A nonconvex problem with strong duality

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

\(A \not\succeq 0\), hence nonconvex

**dual function:** \(g(\lambda) = \inf_x (x^T (A + \lambda I)x + 2b^T x - \lambda)\)

- unbounded below if \(A + \lambda I \not\succeq 0\) or if \(A + \lambda I \succeq 0\) and \(b \notin \mathcal{R}(A + \lambda I)\)
- minimized by \(x = -(A + \lambda I)^+ b\) otherwise: \(g(\lambda) = -b^T (A + \lambda I)^+ b - \lambda\)

**dual problem** and equivalent SDP:

\[
\begin{align*}
\text{maximize} & \quad -b^T (A + \lambda I)^+ b - \lambda \\
\text{subject to} & \quad A + \lambda I \succeq 0 \\
& \quad b \in \mathcal{R}(A + \lambda I)
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad -t - \lambda \\
\text{subject to} & \quad \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0
\end{align*}
\]

strong duality although primal problem is not convex (not easy to show)
Geometric interpretation

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

interpretation of dual function:

\[
g(\lambda) = \inf_{(u,t) \in \mathcal{G}} (t + \lambda u), \quad \text{where} \quad \mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}
\]

- \( \lambda u + t = g(\lambda) \) is (non-vertical) supporting hyperplane to \( \mathcal{G} \)
- hyperplane intersects \( t \)-axis at \( t = g(\lambda) \)
**epigraph variation:** same interpretation if \( G \) is replaced with

\[
A = \{(u, t) \mid f_1(x) \leq u, f_0(x) \leq t \text{ for some } x \in D\}
\]

**strong duality**

- holds if there is a non-vertical supporting hyperplane to \( A \) at \((0, p^*)\)
- for convex problem, \( A \) is convex, hence has supp. hyperplane at \((0, p^*)\)
- Slater’s condition: if there exist \((\tilde{u}, \tilde{t}) \in A\) with \(\tilde{u} < 0\), then supporting hyperplanes at \((0, p^*)\) must be non-vertical
Complementary slackness

assume strong duality holds, $x^*$ is primal optimal, $(\lambda^*, \nu^*)$ is dual optimal

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

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

$$\leq f_0(x^*)$$

hence, the two inequalities hold with equality

• $x^*$ minimizes $L(x, \lambda^*, \nu^*)$

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

$$\lambda_i^* > 0 \implies f_i(x^*) = 0, \quad f_i(x^*) < 0 \implies \lambda_i^* = 0$$
Karush-Kuhn-Tucker (KKT) conditions

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

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

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

from page 5–17: if strong duality holds and $x, \lambda, \nu$ are optimal, then they must satisfy the KKT conditions
KKT conditions for convex problem

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

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

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

if **Slater’s condition** is satisfied:

\( x \) is optimal if and only if there exist \( \lambda, \nu \) that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition \( \nabla f_0(x) = 0 \) for unconstrained problem
example: water-filling (assume $\alpha_i > 0$)

\[
\begin{align*}
\text{minimize} & \quad -\sum_{i=1}^{n} \log(x_i + \alpha_i) \\
\text{subject to} & \quad x \succeq 0, \quad 1^T x = 1 
\end{align*}
\]

$x$ is optimal iff $x \succeq 0$, $1^T x = 1$, and there exist $\lambda \in \mathbb{R}^n$, $\nu \in \mathbb{R}$ such that

\[
\lambda \succeq 0, \quad \lambda_i x_i = 0, \quad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu
\]

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$
- determine $\nu$ from $1^T x = \sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\} = 1$

interpretation

- $n$ patches; level of patch $i$ is at height $\alpha_i$
- flood area with unit amount of water
- resulting level is $1/\nu^*$