OPER 732: Optimization under Uncertainty
Lecture 7: Regularized Decomposition

Department of Statistical Sciences and Operations Research
Virginia Commonwealth University

Oct. 2, 2017

Criticism of L-shaped method

- Bouncing around effect
- The current relaxation solution may be misleading!
- Cannot take advantage of “good” starting point

Behavior of the L-shaped method with an optimal initial solution:

- Move away from the optimal (or even feasible region).
- Then bounce around a while.
- At last converge back to optimal.

Q: How to fix this?

One cycle of multicut Benders decomposition

Master problem:

$$z_L^* = \min_{x \in X} \hat{F}(x) := c^\top x + \min \sum_{k \in N} p_k \theta^k$$

$$\text{s.t. } \theta^k \geq G^k x + g^k, \ (G^k, g^k) \in \mathcal{G}^k, \ \forall k \in N$$

$$Lx \geq l, \ (L, l) \in \mathcal{L}, \ \forall k \in N.$$

- For each $k \in N$, $\mathcal{G}^k$ collects all the optimality cuts added for scenario k, and $\mathcal{L}$ collects all the feasibility cuts added so far
- $\hat{F}(x)$ gives a piecewise linear relaxation of the true objective function $F(x)$
- Given a solution $(\hat{x}, \hat{\theta})$ of the master problem, check its optimality/feasiblity, and add more cuts if necessary

A couple more disadvantages:

- Too many cuts may be added
- There is no reliable criterion to drop cuts

One cycle of multicut Benders decomposition

Master problem:

$$z_L^* = \min_{x \in X} \hat{F}(x) := c^\top x + \min \sum_{k \in N} p_k \theta^k$$

$$\text{s.t. } \theta^k \geq G^k x + g^k, \ (G^k, g^k) \in \mathcal{G}^k, \ \forall k \in N$$

$$Lx \geq l, \ (L, l) \in \mathcal{L}, \ \forall k \in N.$$

- For each $k \in N$, $\mathcal{G}^k$ collects all the optimality cuts added for scenario k, and $\mathcal{L}$ collects all the feasibility cuts added so far
- $\hat{F}(x)$ gives a piecewise linear relaxation of the true objective function $F(x)$
- Given a solution $(\hat{x}, \hat{\theta})$ of the master problem, check its optimality/feasiblity, and add more cuts if necessary

---

A couple more disadvantages:

- Too many cuts may be added
- There is no reliable criterion to drop cuts

Regularized decomposition (Ruszczyński, QDECOM)

Motivation: let good initial solutions be truely beneficial

- When starting with some good overall feasible first-stage iterates z^i, want the next iterate x^i to be not too far from z^i
  - "overall": both first-stage and second-stage feasible
- Cuts are more likely to be built around the optimal solution x^*
  - Less cuts may be necessary
  - Faster convergence is expected

To do it, introduce a regularization term into the objective:

$$\min \; c^\top x + \sum_{k \in N} p_k \theta^k + \frac{1}{2\rho} \|x - z^i\|^2,$$

with a control parameter $\rho > 0$, and a reference point z^i in each iteration i

Regularized decomposition (Ruszczyński, QDECOM)

Motivation: let good initial solutions be truely beneficial

- When starting with some good overall feasible first-stage iterates z^i, want the next iterate x^i to be not too far from z^i
  - "overall": both first-stage and second-stage feasible
- Cuts are more likely to be built around the optimal solution x^*
  - Less cuts may be necessary
  - Faster convergence is expected

To do it, introduce a regularization term into the objective:

$$\min \; c^\top x + \sum_{k \in N} p_k \theta^k + \frac{1}{2\rho} \|x - z^i\|^2,$$

with a control parameter $\rho > 0$, and a reference point z^i in each iteration i

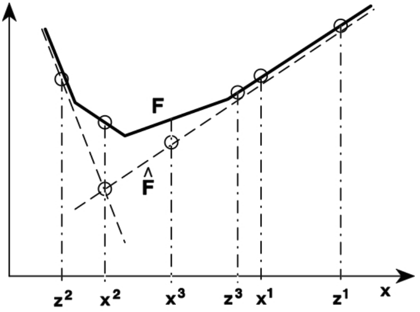
Moving forward

Two cases after solving the regularized master problem (and get the next iterate x^i):

- If x^i is not feasible, add feasibility cuts, and keep the reference point z^k unchanged
- If x^i is feasible, need to decide if we maintain the old reference point z^k or replace it by x^i

---

Three possibilities:



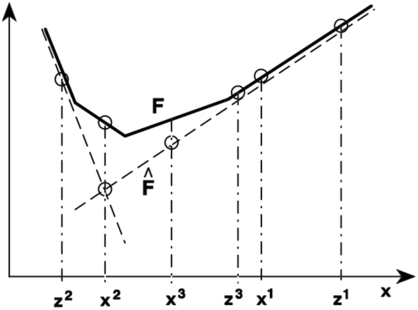
Moving forward

Two cases after solving the regularized master problem (and get the next iterate x^i):

- If x^i is not feasible, add feasibility cuts, and keep the reference point z^k unchanged
- If x^i is feasible, need to decide if we maintain the old reference point z^k or replace it by x^i

---

Three possibilities:



Three possibilities

1. Case 1: $\hat{F}(x^i) = F(x^i)$, the relaxation function is exact at x^i
   - Must move z^i to x^i! Since no cut will be added at x^i, if we do not change z^i, the algorithm will run into an infinite loop
2. Case 2: $\hat{F}(x^i) < F(x^i)$
   1. Case 2.1: x^i is at a “kink” of $\hat{F}$, and $F(x^i) < F(z^i)$, i.e., the upper bound has been “reliably” improved
      - Move z^i to x^i
   2. Case 2.2: Others: keep z^i unchanged

To check if x^i is a “kink” of $\hat{F}$, check if exactly $n_1 + N$ constraints are active at (x^i, θ^i)

Enable cut removal

After solving the regularized master problem, we can delete the inactive cuts from the master!

- Assume nondegeneracy, at the extreme points solution there is no more than $n_1 + N$ active constraints (Q: Why?)
  - So # of cuts are well controlled!
- Why can we delete inactive cuts?

Algorithm: QDECOM

1. Solve the QP master problem at z^i, get (x^i, θ^i). If $\hat{F}_i := c^\top x^i + \sum_i p_i \theta^i$ is very close to $F(z^i)$ (the current upper bound), stop.
2. Remove inactive optimality cuts at (x^i, θ^i).
3. Solve the second stage problems at x^i, and add feasibility/optimality cuts if necessary.
4. If at least one feasibility cut is added, put $z^{i+1} = z^i$, skip the next step.
5. If $F(x^i) = \hat{F}_i$, or else if $F(x^i) < F(z^i)$ and exactly $n + N$ constraints are active at (x^i, θ^i), put $z^{i+1} = x^i$. Otherwise, put $z^{i+1} = z^i$.
6. Let $i := i + 1$, go to Step 1.

Summary

Benefits of regularized decomposition:

- Stablize the master problem
- Make cut removal possible
- Facilitate good initial feasible solutions

- Readings:
  - Stochastic Programming book by Kall and Wallace, Section 3.3 on regularized decomposition
  - Ruszczyński's paper on regularized decomposition
- Homework 3