Cutting Planes

Lecturer: Pradeep Ravikumar
Co-instructor: Aarti Singh

Convex Optimization 10-725/36-725

Dual Derivatives

- Let

$$x_\mu = \arg\min_{x \in X} L(x, \mu) = \arg\min_{x \in X} \Big\{ f(x) + \mu' g(x) \Big\}.$$

Dual Derivatives

- Let

$$x_\mu = \arg\min_{x\in X} L(x,\mu) = \arg\min_{x\in X}\Big\{f(x) + \mu' g(x)\Big\}.$$

Then for all $\tilde{\mu} \in \Re^r$,

$$\begin{aligned} q(\tilde{\mu}) &= \inf_{x\in X}\Big\{f(x) + \tilde{\mu}' g(x)\Big\} \\ &\le f(x_\mu) + \tilde{\mu}' g(x_\mu) \\ &= f(x_\mu) + \mu' g(x_\mu) + (\tilde{\mu} - \mu)' g(x_\mu) \\ &= q(\mu) + (\tilde{\mu} - \mu)' g(x_\mu). \end{aligned}$$

Dual Derivatives

- Let

$$x_\mu = \arg\min_{x \in X} L(x, \mu) = \arg\min_{x \in X} \Big\{ f(x) + \mu' g(x) \Big\}.$$

Then for all $\tilde{\mu} \in \Re^r$,

$$\begin{aligned} q(\tilde{\mu}) &= \inf_{x \in X} \Big\{ f(x) + \tilde{\mu}' g(x) \Big\} \\ &\leq f(x_\mu) + \tilde{\mu}' g(x_\mu) \\ &= f(x_\mu) + \mu' g(x_\mu) + (\tilde{\mu} - \mu)' g(x_\mu) \\ &= q(\mu) + (\tilde{\mu} - \mu)' g(x_\mu). \end{aligned}$$

- Thus $g(x_\mu)$ *is a subgradient of* q *at* μ.

Example: Polyhedral, Non-differentiable Dual

$$q(\mu) = \min_{i \in I} \left\{ a_i' \mu + b_i \right\},$$

where I is a finite index set, and $a_i \in \Re^r$ and b_i are given (arises when X is a discrete set, as in integer programming).

Example: Polyhedral, Non-differentiable Dual

$$q(\mu) = \min_{i \in I} \{ a_i' \mu + b_i \},$$

where I is a finite index set, and $a_i \in \Re^r$ and b_i are given (arises when X is a discrete set, as in integer programming).

- **Proposition:** Let q be polyhedral as above, and let I_μ be the set of indices attaining the minimum

$$I_\mu = \{ i \in I \mid a_i' \mu + b_i = q(\mu) \}.$$

The set of all subgradients of q at μ is

$$\partial q(\mu) = \left\{ g \;\middle|\; g = \sum_{i \in I_\mu} \xi_i a_i,\ \xi_i \geq 0,\ \sum_{i \in I_\mu} \xi_i = 1 \right\}.$$

Primal Problem

minimize $f(x)$

subject to $x \in X, \qquad g_j(x) \leq 0, \qquad j = 1, \ldots, r,$

assuming $-\infty < f^* < \infty.$

Dual Problem

minimize $f(x)$

subject to $x \in X, \qquad g_j(x) \leq 0, \qquad j = 1, \ldots, r,$

assuming $-\infty < f^* < \infty$.

- Dual problem: Maximize

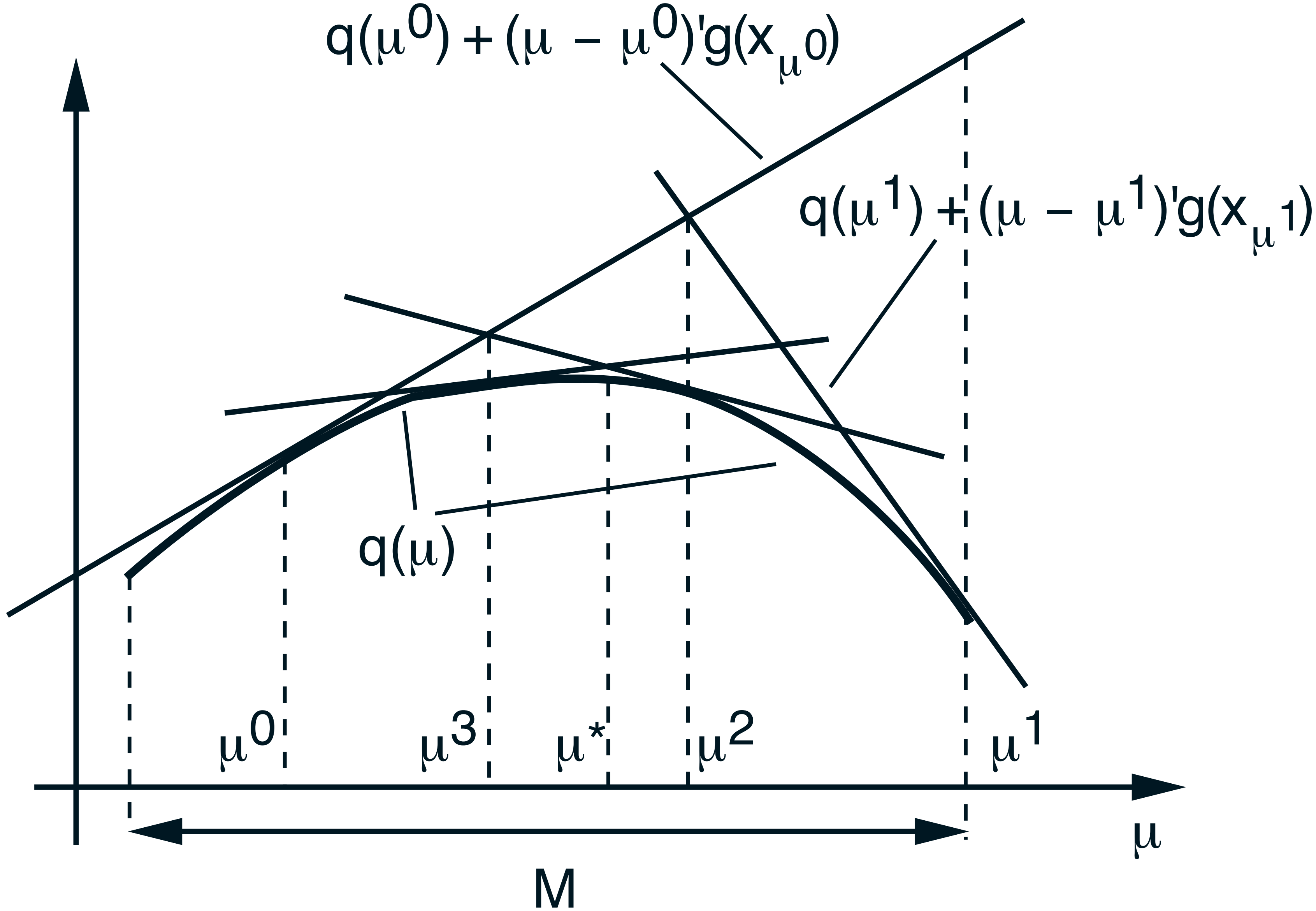
$$q(\mu) = \inf_{x \in X} L(x, \mu) = \inf_{x \in X} \{f(x) + \mu' g(x)\}$$

subject to $\mu \in M = \{\mu \mid \mu \geq 0,\ q(\mu) > -\infty\}$.

Cutting Plane Algorithms

- Cutting Plane Algorithms: iteratively refine the constraint set, or objective function by means of linear inequalities
- **Constraint Set:** For integer linear programs, iterate: if LP relaxation optimal point is not integral, refine the LP constraint set by a linear inequality separating non-integral point from integer constraint set
- **Objective Function:** Useful for convex, but non-differentiable programs
  - Iteratively approximate objective via piece-wise linear function
  - Popular use: for solving non-differentiable dual programs
  - We will be focusing on this class of cutting plane algorithms

Cutting Plane Method



- Solve piece-wise linear approximation to dual

Piece-wise Linear Approx.

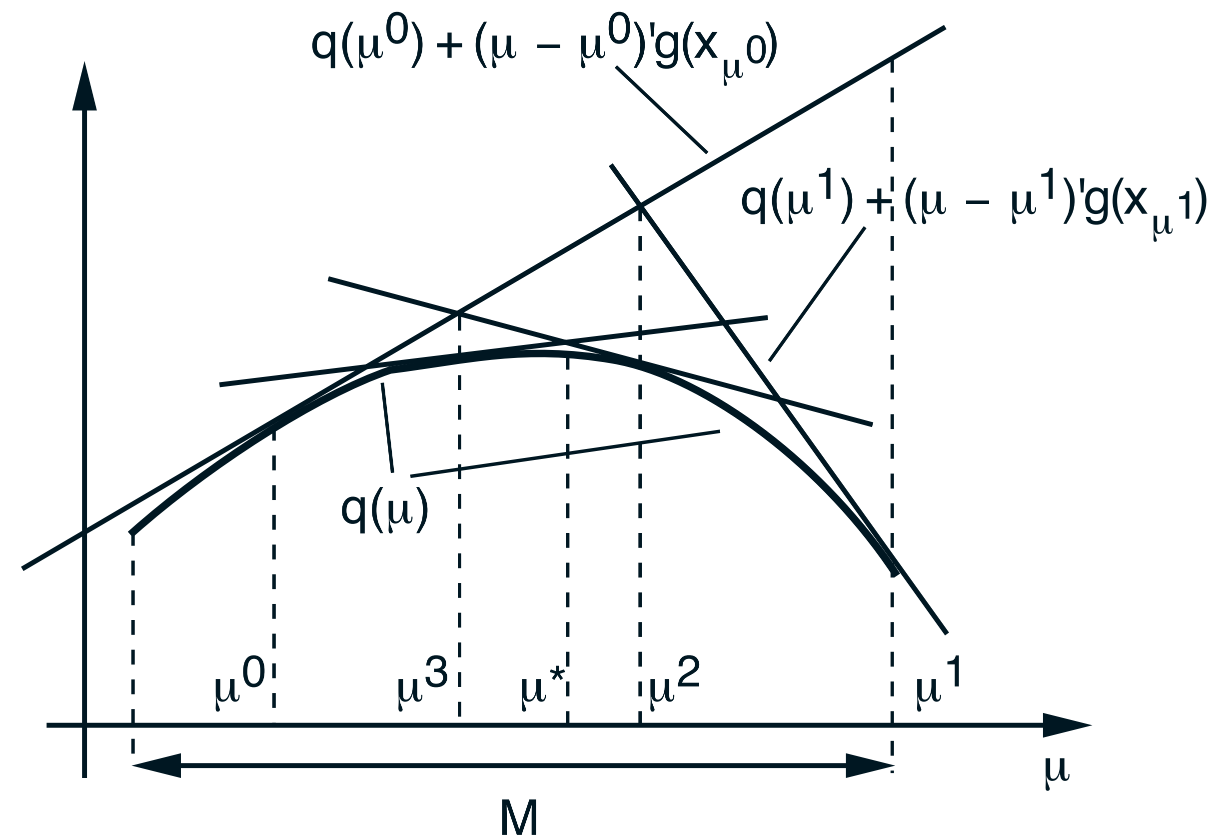
- kth iteration, after μ^i and $g^i = g\left(x_{\mu^i}\right)$ have been generated for $i = 0, \ldots, k-1$:

$$Q^k(\mu) = \min_{i=0,\ldots,k-1} \left\{ q(\mu^i) + (\mu - \mu^i)' g^i \right\}.$$

Piece-wise Linear Approx.

- kth iteration, after μ^i and $g^i = g\big(x_{\mu^i}\big)$ have been generated for $i = 0, \ldots, k-1$:

$$Q^k(\mu) = \min_{i=0,\ldots,k-1} \Big\{ q(\mu^i) + (\mu - \mu^i)' g^i \Big\}.$$

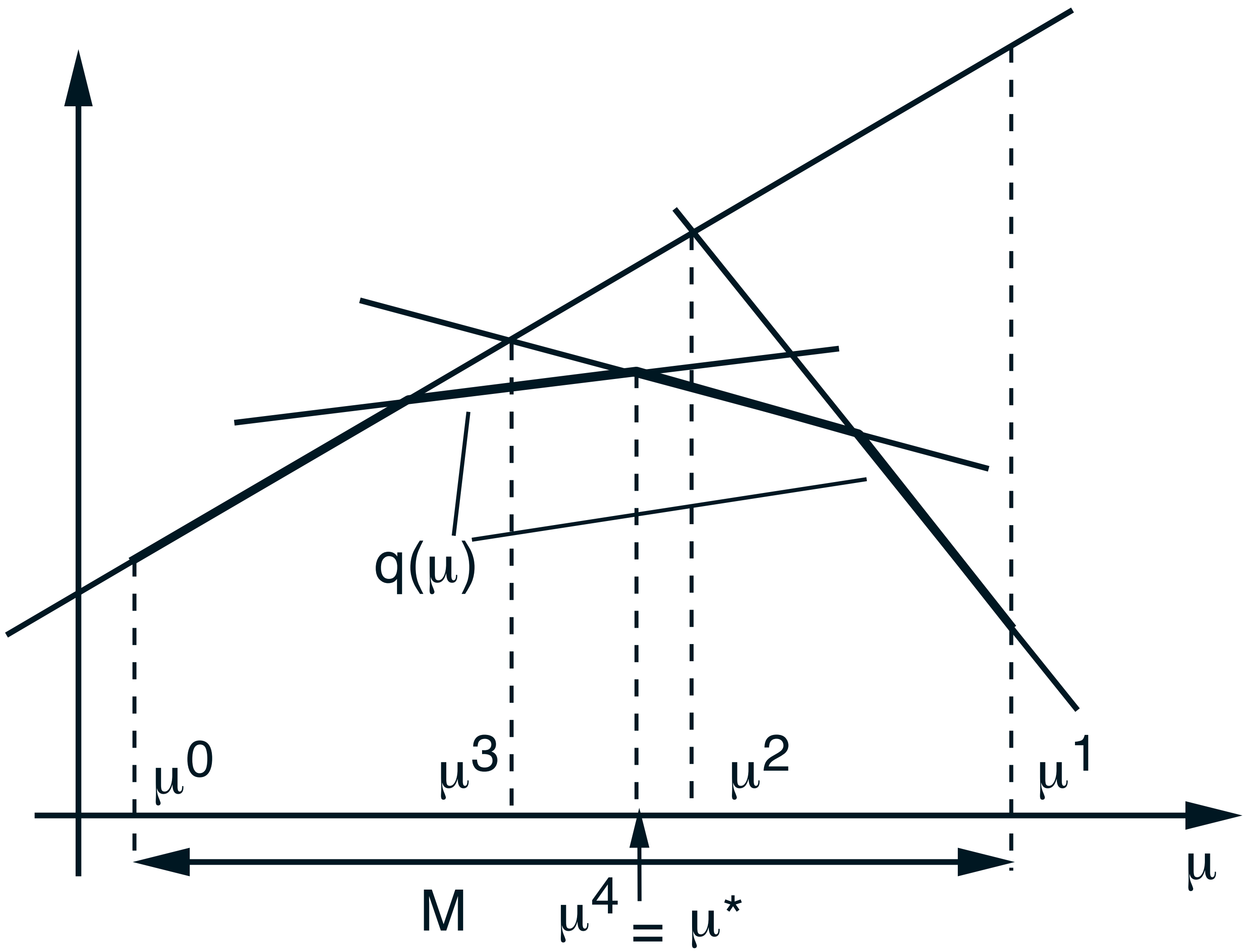


Cutting Plane Method

- kth iteration, after μ^i and $g^i = g\left(x_{\mu^i}\right)$ have been generated for $i = 0, \ldots, k-1$: Solve

$$\max_{\mu \in M} Q^k(\mu)$$

Polyhedral Case



- Where dual objective is already a piece-wise linear function

Polyhedral Case

$$q(\mu) = \min_{i \in I} \left\{ a_i' \mu + b_i \right\}$$

where I is a finite index set, and $a_i \in \Re^r$ and b_i are given.

Polyhedral Case

$$q(\mu) = \min_{i \in I} \left\{ a_i' \mu + b_i \right\}$$

where I is a finite index set, and $a_i \in \Re^r$ and b_i are given.

- Then subgradient g^k in the cutting plane method is a vector a_{i^k} for which the minimum is attained.

Polyhedral Case

$$q(\mu) = \min_{i \in I} \left\{ a_i' \mu + b_i \right\}$$

where I is a finite index set, and $a_i \in \Re^r$ and b_i are given.

- Then subgradient g^k in the cutting plane method is a vector a_{i^k} for which the minimum is attained.
- Finite termination expected.

Convergence

- **Proposition:** Assume that the max of Q_k over M is attained and that q is real-valued. Then every limit point of a sequence $\{\mu^k\}$ generated by the cutting plane method is a dual optimal solution.

Proof of Convergence

Proof: g^i is a subgradient of q at μ^i, so

$$q(\mu^i) + (\mu - \mu^i)' g^i \geq q(\mu), \qquad \forall \, \mu \in M,$$

$$Q^k(\mu^k) \geq Q^k(\mu) \geq q(\mu), \qquad \forall \, \mu \in M. \qquad (1)$$

Proof of Convergence

Proof: g^i is a subgradient of q at μ^i, so

$$q(\mu^i) + (\mu - \mu^i)' g^i \geq q(\mu), \qquad \forall\ \mu \in M,$$

$$Q^k(\mu^k) \geq Q^k(\mu) \geq q(\mu), \qquad \forall\ \mu \in M. \qquad (1)$$

- Suppose $\{\mu^k\}_K$ converges to $\bar{\mu}$. Then, $\bar{\mu} \in M$, and by Eq. (1) and continuity of Q^k and q (real-valued assumption), $Q^k(\bar{\mu}) \geq q(\bar{\mu})$. Using this and Eq. (1), we obtain for all k and $i < k$,

$$q(\mu^i) + (\mu^k - \mu^i)' g^i \geq Q^k(\mu^k) \geq Q^k(\bar{\mu}) \geq q(\bar{\mu}).$$

Proof of Convergence

Proof: g^i is a subgradient of q at μ^i, so

$$q(\mu^i) + (\mu - \mu^i)' g^i \geq q(\mu), \qquad \forall\ \mu \in M,$$

$$Q^k(\mu^k) \geq Q^k(\mu) \geq q(\mu), \qquad \forall\ \mu \in M. \qquad (1)$$

- Suppose $\{\mu^k\}_K$ converges to $\bar{\mu}$. Then, $\bar{\mu} \in M$, and by Eq. (1) and continuity of Q^k and q (real-valued assumption), $Q^k(\bar{\mu}) \geq q(\bar{\mu})$. Using this and Eq. (1), we obtain for all k and $i < k$,

$$q(\mu^i) + (\mu^k - \mu^i)' g^i \geq Q^k(\mu^k) \geq Q^k(\bar{\mu}) \geq q(\bar{\mu}).$$

- Take the limit as $i \to \infty$, $k \to \infty$, $i \in K$, $k \in K$,

$$\lim_{k \to \infty,\, k \in K} Q^k(\mu^k) = q(\bar{\mu}).$$

Proof of Convergence

Proof: g^i is a subgradient of q at μ^i, so

$$q(\mu^i) + (\mu - \mu^i)'g^i \geq q(\mu), \qquad \forall\ \mu \in M,$$

$$Q^k(\mu^k) \geq Q^k(\mu) \geq q(\mu), \qquad \forall\ \mu \in M. \qquad (1)$$

- Suppose $\{\mu^k\}_K$ converges to $\bar{\mu}$. Then, $\bar{\mu} \in M$, and by Eq. (1) and continuity of Q^k and q (real-valued assumption), $Q^k(\bar{\mu}) \geq q(\bar{\mu})$. Using this and Eq. (1), we obtain for all k and $i < k$,

$$q(\mu^i) + (\mu^k - \mu^i)'g^i \geq Q^k(\mu^k) \geq Q^k(\bar{\mu}) \geq q(\bar{\mu}).$$

- Take the limit as $i \to \infty$, $k \to \infty$, $i \in K$, $k \in K$,

$$\lim_{k \to \infty,\, k \in K} Q^k(\mu^k) = q(\bar{\mu}).$$

Combining with (1), $q(\bar{\mu}) = \max_{\mu \in M} q(\mu)$.

Separable Problem

minimize $\displaystyle\sum_{j=1}^{J} f_j(x_j)$

subject to $x_j \in X_j, \quad j = 1, \ldots, J, \quad \displaystyle\sum_{j=1}^{J} A_j x_j = b.$

Separable Problem

minimize $\sum_{j=1}^{J} f_j(x_j)$

subject to $x_j \in X_j, \quad j = 1, \ldots, J, \qquad \sum_{j=1}^{J} A_j x_j = b.$

- Solving for its dual:

$$q(\lambda) = \sum_{j=1}^{J} \min_{x_j \in X_j} \left\{ f_j(x_j) + \lambda' A_j x_j \right\} - \lambda' b$$

$$= \sum_{j=1}^{J} \left\{ f_j\big(x_j(\lambda)\big) + \lambda' A_j x_j(\lambda) \right\} - \lambda' b$$

where $x_j(\lambda)$ attains the min.

Separable Problem

minimize $\sum_{j=1}^{J} f_j(x_j)$

subject to $x_j \in X_j, \quad j = 1, \ldots, J, \qquad \sum_{j=1}^{J} A_j x_j = b.$

- Solving for its dual:

$$q(\lambda) = \sum_{j=1}^{J} \min_{x_j \in X_j} \left\{ f_j(x_j) + \lambda' A_j x_j \right\} - \lambda' b$$

$$= \sum_{j=1}^{J} \left\{ f_j\big(x_j(\lambda)\big) + \lambda' A_j x_j(\lambda) \right\} - \lambda' b$$

where $x_j(\lambda)$ attains the min.

subgradient at λ: $\quad g_\lambda = \sum_{j=1}^{J} A_j x_j(\lambda) - b.$

Dantzig-Wolfe Decomposition

- D-W decomposition method is just the cutting plane applied to the dual problem $\max_\lambda q(\lambda)$.
- At the kth iteration, we solve the “approximate dual”

$$\lambda^k = \arg\max_{\lambda \in \Re^r} Q^k(\lambda) \equiv \min_{i=0,\ldots,k-1} \left\{ q(\lambda^i) + (\lambda - \lambda^i)' g^i \right\}.$$

Dantzig-Wolfe Decomposition

- At the kth iteration, we solve the “approximate dual”

$$\lambda^k = \arg\max_{\lambda \in \Re^r} Q^k(\lambda) \equiv \min_{i=0,\ldots,k-1} \Big\{ q(\lambda^i) + (\lambda - \lambda^i)' g^i \Big\}.$$

- Equivalent linear program in v and λ

maximize v

subject to $v \leq q(\lambda^i) + (\lambda - \lambda^i)' g^i, \quad i = 0, \ldots, k-1$

Dantzig-Wolfe Decomposition

- At the kth iteration, we solve the "approximate dual"

$$\lambda^k = \arg \max_{\lambda \in \Re^r} Q^k(\lambda) \equiv \min_{i=0,\ldots,k-1} \Big\{ q(\lambda^i) + (\lambda - \lambda^i)' g^i \Big\}.$$

- Equivalent linear program in v and λ

maximize v

subject to $v \leq q(\lambda^i) + (\lambda - \lambda^i)' g^i, \quad i = 0, \ldots, k-1$

The dual of this (called $master\ problem$) is

$$\text{minimize} \quad \sum_{i=0}^{k-1} \xi^i \Big(q(\lambda^i) - \lambda^{i\prime} g^i \Big)$$

$$\text{subject to} \quad \sum_{i=0}^{k-1} \xi^i = 1, \qquad \sum_{i=0}^{k-1} \xi^i g^i = 0,$$

$$\xi^i \geq 0, \quad i = 0, \ldots, k-1,$$

- The master problem is written as

$$\text{minimize} \quad \sum_{j=1}^{J} \left(\sum_{i=0}^{k-1} \xi^i f_j\big(x_j(\lambda^i)\big) \right)$$

$$\text{subject to} \quad \sum_{i=0}^{k-1} \xi^i = 1, \qquad \sum_{j=1}^{J} A_j \left(\sum_{i=0}^{k-1} \xi^i x_j(\lambda^i) \right) = b,$$

$$\xi^i \geq 0, \quad i = 0, \ldots, k-1.$$

- The master problem is written as

$$\text{minimize} \quad \sum_{j=1}^{J} \left(\sum_{i=0}^{k-1} \xi^i f_j\big(x_j(\lambda^i)\big) \right)$$

$$\text{subject to} \quad \sum_{i=0}^{k-1} \xi^i = 1, \qquad \sum_{j=1}^{J} A_j \left(\sum_{i=0}^{k-1} \xi^i x_j(\lambda^i) \right) = b,$$

$$\xi^i \geq 0, \quad i = 0, \ldots, k-1.$$

- The primal cost function terms $f_j(x_j)$ are approximated by

$$\sum_{i=0}^{k-1} \xi^i f_j\big(x_j(\lambda^i)\big)$$

- The master problem is written as

$$\text{minimize} \quad \sum_{j=1}^{J} \left(\sum_{i=0}^{k-1} \xi^i f_j\big(x_j(\lambda^i)\big) \right)$$

$$\text{subject to} \quad \sum_{i=0}^{k-1} \xi^i = 1, \qquad \sum_{j=1}^{J} A_j \left(\sum_{i=0}^{k-1} \xi^i x_j(\lambda^i) \right) = b,$$

$$\xi^i \geq 0, \quad i = 0, \ldots, k-1.$$

- The primal cost function terms $f_j(x_j)$ are approximated by

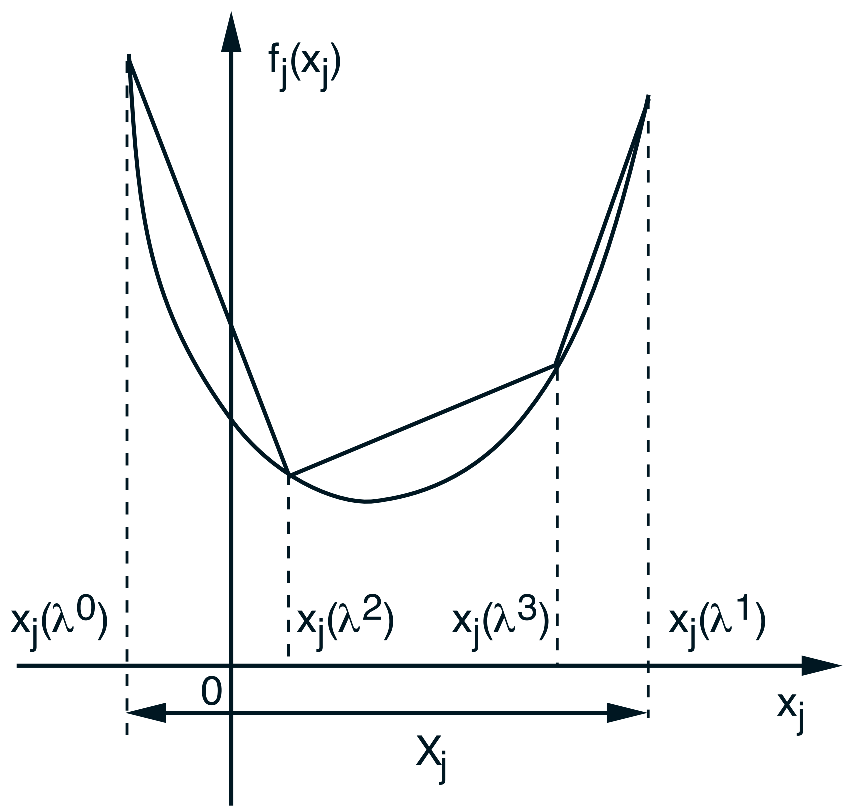
$$\sum_{i=0}^{k-1} \xi^i f_j\big(x_j(\lambda^i)\big)$$

- Vectors x_j are expressed as

$$\sum_{i=0}^{k-1} \xi^i x_j(\lambda^i)$$

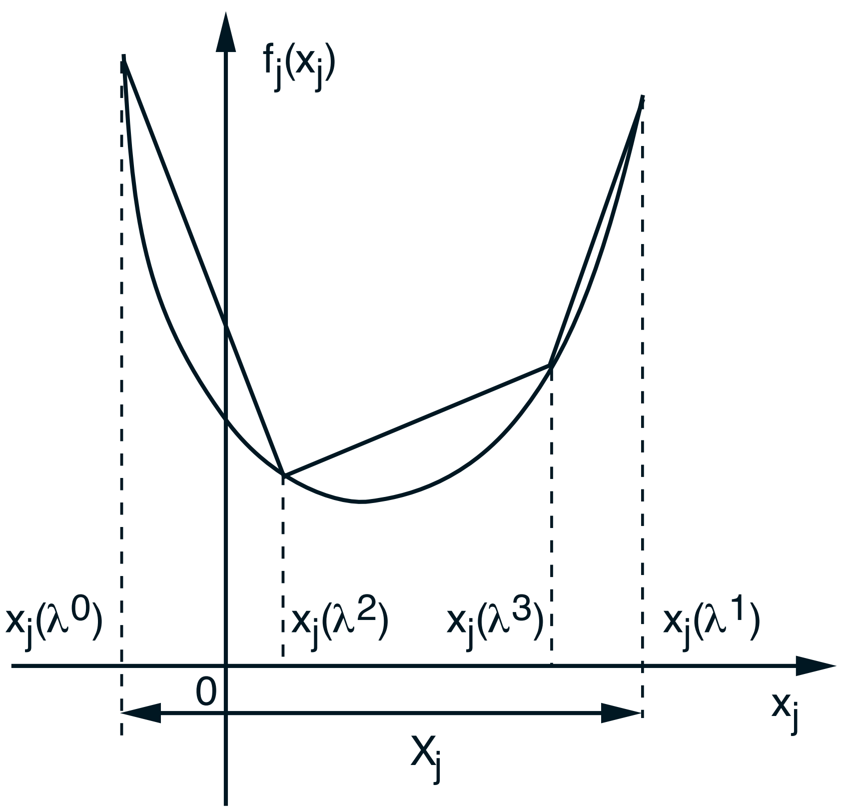
Dantzig-Wolfe Decomposition: Geometric Intuition

- Geometric interpretation of the master problem (the dual of the approximate dual solved in the cutting plane method) is $inner\ linearization$.



Dantzig-Wolfe Decomposition: Geometric Intuition

- Geometric interpretation of the master problem (the dual of the approximate dual solved in the cutting plane method) is *inner linearization*.



- This is a "dual" operation to the one involved in the cutting plane approximation, which can be viewed as *outer linearization*.