A short review of convex analysis and optimization

Guillaume Obozinski

Ecole des Ponts - ParisTech

Master MVA 2014-2015

Review: convex analysis

Convex function

$$\forall\lambda \in [0,1], \qquad f(\lambda\,\mathbf{x} + (1-\lambda)\,\boldsymbol{y}) \leq \lambda\, f(\mathbf{x}) + (1-\lambda)\, f(\boldsymbol{y})$$

Strictly convex function

$$\forall\lambda \in \,]0,1[, \qquad f(\lambda\,\mathbf{x} + (1-\lambda)\,\boldsymbol{y}) < \lambda\, f(\mathbf{x}) + (1-\lambda)\, f(\boldsymbol{y})$$

Strongly convex function

$$\exists\mu > 0, \text{ s.t.} \quad \mathbf{x} \mapsto f(\mathbf{x}) - \mu\|\mathbf{x}\|^2 \quad \text{is convex}$$

Equivalently:

$$\forall\lambda \in [0,1], \quad f(\lambda\,\mathbf{x}+(1-\lambda)\,\boldsymbol{y}) \leq \lambda\, f(\mathbf{x})+(1-\lambda)\, f(\boldsymbol{y})-\mu\,\lambda(1-\lambda)\|\mathbf{x}-\boldsymbol{y}\|^2$$

The largest possible μ is called the strong convexity constant.

Minima of convex functions

Proposition (Supporting hyperplane)

If f is convex and differentiable at $\mathbf{x}$ *then*

$$f(\boldsymbol{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\boldsymbol{y} - \mathbf{x})$$

Convex function
All local minima are global minima.

Strictly convex function
If there is a local minimum, then it is unique and global.

Strongly convex function
There exists a unique local minimum which is also global.

Minima and stationary points of differentiable functions

Definition (Stationary point)

For f differentiable, we say that $\mathbf{x}$ is a stationary point if $\nabla f(\mathbf{x}) = 0$.

Theorem (Fermat)

If f is differentiable at $\mathbf{x}$ and $\mathbf{x}$ is a local minimum, then $\mathbf{x}$ is stationary.

Theorem (Stationary point of a convex differentiable function)

If f is convex and differentiable at $\mathbf{x}$ and $\mathbf{x}$ is stationary then $\mathbf{x}$ is a minimum.

Theorem (Stationary points of a twice differentiable functions)

For f twice differentiable at $\mathbf{x}$

- *if $\mathbf{x}$ is a local minimum then $\nabla f(\mathbf{x}) = 0$ and $\nabla^2 f(\mathbf{x}) \succeq 0$.*
- *conversely if $\nabla f(\mathbf{x}) = 0$ and $\nabla^2 f(\mathbf{x}) \succ 0$ then $\mathbf{x}$ is a strict local minimum.*

Brief review of Lagrange duality

Convex optimization problem with linear constraints

For

- f a convex function,
- $\mathcal{X} \subset \mathbb{R}^p$ a convex set included in the domain of f,
- $\mathbf{A} \in \mathbb{R}^{n \times p}, \quad \mathbf{b} \in \mathbb{R}^n$,

$$\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b} \tag{P}$$

Lagrangian

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

with $\boldsymbol{\lambda} \in \mathbb{R}^n$ the *Lagrange multiplier*.

Properties of the Lagrangian

Link between primal and Lagrangian

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^n} L(\mathbf{x}, \boldsymbol{\lambda}) = \begin{cases} f(\mathbf{x}) & \text{if } \mathbf{Ax} = \mathbf{b} \\ +\infty & \text{otherwise.} \end{cases}$$

So that

$$\min_{\mathbf{x} \in \mathcal{X}:\, \mathbf{Ax}=\mathbf{b}} f(\mathbf{x}) = \min_{\mathbf{x} \in \mathcal{X}} \max_{\boldsymbol{\lambda} \in \mathbb{R}^n} L(\mathbf{x}, \boldsymbol{\lambda})$$

Lagrangian dual objective function

$$g(\boldsymbol{\lambda}) = \min_{\mathbf{x} \in \mathcal{X}} L(\mathbf{x}, \boldsymbol{\lambda})$$

Dual optimization problem

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^n} g(\boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda} \in \mathbb{R}^n} \min_{\mathbf{x} \in \mathcal{X}} L(\mathbf{x}, \boldsymbol{\lambda}) \qquad (D)$$

Maxmin-minmax inequality, weak and strong duality

For any $f : \mathbb{R}^n \times \mathbb{R}^m$ and any $w \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$, we have

$$\boxed{\max_{z \in Z} \min_{w \in W} f(w, z) \leq \min_{w \in W} \max_{z \in Z} f(w, z).}$$

Weak duality

$$d^* := \max_{\boldsymbol{\lambda} \in \mathbb{R}^n} g(\boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda}} \min_{\mathbf{x} \in \mathcal{X}} L(\mathbf{x}, \boldsymbol{\lambda}) \leq \min_{\mathbf{x}\mathbf{x} \in \mathcal{X}} \max_{\boldsymbol{\lambda} \in \mathbb{R}^n} L(\mathbf{x}, \boldsymbol{\lambda}) = \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) =: p^*$$

So that in general, we have $d^* \leq p^*$. This is called weak duality

Strong duality

In some cases, we have strong duality:

- $d^* = p^*$
- Solutions to (P) and (D) are the same

Slater's qualification condition

Slater's qualification condition is a condition on the constraints of a convex optimization problem that guarantees that strong duality holds.

For linear constraints, Slater's condition is very simple:

Slater's condition for a cvx opt. pb with lin. constraints

If there exists an $\mathbf{x}$ *in the relative interior of* $\mathcal{X} \cap \{\mathbf{Ax} = \mathbf{b}\}$ *then strong duality holds.*