Convex Optimization & Lagrange Duality

Chee Wei Tan

Convex Optimization and its Applications to Computer Science

Outline

- Convex optimization
- Optimality condition
- Lagrange duality
- KKT optimality condition
- Sensitivity analysis

Convex Optimization

A convex optimization problem with variables x:

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq 0, \quad i = 1, 2, \ldots, m \\
& a_i^T x = b_i, \quad i = 1, 2, \ldots, p
\end{array}
$$

where $f_0, f_1, \ldots, f_m$ are convex functions.

- Minimize convex objective function (or maximize concave objective function)
- Upper bound inequality constraints on convex functions ($\Rightarrow$ Constraint set is convex)
- Equality constraints must be affine

Convex Optimization

- Epigraph form:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & f_0(x) - t \leq 0 \\ & f_i(x) \leq 0, \quad i = 1, 2, \ldots, m \\ & a_i^T x = b_i, \quad i = 1, 2, \ldots, p \end{array}$$

- Can you reformulate the Illumination Problem in Lecture 1?

- Can you reformulate the following problem (not in convex optimization form):

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & \frac{x_1}{1+x_2^2} \le 0 \\ & (x_1 + x_2)^2 = 0 \end{array}$$

Now transformed into a convex optimization problem:

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & x_1 \le 0 \\ & x_1 + x_2 = 0 \end{array}$$

Locally Optimal $\Rightarrow$ Globally Optimal

Given x is locally optimal for a convex optimization problem, $i.e.$, x is feasible and for some $R > 0$,

$$f_0(x) = \inf\{f_0(z) | z \text{ is feasible }, \|z - x\|_2 \leq R\}$$

Suppose x is not globally optimal, $i.e.$, there is a feasible y such that $f_0(y) < f_0(x)$

Since $\|y - x\|_2 > R$, we can construct a point $z = (1 - \theta)x + \theta y$ where $\theta = \frac{R}{2\|y-x\|_2}$. By convexity of feasible set, z is feasible. By convexity of f_0, we have

$$f_0(z) \leq (1 - \theta) f_0(x) + \theta f_0(y) < f_0(x)$$

which contradicts locally optimality of x

Therefore, there exists no feasible y such that $f_0(y) < f_0(x)$

Optimality Condition for Differentiable f_0

x is optimal for a convex optimization problem iff x is feasible and for all feasible y:

$$\nabla f_0(x)^T(y - x) \geq 0$$

$-\nabla f_0(x)$ is supporting hyperplane to feasible set

Unconstrained convex optimization: condition reduces to:

$$\nabla f_0(x) = 0$$

Proof: take $y = x - t\nabla f_0(x)$ where $t \in \mathbf{R}_+$. For small enough t, y is feasible, so $\nabla f_0(x)^T(y - x) = -t\|\nabla f_0(x)\|_2^2 \geq 0$. Thus $\nabla f_0(x) = 0$

Unconstrained Quadratic Optimization

Minimize $f_0(x) = \frac{1}{2}x^T P x + q^T x + r$

P is positive semidefinite. So it's a convex optimization problem

x minimizes f_0 iff (P, q) satisfy this linear equality:

$$\nabla f_0(x) = Px + q = 0$$

- If $q \notin \mathcal{R}(P)$, no solution. f_0 unbounded below
- If $q \in \mathcal{R}(P)$ and $P \succ 0$, there is a unique minimizer $x^* = -P^{-1}q$
- If $q \in \mathcal{R}(P)$ and P is singular, set of optimal x: $-P^{\dagger}q + \mathcal{N}(P)$

Equality Constrained Convex Optimization

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax = b \end{array}$$

Assume nonempty feasible set. Since $\nabla f_0(x)^T(y-x) \geq 0,\ \forall y : Ay = b$, and every feasible y can be written as $y = x + v$ for some $v \in \mathcal{N}(A)$, optimality condition is:

$$\nabla f_0(x)^T v \geq 0, \quad \forall v \in \mathcal{N}(A)$$

which implies $\nabla f_0(x)^T v = 0$, $i.e.$, $\nabla f_0(x)$ is orthogonal to $\mathcal{N}(A)$. Thus $\nabla f_0(x) \in \mathcal{R}(A^T)$, $i.e.$, there exists ν such that

$$\nabla f_0(x) + A^T \nu = 0$$

Conclusion: x is optimal iff $Ax = b$ and $\exists \nu$ s.t. $\nabla f_0(x) + A^T\nu = 0$

Duality Mentality

Bound or solve an optimization problem via a different optimization problem!

We'll develop the basic Lagrange duality theory for a general optimization problem, then specialize for convex optimization

Lagrange Dual Function

An optimization problem in standard form:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, 2, \ldots, m \\ & h_i(x) = 0, \quad i = 1, 2, \ldots, p \end{array}$$

Variables: $x \in \mathbf{R}^n$. Assume nonempty feasible set

Optimal value: p^*. Optimizer: x^*

Idea: augment objective with a weighted sum of constraints

Lagrangian $L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$

Lagrange multipliers (dual variables): $\lambda \succeq 0, \nu$

Lagrange dual function: $g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$

Lower Bound on Optimal Value

Claim: $g(\lambda, \nu) \le p^*, \quad \forall \lambda \succeq 0, \nu$

Proof: Consider feasible $\tilde{x}$:

$$L(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \le f_0(\tilde{x})$$

since $f_i(\tilde{x}) \le 0$ and $\lambda_i \ge 0$

Hence, $g(\lambda, \nu) \le L(\tilde{x}, \lambda, \nu) \le f_0(\tilde{x})$ for all feasible $\tilde{x}$

Therefore, $g(\lambda, \nu) \le p^*$

Lagrange Dual Function and Conjugate Function

- Lagrange dual function $g(\lambda, \nu)$

- Conjugate function: $f^*(y) = \sup_{x \in \mathbf{dom}\, f}(y^T x - f(x))$

  Consider linearly constrained optimization:

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & Ax \preceq b \\
 & Cx = d
\end{array}
$$

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_x \left(f_0(x) + \lambda^T (Ax - b) + \nu^T (Cx - d) \right) \\
&= -b^T \lambda - d^T \nu + \inf_x \left(f_0(x) + (A^T \lambda + C^T \nu)^T x \right) \\
&= -b^T \lambda - d^T \nu - f_0^*(-A^T \lambda - C^T \nu)
\end{aligned}
$$

Example

We'll use the simplest version of entropy maximization as our example for the rest of this lecture on duality. Entropy maximization is an important basic problem in information theory:

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Ax \preceq b \\ & \mathbf{1}^T x = 1 \end{array}$$

Since the conjugate function of $u \log u$ is e^{y-1}, by independence of the sum, we have

$$f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

Therefore, dual function of entropy maximization is

$$g(\lambda, \nu) = -b^T \lambda - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-a_i^T \lambda}$$

where a^i are columns of A

Lagrange Dual Problem

Lower bound from Lagrange dual function depends on (λ, ν).
What's the best lower bound that can be obtained from Lagrange dual function?

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

This is the Lagrange dual problem with dual variables (λ, ν)

Always a convex optimization! (Dual objective function always a concave function since it's the infimum of a family of affine functions in (λ, ν))

Denote the optimal value of Lagrange dual problem by d^*

Weak Duality

What's the relationship between d^* and p^*?

Weak duality always hold (even if primal problem is not convex):

$$d^* \leq p^*$$

Optimal duality gap:

$$p^* - d^*$$

Efficient generation of lower bounds through (convex) dual problem

Strong Duality

Strong duality (zero optimal duality gap):

$$d^* = p^*$$

If strong duality holds, solving dual is 'equivalent' to solving primal.
But strong duality does not always hold

Convexity and constraint qualifications $\Rightarrow$ Strong duality

A simple constraint qualification: Slater's condition (there exists strictly feasible primal variables $f_i(x) < 0$ for non-affine f_i)

Another reason why convex optimization is 'easy'

Example: Entropy Maximization

Primal optimization problem (variables x):

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Ax \preceq b,\ \mathbf{1}^T x = 1 \end{array}$$

Dual optimization problem (variables λ, ν):

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T \lambda} \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

Analytically maximize over the unconstrained $\nu \Rightarrow$ Simplified dual optimization problem (variables λ) and strong duality holds:

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \log \sum_{i=1}^n \exp(-a_i^T \lambda) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

Saddle Point Interpretation

Assume no equality constraints. We can express primal optimal value as

$$p^* = \inf_x \sup_{\lambda \succeq 0} L(x, \lambda)$$

By definition of dual optimal value:

$$d^* = \sup_{\lambda \succeq 0} \inf_x L(x, \lambda)$$

Weak duality (max min inequality):

$$\sup_{\lambda \succeq 0} \inf_x L(x, \lambda) \leq \inf_x \sup_{\lambda \succeq 0} L(x, \lambda)$$

Strong duality (saddle point property):

$$\sup_{\lambda \succeq 0} \inf_{x} L(x, \lambda) = \inf_{x} \sup_{\lambda \succeq 0} L(x, \lambda)$$

Complementary Slackness

Assume strong duality holds:

$$
\begin{aligned}
f_0(x^*) &= g(\lambda^*, \nu^*) \\
&= \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right) \\
&\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \\
&\leq f_0(x^*)
\end{aligned}
$$

Complementary Slackness

So the two inequalities must hold with equality. This implies:

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, 2, \ldots, m$$

Complementary Slackness Property:

$$\begin{aligned} \lambda_i^* > 0 \quad &\Rightarrow \quad f_i(x^*) = 0 \\ f_i(x^*) < 0 \quad &\Rightarrow \quad \lambda_i^* = 0 \end{aligned}$$

KKT Optimality Conditions

Since x^* minimizes $L(x, \lambda^*, \nu^*)$ over x, we have

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0$$

Karush-Kuhn-Tucker optimality conditions:

$$f_i(x^*) \leq 0, \quad h_i(x^*) = 0, \quad \lambda_i^* \succeq 0$$

$$\lambda_i^* f_i(x^*) = 0$$

$$\nabla f_0(x^*) + \textstyle\sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0$$

KKT Optimality Conditions

- Any optimization (with differentiable objective and constraint functions) with strong duality, KKT condition is necessary condition for primal-dual optimality

- Convex optimization (with differentiable objective and constraint functions) with Slater's condition, KKT condition is also sufficient condition for primal-dual optimality (useful for theoretical and numerical purposes)

Example: Entropy Maximization

Primal optimization problem (variables x):

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Ax \preceq b, \quad \mathbf{1}^T x = 1 \end{array}$$

Dual optimization problem (variables λ, ν):

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T\lambda} \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

Having solved dual problem, recover optimal primal variables as minimizer of

$$L(x, \lambda^*, \nu^*) = \sum_{i=1}^n x_i \log x_i + \lambda^{*T}(Ax - b) + \nu^*(\mathbf{1}^T x - 1)$$

$$i.e., \ x_i^* = 1/\exp(a_i^T \lambda^* + \nu^* + 1)$$

If x^* above is primal feasible, it's the optimal primal. Otherwise, the primal optimum is not attained

Example: Maximizing Channel Capacity as Water-filling

$$
\begin{array}{ll}
\text{maximize} & \displaystyle\sum_{i=1}^{n} \log\left(1 + \frac{x_i}{\alpha_i}\right) \\
\text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1
\end{array}
$$

Variables: x (powers). Constants: $\alpha_i > 0$ (channel noise)

KKT conditions:

$$
x^* \succeq 0, \quad \mathbf{1}^T x^* = 1, \quad \lambda^* \succeq 0
$$

$$
\lambda_i^* x_i^* = 0, \quad -1/(\alpha_i + x_i^*) - \lambda_i^* + \nu^* = 0
$$

Since λ^* are slack variables, reduce to

$$x^* \succeq 0, \quad \mathbf{1}^T x^* = 1$$

$$x_i^*(\nu^* - 1/(\alpha_i + x_i^*)) = 0, \quad \nu^* \geq 1/(\alpha_i + x_i^*)$$

If $\nu^* < 1/\alpha_i$, $x_i^* > 0$. So $x_i^* = 1/\nu^* - \alpha_i$. Otherwise, $x_i^* = 0$

Thus, $x_i^* = [1/\nu^* - \alpha_i]^+$ where ν^* is such that $\sum_i x_i^* = 1$

- Draw a picture to interpret this optimal solution

- Design an algorithm to compute this optimal solution

Global Sensitivity Analysis

Perturbed optimization problem:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq u_i, \quad i = 1, 2, \ldots, m \\ & h_i(x) = v_i \;\; i = 1, 2, \ldots, p \end{array}$$

Optimal value $p^*(u, v)$ as a function of parameters (u, v)

Assume strong duality and that dual optimum is attained:

$$\begin{aligned} p^*(0,0) &= g(\lambda^*, \nu^*) \\ &\leq f_0(x) + \textstyle\sum_i \lambda_i^* f_i(x) + \sum_i \nu_i^* h_i(x) \\ &\leq f_0(x) + \lambda^{*T} u + \nu^{*T} v \end{aligned}$$

Global Sensitivity Analysis

$$p^*(u,v) \geq p^*(0,0) - \lambda^{*T} u - \nu^{*T} v$$

- If λ_i^* is large, tightening ith constraint $(u_i < 0)$ will increase optimal value greatly

- If λ_i^* is small, loosening ith constraint $(u_i > 0)$ will reduce optimal value only slightly

Local Sensitivity Analysis

Assume that $p^*(u, v)$ is differentiable at $(0, 0)$:

$$\lambda_i^* = -\frac{\partial p^*(0,0)}{\partial u_i}, \quad \nu_i^* = -\frac{\partial p^*(0,0)}{\partial v_i}$$

Shadow price interpretation of Lagrange dual variables

Small λ_i^* means tightening or loosening ith constraint will not change optimal value by much

Summary

- Convexity mentality. Convex optimization is 'nice' for several reasons: local optimum is global optimum, zero optimal duality gap (under mild conditions), KKT optimality conditions are necessary and sufficient

- Duality mentality. Can always bound primal through dual, sometimes indirectly solve primal through dual

- Primal-dual: where is the optimum, how sensitive is it to perturbation in problem parameters?

Reading assignment: Sections 4.1-4.2 and 5.1-5.7 of textbook.

Economics Interpretation

- Primal objective: cost of operation
- Primal constraints: can be violated
- Dual variables: price for violating the corresponding constraint (dollar per unit violation). For the same price, can sell 'unused violation' for revenue
- Lagrangian: total cost
- Lagrange dual function: optimal cost as a function of violation prices
- Weak duality: optimal cost when constraints can be violated is less

than or equal to optimal cost when constraints cannot be violated, for any violation prices

- Duality gap: minimum possible arbitrage advantage

- Strong duality: can price the violations so that there is no arbitrage advantages