Introduction to Stochastic Optimization in Supply Chain and Logistic Optimization

John R. Birge
Northwestern University

Outline

• Overview
• Part I - Models
  • Vehicle allocation (integer linear)
  • Financial plans (continuous nonlinear)
  • Manufacturing and real options (integer nonlinear)
• Part II – Optimization Methods
Overview

• Stochastic optimization
  • Traditional
    • Small problems
    • Impractical
  • Current
    • Integrate with large-scale optimization (stochastic programming)
    • Practical examples
    • Expanding rapidly
    • Integration of financial and operation considerations

Vehicle Allocation

• Decision:
  • How to position empty freight cars?

NOW:      DAY 1:      DAY 2:

5 cars A ? A 2 A

0 cars B ? B 2 B

DEMAND:  DAY 1: B to A: Mean Value=2
          DAY 1: A to B: Mean Value=2
Vehicle Allocation: Mean Value Solution

Parameters: COST: 0.5 per empty car from A to B  
REVENUE: 1.5 per full car from B to A, 1 from A to B

• **Maximize**: Revenue-Cost
  
  » MOVE TWO EMPTY CARS FROM A to B

NOW:

<table>
<thead>
<tr>
<th></th>
<th>Day 1:</th>
<th>Day 2:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

5 cars: A  
0 cars: B

RESULT: Net 2: A to B; Net 2: B to A  
**TOTAL(MV) = 4**

Expectation of Mean Value

Suppose: Demand is Random (Expectation from A to B=2)

• 0 from A to B with prob. 1/3  
• 3 from A to B with prob. 2/3

• **Find**: Expected (Revenue-Cost)
  
  » MOVE Two EMPTY CARS FROM A to B

NOW:

<table>
<thead>
<tr>
<th></th>
<th>Day 1:</th>
<th>Day 2:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

5 cars: A  
0 cars: B

Expected Value:

Net 2: A to B;  
Net 2: B to A (w.p. 2/3)  
-1: B to A (w.p. 1/3)

**TOTAL (EMV): 3**
**Stochastic Program Solution**

Suppose: Demand is Random (as before)
GOAL: A solution to obtain highest expected value

- **Maximize: Expected (Revenue-Cost)**

```
Suppose: Demand is Random (as before)
GOAL: A solution to obtain highest expected value

• Maximize: Expected (Revenue-Cost)

NOW:
A
5 cars

MOVE Three EMPTY CARS FROM A to B
DAY 1:
A
2

DAY 2:
A
1

Expected Value:
Net 2: A to B;
Net 3: B to A (w.p. 2/3)
-1.5 : B to A (w.p. 1/3)
TOTAL (RP): 3.5
RP=Recourse Problem

Suppose: Demand is Random (as before)
GOAL: A solution to obtain highest expected value

• Maximize: Expected (Revenue-Cost)

NOW:
A
5 cars

MOVE Three EMPTY CARS FROM A to B
DAY 1:
A
2

DAY 2:
A
1

Expected Value:
Net 2: A to B;
Net 3: B to A (w.p. 2/3)
-1.5 : B to A (w.p. 1/3)
TOTAL (RP): 3.5
RP=Recourse Problem
```

**INFORMATION and MODEL VALUE**

• INFORMATION VALUE:
  • FIND Expected Value with Perfect Information or Wait-and-See (WS) solution:
    • Know demand: if 3, send 3 from A to B; If 0, send 0 from A to B:
      • Earn: 2 (AtoB) + (2/3) (3) + (1/3)0= 4 = WS
      • Expected Value of Perfect Information (EVPI):
        • EVPI = WS - RP = 4 - 3.5 = 0.5
    • Value of knowing future demand precisely
  • MODEL VALUE:
    • FIND EMV, RP
    • Value of the Stochastic Solution (VSS):
      • VSS = RP - EMV=3.5 - 3 = 0.5
      • Value of using the correct optimization model
INFORMATION/MODEL OBSERVATIONS

- EVPI and VSS:
  - ALWAYS >= 0 (WS >= RP >= EMV)
  - OFTEN DIFFERENT (WS=RP but RP > EMV and vice versa)
- FIT CIRCUMSTANCES:
  - COST TO GATHER INFORMATION
  - COST TO BUILD MODEL AND SOLVE PROBLEM
- MEAN VALUE PROBLEMS:
  - MV IS OPTIMISTIC (MV=4 BUT EMV=3, RP=3.5)
  - ALWAYS TRUE IF CONVEX AND RANDOM
  - CONSTRAINT PARAMETERS
  - VSS LARGER FOR SKEWED DISTRIBUTIONS/COSTS

STOCHASTIC PROGRAM

- ASSUME: Random demand on AB and BA
- GOAL: maximize expected profits
  - (risk neutral)
- DECISIONS: xij - empty from i to j
  - yij(s) - full from i to j in scenario s (RECOUSE)
    - (prob. p(s))
- FORMULATION:
  \[
  \text{Max} \ -0.5 \times AB + \sum_{s=1}^{s=2} p(s) \ (1.5 \times yAB(s) + 1.5 \times yBA(s))
  \]
  \[
  \text{s.t.} \quad x_{AB} + x_{AA} = 5 \quad \text{(Initial)}
  \]
  \[
  -x_{AB} + y_{BA}(s) \leq 0 \quad \text{(Limit BA)}
  \]
  \[
  -x_{AA} + y_{AB}(s) \leq 0 \quad \text{(Limit AB)}
  \]
  \[
  y_{BA}(s) \leq D_{BA}(s) \quad \text{(Demand BA)}
  \]
  \[
  y_{AB}(s) \leq D_{AB}(s) \quad \text{(Demand AB)}
  \]
  \[
  x_{AA}, x_{AB}, y_{AA}(s), y_{AB}(s) \geq 0
  \]
- EXTENSIONS: Multiple stages; Constraint/objective complexity (Powell et al.)
Financial Planning

- **GOAL:** Accumulate $G$ for tuition $Y$ years from now
- **Assume:**
  - $W(0)$ - initial wealth
  - $K$ - investments
  - concave utility (piecewise linear)

Utility

\[
\begin{array}{c}
\text{G} \\
\text{W(Y)}
\end{array}
\]

**RANDOMNESS:** returns $r(k,t)$ - for $k$ in period $t$

where $Y \rightarrow T$ decision periods

FORMULATION

- **SCENARIOS:** $\sigma \in \Sigma$
  - Probability, $p(\sigma)$
  - Groups, $S^1, ..., S^t$, at $t$
- **MULTISTAGE STOCHASTIC NLP FORM:**

\[
\begin{align*}
\max & \quad \sum_{\sigma} p(\sigma) \left( U(W(\sigma), T) \right) \\
\text{s.t.} \quad & (\text{for all } \sigma): \sum_{\sigma} x(k,1, \sigma) = W(\sigma) \text{ (initial)} \\
& \sum_{\sigma} r(k,t-1, \sigma) x(k,t-1, \sigma) - \sum_{\sigma} x(k,t, \sigma) = 0, \text{ all } t > 1; \\
& \sum_{\sigma} r(k,T-1, \sigma) x(k,T-1, \sigma) - W(\sigma, T) = 0, \text{ (final)}; \\
& x(k,t, \sigma) \geq 0, \text{ all } k,t;
\end{align*}
\]

Nonanticipativity:

\[
x(k,t, \sigma') - x(k,t, \sigma) = 0 \text{ if } \sigma', \sigma \in S^i \text{ for all } t, i, \sigma', \sigma
\]

This says decision cannot depend on future.
**DATA and SOLUTIONS**

- **ASSUME:**
  - \( Y = 15 \) years
  - \( G = 80,000 \)
  - \( T = 3 \) (5 year intervals)
  - \( k = 2 \) (stock/bonds)

- **Returns (5 year):**
  - Scenario A: \( r(\text{stock}) = 1.25 \) \( r(\text{bonds}) = 1.14 \)
  - Scenario B: \( r(\text{stock}) = 1.06 \) \( r(\text{bonds}) = 1.12 \)

- **Solution:**

<table>
<thead>
<tr>
<th>PERIOD</th>
<th>SCENARIO</th>
<th>STOCK</th>
<th>BONDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-8</td>
<td>41.5</td>
<td>13.5</td>
</tr>
<tr>
<td>2</td>
<td>1-4</td>
<td>65.1</td>
<td>2.17</td>
</tr>
<tr>
<td>2</td>
<td>5-8</td>
<td>36.7</td>
<td>22.4</td>
</tr>
<tr>
<td>3</td>
<td>1-2</td>
<td>83.8</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3-4</td>
<td>0</td>
<td>71.4</td>
</tr>
<tr>
<td>3</td>
<td>5-6</td>
<td>0</td>
<td>71.4</td>
</tr>
<tr>
<td>3</td>
<td>7-8</td>
<td>64.0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Manufacturing Planning**

- **Goal:**
  - Decide on coordinated production, distribution capacity and vendor contracts for multiple models in multiple markets (e.g., NA, Eur, LA, Asia)

- **Traditional approach**
  - Forecast demand for each model/market
  - Forecast costs
  - Obtain piece rates and proposals
  - Construct cash flows and discount

  ➡️ **Optimize for a single-point forecast**
  ➡️ **Missing option value of flexible capacity**
Real Options

• Idea: Assets that are not fully used may still have option value (includes contracts, licenses)
• Value may be lost when the option is exercised (e.g., developing a new product, invoking option for second vendor)
• Traditional NPV analyses are flawed by missing the option value
• Missing parts:
  • Value to delay and learn
  • Option to scale and reuse
  • Option to change with demand variation (uncertainty)
  • Not changing discount rates for varying utilizations

Real Option Valuation for Capacity

• Goal: Production value with capacity $K$
  • Compute uncapacitated value based on Capital Asset Pricing Model:
    • $S = e^{r(T-t)}c_T S_T dF(S_T)$
    • where $c_T$=margin,$F$ is distribution (with risk aversion),
    • $r$ is rate from CAPM (with risk aversion)
  • Assume $S_t$ now grows at riskfree rate, $r_f$; evaluate as if risk neutral:
    • Production value = $S_t - C_t = e^{r_f(T-t)}c_T \min(S_T,K) dF_f(S_T)$
    • where $F_f$ is distribution (with risk neutrality)
Generalizations for Other Long-term Decisions

- Model: period $t$ decisions: $x_t$
- START: Eliminate constraints on production
  - Demand uncertainty remains
  - Can value unconstrained revenue with market rate, $r$
    \[ \frac{1}{1+r^t} c_t x_t \]

**IMPLICATIONS OF RISK NEUTRAL HEDGE:**
Can model as if investors are risk neutral
$\Rightarrow$ value grows at riskfree rate, $r_f$

**Future value:** \( \left[ \frac{1}{1+r^t} c_t (1+r_f^t) x_t \right] \)

**BUT:** This new quantity is constrained

---

New Period $t$ Problem: Linear Constraints on Production

- WANT TO FIND (present value):
  \[ 1/ \left(1+r_f^t \right) \max \left[ c_t x_t (1+r_f^t)/(1+r^t) \mid A_t x_t (1+r_f^t)/(1+r^t) \leq b \right] \]

**EQUIVALENT TO:**

\[ 1/ \left(1+r^t \right) \max \left[ c_t x \mid A_t x \leq b (1+r_f^t)/(1+r^t) \right] \]

**MEANING:** To compensate for lower risk with constraints, constraints expand and risky discount is used
Constraint Modification

- FORMER CONSTRAINTS: $A_t x_t \leq b_t$
- NOW: $A_t x_t (1+r_f)^t/(1+r)^t \leq b_t$

EXTREME CASES

All slack constraints:

\[
\frac{1}{(1+r)^t} \max \{ c_t x \mid A_t x \leq b (1+r)^t/(1+r_f)^t \}
\]

becomes equivalent to:

\[
\frac{1}{(1+r)^t} \max \{ c_t x \mid A_t x \leq b \}
\]

i.e. same as if unconstrained - risky rate

NO SLACK:

becomes equivalent to:

\[
\frac{1}{(1+r)^t} \{ c_t x = B^{-1}b (1+r)^t/(1+r_f)^t = c_t B^{-1}b/(1+r_f)^t \}
\]

i.e. same as if deterministic - riskfree rate
Example: Capacity Planning

- What to produce?
- Where to produce? (When?)
- How much to produce?

**EXAMPLE: Models 1, 2, 3; Plants A, B**

Should B also build 2?

Result: Stochastic Linear Programming Model

- Key: Maximize the Added Value with Installed Capacity
  - Must choose best mix of models assigned to plants
  - Maximize Expected Value over $s[\Sigma_i e^{rt} \text{Profit}(i) \cdot \text{Production}(i,t,s) - \text{CapCost}(i, t) \cdot \text{Capacity}(i, t)]$
  - subject to: $\Sigma_j \text{Production}(i, j, t) >= \Sigma_j \text{Production}(i, j, t, s)$
  - $\Sigma_j \text{Production}(i, j, t) <= e^{(r-f)b} \cdot \text{Capacity}(i, t)$
  - $\text{Production}(i, j, t) <= e^{(r-f)b} \cdot \text{Capacity}(i, t)$
  - $\text{Production}(i, j, t) >= 0$
- Need $\text{MaxSales}(i, t, s)$ - random
- $\text{Capacity}(i, j, 0)$ - Decision in First Stage (now)

**NOTE:** Linear model that incorporates risk
Result with Option Approach

- Can include risk attitude in linear model
- Simple adjustment for the uncertainty in demand
- **Requirement 1**: correlation of all demand to market
- **Requirement 2**: assumptions of market completeness

Outline

- Overview
- Part I - Models
  - Vehicle allocation (integer linear)
  - Financial plans (continuous nonlinear)
  - Manufacturing and real options (integer nonlinear)
- Part II – Optimization Methods
General Stochastic Programming Model: Discrete Time

• Find $x=(x_1, x_2, \ldots, x_T)$ and $p$ (unknown distribution) to

\[
\begin{align*}
\text{minimize} & \quad \mathbb{E}_p \left[ \sum_{t=1}^{T} f_t(x_t, x_{t+1}, p) \right] \\
\text{s.t.} & \quad x_t \in X_t, \quad \text{nonanticipative } p \text{ in } P \text{ (distribution class)} \\
& \quad P[ h_t(x_t, x_{t+1}, p_{t+1}) \leq 0 ] \geq a \text{ (chance constraint)}
\end{align*}
\]

General Approaches:
• Simplify distribution (e.g., sample) and form a mathematical program:
  • Solve step-by-step (dynamic program)
  • Solve as single large-scale optimization problem
  • Use iterative procedure of sampling and optimization steps

Simplified Finite Sample Model

• Assume $p$ is fixed and random variables represented by sample $\xi_i^t$ for $t=1,2,\ldots,T$, $i=1,\ldots,N_t$ with probabilities $p^i_t$, $a(i)$ an ancestor of $i$, then model becomes (no chance constraints):

\[
\begin{align*}
\text{minimize} & \quad \sum_{t=1}^{T} \sum_{i=1}^{N_t} p^i_t f_i(x_{a(i)}^t, x_i^{t+1}, \xi_i^t) \\
\text{s.t.} & \quad x_i^t \in X_i^t
\end{align*}
\]

Observations?
• Problems for different $i$ are similar – solving one may help to solve others
• Problems may decompose across $i$ and across $t$ yielding
  • smaller problems (that may scale linearly in size)
  • opportunities for parallel computation.
Outline

• Overview
• Part I - Models
• Part II – Optimization Methods
  • Factorization/sparsity (interior point/barrier)
  • Decomposition
  • Lagrangian methods
• Conclusions

SOLVING AS LARGE-SCALE MATHEMATICAL PROGRAM

• PRINCIPLES:
  • DISCRETIZATION LEADS TO MATHEMATICAL PROGRAM BUT LARGE-SCALE
  • USE STANDARD METHODS BUT EXPLOIT STRUCTURE
• DIRECT METHODS
  • TAKE ADVANTAGE OF SPARSITY STRUCTURE
    • SOME EFFICIENCIES
  • USE SIMILAR SUBPROBLEM STRUCTURE
    • GREATER EFFICIENCY
• SIZE
  • UNLIMITED (INFINITE NUMBERS OF VARIABLES)
  • STILL SOLVABLE (CAUTION ON CLAIMS)
STANDARD APPROACHES

• Sparsity Structure Advantage
  • PARTITIONING
  • BASIS FACTORIZATION
  • INTERIOR POINT FACTORIZATION

• Similar/Small Problem Advantage
  • DP APPROACHES: DECOMPOSITION
    • BENDERS, L-SHAPED (VAN SLYKE – WETS)
    • DANTZIG-WOLFE (PRIMAL VERSION)
    • REGULARIZED (RUSZCZYNSKI)
    • VARIOUS SAMPLING SCHEMES (HIGLE/SEN Stochastic Decomposition, Abridge Nested Decomposition)

• LAGRANGIAN METHODS

Sparsity Methods: Stochastic Linear Program Example

• Two-stage Linear Model:
  \[ X_1 = \{ x_1 | A x_1 = b, x_1 \geq 0 \} \]
  \[ f_0(x_0,x_1) = c x_1 \]
  \[ f_1( x_1, x_2^i, \xi_2^i ) = q x_2^i \text{ if } T x_1 + W x_2^i = \xi_2^i, \]
  \[ x_2^i \geq 0; + \infty \text{ otherwise} \]

• Result:
  \[ \text{min } c x_1 + \sum_{i=1}^{N1} p_2^i q x_2^i \]
  s. t. \[ A x_1 = b, x_1 \geq 0 \]
  \[ T x_1 + W x_2^i = \xi_2^i, x_2^i \geq 0 \]
LP-BASED METHODS

• USING BASIS STRUCTURE

\[
\begin{align*}
A & \quad W \\
T & \quad W \\
T & \quad W \\
T & \quad W
\end{align*}
\]

PERIOD 1 \quad PERIOD 2

INTERIOR POINT MATRIX STRUCTURE

\[
A'D^2A'T = A'
\]

COMPLETE FILL-IN

• MODEST GAINS FOR SIMPLEX

ALTERNATIVES FOR INTERIOR POINTS

• VARIABLE SPLITTING (MULVEY ET AL.)

• PUT IN EXPLICIT NONANTICIPATIVITY CONSTRAINTS

\[
\begin{align*}
\text{NEW} & \quad \text{NEW} \\
& \quad \text{NEW}
\end{align*}
\]

\[
= A'
\]

•RESULT

•REDUCED FILL-IN BUT LARGER MATRIX
OTHER INTERIOR POINT APPROACHES

• USE OF DUAL FACTORIZATION OR MODIFIED SCHUR COMPLEMENT

\[ A^T D^2 A' = \]

RESULTS:
• SPEEDUPS OF 2 TO 20
• SOME INSTABILITY ⇔ INDEFINITE SYSTEM (VANDERBEI ET AL., CZYZYK ET AL.)

Outline

• Overview
• Part I - Models
• Part II – Optimization Methods
  • Factorization/sparsity (interior point/barrier)
  • Decomposition
  • Lagrangian methods
• Conclusions
SIMILAR/SMALL PROBLEM STRUCTURE: DYNAMIC PROGRAMMING VIEW

- **STAGES**: \( t=1,\ldots,T \)
- **STATES**: \( x_t \rightarrow B_t x_t \) (or other transformation)
- **VALUE FUNCTION**:
  \[
  Q_t(x_t) = E[Q_{t+1}(x_{t+1})] \]
  where \( \xi_t \) is the random element and
  \[
  Q_t(x_t, \xi_t) = \min f_t(x_t, x_{t+1}, \xi_t) + Q_{t+1}(x_{t+1})
  \]
  subject to \( x_{t+1} \in X_{t+1}(\xi_t) \) \( x_t \) given
- **SOLVE**: iterate from \( T \) to \( 1 \)

LINEAR MODEL STRUCTURE

Stage 1 \quad Stage 2 \quad Stage 3

\[
\begin{align*}
\min & \quad c_1 x_1 + Q_2(x_2) \\
\text{s.t.} & \quad W_1 x_1 = h_1 \\
& \quad x_1 \geq 0 \\
Q_t(x_{t-1, a(k)}) &= \sum_{\xi, \xi\in \Xi_t} \text{prob} \left( \xi_{t, k} \mid Q_{t+1}(x_{t+1}) \right) \\
Q_{t, a(k)} &\left( x_{t-1, a(k)}, \xi_{t, k} \right) = \min \frac{c_t \left( \xi_{t, k} \right) x_{t, k} + Q_{t+1} \left( x_{t+1} \right) }{x_{t, k}} \\
\text{s.t.} & \quad W_{t, a(k)} x_{t, k} = h_t \left( \xi_{t, k} \right) - T_{t, a(k)} \left( \xi_{t, k} \right) x_{t-1, a(k)} \\
& \quad x_{t, k} \geq 0
\end{align*}
\]

- \( Q_{N+1}(x_N) = 0 \), for all \( x_N \),
- \( Q_{t, a(k)}(x_{t-1, a(k)}) \) is a piecewise linear, convex function of \( x_{t-1, a(k)} \)
DECOMPOSITION METHODS

- BENDERS IDEA
  - FORM AN OUTER LINEARIZATION OF $Q_t$
    - ADD CUTS ON FUNCTION:
      - Feasible region
      - (feasibility cuts)
    - LINEARIZATION AT ITERATION $k$
    - NEW CUT (OPTIMALITY CUT)
    - MIN AT $k : < Q_t$

- USE AT EACH STAGE TO APPROX. VALUE FUNCTION
- ITERATE BETWEEN STAGES UNTIL ALL MIN = $Q_t$

Nested Decomposition

- In each subproblem, replace expected recourse function $Q_{t,k}(x_{t-1,a(t)})$ with unrestricted variable $\theta_{t,k}$
  - Forward Pass:
    - Starting at the root node and proceeding forward through the scenario tree, solve each node subproblem
      - $\hat{Q}_{t,k}(x_{t-1,a(t)}, \xi_{t,k}) = \min_{x,k} c_j(\xi_{t,k})x_{t,k} + \theta_{t,k}$
        - s.t. $W_{t,k}x_{t,k} = h_j(\xi_{t,k}) - T_{t-1,1}(\xi_{t,k})x_{t-1,a(t)}$ (optimality cuts)
        - $E_{t,k}x_{t,k} + \theta_{t,k} \geq e_{t,k}$ (feasibility cuts)
        - $D_{t,k}x_{t,k} \geq d_{t,k}$
    - Add feasibility cuts as infeasibilities arise
  - Backward Pass
    - Starting in top node of Stage $t = N-1$, use optimal dual values in descendant Stage $t+1$ nodes to construct new optimality cut. Repeat for all nodes in Stage $t$, resolve all Stage $t$ nodes, then $t \rightarrow t-1$.
  - Convergence achieved when $\theta_t = Q_t(x_t)$
SAMPLE RESULTS

- SCAGR7 PROBLEM SET

  ![Graph showing CPU vs. Variables for Standard LP and Nested Decomposition]

  **Log (CPU)**
  **Log (No. of Variables)**
  
  **Parallel:** 60-80% Efficiency in Speedup

  **Other Problems:** Similar Results
  - Only < Order of Magnitude Speedup with Storm
    - Two-stages, Little Commonality in Subproblems
    - Still able to solve Order of Magnitude Larger Problems

Decomposition Enhancements

- Optimal basis repetition
  - Take advantage of having solved one problem to solve others
  - Use *bunching* to solve multiple problems from root basis
  - *Share* bases across levels of the scenario tree
  - Use solution of single scenario as *hot start*

- Multicuts
  - Create cuts for each descendant scenario

- Regularization
  - Add quadratic term to keep close to previous solution

- Sampling
  - Stochastic decomposition (Higle/Sen)
  - Importance sampling (Infanger/Dantzig/Glynn)
  - Multistage (Pereira/Pinto, Abridged ND)
Pereira-Pinto Method

- Incorporates sampling into the general framework of the Nested Decomposition algorithm
- Assumptions:
  - relatively complete recourse
  - no feasibility cuts needed
  - serial independence
    - an optimality cut generated for any Stage $t$ node is valid for all Stage $t$ nodes
- Successfully applied to multistage stochastic water resource problems

1. Randomly select $H N$-Stage scenarios
2. Starting at the root, a forward pass is made through the sampled portion of the scenario tree (solving ND subproblems)
3. A statistical estimate of the first stage objective value is calculated using the total objective value obtained in each sampled scenario
   - the algorithm terminates if current first stage objective value $c_j x_j + \theta_j$ is within a specified confidence interval of
4. Starting in sampled node of Stage $t = N - 1$, solve all Stage $t + 1$ descendant nodes and construct new optimality cut.
   - Repeat for all sampled nodes in Stage $t$, then repeat for $t = t - 1$
Pereira-Pinto Method

• Advantages
  • significantly reduces computation by eliminating a large portion of the scenario tree in the forward pass

• Disadvantages
  • requires a complete backward pass on all sampled scenarios
    • not well designed for bushier scenario trees

Abridged Nested Decomposition

• Also incorporates sampling into the general framework of Nested Decomposition
• Also assumes relatively complete recourse and serial independence
• Samples both the subproblems to solve and the solutions to continue from in the forward pass
Abridged Nested Decomposition

Forward Pass
1. Solve root node subproblem
2. Sample Stage 2 subproblems and solve selected subset
3. Sample Stage 2 subproblem solutions and branch in Stage 3 only from selected subset (i.e., nodes 1 and 2)
4. For each selected Stage t-1 subproblem solution, sample Stage t subproblems and solve selected subset
5. Sample Stage t subproblem solutions and branch in Stage t+1 only from selected subset

Convergence Test
1. Randomly select \( H \) N-Stage scenarios. For each sampled scenario, solve subproblems from root to leaf to obtain total objective value for scenario
2. Calculate statistical estimate of the first stage objective value \( \bar{\xi} \)
   • algorithm terminates if current first stage objective value \( c_j x_j + \theta_j \) is within a specified confidence interval of \( \bar{\xi} \) else, a new forward pass begins

IMA Tutorial, Stochastic Optimization, September 2002

Backward Pass
1. Starting in first branching node of Stage \( t = N-1 \), solve all Stage \( t+1 \) descendant nodes and construct new optimality cut for all stage \( t \) subproblems. Repeat for all sampled nodes in Stage \( t \), then repeat for \( t = t - 1 \)

IMA Tutorial, Stochastic Optimization, September 2002
Sample Computational Results

• Test Problems
  • Dynamic Vehicle Allocation (DVA) problems of various sizes
    • set of homogeneous vehicles move full loads between set of sites
    • vehicles can move empty or loaded, remain stationary
    • demand to move load between two sites is stochastic
  • DVA_{x,y,z}
    • x: number of sites (8, 12, 16)
    • y: number of stages (4, 5)
    • z: number of distinct realizations per stage (30, 45, 60, 75)
  • largest problem has > 30 million scenarios

Computational Results (DVA.8)
Outline

• Overview
• Part I - Models
• Part II – Optimization Methods
  • Factorization/sparsity (interior point/barrier)
  • Decomposition
  • Lagrangian methods
• Conclusions

Lagrangian-based Approaches

• General idea:
  • Relax nonanticipativity
  • Place in objective
  • Separable problems

\[
\begin{align*}
\text{MIN} \quad & E \left[ \sum_{t=1}^{T} f_t(x_t, x_{t+1}) \right] \\
\text{s.t.} \quad & x_t \in X_t \\
\quad & x_t \text{ nonanticipative}
\end{align*}
\]

\[
\begin{align*}
\text{MIN} \quad & E \left[ \sum_{t=1}^{T} f_t(x_t, x_{t+1}) \right] \\
\quad & x_t \in X_t \\
\quad & x_t \text{ nonanticipative}
\end{align*}
\]

\[
E[wx] + r/2||x-x||^2
\]

Update: \( w_i \); Project: \( x \) into \( N \) - nonanticipative space as \( x \)

Convergence: Convex problems - Progressive Hedging Alg. (Rockafellar and Wets)
Advantage: Maintain problem structure (networks)
Lagrangian Methods and Integer Variables

- **Idea:** Lagrangian dual provides bound for primal but
  - Duality gap
  - PHA may not converge
- **Alternative:** standard augmented Lagrangian
  - Convergence to dual solution
  - Less separability
  - May obtain simplified set for branching to integer solutions
- **Problem structure:** Power generation problems
  - Especially efficient on parallel processors
  - Decreasing duality gap in number of generation units

Outline

- **Overview**
- **Part I - Models**
- **Part II – Optimization Methods**
  - Factorization/sparsity (interior point/barrier)
  - Decomposition
  - Lagrangian methods
- **Conclusions**
SOME OPEN ISSUES

• MODELS
  • IMPACT ON METHODS
  • RELATION TO OTHER AREAS
• APPROXIMATIONS
  • USE WITH SAMPLING METHODS
  • COMPUTATION CONSTRAINED BOUNDS
  • SOLUTION BOUNDS
• SOLUTION METHODS
  • EXPLOIT SPECIFIC STRUCTURE
  • MASSIVELY PARALLEL ARCHITECTURES
  • LINKS TO APPROXIMATIONS

CRITICISMS

• UNKNOWN COSTS OR DISTRIBUTIONS
  • FIND ALL AVAILABLE INFORMATION
  • CAN CONSTRUCT BOUNDS OVER ALL DISTRIBUTIONS
    • FITTING THE INFORMATION
  • STILL HAVE KNOWN ERRORS BUT ALTERNATIVE SOLUTIONS
• COMPUTATIONAL DIFFICULTY
  • FIT MODEL TO SOLUTION ABILITY
  • SIZE OF PROBLEMS INCREASING RAPIDLY
View Ahead

• New Trends
  • Methods for integer variables
    • Capacity, suppliers, contracts
    • Vehicle routing
  • Integrating simulation
    • Sampling with optimization
    • On-line optimization
    • Low-discrepancy methods

More Trends

• Modeling languages
  • Ability to build stochastic programs directly
  • Integrating across systems
• Using application structure
  • Separation of problem (dimension reduction)
  • Network properties
  • Generalized versions of convexity
Summary

- Increasing application base
- Value for solving the stochastic problem
- Efficient implementations
- Opportunities for new results