KKT conditions, Lagrangian and duality

Alexandre Gramfort

alexandre.gramfort@inria.fr

Master 2 Data Science, Univ. Paris Saclay
Optimisation for Data Science

Constrained Convex optimization problem

Definition (Convex optimization problem)

$$(\mathcal{P}) : \min f(x), x \in \mathcal{C} \subset \mathbb{R}^n$$

with f convex, $\mathcal{C} = \{x / h(x) = 0, g(x) \preceq 0\}$, where

1. $h = [h_1 \dots h_p]$ are affine functions
2. $g = [g_1 \dots g_q]$ are convex functions.

Remark: equality constraints must be affine and inequality constraints must be convex but one can have many.

Example

$\min_{x \in \mathbb{R}^n} \|b - Ax\|_2^2, \mathrm{s.t.} \|x\|_2 \leq 1$

Lagrangian

Definition (Lagrangian)

- We call the Lagrangian of problem $(\mathcal{P})$ the function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p \times (\mathbb{R}^+)^q \to \mathbb{R}$ defined by

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{p} \lambda_i h_i(x) + \sum_{j=1}^{q} \mu_j g_j(x)$$

Remark: λ and μ are commonly referred to as Lagrange multipliers (or dual variables, cf. below).

Remark: What is the Lagrangian for our constrained least square problem?

KKT conditions

Definition (Karush, Kuhn et Tucker (KKT) conditions)

- $h(x) = 0$ and $g(x) \leq 0$ (primal constraints)
- $\mu \in \mathbb{R}_+^q$ (dual constraints)
- $\forall j \in \{1, \dots, q\}$, $\mu_j g_j(x) = 0$ (complementarity)
- $\nabla f(x) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x) + \sum_{j=1}^{q} \mu_j \nabla g_j(x) = 0$ (stationarity)

Optimality Conditions (convex case)

Theorem (First order necessary and sufficient conditions (KKT))

- *Let us suppose that:*
  - *f, h and g are $\mathcal{C}^1$ (smooth),*
  - *f and g are convex and h is affine,*
  - *problem is stricly feasible (cf. Slater's conditions).*
- *Then x^* is solution of $(\mathcal{P})$ iff $\exists \lambda^* \in \mathbb{R}^p$ and $\mu^* \in \mathbb{R}_+^q$ such that KKT conditions are satisfied.*

Remark: Write the KKT conditions for our constrained least square problem. Express the Lagrange multiplier $\mu \in \mathbb{R}_+$ as a function of x.

Duality

Definition (Saddle point)

A saddle point of $\mathcal{L}$ is a triplet (x^*, λ^*, μ^*) such that

$$\forall x, \lambda, \mu,\ \mathcal{L}(x^*, \lambda, \mu) \leq \mathcal{L}(x^*, \lambda^*, \mu^*) \leq \mathcal{L}(x, \lambda^*, \mu^*)$$

Duality

Definition (Saddle point)

A saddle point of $\mathcal{L}$ is a triplet (x^*, λ^*, μ^*) such that

$$\forall x, \lambda, \mu, \; \mathcal{L}(x^*, \lambda, \mu) \leq \mathcal{L}(x^*, \lambda^*, \mu^*) \leq \mathcal{L}(x, \lambda^*, \mu^*)$$

Theorem (Property of saddle points for a convex problem)

- *We suppose that f, h and g are $\mathcal{C}^1$ (smooth), f, g are convex and h is affine.*
- *The triplet (x^*, λ^*, μ^*) is a saddle point of $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p \times (\mathbb{R}^+)^q \rightarrow \mathbb{R}$ iff it satisfies the KKT conditions:*
    - $h(x^*) = 0$ *and* $g(x^*) \leq 0$
    - $\forall j \in \{1 \dots q\}$, $\mu_j^* \geq 0$ *and* $\mu_j^* g_j(x^*) = 0$
    - $\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{q} \mu_j^* \nabla g_j(x^*) = 0$

Dual problem

- dual function

$$g(\lambda, \mu) = \inf_x \mathcal{L}(x, \lambda, \mu)$$

- for $\mu_j \geq 0, \forall j$

$$g(\lambda, \mu) \leq \inf_{x \in \mathcal{C}} f(x)$$

- dual problem $(\mathcal{D})$ associated to the primal problem $(\mathcal{P})$

$$\max_{\lambda \in \mathbb{R}^p, \mu \in \mathbb{R}_+^q} g(\lambda, \mu)$$

- duality gap : $\inf_{x \in \mathcal{C}} f(x) - \max_{\mu \geq 0} g(\lambda, \mu)$
- If duality gap at optimum is 0 we say that "strong duality holds".

Remark: Write in the dual to solve our constrained least square problem .

References

- Convex Optimization, Stephen Boyd and Lieven Vandenberghe, Cambridge University Press
- `https://web.stanford.edu/~boyd/cvxbook/`
- Read chapter 5.