Optimization (Repetition)

Convexity

- Convex set $S \quad \Leftrightarrow \quad \lambda x_1 + (1-\lambda)x_2 \in S$, $\forall x_1, x_2 \in S$, $\forall \lambda \in [0,1]$.
- Convex function $f \Leftrightarrow D_f$ convex and
  $f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$, $\forall x_1, x_2 \in D_f$, $\forall \lambda \in [0,1]$.
- f convex $\;\Leftrightarrow\;$ epi(f) convex $\;\Leftrightarrow\;$ $f(x+\lambda d)$ convex $\forall x, d$: $x + \lambda d \in D_f$.

Why convexity is good?

- If f convex then
  - loc. min. $\;\Rightarrow\;$ glob. min.
  - stat.point $\;\Rightarrow\;$ glob. min.

  Note that glob. min. does not always exist for convex functions (i.e. $y = e^x$).
- For convex $\min_{x \in S} f$: a is a glob. min. $\;\Leftrightarrow\;$ $\nabla f(a)^T(x-a) \ge 0$, $\forall x \in S$.
- If $S = \{x \in X \mid g(x) \le 0,\; h(x) = 0\}$ and X convex, f, g convex, h affine $\Rightarrow$ *convex* problem.
- For convex problems:
  - KKT $\;\Rightarrow\;$ saddle point $\;\Rightarrow\;$ glob. min.
  - Slater condition: $\exists x_0 \in S\colon g(x_0) < 0 \;\Rightarrow\;$ no duality gap/saddle point.

How to check that a set S is convex?

- Picture ($n \le 3$) or definition.
- S_1, S_2 convex $\;\Rightarrow\;$ $S_1 \cap S_2$ convex.
- f convex $\;\Rightarrow\;$ $\{x \mid f(x) \le \text{const}\}$ convex.

How to check that a function f is convex?

- Graph ($n \le 2$) or definition.
- f_1, f_2 convex $\;\Rightarrow\;$ $f_1 + f_2$ convex and $\max\{f_1, f_2\}$ convex.
- g convex $\nearrow$ and h convex $\;\Rightarrow\;$ $g(h(x))$ convex.
- g convex and h affine $\;\Rightarrow\;$ $g(h(x))$ convex.
- f convex $\;\Leftrightarrow\;$ $\nabla^2 f$ pos.-semidef.

Positive-definite and positive-semidefinite

- H pos.-def. $\Leftrightarrow$ $x^T Hx > 0, \forall x \neq 0$.
- H pos.-def. $\Rightarrow$ $x^T Hx + c^T x + q$ strictly convex $\Rightarrow$
  $\Rightarrow$ glob. min. unique (if exists).
- H pos.-def. $\Rightarrow$ $-H\nabla f$ is a descent direction.
- Loc. min. $\Rightarrow$ $\nabla f = 0$, $\nabla^2 f$ pos.-semidef.
- $\nabla f = 0$, $\nabla^2 f$ pos.-def. $\Rightarrow$ loc. min.
- $\nabla^2 f$ pos.-semidef. on S $\Leftrightarrow$ f convex on S.

How to check positive-definiteness?

- Sylvester: H pos.-def. $\Leftrightarrow$ $\det(H_k) > 0, \forall k = 1, \ldots, n$.
- H pos.-def. $\Leftrightarrow$ all eigenvalues > 0.

How to check positive-semidefiniteness?

- Necessary: H pos.-semidef. $\Rightarrow$ $\det(H_k) \geq 0, \forall k = 1, \ldots, n$.
- Sufficient: modified Sylvester
  $\det(H_k) > 0, \forall k = 1, \ldots, n-1$ and $\det(H) \geq 0$ $\Rightarrow$ H pos.-semidef.
- Completing the squares: H pos.-semidef. $\Leftrightarrow$ $f(x) = x^T Hx =$ sum of squares.
- H pos.-semidef. $\Leftrightarrow$ $H + \epsilon I$ pos.-def. $\forall \epsilon > 0$.
- H pos.-semidef. $\Leftrightarrow$ all eigenvalues ≥ 0.

Factorizations

- $H = C^T C$ $\Rightarrow$ H pos.-semidef.
- $H = C^T C$ and $\det(H) \neq 0$ $\Rightarrow$ H pos.-def.
- Cholesky: H pos.-def. $\Leftrightarrow$ $H = LL^T$, L low-triang., $\det(L) \neq 0$.
- H pos.-def. $\Leftrightarrow$ $H = LDL^T$, L low. triang., $L_{kk} = 1$, $D = \text{diag} > 0$.

Search

Dichotomous vs. Golden section:

- GS: fewer function evaluations.
- Unimodal $\Rightarrow$ glob. min.

Armijo: fast but inexact (normally used in multi-dim.)
Newton vs. Modified Newton:

- Newton: faster
- Modified: always descent direction, better convergence

Newton vs. Quasi-Newton:

- Newton: uses 2d derivative
- Quasi-Newton: only 1st derivative

Conj. dir. vs. Quasi-Newton (DFP, BFGS):

- CD: $d_{new} = -\nabla f + \beta d_{old}$, β updates.
- Quasi-Newton: $d = -D\nabla f$, D updates, lots of memory.

Steepest decent vs. Conj. dir.

- SD: zigzagging
- CD: faster

Convergence for quadratic functions:

- Newton: in one step
- CD = quasi-Newton: in n steps of inner loop (= one outer loop)

LP and Duality

- Particular case:
  - P: $\min c^T x \mid Ax \geq b,\ x \geq 0$
  - D: $\max b^T y \mid A^T y \leq c,\ y \geq 0$
- General case:
  - P: "=" in row k $\Leftrightarrow$ D: y_k free
  - P: x_k free $\Leftrightarrow$ D: "=" in row k
- Easy to get from $c^T x - b^T y = (c - A^T y)^T x + y^T (Ax - b) \geq 0$
- CSP: "=" instead of "$\geq$" above
- Strong duality: finite min in P $\Rightarrow$ finite max in D and min = max.
- $\bar{x}$ primal feasible, $\bar{y}$ dual feasible + CSP $\Rightarrow$ both are the optimal solutions.

Constrained Optimization

- Necessary: loc. min. $\Rightarrow$ CQ point or KKT point
- Sufficient:
  - KKT + convex $\Rightarrow$ glob. min.
  - KKT + 2d order cond. $\Rightarrow$ loc. min.
  - Saddle point $\Rightarrow$ glob. min.
- Numerical solution via penalty/barrier function methods.
  - Penalty: unfeasible approximations.
  - Barrier: feasible, cannot handle equalities.

To check saddle point via Duality:

P: $\min f(x) \mid x \in X,\ g(x) \leq 0,\ h(x) = 0.$

D: $\max \Theta(u,v) \mid u \geq 0,$ where $\Theta(u,v) = \inf_{x \in X} L(x,u,v).$

1. Find $\Theta(u,v)$ and get (if possible) the optimal $x = x(u,v)$.
2. Find $\max \Theta(u,v)$ and get the optimal $\bar{u}$, $\bar{v}$.
3. Put $\bar{x} = x(\bar{u},\bar{v})$ (or calculate $\bar{x}$ as the optimal x on Step 1 for given $\bar{u}$, $\bar{v}$).
4. If $\Theta(\bar{u},\bar{v}) = f(\bar{x})$ then $\bar{x}$ is glob. min.

The course contents:

- Ch 2,3,9: Numerical methods (except Nelder-Mead simplex method).
- Ch 4: Convex sets.
- Ch 5: LP (except the Simplex method).
- Ch 6: Convex functions (except Subgradient and Maximization).
- Ch 7: KKT necessary/sufficient conditions (no Quadratic Programming).
- Ch 8: Duality.