha \nabla f(x^*)\big) \qquad \text{for all/any } \alpha > 0.$$

# Outline

# Primal and dual problems

- Can equivalently write the problem $(P)$ as

$$\min_{x} \max_{\substack{\lambda \\ \lambda_i \geq 0,\ i \in \mathcal{I}}} \mathcal{L}(x, \lambda). \tag{P}$$

- Define the dual problem as

$$\max_{\substack{\lambda \\ \lambda_i \geq 0,\ i \in \mathcal{I}}} \min_{x} \mathcal{L}(x, \lambda). \tag{D}$$

- E.g.: Linear programmes:

Primal problem:

$$c^T x \to \min \text{ s.t. } Ax \geq b.$$

Dual problem:

$$b^T \lambda \to \max \text{ s.t. } \begin{cases} A^T \lambda = c, \\ \lambda \geq 0. \end{cases}$$

# Strong duality in convex optimisation

Assume that the function $f$ is convex, that the functions $c_i$, $i \in \mathcal{E}$, are linear, the functions $c_i$, $i \in \mathcal{I}$, are concave, and that Slater's constraint qualification holds.

- If $(P)$ is bounded below, then $(D)$ admits a solution $\lambda^*$.
- A saddle point of the Lagrangian (or: *primal–dual solution* of $(P)$) is a pair $(x^*, \lambda^*)$ such that

$$\mathcal{L}(x^*, \lambda) \le \mathcal{L}(x^*, \lambda^*) \le \mathcal{L}(x, \lambda^*)$$

  for all $x$ and $\lambda$ with $\lambda_i \ge 0$, $i \in \mathcal{I}$.
- If $(P)$ has a solution, then the Lagrangian has a saddle point $(x^*, \lambda^*)$. Moreover, $x^*$ is a global solution of $(P)$ with Lagrange multiplier $\lambda^*$.

# Outline

# Pareto-optimality

We want to solve a multi-criteria optimisation problem

$$\min_x \bigl(f_1(x), \ldots, f_m(x)\bigr) \qquad \text{s.t. } x \in \Omega. \qquad (M)$$

A point $x^* \in \Omega$ is a Pareto-optimal solution of $(M)$, if there does *not* exist any point $\hat{x} \in \Omega$ such that

$$f_i(\hat{x}) \le f_i(x^*) \qquad \text{for all } 1 \le i \le m,$$

and

$$f_j(\hat{x}) < f_j(x^*) \qquad \text{for at least one } 1 \le j \le m.$$

# Weighted sum method

- Assume that $0 < \lambda_i \leq 1$ such that $\sum_i \lambda_i = 1$. If $x^*$ is a (global) solution of the problem

$$\min_x \sum_{i=1}^{m} \lambda_i f_i(x),$$

then $x^*$ is a Pareto-optimal solution of $(M)$.

- Assume that $f_i$, $1 \leq i \leq m$, is strictly convex. Then $x^*$ is a Pareto-optimal solution of $(M)$ if and only if there exist $0 \leq \lambda_i \leq 1$ such that $\sum_i \lambda_i = 1$ and $x^*$ solves

$$\min_x \sum_{i=1}^{m} \lambda_i f_i(x).$$

# Outline

# Basic setup

- Iterative minimisation of an unconstrained optimisation problem of the form
$$\min_{x \in \mathbb{R}^d} f(x).$$
- Initialise $x_0 \in \mathbb{R}^d$ (close to the expected solution $x^*$).
- In each step, compute first a search direction $p_k$.
- Then choose a reasonable step length $\alpha_k > 0$ and set $x_{k+1} = x_k + \alpha_k p_k$.

# Wolfe conditions

- Armijo condition (sufficient decrease):

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \langle \nabla f(x_k), p_k \rangle.$$

- Weak curvature condition:

$$\langle \nabla f(x_k + \alpha_k p_k), p_k \rangle \geq c_2 \langle \nabla f(x_k), p_k \rangle.$$

- Strong curvature condition:

$$\big| \langle \nabla f(x_k + \alpha_k p_k), p_k \rangle \big| \leq c_2 \big| \langle \nabla f(x_k), p_k \rangle \big|.$$

Note:

- Parameters have to satisfy $0 < c_1 < c_2 < 1$.
- Conditions only make sense, if $p_k$ is a descent direction for $f$, that is, if $\langle \nabla f(x_k), p_k \rangle < 0$.
- Armijo condition provides an upper bound for suitable step lengths and guarantees that the function values decrease.
- Weak curvature condition gives a lower bound for the step lengths.

# Backtracking gradient descent method

Choose parameters $0 < c_1 < 1$, $0 < \rho < 1$, $\hat{\alpha} > 0$.

- In each iteration:
- Choose the search direction
$$p_k = -\nabla f(x_k).$$
- Initialise $\alpha = \hat{\alpha}$.
- While
$$f(x_k + \alpha p_k) > f(x_k) - c_1 \alpha \|\nabla f(x_k)\|^2 :$$
  Set $\alpha \leftarrow \rho\alpha$.
- Set $\alpha_k := \alpha$ and $x_{k+1} = x_k + \alpha_k p_k$.

# Newton's method with backtracking line search

Choose parameters $0 < c_1 < 1$, $0 < \rho < 1$, $\hat{\alpha} > 0$.

- In each iteration:
- Choose the search direction $p_k$ as the solution of

$$H_f(x_k)p = -\nabla f(x_k).$$

- Ensure that $p_k$ is a descent direction.[2]
- Initialise $\alpha = 1$.
- While

$$f(x_k + \alpha p_k) > f(x_k) + c_1 \alpha \langle \nabla f(x_k), p_k \rangle :$$

  Set $\alpha \leftarrow \rho\alpha$.
- Set $\alpha_k := \alpha$ and $x_{k+1} = x_k + \alpha_k p_k$.

[2] If not, e.g. modify the Hessian matrix to guarantee positive definiteness/descent directions.

# Convergence of line search methods

- Consider general line search method with search direction

$$p_k = -B_k \nabla f(x_k)$$

with $B_k$ positive definite and symmetric.

- Armijo condition implies that

$$c_1 \alpha_k \langle \nabla f(x_k), B_k \nabla f(x_k) \rangle \le f(x_k) - f(x_{k+1}).$$

- Summation over all $k$ implies

$$c_1 \sum_k \alpha_k \langle \nabla f(x_k), B_k \nabla f(x_k) \rangle \le f(x_0) - \inf_x f(x) < \infty.$$

- If singular values of $B_k$ are bounded above and step lengths $\alpha_k$ are bounded below, the iteration will converge in the sense that

$$\|\nabla f(x_k)\| \to 0.$$

# Convergence of line search methods – II

- Gradient descent method will converge with either backtracking Armijo line search or line search with Wolfe conditions.
- Newton's method will converge with either backtracking line search or line search with Wolfe conditions provided that the function $f$ is $C^2$ and strongly convex.

- Convergence speed of gradient descent method is typically linear.
- One can expect very slow convergence if $H_f(x^*)$ is ill-conditioned.
- Convergence speed of Newton's method is quadratic for either backtracking Armijo line search or line search with Wolfe conditions with $c_1 < 1/2$, provided that $H_f(x^*)$ is non-singular.

# Outline

# Linear CG methods

- Method for the solution of linear systems

$$Qx = b$$

with $Q$ positive definite and symmetric.

- Equivalent: Minimisation of the quadratic functional

$$\Phi(x) = \frac{1}{2}\langle x, Qx\rangle - \langle b, x\rangle.$$

- Main idea: Perform exact line search for the minimisation of $\Phi$ with search direction

$$p_k = -\nabla\Phi(x_k) + \beta_k p_{k-1},$$

where $\beta_k$ is chosen in such a way that

$$\langle p_k, Qp_{k-1}\rangle = 0.$$

# Non-linear CG methods

- Try to mimic linear CG methods as close as possible while minimising a non-quadratic functional $f(x)$.
- Instead of exact line search, use inexact line search with (strong) Wolfe conditions.
- Define the search directions as

$$p_k = -\nabla f(x_k) + \beta_k p_{k-1}$$

  where $\beta_k$ is chosen in such a way that the method reduces to linear CG for quadratic functionals.
- Different choices of $\beta_k$ possible, yielding slightly different algorithms.

# Quasi-Newton method

- Use a Newton-like iteration but with search directions

$$p_k = -H_k \nabla f(x_k),$$

  where $H_k$ is a positive definite approximation to the inverse of $H_f(x_k)$.
- Main condition: matrices $H_k$ should satisfy the secant equation

$$H_{k+1} y_k = s_k \text{ with } s_k = x_{k+1} - x_k \text{ and } y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

- Main update methods:
  - SR1: symmetric rank one update that satisfies the secant equation.
  - BFGS and DFP: symmetric rank two updates that are minimal in a certain norm.
- Typically superlinear convergence.

# Outline

1. Theory of optimisation
   - Basics
   - Unconstrained problems
   - Constrained optimisation
   - Convex optimisation
   - Lagrangian duality
   - Multi-criteria optimisation

2. Numerical optimisation
   - Line search methods
   - CG and quasi-Newton
   - Trust region methods
   - Non-linear least squares methods
   - Penalty and barrier methods
   - Linear and quadratic programming
   - Sequential quadratic programming

# Idea of trust region methods

In each step consider a model function

$$m_k(p) = f(x_k) + \langle \nabla f(x_k), p \rangle + \frac{1}{2} \langle p, B_k p \rangle.$$

Choose a trust region radius $\Delta_k > 0$ and find $p_k$ solving

$$\min_p m_k(p) \qquad \text{s.t. } \|p\| \leq \Delta_k.$$

Choice/adaptation of the trust region radius by Armijo like condition:
Decrease the radius if

$$\frac{f(x_k + p_k) - f(x_k)}{m_k(p_k) - m_k(0)} < \eta$$

for some parameter $0 < \eta < 1$.

# Analytic computation of the trust region step

The vector $p^*$ solves the problem

$$\min_p c + \langle g, p\rangle + \frac{1}{2}\langle p, Bp\rangle \qquad \text{s.t. } \|p\| \le \Delta$$

if and only if there exists $\lambda^* \ge 0$ such that

$$\begin{aligned}
(\lambda^* I + B)p^* &= -g, \\
\|p^*\| &\le \Delta, \\
\lambda^*(\Delta - \|p^*\|) &= 0, \\
(\lambda^* I + B) &\text{ is positive semi-definite.}
\end{aligned}$$

(Approximations are possible, e.g. the dog-leg method.)

# Outline

# Setup of non-linear least squares

Want to solve an overdetermined (possibly inconsistent) non-linear system

$$g(x) = y.$$

Least squares idea: Define the residual

$$r(x) := g(x) - y$$

and minimize the squared residual:

$$f(x) := \frac{1}{2}\|r(x)\|^2 \to \min.$$

Note: With $J(x)$ ... Jacobian of $r$ at $x$ we have

$$\nabla f(x) = J(x)^T r(x).$$

# Gauß–Newton and Levenberg–Marquardt methods

Approximate

$$H_f(x) \approx J(x)^T J(x)$$

with $J(x)$ ... Jacobian of $r$ at $x$.

- Gauß–Newton method: Line search method for the minimisation of $f$ with search direction $p_k$ solving

$$J(x_k)^T J(x_k) p = -J(x_k)^T r(x_k).$$

- Levenberg–Marquardt method: Trust region method for the minimisation of $f$ with model function

$$m_k(p) = \frac{1}{2} \left\| r(x_k) + J(x_k) p \right\|^2.$$

# Outline

# Quadratic penalty method

Want to solve the problem

$$f(x) \to \min \qquad \text{s.t. } c_i(x) = 0, \; i \in \mathcal{E}.$$

Approximate solution by minimising the quadratic penalty functional

$$Q(x, \mu) = f(x) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i(x)^2.$$

Solutions $x_\mu$ expected to converge to $x^*$ as $\mu \to \infty$.

- Ill-conditioned optimisation problem for large $\mu$.
- Potential existence/divergence problems if $f$ is not coercive.

# Augmented Lagrangian method

Define the augmented Lagrangian as

$$\mathcal{L}_A(x, \lambda; \mu) = \mathcal{L}(x, \lambda) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i(x)^2.$$

Solutions $x^*$ with Lagrange multiplier $\lambda^*$ are also local minimisers of $\mathcal{L}_A(\cdot, \lambda^*; \mu)$ if $\mu$ is sufficiently large (but finite).

Fixed point iteration (for $\lambda$):

- Compute a solution $x_k$ of

$$\min_x \mathcal{L}_A(x, \lambda_k; \mu)$$

- Update

$$\lambda_{k+1} = \lambda_k - \mu c(x_k).$$

# Logarithmic barrier method

Consider inequality constrained problem

$$f(x) \to \min \qquad \text{s.t. } c_i(x) \geq 0, \; i \in \mathcal{I}.$$

Define the barrier functional

$$B(x, \beta) = f(x) - \beta \sum_{i \in \mathcal{I}} \log(c_i(x)).$$

Minimisers of $B(\cdot, \beta)$ expected to converge to solutions of the constrained problem as $\beta \to 0$.

- Combination with quadratic penalisation/augmented Lagrangian for equality constraints possible.
- Usage of slack variables recommended in the case of “complicated” constraints $c_i$.

# Outline

# Linear programmes in standard form

A linear programme in standard form reads

$$\min_x c^T x \qquad \text{s.t.} \qquad \begin{aligned} Ax &= b, \\ x &\geq 0. \end{aligned} \tag{P}$$

Every linear programme can be brought in standard form by:

- introducing slack variables,
- splitting up variables into positive and negative parts.

The dual programme to $(P)$ has the form

$$\max_\lambda b^T \lambda \qquad \text{s.t.} \qquad \begin{aligned} A^T\lambda + s &= c, \\ s &\geq 0. \end{aligned} \tag{D}$$

# Interior point methods for linear programmes

Approximate the KKT conditions for a programme in standard form by

$$\begin{aligned} Ax &= b, \\ A^T\lambda + s &= c, \qquad \text{(IP)} \\ x_i s_i &= \tau, \end{aligned}$$

with the implicit additional constraints $x_i > 0$, $s_i > 0$ for all $i$.

Basic setup of interior point methods:

- Make a single Newton step for (IP).
- Choose a suitable step length $0 < \alpha \leq 1$.
- Decrease $\tau$.

# Quadratic programmes

Consider quadratic optimisation problems with linear constraints

$$q(x) = \frac{1}{2}\langle x, Gx\rangle + \langle c, x\rangle \to \min$$

s.t.

$$\begin{aligned} \langle a_i, x\rangle &= b_i \quad i \in \mathcal{E}, \\ \langle a_i, x\rangle &\geq b_i \quad i \in \mathcal{I}. \end{aligned}$$

Assume in addition that $G$ is symmetric and positive definite.

- Solution is uniquely characterised by the KKT conditions.
- In the case of equality constraints only, the KKT conditions reduce to a linear system.

# Active set methods

Solution of quadratic programme with both equality and inequality constraints:

Main goal: Find the true active set $\mathcal{A}(x^*)$ by starting with a working set $\mathcal{W}_0 \supset \mathcal{E}$ and adding/removing indices.

- Compute the solution $p_k$ of

$$\min_p q(x_k + p) \qquad \text{s.t. } \langle a_i, p \rangle = 0, \ i \in \mathcal{W}_k.$$

- If $p_k = 0$, compute Lagrange parameters $\lambda_i$. If $\lambda_i < 0$ for some $i \in \mathcal{W}_k \cap \mathcal{I}$, set $x_{k+1} = x_k$, $\mathcal{W}_{k+1} = \mathcal{W}_k \setminus \{i\}$.
- Else: choose $0 < \alpha_k \leq 1$ maximal such that

$$\langle a_i, x_k + \alpha_k p_k \rangle \geq b_i \qquad \text{for all } i \notin \mathcal{W}_k,$$

set $x_{k+1} = x_k + \alpha_k p_k$ and

$$\mathcal{W}_{k+1} = \big\{ i \in \mathcal{E} \cup \mathcal{I} : \langle a_i, x_{k+1} \rangle = b_i \big\}.$$

# Outline

# Newton's method for the KKT system

In the case of optimisation with equality constraints, apply Newton's method for the solution of the KKT system

$$\begin{aligned} \nabla f(x) - A(x)^T \lambda &= 0, \\ c(x) &= 0, \end{aligned}$$

with $A(x)$ ... Jacobian of $c$.

- Newton steps require the solution of a linear system with an indefinite matrix.
- Stabilisation of the iteration by backtracking Armijo line search using a *merit function*, e.g. the $\ell^1$ merit function

$$\Phi_1(x, \mu) = f(x) + \mu \|c(x)\|_1.$$

# Sequential quadratic programming

In the case of equality and inequality constraints, compute in step $k$ the primal–dual solution $(p_k, \hat{\lambda}_{k+1})$ of

$$\min_p \langle \nabla f(x_k), p\rangle + \frac{1}{2}\langle p, H_{\mathcal{L}}(x_k, \lambda_k) p\rangle$$

subject to

$$\begin{aligned} c_i(x_k) + \langle \nabla c_i(x_k), p\rangle &= 0, \quad i \in \mathcal{E}, \\ c_i(x_k) + \langle \nabla c_i(x_k), p\rangle &\ge 0, \quad i \in \mathcal{I}. \end{aligned}$$

Perform backtracking Armijo line search for an $\ell^1$-like merit function and define

$$x_{k+1} = x_k + \alpha_k p_k, \qquad \lambda_{k+1} = \lambda_k + \alpha_k(\hat{\lambda}_{k+1} - \lambda_k).$$

# Exam information

- Date: May 09 2023, 15:00–19:00.
- Place: TBA.
- Permitted aids:
  - Approved calculator.
- A formula sheet will be attached to the exam (see the wiki page).
- On the wiki page, you can find a detailed overview of the pensum.
- Question/help session on Friday, May 05, 14:15– in EL1.