Convex Optimization 2025/26

Practical session # 11

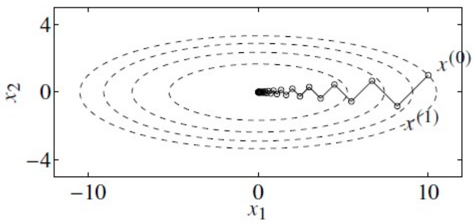
December 11, 2025

1. Let us study the gradient descent (with exact line search) for the problem:

$$\text{minimize } x_1^2 + \gamma x_2^2,$$

for $\gamma > 0$.

- Show that, for $\gamma = 1$, gradient descent solves the problem in one step.
- For arbitrary γ, discuss one step of gradient decent for the points $(\gamma, \pm 1)$.
- From the previous point, conclude that from starting point $x^{(0)} = (\gamma, 1)$ we get the approximation sequence $x^{(0)}, x^{(1)}, x^{(2)}, \dots$ with $x_1^{(k)} = c^k\gamma$, $x_2^{(k)} = (-c)^k$, for $c = \frac{\gamma-1}{\gamma+1}$. For which values of γ does this converge faster/slower? Do we always get 'zig-zagging'?



2. Generalizing Exercise 1 let us next consider a quadratic problem

$$\text{minimize } \frac{1}{2}x^T P x + q^T x + r,$$

for $P \in S^n_{++}, q \in \mathbf{R}^n, r \in \mathbf{R}$.

- By the linear coordinate transformation $y = P^{\frac{1}{2}}x$, the problem is equivalent to

$$\text{minimize } \frac{1}{2}y^T y + \tilde{q}^T y + r,$$

with $\tilde{q} = P^{-\frac{1}{2}}x$. Show that the gradient method solves this transformed problem in one step.
- Based on the descent direction Δy of the transformed problem, which descent direction $\Delta x = P^{-\frac{1}{2}}\Delta y$ do we get for the original problem?

3. Given an arbitrary convex function $f(x)$ that is twice differentiable, we can approximate it locally around some x by the quadratic function

$$f(x + \Delta x) \approx f(x) + \nabla f(x)^T \Delta x + \frac{1}{2}(\Delta x)^T \nabla^2 f(x) \Delta x.$$

By minimizing this quadratic function, we get the *Newton descent direction* $\Delta_{NT}x$. Compute it, and compare it with the result of Exercise 2.

4. In the lecture, we mentioned that $K(x, y) = (x^T y + c)^d$, for $c > 0$ is a kernel for support vector machines, whose feature space are polynomials of degree $\leq d$.

Verify that for $d = 2$ this is indeed the case under the feature map $\phi\colon \mathbf{R}^n \to \mathbf{R}^m$:

$$\phi(x) = (x_1^2, \dots, x_n^2, \sqrt{2}x_1x_2, \dots, \sqrt{2}x_{n-1}x_n, \sqrt{2}cx_1, \dots, \sqrt{2}cx_n, c).$$