Exercises in Convex Optimization
Lecture 6, second week

Michel Baes, Patrick Cheridito

November 1, 2018

KKT Conditions with Equality Constraints In this exercise we want to extend the KKT Theorem seen in the class to the more general setting where the optimization problem also possesses linear equality constraints. More concretely, following a similar idea of the proof seen in class, prove the following theorem:

Theorem (KKT Conditions for Convex Optimization III):
Consider the convex optimization problem

$$f^* := \min\{f(x) \mid g(x) \succcurlyeq b, l(x) = d\},$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ is convex, $g_i : \mathbb{R}^n \to \mathbb{R}$ are concave $(i = 1, \ldots, m)$,and l_j are linear $(j = 1, \ldots, M)$.
Assume that *Slater's condition* holds, i.e.:

$$\exists \bar{x} \text{ s.t. } \; g_i(\bar{x}) > b_i \;, \; l_j(\bar{x}) = d_j, \forall i, j.$$

Then x^* is an optimal solution of the problem iff

- *(Feasibility)* $g(x^*) \succcurlyeq b$ and $l(x^*) = d$,
- *(KKT Condition)* There exist $h_0 \in \partial f(x^*), h_i \in \partial(-g_i(x^*)), \xi_j \in \partial l_j(x^*)$ as well as $\lambda_i^* \geq 0, \nu_j^* \in \mathbb{R}$ such that
$$h_0 + \sum_{i=1}^{m} \lambda_i^* h_i + \sum_{j=1}^{M} \nu_j^* \xi_j = 0,$$
- *(Complementarity)* $\lambda_i^*(b_i - g_i(x^*)) = 0, \forall i.$

(*Remark:* The existence of a Slater point can be replaced by many other requirements, listed in the literature as "Regularity Conditions" or "Constraint Qualifications").

Verify the theorem by proving the following claims:

1. x^* is an optimum of the problem iff $x^* = \arg\min_x \Phi(x)$, where

$$\Phi(x) := \max\{f(x) - f^*, b_1 - g_1(x), \ldots, b_m - g_m(x)\} + \sum_{j=1}^{M} \chi_{L_j}(x),$$

$L_j := \{x \mid l_j(x) = d_j\}$, and χ_{L_j} is the characteristic function of the set L_j for all $j = 1, \ldots, M$.

2. $x^* = \arg\min_x \Phi(x)$ iff $0 \in \partial\Phi(x^*)$.
3. $0 \in \partial\Phi(x^*)$ iff there are $h_0 \in \partial f(x^*), h_i \in \partial(-g_i(x^*))$ as well as $\beta_j \in \mathbb{R}, \alpha_0, \alpha_i \geq 0$ with $\alpha_0 + \sum_{i \in I(x^*)} \alpha_i = 1$ such that

$$0 = \alpha_0 h_0 + \sum_{i \in I(x^*)} \alpha_i h_i + \sum_{j=1}^{M} \beta_j \xi_j.$$

4. Prove that $\alpha_0 \neq 0$.
5. Reformulate $0 = \alpha_0 h_0 + \sum_{i \in I(x^*)} \alpha_i h_i + \sum_{j=1}^{M} \beta_j \xi_j$ as $0 = h_0 + \sum_{i=1}^{m} \lambda_i^* h_i + \sum_{j=1}^{M} \nu_j^* \xi_j$ and derive the complementary constraint.

`Lecture_7/KKTWithEquality`

Expected Shortfall as a linear optimization problem We designate by $x \in \mathbb{R}^d$ a decision vector and by Y a random vector taking values in $\mathbb{R}^m$, with probability density p. The loss incurred by the decision x under the circumstance y (a realization of the random vector Y) is written $f(x, y)$. The set of all events that make us facing a loss larger than L when we take the decision x is therefore $E(x, L) := \{y : f(x, y) \geq L\}$, and its probability is:

$$\psi(x, L) := \int_{E(x,L)} p(y) dy.$$

Observe that $\psi(x, L)$ decreases to 0 as L increases.

We assume that ψ is continuous in L for every x.

The *Value-at-Risk* of x at probability level $\alpha \in]0, 1[$ is defined in this notation as:

$$\mathrm{VaR}_\alpha(x) = \min\{L : \psi(x, L) \leq \alpha\}.$$

(We would use this definition with small values of α, such as $\alpha = 0.005$). The expected shortfall at level $\alpha \in]0, 1[$ is defined as:

$$\mathrm{ES}_\alpha(x) := \frac{1}{\alpha} \int_{E(x, \mathrm{VaR}_\alpha(x))} f(x, y) p(y) dy.$$

This formula can be interpreted as follows: we *average* all losses that exceed $L = \mathrm{VaR}_\alpha(x)$. Note that then the correct probability distribution is $p(y)/\alpha$ and not $p(y)$, because we need its integral on its domain $[L, \infty[$ to equal 1.

Let

$$F_\alpha(x, L) := L + \frac{1}{\alpha} \int [f(x, y) - L]^+ p(y) dy.$$

Here, $[t]^+ := \max\{t, 0\}$.

1. Show that $L \mapsto F_\alpha(x, L)$ is convex on $\mathbb{R}$ for every x.
2. Define $G(L) := \int [f(x, y) - L]^+ p(y) dy$. Observe that G is continuously differentiable. Prove that the subgradient of G is $\partial G(L) := \{-\psi(x, L)\}$. Hint: consider the function $z \mapsto \varphi(z) := (z - \bar{L})^+ - (z - L)^+$ for $\bar{L} > L$, then for $\bar{L} < L$ and observe that an integral of the form $\int \varphi(z) h(z) dz$ can be split into three simple terms.
3. Prove now that $\mathrm{ES}_\alpha(x) = \inf_{L \in \mathbb{R}} F_\alpha(x, L)$, that the minimum is attained, and that $\mathrm{VaR}_\alpha(x)$ is one of them.
4. Verify that $\mathrm{ES}_\alpha(x) = F_\alpha(x, \mathrm{VaR}_\alpha(x))$.
   (It is trivial from the above, but it is nice to see all the ingredients united in a single simple formula.)
5. The integral in $F_\alpha(x, L)$ might be difficult to evaluate. Suggest a way to discretize the integral that makes the resulting optimization problem linear.
6. Suppose that $f(x, y)$ is also convex in x for every y. Prove that $F_\alpha(x, L)$ is convex in (x, L) and that $\mathrm{ES}_\alpha(x)$ is convex in x.

(Note: Rockafellar and Uryasev defined the function $F_\alpha(x, L)$ and studied its properties as above. Their results, rediscovered by yourself, have had and still have a tremendous impact in Quantitative Risk Management.)

`Lecture_7/CVaRLinear`