EE364: Review Session 6

Outline:

• Variable bounds and dual feasibility

• SDP relaxation

• Monotone transformation of the objective

• Homework hints
Variable bounds and dual feasibility

In many problems the constraints include *variable bounds*, as in

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

the Lagrangian is

\[
L(x, \lambda, \mu, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \mu^T (x - u) + \nu^T (l - x)
\]
for any \( x \in \mathbb{R}^n \) and any \( \lambda \), we can choose \( \mu \geq 0 \) and \( \nu \geq 0 \) so that \( x \) minimizes \( L(x, \lambda, \mu, \nu) \)

we have

\[
\nabla_x L(x, \lambda, \mu, \nu) = \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + (\mu - \nu)
\]

if \( x \) minimizes \( L \), we have \( \nabla_x L = 0 \) and therefore

\[
\nu - \mu = \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x)
\]
\[ \nu = \left[ \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) \right]^+ \]

\[ = \frac{1}{2} \left( \left| \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) \right| + \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) \right) \]

and

\[ \mu = \left[ \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) \right]^- \]

\[ = \frac{1}{2} \left( \left| \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) \right| - \nabla f_0(x) - \sum_{i=1}^{m} \lambda_i \nabla f_i(x) \right) \]

where \(| \cdot |\) is componentwise
• therefore, if $\lambda \geq 0$ then $(\lambda, \mu, \nu)$ is dual feasible

• we can obtain a lower bound for any $\lambda \geq 0$
Example

with $x = (l + u)/2$ and $\lambda = 0$ we can find a dual feasible point and a lower bound on $f^*$

we have

\[ \nu = \frac{1}{2} (\nabla f_0((l + u)/2) + |\nabla f_0((l + u)/2)|) \]

\[ \mu = \frac{1}{2} (-\nabla f_0((l + u)/2) + |\nabla f_0((l + u)/2)|) \]
and therefore the lower bound becomes

\[
L(x, 0, \mu, \nu) = f_0\left(\frac{l + u}{2}\right) + \frac{1}{2} \left(\nabla f_0\left(\frac{l + u}{2}\right) + \left|\nabla f_0\left(\frac{l + u}{2}\right)\right|\right)^T \left(\frac{l + u}{2} - u\right) + \frac{1}{2} \left(\nabla f_0\left(\frac{l + u}{2}\right) + \left|\nabla f_0\left(\frac{l + u}{2}\right)\right|\right)^T \left(l - \frac{l + u}{2}\right) = f_0\left(\frac{l + u}{2}\right) - \left(\frac{u - l}{2}\right)^T \left|\nabla f_0\left(\frac{l + u}{2}\right)\right|\]
this bound can also be derived directly

since \( f_0 \) is convex

\[
f^* \geq f_0\left(\frac{u + l}{2}\right) + \nabla f_0\left(\frac{u + l}{2}\right)^T (x^* - \frac{u + l}{2})
\]

\[
\geq f_0\left(\frac{u + l}{2}\right) + \inf_{l \leq x \leq u} \nabla f_0\left(\frac{u + l}{2}\right)^T (x - \frac{u + l}{2})
\]

but \( \inf_{l \leq x \leq u} \nabla f_0((u + l)/2)^T(x - (l + u)/2) \) is obtained for

- \( x_i = (u_i - l_i)/2 \) if \( \nabla f_0((l + u)/2)_i \leq 0 \),
- \( x_i = (l_i + u_i)/2 \) if \( \nabla f_0((l + u)/2)_i > 0 \)

therefore,
\[
\inf_{l \leq x \leq u} \nabla f_0((u + l)/2)^T(x - (u + l)/2) = -\left|\nabla f_0((u + l)/2)\right|^T (u - l)/2
\]

we get
\[
f^* \geq f_0((u + l)/2) - \left|\nabla f_0((u + l)/2)\right|^T (u - l)/2
\]
SDP relaxations of two-way partitioning problem

consider the problem

\[ \text{minimize} \quad x^T W x \]
\[ \text{subject to} \quad x_i^2 = 1, \quad i = 1, \ldots, n, \]

- \( x_i = 1 \) if it belongs to the one partition
- \( x_i = -1 \) if it belongs to the other partition
- \( W_{ij} \) is the cost of having \( i \) and \( j \) in the same partition
the Lagrangian is

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

\[ = x^T(W + \text{diag}(\nu))x - 1^T \nu \]

and therefore the dual problem is

maximize \[ -1^T \nu \]

subject to \[ W + \text{diag}(\nu) \succeq 0 \]

the optimal value of the dual is a lower bound on the optimal value of the partitioning problem
Another bound

since

\[ x^T W x = \text{tr}(x^T W x) = \text{tr}(W xx^T) \]

and

\[ (x x^T)_{ii} = x_i^2 \]

we can write the original problem as

\begin{align*}
\text{minimize} & \quad \text{tr}(W X) \\
\text{subject to} & \quad X \succeq 0, \quad \text{rank} X = 1 \\
& \quad X_{ii} = 1, \quad i = 1, \ldots, n,
\end{align*}
the problem

\[
\begin{align*}
\text{minimize} & \quad \text{tr}(WX) \\
\text{subject to} & \quad X \succeq 0, \quad \text{rank} \ X = 1 \\
& \quad X_{ii} = 1, \quad i = 1, \ldots, n,
\end{align*}
\]

is not convex but we can write a relaxation by removing the rank constraint

\[
\begin{align*}
\text{minimize} & \quad \text{tr}(WX) \\
\text{subject to} & \quad X \succeq 0, \\
& \quad X_{ii} = 1, \quad i = 1, \ldots, n,
\end{align*}
\]

- this problem is convex (SDP) and gives a lower bound on the original problem
- if the solution has rank 1 we solved the original problem
we now find the dual of the dual problem

\[
\begin{align*}
\text{minimize} & \quad 1^T \nu \\
\text{subject to} & \quad W + \text{diag}(\nu) \succeq 0
\end{align*}
\]

introducing a Lagrange multiplier \( X \in \mathbb{S}^n \) for the matrix inequality

\[
L(\nu, X) = 1^T \nu - \text{tr}(X(W + \text{diag}(\nu)))
\]

\[
= 1^T \nu - \text{tr}(XW) - \sum_{i=1}^{n} \nu_i X_{ii}
\]

\[
= -\text{tr}(XW) + \sum_{i=1}^{n} \nu_i (1 - X_{ii})
\]

this is bounded below as a function of \( \nu \) only if \( X_{ii} = 1 \) for all \( i \), so we
obtain the dual problem

\[
\begin{align*}
\text{maximize} & \quad - \text{tr}(WX) \\
\text{subject to} & \quad X \succeq 0 \\
& \quad X_{ii} = 1, \quad i = 1, \ldots, n
\end{align*}
\]

this is the same as the relaxation problem
Monotone transformation of the objective

consider the convex optimization problem

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

suppose \( \phi : \mathbb{R} \to \mathbb{R} \) is increasing and convex
then the problem

\[
\begin{align*}
\text{minimize} & \quad \tilde{f}_0(x) = \phi(f_0(x)) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m
\end{align*}
\]

is convex and equivalent to it

- we are interested in the connection between the duals
- we consider \( \phi(a) = \exp a \)
suppose $\lambda$ is feasible for the dual of the first problem and $\bar{x}$ minimizes

$$f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x)$$

it can be shown that $\bar{x}$ also minimizes

$$\exp f_0(x) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(x)$$

for appropriate choice of $\tilde{\lambda}$

thus, $\tilde{\lambda}$ is dual feasible for the second problem
since \( \bar{x} \) minimizes \( f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \) we have

\[
\nabla f_0(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\bar{x}) = 0
\]

but

\[
\frac{\partial}{\partial x} \left[ \exp f_0(x) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\bar{x}) \right]_{x=\bar{x}} = \exp f_0(\bar{x}) \nabla f_0(\bar{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\bar{x})
\]

\[
= \exp f_0(x) \left[ \nabla f_0(\bar{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i e^{-f_0(\bar{x})} \nabla f_i(\bar{x}) \right]
\]

if we take \( \tilde{\lambda}_i = \exp f_0(\bar{x}) \lambda_i \geq 0 \)

\[
\frac{\partial}{\partial x} \left[ \exp f_0(x) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\bar{x}) \right]_{x=\bar{x}} = 0
\]
if $p^*$ denote the optimal value of the first problem 
the optimal value of the second is $\exp p^*$ 
we have bound 

$$ p^* \geq g(\lambda), $$

where $g$ is the dual function of the first problem and 

$$ \exp p^* \geq \tilde{g}(\tilde{\lambda}) $$

where $\tilde{g}$ is the dual function of the second problem or equivalently 

$$ p^* \geq \log \tilde{g}(\tilde{\lambda}) $$
we have $\tilde{g}(\tilde{\lambda}) = e^{f_0(\bar{x})} + \sum_{i=1}^{m} e^{f_0(\bar{x})} \lambda_i f_i(\bar{x})$ and therefore

$$\log \tilde{g}(\tilde{\lambda}) = \log \left( e^{f_0(\bar{x})} + \sum_{i=1}^{m} e^{f_0(\bar{x})} \lambda_i f_i(\bar{x}) \right)$$

$$= f_0(\bar{x}) + \log \left( 1 + \sum_{i=1}^{m} \lambda_i f_i(\bar{x}) \right)$$

the bound we from the modified problem is always worse, i.e., $\log \tilde{g}(\tilde{\lambda}) \leq g(\lambda)$ in fact

$$\log \tilde{g}(\tilde{\lambda}) - g(\lambda) = \log \left( 1 + \sum_{i=1}^{m} \lambda_i f_i(\bar{x}) \right) - \sum_{i=1}^{m} \lambda_i f_i(\bar{x})$$

from the identity $\log(1 + y) - y \leq 0$ we conclude
Additional problem hint

\[
\text{square}( \text{square}( x + y ) ) \leq x - y
\]

- the problem is that \text{square}() can only accept affine arguments, because it is convex, but not increasing

- we can restrict \text{square}() to \( \mathbb{R}_+ \) so that it’s convex and increasing

- we use \text{square}_\text{pos}() instead:

\[
\text{square}_\text{pos}( \text{square}( x + y ) ) \leq x - y
\]

- we can introduce additional variable

\begin{verbatim}
variable t
square( x+y ) <= t;
square( t ) <= x - y
\end{verbatim}