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 optimizations problems of the form

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

where we make the following assumptions:

- the objective $f_0 : \mathcal{D}_0 \subseteq \mathbb{R}^N \to \mathbb{R}$ 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

$$\Omega = \left\{ \mathbf{x} \in \mathbb{R}^N \;\middle|\; \begin{array}{lcl} f_i(\mathbf{x}) \leq 0 & , & i = 1, \dots, M \\ \mathbf{a}_\imath^\top \mathbf{x} = b_\imath & , & \imath = 1, \dots, P \end{array} \right\}$$

  where the functions $f_i : \mathbb{R}^N \to \mathbb{R}$ 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 $f : \mathcal{D} \subseteq \mathbb{R}^N \to \mathbb{R}$ 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 $\mathbf{x} \in \mathcal{D}$, $\nabla^2 f(\mathbf{x}) \succeq 0$, i.e. the Hessian is positive semidefinite.*

Consider the unconstrained convex optimization problem

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

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

Theorem (Optimality condition for unconstrained convex problems)

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

Terminology for constrained problems (I)

Constrained convex optimization problem

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

Feasible and infeasible problems

- a point $\mathbf{x}$ is said to be feasible if $x \in \Omega \cap \mathcal{D}_0$.
- 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 $\mathcal{F} = \varnothing$

Terminology for constrained problems (I)

Constrained convex optimization problem

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

Feasible and infeasible problems

- a point $\mathbf{x}$ is said to be feasible if $x \in \Omega \cap \mathcal{D}_0$.
- 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 $\mathcal{F} = \varnothing$

Active and inactive constraints

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

- 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

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

Optimal value

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

Terminology for constrained problems (II)

Constrained convex optimization problem

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

Optimal value

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

Optimal points

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

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

NB: Terminology is identical for non-convex problems.

First order characterization of optimal points

Constrained convex optimization problem

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

Assumption: f_0 is differentiable

Proposition

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

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

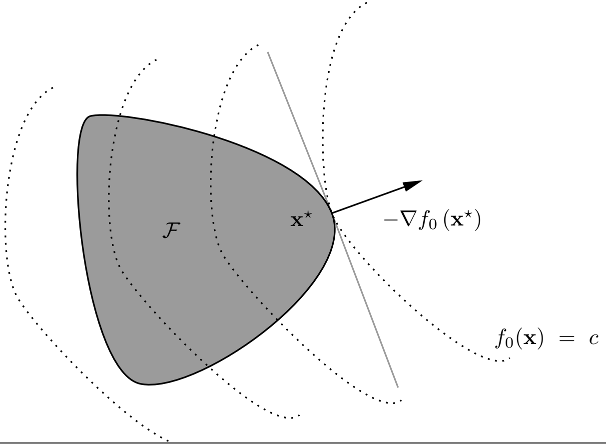
- generalizes the necessary and sufficient condition $(\nabla f_0(\mathbf{x}^\star) = 0 \Leftrightarrow \mathbf{x}^\star$ is solution) encountered for unconstrained convex problems;
- if $\nabla f_0(\mathbf{x}^\star) \neq 0$, the vector $-\nabla f_0(\mathbf{x}^\star)$ defines a *supporting hyperplane*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

Geometric interpretation

Proposition

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

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



Proof of optimality condition

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

$\boxed{\Rightarrow}$ Suppose $\mathbf{x}^\star \in \mathcal{F}$ is optimal but the conditions does not hold. There is some $\mathbf{y} \in \mathcal{F}$ for which $\nabla f_0(\mathbf{x}^\star)^\top(\mathbf{y} - \mathbf{x}^\star) < 0$.
Consider $\mathbf{z}(t) = t\mathbf{y} + (1-t)\mathbf{x}^\star$ for $t \in [0,1]$. Since $\mathcal{F}$ is convex, $\mathbf{z}(t) \in \mathcal{F}$.

Moreover, observe that $f_0'(\mathbf{z}(t)) = \nabla f_0(\mathbf{z}(t))^\top(\mathbf{y} - \mathbf{x}^\star)$ and thus for $t = 0$ one has $f_0'(\mathbf{z}(t))|_{t=0} = \nabla f_0(\mathbf{x}^\star)^\top(\mathbf{y} - \mathbf{x}^\star) < 0$.

For small positive t, we thus have $f_0(\mathbf{z}(t)) \leq f_0(\mathbf{x}^\star)$ which shows that $\mathbf{x}^\star$ is not optimal. Hence we have a contradiction.

Outline

The Lagrangian for a general convex problem

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

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 $f_i(\mathbf{x}) \le 0$;
- 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 $f_i, i = 0, \ldots, M$. 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 $p^\star$ 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

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

Proof. Suppose that $\tilde{\mathbf{x}}$ is feasible and that $\boldsymbol{\lambda} \geq 0$. Hence $f_i(\tilde{\mathbf{x}}) \leq 0$ for $i = 1, \ldots, M$ and $\mathbf{a}_i^\top \tilde{\mathbf{x}} - b_i = 0$ for $i = 1, \ldots, P$. 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 $\tilde{\mathbf{x}} = \mathbf{x}^\star$, one gets $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^\star$.

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

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

2. linear program in standard form

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

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)

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

Lagrange dual problem

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

Comments

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

Weak and strong duality (I)

Let $d^\star = \sup_{\boldsymbol{\lambda} \geq 0} g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ 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)

$$d^\star \leq p^\star$$

Comments

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

Weak and strong duality (II)

Strong duality (not always holds)

when $d^\star = p^\star$, 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

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, M \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

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

Examples

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

❶ minimum norm solution of least-squares equations

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

❷ quadratic constrained quadratic program

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

where $\mathbf{Q}_0 \succ 0$ and $\mathbf{Q}_1 \succeq 0$.

Saddle point interpretation

Consider the optimization problem without equality constraints

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

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

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

Proof. If $\mathbf{x}$ is feasible, then $f_i(\mathbf{x}) \le 0$ 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 $L(x, \boldsymbol{\lambda})$ is unbounded above as a function of $\boldsymbol{\lambda}$ and $\sup_{\boldsymbol{\lambda} \ge 0} L(\mathbf{x}, \boldsymbol{\lambda}) = \infty$.

Rewriting primal and dual optimal values

$$p^\star = \inf_{\mathbf{x}} \sup_{\boldsymbol{\lambda} \ge 0} L(\mathbf{x}, \boldsymbol{\lambda}), \qquad d^\star = \sup_{\boldsymbol{\lambda} \ge 0} \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda})$$

Saddle point interpretation

When strong duality holds $p^\star = d^\star$ 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 $\mathbf{x}^\star$ and $\boldsymbol{\lambda}^\star$ are primal and dual optimal (and strong duality holds) one has

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

i.e., $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$ 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

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

with Lagrange dual problem

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

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: $p^\star, d^\star$ are attained and $p^\star = d^\star$.

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

$$
\begin{aligned}
f_0(\mathbf{x}^\star) &= g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) && \text{(strong duality)} \\
&= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{M} \lambda_i^\star f_i(\mathbf{x}) + \sum_{i=1}^{P} \nu_i^\star (\mathbf{a}_i^\top \mathbf{x} - b_i) \right) && \text{(def of } g\text{)} \\
&\leq f_0(\mathbf{x}^\star) + \sum_{i=1}^{M} \lambda_i^\star f_i(\mathbf{x}^\star) + \sum_{i=1}^{P} \nu_i^\star (\mathbf{a}_i^\top \mathbf{x}^\star - b_i) && \text{(inf property)} \\
&\leq f_0(\mathbf{x}^\star) && (\mathbf{x}^\star, \boldsymbol{\lambda}^\star \text{ are feasible})
\end{aligned}
$$

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

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

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

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: $f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, M \quad$ and $\quad \mathbf{A}\mathbf{x} = \mathbf{b}$

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

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

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

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

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:

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

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

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!