Duality (II)

Lijun Zhang
zlj@nju.edu.cn
http://cs.nju.edu.cn/zlj

Outline

- Saddle-point Interpretation
  - Max-min Characterization of Weak and Strong Duality
  - Saddle-point Interpretation
  - Game Interpretation
- Optimality Conditions
  - Certificate of Suboptimality and Stopping Criteria
  - Complementary Slackness
  - KKT Optimality Conditions
  - Solving the Primal Problem via the Dual
- Examples
- Generalized Inequalities

Outline

- Saddle-point Interpretation
  - Max-min Characterization of Weak and Strong Duality
  - Saddle-point Interpretation
  - Game Interpretation
- Optimality Conditions
  - Certificate of Suboptimality and Stopping Criteria
  - Complementary Slackness
  - KKT Optimality Conditions
  - Solving the Primal Problem via the Dual
- Examples
- Generalized Inequalities

More Symmetric Form

- Assume no equality constraint

$$\sup_{\lambda \succcurlyeq 0} L(x,\lambda) = \sup_{\lambda \succcurlyeq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right)$$

$$= \begin{cases} f_0(x) & f_i(x) \le 0, \quad i = 1, \dots, m \\ \infty & \text{otherwise} \end{cases}$$

  - Suppose $f_i(x) > 0$ for some i. Then, $\sup_{\lambda \succcurlyeq 0} L(x,\lambda) = \infty$ by $\lambda_j = 0, j \neq i$ and $\lambda_i \to \infty$
  - If $f_i(x) \le 0, i = 1, \dots, m$, then the optimal choice of λ is $\lambda = 0$ and $\sup_{\lambda \succcurlyeq 0} L(x,\lambda) = f_0(x)$

More Symmetric Form

- Assume no equality constraint

$$\sup_{\lambda \succcurlyeq 0} L(x, \lambda) = \sup_{\lambda \succcurlyeq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right)$$

$$= \begin{cases} f_0(x) & f_i(x) \leq 0, \quad i = 1, \dots, m \\ \infty & \text{otherwise} \end{cases}$$

- Optimal Value of Primal Problem

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

More Symmetric Form

- Optimal Value of Primal Problem

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

- Optimal Value of Dual Problem

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

- Weak Duality

$$\sup_{\lambda \succcurlyeq 0} \inf_x L(x, \lambda) \le \inf_x \sup_{\lambda \succcurlyeq 0} L(x, \lambda)$$

- Strong Duality

$$\sup_{\lambda \succcurlyeq 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \succcurlyeq 0} L(x, \lambda)$$

  - Min and Max can be switched

A More General Form

- Max-min Inequality

$$\sup_{z\in Z}\inf_{w\in W} f(w,z) \leq \inf_{w\in W}\sup_{z\in Z} f(w,z)$$

  - For any $f: \mathbf{R}^n \times \mathbf{R}^m \rightarrow \mathbf{R}$ and any $W \subseteq \mathbf{R}^n, Z \subseteq \mathbf{R}^m$

- Strong Max-min Property

$$\sup_{z\in Z}\inf_{w\in W} f(w,z) = \inf_{w\in W}\sup_{z\in Z} f(w,z)$$

  - Hold only in special cases

Outline

- Saddle-point Interpretation
  - Max-min Characterization of Weak and Strong Duality
  - Saddle-point Interpretation
  - Game Interpretation
- Optimality Conditions
  - Certificate of Suboptimality and Stopping Criteria
  - Complementary Slackness
  - KKT Optimality Conditions
  - Solving the Primal Problem via the Dual
- Examples
- Generalized Inequalities

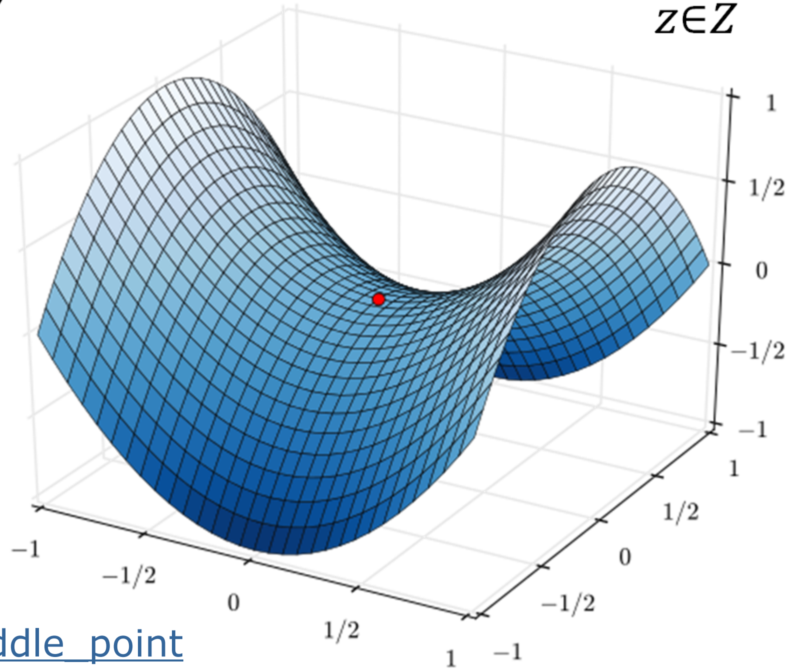
Saddle-point Interpretation

- $\widetilde{w} \in W, \tilde{z} \in Z$ is a saddle point for f

$$f(\widetilde{w}, z) \le f(\widetilde{w}, \tilde{z}) \le f(w, \tilde{z}), \qquad \forall w \in W, z \in Z$$

  - $\widetilde{w}$ minimizes $f(w, \tilde{z})$, $\tilde{z}$ maximizes $f(\widetilde{w}, z)$

$$f(\widetilde{w}, \tilde{z}) = \inf_{w \in W} f(w, \tilde{z}), \qquad f(\widetilde{w}, \tilde{z}) = \sup_{z \in Z} f(\widetilde{w}, z)$$



https://en.wikipedia.org/wiki/Saddle_point

Saddle-point Interpretation

- $\tilde{w} \in W, \tilde{z} \in Z$ is a saddle point for f

$$f(\tilde{w}, z) \leq f(\tilde{w}, \tilde{z}) \leq f(w, \tilde{z}), \qquad \forall w \in W, z \in Z$$

  - $\tilde{w}$ minimizes $f(w, \tilde{z})$, $\tilde{z}$ maximizes $f(\tilde{w}, z)$

$$f(\tilde{w}, \tilde{z}) = \inf_{w \in W} f(w, \tilde{z}), \qquad f(\tilde{w}, \tilde{z}) = \sup_{z \in Z} f(\tilde{w}, z)$$

- Imply the strong max-min property

$$\left.\begin{aligned} \sup_{z \in Z} \inf_{w \in W} f(w, z) \geq \inf_{w \in W} f(w, \tilde{z}) = f(\tilde{w}, \tilde{z}) \\ f(\tilde{w}, \tilde{z}) = \sup_{z \in Z} f(\tilde{w}, z) \geq \inf_{w \in W} \sup_{z \in Z} f(w, z) \end{aligned}\right\}$$

$$\Rightarrow \sup_{z \in Z} \inf_{w \in W} f(w, z) \geq \inf_{w \in W} \sup_{z \in Z} f(w, z)$$

$$\Rightarrow \sup_{z \in Z} \inf_{w \in W} f(w, z) = \inf_{w \in W} \sup_{z \in Z} f(w, z)$$

Saddle-point Interpretation

- $\tilde{w} \in W, \tilde{z} \in Z$ is a saddle point for f

$$f(\tilde{w}, z) \le f(\tilde{w}, \tilde{z}) \le f(w, \tilde{z}), \qquad \forall w \in W, z \in Z$$

  - $\tilde{w}$ minimizes $f(w, \tilde{z})$, $\tilde{z}$ maximizes $f(\tilde{w}, z)$

$$f(\tilde{w}, \tilde{z}) = \inf_{w \in W} f(w, \tilde{z}), \qquad f(\tilde{w}, \tilde{z}) = \sup_{z \in Z} f(\tilde{w}, z)$$

  - If $x^\star, \lambda^\star$ are primal and dual optimal points and strong duality holds, $x^\star, \lambda^\star$ form a saddle-point.
  - If x, λ is saddle-point, then x is primal optimal, λ is dual optimal, and the duality gap is zero.

Outline

- Saddle-point Interpretation
  - Max-min Characterization of Weak and Strong Duality
  - Saddle-point Interpretation
  - Game Interpretation
- Optimality Conditions
  - Certificate of Suboptimality and Stopping Criteria
  - Complementary Slackness
  - KKT Optimality Conditions
  - Solving the Primal Problem via the Dual
- Examples
- Generalized Inequalities

Continuous Zero-sum Game

- Two players
  - The 1st player chooses $w \in W$, and the 2nd player selects $z \in Z$
  - Player 1 pays an amount $f(w,z)$ to player 2
- Goals
  - Player 1 wants to minimize f
  - Player 2 wants to maximize f
- Continuous game
  - The choices are vectors, and not discrete

Continuous Zero-sum Game

- Player 1 makes his choice first
  - Player 2 wants to maximize payoff $f(w,z)$ and the resulting payoff is $\sup_{z \in Z} f(w,z)$
  - Player 1 knows that player 2 will follow this strategy, and so will choose $w \in W$ to make $\sup_{z \in Z} f(w,z)$ as small as possible
  - Thus, player 1 chooses

$$\operatorname*{argmin}_{w \in W} \sup_{z \in Z} f(w,z)$$

  - The payoff

$$\inf_{w \in W} \sup_{z \in Z} f(w,z)$$

Continuous Zero-sum Game

- Player 2 makes his choice first
  - Player 1 wants to minimize payoff $f(w,z)$ and the resulting payoff is $\inf_{w \in W} f(w,z)$
  - Player 2 knows that player 1 will follow this strategy, and so will choose $z \in Z$ to make $\inf_{w \in W} f(w,z)$ as large as possible
  - Thus, player 2 chooses

$$\operatorname*{argmax}_{z \in Z} \inf_{w \in W} f(w,z)$$

  - The payoff

$$\sup_{z \in Z} \inf_{w \in W} f(w,z)$$

Continuous Zero-sum Game

Max-min Inequality

$$\sup_{z \in Z} \inf_{w \in W} f(w,z) \leq \inf_{w \in W} \sup_{z \in Z} f(w,z)$$

Player 2 plays first   Player 1 plays first

- Player 1 wants to minimize f
- Player 2 wants to maximize f

It is better for a
player to go second

Continuous Zero-sum Game

Strong Max-min Property

$$\sup_{z\in Z}\inf_{w\in W} f(w,z) = \inf_{w\in W}\sup_{z\in Z} f(w,z)$$

Player 2 plays first — Player 1 plays first

- Player 1 wants to minimize f
- Player 2 wants to maximize f

There is no advantage to playing second

Continuous Zero-sum Game

Strong Max-min Property

$$\sup_{z \in Z} \inf_{w \in W} f(w,z) = \inf_{w \in W} \sup_{z \in Z} f(w,z)$$

Player 2 plays first　　Player 1 plays first

Saddle-point Property

- If $\tilde{w}, \tilde{z}$ is a saddle-point for f (and W, Z), then it is called a solution of the game
  - ✓ $\tilde{w}$: the optimal strategy for player 1
  - ✓ $\tilde{z}$: the optimal strategy for player 2
  - ✓ No advantage to playing second

A Special Case

- Payoff is the Lagrangian; $W = \mathbf{R}^n, Z = \mathbf{R}_+^m$
  - Player 1 chooses the primal variable x while player 2 chooses the dual variable $\lambda \succcurlyeq 0$
  - The optimal choice for player 2, if she must choose first, is any dual optimal $\lambda^\star$
    - ✓ The resulting payoff: $d^\star$
  - Conversely, if player 1 chooses first, his optimal choice is any primal optimal $x^\star$
    - ✓ The resulting payoff: $p^\star$
  - Duality gap: advantage of going second

Outline

- Saddle-point Interpretation
  - Max-min Characterization of Weak and Strong Duality
  - Saddle-point Interpretation
  - Game Interpretation
- Optimality Conditions
  - Certificate of Suboptimality and Stopping Criteria
  - Complementary Slackness
  - KKT Optimality Conditions
  - Solving the Primal Problem via the Dual
- Examples
- Generalized Inequalities

Certificate of Suboptimality

- Dual Feasible (λ, ν)
  - A lower bound on the optimal value of the primal problem

    $$p^\star \geq g(\lambda, \nu)$$

  - Provides a proof or certificate
  - Bound how suboptimal a given feasible point x is, without knowing the value of $p^\star$

    $$f_0(x) - p^\star \leq f_0(x) - g(\lambda, \nu) = \epsilon$$

    ✓ x is ϵ-suboptimal for primal problem

Certificate of Suboptimality

- Dual Feasible (λ, ν)
  - A lower bound on the optimal value of the primal problem
  $$p^\star \geq g(\lambda, \nu)$$
  - Provides a proof or certificate
  - Bound how suboptimal a dual feasible point (λ, ν) is, without knowing the value of $d^\star$
  $$d^* - g(\lambda, \nu) \leq f_0(x) - g(\lambda, \nu) = \epsilon$$
    - ✓ (λ, ν) is ϵ-suboptimal for dual problem

Certificate of Suboptimality

Gap between Primal & Dual Objectives

$$f_0(x) - g(\lambda, \nu)$$

- Referred to as duality gap associated with primal feasible x and dual feasible (λ, ν)

- $x, (\lambda, \nu)$ localizes the optimal value of the primal (and dual) problems to an interval

$$p^\star \in [g(\lambda, \nu), f_0(x)], \qquad d^\star \in [g(\lambda, \nu), f_0(x)]$$

  - ✓ The width of the interval is the duality gap

- If duality gap of $x, (\lambda, \nu)$ is 0, then x is primal optimal and (λ, ν) is dual optimal

Stopping Criteria

- Optimization algorithms produce a sequence of primal feasible $x^{(k)}$ and dual feasible $(\lambda^{(k)}, \nu^{(k)})$ for $k = 1,2,\dots,$
- Required absolute accuracy: ϵ_{abs}
- A Nonheuristic Stopping Criterion

$$f_0\left(x^{(k)}\right) - g\left(\lambda^{(k)}, \nu^{(k)}\right) \le \epsilon_{\text{abs}}$$

  - Guarantees when algorithm terminates, $x^{(k)}$ is ϵ_{abs}-suboptimal

Stopping Criteria

- A Relative Accuracy ϵ_{rel}
- Nonheuristic Stopping Criteria
  - If

$$g\left(\lambda^{(k)}, \nu^{(k)}\right) > 0, \qquad \frac{f_0\left(x^{(k)}\right) - g\left(\lambda^{(k)}, \nu^{(k)}\right)}{g\left(\lambda^{(k)}, \nu^{(k)}\right)} \leq \epsilon_{\mathrm{rel}}$$

or

$$f_0\left(x^{(k)}\right) < 0, \qquad \frac{f_0\left(x^{(k)}\right) - g\left(\lambda^{(k)}, \nu^{(k)}\right)}{-f_0\left(x^{(k)}\right)} \leq \epsilon_{\mathrm{rel}}$$

  - Then $p^\star \neq 0$, and the relative error satisfies

$$\frac{f_0\left(x^{(k)}\right) - p^\star}{|p^\star|} \leq \epsilon_{\mathrm{rel}}$$

Outline

- Saddle-point Interpretation
  - Max-min Characterization of Weak and Strong Duality
  - Saddle-point Interpretation
  - Game Interpretation
- Optimality Conditions
  - Certificate of Suboptimality and Stopping Criteria
  - Complementary Slackness
  - KKT Optimality Conditions
  - Solving the Primal Problem via the Dual
- Examples
- Generalized Inequalities

Complementary Slackness

- Suppose Strong Duality Holds
  - For primal optimal $x^\star$ & dual optimal $(\lambda^\star, \nu^\star)$

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

  - ✓ First line: the optimal duality gap is zero
  - ✓ Second line: definition of the dual function
  - ✓ Third line: infimum of Lagrangian over x is less than or equal to its value at $x = x^\star$

Complementary Slackness

- Suppose Strong Duality Holds
  - For primal optimal $x^\star$ & dual optimal $(\lambda^\star, \nu^\star)$

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

    - ✓ Last line: $\lambda_i^\star \geq 0, f_i(x^\star) \leq 0, i = 1, \ldots, m$ and $h_i(x^\star) = 0, i = 1, \ldots, p$
    - ✓ We conclude that the two inequalities in this chain hold with equality

Complementary Slackness

- Suppose Strong Duality Holds
  - For primal optimal $x^\star$ & dual optimal $(\lambda^\star, \nu^\star)$

$$f_0(x^\star) = g(\lambda^\star, \nu^\star)$$

$$= \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{i=1}^{p} \nu_i^\star h_i(x) \right)$$

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

$$= f_0(x^\star)$$

- ✓ Equality in the third line implies $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
- ✓ Equality in the last line implies $\sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) = 0$

Complementary Slackness

Complementary Slackness

$$\lambda_i^\star f_i(x^\star) = 0, \qquad i = 1, \dots, m$$

- Derived from $\sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) = 0$
- Holds for any primal optimal $x^\star$ and dual optimal $\lambda^\star, \nu^\star$ (when strong duality holds)

- Other expressions

$$\lambda_i^\star > 0 \Rightarrow f_i(x^\star) = 0$$
$$f_i(x^\star) < 0 \Rightarrow \lambda_i^\star = 0$$

  - ✓ i-th optimal Lagrange multiplier is 0 unless i-th constraint is active at the optimum $f_i(x^\star) = 0$

Outline

- Saddle-point Interpretation
  - Max-min Characterization of Weak and Strong Duality
  - Saddle-point Interpretation
  - Game Interpretation
- Optimality Conditions
  - Certificate of Suboptimality and Stopping Criteria
  - Complementary Slackness
  - KKT Optimality Conditions
  - Solving the Primal Problem via the Dual
- Examples
- Generalized Inequalities

KKT Conditions for Nonconvex Problems

- $x^\star$ and $(\lambda^\star, \nu^\star)$: any primal and dual optimal points with zero duality gap
  - $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$

$$\Rightarrow \nabla L(x^\star, \lambda^\star, \nu^\star) = 0$$

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

KKT Conditions for Nonconvex Problems

- $x^\star$ and $(\lambda^\star, \nu^\star)$: any primal and dual optimal points with zero duality gap

$$
\begin{aligned}
f_i(x^\star) &\le 0, & i &= 1, \dots, m \\
h_i(x^\star) &= 0, & i &= 1, \dots, p \\
\lambda_i^\star &\ge 0, & i &= 1, \dots, m \\
\lambda_i^\star f_i(x^\star) &= 0, & i &= 1, \dots, m \\
\nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) &+ \sum_{i=1}^{p} \nu_i^\star \nabla h_i(x^\star) = 0
\end{aligned}
$$

  - Karush-Kuhn-Tucker (KKT) conditions

Necessary Condition

For optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal must satisfy KKT conditions.

KKT Conditions for Convex Problems

- If f_i are convex, h_i are affine, $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ satisfy

$$
\begin{aligned}
f_i(\tilde{x}) &\le 0, \qquad & i = 1, \dots, m \\
h_i(\tilde{x}) &= 0, \qquad & i = 1, \dots, p \\
\tilde{\lambda}_i &\ge 0, \qquad & i = 1, \dots, m \\
\tilde{\lambda}_i f_i(\tilde{x}) &= 0, \qquad & i = 1, \dots, m
\end{aligned}
$$

$$
\nabla f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_{\mathrm{i}} \nabla h_i(\tilde{x}) = 0
$$

- Then, $\tilde{x}$ and $\tilde{\lambda}, \tilde{\nu}$ are primal and dual optimal, with zero duality gap.

Sufficient Condition

For any convex optimization problem with differentiable objective and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap.

KKT Conditions for Convex Problems

- For convex problem satisfying Slater's condition, KKT conditions provide necessary and sufficient conditions for optimality.
  - Slater's condition implies that optimal duality gap is zero and dual optimum is attained
  - x is optimal if and only if there are (λ, ν) that, together with x, satisfy the KKT conditions

KKT Conditions for Convex Problems

- The KKT conditions play an important role in optimization.
  - In a few special cases it is possible to solve the KKT conditions.
  - More generally, many algorithms for convex optimization can be interpreted as methods for solving the KKT conditions

Example

- Equality Constrained Convex Quadratic Minimization
  - Primal Problem (with $P \in \mathbf{S}_+^n$)

$$\begin{aligned} \min \quad & (1/2)x^\top P x + q^\top x + r \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

  - KKT conditions

$$Ax^\star = b, Px^\star + q + A^\top \nu^\star = 0$$

$$\Leftrightarrow \begin{bmatrix} P & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

    - ✓ Solving this set of $m+n$ equations in $m+n$ variables $x^\star, \nu^\star$ gives optimal primal and dual variables

Outline

- Saddle-point Interpretation
  - Max-min Characterization of Weak and Strong Duality
  - Saddle-point Interpretation
  - Game Interpretation
- Optimality Conditions
  - Certificate of Suboptimality and Stopping Criteria
  - Complementary Slackness
  - KKT Optimality Conditions
  - Solving the Primal Problem via the Dual
- Examples
- Generalized Inequalities

Solving the Primal Problem via the Dual

- If strong duality holds and a dual optimal solution $(\lambda^\star, \nu^\star)$ exists, any primal optimal point is also a minimizer of $L(x, \lambda^\star, \nu^\star)$
  - Suppose the minimizer of $L(x, \lambda^\star, \nu^\star)$ below is unique

$$\min \quad f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{i=1}^{p} \nu_i^\star h_i(x)$$

    - ✓ If solution is primal feasible, it's primal optimal
    - ✓ If not primal feasible, no optimal point exists

Example

- Entropy Maximization
  - Primal Problem (with domain $\mathbb{R}^n_{++}$)

$$\begin{array}{ll} \min & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{s.t.} & Ax \preccurlyeq b \\ & \mathbf{1}^\top x = 1 \end{array}$$

  - Dual Problem (a_i: the i-th column of A)

$$\begin{array}{ll} \max & -b^\top \lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^\top \lambda} \\ \text{s.t.} & \lambda \succcurlyeq 0 \end{array}$$

  - Assume weak Slater's condition holds
    - ✓ There exists an $x \succ 0$ with $Ax \preccurlyeq b, \mathbf{1}^\top x = 1$
    - ✓ So strong duality holds and an optimal solution $(\lambda^\star, \nu^\star)$ exists

Example

Entropy Maximization

- Suppose we have solved the dual problem
- The Lagrangian at $(\lambda^\star, \nu^\star)$ is

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

  - ✓ Strictly convex on $\mathcal{D}$ and bounded below
  - ✓ So it has a unique solution
    $x_i^\star = 1/\exp\left(a_i^\top \lambda^\star + \nu^\star + 1\right), \qquad i = 1, \dots, n$
  - ✓ If $x^\star$ is primal feasible, it must be the optimal solution of the primal problem
  - ✓ If $x^\star$ is not primal feasible, we can conclude that the primal optimum is not attained

Outline

- Saddle-point Interpretation
  - Max-min Characterization of Weak and Strong Duality
  - Saddle-point Interpretation
  - Game Interpretation
- Optimality Conditions
  - Certificate of Suboptimality and Stopping Criteria
  - Complementary Slackness
  - KKT Optimality Conditions
  - Solving the Primal Problem via the Dual
- Examples
- Generalized Inequalities

Examples

- Introduce New Variables and Equality Constraints
- Transform the Objective
- Implicit Constraints

Introduce New Variables and Equality Constraints

Unconstrained Problem

$$\min \quad f_0(Ax+b)$$

- Lagrange dual function: constant $p^\star$
  - ✓ strong duality holds ($p^\star = d^\star$), but it is not useful

Reformulation

$$\begin{aligned} \min \quad & f_0(y) \\ \text{s.t.} \quad & Ax+b=y \end{aligned}$$

- Lagrangian of the reformulated problem

$$L(x,y,\nu) = f_0(y) + \nu^\top (Ax+b-y)$$

Introduce New Variables and Equality Constraints

Unconstrained Problem

- Find dual function by minimizing L
  - ✓ Minimizing over x, $g(\nu) = -\infty$ unless $A^\top \nu = 0$
- When $A^\top \nu = 0$, minimizing L gives

$$g(\nu) = b^\top \nu + \inf_y (f_0(y) - \nu^\top y) = b^\top \nu - f_0^*(\nu)$$

  - ✓ f_0^*: conjugate of f_0
- Dual problem

$$\begin{aligned} \max \quad & b^\top \nu - f_0^*(\nu) \\ \text{s.t.} \quad & A^\top \nu = 0 \end{aligned}$$

  - ✓ More useful

Example

Unconstrained Geometric Program

- Problem

$$\min \quad \log\left(\sum_{i=1}^{m} \exp\left(a_i^\top x + b_i\right)\right)$$

- Add new variables & equality constraints

$$\begin{aligned} \min \quad & f_0(y) = \log\left(\sum_{i=1}^{m} \exp y_i\right) \\ \text{s.t.} \quad & Ax + b = y \end{aligned}$$

✓ $a_i^\top$: i-th row of A

- Conjugate of the log-sum-exp function

$$f_0^*(\nu) = \begin{cases} \sum_{i=1}^{m} \nu_i \log \nu_i & \nu \succcurlyeq 0, \mathbf{1}^\top \nu = 1 \\ \infty & \text{otherwise} \end{cases}$$

Introduce New Variables and Equality Constraints

Unconstrained Geometric Program

- Primal Problem

$$\begin{aligned} \min \quad & f_0(y) = \log\left(\sum_{i=1}^{m} \exp y_i\right) \\ \text{s.t.} \quad & Ax + b = y \end{aligned}$$

- Dual of the reformulated problem

$$\begin{aligned} \max \quad & b^\top \nu - \sum_{i=1}^{m} \nu_i \log \nu_i \\ \text{s.t.} \quad & \mathbf{1}^\top \nu = 1 \\ & A^\top \nu = 0 \\ & \nu \succcurlyeq 0 \end{aligned}$$

✓ An entropy maximization problem

Example

Norm Approximation Problem

- Problem (with any norm $\|\cdot\|$)

$$\min \quad \|Ax - b\|$$

  - ✓ Constant Lagrange dual function (not useful)

- Reformulate the problem

$$\begin{aligned} \min \quad & \|y\| \\ \text{s.t.} \quad & Ax - b = y \end{aligned}$$

- Lagrange dual problem

$$\begin{aligned} \max \quad & b^\top \nu \\ \text{s.t.} \quad & \|\nu\|_* \le 1, A^\top \nu = 0 \end{aligned}$$

  - ✓ The conjugate of a norm is the indicator function of the dual norm unit ball

Introduce New Variables and Equality Constraints

Constraint Functions

$$\begin{aligned} \min \quad & f_0(A_0 x + b_0) \\ \text{s.t.} \quad & f_i(A_i x + b_i) \le 0, \qquad i = 1, \ldots, m \end{aligned}$$

- $A_i \in \mathbf{R}^{k_i \times n}; f_i : \mathbf{R}^{k_i} \to \mathbf{R}$
- Introduce $y_i \in \mathbf{R}^{k_i}, i = 0, \ldots, m$

$$\begin{aligned} \min \quad & f_0(y_0) \\ \text{s.t.} \quad & f_i(y_i) \le 0, \qquad i = 1, \ldots, m \\ & A_i x + b_i = y_i, \qquad i = 0, \ldots, m \end{aligned}$$

- The Lagrangian for the above problem

$$L(x, y_0, \ldots, y_m, \lambda, \nu_0, \ldots, \nu_m)$$
$$= f_0(y_0) + \sum_{i=1}^{m} \lambda_i f_i(y_i) + \sum_{i=0}^{m} \nu_i^{\top} (A_i x + b_i - y_i)$$

Introduce New Variables and Equality Constraints

Constraint Functions

- Dual function (by minimizing over x & y_i)
  - ✓ Minimum over x is $-\infty$ unless $\sum_{i=0}^{m} A_i^\top \nu_i = 0$

In this case, for $\lambda \succ 0$, $g(\lambda, \nu_0, \dots, \nu_m)$

$$= \sum_{i=0}^{m} \nu_i^\top b_i + \inf_{y_0,\dots,y_m} \left(f_0(y_0) + \sum_{i=1}^{m} \lambda_i f_i(y_i) - \sum_{i=0}^{m} \nu_i^\top y_i \right)$$

$$= \sum_{i=0}^{m} \nu_i^\top b_i + \inf_{y_0} \left(f_0(y_0) - \nu_0^\top y_0 \right) + \sum_{i=1}^{m} \lambda_i \inf_{y_i} \left(f_i(y_i) - (\nu_i/\lambda_i)^\top y_i \right)$$

$$= \sum_{i=0}^{m} \nu_i^\top b_i - f_0^*(\nu_0) - \sum_{i=1}^{m} \lambda_i f_i^*(\nu_i/\lambda_i)$$

Introduce New Variables and Equality Constraints

Constraint Functions

- What happens when $\lambda \succcurlyeq 0$ (but some $\lambda_i = 0$)
  - ✓ If $\lambda_i = 0 \,\&\, \nu_i \neq 0$, the dual function is $-\infty$
  - ✓ If $\lambda_i = 0 \,\&\, \nu_i = 0$, terms involving y_i, ν_i, λ_i are 0
- The expression for g is valid for all $\lambda \succcurlyeq 0$ if
  - ✓ Take $\lambda_i f_i^*(\nu_i/\lambda_i) = 0$, when $\lambda_i = 0 \,\&\, \nu_i = 0$
  - ✓ Take $\lambda_i f_i^*(\nu_i/\lambda_i) = \infty$, when $\lambda_i = 0 \,\&\, \nu_i \neq 0$
- Dual Problem

$$\begin{array}{ll} \max & \sum_{i=0}^{m} \nu_i^\top b_i - f_0^*(\nu_0) - \sum_{i=1}^{m} \lambda_i f_i^*(\nu_i/\lambda_i) \\ \text{s.t.} & \lambda \succcurlyeq 0, \ \sum_{i=0}^{m} A_i^\top \nu_i = 0 \end{array}$$

Example

- Inequality Constrained Geometric Program
  - Problem

$$\begin{aligned} \min \quad & \log\left(\sum_{k=1}^{K_0} e^{a_{0k}^\top x + b_{0k}}\right) \\ \text{s.t.} \quad & \log\left(\sum_{k=1}^{K_i} e^{a_{ik}^\top x + b_{ik}}\right) \le 0, \ \ i = 1, \dots, m \end{aligned}$$

  ✓ Let $f_i(y) = \log\left(\sum_{k=1}^{K_i} e^{y_k}\right)$

  ✓ Conjugate of f_i

$$f_i^*(\nu) = \begin{cases} \sum_{k=1}^{K_i} \nu_k \log \nu_k & \nu \succcurlyeq 0, \mathbf{1}^\top \nu = 1 \\ \infty & \text{otherwise} \end{cases}$$

Example

- Inequality Constrained Geometric Program
  - Dual problem is

$$\begin{array}{ll} \max & b_0^\top \nu_0 - \sum_{k=1}^{K_0} \nu_{0k} \log \nu_{0k} + \sum_{i=1}^{m} \left(b_i^\top \nu_i - \sum_{k=1}^{K_i} \nu_{ik} \log(\nu_{ik}/\lambda_i) \right) \\ \text{s.t.} & \nu_0 \succcurlyeq 0, \quad \mathbf{1}^\top \nu_0 = 1 \\ & \nu_i \succcurlyeq 0, \quad \mathbf{1}^\top \nu_i = \lambda_i, \qquad i = 1, \dots, m \\ & \lambda_i \geq 0, \quad i = 1, \dots, m \\ & \sum_{i=0}^{m} A_i^\top \nu_i = 0 \end{array}$$

Transform the Objective

- Replace the Objective f_0 by an Increasing Function of f_0
  - The resulting problem is equivalent
  - The dual of this equivalent problem can be very different from dual of original problem

Example

- Minimum Norm Problem

$$\min \quad \|Ax - b\|$$

- Reformulate this problem as

$$\begin{aligned} \min \quad & (1/2)\|y\|^2 \\ \text{s.t.} \quad & Ax - b = y \end{aligned}$$

  - ✓ Introduce new variables and replace the objective by half its square
  - ✓ Equivalent to the original problem

- Dual of the reformulated problem

$$\begin{aligned} \max \quad & -\left(\frac{1}{2}\right)\|\nu\|_*^2 + b^\top \nu \\ \text{s.t.} \quad & A^\top \nu = 0 \end{aligned}$$

Example

Norm Approximation Problem

- Problem (with any norm $\|\cdot\|$)

$$\min \quad \|Ax - b\|$$

  - ✓ Constant Lagrange dual function (not useful)

- Reformulate the problem

$$\begin{aligned} \min \quad & \|y\| \\ \text{s.t.} \quad & Ax - b = y \end{aligned}$$

- Lagrange dual problem

$$\begin{aligned} \max \quad & b^\top \nu \\ \text{s.t.} \quad & \|\nu\|_* \leq 1, A^\top \nu = 0 \end{aligned}$$

  - ✓ The conjugate of a norm is the indicator function of the dual norm unit ball

Implicit Constraints

- Include Some of the Constraints in the Objective Function
  - Modifying the objective function to be infinite when the constraint is violated

Example

Linear Program with Box Constraints

- Problem

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax = b \\ & l \preccurlyeq x \preccurlyeq u \end{aligned}$$

  - ✓ $A \in \mathbf{R}^{p\times n}$ and $l \prec u$
  - ✓ $l \preccurlyeq x \preccurlyeq u$ are called box constraints

- Derive the dual of this linear program

$$\begin{aligned} \min \quad & -b^\top \nu - \lambda_1^\top u + \lambda_2^\top l \\ \text{s.t.} \quad & A^\top \nu + \lambda_1 - \lambda_2 + c = 0 \\ & \lambda_1 \succcurlyeq 0, \qquad \lambda_2 \succcurlyeq 0 \end{aligned}$$

Example

Linear Program with Box Constraints

- Problem

$$
\begin{aligned}
\min \quad & c^\top x \\
\text{s.t.} \quad & Ax = b \\
& l \preccurlyeq x \preccurlyeq u
\end{aligned}
$$

  - ✓ $A \in \mathbf{R}^{p \times n}$ and $l \prec u$
  - ✓ $l \preccurlyeq x \preccurlyeq u$ are called box constraints

- Reformulate the problem as

$$
\begin{aligned}
\min \quad & f_0(x) \\
\text{s.t.} \quad & Ax = b
\end{aligned}
$$

  - ✓ Here, we define $f_0(x) = \begin{cases} c^\top x & l \preccurlyeq x \preccurlyeq u \\ \infty & \text{otherwise} \end{cases}$

Implicit Constraints

- Linear Program with Box Constraints
  - Dual function

$$g(\nu) = \inf_{l \preccurlyeq x \preccurlyeq u} \left(c^\top x + \nu^\top (Ax - b) \right)$$
$$= -b^\top \nu - u^\top (A^\top \nu + c)^- + l^\top (A^\top \nu + c)^+$$

    - ✓ $y_i^+ = \max\{y_i, 0\}$, $y_i^- = \max\{-y_i, 0\}$
    - ✓ We can derive an analytical formula for g, which is a concave piecewise-linear function
  - Dual problem

$$\max \quad -b^\top \nu - u^\top (A^\top \nu + c)^- + l^\top (A^\top \nu + c)^+$$

    - ✓ Unconstrained problem
    - ✓ Different form from the dual of original problem

Outline

- Saddle-point Interpretation
  - Max-min Characterization of Weak and Strong Duality
  - Saddle-point Interpretation
  - Game Interpretation
- Optimality Conditions
  - Certificate of Suboptimality and Stopping Criteria
  - Complementary Slackness
  - KKT Optimality Conditions
  - Solving the Primal Problem via the Dual
- Examples
- Generalized Inequalities

Generalized Inequalities

Problems with Generalized Inequality Constraints

- Primal Problem

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \preccurlyeq_{K_i} 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{aligned}$$

  - ✓ $K_i \subseteq \mathbf{R}^{k_i}$ are proper cones
  - ✓ Do not assume convexity of the problem
  - ✓ Assume the domain is nonempty

The Lagrange Dual

Lagrangian

$$L(x, \lambda, \nu) = f_0(x) + \lambda_1^\top f_1(x) + \cdots + \lambda_m^\top f_m(x) + \nu_1 h_1(x) + \cdots + \nu_p h_p(x)$$

- ✓ $\lambda = (\lambda_1, \dots, \lambda_m), \lambda_i \in \mathbf{R}^{k_i}, \nu = (\nu_1, \dots, \nu_p)$

Dual Function

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu)$$

$$= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^\top f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

- ✓ Lagrangian is affine in dual variables
- ✓ Dual function is pointwise infimum of Lagrangian; So, dual function is concave

The Lagrange Dual

- Nonnegativity on dual variables

$$\lambda_i \succcurlyeq_{K_i^*} 0, \qquad i = 1, \dots, m$$

  - K_i^*: the dual cone of K_i
  - Lagrange multipliers must be dual nonnegative

- Weak Duality
  - If $\lambda_i \succcurlyeq_{K_i^*} 0$ and $f_i(\tilde{x}) \preccurlyeq_{K_i} 0$, then $\lambda_i^\top f_i(\tilde{x}) \leq 0$
  - So, for any primal feasible $\tilde{x}$ and $\lambda_i \succcurlyeq_{K_i^*} 0$,

$$f_0(\tilde{x}) + \sum_{i=1}^m \lambda_i^\top f_i(\tilde{x}) + \sum_{i=1}^p \nu_i h_i(\tilde{x}) \leq f_0(\tilde{x})$$

  - Taking the infimum over $\tilde{x}$ yields $g(\lambda, \nu) \leq p^\star$

The Lagrange Dual

- Lagrange dual optimization problem

$$\begin{aligned} \max \quad & g(\lambda, \nu) \\ \text{s.t.} \quad & \lambda_i \succcurlyeq_{K_i^*} 0, \quad i = 1, \dots, m \end{aligned}$$

  - Always have weak duality ($d^\star \leq p^\star$) whether or not the primal problem is convex

- Primal Problem

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \preccurlyeq_{K_i} 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

The Lagrange Dual

Slater's Condition and Strong Duality

- Strong duality: $d^\star = p^\star$
  - ✓ Holds when primal problem is convex and satisfies appropriate constraint qualifications
- For problem (convex f_0 and K_i-convex f_i)

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \preccurlyeq_{K_i} 0, \quad i = 1, \dots, m \\ & Ax = b \end{aligned}$$

- Generalized version of Slater's condition
  - ✓ $\exists x \in \operatorname{relint} \mathcal{D}, Ax = b, f_i(x) \prec_{K_i} 0, i = 1, \dots, m$
  - ✓ Implies strong duality and the dual optimum is attained

Example

Lagrange Dual of Cone Program in Standard Form

- Primal Problem

$$\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & Ax = b \\ & x \succcurlyeq_K 0 \end{array}$$

  ✓ $A \in \mathbf{R}^{m\times n}, b \in \mathbf{R}^m$ and $K \subseteq \mathbf{R}^n$ is a proper cone

- Lagrangian: $L(x, \lambda, \nu) = c^\top x - \lambda^\top x + \nu^\top (Ax - b)$
- Dual function

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^\top \nu & A^\top \nu - \lambda + c = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

Example

- Lagrange Dual of Cone Program in Standard Form
  - Dual problem

$$\begin{aligned} \max \quad & -b^\top \nu \\ \mathrm{s.t.} \quad & A^\top \nu + c = \lambda \\ & \lambda \succcurlyeq_{K^*} 0 \end{aligned}$$

  - Eliminate λ and define $y = -\nu$ gives

$$\begin{aligned} \max \quad & b^\top y \\ \mathrm{s.t.} \quad & A^\top y \preccurlyeq_{K^*} c \end{aligned}$$

    - ✓ A cone program in inequality form
    - ✓ Involve the dual generalized inequality
    - ✓ Strong duality (Slater condition): $x \succ_K 0, Ax = b$

Optimality Conditions

Complementary Slackness

- Assume primal and dual optimal values are equal, and attained at $x^\star, \lambda^\star, \nu^\star$
- Complementary slackness

$$f_0(x^\star) = g(\lambda^\star, \nu^\star)$$

$$\leq f_0(x^\star) + \sum_{i=1}^{m} {\lambda_i^\star}^\top f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star h_i(x^\star)$$

$$\leq f_0(x^\star)$$

  - ✓ $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
  - ✓ The two sums in the second line are zero
  - ✓ The second sum is zero $\Rightarrow \sum_{i=1}^{m} {\lambda_i^\star}^\top f_i(x^\star) = 0 \Rightarrow$

$${\lambda_i^\star}^\top f_i(x^\star) = 0, \qquad i = 1, \ldots, m$$

Optimality Conditions

Complementary Slackness

- Assume primal and dual optimal values are equal, and attained at $x^\star, \lambda^\star, \nu^\star$
- Complementary slackness

$$f_0(x^\star) = g(\lambda^\star, \nu^\star)$$

$$\leq f_0(x^\star) + \sum_{i=1}^{m} {\lambda_i^\star}^\top f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star h_i(x^\star)$$

$$\leq f_0(x^\star)$$

✓ From ${\lambda_i^\star}^\top f_i(x^\star) = 0$, we can conclude

$$\lambda_i^\star \succ_{K_i^*} 0 \Rightarrow f_i(x^\star) = 0, \quad f_i(x^\star) \prec_{K_i} 0 \Rightarrow \lambda_i^\star = 0$$

✓ Possible to satisfy ${\lambda_i^\star}^\top f_i(x^\star) = 0$ with $\lambda_i^\star \neq 0$ & $f_i(x^\star) \neq 0$

Optimality Conditions

KKT Conditions

- Additionally assume f_i, h_i are differentiable

- Generalize the KKT conditions to problems with generalized inequalities

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$

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

  ✓ $Df_i(x^\star) \in \mathbf{R}^{k_i \times n}$: derivative of f_i evaluated at $x^\star$

Optimality Conditions

KKT Conditions

- If strong duality holds, any primal optimal $x^\star$ and dual optimal $(\lambda^\star, \nu^\star)$ must satisfy the optimality conditions (or KKT conditions)

$$
\begin{aligned}
f_i(x^\star) &\preccurlyeq_{K_i} 0, & i &= 1, \ldots, m \\
h_i(x^\star) &= 0, & i &= 1, \ldots, p \\
\lambda_i^\star &\succcurlyeq_{K_i^*} 0, & i &= 1, \ldots, m \\
{\lambda_i^\star}^\top f_i(x^\star) &= 0, & i &= 1, \ldots, m
\end{aligned}
$$

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

  - ✓ If the primal problem is convex, the converse also holds

Summary

- Saddle-point Interpretation
  - Max-min Characterization of Weak and Strong Duality
  - Saddle-point Interpretation
  - Game Interpretation
- Optimality Conditions
  - Certificate of Suboptimality and Stopping Criteria
  - Complementary Slackness
  - KKT Optimality Conditions
  - Solving the Primal Problem via the Dual
- Examples
- Generalized Inequalities