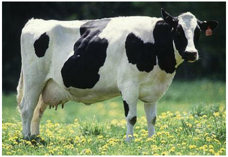
Models and Algorithms for Stochastic Programming

Jeff Linderoth

Dept. of Industrial and Systems Engineering
Univ. of Wisconsin-Madison
linderoth@wisc.edu



Enterprise-Wide Optimization Meeting
Carnegie-Mellon University
March 10th, 2009

Mission Impossible

- Explaining Stochastic Programming in 90 mins
- I will try to give an overview – please interrupt with questions!

What I'll Ramble On

Models

- How to deal with uncertainty
- Why modeling uncertainty is important
- Who has used stochastic programming?
- Why more people don't use stochastic programming

Algorithms

- Extensive Form
- Benders Decomposition (2-stage)
- Sampling
- Nested Benders Decomposition (multistage)

Etymology

- program:
  - **(3)** An ordered list of events to take place or procedures to be followed; a schedule
    - Late Latin programma, public notice, from Greek programma, programmat-, from prographein, to write publicly
- stochastic:
  - **(1b)** Involving chance or probability
    - Greek stokhastikos, from stokhasts, diviner, from stokhazesthai, to guess at, from stokhos, aim, goal.
- Source: The American Heritage Dictionary of the English Language, Fourth Edition.

Sources of Uncertainty

Houston, we have uncertainty!

> *What we anticipate seldom occurs; what we least expected generally happens.*
>
> Benjamin Disraeli (1804 - 1881)

- Financial
  - Market price movements
  - Defaults by a business partner
- Operational
  - Customer demands,
  - Travel times
- Technology related
  - Will a new technology be ready "in time"
- Market Related
  - Shifts in tastes
- Competition
  - What will your competitors strategy be next year?
- Acts of God: Jeff's travel experience yesterday!!!
  - Weather
  - Equipment failure
  - Birds flying into planes

Stochastic Programming

- A tool used in *planning under uncertainty*
- More specifically: Mathematical Programming, or Optimization, in which some of the parameters defining a problem instance are random, or uncertain

---

Optimization

$$\min_{x \in X} f(x)$$

- x: Variables you control

Stochastic Optimization

$$\min_{x \in X(\omega)} F(x, \omega)$$

- ω: Variables you don't control

Stochastic Optimization is UNDEFINED

- You can't possibly choose an x that optimizes for all ω
- More specifiation is required

Jeff's Stochastic Programming Assumptions

- In stochastic programming, we assume that a probability distribution for the uncertainty ω is known or can be approximated.
- We also assume that probabilities are independent of the decisions that are taken.

Decision-dependent uncertainty

- Decisions influence probability distributions
- Decisions influence knowledge discovery

Talk to Ignacio!

Want to know about stochastic programming with decision-dependent uncertainty?

Probability Theory(?)

- This notion of having to know a probability distribution for the randomness is troubling, since in reality, very few people exactly know that
  - Their customer demands follow a log-normal distribution with mean 17.26 and variance 2.88726
  - Their plant will have forced shutdowns following a Weibull distribution with parameters (100.25, 73.7916)
- Instead, you *might* be able to
  - Estimate distributions from historial data (be careful!)
  - Have "qualitative" probability measures ("low/medium/high")
  - Create your own scenarios of interest

The Journey is the Reward?

- Business process people can argue/discuss amongst themselves what the various scenarios might be and the outcomes of those scenarios.
- This process by itself can be very useful

- There is a good amount of frightening-looking mathematical theory and computational evidence that solutions obtained from stochastic programs are often quite "stable" with respect to changes in the input probability distribution

---

The Upshot

- It doesn't matter "too much" if your numbers aren't quite right
- The insights you gain from considering the uncertainty can still be valuable

A Concrete Example: An Uncertain LP

$$\begin{aligned} \min \quad & cx \\ \text{s.t.} \quad & Ax \geq b \\ & T(\omega)x \geq h(\omega) \\ & x \geq 0 \end{aligned}$$

- $T(\omega)$ and $h(\omega)$ are uncertain: $X(\omega) = \{x \mid Ax \geq b, T(\omega)x \geq h(\omega)\}$
- We must choose x despite this uncertainty
- Examples:
  - Decide production quantities before knowing demands
  - Constraint data includes imprecise measurements

Three Approaches

1. Robust optimization
2. Chance-constrained programming
3. Recourse-based stochastic programming

Robust Optimization

- Uncertain data is assumed to lie in an uncertainty set

$$(T(\omega), h(\omega)) \in \mathcal{U}$$

- Guarantee that constraints be satisfied for all possible realizations

$$\begin{aligned} \min \quad & cx \\ \text{s.t.} \quad & Ax \geq b \\ & Tx \geq h \quad \forall (T, h) \in \mathcal{U} \\ & x \geq 0 \end{aligned}$$

- Tractability depends on structure of $\mathcal{U}$

Robust Optimization

- To control conservatism, uncertainty set can be parameterized by a budget of uncertainty
- Example 1: $T_{ij}(\omega) \in [l_{ij}, u_{ij}]$ (Bertsimas and Sim)
    - At most K of the components in each row can differ from the nominal value
    - Nature can choose which K will differ
    - K large $\Rightarrow$ highly conservative (Soyster)
    - $K = 0 \Rightarrow$ No robustness
    - Can formulate this problem as a linear program
- Example 2: $\mathcal{U}$ is ellipsoidal (Ben-Tal and Nemirovski)

Robust Optimization

Advantages:

- Computationally tractable
- Can yield extremely reliable solutions
- Does not require stochastic model

Disadvantages:

- Does not use a stochastic model
- Although conservatism can be controlled, the control parameter doesn't have meaning to decision makers

Stochastic Programming

Assume uncertain data are random variables with *known* distributions

Two approaches to uncertain constraints:

1. Require constraint to be satisfied with high probability

$$\min\{cx : x \in X, P\{T(\omega)x \geq h(\omega)\} \geq 1 - \epsilon\}$$

   ϵ is a parameter, e.g. $\epsilon = 0.05$ or $\epsilon = 0.01$
   Linear program with probabilistic (chance) constraints

2. Penalize violations of constraints

$$\min\left\{cx + \mathbb{E}[\lambda(h(\omega) - T(\omega)x)^{+}] : x \in X\right\}$$

   Special case of a Two stage stochastic program

Linear Programs with Probabilistic Constraints

- Individual constraints:

$$\min \left\{ cx : x \in X, P\{T(\omega)^i x \geq h(\omega)_i\} \geq 1 - \epsilon_i \ \forall i \right\}$$

- Joint constraints:

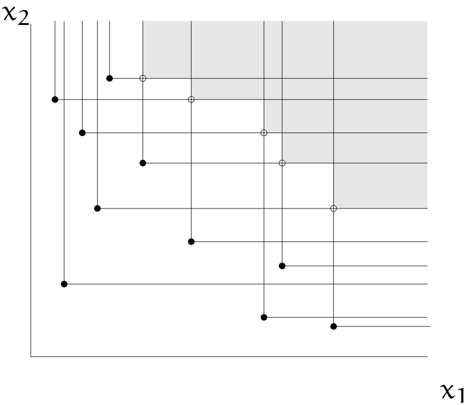
$$\min \{cx : x \in X, P\{T(\omega)x \geq h(\omega)\} \geq 1 - \epsilon\}$$

- Bad news: calculating probability is hard
- Worse news: probabilistic constraints are generally non-convex!

Non-convexity of the feasible region

Consider: $\mathbb{P}\{x_1 \geq \xi_1, x_2 \geq \xi_2\} \geq 0.6$

Each dot: a realization of ξ which occurs with probability 1/10



Two Stage Stochastic Programming

$$(SP) \quad \min \left\{ cx + \mathbb{E}[\lambda(h(\omega) - T(\omega)x)^+] : x \in X \right\}$$

Choose $x \Rightarrow$ Observe $(T(\omega), h(\omega)) \Rightarrow$ Pay penalty

- Good news: (SP) is convex
- Bad news: Calculating expectation is hard

Successful Approach: Sample Average Approximation

- Generate $(T(\omega)^1, h(\omega)^1), \ldots, (T(\omega)^N, h(\omega)^N)$ and solve

$$(SP_N) \quad \min \left\{ cx + \sum_{i=1}^{N} \frac{1}{N} \lambda(h(\omega)^i - T(\omega)^i x)^+ : x \in X \right\}$$

- x_N^* is a often a good approximation to true optimal solution
- We'll see (a lot) more later!

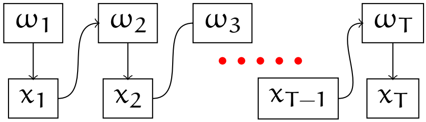
Stochastic Programming vs. Simulation

- Simulation
  - (Pro): Very flexible—System need not be mathematically defined
  - (Pro): Fast
  - (Con): If I run 100 "what-ifs" and get 100 different solutions, how does simulation help me plan for the future?
- Stochastic Programming
  - (Con): More challenging to build and solve models
  - (Pro): SP helps you "optimize" over your "what-ifs".

The Upshot!

Use simulation to generate scenarios. Input the scenarios to a stochastic program to show how to decide how to best hedge against this uncertainty

Multistage Decision Making



- Random vectors $\omega_1 \in \mathbb{R}^{n_1}, \omega_2 \in \mathbb{R}^{n_2}, \ldots, \omega_T \in \mathbb{R}^{n_T}$
- Make sequence of decisions $x_1 \in X_1, x_2 \in X_2, \ldots, x_T \in X_T$.

- The evolution of information is of fundamental importance to the decision-making progress.
- We make a decision now (x_1)
- Nature makes a random decision ω_2: ("stuff" happens)
- We make a second period decision x_2 that attempts to repair the havoc wrought by nature in (recourse).
- Repeat as necessary...
- We make decisions in stages, in between which uncertainty is revealed to us

Hot Off the Presses

- A paperboy (newsvendor) needs to decide how many papers to buy in order to maximize his profit.
- He doesn't know at the beginning of the day how many papers he can sell (his demand).
  - Each newspaper costs c.
  - He can sell each newspaper for a price of q.
  - He can return each unsold newspaper at the end of the day for r. (Obviously $r < c < q$).

The Newsvendor Problem

- Given only knowledge of the probability distribution F of demand, how may papers should the newsvendor buy?

Newsvendor Problem

- Suppose that the newsvendor's goal is to maximize the profits in the long run. (In expectation)...
- Intuitively, it seems that the newsvendor's best strategy is to every purchase the average demand

---

Take Away Message!

- The "optimal" solution is NOT to use the mean demand.
- In fact, the two solution can be far apart. (Depending on the distribution, and parameters r, c, q

Example—The Newsvendor

- $c = 50, q = 70, r = 5$
- Demand: (Truncated) Normal distributed. $\mu = 100, \sigma = 50$
- Mean Value Solution
  - Buy 100. (Duh!)
  - Expect to profit: 2000
  - TRUE long run profit ≈ 650
- Stochastic Solution
  - Buy 75.
  - Expect to profit: 1500
  - TRUE long run profit ≈ 880
- The difference between the two solutions $(880 - 650)$ is called the *value of the stochastic solution*.
  - How much is it worth to you to plan using *full* uncertainty information as opposed to mean-values for the uncertain parameters

A Take Away Message

The "Flaw" of Averages

- The flaw of averages occurs when uncertainties are replaced by single average numbers planning.
  - Did you hear the one about the statistician who drowned fording a river with an average depth of three feet.

Point Estimates

- If you are planning with point estimates for demands, then you are planning sub-optimally
- It doesn't matter how carefully you choose the point estimate – it is impossible to hedge against future uncertainty by considering one realization of the uncertainty in your planning process

Russell-Yasuda Kasai

- Yasuda Kasai: Seventh largest (worldwide) property and casualty insurer.
- Assets of $>$ ¥3.47 trillion
- Liability structure is complex, but want a tool that will allow them to maximize the revenue from these assets in the face of asset management restrictions
- Frank Russell Company hired to develop Asset-Liability Management Model based on (multistage) stochastic programming
- Carino, Myers, Ziemba, Second place in Edelman prize competition of INFORMS.

Asset Allocation Model

- Decisions:
  - Investment amounts for various assets
- Random Events:
  - Return on investment for each asset.
  - Liability payouts
- Constraints:
  - Asset Allocation Constraints (Complex)
  - Loan Model
  - Liability Model

Compared to a performance benchmark established at Yasuda Kasai at the beginning of the Fiscal Year to measure the value added by their use of the model, the new model increased annual income by ￥9.5 billion.

Mr. Kunihiko Sasamoto, Director and Deputy President, Yasuda Kasai.

But Wait There's More!

- Ease of Use
  - Risk is well defined, not using some “abstract” measure like standard deviation
- Improved other systems
  - Other models and IT systems “upgraded” to support new system
- Improved Human Judgement
  - How to think about and incorporate uncertainty into the planning process

Product Portfolio Planning

- Decisions:
  - Invest in various projects (All or nothing investment).
  - Complicated project prerequisite structure
- Random Events:
  - Design-win from customers
  - Technology failures
  - Market forces
- Constraints:
  - Resources
  - Hire-fire costs

(HUGE impact)

Product Portfolio Management at Agere

- We implemented a decision support tool for Agere

1. Optimization Model
2. Simulator of future conditions – (random events were correlated!)

The muckety-mucks loved it!

- They like the ability to talk about the different scenarios.
- Focuses discussion in business planning meetings
- Gives “unbiased” simulator view of potential outcomes of decisions

SP in the Supply Chain

- Decisions:
  - Regular supply chain decision: How much? where? and when?
- Random Elements:
  - Demands, prices, resource capacity.
  - Supply chains going global imply that companies are now more exposed to risky factors such as exchange rates and reliability of transfer channels.
- Constraints:
  - Regular supply chain constraints: Flow balance, material availability, etc.

A Case Study

T. Santoso, S. Ahmed, M. Goetschalckx, and A. Shapiro. "A Stochastic Programming Approach for Supply Chain Network Design under Uncertainty," *European Journal of Operational Research*, vol.167, pp.96-115, 2005.

- Two real supply chains
  - One Domestic (Cardboard packages to breweries and soft drink manufacturers...)
  - One global
- Sizes: Around 100 facilities. Around 100 customers,
- In general, the (sampled) stochastic model was roughly 5% better than using the "mean value" of demand, translating into millions of dollars in potential savings.

Supply Chain Projects

- Bulk Gas Production and Distribution
  - Uncertainty in customer demands, "competitor drain"
  - Built (prototype) optimization model and simulator.
  - They are now(?) doing a real implementation

Lesson Learned

Having a (static) simulation of the production-disribution process is a key component to the project

Other Industrial Applications of SP

- Energy Industry
  - Unit Commitment Problem: Schedule production from power generation units
- Telecommunication
  - Capacity/bandwidth planning: Invest in capacity for the network before you know the true bandwidth demands
- Military
  - Network Interdiction Problem: Where to place "agent" on a network to "interrupt" evil-doers

---

It ain't that rosy

As far as I know, mot implementations are built on a case-by-case basis and are fairly ad-hoc.

Stochastic Programming Objectives—Risk Profile

What is your goal?

1. I want to do well on average
   - Expected Value
2. I want to limit my exposure in the "worst" case or cases
   - Value at Risk/Conditional Value at Risk
3. I want the probability that I achieve a goal to be sufficiently high?
   - Chance constraints
4. I want to achieve a "steady" return?
   - Dispersion-based objectives

- Each of these imply a different notion of risk, and lead to different stochastic optimization problems
- Stochastic Programming isn't about getting a number, it's about getting a distribution that looks good to you

Some SP Objectives

- $\min F(x, \bar{\omega})$ — Mean-Value Problem
- $\min \mathbb{E}_{\omega} F(x, \omega)$ — Risk Neutral
- $\min \mathbb{E}_{\omega} F(x, \omega) - \lambda \rho(F(x, \omega))$ — Risk Measures
  - $\rho(F(x, \omega)) = \mathrm{Var} F(x, \omega)$ — Markowitz
  - $\rho(F(x, \omega)) = \mathbb{E}\left[(\mathbb{E}F(x, \omega) - F(x, \omega))_+\right]$ — Semideviation

Things People Want

Network Problems

Nonlinear Problems

Integer problems

(Conditional) Value at Risk

Scenario Trees

Free Beer

Arbitrary Distributions

Stochastic Dynamic Programming

(Joint) Chance Constraints

Stochastic Dominance

Robust Optimization

Stochastic Control

Joint Distributions

Supporting Stochastic Programs

- I point out all these different flavors of SP to highlight what I think has been one of the hinderances of having a modeling laguage for SP.

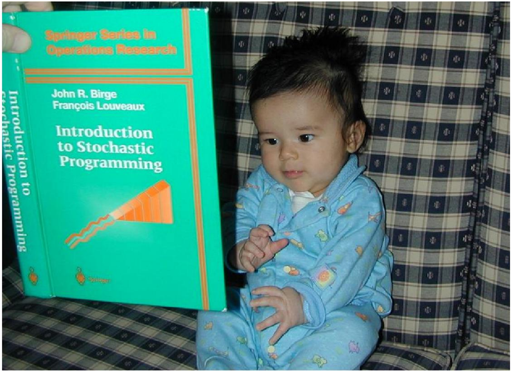
  I don't know the key to success, but the key to failure is trying to please everybody.

  Bill Cosby (1937 -)

- I believe the fact that a "stochastic program" is not a well-defined concept is one of the fundamental reasons why more people don't use stochastic programming
- Other reasons people don't use stochastic programming?

Why Don't More People Use Stochastic Programming

They don't start their training early enough!



- Jacob Linderoth, age 4 months, reading *Introduction to Stochastic Programming*

Why Don't More People Use Stochastic Programming

- Because they don't know the probability distribution?
  - Even crude approximations can help
- Because they can't "solve" them?
  - Linderoth and Wright solve a *10-million* scenario problem
  - Recent theory suggests that you don't need to include many scenarios to get an accurate solution to the true problem
- Because they can't model them?
  - Modeling tools are on the way (more later)
- Because it is hard to verify that the solution is better
  - The same could be said of Deterministic Optimization
  - Use simulation to verify that the solution is better

Probability Management

- A "true believer" is Sam Savage (consulting professor at at Stanford).
- He believes companies should have a comprehensive probability management plan.

Probability Management

- Simulations to generate distributions
- Information systems to hold distributions of key uncertain inputs
- A "Chief probability officer" responsible for signing off on the distributions

---

You can start small...

1. What are your scenarios and distributions?
2. Do you have models that can use this information?

ALGORITHMS

- I focus almost exclusively on two-stage recourse problems

Stochastic Programming

A Stochastic Program

$$\min_{x \in X} \{\mathbb{E}_\omega F(x, \omega)\}$$

---

2 Stage Stochastic LP w/Recourse

$$F(x, \omega) \stackrel{\text{def}}{=} c^T x + Q(x, \omega)$$

- $c^T x$: Pay me now
- $Q(x, \omega)$: Pay me later

The Recourse Problem

$$Q(x, \omega) \stackrel{\text{def}}{=} \min q^T y$$

$$\begin{aligned} Wy &= h(\omega) - T(\omega)x \\ y &\geq 0 \end{aligned}$$

- Expected Recourse Function:

$$\mathcal{Q}(x) \stackrel{\text{def}}{=} \mathbb{E}_\omega[Q(x, \omega)]$$

- Two-Stage Stochastic LP

Extensive Form

- Assume $\Omega = \{\omega_1, \omega_2, \dots \omega_S\} \subseteq \mathbb{R}^r$, $P(\omega = \omega_s) = p_s, \forall s = 1, 2, \dots, S$
- $T_s \equiv T(\omega_s), h_s = h(\omega_s)$
- Then can write extensive form:

$$\begin{array}{ccccccccccc}
c^T x & + & p_1 q^T y_1 & + & p_2 q^T y_2 & + & \cdots & + & p_s q^T y_s & & \\
\text{s.t.} & & & & & & & & & & \\
Ax & & & & & & & & & = & b \\
T_1 x & + & W y_1 & & & & & & & = & h_1 \\
T_2 x & & & + & W y_2 & & & & & = & h_2 \\
\vdots & & & + & & & \ddots & & & & \vdots \\
T_S x & & & & & & & + & W y_s & = & h_s \\
x \in X & & y_1 \in Y & & y_2 \in Y & & & & y_s \in Y & &
\end{array}$$

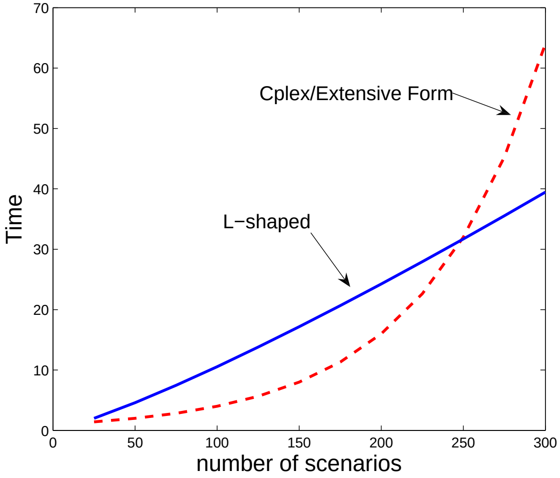
The Upshot!

- This is just a larger linear program
- It is a larger linear program that also has special structure

Best-Known Solution Procedure

M
E
T
H O D

Small SP's are Easy!



Two-Stage Stochastic Linear Programming

- We assume that the P has finite support, so ω has a finite number of possible realizations (*scenarios*):

$$\mathcal{Q}(x) = \sum_{i=1}^{N} p_i Q(x, \omega_i)$$

- For a partition of the N scenarios into sets $\mathcal{N}_1, \mathcal{N}_2, \ldots \mathcal{N}_t$, let $\mathcal{Q}_{[j]}(x)$ be the contribution of the jth set to $\mathcal{Q}(x)$:

$$\mathcal{Q}_{[j]}(x) \stackrel{\text{def}}{=} \sum_{i \in \mathcal{N}_j} p_i Q(x, \omega_i)$$

- so then $\mathcal{Q}(x) = \sum_{j=1}^{t} \mathcal{Q}_{[j]}$

Important (and well-known) Facts

- $Q(x, \omega_i)$, $\mathcal{Q}_{[\cdot]}(x)$, and $\mathcal{Q}(x)$ are piecewise linear convex functions of x.
- If π_i is an optimal dual solution to the linear program corresponding to $Q(\hat{x}, \omega_i)$, then $-T_i^T \pi_i \in \partial Q(\hat{x}, \omega_i)$
    - $g_j(\hat{x}) \stackrel{\text{def}}{=} \sum_{i \in \mathcal{N}_j} -p_i T_i^T \pi_i \in \partial \mathcal{Q}_{[j]}(\hat{x})$.

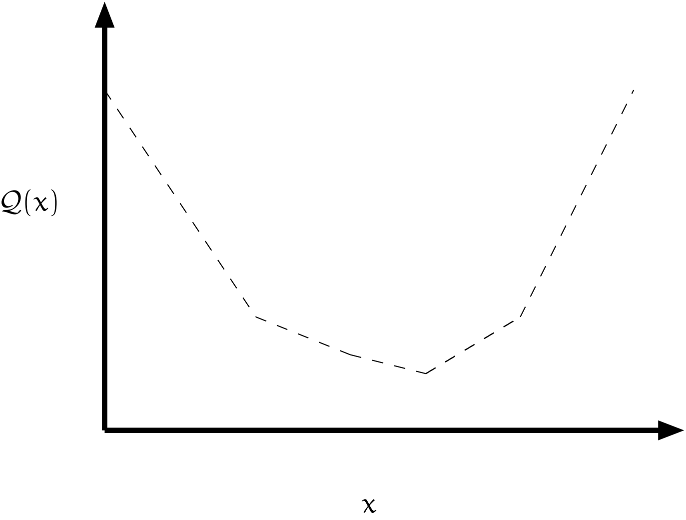
Key Idea

- Represent $\mathcal{Q}_{[j]}(x)$ by an artificial variable θ_j and find supporting planes for θ_j
    - $\theta_j \geq \mathcal{Q}_{[j]}(x^k) + g_j(x^k)^T(x - x^k) \qquad (*)$

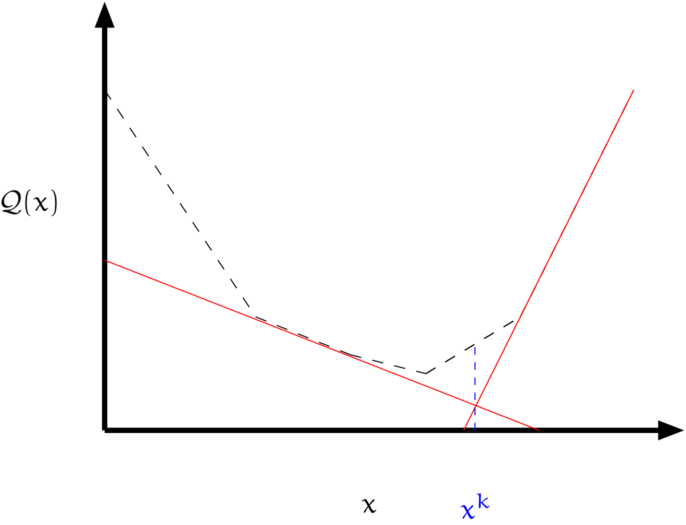
Point of Decomposition

- Evaluation of $\mathcal{Q}(\hat{x})$ is separable
- We can solve linear programs corresponding to each $Q(\hat{x}, \omega_i)$ independently – in parallel!

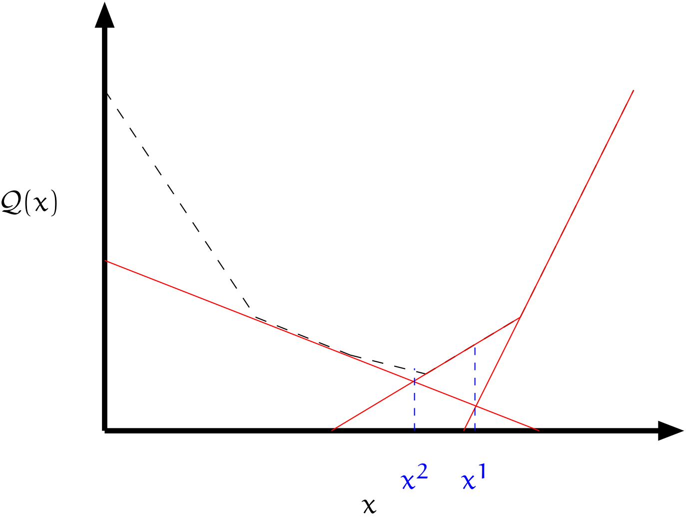
Worth 1000 Words?



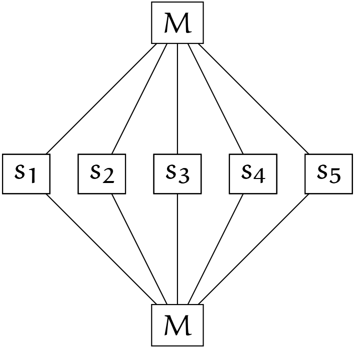
Worth 1000 Words



Worth 1000 Words



(Multicut) L-shaped method



1. Solve the **master problem** M with the current approximation to $\mathcal{Q}(x)$ for x^k.
2. Solve the **subproblems**, (s_j) evaluating $\mathcal{Q}(x^k)$ and obtaining subgradient(s) to update master approximation M
3. $k = k+1$. Goto 1.

Let's Get Parallel!

Of course, solution of s_j can be carried out independently.

Warning!

- If $\mathcal{Q}(x)$ is not convex, then this algorithm doesn't work
- If you have a integer recourse variables $y \in \mathbb{Z}^p \times \mathbb{R}^{n-p}$, the problem becomes significantly more difficult.

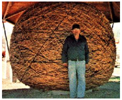
Your Options

- Give your favorite solver the full extensive form (and pray)
  - Weak relaxation
- Decomposition method: Carøe and Schultz, Sen
- Spatial branch and bound:

Want to know about stochastic integer programming/spatial branch and bound?

Talk to Nick!

Does it Work? The World's Largest LP



- Linderoth and Wright built a fancy decomposition-based solver capable of running on "the grid"
- Storm – A stochastic cargo-flight scheduling problem (Mulvey and Ruszczyński)
- We aim to solve an instance with 10,000,000 scenarios
- $x \in \mathbb{R}^{121}, y_k \in \mathbb{R}^{1259}$
- The deterministic equivalent LP is of size

$$A \in \mathbb{R}^{985,032,889 \times 12,590,000,121}$$

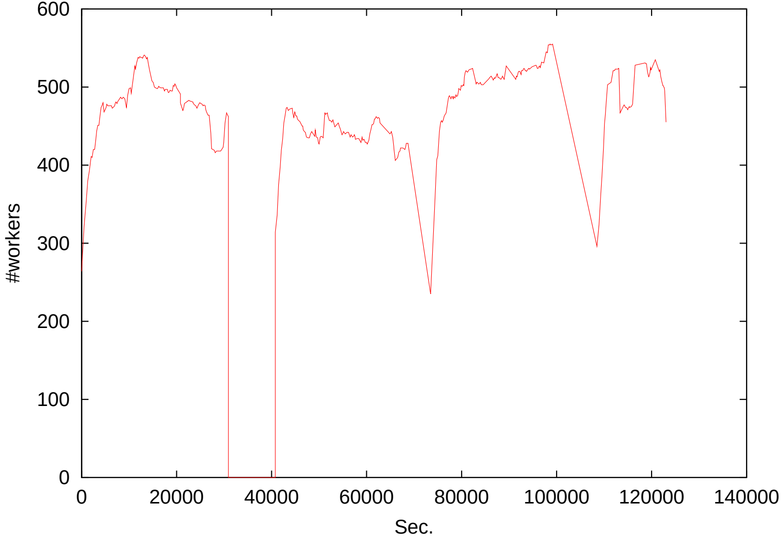
The Super Storm Computer

Number	Type	Location
184	Intel/Linux	Argonne
254	Intel/Linux	New Mexico
36	Intel/Linux	NCSA
265	Intel/Linux	Wisconsin
88	Intel/Solaris	Wisconsin
239	Sun/Solaris	Wisconsin
124	Intel/Linux	Georgia Tech
90	Intel/Solaris	Georgia Tech
13	Sun/Solaris	Georgia Tech
9	Intel/Linux	Columbia U.
10	Sun/Solaris	Columbia U.
33	Intel/Linux	Italy (INFN)
1345		

TA-DA!!!!!

Wall clock time	31:53:37
CPU time	1.03 Years
Avg. # machines	433
Max # machines	556
Parallel Efficiency	67%
Master iterations	199
CPU Time solving the master problem	1:54:37
Maximum number of rows in master problem	39647

Number of Workers



Why Sampling is Necessary

- $y_s \equiv y(\omega_s)$ is the recourse action to take if scenario ω_s occurs.
- Pro: It's a linear program.
- Con: It's a BIG linear program.
- Imagine the following (real) problem. A Telecom company wants to expand its network in a way in which to meet an unknown (random) demand.
- There are 86 unknown demands. Each demand is independent and may take on one of five values.
- $S = |\Omega| = \Pi_{k=1}^{86}(5) = 5^{86} = 4.77 \times 10^{72}$
  - The number of subatomic particles in the universe.
- How do we solve a problem that has more variables and more constraints than the number of subatomic particles in the universe?

But Its Even Worse!

- The answer is we can't!
- If Ω is not a countable set say if it is made up of continuous-valued random variables, our "deterministic equivalent" would have ∞ variables and constraints. :-)
- We solve an approximating problem obtained through sampling.

The Very Good News

- Using Monte-Carlo methods (Sample Average Approximation), we can obtain high-quality solutions
- Even Better: Can obtain (statistical) bounds on the quality of the solution

Sample Average Approximation(SAA)

The Story

- Solving two-stage SP exactly is often impossible
- Solving two-stage SP approximately is often easy: Sample Average Approximation (SAA)

- I view SAA as the Jeff Linderoth of solution methods
  - It ain't smart
  - It ain't sexy
  - But it generally does work!

SAA for Dummies

- Let v^* be the optimal solution to the "true" problem:

$$v^* \stackrel{\text{def}}{=} \min_{x \in X} \left\{ f(x) \stackrel{\text{def}}{=} \mathbb{E}_\omega F(x, \omega) \right\}$$

- Take a sample $(\omega_1, \ldots, \omega_N)$ of N realizations of the vector ω, and form the sample average function

$$\widehat{f}_N(x) \stackrel{\text{def}}{=} N^{-1} \sum_{j=1}^{N} F(x, \omega^j)$$

  - For Stochastic LP w/recourse, evaluate $\widehat{f}_N(x) \Rightarrow$ solve one LP for each of N scenarios
- Optimize sample average function:

$$v_N \stackrel{\text{def}}{=} \min_{x \in X} \left\{ \widehat{f}_N(x) \stackrel{\text{def}}{=} N^{-1} \sum_{j=1}^{N} F(x, \omega^j) \right\}$$

SAA for Dummies, Cont.

- Note that v_N is a random variable, as it depends on the (random) sample of size N
- From this information, we can get bounds on the optimal solution value v^*

All "Good" Talks Contain...

Thm. $\mathbb{E}(v_N) \leq v^* \leq f(x) \; \forall x$

Making SAA Work

- Take a solution $\hat{x}$ from a SAA instance
- We are mostly interested in estimating the quality of a given solution $\hat{x}$. This is $f(\hat{x}) - v^*$.

---

1. Get upper bound on v^* from $f(\hat{x})$. Estimate $f(\hat{x})$ by solving N' (completely independent) linear programs—recourse LP's with $\hat{x}$ fixed.

$$\widehat{f}_{N'}(\hat{x}) \stackrel{\text{def}}{=} (N')^{-1} \sum_{j=1}^{N'} F(\hat{x}, \omega^j)$$

2. Get a lower bound on v^* from $\mathbb{E}(v_N)$. Estimate $\mathbb{E}(v_N)$ by solving M independent stochastic LPs, giving optimal values $v_N^1, v_N^2, \ldots v_N^M$

$$\widehat{\mathbb{E}(v_N)} \stackrel{\text{def}}{=} M^{-1} \sum_{j=1}^{M} v_N^j$$

- Independent $\Rightarrow$ no synchronization $\Rightarrow$ good for the Grid

More Theory

- A *very interesting* result of Shapiro and Homem-de-Mello says the following:
- Suppose that $x^\star$ is the unique optimal solution to the "true" problem
- Let $\hat{x}_N$ be the solution to the sampled approximating problem
- Under certain conditions, the event $(\hat{x}_N = x^\star)$ happens with probability 1 for N large enough.
- The probability of this event approaches 1 exponentially fast as $N \to \infty$!!
- There exists a constant β such that

$$\lim_{N \to \infty} N^{-1} \log[1 - P(\hat{x} = x^*)] \leq -\beta.$$

- This is a qualitative result indicating that it might not be necessary to have a large sample size in order to solve the true problem *exactly*.
- For a problem with 5^{1000} scenarios a sample of size $N \approx 400$ is required in order to find the true optimal solution with probability 95%!!!

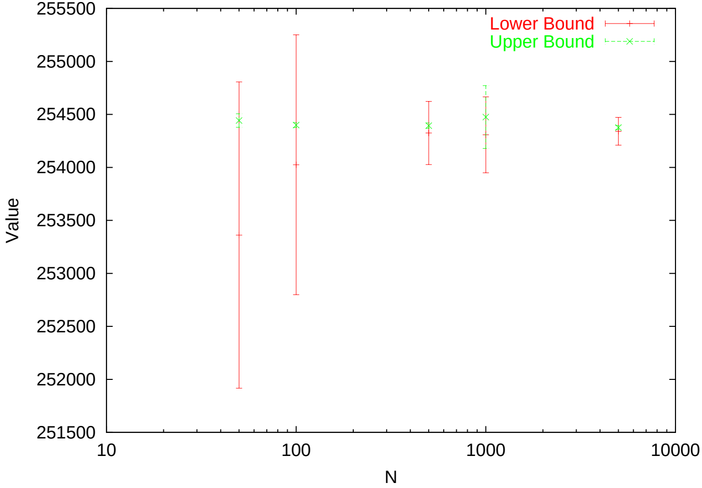
Does SAA Work on "Real" Problems?

- $M = 10$ times – Solve a stochastic sampled approximation of size N. Compute confidence interval on lower bound estimate $\widehat{\mathbb{E}(v_N)}$
- Choose one $\hat{x}$ from solution to M SAA instances and compute confidence interval on upper bound estimate $\hat{f}_{N'}(\hat{x})$, with $N' = 10000$

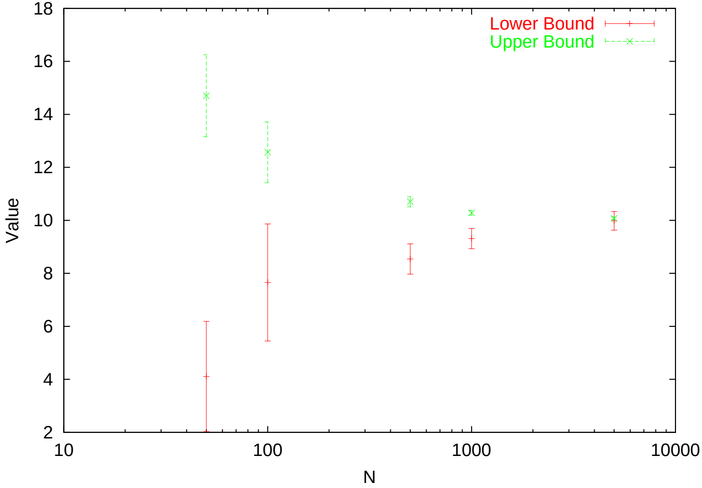
Test Instances

Name	**Application**	$\lvert\Omega\rvert$
LandS	HydroPower Planning	10^6
gbd	Aircraft Allocation	6.46×10^5
storm	Cargo Flight Scheduling	6×10^{81}
20term	Vehicle Assignment	1.1×10^{12}
ssn	Telecom. Network Design	10^{70}

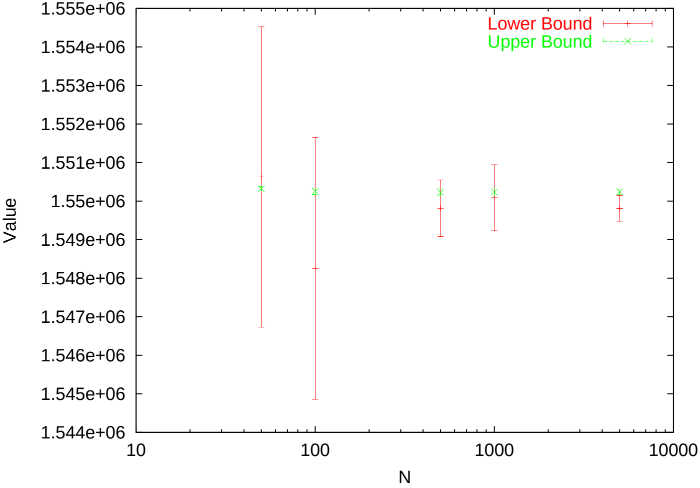
20term Convergence



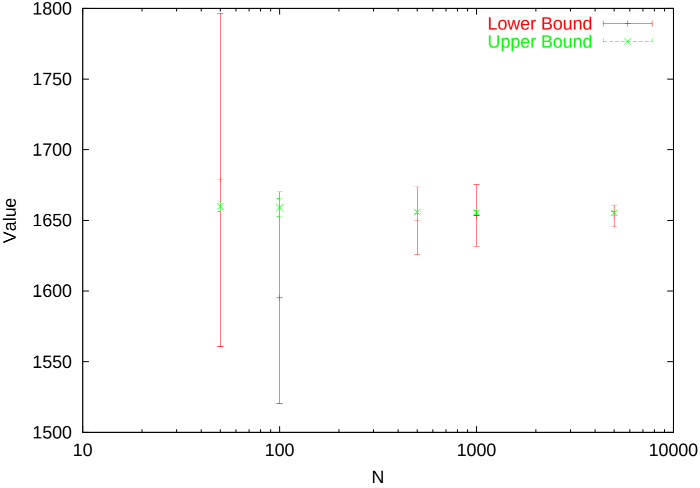
ssn Convergence



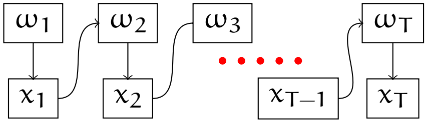
storm Convergence



gbd Convergence



Multistage Stochastic LP



- Random vectors $\omega_1 \in \mathbb{R}^{n_1}, \omega_2 \in \mathbb{R}^{n_2}, \ldots, \omega_T \in \mathbb{R}^{n_T}$
- Make sequence of decisions $x_1 \in X_1, x_2 \in X_2, \ldots, x_T \in X_T$.

- Risk Neutral: We always aim to optimize the expected value of our current decision x_t
- Linear: Assume X_t are polyhedra
- Discrete: Assume ω_t are drawn from a discrete distribution.

The Hard Part

Decisions made at period t (x_t) must only depend on events and decisions up to period t

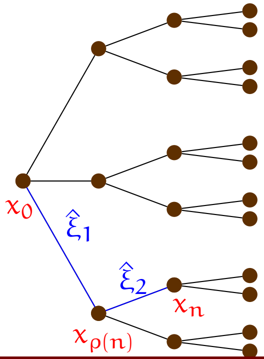
The Stickler. My Favorite Eight Syllable Word.

- We need to enforce *nonanticipativity*.
- Other eight-syllable words...
  - autosuggestibility, incommensurability, electroencephalogram, unidirectionality
- At any point in time, different scenarios "look the same"
  - We can't allow different decisions for these scenarios.
  - We are not allowed to anticipate the outcome of future random events when making our decision now.

How to do it?

1. Use Tree Structure (Nested Decomposition)
2. Create (extra) variables for all possible scenarios, and enfroce equality between decisions that should be nonanticipative (Progressive Hedging)

Scenario Tree



- N: Set of nodes in the tree
- $\rho(n)$: Unique predecessor of node n in the tree
- $\mathcal{S}(n)$: Set of successor nodes of n
- q_n: Probability that the sequence of events leading to node n occurs
- x_n: Decision taken at node n

Warning!

- Scenario Trees can get big
- There are some tools that try and "prune" the tree while keeping similar statistical properties in the stochastic process

Multistage Stochastic Programming

Entensive Form

$$z_{SP} = \min \left\{ \sum_{n \in N} q_n c_n^T x_n \mid T_n x_{\rho(n)} + W_n x_n = h_n \ \forall n \in N \right\}$$

Value Function of node n

$$\mathcal{Q}_n(x_{\rho(n)}) \stackrel{\text{def}}{=} \min_{x_n} \left\{ c_n^T x_n + \sum_{m \in \mathcal{S}(n)} \hat{q}_{mn} \mathcal{Q}_m(x_n) \mid W_n x_n = h_n - T_n x_{\rho(n)} \right\}$$

- $\hat{q}_{mn}$: conditional probability of node n given node m

- Tree structure encodes nonanticipativity

Nested Decomposition

- 0: Root node of the scenario tree
- x_0: Initial state of the system

Recursive Formulation

$$z_{SP} = \mathcal{Q}_0(x_0)$$

- Cost to go: $\mathcal{G}_n(x) \stackrel{\text{def}}{=} \sum_{m \in \mathcal{S}(n)} \hat{q}_{mn} \mathcal{Q}_m(x)$
- $M_n^k(x)$: Lower bound on $\mathcal{G}_n(x)$ in iteration k

$$\mathcal{Q}_n(x_{\rho(n)}) \geq \min_{x_n} \left\{ c_n^T x_n + M_n^k(x_n) \;\middle|\; W_n x_n = h_n - T_n x_{\rho(n)} \right\} \qquad ((\text{MLP}_n))$$

Building $M_n^k(x)$

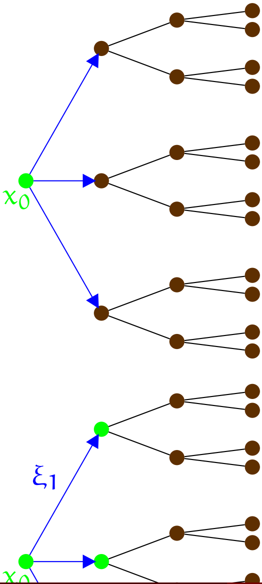
- Create a partition (or clustering $\mathcal{C}_n$) of $\mathcal{S}(n)$
- A lower bound $m_{n[j]}^k$ for each element of the partition (each cluster) is created independently

$$M_n^k(x) \stackrel{\text{def}}{=} \sum_{j \in \mathcal{C}_n} m_{n[j]}^k(x)$$

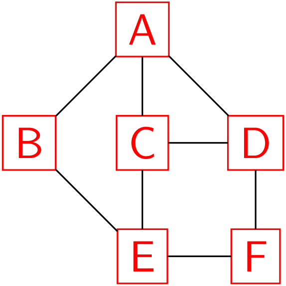
$$m_{n[j]}^k(x) \stackrel{\text{def}}{=} \inf\left\{\theta_j \;\middle|\; \theta_j e \geq F_{n[j]}^k x + f_{n[j]}^k\right\}$$

- $F_{n[j]}^k, f_{n[j]}$ obtained from dual solutions (to form subgradients) of linear programs of nodes within cluster $[j]$
- $M_n^k(x^*) \rightarrow \mathcal{G}_n(x^*)$

Action Pictures



A small Multistage Telecom Problem



- Set of stages T
- Set J of links
- Sets I_t of demands
- Random demand $d_t(\xi) \in \mathbb{R}^{|I_t|}$
- Budget each period
- Install capacity on links each period to minimize the total expected unserved demand

Some (Limited) Computational Results

- $T = 5$
- K: Realizations/Period
- N: Number of scenarios
- DE: Size of deterministic equivalent

K	N	DE Size
30	0.81M	18M * 31M
50	6.25M	140M * 236M
60	12.9M	290M * 488M

Computational Results

- It: Number of iterations (Times MLP_0 was solved)
- E: Parallel efficiency.

$$\frac{\text{Time machines solving } MLP_n}{\text{Time machines available}}$$

K	It	Avg Workers	Wall Time	CPU Time	E
30	9	62	2:34:21	6:15:15:10	67
50	7	75	1:12:49:27	85:20:24:15	77
60	11	162	3:16:51:00	431:12:15:37	73

Existing Modeling Tools

- Many stochastic programming implementations I'm aware of have been built from scratch
- But there are some modeling tools on the way

Name	Author(s)	Comment
AIMMS	AIMMS Team	Commercial
Gams	Gams Team	Commercial
MPL	Kristjensen	Commercial
XPRESS-SP	Verma, Dash Opt.	Commercial, Beta
SPiNE	Valente, CARISMA	
STRUMS	Fourer and Lopes	Prototype(?)
SUTIL	Czyzyk and Linderoth	C++ classes
SLPLib	Felt, Sarich, Ariyawansa	Open Source C Routines
COIN-Smi, SP/OSL	COIN, IBM	C++ methods

Existing Solution Tools

- Most stochastic programming implementations of which I'm aware, merely form and solve extensive form
- Other software:

Name	Author(s)	Comment
AIMMS	AIMMS Team	Commercial, LShaped method
SLP-IOR	Kall, Mayer	LShaped, Stochastic Decomposition, others
MSLiP	Gassmann	Nested LShaped
SPInE	Valente, CARISMA	Commercial, LShaped method, may not exist anymore
BNBS	Altenstedt	Nested LShaped method, Open source
ATR	Linderoth, Wright	Design to run in parallel. Not simple to build and run

Conclusions

Stochastic Programming

- A tool for decision making under uncertainty
  - Considers the impact of recourse decisions
  - It may not be the answer, but it does help you hedge against upcoming uncertainty
  - More importantly, it gets people talking about the impact of uncertainty in the decision making process
- Planning with "mean-value" estimates will not lead to an optimal policy

---

- Used with some success in industry
  - Financial Services (Many successes)
  - Logistics and Supply Chain (Fewer successes, but coming!)
- Tools and algorithms are "on the way"

We Want YOU!

To consider using Stochastic Programming as a decision support tool to help manage in turbulent times!

Thanks!

- I am happy to help. email: linderoth@wisc.edu
- http://www.stoprog.org/

Some Take Away Quotes

"If a man will begin with certainty, he shall end in doubts, but if he will be content to beign with doubts, he shall end in certainties"

— Francis Bacon

"It is a good thing for the uneducated person to read books of quotations"

—Winston Churchill