EE/ACM 150 - Applications of Convex Optimization in Signal Processing and Communications
Lecture 12

Andre Tkacenko

Signal Processing Research Group
Jet Propulsion Laboratory

May 10, 2012

Outline

1. Optimality Conditions
   - Certificate of Suboptimality and Stopping Criteria
   - Karush-Kuhn-Tucker Conditions
2. Perturbation and Sensitivity Analysis
   - Global Sensitivity Results
   - Local Sensitivity Results
3. Duality and Problem Reformulations
4. Lagrangian Duality for Problems with Generalized Inequalities
   - The Lagrangian, Dual Function, and Dual Problem
   - Optimality Conditions
   - Perturbation and Sensitivity Analysis

Certificate of Suboptimality/Optimality

Recall that for any primal feasible point $\mathbf{x}$ and dual feasible point $(\boldsymbol{\lambda}, \boldsymbol{\nu})$, we have

$$f_0(\mathbf{x}) \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \Longleftrightarrow G(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \triangleq f_0(\mathbf{x}) - g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \geq 0 \,.$$

Here, the quantity $G(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is called the *duality gap* associated with $\mathbf{x}$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu})$. The *optimal duality gap* γ (i.e., the minimum value of $G(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ over all feasible $\mathbf{x}$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu})$), is given by

$$\gamma \triangleq \inf_{\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}} \left\{ G(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \right\} = \inf_{\mathbf{x}} f_0(\mathbf{x}) - \sup_{\boldsymbol{\lambda}, \boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = p^\star - d^\star \,.$$

For cases in which the primal problem is difficult to solve, we can always bound the suboptimality corresponding to a particular primal feasible point $\widetilde{\mathbf{x}}$ by the duality gap. Specifically, we have

$$f_0(\widetilde{\mathbf{x}}) - p^\star \leq G\left(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\nu}}\right) \,,$$

for any dual feasible $\left(\widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\nu}}\right)$ and so $\widetilde{\mathbf{x}}$ is ϵ-suboptimal with $\epsilon = G\left(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\nu}}\right)$. Thus, $\left(\widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\nu}}\right)$ is a *certificate of* $G\left(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\nu}}\right)$*-suboptimality* for $\widetilde{\mathbf{x}}$. As this also holds for any dual optimal point $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ for which $g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = d^\star$, we have the sharper bound

$$f_0(\widetilde{\mathbf{x}}) - p^\star \leq G(\widetilde{\mathbf{x}}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) \,,$$

and so $\widetilde{\mathbf{x}}$ is, in fact, $G(\widetilde{\mathbf{x}}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$-optimal. If the optimal duality gap $\gamma = 0$ (i.e., strong duality holds), then any $\widetilde{\mathbf{x}}$ for which $G(\widetilde{\mathbf{x}}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = 0$ means that $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ is a *certificate of optimality* for $\widetilde{\mathbf{x}} = \mathbf{x}^\star$.

Stopping Criteria for Optimization Algorithms

Note that the bound on the suboptimality of a given primal feasible point $\mathbf{x}$ and dual feasible point $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ given by $G(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ can be computed in practice. This can be used to provide nonheuristic stopping criteria for optimization algorithms that produce a sequence of primal feasible points $\mathbf{x}^{(k)}$ and dual feasible points $\left(\boldsymbol{\lambda}^{(k)}, \boldsymbol{\nu}^{(k)}\right)$, for $k \in \mathbb{N}$.

Absolute accuracy stopping criterion: Suppose we choose an absolute accuracy $\epsilon_{\mathrm{abs}} > \gamma$ (assuming that we can bound the optimal duality gap γ). Then, we terminate the algorithm when

$$f_0\left(\mathbf{x}^{(k)}\right) - g\left(\boldsymbol{\lambda}^{(k)}, \boldsymbol{\nu}^{(k)}\right) \leq \epsilon_{\mathrm{abs}} \, .$$

This *guarantees* that $\mathbf{x}^{(k)}$ is at least ϵ_{abs}-suboptimal, and $\left(\boldsymbol{\lambda}^{(k)}, \boldsymbol{\nu}^{(k)}\right)$ is a certificate proving it.

Relative accuracy stopping criterion: Suppose we choose a relative accuracy $\epsilon_{\mathrm{rel}} > \gamma / |p^\star|$ (assuming that we can bound γ and that $p^\star \neq 0$). Then, we terminate the algorithm if

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

holds or

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

holds. In this case, we are guaranteed to have $\frac{f_0\left(\mathbf{x}^{(k)}\right) - p^\star}{|p^\star|} \leq \epsilon_{\mathrm{rel}}$.

Karush-Kuhn-Tucker Conditions for Optimality

For the problem

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

with variable $\mathbf{x} \in \mathbb{R}^n$, assume that strong duality holds (i.e., the optimal duality gap is zero) and that $\mathbf{x}^\star$ and $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ are primal and dual optimal points, respectively. Then, we have the following *necessary* conditions for optimality, assuming $f_0, \ldots, f_m, h_1, \ldots, h_p$ are differentiable.

Karush-Kuhn-Tucker (KKT) Conditions:

1. *Primal feasibility:*

$$f_i(\mathbf{x}^\star) \leq 0\,,\ i = 1, \ldots, m\,,\ h_i(\mathbf{x}^\star) = 0\,,\ i = 1, \ldots, p\,.$$

2. *Dual feasibility:*

$$\lambda_i^\star \geq 0\,,\ i = 1, \ldots, m\,.$$

3. *Complementary slackness:*

$$\lambda_i^\star f_i\left(\mathbf{x}^\star\right) = 0\,,\ i = 1, \ldots, m\,.$$

4. *Stationarity:*

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

Karush-Kuhn-Tucker Conditions for Convex Problems

For the convex problem

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

with variable $\mathbf{x} \in \mathbb{R}^n$, where $f_0, \ldots, f_m$ are convex, let $\widetilde{\mathbf{x}}$ be any primal point and $\left(\widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\nu}}\right)$ be any dual point. Then, we have the following *necessary and sufficient* conditions for optimality of $\widetilde{\mathbf{x}}$ and $\left(\widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\nu}}\right)$, assuming $f_0, \ldots, f_m$ are differentiable.

Karush-Kuhn-Tucker (KKT) Conditions for Convex Optimization Problems:

1. *Primal feasibility:* $f_i(\widetilde{\mathbf{x}}) \leq 0\,,\ i = 1, \ldots, m\,,\ \mathbf{A}\widetilde{\mathbf{x}} = \mathbf{b}$.
2. *Dual feasibility:* $\widetilde{\lambda}_i \geq 0\,,\ i = 1, \ldots, m$.
3. *Complementary slackness:* $\widetilde{\lambda}_i f_i(\widetilde{\mathbf{x}}) = 0\,,\ i = 1, \ldots m$.
4. *Stationarity:* $\nabla f_0(\widetilde{\mathbf{x}}) + \sum_{i=1}^{m} \widetilde{\lambda}_i \nabla f_i(\widetilde{\mathbf{x}}) + \mathbf{A}^T \widetilde{\boldsymbol{\nu}} = \mathbf{0}$.

Sufficiency follows from the fact that $f_0(\widetilde{\mathbf{x}}) = L\left(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\nu}}\right)$, by complementary slackness, and that $g\left(\widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\nu}}\right) = L\left(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\nu}}\right)$, by convexity of $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ in $\mathbf{x}$ and stationarity. Hence, $f_0(\widetilde{\mathbf{x}}) = g\left(\widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\nu}}\right)$.

If Slater's condition is satisfied, then $\mathbf{x}$ is optimal if and only if there exist $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ that satisfy the KKT conditions. This is because it implies strong duality and that the dual optimum is attained.

Water-Filling Example

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

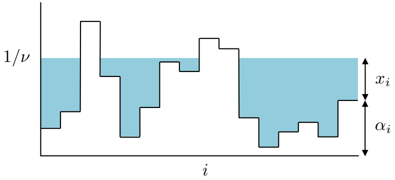
where $\alpha_i > 0$. By the KKT conditions, $\mathbf{x}$ is optimal if and only if $\mathbf{x} \succeq \mathbf{0}$, $\mathbf{1}^T \mathbf{x} = 1$, and there exist $\boldsymbol{\lambda} \in \mathbb{R}^n$ and $\nu \in \mathbb{R}$ such that

$$\lambda_i \geq 0\,,\ \lambda_i x_i = 0\,,\ -\frac{1}{x_i + \alpha_i} - \lambda_i + \nu = 0\,,\ i = 1, \ldots, n\,.$$

- If $\nu < 1/\alpha_i$, then $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$.
- If $\nu \geq 1/\alpha_i$, then $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$.
- We can determine ν from the constraint $\mathbf{1}^T \mathbf{x} = \sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\} = 1$.

Interpretation:

- There are n patches of land, where the level of the i-th patch is at height α_i.
- We flood the area with a unit amount of water.
- The resulting level is $1/\nu$.



Perturbed Optimization Problem and its Dual

For the unperturbed problem and its dual

Primal:

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

Dual:

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

we consider the following perturbed problem and its dual.

Perturbed Problem and its Dual:

Primal:

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

Dual:

$$\begin{array}{ll} \text{maximize} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) - \mathbf{u}^T \boldsymbol{\lambda} - \mathbf{v}^T \boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{\lambda} \succeq \mathbf{0} \end{array}.$$

- Here, $\mathbf{x}$ is the primal variable, while $\mathbf{u}$ and $\mathbf{v}$ are parameters.
- We denote $p^\star(\mathbf{u}, \mathbf{v})$ as the optimal value as a function of $\mathbf{u}$ and $\mathbf{v}$.
- We are interested in information about $p^\star(\mathbf{u}, \mathbf{v})$ that we can obtain from the solution of the unperturbed problem and its dual.

Global Sensitivity Inequality

Assume that strong duality holds for the unperturbed problem, and that $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ are dual optimal for the unperturbed problem.

Applying weak duality to the perturbed problem yields

$$\begin{aligned} p^\star(\mathbf{u}, \mathbf{v}) &\geq g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) - \mathbf{u}^T \boldsymbol{\lambda}^\star - \mathbf{v}^T \boldsymbol{\nu}^\star \,, \\ &= p(\mathbf{0}, \mathbf{0}) - \mathbf{u}^T \boldsymbol{\lambda}^\star - \mathbf{v}^T \boldsymbol{\nu}^\star \,. \end{aligned}$$

Sensitivity Interpretation:

- If $\lambda_i^\star$ is large, then $p^\star$ increases greatly if we tighten the i-th constraint (i.e., $u_i < 0$).
- If $\lambda_i^\star$ is small, then $p^\star$ does not decrease much if we loosen the i-th constraint (i.e., $u_i > 0$).
- If $\nu_i^\star$ is large and positive, then $p^\star$ increases greatly if we take $v_i < 0$.
  If $\nu_i^\star$ is large and negative, then $p^\star$ increases greatly if we take $v_i > 0$.
- If $\nu_i^\star$ is small and positive, then $p^\star$ does not decrease much if we take $v_i > 0$.
  If $\nu_i^\star$ is small and negative, then $p^\star$ does not decrease much if we take $v_i < 0$.

Local Sensitivity Analysis

Assume as before that strong duality holds for the unperturbed problem. If, in addition, $p^\star(\mathbf{u}, \mathbf{v})$ is differentiable at $(\mathbf{0}, \mathbf{0})$, then we have

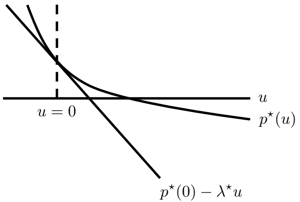
$$\lambda_i^\star = -\frac{\partial p^\star(\mathbf{0}, \mathbf{0})}{\partial u_i}\,,\ i = 1, \ldots, m\,,\ \nu_i^\star = -\frac{\partial p^\star(\mathbf{0}, \mathbf{0})}{\partial v_i}\,,\ i = 1, \ldots, p\,.$$

Proof: For $\lambda_i^\star$, from the global sensitivity result, we have

$$\frac{\partial p^\star(\mathbf{0}, \mathbf{0})}{\partial u_i} = \lim_{t \searrow 0} \frac{p^\star(t\mathbf{e}_i, \mathbf{0}) - p^\star(\mathbf{0}, \mathbf{0})}{t} \geq -\lambda_i^\star\,,$$
$$\frac{\partial p^\star(\mathbf{0}, \mathbf{0})}{\partial u_i} = \lim_{t \nearrow 0} \frac{p^\star(t\mathbf{e}_i, \mathbf{0}) - p^\star(\mathbf{0}, \mathbf{0})}{t} \leq -\lambda_i^\star\,.$$

As both of these inequalities hold, they must with equality. A similar proof follows for $\nu_i^\star$.

Visual Example: Sketch of $p^\star(u)$ for a problem with one inequality constraint.



Duals of Equivalent Problem Formulations

- 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 is *uninteresting*.

Common Reformulations:

- Introduce new variables and equality constraints.
- Make explicit constraints implicit or vice-versa.
- Carry out a change of variables, such as $\mathbf{x} = \phi(\mathbf{z})$.
- Transform the objective or constraint functions.
  For example, we can replace $f_0(\mathbf{x})$ with $\psi(f_0(\mathbf{x}))$, where ψ is convex and increasing.

Introducing New Variables and Equality Constraints

As an example, consider the problem

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

- The dual function is constant as we have $g = \inf_{\mathbf{x}} L(\mathbf{x}) = \inf_{\mathbf{x}} f_0(\mathbf{Ax}+\mathbf{b}) = p^\star$.
- We have strong duality always, but the dual is rather useless.

Reformulated problem and its dual:

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

The dual function follows from

$$\begin{aligned} g(\boldsymbol{\nu}) &= \inf_{\mathbf{x},\mathbf{y}} \left\{ f_0(\mathbf{y}) + \boldsymbol{\nu}^T(\mathbf{Ax}+\mathbf{b}-\mathbf{y}) \right\} , \\ &= \inf_{\mathbf{x},\mathbf{y}} \left\{ -\left(\boldsymbol{\nu}^T\mathbf{y} - f_0(\mathbf{y})\right) + \left(\mathbf{A}^T\boldsymbol{\nu}\right)^T\mathbf{x} + \mathbf{b}^T\boldsymbol{\nu} \right\} , \\ &= \mathbf{b}^T\boldsymbol{\nu} - \sup_{\mathbf{y}} \left\{ \boldsymbol{\nu}^T\mathbf{y} - f_0(\mathbf{y}) \right\} + \inf_{\mathbf{x}} \left\{ \left(\mathbf{A}^T\boldsymbol{\nu}\right)^T\mathbf{x} \right\} , \\ &= \begin{cases} \mathbf{b}^T\boldsymbol{\nu} - f_0^{\circledast}(\boldsymbol{\nu}) , & \mathbf{A}^T\boldsymbol{\nu}=\mathbf{0} \\ -\infty , & \text{otherwise} \end{cases} . \end{aligned}$$

Examples with New Variables and Equality Constraints

Norm approximation problem:

$$\text{minimize} \quad ||\mathbf{A}\mathbf{x} - \mathbf{b}|| \quad .$$

Using the trick of introducing new variables and new equality constraints, this becomes

$$\begin{array}{ll} \text{minimize} & ||\mathbf{y}|| \\ \text{subject to} & \mathbf{y} = \mathbf{A}\mathbf{x} - \mathbf{b} \end{array} .$$

We can exploit the fact that

$$f(\mathbf{x}) = ||\mathbf{x}|| \iff f^{\circledast}(\mathbf{y}) = \begin{cases} 0\,, & ||\mathbf{y}||_{\circledast} \leq 1 \\ \infty\,, & \text{otherwise} \end{cases} ,$$

or derive the dual function directly as follows.

$$\begin{aligned} g(\boldsymbol{\nu}) &= \inf_{\mathbf{x},\mathbf{y}} \left\{ ||\mathbf{y}|| + \boldsymbol{\nu}^T \left(\mathbf{y} - \mathbf{A}\mathbf{x} + \mathbf{b}\right) \right\} = \mathbf{b}^T\boldsymbol{\nu} + \inf_{\mathbf{y}} \left\{ ||\mathbf{y}|| + \boldsymbol{\nu}^T\mathbf{y} \right\} - \sup_{\mathbf{x}} \left\{ \left(\mathbf{A}^T\boldsymbol{\nu}\right)^T \mathbf{x} \right\} , \\ &= \begin{cases} \mathbf{b}^T\boldsymbol{\nu} + \inf_{\mathbf{y}} \left\{ ||\mathbf{y}|| + \boldsymbol{\nu}^T\mathbf{y} \right\} , & \mathbf{A}^T\boldsymbol{\nu} = \mathbf{0} \\ -\infty\,, & \text{otherwise} \end{cases} , \\ &= \begin{cases} \mathbf{b}^T\boldsymbol{\nu}\,, & \mathbf{A}^T\boldsymbol{\nu} = \mathbf{0}\,,\ ||\boldsymbol{\nu}||_{\circledast} \leq 1 \\ -\infty\,, & \text{otherwise} \end{cases} . \end{aligned}$$

Dual of norm approximation problem:

$$\begin{array}{ll} \text{maximize} & \mathbf{b}^T\boldsymbol{\nu} \\ \text{subject to} & \mathbf{A}^T\boldsymbol{\nu} = \mathbf{0}\,,\ ||\boldsymbol{\nu}||_{\circledast} \leq 1 \end{array} .$$

Implicit Constraints

LP with box constraints:

Primal:		*Dual:*	
minimize	$\mathbf{c}^T\mathbf{x}$	maximize	$-\mathbf{b}^T\boldsymbol{\nu} - \mathbf{u}^T\boldsymbol{\lambda}_{\text{u}} + \mathbf{l}^T\boldsymbol{\lambda}_{\text{l}}$
subject to	$\mathbf{A}\mathbf{x} = \mathbf{b}$	subject to	$\mathbf{A}^T\boldsymbol{\nu} + \boldsymbol{\lambda}_{\text{u}} - \boldsymbol{\lambda}_{\text{l}} + \mathbf{c} = \mathbf{0}$,
	$\mathbf{l} \preceq \mathbf{x} \preceq \mathbf{u}$		$\boldsymbol{\lambda}_{\text{u}} \succeq \mathbf{0}\,, \boldsymbol{\lambda}_{\text{l}} \succeq \mathbf{0}$

Reformulation with box constraints made implicit:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) = \begin{cases} \mathbf{c}^T\mathbf{x}\,, & \mathbf{l} \preceq \mathbf{x} \preceq \mathbf{u} \\ \infty\,, & \text{otherwise} \end{cases} \; . \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array}$$

With this reformulation, the dual function is

$$\begin{aligned} g(\boldsymbol{\nu}) &= \inf_{\mathbf{l} \preceq \mathbf{x} \preceq \mathbf{u}} \left\{ \mathbf{c}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) \right\} = -\mathbf{b}^T\boldsymbol{\nu} + \inf_{\mathbf{l} \preceq \mathbf{x} \preceq \mathbf{u}} \left\{ \left(\mathbf{A}^T\boldsymbol{\nu} + \mathbf{c}\right)^T \mathbf{x} \right\} , \\ &= -\mathbf{b}^T\boldsymbol{\nu} - \mathbf{u}^T\left(\mathbf{A}^T\boldsymbol{\nu} + \mathbf{c}\right)^- + \mathbf{l}^T\left(\mathbf{A}^T\boldsymbol{\nu} + \mathbf{c}\right)^+ , \end{aligned}$$

where $y_i^- \triangleq \max\{-y_i, 0\}$ and $y_i^+ \triangleq \max\{y_i, 0\}$.

Dual problem:

$$\text{maximize} \quad -\mathbf{b}^T\boldsymbol{\nu} - \mathbf{u}^T\left(\mathbf{A}^T\boldsymbol{\nu} + \mathbf{c}\right)^- + \mathbf{l}^T\left(\mathbf{A}^T\boldsymbol{\nu} + \mathbf{c}\right)^+ \; .$$

The Lagrangian and Dual Function

Consider the problem with generalized inequalities given by

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & \mathbf{f}_i(\mathbf{x}) \preceq_{\mathcal{K}_i} \mathbf{0}\,,\ i = 1, \ldots, m \quad , \\ & h_i(\mathbf{x}) = 0\,,\ i = 1, \ldots, p \end{array}$$

where $\mathcal{K}_i \subseteq \mathbb{R}^{k_i}$ are proper cones. We do not assume convexity here. The definitions of the Lagrangian and the dual function parallel those in the scalar case, except now, for each constraint $\mathbf{f}_i(\mathbf{x}) \preceq_{\mathcal{K}_i} \mathbf{0}$, we associate the Lagrange multiplier vector $\boldsymbol{\lambda}_i \in \mathbb{R}^{k_i}$.

The Lagrangian:

The Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \to \mathbb{R}$ is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\nu}) \triangleq f_0(\mathbf{x}) + \sum_{i=1}^{m} \boldsymbol{\lambda}_i^T \mathbf{f}_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \ .$$

Lagrange Dual Function:

The dual function $g : \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \to \mathbb{R}$ is defined as

$$g(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\nu}) \triangleq \inf_{\mathbf{x} \in \mathcal{D}} \left\{ L(\mathbf{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\nu}) \right\} \ .$$

Lower Bound Property and the Dual Problem

Lower Bound Property: If $\boldsymbol{\lambda}_i \succeq_{\mathcal{K}_i^{\circledast}} \mathbf{0}$ for $i = 1, \ldots, m$, then $g(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\nu}) \leq p^\star$.

Proof: If $\widetilde{\mathbf{x}}$ is feasible and $\boldsymbol{\lambda}_i \succeq_{\mathcal{K}_i^{\circledast}} \mathbf{0}$ for $i = 1, \ldots, m$, then we have

$$\begin{aligned} g(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\nu}) &= \inf_{\mathbf{x} \in \mathcal{D}} \left\{ L(\mathbf{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\nu}) \right\} , \\ &\leq f_0(\widetilde{\mathbf{x}}) + \sum_{i=1}^{m} \boldsymbol{\lambda}_i^T \mathbf{f}_i(\widetilde{\mathbf{x}}) + \sum_{i=1}^{p} \nu_i h_i(\widetilde{\mathbf{x}}) , \\ &\leq f_0(\widetilde{\mathbf{x}}) . \end{aligned}$$

Minimizing over all feasible $\widetilde{\mathbf{x}}$ gives $g(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\nu}) \leq p^\star$.

Lagrange Dual Problem:

$$\begin{array}{ll} \text{maximize} & g(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\nu}) \\ \text{subject to} & \boldsymbol{\lambda}_i \succeq_{\mathcal{K}_i^{\circledast}} \mathbf{0} \, , \; i = 1, \ldots, m \end{array} .$$

- *Weak duality:* $p^\star \geq d^\star$ always.
- *Strong duality:* $p^\star = d^\star$ for a convex problems with a constraint qualification. For example, if Slater's condition holds, meaning that the primal problem is strictly feasible, then we have strong duality.

Inequality Form SDP and its Dual

Primal SDP: $(\mathbf{F}_i, \mathbf{G} \in \mathbb{S}^k)$

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T \mathbf{x} \\ \text{subject to} & x_1 \mathbf{F}_1 + \cdots + x_n \mathbf{F}_n \preceq \mathbf{G} \end{array} .$$

- The Lagrange multiplier is a matrix $\mathbf{Z} \in \mathbb{S}^k$.
- The Lagrangian is

$$L(\mathbf{x}, \mathbf{Z}) = \mathbf{c}^T \mathbf{x} + \mathrm{tr}(\mathbf{Z}\,(x_1 \mathbf{F}_1 + \cdots + x_n \mathbf{F}_n - \mathbf{G}))\ .$$

- The dual function is

$$g(\mathbf{Z}) = \inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{Z}) = \begin{cases} -\mathrm{tr}(\mathbf{G}\mathbf{Z})\,, & \mathrm{tr}(\mathbf{F}_i \mathbf{Z}) + c_i = 0\,,\ i = 1, \ldots, n \\ -\infty\,, & \text{otherwise} \end{cases} .$$

Dual SDP:

$$\begin{array}{ll} \text{maximize} & -\mathrm{tr}(\mathbf{G}\mathbf{Z}) \\ \text{subject to} & \mathrm{tr}(\mathbf{F}_i \mathbf{Z}) + c_i = 0\,,\ i = 1, \ldots, n \\ & \mathbf{Z} \succeq \mathbf{0} \end{array} .$$

We have $p^\star = d^\star$ if the primal SDP is strictly feasible, meaning that there exists an $\mathbf{x}$ such that $x_1 \mathbf{F}_1 + \cdots + x_n \mathbf{F}_n \prec \mathbf{G}$.

Complementary Slackness

Assume that strong duality holds (i.e., $p^\star = d^\star$), $\mathbf{x}^\star$ is primal optimal, and $(\boldsymbol{\lambda}_1^\star, \ldots, \boldsymbol{\lambda}_m^\star, \boldsymbol{\nu}^\star)$ is dual optimal. Then, we have the following.

$$
\begin{aligned}
f_0(\mathbf{x}^\star) &= g(\boldsymbol{\lambda}_1^\star, \ldots, \boldsymbol{\lambda}_m^\star, \boldsymbol{\nu}^\star) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}_1^\star, \ldots, \boldsymbol{\lambda}_m^\star, \boldsymbol{\nu}^\star) \\
&= \inf_{\mathbf{x} \in \mathcal{D}} \left\{ f_0(\mathbf{x}) + \sum_{i=1}^{m} \left(\boldsymbol{\lambda}_i^\star\right)^T \mathbf{f}_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^\star h_i(\mathbf{x}) \right\} , \\
&\leq f_0(\mathbf{x}^\star) + \sum_{i=1}^{m} \left(\boldsymbol{\lambda}_i^\star\right)^T \mathbf{f}_i(\mathbf{x}^\star) + \sum_{i=1}^{p} \nu_i^\star h_i(\mathbf{x}^\star) , \\
&\leq f_0(\mathbf{x}^\star) .
\end{aligned}
$$

Thus, the two inequalities hold with equality. From this, we ascertain the following.

- The primal optimal point $\mathbf{x}^\star$ minimizes $L(\mathbf{x}, \boldsymbol{\lambda}_1^\star, \ldots, \boldsymbol{\lambda}_m^\star, \boldsymbol{\nu}^\star)$. Note that the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda}_1^\star, \ldots, \boldsymbol{\lambda}_m^\star, \boldsymbol{\nu}^\star)$ may have other minimizers besides $\mathbf{x}^\star$.
- We have the *complementary slackness* condition $\left(\boldsymbol{\lambda}_i^\star\right)^T \mathbf{f}_i(\mathbf{x}^\star) = 0$ for $i = 1, \ldots, m$. Specifically, this implies that

$$
\boldsymbol{\lambda}_i^\star \succ_{\mathcal{K}_i^\circledast} \mathbf{0} \Longrightarrow \mathbf{f}_i(\mathbf{x}^\star) = \mathbf{0} , \ \mathbf{f}_i(\mathbf{x}^\star) \prec_{\mathcal{K}_i} \mathbf{0} \Longrightarrow \boldsymbol{\lambda}_i^\star = \mathbf{0} .
$$

  In contrast to problems with scalar inequalities, it is possible to satisfy $\left(\boldsymbol{\lambda}_i^\star\right)^T \mathbf{f}_i(\mathbf{x}^\star) = 0$ with $\boldsymbol{\lambda}_i^\star \neq \mathbf{0}$ and $\mathbf{f}_i(\mathbf{x}^\star) \neq \mathbf{0}$.

KKT Conditions for Optimality

Under the assumption that the functions $f_0, \mathbf{f}_1, \ldots, \mathbf{f}_m, h_1, \ldots, h_p$ are differentiable, the KKT conditions become the following.

KKT Conditions for Problems with Generalized Inequalities:

- *Primal feasibility:*
$$\mathbf{f}_i(\mathbf{x}) \preceq_{\mathcal{K}_i} \mathbf{0}\,,\; i = 1, \ldots, m\,,\; h_i(\mathbf{x}) = 0\,,\; i = 1, \ldots, p\,.$$
- *Dual feasibility:*
$$\boldsymbol{\lambda}_i \succeq_{\mathcal{K}_i^{\circledast}} \mathbf{0}\,,\; i = 1, \ldots, m\,.$$
- *Complementary slackness:*
$$\left(\boldsymbol{\lambda}_i^{\star}\right)^T \mathbf{f}_i(\mathbf{x}^{\star}) = 0\,,\; i = 1, \ldots, m\,.$$
- *Stationarity:*
$$\nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} \left(\frac{d\mathbf{f}_i(\mathbf{x})}{d\mathbf{x}}\right)^T \boldsymbol{\lambda}_i + \sum_{i=1}^{p} \nu_i \nabla h_i(\mathbf{x}) = \mathbf{0}\,.$$

- If strong duality holds, then the KKT conditions are *necessary* for the optimality of any primal optimal $\mathbf{x}^{\star}$ and dual optimal $(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\nu})$.
- If the primal problem is convex, then the KKT conditions are *necessary and sufficient* for the optimality of $\mathbf{x}^{\star}$, $(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\nu})$.

Perturbation and Sensitivity Results

Perturbed Problem with Generalized Inequalities and its Dual:

Primal:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & \mathbf{f}_i(\mathbf{x}) \preceq_{\mathcal{K}_i} \mathbf{u}_i\,,\ i = 1, \ldots, m \\ & h_i(\mathbf{x}) = v_i\,,\ i = 1, \ldots, p \end{array}$$

Dual:

$$\begin{array}{ll} \text{maximize} & g(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\nu}) \\ & - \sum_{i=1}^m \mathbf{u}_i^T \boldsymbol{\lambda}_i - \mathbf{v}^T \boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{\lambda}_i \succeq_{\mathcal{K}_i^{\circledast}} \mathbf{0}\,,\ i = 1, \ldots, m \end{array} .$$

Global Sensitivity Result: Assuming zero duality gap for the unperturbed problem, we have

$$p^\star(\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}) \geq p^\star(\mathbf{0}, \ldots, \mathbf{0}, \mathbf{0}) - \sum_{i=1}^m \mathbf{u}_i^T \boldsymbol{\lambda}_i^\star - \mathbf{v}^T \boldsymbol{\nu}\,.$$

Local Sensitivity Result: Assuming further that $p^\star(\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v})$ is differentiable at $\mathbf{u}_1 = \mathbf{0}, \ldots, \mathbf{u}_m = \mathbf{0}, \mathbf{v} = \mathbf{0}$, we have

$$\boldsymbol{\lambda}_i^\star = -\nabla_{\mathbf{u}_i} p^\star(\mathbf{0}, \ldots, \mathbf{0}, \mathbf{0})\ ,\ i = 1, \ldots, m\,,\ \nu_i^\star = -\frac{\partial p^\star(\mathbf{0}, \ldots, \mathbf{0}, \mathbf{0})}{\partial v_i}\ ,\ i = 1, \ldots, p\,.$$