Recitation #11

Irina Espejo (iem244@nyu.edu)

Center for Data Science

DS-GA 1014: Optimization and Computational Linear Algebra for Data Science

Announcement

Tomorrow is Thanksgiving so there will be no office hours.
Next office hours will be on Monday 20th at 8AM EST.
Recitation on Thursday 3rd at 8AM EST will proceed as usual.

recall: unconstrained optimization

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

- a **critical point** of f if $\nabla f = 0$
- a **global minimizer** of f if for all $x' \in \mathbb{R}^n$ it holds that $f(x) \leq f(x')$
- a **local minimizer** of f if there exists $\delta > 0$ such that for all $x' \in B(x, \delta)$ it holds that $f(x) \leq f(x')$.
  Note that: $B(x', \delta) = \{x' \mid \|x' - x\| \leq \delta\}$ are the all the elements inside the ball centered at x with radius δ

Theorem: First order necessary conditions

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

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

Unconstrained optimization

Theorem: Second order sufficient conditions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function and let $x \in \mathbb{R}^n$ be a critical point of f (that is $\nabla f(x) = 0$). Then,

- If $H_{f(x)}$ is **positive definite** (*all its eigenvalues are strictly positive*), then x is a local **minimizer** of f
- If $H_{f(x)}$ is **negative definite** (*all its eigenvalues are strictly negative*), then x is a local **maximizer** of f
- If $H_{f(x)}$ is **admits positive and negative eigenvalues**, then x is neither a local minimizer nor a local maximizer of f. We call x a **saddle point**

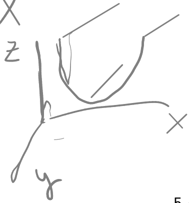
practice: unconstrained optimization

Exercise 1

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

1 $H_{f(x)} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ $f(x,y) = \frac{1}{2}x^2$



Exercise 1

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

2 $\quad H_{f(x)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad f(x,y) = xy$

recall: constrained optimization

The problem with constrained optimization is

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

recall: constrained optimization

Theorem KKT: necessary conditions

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

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

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

Exercise 2

Using the KKT necessary conditions, find the minimum and the minimizers of the following constrained optimization problem

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

practice: constrained optimization

Exercise 2

Using the KKT necessary conditions, find the minimum and the minimizers of the following constrained optimization problem

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

recall: Lagrangian

Lagrangian

The Lagrangian of the constrained optimization problem is defined as:

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

Lagrange dual function

The Lagrange dual function of the constrained optimization problem is defined as:

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

and the dual problem is defined as

$$\begin{aligned} \text{maximize} \ \ & \ell(\lambda, \nu) & \\ \text{subject to} \ \ & \lambda_i(x) \leq 0 & i = 1, .., m \\ & \nu_i(x) \in \mathbb{R} & i = 1, .., p \end{aligned}$$

recall: important theorems

Theorem of weak duality

The optimal value $p^\star$ of the original (primal) constrained problem is larger or equal than the optimal value $d^\star$ of the dual problem.

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

Theorem of Slater's condition and strong duality

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

recall: constrained optimization

KKT necessary and sufficient conditions

Assume:

- functions $f, g_1, .., g_m$ are convex and differentialble and $h_1, .., h_p$ are affine.
- strong duality holds: $p^\star = d^\star$ is finite
- optimal value of the dual problem is attained at some $(\lambda, \nu) \in {\mathbb{R}_+}^m \times \mathbb{R}^p$ (like in Slater's condition)

Then, $x \in \mathbb{R}^n$ is a solution of the primal problem if and only if x is feasible and

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

$$\lambda_i g_i(x) = 0 \quad \text{for all } i = 1..m \text{ (complementarity)}$$

Exercise 3

Consider the primal problem of minimizing $f_0(x) = \frac{1}{2}x^T Qx$ subject to $h(x) = Ax - b = 0$, where Q is a symmetric positive definite $n \times n$ matrix and A is $m \times m$ with full rank m and $m \leq n$. No inequality constraints.

1. Write down the Lagrangian $\mathscr{L}(x, \lambda, \nu)$.

$$\mathcal{L} = \nabla f_0(x) + \nu \nabla h(x)$$

$$= (Q + Q^T)x + \nu A^T$$

Exercise 3

Consider the primal problem of minimizing $f_0(x) = \frac{1}{2}x^T Qx$ subject to $h(x) = Ax - b = 0$, where Q is a symmetric positive definite $n \times n$ matrix and A is $m \times m$ with full rank m and $m \leq n$. No inequality constraints.

2. Since $\mathscr{L}(x, \lambda, \nu)$ is convex, differentiable and bounded below x, set its gradient to 0 to find its minimizer and write down a formula for the Lagrange dual function $\ell(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda, \nu)$ (take the inf as a min)

$$(Q+Q^T)x = -\nu A^T$$

$$x^* = -(Q+Q^T)^{-1}\nu A^T$$

Exercise 3

$$\ell(\lambda, \upsilon) = \min_{x \in \mathbb{R}^n} (Q+Q^T)x + \upsilon A^T$$

subject to $\upsilon \in \mathbb{R}^m$