LECTURE 18

LECTURE OUTLINE

- Generalized polyhedral approximation methods
- Combined cutting plane and simplicial decomposition methods
- Lecture based on the paper

D. P. Bertsekas and H. Yu, "A Unifying Polyhedral Approximation Framework for Convex Optimization," SIAM J. on Optimization, Vol. 21, 2011, pp. 333-360.

All figures are courtesy of Athena Scientific, and are used with permission.

Generalized Polyhedral Approximations in Convex Optimization

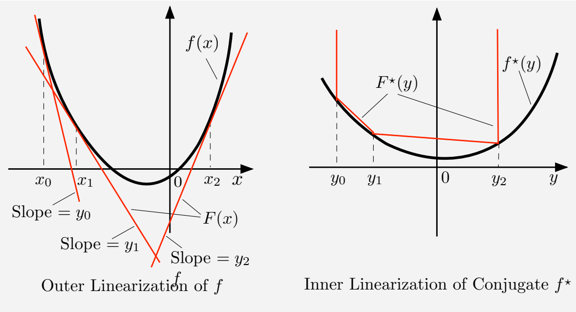
Dimitri P. Bertsekas

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Lecture 18, 6.253 Class

Lecture Summary

- Outer/inner linearization and their duality.



- A unifying framework for polyhedral approximation methods.
- Includes classical methods:

  Cutting plane/Outer linearization
  Simplicial decomposition/Inner linearization

- Includes new methods, and new versions/extensions of old methods.

Vehicle for Unification

- Extended monotropic programming (EMP)

$$\min_{(x_1,\ldots,x_m)\in S} \quad \sum_{i=1}^{m} f_i(x_i)$$

where $f_i : \Re^{n_i} \mapsto (-\infty, \infty]$ is a convex function and S is a subspace.

- The dual EMP is

$$\min_{(y_1,\ldots,y_m)\in S^\perp} \sum_{i=1}^{m} f_i^\star(y_i)$$

where $f_i^\star$ is the convex conjugate function of f_i.

- Algorithmic Ideas:
  - Outer or inner linearization for some of the f_i
  - Refinement of linearization using duality
- Features of outer or inner linearization use:
  - They are combined in the same algorithm
  - Their roles are reversed in the dual problem
  - Become two (mathematically equivalent dual) faces of the same coin

Advantage over Classical Cutting Plane Methods

- The refinement process is much faster.
  - Reason: At each iteration we add multiple cutting planes (as many as one per component function f_i).
  - By contrast a single cutting plane is added in classical methods.
- The refinement process may be more convenient.
  - For example, when f_i is a scalar function, adding a cutting plane to the polyhedral approximation of f_i can be very simple.
  - By contrast, adding a cutting plane may require solving a complicated optimization process in classical methods.

References

- D. P. Bertsekas, “Extended Monotropic Programming and Duality," Lab. for Information and Decision Systems Report 2692, MIT, Feb. 2010; a version appeared in JOTA, 2008, Vol. 139, pp. 209-225.
- D. P. Bertsekas, “Convex Optimization Theory," 2009, www-based “living chapter" on algorithms.
- D. P. Bertsekas and H. Yu, “A Unifying Polyhedral Approximation Framework for Convex Optimization," Lab. for Information and Decision Systems Report LIDS-P-2820, MIT, September 2009; SIAM J. on Optimization, Vol. 21, 2011, pp. 333-360.

Outline

1. Polyhedral Approximation
   - Outer and Inner Linearization
   - Cutting Plane and Simplicial Decomposition Methods
2. Extended Monotropic Programming
   - Duality Theory
   - General Approximation Algorithm
3. Special Cases
   - Cutting Plane Methods
   - Simplicial Decomposition for $\min_{x \in C} f(x)$

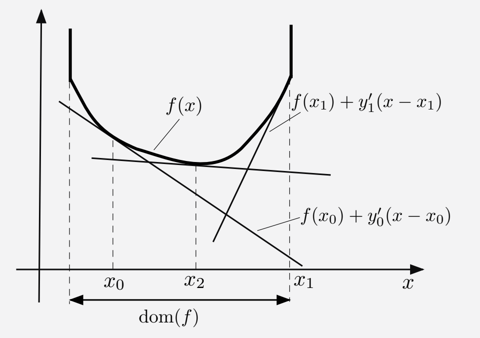
Outer Linearization - Epigraph Approximation by Halfspaces

- Given a convex function $f : \Re^n \mapsto (-\infty, \infty]$.
- Approximation using subgradients:

$$\max\left\{f(x_0) + y_0'(x - x_0), \ldots, f(x_k) + y_k'(x - x_k)\right\}$$

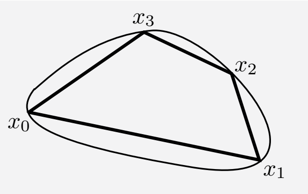
where

$$y_i \in \partial f(x_i), \qquad i = 0, \ldots, k$$

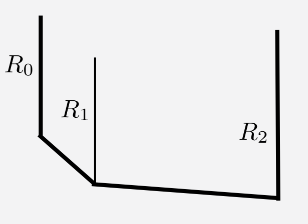


Convex Hulls

- Convex hull of a finite set of points x_i



- Convex hull of a union of a finite number of rays R_i

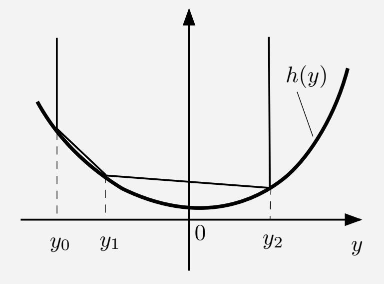


Inner Linearization - Epigraph Approximation by Convex Hulls

- Given a convex function $h : \Re^n \mapsto (-\infty, \infty]$ and a finite set of points

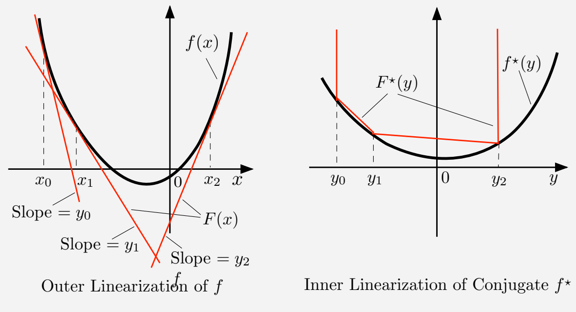
$$y_0, \ldots, y_k \in \mathrm{dom}(h)$$

- Epigraph approximation by convex hull of rays $\{(y_i, w) \mid w \geq h(y_i)\}$



Conjugacy of Outer/Inner Linearization

- Given a function $f : \Re^n \mapsto (-\infty, \infty]$ and its conjugate $f^\star$.
- The conjugate of an outer linearization of f is an inner linearization of $f^\star$.



- Subgradients in outer lin. <==> Break points in inner lin.

Cutting Plane Method for $\min_{x \in C} f(x)$ (C polyhedral)

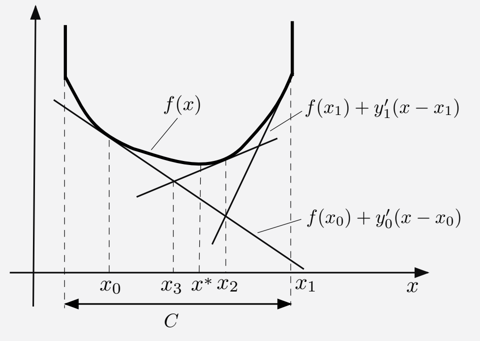
- Given $y_i \in \partial f(x_i)$ for $i = 0, \ldots, k$, form

$$F_k(x) = \max\left\{f(x_0) + y_0'(x - x_0), \ldots, f(x_k) + y_k'(x - x_k)\right\}$$

and let

$$x_{k+1} \in \arg\min_{x \in C} F_k(x)$$

- At each iteration solves LP of large dimension (which is simpler than the original problem).



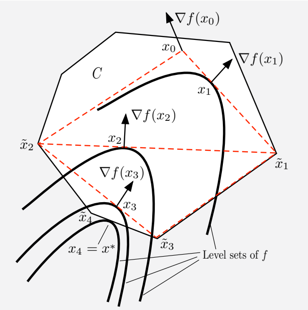
Simplicial Decomposition for $\min_{x \in C} f(x)$ (f smooth, C polyhedral)

- At the typical iteration we have x_k and $X_k = \{x_0, \tilde{x}_1, \ldots, \tilde{x}_k\}$, where $\tilde{x}_1, \ldots, \tilde{x}_k$ are extreme points of C.
- Solve LP of large dimension: Generate

$$\tilde{x}_{k+1} \in \arg\min_{x \in C} \{\nabla f(x_k)'(x - x_k)\}$$

- Solve NLP of small dimension: Set $X_{k+1} = \{\tilde{x}_{k+1}\} \cup X_k$, and generate x_{k+1} as

$$x_{k+1} \in \arg\min_{x \in \text{conv}(X_{k+1})} f(x)$$



- Finite convergence if C is a bounded polyhedron.

Comparison: Cutting Plane - Simplicial Decomposition

- Cutting plane aims to use LP with same dimension and smaller number of constraints.
- Most useful when problem has small dimension and:

  There are many linear constraints, or
  The cost function is nonlinear and linear versions of the problem are much simpler

- Simplicial decomposition aims to use NLP over a simplex of small dimension [i.e., $conv(X_k)$].
- Most useful when problem has large dimension and:

  Cost is nonlinear, and
  Solving linear versions of the (large-dimensional) problem is much simpler (possibly due to decomposition)

- The two methods appear very different, with unclear connection, despite the general conjugacy relation between outer and inner linearization.
- We will see that they are special cases of two methods that are dual (and mathematically equivalent) to each other.

Extended Monotropic Programming (EMP)

$$\min_{(x_1,\ldots,x_m)\in S} \sum_{i=1}^{m} f_i(x_i)$$

where $f_i : \Re^{n_i} \mapsto (-\infty, \infty]$ is a closed proper convex, S is subspace.

- Monotropic programming (Rockafellar, Minty), where f_i: scalar functions.
- Single commodity network flow (S: circulation subspace of a graph).
- Block separable problems with linear constraints.
- Fenchel duality framework: Let $m = 2$ and $S = \{(x, x) \mid x \in \Re^n\}$. Then the problem

  $$\min_{(x_1,x_2)\in S} f_1(x_1) + f_2(x_2)$$

  can be written in the Fenchel format

  $$\min_{x\in\Re^n} f_1(x) + f_2(x)$$

- Conic programs (second order, semidefinite - special case of Fenchel).
- Sum of functions (e.g., machine learning): For $S = \{(x, \ldots, x) \mid x \in \Re^n\}$

$$\min_{x\in\Re^n} \sum_{i=1}^{m} f_i(x)$$

Dual EMP

- Derivation: Introduce $z_i \in \Re^{n_i}$ and convert EMP to an equivalent form

$$\min_{(x_1,\ldots,x_m)\in S} \sum_{i=1}^{m} f_i(x_i) \qquad \equiv \qquad \min_{\substack{z_i=x_i,\ i=1,\ldots,m,\\ (x_1,\ldots,x_m)\in S}} \sum_{i=1}^{m} f_i(z_i)$$

- Assign multiplier $y_i \in \Re^{n_i}$ to constraint $z_i = x_i$, and form the Lagrangian

$$L(x,z,y) = \sum_{i=1}^{m} f_i(z_i) + y_i'(x_i - z_i)$$

  where $y = (y_1, \ldots, y_m)$.
- The dual problem is to maximize the dual function

$$q(y) = \inf_{(x_1,\ldots,x_m)\in S,\ z_i\in\Re^{n_i}} L(x,z,y)$$

- Exploiting the separability of $L(x,z,y)$ and changing sign to convert maximization to minimization, we obtain the dual EMP in symmetric form

$$\min_{(y_1,\ldots,y_m)\in S^{\perp}} \sum_{i=1}^{m} f_i^{\star}(y_i)$$

  where $f_i^{\star}$ is the convex conjugate function of f_i.

Optimality Conditions

- There are powerful conditions for strong duality $q^* = f^*$ (generalizing classical monotropic programming results):
  - Vector Sum Condition for Strong Duality: Assume that for all feasible x, the set
$$S^\perp + \partial_\epsilon(f_1 + \cdots + f_m)(x)$$
is closed for all $\epsilon > 0$. Then $q^* = f^*$.
  - Special Case: Assume each f_i is finite, or is polyhedral, or is essentially one-dimensional, or is domain one-dimensional. Then $q^* = f^*$.
  - By considering the dual EMP, "finite" may be replaced by "co-finite" in the above statement.

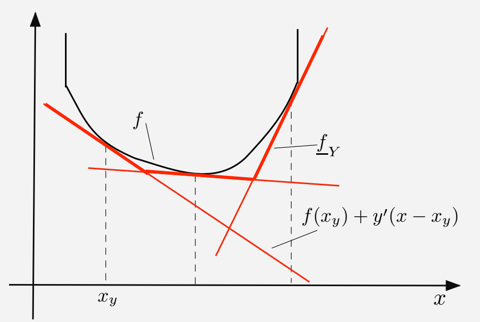
- Optimality conditions, assuming $-\infty < q^* = f^* < \infty$:
  - (x^*, y^*) is an optimal primal and dual solution pair if and only if
$$x^* \in S, \qquad y^* \in S^\perp, \qquad y_i^* \in \partial f_i(x_i^*), \quad i = 1, \ldots, m$$
  - Symmetric conditions involving the dual EMP:
$$x^* \in S, \qquad y^* \in S^\perp, \qquad x_i^* \in \partial f_i^\star(y_i^*), \quad i = 1, \ldots, m$$

Outer Linearization of a Convex Function: Definition

- Let $f : \Re^n \mapsto (-\infty, \infty]$ be closed proper convex.
- Given a finite set $Y \subset \mathrm{dom}(f^\star)$, we define the outer linearization of f

$$\underline{f}_Y(x) = \max_{y \in Y} \left\{ f(x_y) + y'(x - x_y) \right\}$$

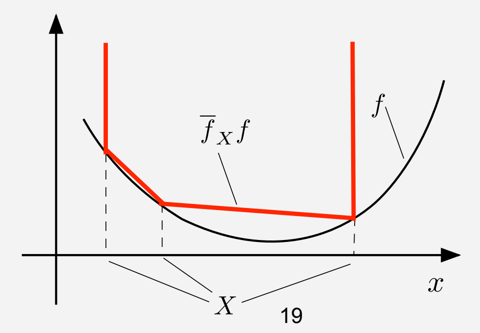
where x_y is such that $y \in \partial f(x_y)$.



Inner Linearization of a Convex Function: Definition

- Let $f : \Re^n \mapsto (-\infty, \infty]$ be closed proper convex.
- Given a finite set $X \subset \mathrm{dom}(f)$, we define the inner linearization of f as the function $\bar{f}_X$ whose epigraph is the convex hull of the rays $\{(x, w) \mid w \geq f(x),\ x \in X\}$:

$$\bar{f}_X(z) = \begin{cases} \min_{\substack{\sum_{x\in X} \alpha_x x = z, \\ \sum_{x\in X} \alpha_x = 1,\ \alpha_x \geq 0,\ x\in X}} \sum_{x\in X} \alpha_x f(z) & \text{if } z \in \mathrm{conv}(X) \\ \infty & \text{otherwise} \end{cases}$$



Polyhedral Approximation Algorithm

- Let $f_i : \Re^{n_i} \mapsto (-\infty, \infty]$ be closed proper convex, with conjugates $f_i^\star$. Consider the EMP

$$\min_{(x_1,\dots,x_m)\in S} \sum_{i=1}^{m} f_i(x_i)$$

- Introduce a fixed partition of the index set:

$$\{1,\dots,m\} = I \cup \underline{I} \cup \bar{I}, \qquad \underline{I} : \text{Outer indices}, \quad \bar{I} : \text{Inner indices}$$

- **Typical Iteration:** We have finite subsets $Y_i \subset \mathrm{dom}(f_i^\star)$ for each $i \in \underline{I}$, and $X_i \subset \mathrm{dom}(f_i)$ for each $i \in \bar{I}$.

  Find primal-dual optimal pair $\hat{x} = (\hat{x}_1,\dots,\hat{x}_m)$, and $\hat{y} = (\hat{y}_1,\dots,\hat{y}_m)$ of the approximate EMP

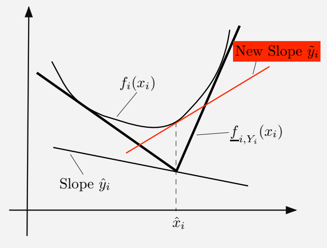
$$\min_{(x_1,\dots,x_m)\in S} \sum_{i\in I} f_i(x_i) + \sum_{i\in \underline{I}} \underline{f}_{i,Y_i}(x_i) + \sum_{i\in \bar{I}} \bar{f}_{i,X_i}(x_i)$$

  Enlarge Y_i and X_i by differentiation:

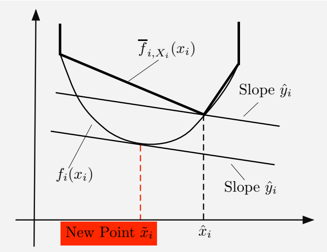
  - For each $i \in \underline{I}$, add $\tilde{y}_i$ to Y_i where $\tilde{y}_i \in \partial f_i(\hat{x}_i)$
  - For each $i \in \bar{I}$, add $\tilde{x}_i$ to X_i where $\tilde{x}_i \in \partial f_i^\star(\hat{y}_i)$.

Enlargement Step for ith Component Function

- Outer: For each $i \in \underline{I}$, add $\tilde{y}_i$ to Y_i where $\tilde{y}_i \in \partial f_i(\hat{x}_i)$.



- Inner: For each $i \in \bar{I}$, add $\tilde{x}_i$ to X_i where $\tilde{x}_i \in \partial f_i^\star(\hat{y}_i)$.



Mathematically Equivalent Dual Algorithm

- Instead of solving the primal approximate EMP

$$\min_{(x_1,\ldots,x_m)\in S} \sum_{i\in I} f_i(x_i) + \sum_{i\in \underline{I}} \underline{f}_{i,Y_i}(x_i) + \sum_{i\in \bar{I}} \bar{f}_{i,X_i}(x_i)$$

  we may solve its dual

$$\min_{(y_1,\ldots,y_m)\in S^\perp} \sum_{i\in I} f_i^\star(y_i) + \sum_{i\in \underline{I}} \underline{f^\star}_{i,Y_i}(y_i) + \sum_{i\in \bar{I}} \bar{f^\star}_{i,X_i}(x_i)$$

  where $\underline{f^\star}_{i,Y_i}$ and $\bar{f^\star}_{i,X_i}$ are the conjugates of $\underline{f}_{i,Y_i}$ and $\bar{f}_{i,X_i}$.

- Note that $\underline{f^\star}_{i,Y_i}$ is an inner linearization, and $\bar{f^\star}_{i,X_i}$ is an outer linearization (roles of inner/outer have been reversed).
- The choice of primal or dual is a matter of computational convenience, but does not affect the primal-dual sequences produced.

Comments on Polyhedral Approximation Algorithm

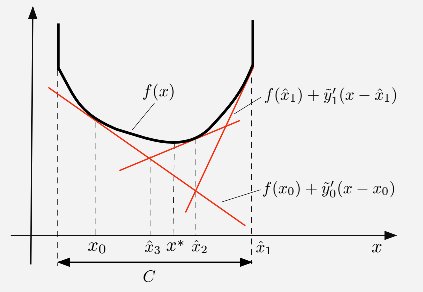
- In some cases we may use an algorithm that solves simultaneously the primal and the dual.
  - **Example:** Monotropic programming, where x_i is one-dimensional.
  - **Special case:** Convex separable network flow, where x_i is the one-dimensional flow of a directed arc of a graph, S is the circulation subspace of the graph.
- In other cases, it may be preferable to focus on solution of either the primal or the dual approximate EMP.
- After solving the primal, the refinement of the approximation ($\tilde{y}_i$ for $i \in \underline{I}$, and $\tilde{x}_i$ for $i \in \bar{I}$) may be found later by differentiation and/or some special procedure/optimization.
  - This may be easy, e.g., in the cutting plane method, or
  - This may be nontrivial, e.g., in the simplicial decomposition method.
- Subgradient duality [$y \in \partial f(x)$ iff $x \in \partial f^\star(y)$] may be useful.

Cutting Plane Method for $\min_{x \in C} f(x)$

- EMP equivalent: $\min_{x_1 = x_2} f(x_1) + \delta(x_2 \mid C)$, where $\delta(x_2 \mid C)$ is the indicator function of C.
- Classical cutting plane algorithm: Outer linearize f only, and solve the primal approximate EMP. It has the form

$$\min_{x \in C} \underline{f}_Y(x)$$

where Y is the set of subgradients of f obtained so far. If $\hat{x}$ is the solution, add to Y a subgradient $\tilde{y} \in \partial f(\hat{x})$.



Simplicial Decomposition Method for $\min_{x \in C} f(x)$

- EMP equivalent: $\min_{x_1 = x_2} f(x_1) + \delta(x_2 \mid C)$, where $\delta(x_2 \mid C)$ is the indicator function of C.
- Generalized Simplicial Decomposition: Inner linearize C only, and solve the primal approximate EMP. In has the form

$$\min_{x \in \bar{C}_X} f(x)$$

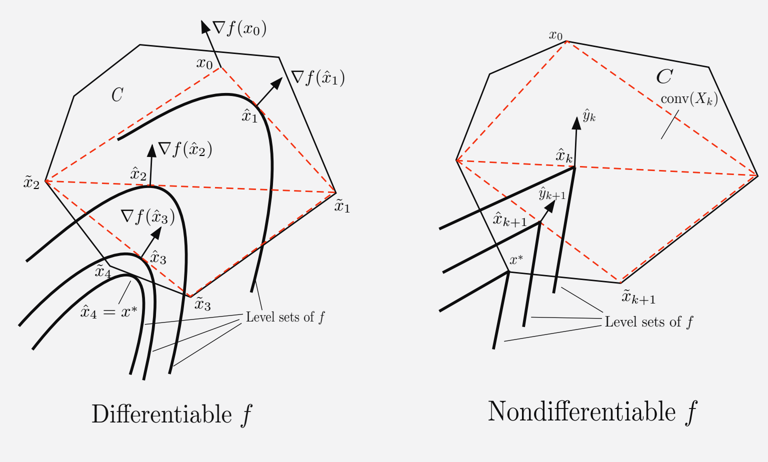
  where $\bar{C}_X$ is an inner approximation to C.
- Assume that $\hat{x}$ is the solution of the approximate EMP.
  - Dual approximate EMP solutions:

$$\left\{ (\hat{y}, -\hat{y}) \mid \hat{y} \in \partial f(\hat{x}),\ -\hat{y} \in (\text{normal cone of } \bar{C}_X \text{ at } \hat{x}) \right\}$$

  - In the classical case where f is differentiable, $\hat{y} = \nabla f(\hat{x})$.
  - Add to X a point $\tilde{x}$ such that $-\hat{y} \in \partial\delta(\tilde{x} \mid C)$, or

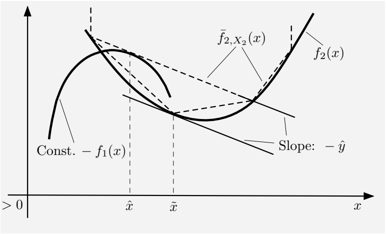
$$\tilde{x} \in \arg\min_{x \in C} \hat{y}' x$$

Illustration of Simplicial Decomposition for $\min_{x \in C} f(x)$

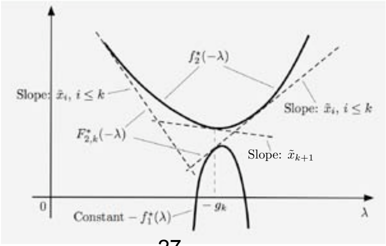


Dual Views for $\min_{x\in\Re^n}\left\{f_1(x)+f_2(x)\right\}$

- Inner linearize f_2



- Dual view: Outer linearize $f_2^\star$



Convergence - Polyhedral Case

- Assume that
  - All outer linearized functions f_i are finite polyhedral
  - All inner linearized functions f_i are co-finite polyhedral
  - The vectors $\tilde{y}_i$ and $\tilde{x}_i$ added to the polyhedral approximations are elements of the finite representations of the corresponding f_i
- Finite convergence: The algorithm terminates with an optimal primal-dual pair.
- Proof sketch: At each iteration two possibilities:
  - Either $(\hat{x}, \hat{y})$ is an optimal primal-dual pair for the original problem
  - Or the approximation of one of the f_i, $i \in \underline{I} \cup \bar{I}$, will be refined/improved
- By assumption there can be only a finite number of refinements. □

Convergence - Nonpolyhedral Case

- Convergence, pure outer linearization ($\bar{I}$: Empty). Assume that the sequence $\{\tilde{y}_i^k\}$ is bounded for every $i \in \underline{I}$. Then every limit point of $\{\hat{x}^k\}$ is primal optimal.
- Proof sketch: For all k, $\ell \leq k-1$, and $x \in S$, we have

$$\sum_{i \notin \underline{I}} f_i(\hat{x}_i^k) + \sum_{i \in \underline{I}} \left(f_i(\hat{x}_i^\ell) + (\hat{x}_i^k - \hat{x}_i^\ell)' \tilde{y}_i^\ell \right) \leq \sum_{i \notin \underline{I}} f_i(\hat{x}_i^k) + \sum_{i \in \underline{I}} \underline{f}_{i, Y_i^{k-1}}(\hat{x}_i^k) \leq \sum_{i=1}^m f_i(x_i)$$

- Let $\{\hat{x}^k\}_{\mathcal{K}} \to \bar{x}$ and take limit as $\ell \to \infty$, $k \in \mathcal{K}$, $\ell \in \mathcal{K}$, $\ell < k$. □
- Exchanging roles of primal and dual, we obtain a convergence result for pure inner linearization case.
- Convergence, pure inner linearization ($\underline{I}$: Empty). Assume that the sequence $\{\tilde{x}_i^k\}$ is bounded for every $i \in \bar{I}$. Then every limit point of $\{\hat{y}^k\}$ is dual optimal.
- General mixed case: Convergence proof is more complicated (see the Bertsekas and Yu paper).

Concluding Remarks

- A unifying framework for polyhedral approximations based on EMP.
- Dual and symmetric roles for outer and inner approximations.
- There is option to solve the approximation using a primal method or a dual mathematical equivalent - whichever is more convenient/efficient.
- Several classical methods and some new methods are special cases.
- Proximal/bundle-like versions:
  - Convex proximal terms can be easily incorporated for stabilization and for improvement of rate of convergence.
  - Outer/inner approximations can be carried from one proximal iteration to the next.

MIT OpenCourseWare
http://ocw.mit.edu

6.253 Convex Analysis and Optimization
Spring 2012

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.