Lagrange Duality

Prof. Daniel P. Palomar

ELEC5470/IEDA6100A - Convex Optimization
The Hong Kong University of Science and Technology (HKUST)
Fall 2020-21
Outline of Lecture

• Lagrangian
• Dual function
• Dual problem
• Weak and strong duality
• KKT conditions
• Summary

(Acknowledgement to Stephen Boyd for material for this lecture.)
Lagrangian

• Consider an optimization problem in standard form (not necessarily convex)

\[
\begin{align*}
\text{minimize} & \quad f_0 (x) \\
\text{subject to} & \quad f_i (x) \leq 0 \quad i = 1, \ldots, m \\
& \quad h_i (x) = 0 \quad i = 1, \ldots, p
\end{align*}
\]

with variable \( x \in \mathbb{R}^n \), domain \( D \), and optimal value \( p^* \).

• The *Lagrangian* is a function \( L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \), with \( \text{dom} \ L = D \times \mathbb{R}^m \times \mathbb{R}^p \), defined as

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

where \( \lambda_i \) is the Lagrange multiplier associated with \( f_i (x) \leq 0 \) and \( \nu_i \) is the Lagrange multiplier associated with \( h_i (x) = 0 \).
Lagrange Dual Function

• The *Lagrange dual function* is defined as the infimum of the Lagrangian over $x$: $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$,

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

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

• Observe that:
  – the infimum is unconstrained (as opposed to the original constrained minimization problem)
  – $g$ is concave regardless of original problem (infimum of affine functions)
  – $g$ can be $-\infty$ for some $\lambda, \nu$
**Lower bound property**: if \( \lambda \geq 0 \), then \( g(\lambda, \nu) \leq p^* \).

**Proof.** Suppose \( \tilde{x} \) is feasible and \( \lambda \geq 0 \). Then,

\[
f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu).
\]

Now choose minimizer of \( f_0(\tilde{x}) \) over all feasible \( \tilde{x} \) to get \( p^* \geq g(\lambda, \nu) \). \( \square \)

- We could try to find the best lower bound by maximizing \( g(\lambda, \nu) \). This is in fact the dual problem.
**Dual Problem**

- The *Lagrange dual problem* is defined as

  $$\max_{\lambda, \nu} g(\lambda, \nu)$$

  subject to \( \lambda \geq 0 \).

- This problem finds the best lower bound on \( p^* \) obtained from the dual function.

- It is a convex optimization (maximization of a concave function and linear constraints).

- The optimal value is denoted \( d^* \).

- \( \lambda, \nu \) are dual feasible if \( \lambda \geq 0 \) and \((\lambda, \nu) \in \text{dom } g \) (the latter implicit constraints can be made explicit in problem formulation).
Example: Least-Norm Solution of Linear Equations

• Consider the problem

$$\begin{align*}
\text{minimize} \quad & x^T x \\
\text{subject to} \quad & Ax = b.
\end{align*}$$

• The Lagrangian is

$$L(x, \nu) = x^T x + \nu^T (Ax - b).$$

• To find the dual function, we need to solve an unconstrained minimization of the Lagrangian. We set the gradient equal to zero

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \implies x = -(1/2) A^T \nu$$
and we plug the solution in $L$ to obtain $g$:

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

- The function $g$ is, as expected, a concave function of $\nu$.

- From the lower bound property, we have

$$p^* \geq -\frac{1}{4} \nu^T A A^T \nu - b^T \nu \text{ for all } \nu.$$

- The dual problem is the QP

$$\max_{\nu} -\frac{1}{4} \nu^T A A^T \nu - b^T \nu.$$
Example: Standard Form LP

- Consider the problem

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b, \quad x \geq 0.
\end{align*}
\]

- The Lagrangian is

\[
L(x, \lambda, \nu) = c^T x + \nu^T (Ax - b) - \lambda^T x
\]

\[
= (c + A^T \nu - \lambda)^T x - b^T \nu.
\]

- \( L \) is a linear function of \( x \) and it is unbounded if the term multiplying \( x \) is nonzero.
• Hence, the dual function is

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

• The function \( g \) is a concave function of \((\lambda, \nu)\) as it is linear on an affine domain.

• From the lower bound property, we have

\[ p^* \geq -b^T \nu \quad \text{if } c + A^T \nu \geq 0. \]

• The dual problem is the LP

\[
\begin{align*}
\text{maximize} & \quad -b^T \nu \\
\text{subject to} & \quad c + A^T \nu \geq 0.
\end{align*}
\]
Example: Two-Way Partitioning

• Consider the problem

\[
\begin{align*}
\text{minimize} & \quad x^T W x \\
\text{subject to} & \quad x_i^2 = 1, \quad i = 1, \ldots, n.
\end{align*}
\]

• It is a nonconvex problem (quadratic equality constraints). The feasible set contains \(2^n\) discrete points.

• The Lagrangian is

\[
L (x, \nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) = x^T (W + \text{diag} (\nu)) x - 1^T \nu.
\]

• \(L\) is a quadratic function of \(x\) and it is unbounded if the matrix \(W + \text{diag} (\nu)\) has a negative eigenvalue.
• Hence, the dual function is

\[
g(\nu) = \inf_x L(x, \nu) = \begin{cases} -1^T \nu & W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise.} \end{cases}
\]

• From the lower bound property, we have

\[
p^* \geq -1^T \nu \quad \text{if } W + \text{diag}(\nu) \succeq 0.
\]

• As an example, if we choose \( \nu = -\lambda_{\text{min}}(W) 1 \), we get the bound

\[
p^* \geq n \lambda_{\text{min}}(W).
\]

• The dual problem is the SDP

\[
\begin{align*}
\text{maximize} & \quad -1^T \nu \\
\text{subject to} & \quad W + \text{diag}(\nu) \succeq 0.
\end{align*}
\]
Weak and Strong Duality

• From the lower bound property, we know that $g(\lambda, \nu) \leq p^*$ for feasible $(\lambda, \nu)$. In particular, for a $(\lambda, \nu)$ that solves the dual problem.

• Hence, weak duality always holds (even for nonconvex problems):

\[ d^* \leq p^*. \]

• The difference $p^* - d^*$ is called duality gap.

• Solving the dual problem may be used to find nontrivial lower bounds for difficult problems.
• Even more interesting is when equality is achieved in weak duality. This is called *strong duality*:

\[ d^* = p^*. \]

• Strong duality means that the duality gap is zero.

• Strong duality:
  – is very desirable (we can solve a difficult problem by solving the dual)
  – 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

- Slater’s constraint qualification is a very simple condition that is satisfied in most cases and ensures strong duality for convex problems.

- Strong duality holds for a convex problem

\[
\begin{align*}
\text{minimize} & \quad f_0 (x) \\
\text{subject to} & \quad f_i (x) \leq 0 \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

if it is strictly feasible, i.e.,

\[
\exists x \in \text{int} \mathcal{D} : \quad f_i (x) < 0 \quad i = 1, \ldots, m, \quad Ax = b.
\]
• It can be relaxed by using $\text{relint} \mathcal{D}$ (interior relative to affine hull) instead of $\text{int} \mathcal{D}$; linear inequalities do not need to hold with strict inequality, ...

• There exist many other types of constraint qualifications.
Example: Inequality Form LP

- Consider the problem

  $$\begin{align*}
  \text{minimize} & \quad c^T x \\
  \text{subject to} & \quad Ax \leq b.
  \end{align*}$$

- The dual problem is

  $$\begin{align*}
  \text{maximize} & \quad -b^T \lambda \\
  \text{subject to} & \quad A^T \lambda + c = 0, \quad \lambda \geq 0.
  \end{align*}$$

- From Slater’s condition: \( p^* = d^* \) if \( A\tilde{x} < b \) for some \( \tilde{x} \).

- In this case, in fact, \( p^* = d^* \) except when primal and dual are infeasible.
Example: Convex QP

• Consider the problem (assume \( P \succeq 0 \))

\[
\begin{align*}
\text{minimize} & \quad x^T P x \\
\text{subject to} & \quad Ax \leq b.
\end{align*}
\]

• The dual problem is

\[
\begin{align*}
\text{maximize} & \quad - \frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\
\text{subject to} & \quad \lambda \geq 0.
\end{align*}
\]

• From Slater’s condition: \( p^* = d^* \) if \( A \tilde{x} < b \) for some \( \tilde{x} \).

• In this case, in fact, \( p^* = d^* \) always.
Example: Nonconvex QP

• Consider the problem

$$\begin{align*}
\text{minimize} & \quad x^T Ax + 2b^T x \\
\text{subject to} & \quad x^T x \leq 1
\end{align*}$$

which is nonconvex in general as $A \not\succeq 0$.

• The dual problem is

$$\begin{align*}
\text{maximize} & \quad -b^T (A + \lambda I)^\# b - \lambda \\
\text{subject to} & \quad A + \lambda I \succeq 0 \\
& \quad b \in \mathcal{R}(A + \lambda I)
\end{align*}$$
which can be rewritten as

$$\max_{t, \lambda} -t - \lambda$$

subject to

$$\begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0.$$ 

• In this case, strong duality holds even though the original problem is nonconvex (not trivial).
Complementary Slackness

• Assume strong duality holds, \( x^* \) is primal optimal and \((\lambda^*, \nu^*)\) is dual optimal. Then,

\[
f_0(x^*) = g(\lambda^*, \nu^*) = \inf_x \left( f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right)
\]

\[
\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*)
\]

\[
\leq f_0(x^*)
\]

• Hence, the two inequalities must hold with equality. Implications:
  – \( x^* \) minimizes \( L(x, \lambda^*, \nu^*) \)
  – \( \lambda_i^* f_i(x^*) = 0 \) for \( i = 1, \ldots, m \); this is called complementary slackness:
    \[
    \lambda_i^* > 0 \implies f_i(x^*) = 0, \quad f_i(x^*) < 0 \implies \lambda_i^* = 0.
    \]
Karush-Kuhn-Tucker (KKT) Conditions

KKT conditions (for differentiable $f_i, h_i$):

1. primal feasibility: $f_i(x) \leq 0, \ i = 1, \ldots, m, \ h_i(x) = 0, \ i = 1, \ldots, p$

2. dual feasibility: $\lambda \geq 0$

3. complementary slackness: $\lambda_i^* f_i(x^*) = 0$ for $i = 1, \ldots, m$

4. zero gradient of Lagrangian with respect to $x$:

$$
\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0
$$
We already known that if strong duality holds and \( x, \lambda, \nu \) are optimal, then they must satisfy the KKT conditions.

What about the opposite statement?

If \( x, \lambda, \nu \) satisfy the KKT conditions for a convex problem, then they are optimal.

\[ f_0(x) = L(x, \lambda, \nu) \]

From complementary slackness, \( f_0(x) = L(x, \lambda, \nu) \) and, from 4th KKT condition and convexity, \( g(\lambda, \nu) = L(x, \lambda, \nu) \). Hence, \( f_0(x) = g(\lambda, \nu) \).

**Theorem.** If a problem is convex and Slater’s condition is satisfied, then \( x \) is optimal if and only if there exists \( \lambda, \nu \) that satisfy the KKT conditions.
Perturbation and Sensitivity Analysis

- Recall the original (unperturbed) optimization problem and its dual:

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0 \quad \forall i \\
& \quad h_i(x) = 0 \quad \forall i
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad g(\lambda, \nu) \\
\text{subject to} & \quad \lambda \geq 0
\end{align*}
\]

- Define the perturbed problem and dual as

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq u_i \quad \forall i \\
& \quad h_i(x) = v_i \quad \forall i
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad g(\lambda, \nu) - u^T \lambda - v^T \nu \\
\text{subject to} & \quad \lambda \geq 0
\end{align*}
\]

- \( x \) is primal variable and \( u, v \) are parameters
- Define \( p^*(u, v) \) as the optimal value as a function of \( u, v \).
• **Global sensitivity**: Suppose strong duality holds for unperturbed problem and $\lambda^*, \nu^*$ are dual optimal for unperturbed problem. Then, from weak duality:

$$p^*(u, v) \geq g(\lambda^*, \nu^*) - u^T \lambda^* - v^T \nu^*$$

$$= p^*(0, 0) - u^T \lambda^* - v^T \nu^*$$

• Interpretation:

  – if $\lambda_i^*$ large: $p^*$ increases a lot if we tighten constraint $i$ ($u_i < 0$)
  – if $\lambda_i^*$ small: $p^*$ does not decrease much if we loosen constraint $i$ ($u_i > 0$)
  – if $\nu_i^*$ large and positive: $p^*$ increases a lot if we take $\nu_i < 0$
  – if $\nu_i^*$ large and negative: $p^*$ increases a lot if we take $\nu_i > 0$
  – etc.
• **Local sensitivity**: Suppose strong duality holds for unperturbed problem, \( \lambda^*, \nu^* \) are dual optimal for unperturbed problem, and \( p^*(u,v) \) is differentiable at \((0,0)\). Then,

\[
\frac{\partial p^*(0,0)}{\partial u_i} = -\lambda^*_i, \quad \frac{\partial p^*(0,0)}{\partial v_i} = -\nu^*_i.
\]

**Proof.** (for \( \lambda^*_i \)) From the global sensitivity result, we have

\[
\frac{\partial p^*(0,0)}{\partial u_i} = \lim_{\epsilon \downarrow 0} \frac{p^*(te_i,0) - p^*(0,0)}{t} \geq \lim_{\epsilon \downarrow 0} \frac{-t\lambda^*_i}{t} = -\lambda^*_i.
\]

\[
\frac{\partial p^*(0,0)}{\partial u_i} = \lim_{\epsilon \uparrow 0} \frac{p^*(te_i,0) - p^*(0,0)}{t} \leq \lim_{\epsilon \uparrow 0} \frac{-t\lambda^*_i}{t} = -\lambda^*_i.
\]

Hence, the equality. \( \square \)
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 tricks:
  – 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 $\phi$ convex and increasing).
Example: Introducing New Variables

• Consider the problem

\[
\min_x \| Ax - b \|_2.
\]

• We can rewrite it as

\[
\begin{align*}
\min_{x,y} & \quad \| y \|_2 \\
\text{subject to} & \quad y = Ax - b.
\end{align*}
\]

• We can then derive the dual problem:

\[
\begin{align*}
\max_{\nu} & \quad b^T \nu \\
\text{subject to} & \quad A^T \nu = 0, \quad \| \nu \|_2 \leq 1.
\end{align*}
\]
Example: Implicit Constraints

• Consider the following LP with box constrains:

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b \\
& \quad -1 \leq x \leq 1
\end{align*}
\]

• The dual problem is

\[
\begin{align*}
\text{maximize} & \quad -b^T \nu - 1^T \lambda_1 - 1^T \lambda_2 \\
\text{subject to} & \quad c + A^T \nu + \lambda_1 - \lambda_2 = 0 \\
& \quad \lambda_1 \geq 0, \quad \lambda_2 \geq 0,
\end{align*}
\]

which does not give much insight.
If, instead, we rewrite the primal problem as

\[
\begin{align*}
\text{minimize} \quad & f_0(x) = \begin{cases} 
  c^T x & -1 \leq x \leq 1 \\
  \infty & \text{otherwise}
\end{cases} \\
\text{subject to} \quad & Ax = b
\end{align*}
\]

then the dual becomes way more insightful:

\[
\begin{align*}
\text{maximize} \quad & -b^T \nu - \|A^T \nu + c\|_1
\end{align*}
\]
Duality for Problems with Generalized Inequalities

- The Lagrange duality can be naturally extended to generalized inequalities of the form

\[ f_i(x) \succeq_{K_i} 0 \]

where \( \succeq_{K_i} \) is a generalized inequality on \( \mathbb{R}^{k_i} \) with respect to the cone \( K_i \).

- The corresponding dual variable has to satisfy

\[ \lambda_i \succeq_{K_i^*} 0 \]

where \( K_i^* \) is the dual cone of \( K_i \).
Semidefinite Programming (SDP)

• Consider the following SDP \((F_i, G \in \mathbb{R}^{k \times k})\):

\[
\begin{aligned}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad x_1 F_1 + \cdots + x_n F_n \preceq G.
\end{aligned}
\]

• The Lagrange multiplier is a matrix \(\Psi \in \mathbb{R}^{k \times k}\) and the Lagrangian

\[
L(x, \Psi) = c^T x + \text{Tr}(\Psi (x_1 F_1 + \cdots + x_n F_n - G))
\]

• The dual problem is

\[
\begin{aligned}
\text{maximize} & \quad -\text{Tr}(\Psi G) \\
\text{subject to} & \quad \text{Tr}(\Psi F_i) + c_i = 0, \ i = 1, \ldots, n \\
& \quad \Psi \succeq 0.
\end{aligned}
\]
Application: Waterfilling Solution

• Consider the maximization of the mutual information in a MIMO channel under Gaussian noise:

$$\begin{align*}
\text{maximize} \quad & \log \det (R_n + HQH^\dagger) \\
\text{subject to} \quad & \text{Tr}(Q) \leq P \\
& Q \succeq 0.
\end{align*}$$

• This problem is convex: the logdet function is concave, the trace constraint is just a linear constraint, and the positive semidefiniteness constraint is an LMI.

• Hence, we can use a general-purpose method such as an interior-point method to solve it in polynomial time.
• However, this problem admits a closed-form solution as can be derived from the KKT conditions.

• The Lagrangian is

\[ L(Q; \mu, \Psi) = -\log \det (R_n + HQH^\dagger) + \mu (\text{Tr}(Q) - P) - \text{Tr}(\Psi Q). \]

• The gradient of the Lagrangian is

\[ \nabla_Q L = -H^\dagger (R_n + HQH^\dagger)^{-1} H + \mu I - \Psi. \]
• The KKT conditions are

\[ \text{Tr} (Q) \leq P, \quad Q \succeq 0 \]
\[ \mu \geq 0, \quad \Psi \succeq 0 \]
\[ H^\dagger (R_n + HQH^\dagger)^{-1} H + \Psi = \mu I \]
\[ \mu (\text{Tr} (Q) - P) = 0, \quad \Psi Q = 0. \]

• Can we find a \( Q \) that satisfies the KKT conditions (together with some dual variables)?
• First, let’s simplify the KKT conditions by defining the so-called whitened channel: \( \tilde{H} = R_n^{-1/2}H \).

• Then, the third KKT condition becomes:

\[
\tilde{H}^\dagger \left( I + \tilde{H}Q\tilde{H}^\dagger \right)^{-1} \tilde{H} + \Psi = \mu I.
\]

• To simplify even further, let’s write the SVD of the channel matrix as \( \tilde{H} = U\Sigma V^\dagger \) (denote the eigenvalues \( \sigma_i \)), obtaining:

\[
\Sigma^\dagger \left( I + \Sigma\tilde{Q}\Sigma^\dagger \right)^{-1} \Sigma + \tilde{\Psi} = \mu I.
\]

where \( \tilde{Q} = V^\dagger QV \) and \( \tilde{\Psi} = V^\dagger \Psi V \).
• The KKT conditions are:

\[
\begin{align*}
\text{Tr}(\tilde{Q}) & \leq P, \quad \tilde{Q} \succeq 0 \\
\mu & \geq 0, \quad \tilde{\Psi} \succeq 0 \\
\Sigma^{\dagger} \left( I + \Sigma \tilde{Q} \Sigma^{\dagger} \right)^{-1} \Sigma + \tilde{\Psi} & = \mu I \\
\mu \left( \text{Tr}(\tilde{Q}) - P \right) & = 0, \quad \tilde{\Psi} \tilde{Q} = 0.
\end{align*}
\]

• At this point, we can make a guess: perhaps the optimal \( \tilde{Q} \) and \( \tilde{\Psi} \) are diagonal? Let’s try ...
• Define $\tilde{Q} = \text{diag}(p)$ ($p$ is the power allocation) and $\tilde{Ψ} = \text{diag}(ψ)$.

• The KKT conditions become:

$$
\sum_{i} p_i \leq P, \quad p_i \geq 0
$$

$$
\mu \geq 0, \quad ψ_i \geq 0
$$

$$
\frac{σ^2_i}{1 + σ^2_i p_i} + ψ_i = \mu
$$

$$
μ \left( \sum_{i} p_i - P \right) = 0, \quad ψ_i p_i = 0.
$$

• Let’s now look into detail at the KKT conditions.
First of all, observe that \( \mu > 0 \), otherwise we would have \( \frac{\sigma_i^2}{1+\sigma_i^2 p_i} + \psi_i = 0 \) which cannot be satisfied.

Let’s distinguish two cases in the power allocation:

- if \( p_i > 0 \), then \( \psi_i = 0 \Rightarrow \frac{\sigma_i^2}{1+\sigma_i^2 p_i} = \mu \Rightarrow p_i = \mu^{-1} - 1/\sigma_i^2 \) (also note that \( \mu = \frac{\sigma_i^2}{1+\sigma_i^2 p_i} < \sigma_i^2 \))

- if \( p_i = 0 \), then \( \sigma_i^2 + \psi_i = \mu \) (note that \( \mu = \sigma_i^2 + \psi_i \geq \sigma_i^2 \)).

Equivalently,

- if \( \sigma_i^2 > \mu \), then \( p_i = \mu^{-1} - 1/\sigma_i^2 \)

- if \( \sigma_i^2 \leq \mu \), then \( p_i = 0 \).
• More compactly, we can write the well-known \textit{waterfilling solution}:

\[
p_i = (\mu^{-1} - 1/\sigma_i^2)^+\]

where $\mu^{-1}$ is called water-level and is chosen to satisfy $\sum_i p_i = P$ (so that all the KKT conditions are satisfied).

• Therefore, the optimal solution is given by

\[
Q^* = V \text{diag}(\mathbf{p}) V^\dagger
\]

where

– the optimal transmit directions are matched to the channel matrix
– the optimal power allocation is the waterfilling.
Summary

- We have introduced the Lagrange duality theory: Lagrangian, dual function, and dual problem.

- We have developed the optimality conditions for convex problems: the KKT conditions.

- We have illustrated the use of the KKT conditions to find the closed-form solution to a problem.

- We have overviewed some additional concepts such as duals of reformulations of problems, sensitivity analysis, generalized inequalities, and SDP.
References

Chapter 5 of
