# SI251 Convex Optimization
# Homework 2

Instructor: Professor Ye Shi

Due on 3 Nov 23:59 UTC+8

Note:

- Please provide enough calculation process to get full marks.
- Please submit your homework to Gradescope with entry code: **J7DK2D**.
- Please check carefully whether the question number on the gradescope corresponds to each question.

## Exercise 1. Convex optimization problem and KKT conditions

(30-pts) Allocating asset portfolios is a core task in financial investment. The goal is to maximize investment returns while keeping risk under control. Consider an investor who wants to allocate capital among $n$ assets, each with different expected returns and risk characteristics. In this scenario, each asset $i$ has a basic positive rate of return $r_i$, where $r_i > 0$ and $r_i$ is a known quantity. And $p_i$ represents the proportion of capital invested in the asset. The investor's goal is to optimize the overall benefit of his portfolio, which can be simulated by the utility function $\log(r_i+p_i)$, where the log function reflects the property of diminishing marginal utility, that is, as the investment amount increases, the additional benefits brought by each unit of additional investment decrease. The investor's total capital is fixed, so the sum of all investment proportions must be equal to 1.

- Based on the topic background, construct a **minimize** convex optimization problem. [Hint: Pay attention to the elements of the optimization problem]
- Get the best investment plan based on KKT conditions.

Solution:

$$\min_{\mathbf{p}} -\sum_{i=1}^{n} \log(r_i + p_i)$$

$$\text{subject to } \mathbf{p} \succeq 0, \mathbf{1}^T\mathbf{p} = 1$$

Introduce Lagrange multiplier $\lambda^* \in \mathbb{R}^n$ for inequality constraints $\mathbf{p} \succeq 0$ and multiplier $v^* \in \mathbb{R}$ for equality constraints $\mathbf{1}^T\mathbf{p} = 1$.

The Lagrange function is

$$L(\mathbf{p}, \lambda^*, v^*) = -\sum_{i=1}^{n} \log(r_i + p_i) - \lambda^*\mathbf{p}^* + v^*(\mathbf{1}^T\mathbf{p} - 1)$$

We obtain the KKT conditions

1. primal feasibility:
$$\mathbf{p}^* \succeq 0$$
2. dual feasibility:
$$\lambda^* \succeq 0$$
3. complementary slackness:
$$\lambda_i^* p_i^* = 0, i = 1, \cdots, n$$
4. zero gradient of Lagrangian with respect to $x$:
$$f_i(x) = p_i - 0, h_i(x) = \mathbf{1}^T\mathbf{p} - 1, \nabla f_i(x) = 1, \nabla h_i(x) = 1$$
$$-\frac{1}{r_i + p_i^*} - \lambda_i^* + v^* = 0, i = 1, \cdots, n$$

According to the above four conditions, we can solve the equation to get

$$\mathbf{p}^* \succeq 0, \quad \mathbf{1}^T\mathbf{p}^* = 1, \quad p_i^*(v^* - \frac{1}{r_i + p_i^*}) = 0, \quad v^* \geq \frac{1}{r_i + p_i^*} \text{ for } i = 1, \cdots, n$$

If $v^* < 1/r_i$, the last condition can only hold if $p_i^* > 0$, which by the third condition implies that $v^* = 1/(r_i + p_i^*)$. Solving for $p_i^*$, we conclude that $p_i^* = 1/v^* - r_i$ if $v^* < 1/r_i$.

If $v^* \geq 1/r_i$, then $p_i^* > 0$ is impossible, because it would imply that $v^* > 1/r_i > 1/(r_i + p_i)$, which violates the complementary slackness condition. Therefore, $p_i^* = 0$ if $v^* \geq 1/r_i$. Thus we have

$$p_i^* = \begin{cases} 1/v^* - r_i, & \text{if } v^* < 1/r_i \\ 0, & \text{if } v^* \geq 1/r_i \end{cases}$$

or, put more simply, $p_i^* = \max\{0, 1/v^* - r_i\}$. Substituting this expression for $p_i^*$ into the condition $\mathbf{1}^T \mathbf{p}^* = 1$ we obtain

$$\sum_{i=1}^{n} \max\{0, 1/v^* - r_i\} = 1$$

The lefthand side is a piecewise-linear increasing function of $1/v^*$, with breakpoints at $r_i$, so the equation has a unique solution which is readily determined.

## Exercise 2. L-smooth

(30-pts) Suppose the function $g : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable. Please prove that the following relations holds for all $x, y \in \mathbb{R}$ if $g$ with an $L$-Lipschitz continuous conditions,

$$[1] \Rightarrow [2]$$

**[1]** $\langle \nabla g(x) - \nabla g(y), x - y \rangle \leq L\|x - y\|^2$,

**[2]** $g(y) \leq g(x) + \nabla g(x)^T (y - x) + \frac{L}{2}\|y - x\|^2$,

Solution:
Define the function $G$: $[0, 1] \to \mathbb{R}$

$$G(t) := g(x + t(y - x)) - g(x) - \langle \nabla g(x), t(y - x) \rangle$$

so that $G(0) = 0$ and $G(1) = g(y) - g(x) - \langle \nabla g(x), t(y - x) \rangle$. By the fundamental theorem of calculus, we have

$$\begin{aligned} G(1) - G(0) = \int_0^1 G'(t) dt &= \int_0^1 \langle \nabla g(x + t(y - x)) - \nabla g(x), y - x \rangle dt \\ &= \int_0^1 \langle \nabla g(x + t(y - x)) - \nabla g(x), t(y - x) \rangle \frac{1}{t} dt \\ &\leq L\|y - x\|_2^2 \int_0^1 t dt \\ &= \frac{L}{2}\|y - x\|_2^2 \end{aligned}$$

Thus, we can get $f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}||y - x||^2$

## Exercise 3. backtracking line search

(40-pts) Please show the convergence of backtracking line search on a $\mu$-strongly convex and $L$-smooth objective function $f$ as

$$f\left(x^{(k)}\right)-p^{\star} \leq c^{k}\left(f\left(x^{(0)}\right)-p^{\star}\right)$$

where $c=1-\min\{2\mu\alpha, 2\beta\alpha\mu/L\}<1$.

---

**Algorithm 1** backtracking line search algorithm.

---

1: **given** a descent direction $\Delta x$ for $f$ at $x \in \mathbf{dom}\ f, \alpha \in (0, 0.5), \beta \in (0,1). t := 1.$
2: **while** $f(x+t\Delta x) > f(x) + \alpha t \nabla f(x)^{T} \Delta x, \quad t := \beta t$
3: **return** $f(x)$;

---

Solution:

Now we consider the case where a backtracking line search is used in the gradient descent method. We will show that the backtracking exit condition,

$$\tilde{f}(t) \leq f(x)-\alpha t\|\nabla f(x)\|_{2}^{2},$$

is satisfied whenever $0 \leq t \leq 1/L$. First note that

$$0 \leq t \leq 1/L \Longrightarrow -t+\frac{Lt^{2}}{2} \leq -t/2$$

which follows from convexity of $-t+Lt^{2}/2$. Using this result and the bound

$$f(y) \leq f(x)+\nabla f(x)^{T}(y-x)+\frac{L}{2}\|y-x\|_{2}^{2},$$

we have, for $0 \leq t \leq 1/L$,

$$\begin{aligned}
\tilde{f}(t) & \leq f(x)-t\|\nabla f(x)\|_{2}^{2}+\frac{Lt^{2}}{2}\|\nabla(f(x))\|_{2}^{2} \\
& \leq f(x)-(t/2)\|\nabla f(x)\|_{2}^{2} \\
& \leq f(x)-\alpha t\|\nabla f(x)\|_{2}^{2},
\end{aligned}$$

since $\alpha < 1/2$. Therefore the backtracking line search terminates either with $t=1$ or with a value $t \geq \beta/L$. This provides a lower bound on the decrease in the objective function. In the first case we have

$$f\left(x^{+}\right) \leq f(x)-\alpha\|\nabla f(x)\|_{2}^{2},$$

and in the second case we have

$$f\left(x^{+}\right) \leq f(x)-(\beta\alpha/L)\|\nabla f(x)\|_{2}^{2}.$$

Putting these together, we always have

$$f\left(x^{+}\right) \leq f(x)-\min\{\alpha, \beta\alpha/L\}\|\nabla f(x)\|_{2}^{2}.$$

Now we can proceed exactly as in the case of exact line search. We subtract $p^{\star}$ from both sides to get

$$f\left(x^{+}\right)-p^{\star} \leq f(x)-p^{\star}-\min\{\alpha, \beta\alpha/L\}\|\nabla f(x)\|_{2}^{2},$$

and combine this with $\|\nabla f(x)\|_{2}^{2} \geq 2\mu\left(f(x)-p^{\star}\right)$ to obtain

$$f\left(x^{+}\right)-p^{\star} \leq (1-\min\{2\mu\alpha, 2\beta\alpha\mu/L\})\left(f(x)-p^{\star}\right).$$

From this we conclude

$$f\left(x^{(k)}\right) - p^{\star} \leq c^{k}\left(f\left(x^{(0)}\right) - p^{\star}\right)$$

where

$$c = 1 - \min\{2\mu\alpha, 2\beta\alpha\mu/L\} < 1.$$

In particular, $f\left(x^{(k)}\right)$ converges to $p^{\star}$ at least as fast as a geometric series with an exponent that depends (at least in part) on the condition number bound $M/m$. In the terminology of iterative methods, the convergence is at least linear.