Recitation 12

Carles Domingo

Unconstrained optimization

We want to minimize a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$.
Remember: We say that $x \in \mathbb{R}^n$ is

- a critical point of f if $\nabla f(x) = 0$.
- a global minimizer of f if $\forall x' \in \mathbb{R}^n$, $f(x) \leq f(x')$.
- a local minimizer of f if there exists $\delta > 0$ such that for all $x' \in B(x, \delta)$, $f(x) \leq f(x')$, where $B(x, \delta) = \{x' | \|x' - x\| \leq \delta\}$.

Theorem (First order necessary conditions)

Let $x \in \mathbb{R}^n$ be a point at which f is differentiable. Then,

$$x \text{ is a local minimizer of } f \implies \nabla f(x) = 0$$

Unconstrained optimization

Theorem (Second order sufficient conditions)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice-differentiable function. Let $x \in \mathbb{R}^n$ be a critical point of f, i.e. $\nabla f(x) = 0$. Then,

- *If $H\,f(x)$ is positive definite (that is, if all the eigenvalues of $H\,f(x)$ are strictly positive), then x is a local minimizer of f.*
- *If $H\,f(x)$ is negative definite (that is, if all the eigenvalues of $H\,f(x)$ are strictly negative), then x is a local maximizer of f.*
- *If $H\,f(x)$ admits strictly positive eigenvalues and strictly negative eigenvalues, then x is neither a local maximum nor a local minimum. We call x a saddle point.*

Unconstrained optimization

The theorem in the previous slide begs for a study of the case in which $x \in \mathbb{R}^n$ is a critical point of f such that $Hf(x)$ is positive semidefinite (and similarly negative semidefinite).

1. Give an example of a twice-differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ with a local minimizer x such that $Hf(x)$ is positive semidefinite.
2. Give an example of a twice-differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ with a critical point x of f that is not a local minimizer and such that $Hf(x)$ is positive semidefinite.

Unconstrained optimization

1. Give an example of a twice-differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ with a local minimizer x such that $H f(x)$ is positive semidefinite.

Unconstrained optimization

2. Give an example of a twice-differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ with a critical point x of f that is not a local minimizer and such that $Hf(x)$ is positive semidefinite.

Unconstrained optimization

2. Give an example of a twice-differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ with a critical point x of f that is not a local minimizer and such that $Hf(x)$ is positive semidefinite.

Constrained optimization

Remember: The problem we consider is

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m, \\ & h_i(x) = 0, \quad i = 1, \ldots, p. \end{aligned} \tag{1}$$

with a variable $x \in \mathbb{R}^n$.

Theorem (KKT necessary conditions)

Assume that the functions $f, g_1, \ldots, g_m, h_1, \ldots, h_p$ *are continuously differentiable. Assume that* x *is a solution of* (1) *with* $\{\nabla g_i(x) | g_i(x) = 0\} \cup \{\nabla f_i(x) | i \in \{1, \ldots, p\}\}$ *a linearly independent set of vectors. Then there exists* $\lambda_1, \ldots, \lambda_m \geq 0$ *and* $\nu_1, \ldots, \nu_p \in \mathbb{R}$ *such that:*

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

and for all $i \in \{1, \ldots, m\}$, $\lambda_i = 0$ *if* $g_i(x) < 0$.

Constrained optimization

Using the KKT necessary conditions, find the minimum and the minimizer(s) of the following problem:

$$\begin{aligned} \text{minimize} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & 4 - (x_1 + 1)^2 - x_2^2 \leq 0 \end{aligned}$$

Constrained optimization

Using the KKT necessary conditions, find the minimum and the minimizer(s) of the following problem:

$$\begin{aligned} \text{minimize} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & 4 - (x_1 + 1)^2 - x_2^2 \leq 0 \end{aligned}$$

Constrained optimization

Using the KKT necessary conditions, find the minimum and the minimizer(s) of the following problem:

$$\begin{aligned} \text{minimize} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & 4 - (x_1 + 1)^2 - x_2^2 \leq 0 \end{aligned}$$

Constrained optimization

Definition (Lagrangian)

The Lagrangian of problem (1) is defined as:

$$L(x,\mu,\nu) = f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x)$$

Definition (Lagrange dual function and dual problem)

The Lagrange dual function is defined as

$$\ell(\mu,\nu) = \inf_{x\in\mathbb{R}^n} L(x,\mu,\nu)$$

and the dual problem is defined as

$$\begin{array}{ll} \text{maximize} & \ell(\mu,\nu) \\ \text{subject to} & \lambda_i \geq 0, \quad i = 1,\dots,m \\ & \nu_i \in \mathbb{R}, \quad i = 1,\dots,p. \end{array} \tag{2}$$

Constrained optimization

Theorem (Weak duality)

The optimal value $p^\star$ of the primal problem (1) *is larger or equal than the optimal value $d^\star$ of the dual problem* (2)*:*

$$d^\star = \sup_{\lambda \geq 0, \nu} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu) \leq \inf_{x \in \mathbb{R}^n} \sup_{\lambda \geq 0, \nu} L(x, \lambda, \nu) = p^\star$$

Theorem (Slater's condition and strong duality)

We say that the problem (1) *verifies Slater's condition if there exists a feasible point x such that $g_i(x) < 0$ for all $i \in \{1, ..., m\}$. If the problem* (1) *is convex and verifies Slater's condition, then strong duality holds: $p^\star = d^\star$. Moreover if $p^\star = d^\star$ is finite then the optimal value of the dual problem is attained at some $(\lambda, \nu) \in \mathbb{R}^m_{\geq 0} \times \mathbb{R}^p$.*

Constrained optimization

Theorem (KKT necessary and sufficient conditions)

Assume that the functions $f, g_1, \dots, g_m$ *are convex, differentiable and that* $h_1, \dots, h_p$ *are affine. Assume that strong duality holds, that* $p^\star = d^\star$ *is finite, and that the optimal value of the dual problem is attained at some* $(\lambda, \nu) \in \mathbb{R}^m_{\geq 0} \times \mathbb{R}^p$*. (This is for instance the case under Slater's condition).*
Then $x \in \mathbb{R}^n$ *is a solution of* (1) *if and only if* x *is feasible and*

$$\begin{cases} \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{i=1}^p \nabla h_i(x) = 0 & \textit{(stationarity)} \\ \lambda_i g_i(x) = 0 \textit{ for all } i = 1, \dots, m & \textit{(complementarity)} \end{cases}$$

Duality

Consider the primal problem of minimizing $f_0(x) = \frac{1}{2}x^\top Qx$ subject to $h(x) = Ax - b$, where Q is a symmetric positive definite $n \times n$ matrix and A is $m \times n$, with $m \leq n$, with full rank m (in other words, A has m linearly independent rows. There are no inequality constraints.

1. Write down the Lagrangian $L(x, \nu)$.
2. Since $L(x, \nu)$ is convex, differentiable and bounded below in x, set its gradient to zero to find its minimizer and write down a formula for the Lagrange dual function $g(\nu) = \inf_x L(x, \nu)$ (as $\inf$ can be replaced by $\min$, in this case).
3. Find the maximizer ν^* of the Lagrange dual function $g(\nu)$ (which is concave) by setting its gradient to zero. What is the dual optimal value $d^* = g(\nu^*)$?
4. Find the associated $\hat{x}$ attaining the minimizer of the Lagrangian $L(x, \nu^*)$.

Duality

Consider the primal problem of minimizing $f_0(x) = \frac{1}{2}x^\top Qx$ subject to $h(x) = Ax - b$, where Q is a symmetric positive definite $n \times n$ matrix and A is $m \times n$, with $m \leq n$, with full rank m (in other words, A has m linearly independent rows. There are no inequality constraints.

1. Write down the Lagrangian $L(x, \nu)$.
2. Since $L(x, \nu)$ is convex, differentiable and bounded below in x, set its gradient to zero to find its minimizer and write down a formula for the Lagrange dual function $g(\nu) = \inf_x L(x, \nu)$ (as $\inf$ can be replaced by $\min$, in this case).
3. Find the maximizer ν^* of the Lagrange dual function $g(\nu)$ (which is concave) by setting its gradient to zero. What is the dual optimal value $d^* = g(\nu^*)$?
4. Find the associated $\hat{x}$ attaining the minimizer of the Lagrangian $L(x, \nu^*)$.

Duality

5. Check whether $\hat{x}$ is feasible for the primal problem (whether it satisfies $Ax = b$).
6. Find the primal value $f_0(\hat{x})$. If $\hat{x}$ is primal feasible, then the optimal primal value $p^* \leq f_0(\hat{x})$.
7. Do you conclude that there is no duality gap, i.e., that $d^* = p^*$?
8. Is Slater's condition verified?

Duality

5. Check whether $\hat{x}$ is feasible for the primal problem (whether it satisfies $Ax = b$).
6. Find the primal value $f_0(\hat{x})$. If $\hat{x}$ is primal feasible, then the optimal primal value $p^* \leq f_0(\hat{x})$.
7. Do you conclude that there is no duality gap, i.e., that $d^* = p^*$?
8. Is Slater's condition verified?

Duality

1. Write down the Lagrangian $L(x, \nu)$.

Duality

2. Since $L(x, \nu)$ is convex, differentiable and bounded below in x, set its gradient to zero to find its minimizer and write down a formula for the Lagrange dual function $g(\nu) = \inf_x L(x, \nu)$ (as $\inf$ can be replaced by $\min$, in this case).

Duality

3. Find the maximizer ν^* of the Lagrange dual function $g(\nu)$ (which is concave) by setting its gradient to zero. What is the dual optimal value $d^* = g(\nu^*)$?

Duality

4. Find the associated $\hat{x}$ attaining the minimizer of the Lagrangian $L(x, \nu^*)$.

Duality

5. Check whether $\hat{x}$ is feasible for the primal problem (whether it satisfies $Ax = b$).

Duality

6. Find the primal value $f_0(\hat{x})$. If $\hat{x}$ is primal feasible, then the optimal primal value $p^* \leq f_0(\hat{x})$.

Duality

7. Do you conclude that there is no duality gap, i.e., that $d^* = p^*$?

Duality

8. Is Slater's condition verified?

Penalized & constrained problems

For $\lambda > 0$, consider the ridge regression problem we saw in the previous recitation:

$$\min_{x \in \mathbb{R}^n} \{\|Ax - y\|^2 + \lambda \|x\|^2\} \tag{3}$$

For $t > 0$, consider as well the constrained optimization problem

$$\begin{aligned} \min \quad & \|Ax - y\|^2 \\ \text{st} \quad & \|x\|^2 \leq t \end{aligned} \tag{4}$$

Use the KKT conditions to show that when $y \notin \text{Im}(A)$, the two problems are "equivalent" in the following sense: for all $\lambda > 0$ any solution of (3) is a solution of (4) for some $t > 0$, and vice versa, for all $t > 0$ any solution of (4) is a solution of (3) for some $\lambda > 0$.