

Numerical optimization

Mines Nancy – Fall 2024

session 7 – constrained convex optimization I
theory: Lagrangian duality and optimality conditions

Lecturer: Julien Flamant (CNRS, CRAN)

✉ julien.flamant@univ-lorraine.fr

📍 Office 425, FST 1er cycle

Course material:

🌐 arche.univ-lorraine.fr/course/view.php?id=74098

🐙 github.com/jflamant/mines-nancy-fall24-optimization

Setting: constrained convex optimization

We consider optimization problems of the form

$$\min_{\mathbf{x} \in \Omega} f_0(\mathbf{x}) \quad (\mathcal{P})$$

where we make the following assumptions:

- the objective $f_0 : \mathcal{D}_0 \subseteq \mathbb{R}^N \rightarrow \mathbb{R}$ is convex on its domain \mathcal{D}_0
- the set of constraints $\Omega \subset \mathbb{R}^N$ is convex. We assume that Ω is described by M inequality constraints and P equality constraints such that

$$\Omega = \left\{ \mathbf{x} \in \mathbb{R}^N \left| \begin{array}{ll} f_i(\mathbf{x}) \leq 0 & , \quad i = 1, \dots, M \\ \mathbf{a}_i^\top \mathbf{x} = b_i & , \quad i = 1, \dots, P \end{array} \right. \right\}$$

where the functions $f_i : \mathbb{R}^N \rightarrow \mathbb{R}$ are convex, hence Ω is a convex set. (prove it)

(\mathcal{P}) is a constrained convex optimization problem

Agenda

Session 7 - 03/12/24: Theory

- Characterization of solutions
- Lagrangian duality
- Karush-Kuhn-Tucker (KKT) conditions

Session 8 - 10/12/24: Algorithms

- Penalty methods
- Projected gradient
- Dual methods: Uzawa's algorithm
- Disciplined convex programming

Main reference

Convex Optimization, Boyd and Vanderberghe (2004) - Chapters 4, 5

Outline

- ① Optimality conditions for constrained convex problems
- ② Lagrangian duality
- ③ Karush-Kuhn-Tucker conditions

Reminder from lecture 2

Theorem (Characterization of convex functions)

Let $f : \mathcal{D} \subseteq \mathbb{R}^N \rightarrow \mathbb{R}$ be twice differentiable. Suppose that \mathcal{D} is convex. These are equivalent:

- ① f is convex on \mathcal{D}
- ② for all $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x})$
- ③ for all $\mathbf{x} \in \mathcal{D}$, $\nabla^2 f(\mathbf{x}) \geq 0$, i.e. the Hessian is positive semidefinite.

Consider the unconstrained convex optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}) \quad (\mathcal{P}_u)$$

where $f : \mathbb{R}^N \rightarrow \mathbb{R}$ is convex (and differentiable).

Theorem (Optimality condition for unconstrained convex problems)

A point \mathbf{x}^* is a solution of (\mathcal{P}_u) if and only if \mathbf{x}^* is a stationary point of f , i.e., such that $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Terminology for constrained problems (I)

Constrained convex optimization problem

$$\min_{\mathbf{x} \in \Omega} f_0(\mathbf{x}) \quad \text{with } f_0 : \mathcal{D}_0 \rightarrow \mathbb{R} \text{ convex and } \Omega \subset \mathbb{R}^N \text{ convex} \quad (\mathcal{P})$$

Feasible and infeasible problems

- a point \mathbf{x} is said to be feasible if $\mathbf{x} \in \Omega \cap \mathcal{D}_0$.
- the set of feasible points is noted $\mathcal{F} = \Omega \cap \mathcal{D}_0$
- (\mathcal{P}) is said to be feasible if there is at least one feasible point
- (\mathcal{P}) is said to be infeasible if $\mathcal{F} = \emptyset$

Terminology for constrained problems (I)

Constrained convex optimization problem

$$\min_{\mathbf{x} \in \Omega} f_0(\mathbf{x}) \quad \text{with } f_0 : \mathcal{D}_0 \rightarrow \mathbb{R} \text{ convex and } \Omega \subset \mathbb{R}^N \text{ convex} \quad (\mathcal{P})$$

Feasible and infeasible problems

- a point \mathbf{x} is said to be feasible if $\mathbf{x} \in \Omega \cap \mathcal{D}_0$.
- the set of feasible points is noted $\mathcal{F} = \Omega \cap \mathcal{D}_0$
- (\mathcal{P}) is said to be feasible if there is at least one feasible point
- (\mathcal{P}) is said to be infeasible if $\mathcal{F} = \emptyset$

Active and inactive constraints

Let \mathbf{x} be a feasible point and $f_i(\mathbf{x}) \leq 0$ an inequality constraint ($i = 1, \dots, P$).

- if $f_i(\mathbf{x}) < 0$ the constraint is said to be **inactive** at \mathbf{x} ;
- if $f_i(\mathbf{x}) = 0$ the constraint is said to be **active** at \mathbf{x} .

Equality constraints are always active at feasible points.

Terminology for constrained problems (II)

Constrained convex optimization problem

$$\min_{\mathbf{x} \in \Omega} f_0(\mathbf{x}) \quad \text{with } f_0 : \mathcal{D}_0 \rightarrow \mathbb{R} \text{ convex and } \Omega \subset \mathbb{R}^N \text{ convex} \quad (\mathcal{P})$$

Optimal value

- $p^* = \inf \{ f_0(\mathbf{x}) \mid \mathbf{x} \in \mathcal{F} \}$
- $p^* = \infty$ if (\mathcal{P}) is infeasible
- $p^* = -\infty$ if (\mathcal{P}) is unbounded below
i.e. there exists a sequence of feasible points $\{\mathbf{x}_k\}$ s.t. $f_0(\mathbf{x}_k) \xrightarrow[k \rightarrow \infty]{} -\infty$.

Terminology for constrained problems (II)

Constrained convex optimization problem

$$\min_{\mathbf{x} \in \Omega} f_0(\mathbf{x}) \quad \text{with } f_0 : \mathcal{D}_0 \rightarrow \mathbb{R} \text{ convex and } \Omega \subset \mathbb{R}^N \text{ convex} \quad (\mathcal{P})$$

Optimal value

- $p^* = \inf \{f_0(\mathbf{x}) \mid \mathbf{x} \in \mathcal{F}\}$
- $p^* = \infty$ if (\mathcal{P}) is infeasible
- $p^* = -\infty$ if (\mathcal{P}) is unbounded below
i.e. there exists a sequence of feasible points $\{\mathbf{x}_k\}$ s.t. $f_0(\mathbf{x}_k) \xrightarrow[k \rightarrow \infty]{} -\infty$.

Optimal points

A point \mathbf{x}^* is optimal, or is a solution to (\mathcal{P}) if

- \mathbf{x}^* is feasible;
- $f_0(\mathbf{x}^*) = p^*$.

NB: Terminology is identical for non-convex problems.

First order characterization of optimal points

Constrained convex optimization problem

$$\min_{\mathbf{x} \in \Omega} f_0(\mathbf{x}) \quad \text{with } f_0 : \mathcal{D}_0 \rightarrow \mathbb{R} \text{ convex and } \Omega \subset \mathbb{R}^N \text{ convex} \quad (\mathcal{P})$$

Assumption: f_0 is differentiable

Proposition

A point \mathbf{x}^* is a solution of (\mathcal{P}) if and only if $\mathbf{x}^* \in \mathcal{F}$ and

$$\nabla f_0(\mathbf{x}^*)^\top (\mathbf{y} - \mathbf{x}^*) \geq 0 \text{ for all } \mathbf{y} \in \mathcal{F}$$

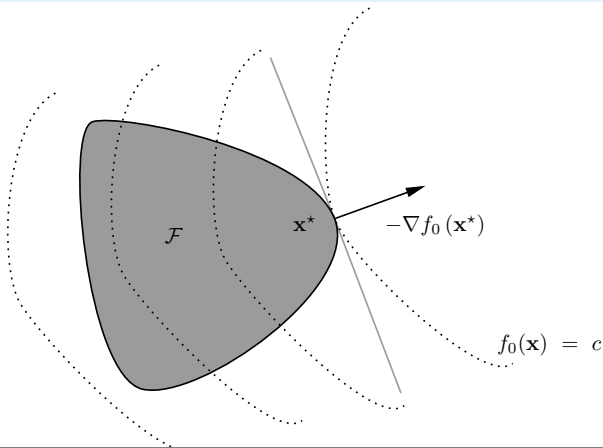
- generalizes the necessary and sufficient condition ($\nabla f_0(\mathbf{x}^*) = 0 \Leftrightarrow \mathbf{x}^*$ is solution) encountered for unconstrained convex problems;
- if $\nabla f_0(\mathbf{x}^*) \neq 0$, the vector $-\nabla f_0(\mathbf{x}^*)$ defines a *supporting hyperplane* to \mathcal{F} at \mathbf{x}^* .
letting $\mathbf{a} = -\nabla f_0(\mathbf{x}^*)$, this means that the hyperplane $\{\mathbf{y} \mid \mathbf{a}^\top \mathbf{y} = \mathbf{a}^\top \mathbf{x}^*\}$ is tangent to \mathcal{F} at \mathbf{x}^* and the halfspace $\{\mathbf{y} \mid \mathbf{a}^\top \mathbf{y} \leq \mathbf{a}^\top \mathbf{x}^*\}$ contains \mathcal{F} .

Geometric interpretation

Proposition

A point \mathbf{x}^* is a solution of (\mathcal{P}) if and only if $\mathbf{x}^* \in \mathcal{F}$ and

$$\nabla f_0(\mathbf{x}^*)^\top (\mathbf{y} - \mathbf{x}^*) \geq 0 \text{ for all } \mathbf{y} \in \mathcal{F}$$



Proof of optimality condition

\Leftarrow Let $\mathbf{x}^* \in \mathcal{F}$ such that the condition is satisfied. Then by convexity of f_0 , one has for $\mathbf{y} \in \mathcal{F}$, $f_0(\mathbf{y}) \geq f_0(\mathbf{x}^*) + \nabla f_0(\mathbf{x}^*)^\top (\mathbf{y} - \mathbf{x}^*) \geq f_0(\mathbf{x}^*)$. Thus \mathbf{x}^* is optimal.

\Rightarrow Suppose $\mathbf{x}^* \in \mathcal{F}$ is optimal but the conditions does not hold. There is some $\mathbf{y} \in \mathcal{F}$ for which $\nabla f_0(\mathbf{x}^*)^\top (\mathbf{y} - \mathbf{x}^*) < 0$. Consider $\mathbf{z}(t) = t\mathbf{y} + (1-t)\mathbf{x}^*$ for $t \in [0, 1]$. Since \mathcal{F} is convex, $\mathbf{z}(t) \in \mathcal{F}$. Moreover, observe that $f'_0(\mathbf{z}(t)) = \nabla f_0(\mathbf{z}(t))^\top (\mathbf{y} - \mathbf{x}^*)$ and thus for $t = 0$ one has $f'_0(\mathbf{z}(t))|_{t=0} = \nabla f_0(\mathbf{x}^*)^\top (\mathbf{y} - \mathbf{x}^*) < 0$. For small positive t , we thus have $f_0(\mathbf{z}(t)) \leq f_0(\mathbf{x}^*)$ which shows that \mathbf{x}^* is not optimal. Hence we have a contradiction.

Outline

- ① Optimality conditions for constrained convex problems
- ② Lagrangian duality
- ③ Karush-Kuhn-Tucker conditions

The Lagrangian for a general convex problem

$$\begin{aligned} & \text{minimize} && f_0(\mathbf{x}) \\ & \text{subject to} && f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, M \\ & && \mathbf{a}_i^\top \mathbf{x} - b_i = 0, \quad i = 1, \dots, P \end{aligned} \tag{\mathcal{P}}$$

The **Lagrangian of the problem** (\mathcal{P}) is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^M \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^P \nu_i (\mathbf{a}_i^\top \mathbf{x} - b_i)$$

where $\boldsymbol{\lambda} \in \mathbb{R}^M$ and $\boldsymbol{\nu} \in \mathbb{R}^P$ are Lagrange multiplier vectors.

- λ_i is the Lagrange multiplier associated to the i -th inequality constraint $f_i(\mathbf{x}) \leq 0$;
- ν_i is the Lagrange multiplier associated to the i -th equality constraint $\mathbf{a}_i^\top \mathbf{x} = b_i$

Lagrange dual function

Let $\mathcal{D} = \cap_{i=0}^M \mathcal{D}_i$ be the joint domain of the functions $f_i, i = 0, \dots, M$.
By minimizing over all $\mathbf{x} \in \mathcal{D}$, we get a very useful function, called the **(Lagrange) dual function**.

Lagrange dual function

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^M \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^P \nu_i (\mathbf{a}_i^\top \mathbf{x} - b_i) \right)$$

Comments

- g is concave, even if the problem (\mathcal{P}) is not convex (**Prove it!**)
- if L is unbounded below in \mathbf{x} , g takes on the value $-\infty$
- the dual function provides a lower bound of the optimal value p^* of the problem (\mathcal{P}) .

Lower bound property

For any $\boldsymbol{\lambda} \in \mathbb{R}^M$ such that $\boldsymbol{\lambda} \geq 0$ and any $\boldsymbol{\nu} \in \mathbb{R}^P$, we have

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^* = \inf \{ f_0(\mathbf{x}) \mid \mathbf{x} \in \mathcal{F} \}$$

Proof. Suppose that $\tilde{\mathbf{x}}$ is feasible and that $\boldsymbol{\lambda} \geq 0$. Hence $f_i(\tilde{\mathbf{x}}) \leq 0$ for $i = 1, \dots, M$ and $\mathbf{a}_i^\top \tilde{\mathbf{x}} - b_i = 0$ for $i = 1, \dots, P$. It results that $\sum_{i=1}^M \lambda_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^P \nu_i (\mathbf{a}_i^\top \tilde{\mathbf{x}} - b_i) \leq 0$. Therefore

$$L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^M \lambda_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^P \nu_i (\mathbf{a}_i^\top \tilde{\mathbf{x}} - b_i) \leq f_0(\tilde{\mathbf{x}})$$

Hence,

$$f_0(\tilde{\mathbf{x}}) \geq L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \geq \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu})$$

The inequality holds for all feasible $\tilde{\mathbf{x}}$ and in particular for $\tilde{\mathbf{x}} = \mathbf{x}^*$, one gets $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$.

Examples

For the two problems below, write the Lagrangian, Lagrange dual function and obtain a lower bound on the optimal value of the problem.

- ① minimum norm solution of least-squares equations

$$\begin{array}{ll}\text{minimize} & \|\mathbf{x}\|^2 \\ \text{subject to} & \mathbf{Ax} = \mathbf{b}, \quad \text{where } \mathbf{A} \in \mathbb{R}^{P \times N}, \mathbf{b} \in \mathbb{R}^P\end{array}$$

- ② linear program in standard form

$$\begin{array}{ll}\text{minimize} & \mathbf{c}^\top \mathbf{x} \\ \text{subject to} & \mathbf{Ax} = \mathbf{b}, \quad \text{where } \mathbf{A} \in \mathbb{R}^{P \times N}, \mathbf{b} \in \mathbb{R}^P \\ & \mathbf{x} \geq 0\end{array}$$

The Lagrange dual problem

Idea: exploit the lower bound property of the dual function to obtain the *best* lower bound possible on the original problem (called primal)

$$\begin{aligned} & \text{minimize} && f_0(\mathbf{x}) \\ & \text{subject to} && f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, M \\ & && \mathbf{a}_i^\top \mathbf{x} - b_i = 0, \quad i = 1, \dots, P \end{aligned} \tag{\mathcal{P}}$$

Lagrange dual problem

$$\begin{aligned} & \text{maximize} && g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ & \text{subject to} && \boldsymbol{\lambda} \geq 0 \end{aligned}$$

Comments

- a pair $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is said to be **dual feasible** if $\boldsymbol{\lambda} \geq 0$ and $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$
- dual optimal parameters are denoted by $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$
- the dual problem is always convex (since g is always concave) even if the primal problem isn't!

Weak and strong duality (I)

Let $d^* = \sup_{\lambda \geq 0} g(\lambda, \nu)$ be the optimal value of the dual Lagrange problem. By the lower bound property of g , we have the fundamental property.

Weak duality (always holds, even for nonconvex problems)

$$d^* \leq p^*$$

Comments

- works even if d^* or p^* are infinite. If primal problem is unbounded below, $p^* = -\infty$ and hence $d^* = -\infty$ (dual is infeasible); if the dual problem is unbounded above, $d = \infty$ and thus $p^* = \infty$ (primal is infeasible).
- the quantity $p^* - d^* \geq 0$ is called the duality gap

Weak and strong duality (II)

Strong duality (not always holds)

when $d^* = p^*$, we say that strong duality holds

Comments

- strong duality is equivalent to having zero duality gap;
- strong duality can be ensured by **constraint qualifications**.

Slater's constraint qualifications

If the primal problem is convex

$$\begin{array}{ll}\text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, M \\ & \mathbf{Ax} = \mathbf{b}\end{array}$$

and if there exists a feasible $\tilde{\mathbf{x}}$ s.t. $f_i(\tilde{\mathbf{x}}) < 0, i = 1, \dots, M$ and $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{b}$ (i.e. it is strictly feasible) then strong duality holds. In addition, when $d^* > -\infty$, the dual optimum is attained.

Examples

Using Slater's criterion, discuss the strong duality of the following optimization problems

- ① minimum norm solution of least-squares equations

$$\begin{aligned} & \text{minimize} \quad \|\mathbf{x}\|^2 \\ & \text{subject to} \quad \mathbf{Ax} = \mathbf{b}, \quad \text{where } \mathbf{A} \in \mathbb{R}^{P \times N}, \mathbf{b} \in \mathbb{R}^P \end{aligned}$$

- ② quadratic constrained quadratic program

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2} \mathbf{x}^\top \mathbf{Q}_0 \mathbf{x} - \mathbf{p}_0^\top \mathbf{x} \\ & \text{subject to} \quad \frac{1}{2} \mathbf{x}^\top \mathbf{Q}_1 \mathbf{x} - \mathbf{p}_1^\top \mathbf{x} + \mathbf{r}_1 \leq 0 \end{aligned}$$

where $\mathbf{Q}_0 > 0$ and $\mathbf{Q}_1 \geq 0$.

Saddle point interpretation

Consider the optimization problem without equality constraints

$$\begin{array}{ll}\text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, M\end{array}$$

with Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda}) = f_0(\mathbf{x}) + \sum_{i=1}^M \lambda_i f_i(\mathbf{x})$. Observe that

$$\sup_{\boldsymbol{\lambda} \geq 0} L(\mathbf{x}, \boldsymbol{\lambda}) = \sup_{\boldsymbol{\lambda} \geq 0} \left(f_0(\mathbf{x}) + \sum_{i=1}^M \lambda_i f_i(\mathbf{x}) \right) = \begin{cases} f_0(\mathbf{x}) & \text{if } \mathbf{x} \text{ is feasible} \\ \infty & \text{otherwise} \end{cases}$$

Proof. If \mathbf{x} is feasible, then $f_i(\mathbf{x}) \leq 0$ for all $i = 1, \dots, M$. Hence the optimal choice of $\boldsymbol{\lambda} = 0$. If \mathbf{x} is not feasible, then for some i , $f_i(\mathbf{x}) > 0$. Hence $L(\mathbf{x}, \boldsymbol{\lambda})$ is unbounded above as a function of $\boldsymbol{\lambda}$ and $\sup_{\boldsymbol{\lambda} \geq 0} L(\mathbf{x}, \boldsymbol{\lambda}) = \infty$.

Rewriting primal and dual optimal values

$$p^* = \inf_{\mathbf{x}} \sup_{\boldsymbol{\lambda} \geq 0} L(\mathbf{x}, \boldsymbol{\lambda}), \quad d^* = \sup_{\boldsymbol{\lambda} \geq 0} \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda})$$

Saddle point interpretation

When strong duality holds $p^* = d^*$ and thus

$$\inf_{\mathbf{x}} \sup_{\boldsymbol{\lambda} \geq 0} L(\mathbf{x}, \boldsymbol{\lambda}) = \sup_{\boldsymbol{\lambda} \geq 0} \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda})$$

i.e., the order of maximization/ minimization can be interchanged without changing the result.

This equality also characterizes the [saddle point property](#) of the Lagrangian $L : \mathbb{R}^N \times \mathbb{R}_+^M$; if \mathbf{x}^* and $\boldsymbol{\lambda}^*$ are primal and dual optimal (and strong duality holds) one has

$$L(\mathbf{x}^*, \boldsymbol{\lambda}) \leq L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}, \boldsymbol{\lambda}^*) \text{ for any } \mathbf{x} \text{ and } \boldsymbol{\lambda} \geq 0$$

i.e., $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ define a saddle point of the Lagrangian.

Conversely, if $(\mathbf{x}, \boldsymbol{\lambda})$ is a saddle-point of the Lagrangian, then \mathbf{x} is primal optimal, $\boldsymbol{\lambda}$ is dual optimal with zero duality gap.

Outline

- ① Optimality conditions for constrained convex problems
- ② Lagrangian duality
- ③ Karush-Kuhn-Tucker conditions**

Motivations

Provide necessary and sufficient conditions for optimality of the constrained convex problem

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, M \\ & \mathbf{a}_i^\top \mathbf{x} - b_i = 0, \quad i = 1, \dots, P \end{array} \quad (\mathcal{P})$$

with Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) := \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^M \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^P \nu_i (\mathbf{a}_i^\top \mathbf{x} - b_i) \right) \\ \text{subject to} & \boldsymbol{\lambda} \geq 0 \end{array}$$

goal: generalize the necessary and sufficient optimality condition of unconstrained convex problems $\nabla f(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x}$ is optimal.

Complementary slackness

Assumption: p^*, d^* are attained and $p^* = d^*$.

Let \mathbf{x}^* and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ be primal and dual optimal, respectively. Then

$$f_0(\mathbf{x}^*) = g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \quad (\text{strong duality})$$

$$= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^M \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^P \nu_i^* (\mathbf{a}_i^\top \mathbf{x} - b_i) \right) \quad (\text{def of } g)$$

$$\leq f_0(\mathbf{x}^*) + \sum_{i=1}^M \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^P \nu_i^* (\mathbf{a}_i^\top \mathbf{x}^* - b_i) \quad (\text{inf property})$$

$$\leq f_0(\mathbf{x}^*) \quad (\mathbf{x}^*, \boldsymbol{\lambda}^* \text{ are feasible})$$

We conclude that $\sum_{i=1}^M \lambda_i^* f_i(\mathbf{x}^*) = 0$ and since each term in the sum is negative, one has the **complementary slackness property**

$$\lambda_i^* f_i(\mathbf{x}^*) = 0, \quad i = 1, \dots, M$$

$$\text{i.e., } \lambda_i^* > 0 \Rightarrow f_i(\mathbf{x}^*) = 0 \Leftrightarrow f_i(\mathbf{x}^*) < 0 \Rightarrow \lambda_i^* = 0$$

KKT conditions for convex problems

Theorem (KKT - sufficient conditions)

Let (\mathcal{P}) be a convex differentiable optimization problem. If the following conditions (called KKT conditions) are satisfied:

Primal feasibility: $f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, M \quad \text{and} \quad \mathbf{Ax} = \mathbf{b}$

Dual feasibility: $\lambda_i \geq 0, \quad i = 1, \dots, M$

Complementary slackness: $\lambda_i f_i(\mathbf{x}) = 0, \quad i = 1, \dots, M$

Stationarity of the Lagrangian with respect to \mathbf{x} :

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = 0 \Leftrightarrow \nabla f_0(\mathbf{x}) + \sum_{i=1}^M \lambda_i \nabla f_i(\mathbf{x}) + \mathbf{A}^T \boldsymbol{\nu} = 0$$

then \mathbf{x} and $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ are primal and dual optimal, with zero-duality gap.

Remark

- If a convex problem satisfies Slater's condition, KKT conditions become **necessary and sufficient** for \mathbf{x} and $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ to be primal and dual optimal.

KKT conditions for convex problems

If $(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ satisfy KKT for a convex problem, then they are optimal.

- \mathbf{x} is primal feasible and $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ are dual feasible
- complementary slackness implies that $f_0(\mathbf{x}) = L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$
- stationarity of the Lagrangian with the fact that $L(\cdot, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is convex show that \mathbf{x} minimizes the Lagrangian over \mathbf{x} . Thus by definition $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$.
- hence $f_0(\mathbf{x}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. Strong duality follows and $(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ are optimal.

If Slater's condition is satisfied (and the problem is convex) then \mathbf{x} is optimal **if and only if** there exist $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ that, together with \mathbf{x} , satisfy the KKT conditions.

- Slater's implies strong duality and dual optimum is attained
- This generalizes the necessary and sufficient condition $\nabla f(\mathbf{x}) = 0$ for unconstrained problems

Using KKT conditions to solve optimization problems

Solve the following optimization problem using KKT conditions:

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \mathbf{x}^\top \mathbf{Q} \mathbf{x} - \mathbf{p}^\top \mathbf{x} + \mathbf{r} \\ & \text{subject to} && \mathbf{A} \mathbf{x} = \mathbf{b} \end{aligned}$$

where $\mathbf{Q} \geq 0$, $\mathbf{A} \in \mathbb{R}^{P \times N}$, $\mathbf{b} \in \mathbb{R}^P$ and $\text{rank } \mathbf{A} = P < N$.

Summary

Key notions

- Lagrangian, Lagrange dual function and Lagrange dual problem
- Lagrange dual problem is always convex, even if the primal problem is not
- Dual problem provides a lower bound on the optimal value of the primal problem, both are equal when strong duality holds
- For convex problems, KKT conditions give sufficient optimality conditions (and necessary and sufficient when Slater's condition holds)

Comments

- strong duality (almost always) holds for convex problems
- the dual problem may help to solve the primal problem very efficiently
- *duality* and *dual methods* are abundant in optimization and play a very important role in modern optimization; this course is just a rough sketch of the tip of the iceberg!