8.Types of Convex Optimization

SOCP, SDP and Overview of Algorithms

Sun Woo Lim

Feb 13, 2021

Slater's Condition for Strong Duality

1. Basic Version of the Slater's Condition

✓ A **Convex Problem** is given.
✓ If $\exists x \in relint(D)$ s.t. $f_i(x) < 0$, for $i = 1, 2, ..., m$ and $h_j(x) = 0$, for $j = 1, 2, ..., p$,

then $d^* = p^*$.

Moreover, if $p^* > -\infty$, $\exists$ dual optimal point (λ^*, ν^*).

2. Stronger Version of the Slater's Condition

✓ A **Convex Problem** is given.
✓ If $\exists x \in relint(D)$ s.t. $f_i(x) \leq 0$, for f_i : affine, $f_i(x) < 0$, for f_i not affine, and $h_j(x) = 0$, for $j = 1, 2, ..., p$,

then $d^* = p^*$.

Moreover, if $p^* > -\infty$, $\exists$ dual optimal point (λ^*, ν^*).

Relative Interior

✓$relint(S) := \{x \in S | \exists \epsilon > 0 : B_\epsilon(x) \cap aff(S)\}$: Interior of S as a subset of its affine hull.
OK to just regard this just as the interior of the domain at this level.
✓ What is the $relint(co\{(1.5, 2), (3, 1)\})$?

Supporting Hyperplane Theorem, Separating Hyperplane Theorem

Supporting Hyperplane Theorem
For C, a convex subset of $\mathbb{R}^n$ and $x_0 \in \partial(C)$,
there exists $\vec{a} \in \mathbb{R}^n$ s.t. $C \subseteq \{X \in \mathbb{R}^n | a^T x \leq a^T x_0\}$

Separating Hyperplane Theorem
For C, D convex subsets of $\mathbb{R}^n$ and $C \cap D = \phi$,
there exists $\vec{a} \in \mathbb{R}^n$, $b \in \mathbb{R}$ s.t. $a^T x \leq b$, $\forall x \in C$ and $a^T x \geq b, \forall x \in D$.
$\{x | a^T x = b\}$ works as a separating hyperplane separating C and D.

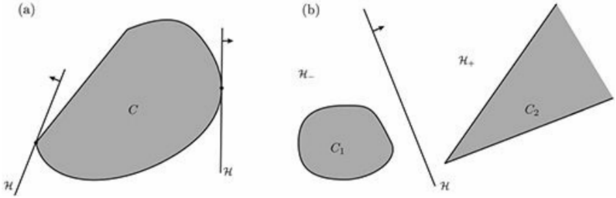


Figure: (a) Supporting Hyperplane; (b) Separating Hyperplane, from Calafiore, El Ghaoui

✓ Important theorems used in proof of **Slater's Condition**.

Geometry of Weak Duality for Non Convex Problem

An Epigraph Viewpoint

Slater's Condition : Proof Idea

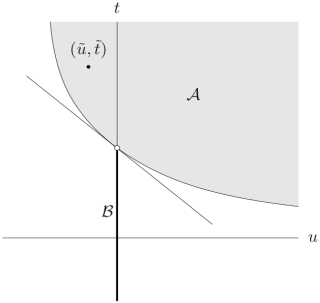


Figure 5.6 Illustration of strong duality proof, for a convex problem that satisfies Slater's constraint qualification. The set $\mathcal{A}$ is shown shaded, and the set $\mathcal{B}$ is the thick vertical line segment, not including the point $(0, p^\star)$, shown as a small open circle. The two sets are convex and do not intersect, so they can be separated by a hyperplane. Slater's constraint qualification guarantees that any separating hyperplane must be nonvertical, since it must pass to the left of the point $(\tilde{u}, \tilde{t}) = (f_1(\tilde{x}), f_0(\tilde{x}))$, where $\tilde{x}$ is strictly feasible.

Figure: Proof Idea of Slater's Condition; from Boyd and Vandenburghe

Types of Convex Optimization Problems : Overview

LP (Linear Programming) $\subseteq QP$ (Quadratic Programming)$\subseteq QCQP$(Quadraticly Constrained Quadratic Programming) $\subseteq SOCP$ (Second Order Cone Programming) $\subseteq SDP$ (Semi-Definite Programming)

Separately, GP (Geometric Programming) can be formed as Convex Optimization Problems.

✓ LP General Form : $min_x c^T x + d$ s.t. $Ax \leq b$, $Gx = h$.

✓ QP Standard Form : $min_x \frac{1}{2} x^T K x + c^T x + d$ s.t. $Ax \leq b$, $Gx = h$ where $K \in \mathbb{S}^n_+$ is a fixed (given) PSD matrix.

✓ QCQP in General Form : $min_x \frac{1}{2} x^T P_0 x + q_0^T x + r_0$ s.t. $\frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0, i = 1, ..., m$, $Ax = b$ where $P_0, P_i, i = 1, 2, ..., m \in \mathbb{S}^n_+$.

✓ SOCP in General Form : $min_x c^T x + d$ s.t. $||A_i x + b_i||_2 \leq c_i^T x + d_i, i = 1, ..., m$.

✓ SDP in inequality Form : $min_x c^T x$ s.t. $F(x) := F_0 + x_1 F_1 + x_m F_m \in \mathbb{S}^n_+$ for $F_0, F_1, ..., F_m \in \mathbb{S}^n$

✓ General Form? Standard Form? Inequality Form? Same **Optimization Class** can be expressed in many forms.

✓ GP : $min_x f_0(x)$ s.t. $f_i(x) \leq 1$, $i = 1, ..., m$, $h_j(x) = 1$, $j = 1, ..., p$ where $f_0, f_1, ..., f_m$ are **posynomials** and $h_1, ..., h_p$ are **monomials**. GP **itself** is not a convex optimization problem.

LP review

✓ LP General Form : $min_x c^T x(+d)$ s.t. $Ax \leq b$, $Gx = h$. vs LP Standard Form : $min_x c^T x(+d)$ s.t. $Gx = h$, $x \geq 0$.

We showed how to interchange two forms

Useful facts of LP

1. Feasible set of LP is a **Polyhedron**, an intersection of finite number of **half spaces**. Usually, deal with polytopes, bounded **polyhedrons**.
2. The solution of LP lies on the **vertex** of a polyhedron : proven last time
3. Dual of LP is LP and unless both the primal and the dual are infesible, $p^* = d^*$: can prove by the Slater's condition.

LP Example : Utility Maximization in Economics (Easy)

Dane can select x_1 days of vacation in the NYC and x_2 days of vacation in a lake house. Solve Dane's constrained utility maximization problem.

$max_{x_1,x_2} x_2(x_1+4)$ s.t. $p_1x_1 + p_2x_2 \leq M$.

$= -min_{x_1,x_2} - x_2(x_1+4)$ s.t. $p_1x_1 + p_2x_2 \leq M$.

$L(x_1, x_2, \lambda) = -x_2(x_1+4) + \lambda(p_1x_1 + p_2x_2 - M)$ s.t. $\lambda \geq 0$.

Use lagrangian stationarity over x_1, x_2, λ and solve $\rightarrow x_1^* = \frac{M-4p_1}{2p_1}, x_2^* = \frac{M+4p_1}{2p_2}$. $p^* = x_2^*(x_1^*+4)$

QP review

✓ QP Standard Form : $min_x \frac{1}{2}x^T Kx + c^T x + d$ s.t. $Ax \leq b$, $Gx = h$ where $K \in \mathbb{S}_+^n$ is a fixed (given) PSD matrix.
✓ Another QP Form : $min_x \frac{1}{2}x^T Kx + c^T x + d$ s.t. $Ax \leq b$, $x \geq 0$ where $K \in \mathbb{S}_+^n$ is a fixed (given) PSD matrix.
Interchange two forms in the same way as in **LP**.

Useful facts of QP

1. Compared to LP where the optimal point lies in the **vertex** of a polytope, not necessarily in QP : QP is more general.
2. Dual of QP is QP and unless both the primal and the dual are infesible, $p^* = d^*$: can prove by the Slater's condition.

QP Examples dealt last time

1. Ordinary Least Squares and Equality Constrained Least Squares
2. Ridge and Lasso. LASSO especially for image compression
3. Optimization over PSD matrices
4. Projecting a point onto a hyperplane
5. Markowitz Portfolio Optimization
6. Hard Margin Support Vector Machine $min_{w,b} \frac{1}{2}||w||_2^2$ s.t. $y_i(w^T x_i + b) \geq 1$.

QCQP in General Form

$min_x \frac{1}{2}x^T P_0 x + q_0^T x + r_0$ s.t. $\frac{1}{2}x^T P_i x + q_i^T x + r_i \leq 0, i = 1, ..., m,\ Ax = b$ where $P_0, P_i, i = 1, 2, ..., m \in \mathbb{S}_+^n$.

QCQP is more general than QP

✓ **Extreme points** : points on a convex set that are only possible to be expressed as convex combination of itself.

Feasible of QP is a **Polyhedron** having only finite number of extreme points.

The feasible set of $QCQP$ can have a continuum of extreme points.

QCQP Dual

QCQP Dual is SOCP, not QCQP!
Instead of directly addressing the QCQP dual, let's rather form QCQP as SOCP and find its dual.

✓ Reminder) SOCP in General Form : $min_x c^T x + d$ s.t. $||A_i x + b_i||_2 \leq c_i^T x + d_i, i = 1, ..., m.$

Forming QCQP as SOCP

QCQP in General Form

$min_x \frac{1}{2}x^T P_0 x + q_0^T x(+r_0)$ s.t. $\frac{1}{2}x^T P_i x + q_i^T x + r_i \leq 0, i = 1, ..., m,\ Ax = b$ where $P_0, P_i, i = 1, 2, ..., m \in \mathbb{S}_+^n$.

Step1) Introduce a slack variable t that represents something bigger than or equal to $\frac{1}{2}x^T P_0 x$.
$min_{x,t} q_0^T x + t$ s.t. $\frac{1}{2}x^T P_0 x \leq t,\ \frac{1}{2}x^T P_i x + q_i^T x + r_i \leq 0,\ i = 1, ..., m,\ Ax = b$.

Step2)
$min_{x,t} q_0^T x + t$ s.t.
$||\begin{bmatrix} \sqrt{2}P_0^{0.5}x \\ t-1 \end{bmatrix}||_2 \leq t+1$ and
$||\begin{bmatrix} \sqrt{2}P_i^{0.5}x \\ \mathsf{q}_i^T x + r_i + 1 \end{bmatrix}||_2 \leq q_i^T x + r_i - 1, i = 1, ..., m$ and $Ax = b$.

✓ SOCP in General Form : $min_x c^T x(+d)$ s.t. $||A_i x + b_i||_2 \leq c_i^T x + d_i, i = 1, ..., m$.

Cone, Second Order Cone, Dual Cone, Norm Cone

✓ $K \subseteq \mathbb{R}^n$ is called a **cone** if $\forall x \in K, \alpha \geq 0, \alpha x \in K$.

✓ Let $K \subseteq \mathbb{R}^n$ be a cone. $K^* := \{y \in \mathbb{R}^n | y^T x \geq 0, \forall x \in K\}$ is called a **Dual Cone** of K.

✓ Let $||\cdot||$ be a norm on $\mathbb{R}^n$ with dual norm $||\cdot||_*$. **Norm Cone** is a cone in $\mathbb{R}^{n+1}$ given by $K := \{(\vec{x}, t) \in \mathbb{R}^{n+1} : ||x|| \leq t\}$.

Then, the "dual of a norm cone" $K^* = \{(\vec{z}, s) \in \mathbb{R}^{n+1} : ||z||_* \leq s\}$.

This information is used in solving SOCP Duality.

Showing the Duality of SOCP using the Dual of Norm Cone

$p^* = min_x c^T x(+d)$ s.t. $||A_i x + b_i||_2 \leq c_i^T x + d_i, i = 1, ..., m.$

$d^* = max_{u,\lambda} \sum_{i=1}^{m} u_i^T b_i - \lambda_i d_i$ s.t. $\sum_{i=1}^{m} (A_i^T u_i - \lambda_i c_i) = -c$, $||u_i||_2 \leq \lambda_i$, $i = 1, ..., m$, $\lambda_i \geq 0$.

There is a special name for this dual, called **Conic Dual**.
Strong duality holds if

SOCP Duality Examples

SOCP Example

SOCP Example

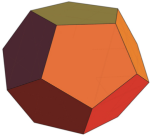
Semi-Definite Programming and Linear Matrix Inequality

SDP in inequality form : : $min_x c^T x$ s.t. $F(x) := F_0 + x_1 F_1 + x_m F_m \succeq 0$ for $F_0, F_1, ..., F_m \in \mathbb{S}^n$.
SDP in standard form : : $min_{X \in \mathbb{S}^n} tr(CX)$ s.t. $tr(A_i X) = b_i$, $i = 1, ..., m$ and $X \succeq 0$
where $C, A_1, ..., A_m \in \mathbb{S}^n$, $b_1, ..., b_m \in \mathbb{R}$

Remember that $\mathbb{S}^n$ is a vector space ($\mathbb{S}^n_+$ is not).
Given fixed symmetric matrices $F_0, F_1, ..., F_m \in \mathbb{S}^n$, $\{F(x) := F_0 + x_1 F_1 + x_m F_m | x \in \mathbb{R}^m\}$ is an **affine subspace** of $\mathbb{S}^n$.
The equation $F_0, F_1, ..., F_m \succeq 0$ is called a **Linear Matrix Inequality**.

If the primal problem is expressed in multiple **LMI**'s, can **incorporate** them into a single **LMI** in **block diagonal form**.

Given $F_0, F_1, ..., F_m \in \mathbb{S}^n$, $\{x \in \mathbb{R}^m | F(x) := F_0 + x_1 F_1 + x_m F_m \succeq 0\}$ is called a **spectrahedron.**



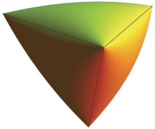


Figure: Left : Polyhedron, Right : Spectrahedron

✓ Every spectrahedron is a convex set, while the converse is not true.

LMI and Spectrahedron Analysis

Spectrahedron example : Cylinder in $\mathbb{R}^3$

Rank Analysis : For what $x, y, z \in \mathbb{R}$, $rank(F(x, y, z)) \leq 3$?

Important Test for Posite Semidefiniteness : $A \in \mathbb{S}^n$ is PSD $\leftrightarrow$ all principal minors are non negative.

LMI Example

LP as SDP

QP as SDP

QCQP as SDP

SOCP as SDP

SDP duality

✓ Even if **both** primal and dual are **feasible**, strong duality may fail.

✓ If **either** the primal and dual is **strictly feasible**, $p^* = d^*$.

✓ If **both** are strictly feasible, both the **primal and dual optimal sets** are nonempty.

SDP duality Example

SDP duality Example

SDP duality Example

$p^* = min_X tr(CX)$ s.t. $tr(X) = b$, $X \in \mathbb{S}^n_+$.
Find p^* and express it in terms of b and eigenvalues of C.

Step 1) Slater's Condition. **I can find** $\frac{b}{n}I$ that **satisfies the equality constraint** and **strictly satisfies the inequality constraint**. Then, $p^* = d^*$.

Step 2) $d^* = max_{\nu \in \mathbb{R}} b\nu$ s.t. $C - \nu I \in \mathbb{S}^n_+$.

The inequality constraint means "All eigenvalues of C are bigger than or equal to ν".

Step 3) $\nu^* = \lambda_{min}(C)$ and $p^* = d^* = b\lambda_{min}(C)$.

Descent Algorithms : General Case

We use **algorithms** to **sequentially** approach the **solution** to an optimization problem : finding the **minimum** of a function.

Here, only deal with **unconstrained minimization problem**.

Why **sequential algorithms**? Usually in **high dimensional setting!**

1. Humans : Hard to analytically find the solution! Cannot **see** the optimum point.
2. Computers : Hard to analytically find the solution! Grid Approximation? Too many Grids.

General Descent Algorithm

1. Start at $x_0 \in dom(f)$.
2. Descent Step : $x_{k+1} = x_k + s_k v_k$, s_k is called a **step size** and v_k is called a **descent direction**.

In order for v_k to be a descent direction, only need $f(x_k + s_k v_k) < f(x_k)$ for all small $s_k > 0$.

Generally, doesn't even need f : differentiable.

For differentiable f, **usually** set $v_k = -\nabla f(x_k)$:
"A ftn **locally decreases** in the direction of **negative gradient**.
If **gradient** is zero, I'm at an **extremum (local minimum / maximum / saddle point)**"

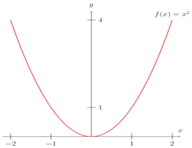
: called **Gradient Descent**.

Gradient Descent : Convex Function

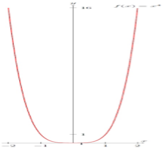
If the objective function is **convex** and $dom(f_0) = \mathbb{R}^n$: if $\nabla f(x^*) = 0, x^*$ is a global minimizer of f.
Why? $f(y) \geq f(x) + \nabla f(x)^T(y - x), \forall y \in \mathbb{R}^n$: **FOC** of convexity for a differentiable function.

However, even for **convex functions**, the claim "the function decreases in the direction of its negative gradient" is only true, **locally**.

Even for a convex functions, selection of **step size** is very important.



(a) $y = x^2$ works for sufficiently small fixed step size



(b) $y = x^4$ may not work even for sufficiently small fixed step size

Practice Question) $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \frac{1}{4}||x||_2^4$. As always, $x^* := argmin_x f(x)$.
Since f is convex w.r.t. x, I use descent algorithms, especially, the gradient descent : $x_{t+1} = x_t - \eta \nabla f(x_t)$ for **fixed** $\eta > 0$.
Find x^* and suppose $||x_0|| = c > 0$. Find the range of η in terms of c s.t. the gradient descent algorithm converge to x^*.