6.2 Lagrange relaxation

Lecture 11, 24.1.13

decomposition setting:

$$S := \left\{ x \in \mathbb{Z}^n : \begin{array}{c} m_1\{ \\ m_2\{ \end{array} \overbrace{\begin{pmatrix} A^1 \\ A^2 \end{pmatrix}}^{n} x \le \begin{pmatrix} b^1 \\ b^2 \end{pmatrix},\ x \ge 0 \right\}$$

with $A^i \in \mathbb{R}^{m_i \times n}$, $b^i \in \mathbb{R}^{m_i}$

$$z_{IP} := \max \{ c^T x : x \in S \}$$

$$Q := \{ x \in \mathbb{Z}^n : A^2 x \le b^2,\ x \ge 0 \}$$

$\boxed{IP(Q):}$ $\quad z_{IP} = \max \{ c^T x : \underbrace{A^1 x \le b^1}_{\text{"difficult" constraints}},\ \underbrace{x \in Q}_{\text{"easy" constraints}} \}$

Lagrange relaxation: Fix $\lambda \in \mathbb{R}^{m_1}_{\ge 0}$ and consider

$\boxed{LR(\lambda):}$ $\quad z_{LR}(\lambda) := \max \{ z(\lambda, x) : x \in Q \}$

where $z(\lambda, x) := c^T x + \lambda^T (b^1 - A^1 x)$

Observation 6.8 If $\bar{x} \in \mathbb{R}^n$ violates $A^1_j x \le b^1_j$,

then $A^1_j x > b^1_j$, hence $\underbrace{b^1_j - A^1_j x < 0}$.

Thus if λ_j is sufficiently large, $A^1_j x \le b^1_j$ is satisfied by an optimal solution of $LR(\lambda)$.

Lemma 6.9 For every $\lambda \in \mathbb{R}^{m_1}_{\geq 0}$,

$LR(\lambda)$ is a relaxation of $IP(Q)$: $z_{IP} \leq z_{LR}(\lambda)$.

Proof: Suppose $\bar{x}$ is optimal for $IP(Q)$.
Then $\bar{x} \in Q$ and $A^1 \bar{x} \leq b^1$. Thus $\forall \lambda \in \mathbb{R}^{m_1}_{\geq 0}$

$$z_{LR}(\lambda) \geq z(\lambda, \bar{x}) = c^T\bar{x} + \underbrace{\lambda^T}_{\geq 0}\underbrace{(b^1 - A^1\bar{x})}_{\geq 0} \geq c^T\bar{x} = z_{IP}.$$

$\square$

The idea of Lagrange relaxation:

Every $\lambda \in \mathbb{R}^{m_1}_{\geq 0}$ defines an upper bound $z_{LR}(\lambda)$ for z_{IP}.
We want to find the best (i.e. smallest) upper bound:

$\boxed{LD}$: $$z_{LD} := \min \{ z_{LR}(\lambda) : \lambda \in \mathbb{R}^{m_1}_{\geq 0} \}$$

is called the Lagrange dual of $IP(Q)$ w.r.t. $A^1 x \leq b^1$.

$\rightarrow$ How do we solve it?
Do optimal solutions of LD satisfy $A^1 x \leq b^1$?

Example 6.10

$IP(Q)$: $\quad z_{IP} = \max\{7x_1 + 2x_2 : \; -x_1 + 2x_2 \le 4, \; x \in Q\}$

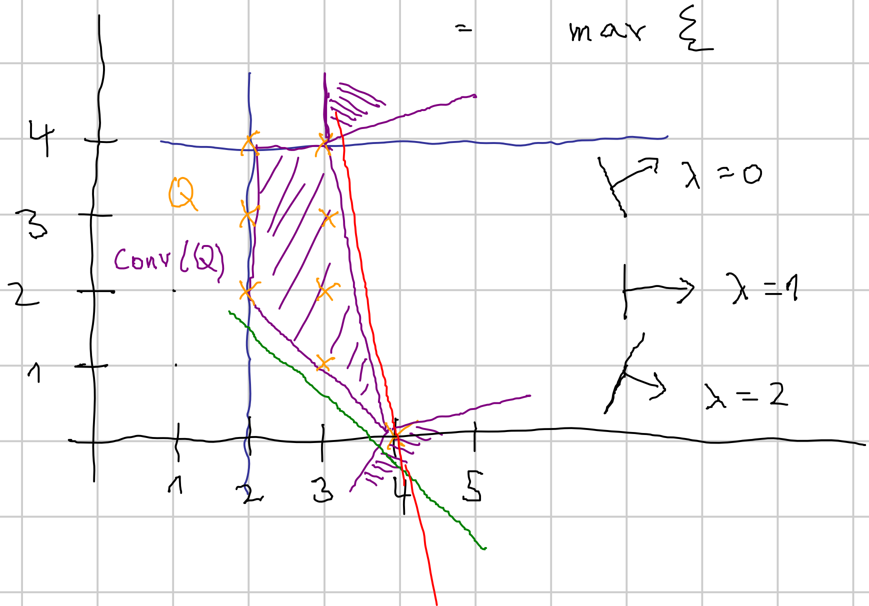
where $Q := \left\{ x \in \mathbb{Z}^2_{\ge 0} : \begin{array}{rcl} 5x_1 + x_2 & \le & 20 \\ -2x_1 - 2x_2 & \le & -7 \\ -x_1 & \le & -2 \\ x_2 & \le & 4 \end{array} \right\}$

$LR(\lambda)$: $\quad z_{LR}(\lambda) = \max\{ z(\lambda, x) : x \in Q\}$

$= \max\{7x_1 + 2x_2 + \lambda(4 + x_1 - 2x_2) : x \in Q\}$

$= \max\{(7+\lambda)x_1 + (2-2\lambda)x_2 + 4\lambda : x \in Q\}$

$= \max\{ \quad \text{''} \quad : x \in \text{conv}(Q)\}$



Prop 6.11

$z_{LR}(\lambda)$ is a piecewise-linear, convex function.

„Proof by example": • For a given λ, there exists a x_λ which is an optimal vertex of conv(Q) for

$\max\{(7+\lambda)x_1 + (2-2\lambda)x_2 + 4\lambda : x \in \text{Conv}(Q)\}$

$= \max\{7x_1 + 2x_2 + \lambda(4 + x_1 - 2x_2) : x \in \text{conv}(Q)\}$

- x_λ remains optimal vertex as long as the objective function $\binom{7+\lambda}{2-2\lambda}$ does not leave the normal cone of Conv(Q) in x_λ.
- as long as this is the case, the objective function is an affine function in λ.

$\Rightarrow$ To compute $z_{LD} = \min_{\lambda \ge 0} z_{LR}(\lambda)$, we only need to minimize a piecewise-linear, convex function.

$\boxed{11-3}$

Theorem 6.12 For the Lagrange dual of

$$z_{IP} = \max\{c^T x : A^1 x \le b^1,\ x \in Q\}$$

we have

$$z_{LD} \overset{\text{def}}{=} \min\{z_{LR}(\lambda) : \lambda \in \mathbb{R}^{m_1}_{\ge 0}\}$$
$$= \max\{c^T x : A^1 x \le b^1,\ x \in \text{conv}(Q)\}.$$

In other words: z_{LD} yields the same value as a vector from $\text{conv}(Q)$ that also satisfies $A^1 x \le b^1$.

Proof: Let $Q = \{x^1, \dots, x^T\}$.

$$\begin{aligned}
z_{LD} &= \min\{z_{LR}(\lambda) : \lambda \ge 0\} \\
&= \min_{\lambda \ge 0} \max_{x^i \in Q} \{z(\lambda, x^i)\} \\
&= \min_{\lambda \ge 0} \max_{x^i \in Q} \{c^T x^i + \lambda^T (b^1 - A^1 x^i)\} \\
&= \min_{\substack{\lambda \ge 0 \\ \gamma \in \mathbb{R}}} \{\gamma : \gamma \ge c^T x^i + \lambda^T (b^1 - A^1 x^i) \quad \forall i \in [T]\} \\
&= \min_{\substack{\lambda \ge 0 \\ \gamma \in \mathbb{R}}} \{\gamma : \lambda^T (A^1 x^i - b^1) + \gamma \ge c^T x^i \quad \forall i \in [T]\}
\end{aligned}$$

$$= \min_{\substack{\lambda \in \mathbb{R}^{m_1} \\ \gamma \in \mathbb{R}}} \Big\{ \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}}_{=:e}^T \lambda + \underbrace{1}_{=:f} \cdot \gamma : \underbrace{\begin{pmatrix} A^1 x^1 - b^1 \\ \vdots \\ A^1 x^T - b^T \end{pmatrix}}_{=:E} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_{m_1} \end{pmatrix} + \underbrace{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}_{=:F} \cdot \gamma \ge \underbrace{\begin{pmatrix} c^T x^1 \\ \vdots \\ c^T x^T \end{pmatrix}}_{=:g},\ \lambda \ge 0 \Big\}$$

Now LP-duality tells us that

$$\min_{\lambda,\gamma} \left\{ e^T\lambda + f^T\gamma : \begin{array}{l} E\lambda + F\gamma \geq g \\ \lambda \geq 0 \end{array} \right\}$$

$$= \max_{u} \left\{ u^T g : u^T E \leq e^T,\ u^T F = f^T,\ u \geq 0 \right\}$$

In our case this implies that

$$z_{LD} = \max_{\alpha_1,\dots,\alpha_T \in \mathbb{R}} \left\{ \sum_{i=1}^{T} \alpha_i (c^T x^i) : \begin{array}{l} \sum_{i=1}^{T} \alpha_i (A^1 x^i - b^1) \leq 0 \\ \sum_{i=1}^{T} \alpha_i = 1 \\ \alpha \geq 0 \end{array} \right\}$$

$$= \max_{\alpha \in \mathbb{R}^T} \left\{ c^T \left(\sum_{i=1}^{T} \alpha_i x^i \right) : \underline{\alpha \geq 0,\ \sum_{i=1}^{T} \alpha_i = 1},\ A^1 \left(\sum_{i=1}^{T} \alpha_i x^i \right) \leq b^1 \right\}$$

$$= \max_{v} \left\{ c^T v : v \in \operatorname{conv}\{x^1,\dots,x^T\},\ A^1 v \leq b^1 \right\}$$

$$= \max_{v} \left\{ c^T v : v \in \operatorname{conv}(Q),\ A^1 v \leq b^1 \right\}.$$

$\square$

In general we have

$$\operatorname{conv}(S) \subseteq \operatorname{conv}(Q) \cap \{x : A^1 x \leq b^1\} \subseteq \{x : Ax \leq b\},$$

thus

$$z_{IP} \leq z_{LD} \leq z_{LP}.$$

* How do we compute $z_{LD} = \min\{ z_{LR}(\lambda) : \lambda \in \mathbb{R}^{m_1}_{\geq 0} \}$?

$z_{LR}(\lambda) \overset{!!}{=} f(\lambda)$

piecewise-linear, convex function

If $f(\lambda)$ were differentiable, we could use the steepest descent method: $\lambda^{k+1} := \lambda^k - \xi \cdot \nabla f(\lambda^k)$,

but we don't always have a gradient.

$\rightsquigarrow$ subgradient method.

<u>Def 6.13</u> Consider $f : \mathbb{R}^n \to \mathbb{R}$.

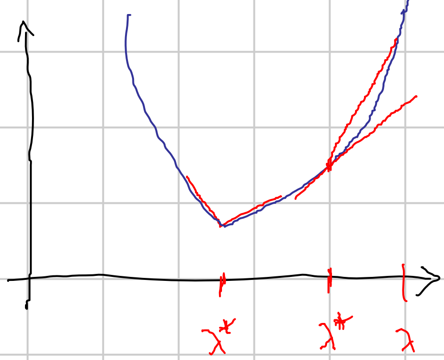
A vector $s \in \mathbb{R}^n$ is called a <u>subgradient</u> of f at λ^* if

$$\boxed{f(\lambda) \geq f(\lambda^*) + s^T(\lambda - \lambda^*) \quad \forall \lambda \in \mathbb{R}^n.}$$

The set of all subgradients of f at λ^* is denoted by $\partial f(\lambda^*)$ and called the <u>subdifferential</u> of f at λ^*

<u>Example</u> $n := 1$

$$\frac{f(\lambda) - f(\lambda^*)}{\lambda - \lambda^*} \geq s$$



<u>Lemma 6.14</u>

a) If a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then

$\forall \lambda^* : \quad \partial f(\lambda^*) \neq \emptyset$

b) If f is differentiable, then $\partial f(\lambda^*) = \{ \nabla f(\lambda^*) \}$.

c) If f is convex, then λ^* minimizes f over $\mathbb{R}^n$ iff $0 \in \partial f(\lambda^*)$.

Algo 6.15 (subgradient method to solve $\min\{f(\lambda) : \lambda \in \mathbb{R}^n\}$.)

1) Let $\lambda^1 \in \mathbb{R}^n$ be arbitrary, set $k := 1$ and choose $\xi_1, \xi_2, \ldots \in \mathbb{R}_{>0}$ such that

a) $\lim_{i \to \infty} \xi_i = 0$

b) $\sum_{i=1}^{\infty} \xi_i = \infty$

2) Compute $s \in \partial f(\lambda^k)$. If $s = 0$ then stop with optimal solution λ^k.

3) $\lambda^{k+1} := \lambda^k - \xi_k \frac{s}{\|s\|}$, $k := k+1$, goto 2).

Thm 6.16 Algo 6.15 converges.

Remark 6.17

When computing $z_{LD} = \min\{ z_{LR}(\lambda) : \lambda \geq 0 \}$, we proceed as follows:

- Suppose λ^k is given
- compute x^k by $\max_{x \in Q} \{ c^T x + (\lambda^k)^T (b^1 - A^1 x) \}$
- now use the fact that $b^1 - A^1 x^k$ is a subgradient:

$$\lambda^{k+1} := \max\{ 0, \lambda^k - \xi_k \cdot (b^1 - A^1 x^k) \}$$