Mathematical Optimization Theory Review

Yinyu Ye

http://www.stanford.edu/~yyye

Chapters 1, 2.1-6, 3.1-6, 6.1-4, 7.2, 11.3, 11.6-8, 14.1-2, Appendix A, B.

Structured/Disciplined Convex Optimization Again: Conic Linear Programming (CLP)

$$
\begin{array}{lll}
(CLP) & \text{minimize} & \mathbf{c} \bullet \mathbf{x} \\
 & \text{subject to} & \mathbf{a}_i \bullet \mathbf{x} = b_i, i = 1, 2, ..., m, \ \mathbf{x} \in K, \\
 & & (\ \mathcal{A}^T \mathbf{x} = \mathbf{b}\ ),
\end{array}
$$

where K is a closed and pointed convex cone.

Linear Programming (LP): $\mathbf{c}, \mathbf{a}_i, \mathbf{x} \in \mathcal{R}^n$ and $K = \mathcal{R}^n_+$

Second-Order Cone Programming (SOCP): $\mathbf{c}, \mathbf{a}_i, \mathbf{x} \in \mathcal{R}^n$ and $K = SOC = \{\mathbf{x} : \ x_1 \geq \|\mathbf{x}_{-1}\|_2\}$; where $\mathbf{x}_{-1}$ is the vector $(x_2; ...; x_n) \in R^{n-1}$.

Semidefinite Programming (SDP): $\mathbf{c}, \mathbf{a}_i, \mathbf{x} \in \mathcal{S}^n$ and $K = \mathcal{S}^n_+$

Cone K can be also a product of different cones, that is, $\mathbf{x} = (\mathbf{x}_1; \mathbf{x}_2; \ldots)$ where $\mathbf{x}_1 \in K_1$, $\mathbf{x}_2 \in K_2$,... and so on with linear constraints:

$$\mathcal{A}_1\mathbf{x}_1 + \mathcal{A}_2\mathbf{x}_2 + ... = \mathbf{b}.$$

Cone, Convex Cone and Dual

- A set K is a cone if $\mathbf{x} \in K$ implies $\alpha\mathbf{x} \in K$ for all $\alpha > 0$
- The intersection of cones is a cone
- A convex cone is a cone and also a convex set
- A pointed cone is a cone that does not contain a line
- Dual of Cone K:

$$K^* := \{\mathbf{y} : \ \mathbf{x} \bullet \mathbf{y} \geq 0 \quad \text{for all} \quad \mathbf{x} \in K\}.$$

Theorem 1 *The dual is always a closed convex cone, and the dual of the dual is the closure of convex hall of* K.

Cone Examples

- Example 1: The n-dimensional non-negative orthant, $\mathcal{R}^n_+ = \{\mathbf{x} \in \mathcal{R}^n : \mathbf{x} \geq \mathbf{0}\}$, is a convex cone. Its dual is itself.

- Example 2: The set of all PSD matrices in $\mathcal{S}^n$, $\mathcal{S}^n_+$, is a convex cone, called the PSD matrix cone. Its dual is itself.

- Example 3: The set $\{(t; \mathbf{x}) \in \mathcal{R}^{n+1} : t \geq \|\mathbf{x}\|_p\}$ for a $p \geq 1$ is a convex cone in $\mathcal{R}^{n+1}$, called the p-order cone. Its dual is the q-order cone with $\frac{1}{p} + \frac{1}{q} = 1$.

- The dual of the second-order cone ($p = 2$) is itself.

Recall LP, SOCP, and SDP Examples

$$(LP) \quad \text{minimize} \quad 2x_1 + x_2 + x_3$$
$$\text{subject to} \quad x_1 + x_2 + x_3 = 1,$$
$$(x_1; x_2; x_3) \geq \mathbf{0}.$$

$$(SOCP) \quad \text{minimize} \quad 2x_1 + x_2 + x_3$$
$$\text{subject to} \quad x_1 + x_2 + x_3 = 1,$$
$$x_1 - \sqrt{x_2^2 + x_3^2} \geq 0.$$

$$(SDP) \quad \text{minimize} \quad 2x_1 + x_2 + x_3$$
$$\text{subject to} \quad x_1 + x_2 + x_3 = 1,$$
$$\begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \succeq \mathbf{0}.$$

(SDP) can be structurally rewritten as

$$
\begin{array}{ll}
\text{minimize} & \begin{pmatrix} 2 & .5 \\ .5 & 1 \end{pmatrix} \bullet \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \\
\text{subject to} & \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix} \bullet \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} = 1, \\
& \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \succeq \mathbf{0},
\end{array}
$$

that is

$$
\mathbf{c} = \begin{pmatrix} 2 & .5 \\ .5 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{a}_1 = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}.
$$

Dual of Conic LP

$$(CLD) \quad \text{maximize} \quad \mathbf{b}^T\mathbf{y}$$
$$\text{subject to} \quad \sum_i^m y_i \mathbf{a}_i + \mathbf{s} = \mathbf{c}, \ (\mathcal{A}^T\mathbf{y} + \mathbf{s} = \mathbf{c}), \ \mathbf{s} \in K^*,$$

where $\mathbf{y} \in \mathcal{R}^m$, $\mathbf{s}$ is called the dual slack vector/matrix, and K^* is the dual cone of K.

Here, operator $\mathcal{A}\mathbf{x}$ and Adjoint-Operator $\mathcal{A}^T\mathbf{y}$ mimic matrix-vector production $A\mathbf{x}$ and its transpose operation $A^T\mathbf{y}$, where

$$\mathcal{A} = (\mathbf{a}_1; \mathbf{a}_2; ...; \mathbf{a}_m), \quad \mathcal{A}\mathbf{x} = (\mathbf{a}_1 \bullet \mathbf{x}; ...; \mathbf{a}_m \bullet \mathbf{x}), \quad \text{and} \quad A^T\mathbf{y} = \sum_i y_i \mathbf{a}_i^T.$$

LP, SOCP, and SDP Examples Again

$$\begin{array}{ll} \min & 2x_1 + x_2 + x_3 \\ \text{s. t.} & x_1 + x_2 + x_3 = 1, \\ & (x_1; x_2; x_3) \geq \mathbf{0}. \end{array} \qquad \begin{array}{ll} \max & y \\ \text{s.t.} & \mathbf{e} \cdot y + \mathbf{s} = (2;\ 1;\ 1), \\ & (s_1; s_2; s_3) \geq \mathbf{0}. \end{array}$$

$$\begin{array}{ll} \min & 2x_1 + x_2 + x_3 \\ \text{s.t.} & x_1 + x_2 + x_3 = 1, \\ & x_1 - \sqrt{x_2^2 + x_3^2} \geq 0. \end{array} \qquad \begin{array}{ll} \max & y \\ \text{s.t.} & \mathbf{e} \cdot y + \mathbf{s} = (2;\ 1;\ 1), \\ & s_1 - \sqrt{s_2^2 + s_3^2} \geq 0. \end{array}$$

For the SOCP case: $2 - y \geq \sqrt{2(1-y)^2}$. Since $y = 1$ is feasible for the dual, $y^* \geq 1$ so that the dual constraint becomes $2 - y \geq \sqrt{2}(y-1)$ or $y \leq \sqrt{2}$. Thus, $y^* = \sqrt{2}$, and there is no duality gap.

$$\begin{array}{ll} \text{minimize} & \begin{pmatrix} 2 & .5 \\ .5 & 1 \end{pmatrix} \bullet \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \\ \text{subject to} & \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix} \bullet \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} = 1, \\ & \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \succeq \mathbf{0}, \end{array}$$

$$\begin{array}{ll} \text{maximize} & y \\ \text{subject to} & \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix} y + \mathbf{s} = \begin{pmatrix} 2 & .5 \\ .5 & 1 \end{pmatrix}, \\ & \mathbf{s} = \begin{pmatrix} s_1 & s_2 \\ s_2 & s_3 \end{pmatrix} \succeq \mathbf{0}. \end{array}$$

CLP Duality Theorems

Theorem 2 *(Weak duality theorem)* $\mathbf{c} \bullet \mathbf{x} - \mathbf{b}^T \mathbf{y} = \mathbf{x} \bullet \mathbf{s} \geq 0$ *for any feasible* $\mathbf{x}$ *of (CLP) and* $(\mathbf{y}, \mathbf{s})$ *of (CLD).*

The weak duality theorem shows that a feasible solution to either problem yields a bound on the value of the other problem. We call $\mathbf{c} \bullet \mathbf{x} - \mathbf{b}^T \mathbf{y}$ the duality gap.

Corollary 1 *Let* $\mathbf{x}^* \in \mathcal{F}_p$ *and* $(\mathbf{y}^*, \mathbf{s}^*) \in \mathcal{F}_d$*. Then,* $\mathbf{c} \bullet \mathbf{x}^* = \mathbf{b}^T \mathbf{y}^*$ *implies that* $\mathbf{x}^*$ *is optimal for (CLP) and* $(\mathbf{y}^*, \mathbf{s}^*)$ *is optimal for (CLD).*

Is the reverse also true? That is, given $\mathbf{x}^*$ optimal for (CLP), then there is $(\mathbf{y}^*, \mathbf{s}^*)$ feasible for (CLD) and $\mathbf{c} \bullet \mathbf{x}^* = \mathbf{b}^T \mathbf{y}^*$?

This is called the Strong Duality Theorem: it is “rue” when $K = \mathcal{R}^n_+$, but not true in general CLP.

LP and LD Relations

Dual \ Primal	F-B	F-UB	IF
F-B	☺		
F-UB			☹
IF		☹	☹

$$
\begin{array}{llllll}
\min & -x_1 & - & x_2 & & \\
\text{s.t.} & x_1 & - & x_2 & = & 1 \\
 & -x_1 & + & x_2 & = & 1 \\
 & x_1, & & x_2 & \geq & 0
\end{array}
$$

$$
\begin{array}{llllll}
\max & y_1 & + & y_2 & & \\
\text{s.t.} & y_1 & - & y_2 & \leq & -1 \\
 & -y_1 & + & y_2 & \leq & -1
\end{array}
$$

A case that neither (LP) nor (LD) is feasible: $\mathbf{c} = (-1;\ 0), \quad A = (0,\ -1), \quad b = 1.$

How to test the LP or LD constraint set is feasible or not using the relation table? The Farkas Lemma!

LP Optimality Conditions and Solution Support

$$\left\{ (\mathbf{x}, \mathbf{y}, \mathbf{s}) \in (\mathcal{R}^n_+, \mathcal{R}^m, \mathcal{R}^n_+) : \begin{array}{rcl} \mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y} & = & \mathbf{0} \\ A\mathbf{x} & = & \mathbf{b} \\ -A^T\mathbf{y} - \mathbf{s} & = & -\mathbf{c} \end{array} \right\}; \quad \text{or}$$

$$\begin{array}{rcl} \mathbf{x}. \cdot \mathbf{s} & = & \mathbf{0} \\ A\mathbf{x} & = & \mathbf{b} \\ -A^T\mathbf{y} - \mathbf{s} & = & -\mathbf{c}. \end{array}$$

Let $\mathbf{x}^*$ and $\mathbf{s}^*$ be optimal solutions with zero duality gap. Then

$$|\text{supp}(\mathbf{x}^*)| + |\text{supp}(\mathbf{s}^*)| \leq n.$$

There always exist $\mathbf{x}^*$ and $\mathbf{s}^*$ such that the sum of support sizes of $\mathbf{x}^*$ and $\mathbf{s}^*$ equal n: called a strict complemetarity pair. Geometrically, they are in the interior of the optimal solution sets.

If there is one $\mathbf{s}^*$ such that $|\text{supp}(\mathbf{s}^*)| \geq n - d$, then the support size for all $\mathbf{x}^*$ is at most d,

The CLP and CLD Relations

Dual \ Primal	F-B	F-UB	IF
F-B	☺		☹
F-UB			☹
IF	☹	☹	☹

$$\begin{array}{lll} \text{min} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \bullet X & \\ \text{s.t.} & \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \bullet X & = 0 \\ & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \bullet X & = 2 \\ & X & \succeq \mathbf{0} \end{array}$$

$$\begin{array}{llll} \text{max} & 2y_2 & & \\ \text{s.t.} & y_1 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + y_2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + S & = & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ & S & \succeq & \mathbf{0} \end{array}$$

The Dual is feasible and bounded, but Primal is infeasible.

Test the CLP or CLD constraint set feasibility?

Optimality and Complementarity Conditions for SDP

$$\begin{aligned} \mathbf{c} \bullet X - \mathbf{b}^T \mathbf{y} &= 0 \\ \mathcal{A}X &= \mathbf{b} \\ -\mathcal{A}^T \mathbf{y} - S &= -\mathbf{c} \\ X, S &\succeq \mathbf{0} \end{aligned}, \quad (1)$$

$$\begin{aligned} XS &= \mathbf{0} \\ \mathcal{A}X &= \mathbf{b} \\ -\mathcal{A}^T \mathbf{y} - S &= -\mathbf{c} \\ X, S &\succeq \mathbf{0} \end{aligned} \quad (2)$$

Transportation Dual: Economic Interpretation

$$
\begin{array}{lll}
\text{min} & \sum_{i=1}^m \sum_{j=1}^n c_{ij}x_{ij} & \\
\text{s.t.} & \sum_{j=1}^n x_{ij} & = s_i, \ \forall i = 1, ..., m \\
 & \sum_{i=1}^m x_{ij} & = d_j, \ \forall j = 1, ..., n \\
 & x_{ij} & \geq 0, \ \forall i, j.
\end{array}
$$

$$
\begin{array}{lll}
\text{max} & \sum_{i=1}^m s_i u_i + \sum_{j=1}^n d_j v_j & \\
\text{s.t.} & u_i + v_j & \leq c_{ij}, \ \forall i, j.
\end{array}
$$

u_i: supply site unit price

v_i: demand site unit price

$u_i + v_j \leq c_{ij}$: incentive/competitiveness

Algorithmic Applications: Optimal Value Function and Shadow Prices

$$z(\mathbf{b}) = \quad \text{minimize} \quad \mathbf{c}^T\mathbf{x}$$
$$\text{subject to} \quad A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}.$$

Suppose a new right-hand-vector $\mathbf{b}^+$ such that

$$b_k^+ = b_k + \delta \quad \text{and} \quad b_i^+ = b_i,\ \forall i \neq k.$$

Then, the optimal dual solution $\mathbf{y}^*$ has a property

$$y_k^* = (z(\mathbf{b}^+) - z(\mathbf{b}))/\delta$$

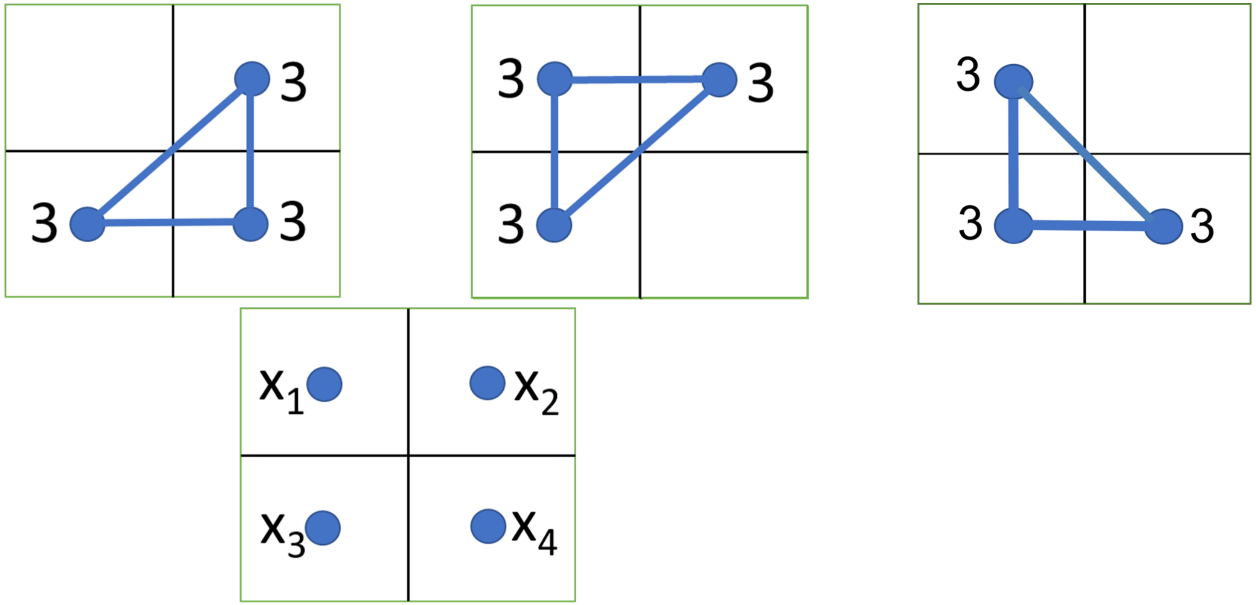
as long as $\mathbf{y}^*$ remains the dual optimal solution for $\mathbf{b}^+$, because

$$z(\mathbf{b}^+) = (\mathbf{b}^+)^T\mathbf{y}^* = z(\mathbf{b}) + \delta \cdot y_k^*.$$

Thus, the optimal dual value is the rate of the net change of the optimal objective value over the net change of an entry of the right-hand-vector resources, i.e.,

$$\nabla z(\mathbf{b}) = \mathbf{y}^*.$$

Application in the Wassestein Barycenter Problem



Find distribution of $x_i, i = 1, 2, 3, 4$ to minimize

$$\text{min} \quad WD_l(\mathbf{x}) + WD_m(\mathbf{x}) + WD_r(\mathbf{x})$$

$$\text{s.t.} \quad x_1 + x_2 + x_3 + x_4 = 9, \qquad x_i \geq 0, \; i = 1, 2, 3, 4.$$

The objective is a nonlinear function, but its gradient vector $\nabla WD_l(\mathbf{x})$, $\nabla WD_m(\mathbf{x})$ and $\nabla WD_l(\mathbf{x})$ are shadow prices of the three sub-transportation problems –popularly used in Hierarchy Optimization.

(WBCgradient3.m of Chapter 8)

The Dual of the Reinforcement Learning LP

Recall the cost-to-go value of the reinforcement learning LP problem:

$$
\begin{array}{lrcl}
\text{maximize}_{\mathbf{y}} & \sum_{i=1}^{m} y_i & & \\
\text{subject to} & y_1 - \gamma \mathbf{p}_j^T \mathbf{y} & \leq & c_j,\ j \in \mathcal{A}_1 \\
 & & \ldots & \\
 & y_i - \gamma \mathbf{p}_j^T \mathbf{y} & \leq & c_j,\ j \in \mathcal{A}_i \\
 & & \ldots & \\
 & y_m - \gamma \mathbf{p}_j^T \mathbf{y} & \leq & c_j,\ j \in \mathcal{A}_m.
\end{array}
$$

$$
\begin{array}{lccccc}
\text{minimize}_{\mathbf{x}} & \sum_{j \in \mathcal{A}_1} c_j x_j + & \ldots & + \sum_{j \in \mathcal{A}_m} c_j x_j & & \\
\text{subject to} & \sum_{j \in \mathcal{A}_1} (\mathbf{e}_1 - \gamma \mathbf{p}_j) x_j + & \ldots & + \sum_{j \in \mathcal{A}_m} (\mathbf{e}_m - \gamma \mathbf{p}_j) x_j & = & \mathbf{e}, \\
 & \ldots & x_j & \ldots & \geq & 0,\ \forall j,
\end{array}
$$

where $\mathbf{e}_i$ is the unit vector with 1 at the ith position and 0 everywhere else.

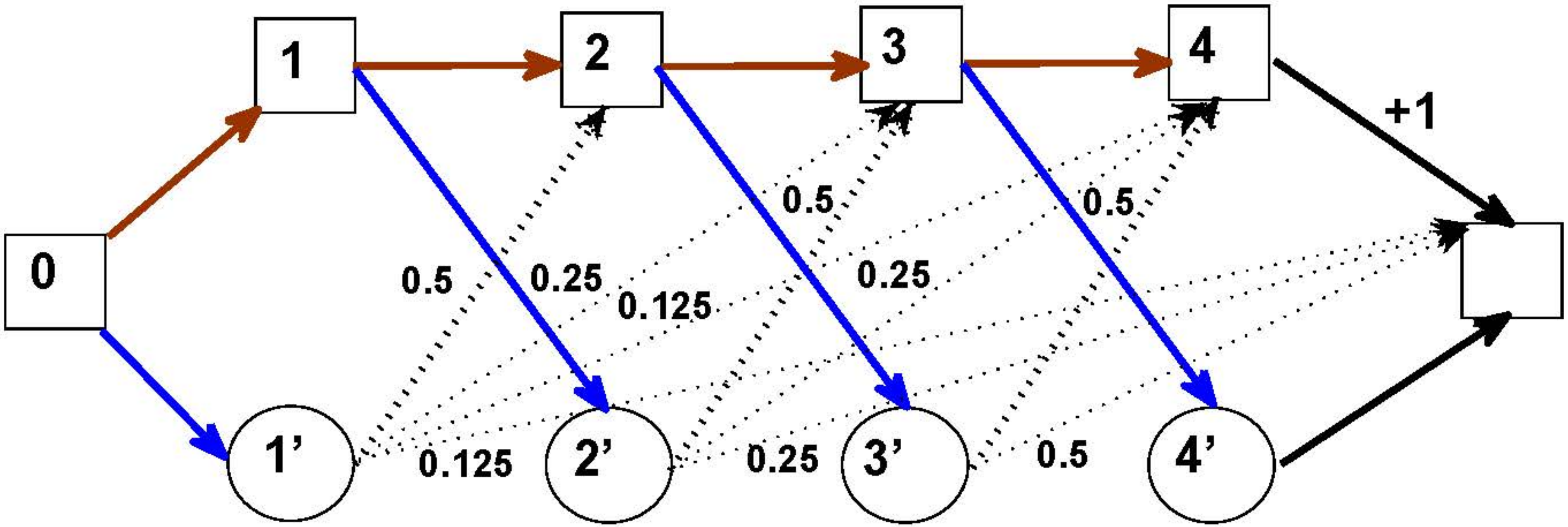
Interpretation of the Dual of the RL-LP

Variable x_j, $j \in \mathcal{A}_i$, is the state-action frequency or called flux, or the expected present value of the number of times that an individual is in state i and takes state-action j.

Thus, solving the problem entails choosing a state-action frequencies/fluxes that minimizes the expected present value of total costs for the infinite horizon, where the RHS is $(1; 1; 1; 1; 1; 1)$:

x:	(0_1)	(0_2)	(1_1)	(1_2)	(2_1)	(2_2)	(3_1)	(3_2)	(4_1)	(5_1)	b
c:	0	0	0	0	0	0	0	0	1	0	
(0)	1	1	0	0	0	0	0	0	0	0	1
(1)	$-\gamma$	0	1	1	0	0	0	0	0	0	1
(2)	0	$-\gamma/2$	$-\gamma$	0	1	1	0	0	0	0	1
(3)	0	$-\gamma/4$	0	$-\gamma/2$	$-\gamma$	0	1	1	0	0	1
(4)	0	$-\gamma/8$	0	$-\gamma/4$	0	$-\gamma/2$	$-\gamma$	0	1	0	1
(5)	0	$-\gamma/8$	0	$-\gamma/4$	0	$-\gamma/2$	0	$-\gamma$	$-\gamma$	$1-\gamma$	1

where state 5 is the absorbing state that has a infinite loops to itself.



The optimal dual solution is

$$x^*_{01} = 1,\ x^*_{11} = 1 + \gamma,\ x^*_{21} = 1 + \gamma + \gamma^2,\ x^*_{32} = 1 + \gamma + \gamma^2 + \gamma^3,\ x^*_{41} = 1,$$
$$x^*_{51} = \frac{1+2\gamma+\gamma^2+\gamma^3+\gamma^4}{1-\gamma}.$$

(Sect2_2MazerunLP.m of Chapter 2)

The Maze Runner Example: Complementarity Condition

The LP optimal Cost-to-Go values are $y_1^* = 0, y_1^* = 0, y_2^* = 0, y_3^* = 0, y_4^* = 1$:

$$
\begin{array}{lrl}
\text{maximize}_{\mathbf{y}} & y_0 + y_1 + y_2 + y_3 + y_4 + y_5 & \\
\text{subject to} & y_0 - \gamma y_1 & \leq 0, \ (x_{01}^* = 1) \\
& y_0 - \gamma(0.5y_2 + 0.25y_3 + 0.125y_4) & \leq 0, \ (x_{02}^* = 0) \\
& y_1 - \gamma y_2 & \leq 0, \ (x_{11}^* = 1 + \gamma) \\
& y_1 - \gamma(0.5y_3 + 0.25y_4) & \leq 0, \ (x_{12}^* = 0) \\
& y_2 - \gamma y_3 & \leq 0, \ (x_{21}^* = 1 + \gamma + \gamma^2) \\
& y_2 - \gamma(0.5y_4) & \leq 0, \ (x_{22}^* = 0) \\
& y_3 - \gamma y_4 & \leq 0, \ (x_{31}^* = 0) \\
& y_3 & \leq 0, \ (x_{32}^* = 1 + \gamma + \gamma^2 + \gamma^3) \\
& y_4 - \gamma y_5 & \leq 1, \ (x_{41}^* = 1) \\
& y_5 - \gamma y_5 & = 0. \ (x_{51}^* = \frac{1+2\gamma+\gamma^2+\gamma^3+\gamma^4}{1-\gamma})
\end{array}
$$

Dual of Information Markets

$$
\begin{array}{lrl}
\text{max} & \pi^T \mathbf{x} - z & \\
\text{s.t.} & A\mathbf{x} - \mathbf{e} \cdot z & \leq \mathbf{0}, \\
& \mathbf{x} & \leq \mathbf{q}, \\
& \mathbf{x} & \geq 0.
\end{array}
$$

$\pi^T \mathbf{x}$: the optimistic amount can be collected.

z: the worst-case amount need to pay to the winning bids.

$$
\begin{array}{lrl}
\text{min} & \mathbf{q}^T \mathbf{y} & \\
\text{s.t.} & A^T \mathbf{p} + \mathbf{y} & \geq \pi, \\
& \mathbf{e}^T \mathbf{p} & = 1, \\
& (\mathbf{p}, \mathbf{y}) & \geq 0.
\end{array}
$$

$\mathbf{p}$ represents the state prices or probability distributions.

Dual Interpretation: Regression using Important Data Sampling

Note that

$$y_j = \max\{0,\ \pi_j - \mathbf{a}_j^T\mathbf{p}\},\ \forall j.$$

so that

$$\begin{array}{lrl} \min & \sum_j \max\{0,\ \pi_j - \mathbf{a}_j^T\mathbf{p}\} & \\ \text{s.t.} & \mathbf{e}^T\mathbf{p} & = 1, \\ & \mathbf{p} & \geq 0. \end{array}$$

The $\max\{0,\ \cdot\}$ is called ReLu function in AI.

Dual Interpretation: Find the probability estimations such that low-bids are automatically uncounted/removed. (Sect2_2WorldcupLP.m of Chapter 2)

World Cup Information Market Result

Order:	#1	#2	#3	#4	#5	State Price
Argentina	1	0	1	1	0	0.2
Brazil	1	0	0	1	1	0.35
Italy	1	0	1	1	0	0.2
Germany	0	1	0	1	1	0.25
France	0	0	1	0	0	0
Bidding Price: π	0.75	0.35	0.4	0.95	0.75	
Quantity limit: $\mathbf{q}$	10	5	10	10	5	
Order fill: $\mathbf{x}^*$	5	5	5	0	5	

Question: How to make the dual prices unique and the market Online?

Recall SNL: SOCP Relaxation for SNL

System of SOCP Feasibility for $\mathbf{x}_i \in R^2$:

$$\|\mathbf{x}_i - \mathbf{x}_j\| \le d_{ij}, \ \forall\,(i,j) \in N_x, \ i < j,$$
$$\|\mathbf{a}_k - \mathbf{x}_j\| \le d_{kj}, \ \forall\,(k,j) \in N_a,$$

where $\mathbf{a}_k$ are points whose locations are known.

Consider the case where a single unknown point $\mathbf{x}_1$ is connected to three anchors $\mathbf{a}_k, \ k = 1, 2, 3$ on R^2:

$$\|\mathbf{a}_k - \mathbf{x}\| \le d_k, \ k = 1, 2, 3$$

Optimality Condition of the SOCP Relaxation: One Sensor and Three Anchors

Then, the optimality conditions would be

$$\mathbf{z}_k = (\lambda_k / d_k)(\mathbf{a}_k - \mathbf{x})$$

and

$$\sum_k (\lambda_k / d_k)(\mathbf{a}_k - \mathbf{x}) = \mathbf{0}$$

where λ_k's are the three dual variables. It represents a positive force in direction $\mathbf{a}_k - \mathbf{x}$, and the total forces should be balanced along the three directions.

If the true location of the sensor, say $\mathbf{b}$, is in the convex-hull of the three anchors, these conditions are achievable so that the optimal solution of the SOCP relaxation is exact, that is, $\mathbf{x}^* = \mathbf{b}$.

What happen if it is NOT?

Recall SDP Relaxation for SNL

Find a symmetric matrix $Z \in \mathbf{R}^{(2+n)\times(2+n)}$ such that

$$
\begin{aligned}
Z_{1:2,1:2} &= I \\
(\mathbf{0}; \mathbf{e}_i - \mathbf{e}_j)(\mathbf{0}; \mathbf{e}_i - \mathbf{e}_j)^T \bullet Z &= d_{ij}^2, \ \forall\, i, j \in N_x, \ i < j, \\
(\mathbf{a}_k; -\mathbf{e}_j)(\mathbf{a}_k; -\mathbf{e}_j)^T \bullet Z &= d_{kj}^2, \ \forall\, k, j \in N_a, \\
Z &\succeq \mathbf{0}.
\end{aligned}
$$

This is semidefinite programming feasibility system (with a null objective).

When this relaxation is exact?

One case is that the single unknown point $\mathbf{x}_1$ is connected to three anchors $\mathbf{a}_k, \ k = 1, 2, 3$.

In general, if the rank of a feasible Z is 2, then it solves the original graph relaxation problem.

Duality Theorem for SNL

Theorem 3 *Let $\bar{Z}$ be a feasible solution for SDP and $\bar{U}$ be an optimal slack matrix of the dual. Then,*

1. *complementarity condition holds: $\bar{Z} \bullet \bar{U} = 0$ or $\bar{Z}\bar{U} = \mathbf{0}$;*
2. *$Rank(\bar{Z}) + Rank(\bar{U}) \leq 2 + n$;*
3. *$Rank(\bar{Z}) \geq 2$ and $Rank(\bar{U}) \leq n$.*

An immediate result from the theorem is the following:

Corollary 2 *If an optimal dual slack matrix has rank n, then every solution of the SDP has rank 2, that is, the SDP relaxation solves the original problem exactly.*

Theoretical Analyses on SNL-SDP Relaxation

A sensor network is 2-universally-localizable (UL) if there is a unique localization in $\mathbf{R}^2$ and there is no $x_j \in \mathbf{R}^h,\ j = 1, ..., n$, where $h > 2$, such that

$$\|x_i - x_j\|^2 = d_{ij}^2,\ \forall\, i, j \in N_x,\ i < j,$$
$$\|(a_k; \mathbf{0}) - x_j\|^2 = \hat{d}_{kj}^2,\ \forall\, k, j \in N_a.$$

The latter says that the problem cannot be localized in a higher dimension space where anchor points are simply augmented to $(a_k; \mathbf{0}) \in \mathbf{R}^h$, $k = 1, ..., m$.

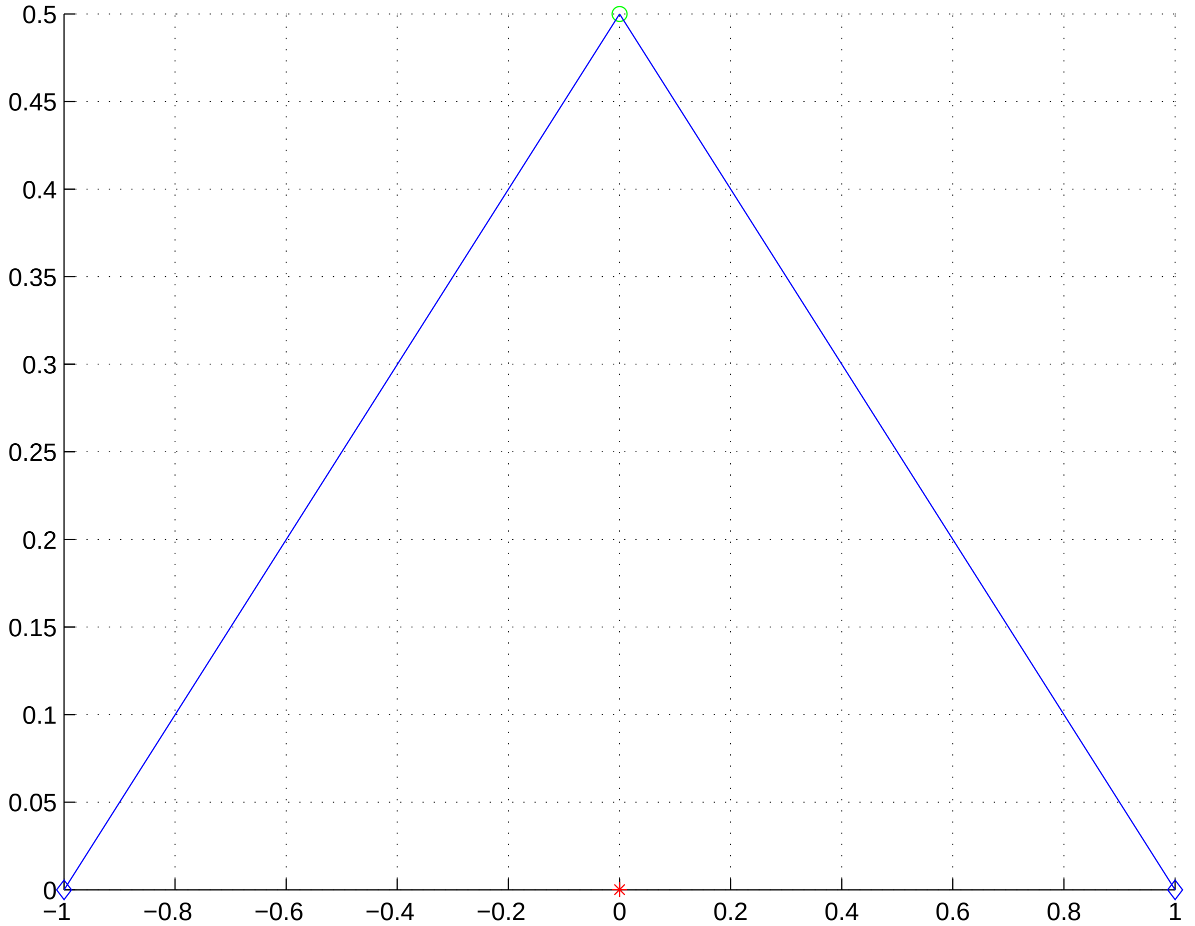


Figure 1: One sensor-Two anchors: Not Localizable

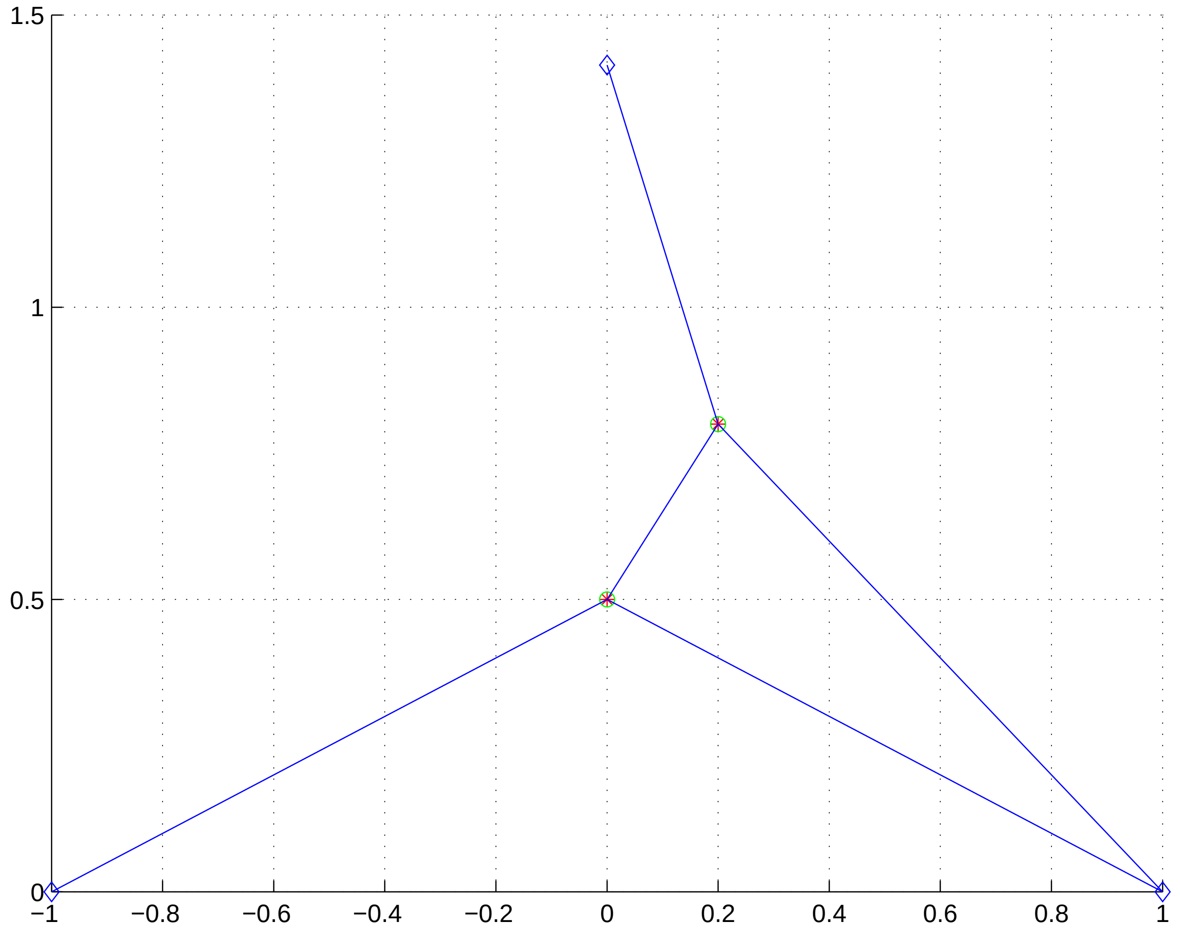


Figure 2: Two sensor-Three anchors: (Strongly) Localizable

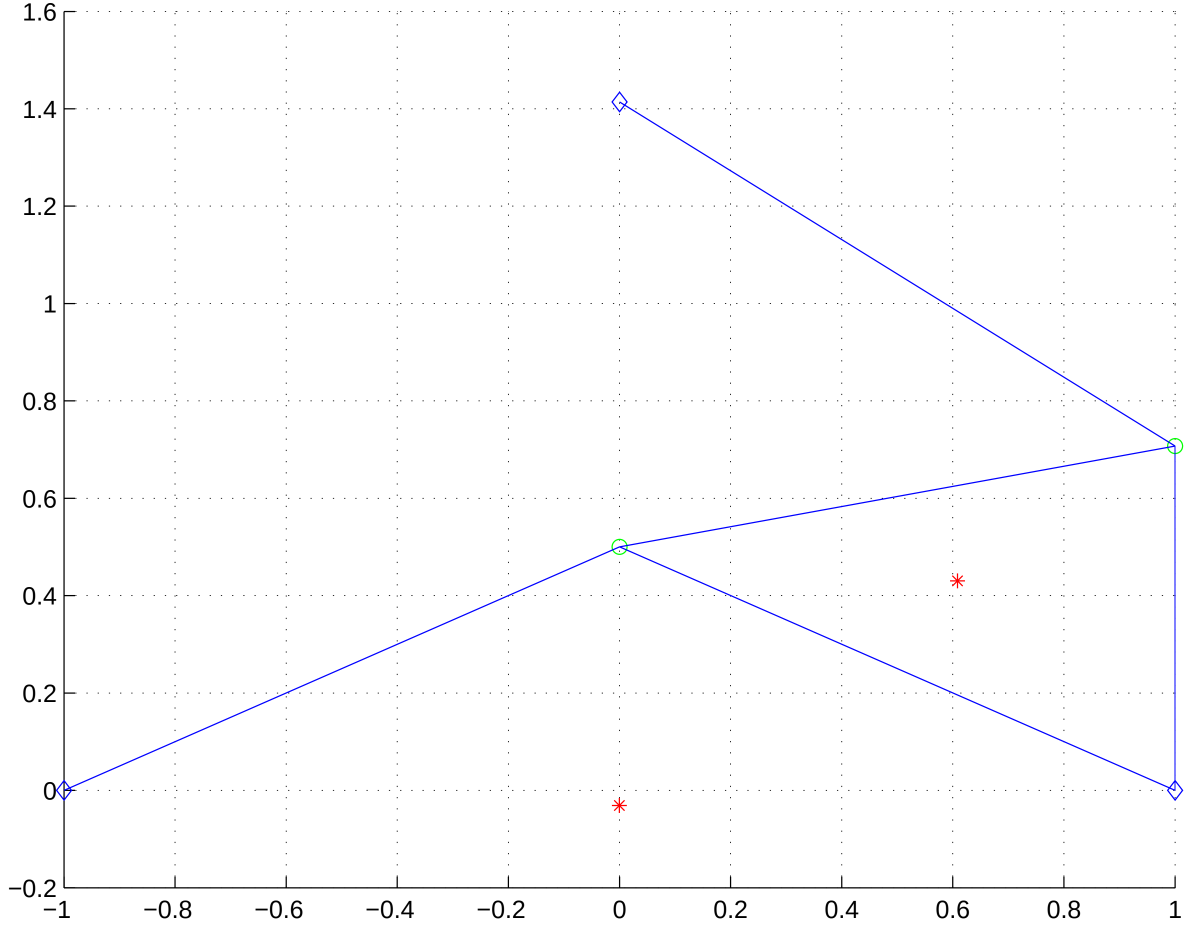


Figure 3: Two sensor-Three anchors: Not Localizable

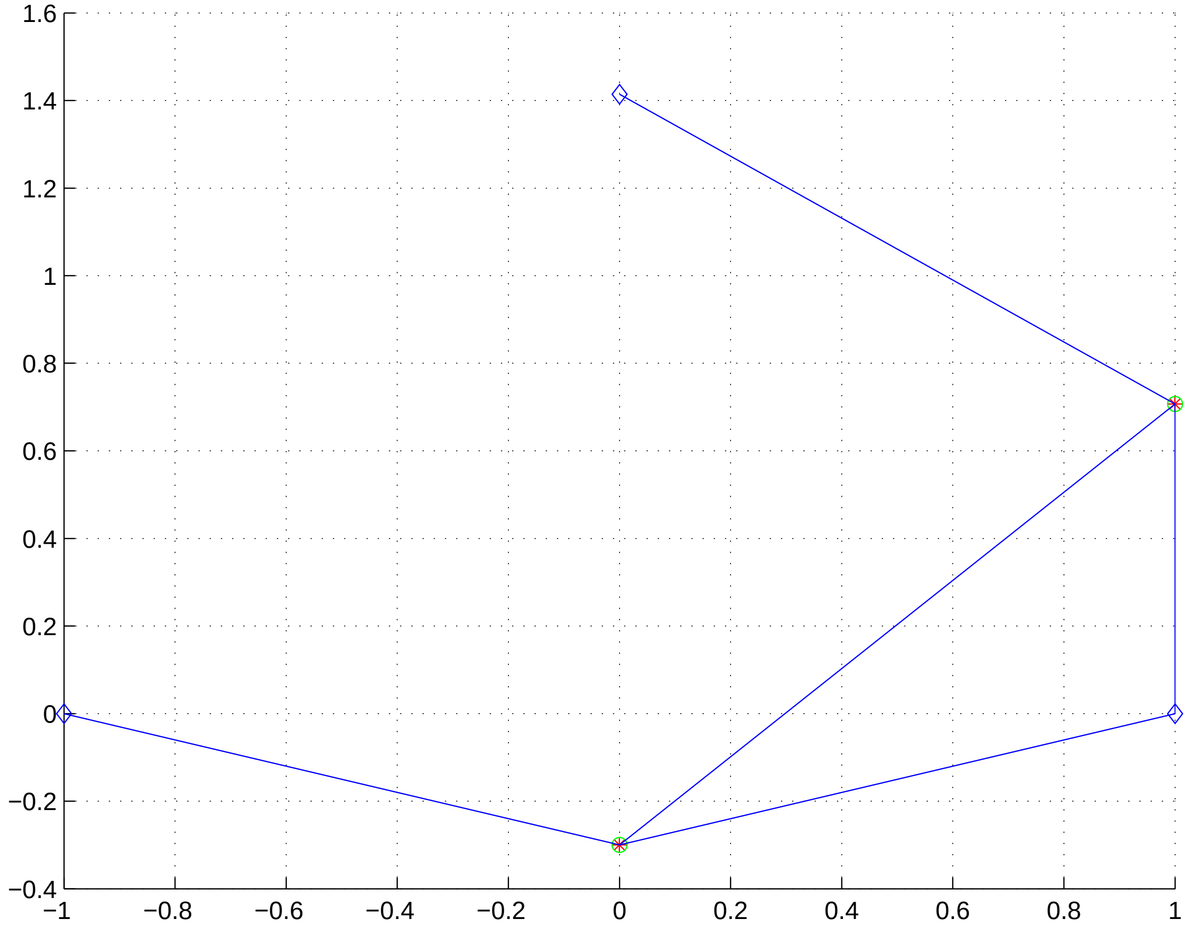


Figure 4: Two sensor-Three anchors: (Strongly) Localizable

Universally-Localizable Problems (ULP)

Theorem 4 *The following SNL problems are Universally-Localizable:*

- *If every edge length is specified, then the sensor network is 2-universally-localizable (Schoenberg 1942).*
- *There is a sensor network (trilateral graph), with $O(n)$ edge lengths specified, that is 2-universally-localizable (So 2007).*
- *If one sensor with its edge lengths to at least three anchors (in general positions) specified, then it is 2-universally-localizable (So and Y 2005).*

(SDPsnldsdp.m of Chapter 6)

ULPs Can be Localized as Convex Optimization

Theorem 5 *(So and Y 2005) The following statements are equivalent:*

1. *The sensor network is 2-universally-localizable;*

2. *The max-rank solution of the SDP relaxation has rank 2;*

3. *The solution matrix has $Y = X^T X$ or $Tr(Y - X^T X) = 0$.*

When an optimal dual (stress) slack matrix has rank n, then the problem is 2-strongly-localizable-problem (SLP). This is a sub-class of ULP.

Example: if one sensor with its edge lengths to three anchors (in general positions) are specified, then it is 2-strongly-localizable.

Unstructured Optimization

Now consider the general (constrained) optimization (GCO) problem:

$$\text{(P)} \quad \text{minimize} \quad f(\mathbf{x})$$

$$\text{subject to} \quad c_i(\mathbf{x}) \; (\leq, = \geq) \; 0 \quad i = 1, ..., m$$

Optimality Condition Theories help to identify and verify when a solution is optimal.

General Optimization: First-Order Necessary Conditions for Constrained Optimization

Consider constraints $\{\mathbf{x} : \mathbf{h}(\mathbf{x}) = \mathbf{0},\ \mathbf{c}(\mathbf{x}) \geq \mathbf{0}.\}$

Theorem 6 *(First-Order or KKT Optimality Condition) Let $\bar{\mathbf{x}}$ be a (local) minimizer of (GCO) and it is a regular point of $\{\mathbf{x} : \mathbf{h}(\mathbf{x}) = \mathbf{0},\ c_i(\mathbf{x}) = 0, i \in \mathcal{A}_{\bar{\mathbf{x}}}\}$. Then, for some multipliers $(\bar{\mathbf{y}}, \bar{\mathbf{s}} \geq \mathbf{0})$*

$$\nabla f(\bar{\mathbf{x}}) = \bar{\mathbf{y}}^T \nabla \mathbf{h}(\bar{\mathbf{x}}) + \bar{\mathbf{s}}^T \nabla \mathbf{c}(\bar{\mathbf{x}}) \tag{3}$$

and (complementarity slackness)

$$\bar{s}_i c_i(\bar{\mathbf{x}}) = 0,\ \forall i.$$

$\bar{\mathbf{x}}$ being a regular point is often referred as a Constraint Qualification condition.

A solution who satisfies these conditions is called an KKT point or solution of (GCO) – any local minimizer $\bar{\mathbf{x}}$, if it is also a regular point, must be an KKT solution; but the reverse may not be true.

KKT via the Lagrangian Function

It is more convenient to introduce the Lagrangian Function associated with generally constrained optimization:

$$L(\mathbf{x}, \mathbf{y}, \mathbf{s}) = f(\mathbf{x}) - \mathbf{y}^T \mathbf{h}(\mathbf{x}) - \mathbf{s}^T \mathbf{c}(\mathbf{x}),$$

where multipliers $\mathbf{y}$ of the equality constraints are "free" and $\mathbf{s} \geq \mathbf{0}$ for the "greater or equal to" inequality constraints, so that the KKT condition (3) can be written as

$$\nabla_{\mathbf{x}} L(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}) = \mathbf{0}.$$

Lagrangian Function can be viewed as a "Penalty" function aggregated with the original objective function plus the penalized terms on constraint violations.

In theory, one can adjust the penalty multipliers $(\mathbf{y}, \mathbf{s} \geq \mathbf{0})$ to repeatedly solve the following so-called Lagrangian Relaxation Problem:

$$(LRP) \quad \min_{\mathbf{x}} \quad L(\mathbf{x}, \mathbf{y}, \mathbf{s}).$$

Summary of KKT Conditions for More General GCO

$$\text{(GCO)} \quad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & c_i(\mathbf{x}) \quad (\leq, =, \geq) \quad 0, \; i = 1, ..., m, \end{array} \quad \text{(Original Problem Constraints (OPC))}$$

the Lagrangian Function is given by

$$L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T \mathbf{c}(\mathbf{x}) = f(\mathbf{x}) - \sum_{i=1}^{m} y_i c_i(\mathbf{x}).$$

For any feasible point $\mathbf{x}$ of (GCO) define the active constraint set by $\mathcal{A}_{\mathbf{x}} = \{i : \; c_i(\mathbf{x}) = 0\}$. Let $\bar{\mathbf{x}}$ be a local minimizer for (GCO) and $\bar{\mathbf{x}}$ is a regular point on the hypersurface of the active constraints Then there exist multipliers $\bar{\mathbf{y}}$ such that

$$\begin{array}{rcll} \nabla f(\bar{\mathbf{x}}) & = & \bar{\mathbf{y}}^T \nabla \mathbf{c}(\bar{\mathbf{x}}) & \text{(Lagrangian Derivative Conditions (LDC))} \\ \bar{y}_i & (\leq, '\text{free}', \geq) & 0, \; i = 1, ..., m, & \text{(Multiplier Sign Constraints (MSC))} \\ \bar{y}_i c_i(\bar{\mathbf{x}}) & = & 0, & \text{(Complementarity Slackness Conditions (CSC))}. \end{array}$$

Second-Order Necessary Conditions for Constrained Optimization

Now in addition we assume all functions are in C^2, that is, twice continuously differentiable. Recall the tangent linear sub-space at $\bar{\mathbf{x}}$:

$$T_{\bar{\mathbf{x}}} := \{\mathbf{z} : \nabla \mathbf{h}(\bar{\mathbf{x}})\mathbf{z} = \mathbf{0},\ \nabla c_i(\bar{\mathbf{x}})\mathbf{z} = 0\ \forall i \in \mathcal{A}_{\bar{\mathbf{x}}}\}.$$

Theorem 7 *Let $\bar{\mathbf{x}}$ be a (local) minimizer of (GCO) and a regular point of hypersurface $\{\mathbf{x} :\ \mathbf{h}(\mathbf{x}) = \mathbf{0},\ c_i(\mathbf{x}) = 0, i \in \mathcal{A}_{\bar{\mathbf{x}}}\}$, and let $\bar{\mathbf{y}}, \bar{\mathbf{s}}$ denote Lagrange multipliers such that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})$ satisfies the (first-order) KKT conditions of (GCO). Then, it is necessary to have*

$$\mathbf{d}^T\ \nabla^2_{\mathbf{x}} L(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})\mathbf{d} \geq 0 \qquad \forall\ \mathbf{d} \in T_{\bar{\mathbf{x}}}.$$

The Hessian of the Lagrangian function need to be positive semidefinite on the tangent-space.

Second-Order Sufficient Conditions for GCO

Theorem 8 *Let $\bar{\mathbf{x}}$ be a regular point of (GCO) with* **equality constraints only** *and let $\bar{\mathbf{y}}$ be the Lagrange multipliers such that $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ satisfies the (first-order) KKT conditions of (GCO). Then, if in addition*

$$\mathbf{d}^T \nabla^2_{\mathbf{x}} L(\bar{\mathbf{x}}, \bar{\mathbf{y}})\mathbf{d} > 0 \qquad \forall\, \mathbf{0} \neq \mathbf{d} \in T_{\bar{\mathbf{x}}},$$

then $\bar{\mathbf{x}}$ is a local minimizer of (GCO).

See the proof in Chapter 11.5 of LY.

The SOSC for general (GCO) is proved in Chapter 11.8 of LY.

min $(x_1)^2 + (x_2)^2$ s.t. $(x_1)^2/4 + (x_2)^2 - 1 = 0$

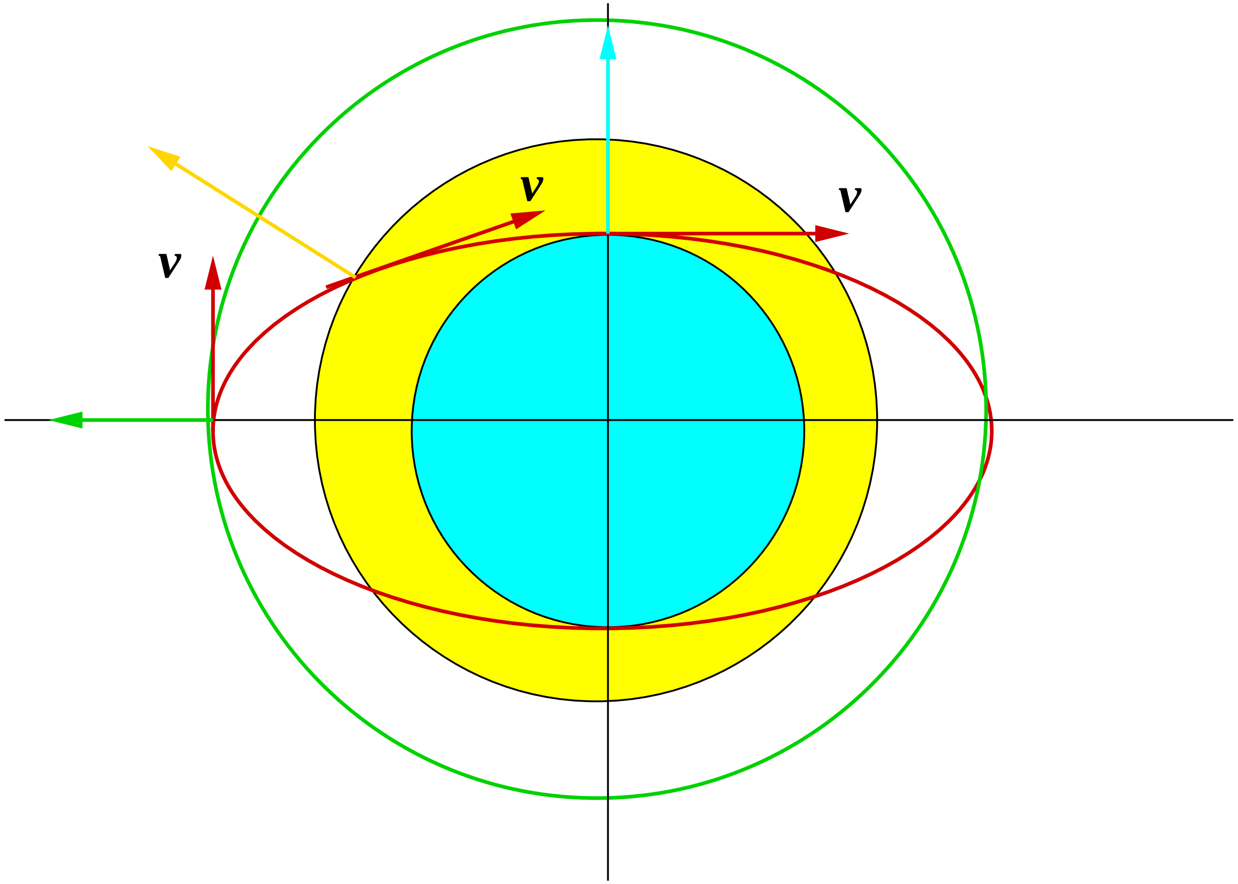


Figure 5: FONC and SONC for Constrained Minimization

More General Lagrangian Functions and The Lagrangian Dual

Consider the general constrained optimization again:

$$
\text{(GCO)} \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & c_i(\mathbf{x}) \quad (\leq, =, \geq) \quad 0, \; i = 1, ..., m, \end{array}
$$

For Lagrange Multipliers.

$$
Y := \{y_i \quad (\leq, '\text{free}', \geq) \quad 0, \; i = 1, ..., m\},
$$

the Lagrangian Function is again given by

$$
L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T \mathbf{c}(\mathbf{x}) = f(\mathbf{x}) - \sum_{i=1}^{m} y_i c_i(\mathbf{x}), \; \mathbf{y} \in Y.
$$

We now develop the Lagrangian Duality theory as an alternative to Conic Duality theory. For general nonlinear constraints, the Lagrangian Duality theory is more applicable.

Toy Example Again

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

$$L(\mathbf{x},\mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T\mathbf{c}(\mathbf{x}) = f(\mathbf{x}) - \sum_{i=1}^{2} y_i c_i(\mathbf{x}) =$$

$$= (x_1-1)^2+(x_2-1)^2 - y_1(x_1+2x_2-1) - y_2(2x_1+x_2-1),\ (y_1;y_2)\le \mathbf{0}$$

where

$$\nabla L_x(\mathbf{x},\mathbf{y}) = \begin{pmatrix} 2(x_1-1)-y_1-2y_2 \\ 2(x_2-1)-2y_1-y_2 \end{pmatrix}$$

The Lagrangian Relaxation Problem

For given multipliers $\mathbf{y} \in Y$, consider problem

$$(LRP) \quad \text{inf} \quad L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T \mathbf{c}(\mathbf{x})$$
$$\text{s.t.} \quad \mathbf{x} \in R^n.$$

Again, $\mathbf{y}_i$ can be viewed as a penalty weight/parameter to penalize constraint violation of $c_i(\mathbf{x})$.

In the toy example, for given $(y_1; y_2) \leq \mathbf{0}$, the LRP is:

$$\text{inf} \quad (x_1 - 1)^2 + (x_2 - 1)^2 - y_1(x_1 + 2x_2 - 1) - y_2(2x_1 + x_2 - 1)$$
$$\text{s.t.} \quad (x_1; x_2) \in R^2,$$

and it has a close form solution $\mathbf{x}$ for any given $\mathbf{y}$:

$$x_1 = \frac{y_1 + 2y_2}{2} + 1 \quad \text{and} \quad x_2 = \frac{2y_1 + y_2}{2} + 1$$

with the minimal or infimum value function $= -1.25y_1^2 - 1.25y_2^2 - 2y_1y_2 - 2y_1 - 2y_2$.

Inf-Value Function as the Dual Objective

For any $\mathbf{y} \in Y$, the minimal value function (including unbounded from below or infeasible cases) and the Lagrangian Dual Problem (LDP) are given by:

$$\phi(\mathbf{y}) := \quad \inf_{\mathbf{x}} \quad L(\mathbf{x}, \mathbf{y}), \quad \text{s.t.} \quad \mathbf{x} \in R^n.$$

$$(LDP) \quad \sup_{\mathbf{y}} \quad \phi(\mathbf{y}), \quad \text{s.t.} \quad \mathbf{y} \in Y.$$

Theorem 9 *The Lagrangian dual objective $\phi(\mathbf{y})$ is a concave function.*

Proof: For any given two multiply vectors $\mathbf{y}^1 \in Y$ and $\mathbf{y}^2 \in Y$,

$$\begin{aligned}
\phi(\alpha\mathbf{y}^1 + (1-\alpha)\mathbf{y}^2) &= \inf_{\mathbf{x}} L(\mathbf{x}, \alpha\mathbf{y}^1 + (1-\alpha)\mathbf{y}^2) \\
&= \inf_{\mathbf{x}}[f(\mathbf{x}) - (\alpha\mathbf{y}^1 + (1-\alpha)\mathbf{y}^2)^T\mathbf{c}(\mathbf{x})] \\
&= \inf_{\mathbf{x}}[\alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{x}) - \alpha(\mathbf{y}^1)^T\mathbf{c}(\mathbf{x}) - (1-\alpha)(\mathbf{y}^2)^T\mathbf{c}(\mathbf{x})] \\
&= \inf_{\mathbf{x}}[\alpha L(\mathbf{x}, \mathbf{y}^1) + (1-\alpha)L(\mathbf{x}, \mathbf{y}^2)] \\
&\geq \alpha[\inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}^1)] + (1-\alpha)[\inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}^2)] \\
&= \alpha\phi(\mathbf{y}^1) + (1-\alpha)\phi(\mathbf{y}^2),
\end{aligned}$$

Dual Objective Establishes a Lower Bound

Theorem 10 *(Weak duality theorem) For every* $\mathbf{y} \in Y$*, the Lagrangian dual function* $\phi(\mathbf{y})$ *is less or equal to the infimum value of the original GCO problem.*

Proof:

$$\begin{aligned} \phi(\mathbf{y}) &= \inf_{\mathbf{x}} \{ f(\mathbf{x}) - \mathbf{y}^T \mathbf{c}(\mathbf{x}) \} \\ &\le \inf_{\mathbf{x}} \{ f(\mathbf{x}) - \mathbf{y}^T \mathbf{c}(\mathbf{x}) \text{ s.t. } \mathbf{c}(\mathbf{x})(\le, =, \ge)\mathbf{0} \} \\ &\le \inf_{\mathbf{x}} \{ f(\mathbf{x}) : \text{ s.t. } \mathbf{c}(\mathbf{x})(\le, =, \ge)\mathbf{0} \}. \end{aligned}$$

The first inequality is from the fact that the unconstrained inf-value is no greater than the constrained one.

The second inequality is from $\mathbf{c}(\mathbf{x})(\le, =, \ge)\mathbf{0}$ and $\mathbf{y}(\le, '\text{free}', \ge)\mathbf{0}$ imply $-\mathbf{y}^T \mathbf{c}(\mathbf{x}) \le 0$.

Lagrangian Strong Duality Theorem

Theorem 11 *Let (GCO) be a convex minimization problem and the infimum* f^* *of (GCO) be finite, and the suprermum of (LDP) be* ϕ^**. In addition, let (GCO) have an interior-point feasible solution with respect to inequality constraints, that is, there is* $\hat{\mathbf{x}}$ *such that all inequality constraints are strictly held. Then,* $f^* = \phi^*$*, and (LDP) admits a maximizer* $\mathbf{y}^*$ *such that*

$$\phi(\mathbf{y}^*) = f^*.$$

Furthermore, if (GCO) admits a minimizer $\mathbf{x}^*$*, then*

$$y_i^* c_i(\mathbf{x}^*) = 0, \ \forall i = 1, ..., m.$$

The assumption of "interior-point feasible solution" is called Constraint Qualification condition, which was also needed as a condition to prove the strong duality theorem for general Conic Linear Optimization.

Note that the problem would be a convex minimization problem if all equality constraints are hyperplane or affine functions $c_i(\mathbf{x}) = \mathbf{a}_i \mathbf{x} - b_i$, all other level sets are convex.

The Lagrangian Dual with Primal Constraints

Consider the constrained problem with convex set constraints

$$
\begin{array}{lll}
(GCO) & \text{inf} & f(\mathbf{x}) \\
& \text{s.t.} & \mathbf{c}_i(\mathbf{x}) \ (\leq, =, \geq) \ 0, \ i = 1, ..., m, \\
& & \mathbf{x} \in \Omega \subset R^n.
\end{array}
$$

Typically, Ω has a simple form such as the cone

$$\Omega = R^n_+ = \{\mathbf{x} : \ \mathbf{x} \geq \mathbf{0}\}.$$

Using the (partial) Lagrangian Function:

$$L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T \mathbf{c}(\mathbf{x}), \ \mathbf{y} \in Y;$$

we can define the dual objective function of $\mathbf{y}$ be

$$
\begin{array}{lll}
\phi(\mathbf{y}) := & \text{inf}_{\mathbf{x}} & L(\mathbf{x}, \mathbf{y}) \\
& \text{s.t.} & \mathbf{x} \in \Omega.
\end{array}
$$

The similar weak and strong duality theorem also holds.

Rules to Construct the Explicit Lagrangian Dual

$$(\text{GCO}) \qquad \begin{array}{ll} \text{min} & f(\mathbf{x}) \\ \text{s.t.} & c_i\mathbf{x}) \quad (\leq, =, \geq) \quad 0, \; i = 1, ..., m, \end{array}$$

- All multipliers are dual variables.
- Derive the LDC

$$\nabla f(\mathbf{x}) = \mathbf{y}^T \nabla \mathbf{c}(\mathbf{x})$$

If no $\mathbf{x}$ appeared in an equation, set it as an equality constraint for the dual; otherwise, express $\mathbf{x}$ in terms of $\mathbf{y}$ and replace $\mathbf{x}$ in the Lagrange function, which becomes the Dual objective. (This may be very difficult ...)

- Add the MSC as dual constraints.

The Lagrangian Dual of LP with the Log-Barrier I

For a fixed $\mu > 0$, consider the problem

$$\begin{array}{ll} \text{min} & \mathbf{c}^T\mathbf{x} - \mu\sum_{j=1}^n \log(x_j) \\ \text{s.t.} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

Again, the non-negativity constraints can be “ignored” if the feasible region has an ”interior”, that is, any minimizer must have $\mathbf{x}(\mu) > \mathbf{0}$. Thus, the Lagrangian function would be simply given by

$$L(\mathbf{x}, \mathbf{y}) = \mathbf{c}^T\mathbf{x} - \mu\sum_{j=1}^n \log(x_j) - \mathbf{y}^T(A\mathbf{x} - \mathbf{b}) = (\mathbf{c} - A^T\mathbf{y})^T\mathbf{x} - \mu\sum_{j=1}^n \log(x_j) + \mathbf{b}^T\mathbf{y}.$$

Then, the Lagrangian dual objective (we implicitly need $\mathbf{x} > \mathbf{0}$ for the function to be defined)

$$\phi(\mathbf{y}) := \inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) = \inf_{\mathbf{x}} \left[(\mathbf{c} - A^T\mathbf{y})^T\mathbf{x} - \mu\sum_{j=1}^n \log(x_j) + \mathbf{b}^T\mathbf{y} \right].$$

The Lagrangian Dual of LP with the Log-Barrier II

First, from the view point of the dual, the dual needs to choose $\mathbf{y}$ such that $\mathbf{c} - A^T\mathbf{y} > \mathbf{0}$, since otherwise the primal can choose $\mathbf{x} > \mathbf{0}$ to make $\phi(\mathbf{y})$ go to $-\infty$.

Now for any given $\mathbf{y}$ such that $\mathbf{c} - A^T\mathbf{y} > \mathbf{0}$, the inf problem has a unique finite close-form minimizer $\mathbf{x}$

$$x_j = \frac{\mu}{(\mathbf{c} - A^T\mathbf{y})_j}, \ \forall j = 1, ..., n.$$

Thus,

$$\phi(\mathbf{y}) = \mathbf{b}^T\mathbf{y} + \mu \sum_{j=1}^{n} \log(\mathbf{c} - A^T\mathbf{y})_j + n\mu(1 - \log(\mu)).$$

Therefore, the dual problem, for any fixed μ, can be written as

$$\max_{\mathbf{y}} \ \phi(\mathbf{y}) = n\mu(1 - \log(\mu)) + \max_{\mathbf{y}} \ [\mathbf{b}^T\mathbf{y} + \mu \sum_{j=1}^{n} \log(\mathbf{c} - A^T\mathbf{y})_j].$$

This is actually the LP dual with the Log-Barrier on dual inequality constraints $\mathbf{c} - A^T\mathbf{y} \geq \mathbf{0}$.

The gradient and Hessian of the Dual Objective ϕ

Let $\mathbf{x}(\mathbf{y})$ be a minimizer. Then

$$\phi(\mathbf{y}) = f(\mathbf{x}(\mathbf{y})) - \mathbf{y}^T \mathbf{h}(\mathbf{x}(\mathbf{y}))$$

Thus,

$$\begin{aligned}
\nabla \phi(\mathbf{y}) &= \nabla f(\mathbf{x}(\mathbf{y}))^T \nabla \mathbf{x}(\mathbf{y}) - \mathbf{y}^T \nabla \mathbf{h}(\mathbf{x}(\mathbf{y})) \nabla \mathbf{x}(\mathbf{y}) - \mathbf{h}(\mathbf{x}(\mathbf{y})) \\
&= (\nabla f(\mathbf{x}(\mathbf{y}))^T - \mathbf{y}^T \nabla \mathbf{h}(\mathbf{x}(\mathbf{y}))) \nabla \mathbf{x}(\mathbf{y}) - \mathbf{h}(\mathbf{x}(\mathbf{y})) \\
&= -\mathbf{h}(\mathbf{x}(\mathbf{y})).
\end{aligned}$$

Similarly, we can derive

$$\nabla^2 \phi(\mathbf{y}) = -\nabla \mathbf{h}(\mathbf{x}(\mathbf{y})) \left(\nabla^2_{\mathbf{x}} L(\mathbf{x}(\mathbf{y}), \mathbf{y})\right)^{-1} \nabla \mathbf{h}(\mathbf{x}(\mathbf{y}))^T,$$

where $\nabla^2_{\mathbf{x}} L(\mathbf{x}(\mathbf{y}), \mathbf{y})$ is the Hessian of the Lagrangian function that is assumed to be positive definite at any (local) minimizer.

The Toy Example

$$\begin{array}{ll} \text{minimize} & (x_1-1)^2+(x_2-1)^2 \\ \text{subject to} & x_1+2x_2-1=0, \quad 2x_1+x_2-1=0. \end{array}$$

$$L(\mathbf{x},\mathbf{y})=(x_1-1)^2+(x_2-1)^2-y_1(x_1+2x_2-1)-y_2(2x_1+x_2-1).$$

$$x_1=0.5y_1+y_2+1, \quad x_2=y_1+0.5y_2+1.$$

$$\phi(\mathbf{y})=-1.25y_1^2-1.25y_2^2-2y_1y_2-2y_1-2y_2.$$

$$\nabla\phi(\mathbf{y})=\begin{pmatrix} 2.5y_1+2y_2+2 \\ 2y_1+2.5y_21+2 \end{pmatrix},$$

$$\nabla^2\phi(\mathbf{y})=-\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1}\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}^T=-\begin{pmatrix} 2.5 & 2 \\ 2 & 2.5 \end{pmatrix}$$

The Fisher Example again

$$\begin{array}{ll} \text{minimize} & -5\log(2x_1 + x_2) - 8\log(3x_3 + x_4) \\ \text{subject to} & x_1 + x_3 = 1, \quad x_2 + x_4 = 1, \quad \mathbf{x} \geq \mathbf{0}. \end{array}$$

$$L(\mathbf{x}(\geq \mathbf{0}), \mathbf{y}) = -5\log(2x_1 + x_2) - 8\log(3x_3 + x_4) - y_1(x_1 + x_3 - 1) - y_2(x_2 + x_4 - 1).$$

Start from $\mathbf{y}^0 > \mathbf{0}$, at the kth step, compute $\mathbf{x}^{k+1}$ from

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x} \geq \mathbf{0}} L(\mathbf{x}(\geq \mathbf{0}), \mathbf{y}^k),$$

then let

$$\mathbf{y}^{k+1} = \mathbf{y}^k - \frac{1}{\beta}(A\mathbf{x}^{k+1} - \mathbf{b}).$$

Note that $\mathbf{x}$ in each iteration has a close-form solution! (FisherexampleLMM.m of Chapter 14)

Infeasibility Certificate (Farkas Lemma) for Nonlinear Constraints I

Consider the convex constrained system:

$$(\text{CCS}) \qquad \begin{array}{lll} \text{min} & \mathbf{0}^T\mathbf{x} & \\ \text{s.t.} & c_i(\mathbf{x}) \geq & 0, \ i = 1, ..., m, \end{array}$$

where $c_i(.)$ are concave functions and the Lagrangian Function is given by

$$L(\mathbf{x}, \mathbf{y}) = -\mathbf{y}^T\mathbf{c}(\mathbf{x}) = -\sum_{i=1}^{m} y_i c_i(\mathbf{x}), \ \mathbf{y} \geq \mathbf{0}.$$

Again, let

$$\phi(\mathbf{y}) := \inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}).$$

Theorem 12 *If there exists* $\mathbf{y} \geq \mathbf{0}$ *such that* $\phi(\mathbf{y}) > 0$*, then (CSS) is infeasible.*

The proof is directly from the dual objective function $\phi(\mathbf{y})$ is a homogeneous function and the dual has its objective value unbounded from above.

Infeasibility Certificate (Farkas Lemma) for Nonlinear Constraints II

Consider the system, for a parameter $b \geq 0$,

$$-x_1^2 - (x_2 - 1)^2 + b \geq 0, \quad (y_1 \geq 0)$$
$$-x_1^2 - (x_2 + 1)^2 + b \geq 0, \quad (y_2 \geq 0)$$

$$L(\mathbf{x}, \mathbf{y}) = y_1(x_1^2 + (x_2 - 1)^2 - b) + y_2(x_1^2 + (x_2 + 1)^2 - b).$$

Then, if $y_1 + y_2 \neq 0$,

$$\phi(\mathbf{y}) = \frac{4y_1y_2 - b(y_1 + y_2)^2}{y_1 + y_2}, \quad (y_1, y_2) \geq 0$$

When $b \geq 1$, $\phi(\mathbf{y}) \leq 0$; and, otherwise, one can choose $y_1 = y_2 = y > 0$ such that

$$\phi(\mathbf{y}) = 2(1 - b)y > 0$$

which implies that the original constrained system is infeasible.

The Augmented Lagrangian Function

For equality constraints $\{\mathbf{x} : \ \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$, in both theory and practice, we can consider an augmented Lagrangian function (ALF)

$$L_a(\mathbf{x}, \mathbf{y}, \mathbf{s}) = f(\mathbf{x}) - \mathbf{y}^T \mathbf{h}(\mathbf{x}) - \mathbf{s}^T \mathbf{c}(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{h}(\mathbf{x})\|^2$$

for some positive parameter ρ, which corresponds to an equivalent problem of (**??**):

$$f^* := \text{min} \quad f(\mathbf{x}) + \tfrac{\beta}{2}\|\mathbf{h}(\mathbf{x})\|^2 \quad \text{s.t.} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}.$$

Note that, although at feasibility the additional square term in objective is redundant, it helps to improve strict convexity of the Lagrangian function.

For the Fisher example:

$$\begin{aligned} & L_a(\mathbf{x}(\geq \mathbf{0}), \mathbf{y}) \\ = \quad & -5\log(2x_1 + x_2) - 8\log(3x_3 + x_4) - y_1(x_1 + x_3 - 1) - y_2(x_2 + x_4 - 1) \\ & + \tfrac{\beta}{2}((x_1 + x_3 - 1)^2 + (x_2 + x_4 - 1)^2). \end{aligned}$$

Fisherexample using ALMM? Not close-form solution anymore - more on this issue latter.

The Augmented Lagrangian Dual

Now the dual function:

$$\phi_a(\mathbf{y}) = \min_{\mathbf{x} \in X} L_a(\mathbf{x}, \mathbf{y}); \tag{4}$$

and the dual problem

$$(f^* \geq) \phi_a^* := \max \quad \phi_a(\mathbf{y}). \tag{5}$$

Note that the dual function approximately satisfies $\frac{1}{\beta}$-Lipschitz condition (see Chapter 14 of L&Y).

For the convex optimization case, say $\mathbf{h}(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$, we have

$$\nabla^2 L_a(\mathbf{x}, \mathbf{y}) = \nabla^2 f(\mathbf{x}) + \beta(A^T A).$$