MATHEMATICS OF DATA SCIENCE

November 7, 2023

Today: Last session on convex optimization
Next up: statistics for data science

SOLVING OPTIMIZATION PROBLEMS (Pt 2)

① Existence principles

② Duality

Domain of (P):
$\mathcal{X} = \{x \in \mathbb{R}^n \mid f(x) < \infty,\ g_i(x) < \infty\ \forall i = 1..m,\ h_i(x) < \infty\ \forall i = 1..\ell\}$

$$(P)\begin{cases} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \quad \text{s.t.} \quad g_i(x) \leq 0 \quad i = 1..m \\ & \qquad\qquad\qquad\ h_i(x) = 0 \quad i = 1..\ell \end{cases}$$

$$f, g_i, h_i : \mathbb{R}^n \longrightarrow (-\infty, +\infty] \qquad g = [g_i]_{i=1..m} \qquad h = [h_i]_{i=1..\ell}$$

The Lagrangian associated with (P) is the function

$$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^\ell \longrightarrow \overline{\mathbb{R}}$$

$$(x, \lambda, \mu) \longmapsto \mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x)$$

$$= f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{\ell} \mu_i h_i(x)$$

Property: Suppose that $x \in \mathbb{R}^n$ is feasible for (P). Then

$$\mathcal{L}(x, \lambda, \mu) \leq f(x) \qquad \forall \lambda \in \mathbb{R}^m_+ \ \big) \Rightarrow \{\lambda_i \geq 0\}_{i=1..m}$$
$$\forall \mu \in \mathbb{R}^\ell$$

Definitions

- Primal function associated with (P)

$$p : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$$

$$x \longmapsto \begin{cases} \sup_{(\lambda,\mu) \in \mathcal{Y}} \mathcal{L}(x, \lambda, \mu) & \text{if } x \in \mathcal{X} \\ +\infty & \text{if } x \notin \mathcal{X} \end{cases}$$

optimal value of a maximization problem

where $\mathcal{X}$ is the domain of (P)

and $\mathcal{Y} = \{ (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^\ell \mid \lambda_i \geq 0 \quad i=1..m \}$

↳ If (P) is feasible (feasible set not empty), then (P) is equivalent to $\underset{x \in \mathbb{R}^n}{\text{minimize}}\ p(x)$

⊕ The new formulation is unconstrained

⊖ the new objective function has a complicated expression (each value $p(x)$ is the optimal value of an optimization problem)

• Dual function (of problem (P))

$$d : \mathbb{R}^m \times \mathbb{R}^\ell \longrightarrow \bar{\mathbb{R}}$$
$$(\lambda, \mu) \longmapsto d(\lambda, \mu) = \begin{cases} \inf_{x \in \mathcal{X}} \mathcal{L}(x, \lambda, \mu) & \text{if } (\lambda, \mu) \in \mathcal{Y} \\ -\infty & \text{if } (\lambda, \mu) \notin \mathcal{Y} \end{cases}$$

↳ The function d is always concave!

Def: Dual optimization problem

The dual problem of (P) is

$$(D) \quad \underset{\substack{\lambda \in \mathbb{R}^m \\ \mu \in \mathbb{R}^\ell}}{\text{maximize}}\ d(\lambda, \mu) \quad \text{s.t.} \quad \lambda \geq 0 \qquad (\Leftrightarrow \lambda_i \geq 0 \quad i=1..m)$$

⇒ This is a convex optimization problem

(Reformulation: $\underset{\substack{\lambda \in \mathbb{R}^m \\ \mu \in \mathbb{R}^\ell}}{\text{minimize}}\ -d(\lambda, \mu) \quad \text{s.t.} \quad -\lambda_i \leq 0 \quad i=1..m$

The dual optimal value of (P) is the optimal value of (D), that is

$$d^* = \sup \{ d(\lambda,\mu) \mid \lambda \geq 0 \}$$

$\mathbb{R}^m \times \mathbb{R}^l \ni (\lambda,\mu)$ is dual feasible if $\lambda \geq 0$ $\Rightarrow$ Feasible set always nonempty

Example: Linear programming

Canonical form $\begin{aligned} Ax-b&=0 \\ -x&\leq 0 \end{aligned}$

$$\underset{x\in\mathbb{R}^m}{\text{minimize}} \quad c^T x \quad \text{s.t.} \quad \boxed{\begin{aligned} Ax&=b \\ x&\geq 0 \end{aligned}}$$

$A \in \mathbb{R}^{l\times m}$
$b \in \mathbb{R}^l$
$c \in \mathbb{R}^n$

$$\mathcal{L}(\underset{\in\mathbb{R}^m}{x,\lambda},\underset{\in\mathbb{R}^l}{\mu}) = c^T x - \lambda^T x + \mu^T(Ax-b)$$

Dual function: $d(\lambda,\mu) = \begin{cases} -b^T\mu & \text{if } A^T\mu - \lambda + c = 0 \text{ and } \lambda \geq 0 \\ -\infty & \text{otherwise} \end{cases}$

by definition

The dual problem can be reformulated as

$$\underset{\lambda\in\mathbb{R}^m,\mu\in\mathbb{R}^l}{\text{maximize}} \quad -b^T\mu \quad \text{s.t.} \quad \begin{aligned} A^T\mu - \lambda + c &= 0 \\ \lambda &\geq 0 \end{aligned}$$

and this a convex optimization problem, equivalent to

$$\underset{\substack{\lambda\in\mathbb{R}^m \\ \mu\in\mathbb{R}^l}}{\text{minimize}} \quad b^T\mu \quad \text{s.t.} \quad \begin{aligned} -\lambda &\leq 0 \\ A^T\mu - \lambda + c &= 0 \end{aligned}$$

convex optimization problem in standard form

Q) Link between the dual problem and (P)?

Th) **Weak duality**

For any (x, λ, μ) such that $x \in \mathcal{X}$ and $(\lambda, \mu) \in \mathcal{Y}$,

$$d(\lambda, \mu) \leq p(x)$$

implying that

optimal value for the primal problem

$$d^* \leq p^* = \inf_{x \in \mathbb{R}^n} \{ p(x) \}$$

$$= \inf_{x \in \mathbb{R}^n} \{ f(x) \mid g(x) \leq 0, h(x) = 0 \}$$

when the feasible set is not empty

optimal value of (P)

Solving the dual gives an underestimate of the optimal value of (P)

NB: If $d^* = +\infty$, then the primal problem has an empty domain

⚠ In general, $d^* < p^*$ (duality gap: $p^* - d^*$)

So the dual is an approximation of the problem, not a reformulation (but it is a convex approximation, hence easier to analyze)

↳ When (P) is convex, then under mild conditions we have $d^* = p^*$, and the dual is a reformulation of (P)

Ex) If all constraints in (P) are linear, then (D) is a reformulation of (P)

Ex) If the problem is a convex optimization problem in standard form

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; f(x) \quad \text{s.t.} \quad \begin{array}{l} g_i(x) \leq 0 \quad i = 1..m \\ a_i^T x - b_i = 0 \quad i = 1..l \end{array}$$

with $f, g_1, \ldots, g_m$ convex, such that

$\exists\, x \in \mathbb{R}^n$ satisfying
- $g_i(x) < 0 \quad \forall i = 1..m$
- $a_i^T x - b_i = 0$
- $f(x) < \infty$

then $d^* = p^*$ also holds.

When $d^* = p^*$, we say that strong duality holds.

3 KKT conditions

Consider again problem (P) and suppose that $f, g_1, \ldots, g_m, h_1, \ldots, h_\ell$ differentiable on $\mathbb{R}^n$ (for simplicity).

$\Rightarrow$ domain of (P) is $\mathbb{R}^n$

Suppose also that strong duality holds.

Then if $x^* \in \mathbb{R}^n$ is a solution of (P), then there exist $(\lambda^*, \mu^*) \in \mathbb{R}^m \times \mathbb{R}^\ell$ that is a solution of (D) and we have

KKT conditions
$$\left\{\begin{array}{ll}
\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^{\ell} \mu_i^* \nabla h_i(x^*) = 0 & \\
g_i(x^*) \leq 0 & \forall i = 1..m \\
h_i(x^*) = 0 & \forall i = 1..\ell \\
\lambda_i^* \geq 0 & \forall i = 1..m \\
g_i(x^*)\, \lambda_i^* = 0 & \forall i = 1..m
\end{array}\right.$$

If (P) is convex, x^* is a solution of (P) if and only if there exist (λ^*, μ^*) such that the KKT conditions hold.

Application: Given $(A, b) \in \mathbb{R}^{m \times n} \times \mathbb{R}^m$, find $x \in \mathbb{R}^n$ such that $Ax = b$

if $n > m$ and $\text{rank}(A) = m$, there exist infinitely many values of x that satisfy $Ax = b$

$\Rightarrow$ In practice, we like solutions that are not too sensitive to changes in the data. One type of such solutions is the minimal-norm solution, obtained by solving

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \underbrace{\frac{1}{2}\|x\|^2}_{\text{convex function}} \quad \text{s.t.} \ \underbrace{Ax = b}_{\text{convex constraints}}$$

$\Rightarrow$ Convex optimization problem

$\Rightarrow$ Linear constraints, so strong duality holds

$\Rightarrow$ KKT equations

$$x^* \text{ is a solution} \iff \exists \mu^* \in \mathbb{R}^m, \quad \begin{aligned} Ax^* &= b \\ x^* + A^T\mu^* &= 0 \end{aligned}$$

$$\iff x^* = A^T(AA^T)^{-1}b$$

$\Rightarrow$ The KKT equations give the solution and μ^* indicates the sensitivity of x^* with respect to each constraint