Solution Methods for Stochastic Programs

Huseyin Topaloglu
School of Operations Research and Information Engineering
Cornell University
ht88@cornell.edu

August 14, 2010

Outline

- Cutting plane methods for convex optimization
- Decomposition methods for two-stage stochastic programs
- Dual methods for two-stage stochastic programs

Cutting Plane Methods for Convex Optimization

A Simple Method for Convex Optimization

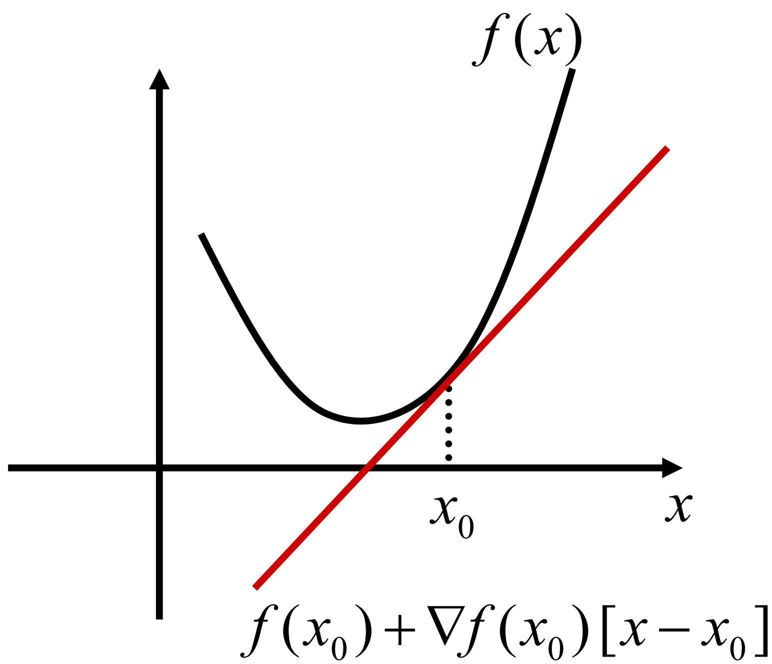
- Consider the optimization problem

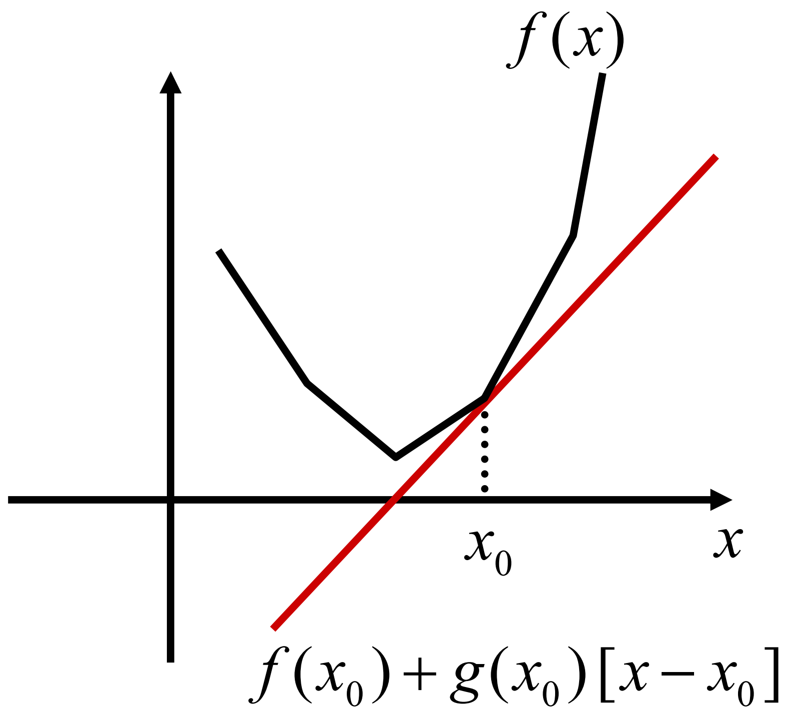
$$z^* = \min_{x \in X} f(x),$$

where f is convex and X is a compact and convex set

- Necessary background is a first course on optimization that includes linear programming duality
- Recall the subgradient inequality, which constructs a cutting plane approximation to f at point x_0 by

$$f(x_0) + \nabla f(x_0)\,[x - x_0] \leq f(x)$$


$f(x)$
x_0
x
$f(x_0)+\nabla f(x_0)[x-x_0]$


$f(x)$
x_0
x
$f(x_0)+g(x_0)[x-x_0]$

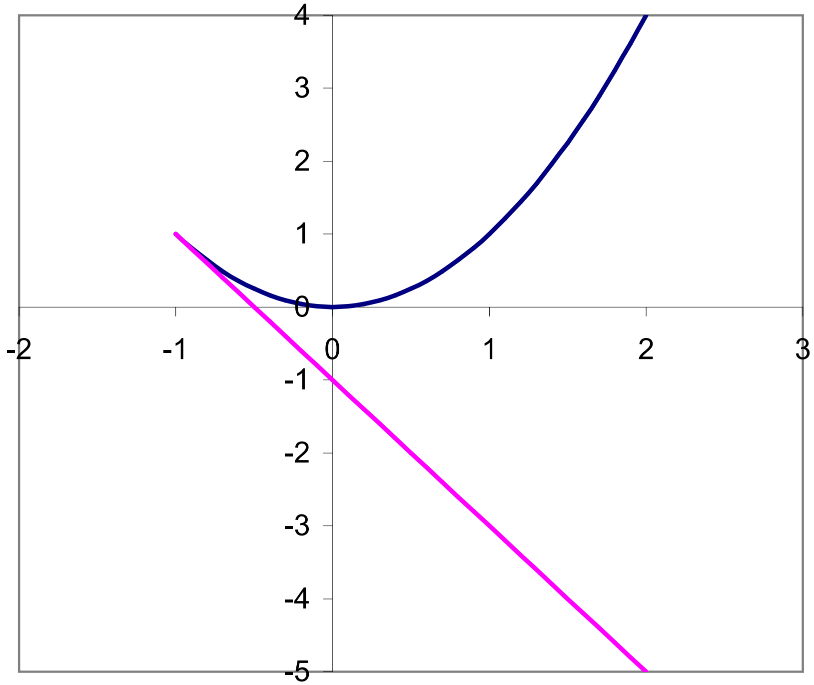
A Simple Method for Convex Optimization

- Consider the optimization problem

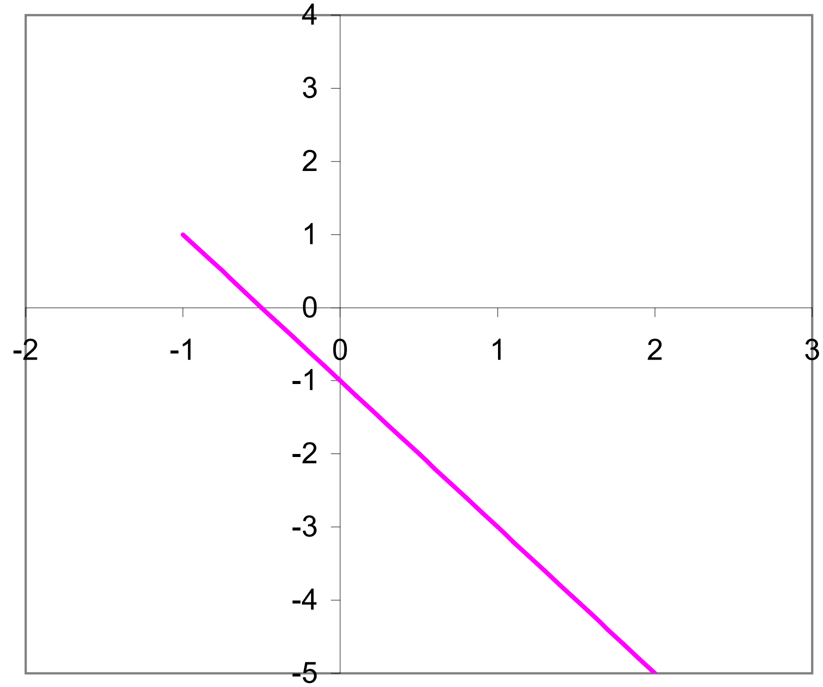
$$z^* = \min_{x \in [-1,2]} (x)^2$$

- Start with an initial guess for the optimal solution $x^1 = -1$
- Use the subgradient inequality to construct a cutting plane approximation to the objective function at x^1

$$(x^1)^2 + 2\,x^1\,(x - x^1) = 1 - 2\,(x + 1) = -2\,x - 1$$



4
3
2
1
0
-1
-2
-3
-4
-5
-2
-1
0
1
2
3



4
3
2
1
0
-1
-2
-3
-4
-5
-2
-1
0
1
2
3

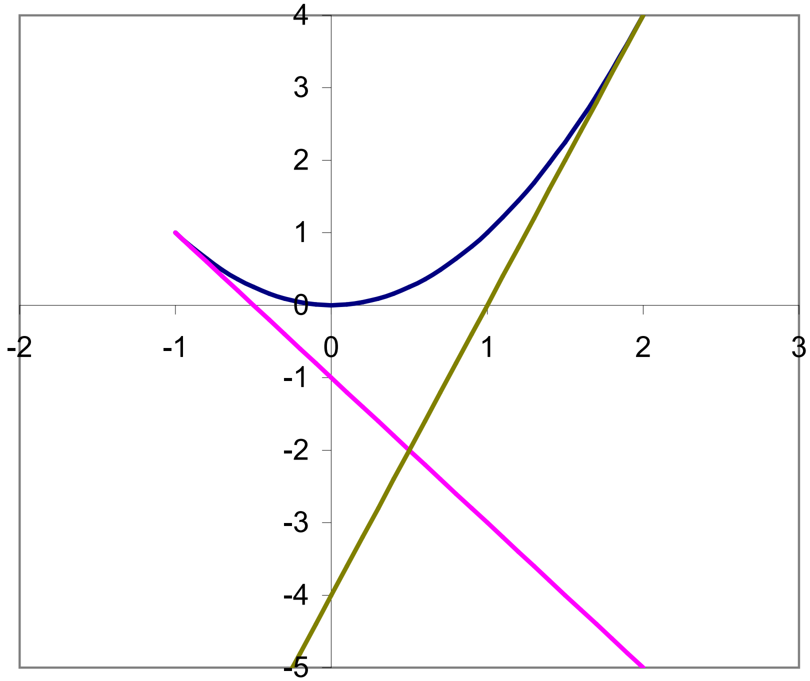
A Simple Method for Convex Optimization

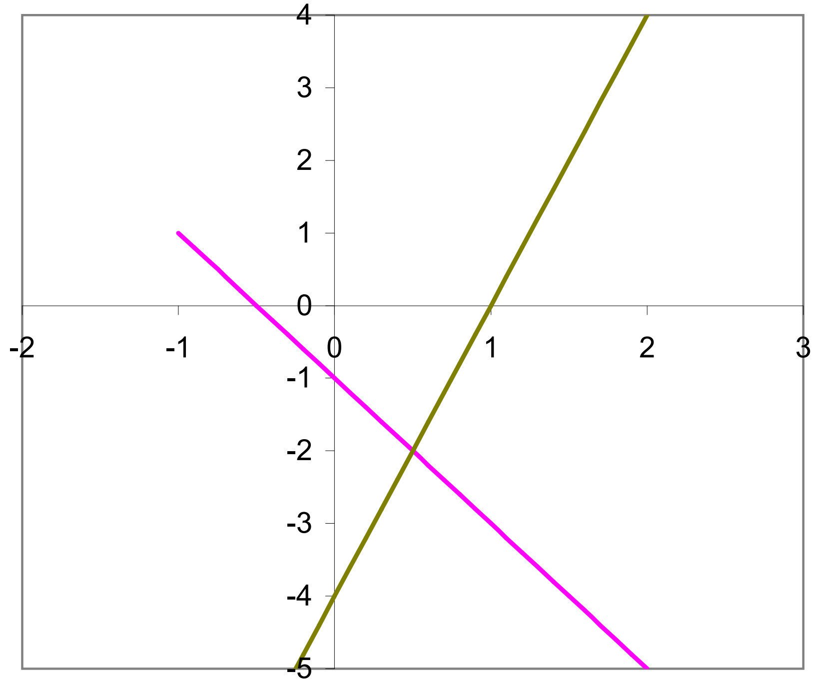
- Minimize the lower bound approximation over the feasible set by solving the problem

$$\begin{aligned} \min \quad & v \\ \text{subject to} \quad & x \in [-1, 2] \\ & v \geq -2\,x - 1 \end{aligned}$$

- Obtain a guess for the optimal solution $x^2 = 2$ and a guess for the optimal objective value $v^2 = -5$
- Use the subgradient inequality to construct a cutting plane approximation to the objective function at x^2

$$(x^2)^2 + 2\,x^2\,(x - x^2) = 4 + 4\,(x - 2) = 4\,x - 4$$


4
3
2
1
0
-2
-1
0
1
2
3
-1
-2
-3
-4
-5

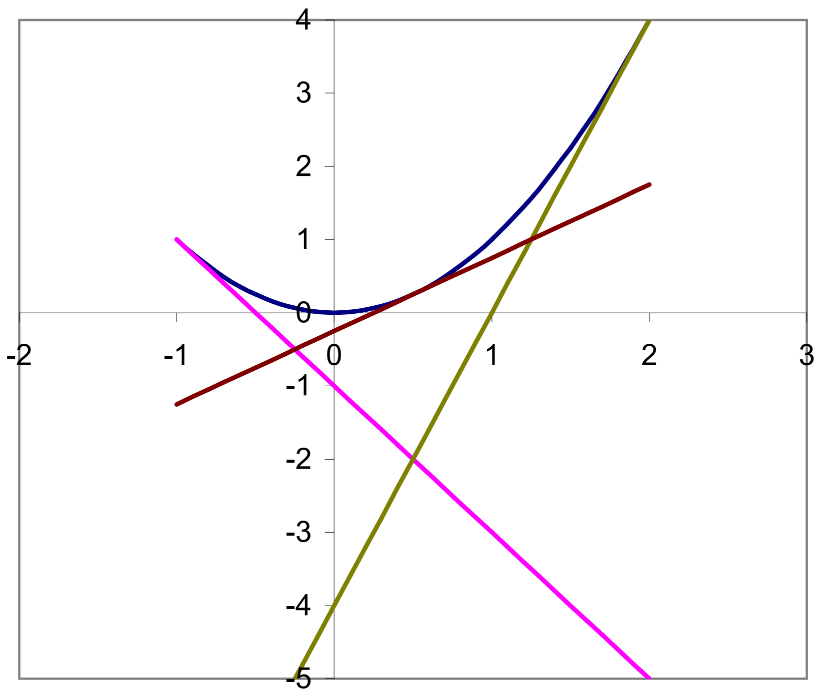

4
3
2
1
0
-2
-1
0
1
2
3
-1
-2
-3
-4
-5

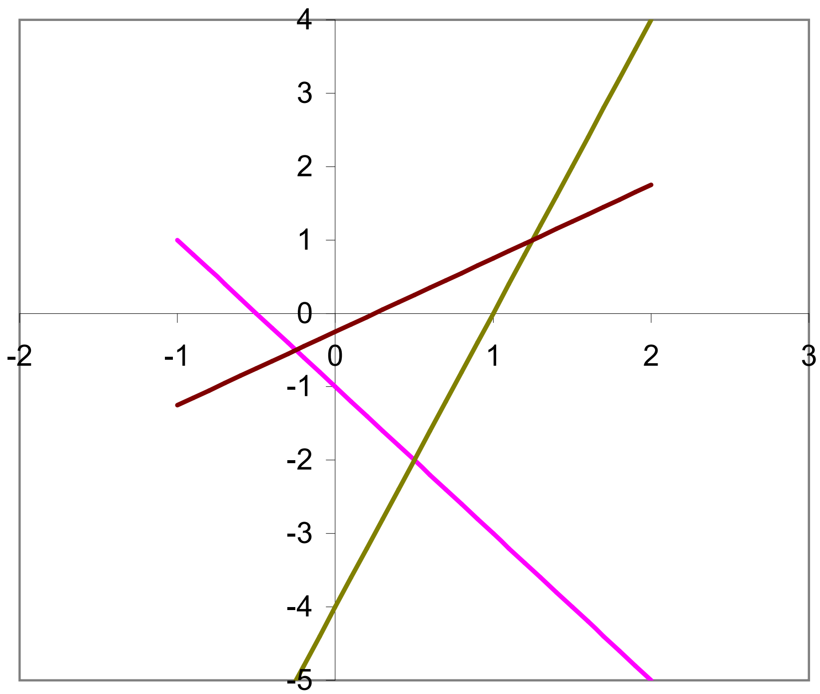
- Minimize the lower bound approximation over the feasible set by solving the problem

$$\begin{aligned} \min \quad & v \\ \text{subject to} \quad & x \in [-1, 2] \\ & v \geq -2\,x - 1,\ v \geq 4\,x - 4 \end{aligned}$$

- Obtain a guess for the optimal solution $x^3 = \frac{1}{2}$ and a guess for the optimal objective value $v^3 = -2$
- Use the subgradient inequality to construct a cutting plane approximation to the objective function at x^3

$$(x^3)^2 + 2\,x^3\,(x - x^3) = \tfrac{1}{4} + (x - \tfrac{1}{2}) = x - \tfrac{1}{4}$$


4
3
2
1
0
-1
-2
-3
-4
-5
-2
-1
0
1
2
3


4
3
2
1
0
-1
-2
-3
-4
-5
-2
-1
0
1
2
3

A Simple Method for Convex Optimization

- Minimize the lower bound approximation over the feasible set by solving the problem

$$\begin{aligned} \min \quad & v \\ \text{subject to} \quad & x \in [-1, 2] \\ & v \geq -2\,x - 1,\ v \geq 4\,x - 4,\ v \geq x - 1/4 \end{aligned}$$

- Obtain a guess for the optimal solution $x^4 = -\frac{1}{4}$ and a guess for the optimal objective value $v^4 = -\frac{1}{2} \dots$

Important Observations

- At each iteration, we minimize a lower bound approximation to the objective function f
- Using f^n to denote the lower bound approximation to the objective function f at iteration n,

$$f^n(x) \leq f^{n+1}(x) \leq f(x)$$

for all $x \in X$

Important Observations

- At iteration n, x^n minimizes the lower bound approximation f^n and

$$v^n = \min_{x \in X} f^n(x) = f^n(x^n)$$

- Note that

$$v^n = \min_{x \in X} f^n(x) \leq \min_{x \in X} f(x) = z^* \leq f(x^n)$$

- If $v^n = f(x^n)$, then $f(x^n) = z^*$ and x^n must be optimal

Cutting Plane Method for Convex Optimization

1. Start with an initial guess for the optimal solution $x^1 \in X$ and set $n = 1$
2. Construct a cutting plane approximation to f at x^n

$$f(x^n) + \nabla f(x^n)[x - x^n] \le f(x)$$

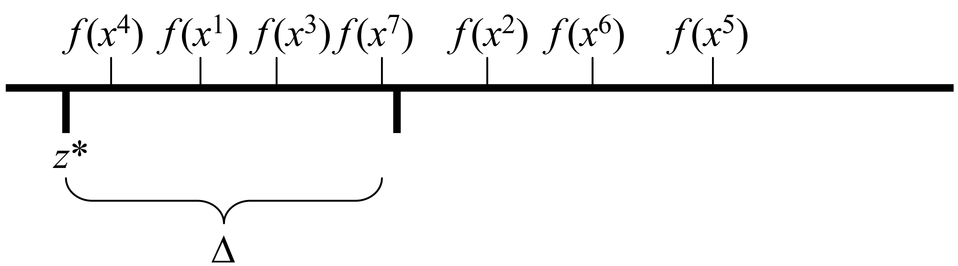
3. Minimize the lower bound approximation by solving

$$\begin{aligned} \min \quad & v \\ \text{subject to} \quad & x \in X \\ & v \ge f(x^i) + \nabla f(x^i)\left[x - x^i\right] \qquad i = 1, \ldots, n \end{aligned}$$

4. Letting (v^{n+1}, x^{n+1}) be an optimal solution, if $v^{n+1} = f(x^{n+1})$, then stop, else increase n by 1 and go to Step 2

Convergence of the Cutting Plane Method

Theorem Letting the sequence of points $\{x^n\}_n$ be generated by the cutting plane method, $\lim_{n\to\infty} f(x^n) = z^*$



- Take the first two points that provide an objective function value that lies to the right of $z^* + \Delta$
- These points are x^2 and x^5

Convergence of the Cutting Plane Method

- Cutting plane approximation that we generate at iteration 2 is

$$f(x^2) + \nabla f(x^2)\,[x - x^2]$$

  and the constraint $v \geq f(x^2) + \nabla f(x^2)\,[x - x^2]$ remains in our lower bound approximation after iteration 2

- At iteration 5, since this constraint is still in our lower bound approximation, it must to be satisfied by the solution (v^5, x^5) obtained at iteration 5 and we have

$$v^5 \geq f(x^2) + \nabla f(x^2)\,[x^5 - x^5]$$

- Since $f(x^5) \geq z^* + \Delta$, we have $f(x^5) - \Delta \geq z^* \geq v^5$

Convergence of the Cutting Plane Method

- Thus,

$$f(x^5) - f(x^2) - \nabla f(x^2)\,[x^5 - x^2] \geq \Delta$$

- Letting $C = \max_{x \in X} \|\nabla f(x)\|$,

$$\|f(x) - f(y)\| \leq C\,\|x - y\|$$

- We obtain

$$\Delta \leq f(x^5) - f(x^2) - \nabla f(x^2)\,[x^5 - x^2] \leq C\,\|x^5 - x^2\| + C\,\|x^5 - x^2\|$$

so that

$$\Delta/2\,C \leq \|x^5 - x^2\|$$

Decomposition Methods for Two-Stage Stochastic Programs

Decomposition Methods for Two-Stage Stochastic Programs

- Assume that there are $K < \infty$ scenarios and the only random component is the right side of the constraints in the second stage
- Use p_k to denote the probability of scenario k and h_k to denote the right side of the constraints in the second stage under scenario k

$$\begin{aligned} \min \quad & c^t x + Q(x) \\ \text{subject to} \quad & Ax = b \\ & x \geq 0, \end{aligned}$$

where the recourse function $Q(x) = \sum_{k=1}^{K} p_k \, Q_k(x)$ is defined as

$$\begin{aligned} Q_k(x) = \min \quad & q^t y \\ \text{subject to} \quad & Wy = h_k - Tx \\ & y \geq 0 \end{aligned}$$

Decomposition Methods for Two-Stage Stochastic Programs

- We can write the two-stage stochastic problem in its deterministic equivalent form

$$
\begin{aligned}
\min \quad & c^t x + \sum_{k=1}^{K} p_k \, q^t \, y_k \\
\text{subject to} \quad & Ax = b \\
& Tx + W y_k = h_k \qquad k = 1, \ldots, K \\
& x \geq 0, \; y_k \geq 0 \qquad k = 1, \ldots, K
\end{aligned}
$$

- This problem has both large number of constraints and large number of decision variables

Convexity of the Recourse Function

- If the recourse function Q is convex and it is tractable to compute its subgradients, then we can use the cutting plane method to solve the two-stage stochastic program

$$\begin{aligned} \min \quad & c^t x + Q(x) \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Convexity of the Recourse Function

- The second stage problem is

$$\begin{aligned} Q_k(x) = \min \quad & q^t y \\ \text{subject to} \quad & W y = h_k - T x \\ & y \geq 0 \end{aligned}$$

- Let $\pi_k(x)$ be an optimal solution to the dual of the second stage problem

$$\begin{aligned} Q_k(x) = \max \quad & [h_k - T\,x]^t\,\pi \\ \text{subject to} \quad & W^t \pi \leq q \end{aligned}$$

- By optimality of $\pi_k(x_0)$, $Q_k(x_0) = [h_k - Tx_0]^t\,\pi_k(x_0)$ and by feasibility, $Q_k(x) \geq [h_k - Tx]^t\,\pi_k(x_0)$

Convexity of the Recourse Function

- The second stage problem is

$$\begin{aligned} Q_k(x) = \min \quad & q^t y \\ \text{subject to} \quad & W y = h_k - T x \\ & y \geq 0 \end{aligned}$$

- Let $\pi_k(x)$ be an optimal solution to the dual of the second stage problem

$$\begin{aligned} Q_k(x_0) = \max \quad & [h_k - T\, x_0]^t\, \pi \\ \text{subject to} \quad & W^t \pi \leq q \end{aligned}$$

- By optimality of $\pi_k(x_0)$, $Q_k(x_0) = [h_k - T x_0]^t\, \pi_k(x_0)$ and by feasibility, $Q_k(x) \geq [h_k - T x]^t\, \pi_k(x_0)$

Convexity of the Recourse Function

- We obtain the subgradient inequality for Q_k

$$Q_k(x) \geq Q_k(x_0) - \pi_k^t(x_0)\, T\, [x - x_0]$$

- Taking expectations,

$$Q(x) \geq Q(x_0) - \sum_{k=1}^{K} p_k\, \pi_k^t(x_0)\, T\, [x - x_0]$$

- Thus, a subgradient of Q at point x_0 is given by

$$g(x_0) = - \sum_{k=1}^{K} p_k\, \pi_k^t(x_0)\, T$$

Decomposition Method for Two-Stage Stochastic Programs

1. Start with an initial guess for the optimal solution x^1 and set $n = 1$
2. Construct a cutting plane approximation to Q at x^n

$$Q(x^n) + g(x^n)\,[x - x^n]$$

3. Minimize the lower bound approximation by solving

$$\begin{aligned} \min \quad & c^t\, x + v \\ \text{subject to} \quad & A\,x = b,\ x \geq 0 \\ & v \geq Q(x^i) + g(x^i)\,[x - x^i] \qquad i = 1, \ldots, n \end{aligned}$$

4. Letting (v^{n+1}, x^{n+1}) be an optimal solution, if $v^{n+1} = Q(x^{n+1})$, then stop, else increase n by 1 and go to Step 2

Dealing with Infeasibility

- Assume that we are at iteration n with the solution x^n and there exists a scenario k under which the second stage problem is infeasible

$$
\begin{aligned}
Q_k(x^n) = \min \quad & q^t y \\
\text{subject to} \quad & Wy = h_k - Tx^n \\
& y \geq 0
\end{aligned}
$$

- We can detect this infeasibility by solving

$$
\begin{aligned}
U_k(x^n) = \min \quad & e^t z_+ + e^t z_- \\
\text{subject to} \quad & Wy + z_+ - z_- = h_k - Tx^n \\
& y \geq 0, z_+ \geq 0, z_- \geq 0
\end{aligned}
$$

and observing that $U_k(x^n) > 0$

Dealing with Infeasibility

- Dual of the infeasibility detection problem is

$$\begin{aligned} U_k(x^n) = \max \quad & [h_k - Tx^n]^t \eta \\ \text{subject to} \quad & W^t \eta \le 0 \\ & -1 \le \eta \le 1 \end{aligned}$$

- Let $\eta_k(x^n)$ be the solution to the dual of the infeasibility detection problem and consider the constraint

$$[h_k - T\,x]^t\, \eta_k(x^n) \le 0$$

- This constraint is not satisfied by the problematic point x^n since

$$[h_k - T\,x^n]^t\, \eta_k(x^n) = U_k(x^n) > 0$$

- This constraint is satisfied by any nonproblematic point x since

$$[h_k - T\,x]^t\, \eta_k(x^n) \le U_k(x) = 0$$

Dealing with Infeasibility

- Dual of the infeasibility detection problem is

$$\begin{aligned} U_k(x) = \max \quad & [h_k - Tx]^t \eta \\ \text{subject to} \quad & W^t \eta \leq 0 \\ & -1 \leq \eta \leq 1 \end{aligned}$$

- Let $\eta_k(x^n)$ be the solution to the dual of the infeasibility detection problem and consider the constraint

$$[h_k - T\,x]^t\,\eta_k(x^n) \leq 0$$

- This constraint is not satisfied by the problematic point x^n since

$$[h_k - T\,x^n]^t\,\eta_k(x^n) = U_k(x^n) > 0$$

- This constraint is satisfied by any nonproblematic point x since

$$[h_k - T\,x]^t\,\eta_k(x^n) \leq U_k(x) = 0$$

Decomposition Method for Two-Stage Stochastic Programs

1. Start with an initial guess for the optimal solution x^1, set $n = 1$, $N^O = \varnothing$ and $N_k^F = \varnothing$ for all $k = 1, \ldots, K$

2. With the solution x^n, solve the second stage problem

$$\begin{aligned} Q_k(x^n) = \min \quad & q^t y \\ \text{subject to} \quad & W y = h_k - T x^n \\ & y \geq 0 \end{aligned}$$

   for all $k = 1, \ldots, K$

3. If there is a scenario k under which the second stage problem is infeasible, then construct the constraint

$$[h_k - T\, x]^t\, \eta_k(x^n) \leq 0$$

   and set $N_k^F \leftarrow N_k^F \cup \{n\}$

4. If the second stage problem is feasible for all scenarios, then construct a cutting plane approximation to Q at x^n

$$Q(x^n) + g(x^n)\,[x - x^n]$$

and set $N^O \leftarrow N^O \cup \{n\}$

5. Minimize the lower bound approximation by solving

$$\begin{aligned} \min \quad & c^t\, x + v \\ \text{subject to} \quad & A\, x = b,\ x \geq 0 \\ & v \geq Q(x^i) + g(x^i)\,[x - x^i] \qquad i \in N^O \\ & 0 \geq [h_k - T\, x]^t\, \eta_k(x^i) \qquad i \in N_k^F,\ k = 1, \ldots, K \end{aligned}$$

6. Letting (v^{n+1}, x^{n+1}) be an optimal solution, if $v^{n+1} = Q(x^{n+1})$, then stop, else increase n by 1 and go to Step 2

Dual Methods for Two-Stage Stochastic Programs

- Consider the deterministic equivalent form

$$
\begin{aligned}
z^* = \min \quad & c^t x + \sum_{k=1}^{K} p_k \, q^t \, y_k \\
\text{subject to} \quad & Ax = b \\
& Tx + W y_k = h_k \qquad k = 1, \ldots, K \\
& x \geq 0, \; y_k \geq 0 \qquad k = 1, \ldots, K
\end{aligned}
$$

- Write the deterministic equivalent form as

$$
\begin{aligned}
z^* = \min \quad & c^t x_0 + \sum_{k=1}^{K} p_k \, q^t \, y_k \\
\text{subject to} \quad & A x_k = b \qquad k = 1, \ldots, K \\
& T x_k + W y_k = h_k \qquad k = 1, \ldots, K \\
& x_k - x_0 = 0 \qquad k = 1, \ldots, K \qquad (\lambda_k) \\
& x_k \geq 0, \; y_k \geq 0 \qquad k = 1, \ldots, K
\end{aligned}
$$

- Relax the constraints that link the scenarios by associating the Lagrange multipliers $\lambda = (\lambda_1, \ldots, \lambda_K)$ with them

$$
\begin{aligned}
L(\lambda) = \min \quad & \left[c^t - \sum_{k=1}^{K} \lambda_k^t \right] x_0 + \sum_{k=1}^{K} \lambda_k^t \, x_k + \sum_{k=1}^{K} p_k \, q^t \, y_k \\
\text{subject to} \quad & A x_k = b \qquad k = 1, \ldots K \\
& T x_k + W y_k = h_k \qquad k = 1, \ldots, K \\
& x_k \geq 0, \; y_k \geq 0 \qquad k = 1, \ldots, K
\end{aligned}
$$

- Relaxed problem decomposes by the scenarios and it can be solved in a tractable fashion by solving one subproblem for each scenario

Dual Methods for Two-Stage Stochastic Programs

- For any λ, $L(\lambda) \leq z^*$
- To obtain the tightest possible lower bound on z^*, solve

$$\max_{\lambda} L(\lambda) \leq z^*$$

- The tightest possible lower bound satisfies

$$\max_{\lambda} L(\lambda) = z^*$$

- When viewed as a function of the Lagrange multipliers, the optimal nonobjective of the relaxed problem is concave in λ
- We can use the cutting plane method to solve the problem

$$\max_{\lambda} L(\lambda)$$

Limitations and Extensions

- Obtaining a subgradient of the recourse function requires solving the second stage problem for all scenarios
  - Stochastic decomposition and cutting plane and partial sampling methods allow solving the second stage problem for only a subset of the scenarios
- Cutting plane methods do not take advantage of a good initial solution
  - Regularized decomposition and trust region methods try to make use of a good initial solution by limiting how much we move towards a promising point
- Cutting plane methods get within the vicinity of the optimal solution quite fast, but can take a lot of iterations to get to the optimal solution

Further Reading

Introduction to Stochastic Programming, J. R. Birge, F. Louveaux

Numerical Techniques for Stochastic Optimization, edited by Yu. Ermoliev, R. J-B Wets

Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming, J. L. Higle, S. Sen

Stochastic Linear Programming: Models, Theory, and Computation, P. Kall, J. Mayer

Stochastic Programming, P. Kall and S. W. Wallace

Handbooks in Operations Research and Management Science: Stochastic Programming, edited by A. Ruszczynski, A. Shapiro

Lectures on Stochastic Programming: Modeling and Theory, A. Shapiro, D. Dentcheva, A. Ruszczynski

Applications of Stochastic Programming, edited by S. W. Wallace, W. T. Ziemba

Further Reading

Nonlinear Programming, D. P. Bertsekas

Convex Analysis and Optimization, D. Bertsekas, A. Nedic, A. Ozdaglar

Convex Analysis, R. T. Rockafellar

Nonlinear Optimization, A. Ruszczynski

(Please excuse any involuntary omissions)