Network Optimization
Chapter 2: Mathematical Tools

Yiyin Wang
yiyinwang@sjtu.edu.cn
Department of Automation
Shanghai Jiao Tong University

Oct. 21, 2014

Acknowledgements

Slides borrow heavily from lectures by Stephen Boyd (Stanford) and Lieven Vandenberghe (UCLA)

Part 1: Fundamentals of convex optimization

- 2.1 Convex sets, convex functions, and convex problems
- 2.2 Duality

Lagrangian

Standard form problem (not necessarily convex)

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

variable $x \in \mathbf{R}^n$, domain D, optimal value p^*
Lagrangian: L: $\mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with $\operatorname{dom} L = D \times \mathbf{R}^m \times \mathbf{R}^p$,

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

- Weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrange dual function

Lagrange dual function: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$

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

g is concave, can be $-\infty$ for some λ, ν

Lower bound property: if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^*$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq 0$, then

$$
f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in D} L(x, \lambda, \nu) = g(\lambda, \nu)
$$

minimizing over all feasible $\tilde{x}$ given $p^* \geq g(\lambda, \nu)$

Least-norm solution of linear equations

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b \end{array}$$

Dual function

- Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$
- To minimize L over x, set gradient equal to zero:

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \rightarrow x = -(1/2) A^T \nu$$

- Plug in x into L to obtain g:

$$g(\nu) = L(((-1/2)A^T \nu, \nu)) = -\frac{1}{4} \nu^T A A^T \nu - b^T \nu$$

a concave function of ν

Lower bound property: $p^* \geq -(1/4)\nu^T A A^T \nu - b^T \nu$ for all ν

Standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, x \succeq 0 \end{array}$$

Dual function

- Lagrangian is

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= -b^T \nu + (c + A^T \nu - \lambda)^T x \end{aligned}$$

- L is affine in x, hence

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

g is linear on affine domain $\{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$, hence concave

Lower bound property: $p^* \geq -b^T \nu$ if $A^T \nu + c \succeq 0$

Two-way partitioning

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, i = 1, \dots, n \end{array}$$

- A nonconvex problem; feasible set contains 2^n discrete points
- Interpretation: partition $\{1, \dots, n\}$ in two sets; W_{ij} is cost of assigning i, j to the same set; $-W_{ij}$ is cost of assigning to different sets

Dual function

$$g(\nu) = \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) = \inf_x \left(x^T (W + \mathrm{diag}(\nu)) x - \mathbf{1}^T \right) \nu$$

$$= \begin{cases} -\mathbf{1}^T \nu & W + \mathrm{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases}$$

g is linear on affine domain $\{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$, hence concave

Lower bound property: $p^* \geq -\mathbf{1}^T \nu$ if $W + \mathrm{diag}(\nu) \succeq 0$

Lagrange dual and conjugate function

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

Dual function

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

- Recall definition of conjugate $f^*(y) = \sup_{x \in \mathrm{dom}\, f}(y^T x - f(x))$
- Simplifies derivation of dual if conjugate of f_0 is known

Example: entropy maximization

$$f_0(x) = \sum_{i=1}^{n} x_i \log x_i, \quad f_0^*(y) = \sum_{i=1}^{n} e^{y_i - 1}$$

The dual problem

Lagrange dual problem

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

- Find best lower bound on p^*, obtained from Lagrange dual function
- A convex optimization problem; optimal value denoted d^*
- λ, ν are dual feasible if $\lambda \succeq 0, (\lambda, \nu) \in \mathrm{dom}\, g$
- Often simplified by making implicit constraint $(\lambda, \nu) \in \mathrm{dom}\, g$ explicit

Example: standard form LP and its dual

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array}$$

$$\begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c \succeq 0 \end{array}$$

Weak and strong duality

Weak duality: $d^* \leq p^*$

- Always holds (for convex and nonconvex problems)
- Can be used to find nontrivial lower bounds for difficult problems for example, solving the SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathrm{diag}(\nu) \succeq 0 \end{array}$$

given a lower bound for the two-way partitioning problem

Strong duality: $d^* = p^*$

- Does not hold in general
- (Usually) holds for convex problems
- Conditions that guarantee strong duality in convex problems are called constraint qualifications

Slater's constraint qualification

Strong duality holds for a convex problem

$$\begin{array}{rl} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, i = 1, \ldots, m \\ & Ax = b \end{array}$$

If it is strictly feasible, *i.e.*,

$$\exists x \in \operatorname{int} D : \quad f_i(x) < 0, i = 1, \ldots, m, Ax = b$$

- Also guarantees that the dual optimum is attained (if $p^* > -\infty$)
- Can be sharpened: e.g., can replace $\operatorname{int} D$ with $\operatorname{relint} D$ (interior relative to affine hull); linear inequalities do not need to hold with strict inequality
- There exist many other types of constraint qualifications

Inequality form LP

Primal problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

Dual function

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

Dual problem

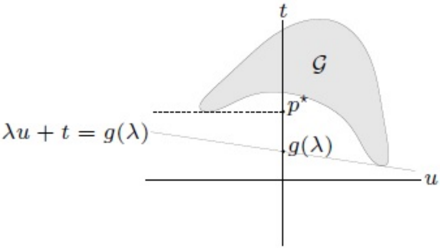
$$\begin{array}{ll} \text{maximize} & -b^T\lambda \\ \text{subject to} & A^T\lambda + c = 0, \lambda \succeq 0 \end{array}$$

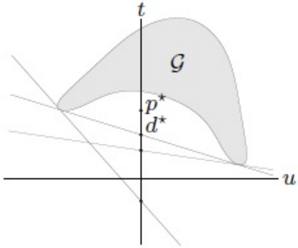
- From Slater's condition: $p^* = d^*$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- In fact, $p^* = d^*$ except when primal and dual are infeasible

Geometric interpretation

For simplicity, consider problem with one constraint $f_1(x) \leq 0$

Interpretation of dual function: $g(\lambda) = \inf_{u,t \in \mathcal{G}}(t + \lambda u)$, where $\mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$





- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$
- Hyperplane intersects t-axis at $t = g(\lambda)$

Complementary slackness

Assume strong duality holds, x^* is primal optimal, (λ^*, ν^*) is dual optimal

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

Hence, the two inequalities hold with equality

- x^* minimizes $L(x, \lambda^*, \nu^*)$
- $\lambda_i^* f_i(x^*) = 0$ for $i = 1, \ldots, m$ (known as complementary slackness):

$$\lambda_i^* > 0 \to f_i(x^*) = 0, \quad f_i(x^*) < 0 \to \lambda_i^* = 0$$

Karush-Kuhn-Tucker (KKT) conditions

The following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i):

1. Primal constraints:
   $f_i(x) \leq 0, i = 1, \ldots, m, h_i(x) = 0, i = 1, \ldots, p$
2. Dual constraints: $\lambda \succeq 0$
3. Complementary slackness: $\lambda_i f_i(x) = 0, i = 1, \ldots, m$
4. Gradient of Lagrangian with respect to x vanishes:

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

If strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

KKT conditions for convex problem

if $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal:

- From complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- From the 4th condition (convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

Hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$

If Slater's condition is satisfied:
x is optimal if and only if there exist λ, ν that satisfy KKT conditions

- Recall that Slater implies strong duality, and dual optimum is attained
- Generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem

Example

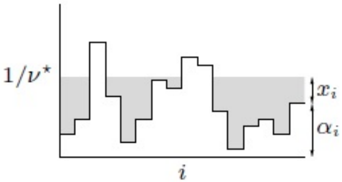
Water-filling (assume $\alpha_i > 0$)

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

KKT conditions:

Interpretation:

- n patches; level of patch i is at height α_i
- Flood area with unit amount of water
- Resulting level is $1/\nu^*$



Duality and problem reformulations

- Equivalent formulations of a problem can lead to very different duals
- Reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting

Common reformulations

- Introduce new variables and equality constraints
- Make explicit constraints implicit or vice-versa
- Transform objective or constraint functions, e.g., replace $f_0(x)$ by $\phi(f_0(x))$ with ϕ convex, increasing

Introducing new variables and equality constraints

$$\text{minimize} \quad f_0(Ax + b)$$

- Dual function is constant:
  $g = \inf_x L(x) = \inf_x f_0(Ax + b) = p^*$
- We have strong duality, but dual is quite useless

Reformulated problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax + b - y = 0 \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T\nu - f_0^*(\nu) \\ \text{subject to} & A^T\nu = 0 \end{array}$$

Dual function follows from

$$\begin{aligned} g(\nu) &= \inf_{x,y}(f_0(y) - \nu^T y + \nu^T Ax + b^T\nu) \\ &= \begin{cases} -f_0^*(\nu) + b^T\nu & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

Implicit constraints

LP with box constraints: primal and dual problem

$$\begin{array}{rl} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -1 \preceq x \preceq 1 \end{array} \qquad \begin{array}{rl} \text{maximize} & -b^T\nu - \mathbf{1}^T\lambda_1 - \mathbf{1}^T\lambda_2 \\ \text{subject to} & c + A^T\nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \lambda_2 \succeq 0 \end{array} \qquad (2)$$

Reformulation with box constraints made implicit

$$\begin{array}{rl} \text{minimize} & f_0(x) = \begin{cases} c^T x & -1 \preceq x \preceq 1 \\ \infty & \text{otherwise} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

Dual function

$$\begin{aligned} g(\nu) &= \inf_{-1 \preceq x \preceq 1} (c^T x + \nu^T (Ax - b)) \\ &= -b^T\nu - \|A^T\nu + c\|_1 \end{aligned}$$

Dual problem: maximize $-b^T\nu - \|A^T\nu + c\|_1$

Summary

- Lagrange dual problem
- Weak and strong duality
- Geometric interpretation
- Optimality conditions
- Duality and problem reformulation