Optimization Tools in an Uncertain Environment

Michael C. Ferris

University of Wisconsin, Madison

Uncertainty Workshop, Chicago: July 21, 2008

Optimization of a model under uncertainty

Modeler: assumes knowledge of distribution

Often formulated mathematically as

$$\min_{x \in X} f(x) = \mathbb{E}[F(x,\xi)] = \int_{\xi} F(x,\xi)p(\xi)d\xi$$

(p is probability distribution).

- Can think of this as optimization with noisy function evaluations
- Traditional Stochastic Optimization approaches: (Robinson/Munro, Keifer/Wolfowitz)
- Often require estimating gradients: IPA, finite differences
- Stochastic neighborhood search

Example: Two stage stochastic LP with recourse

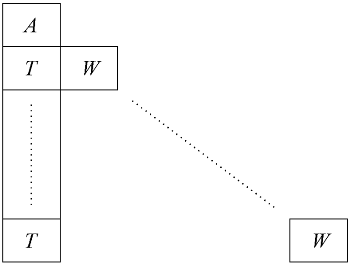
$$\min_{x \in \mathbb{R}^n} c^T x + \mathbb{E}[\mathcal{Q}(x,\xi)] \text{ s.t. } Ax = b, x \geq 0$$

$$\mathcal{Q}(x,\xi) = \min_y q^T y \text{ s.t. } Tx + Wy = h, y \geq 0$$

$\xi = (q, h, T, W)$ (some are random). Expectation wrt ξ.

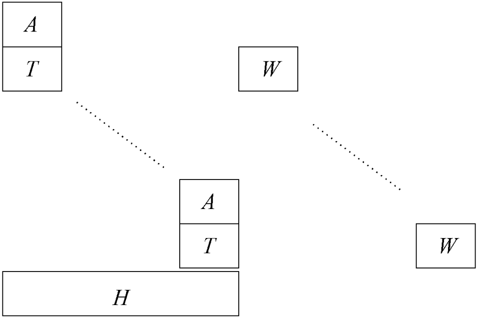
x are first stage vars, y are second stage vars.

Special case: discrete distribution $\Omega = \{\xi_i : i = 1, 2, \ldots, K\}$



Deterministic equivalent problem

Key-idea: Non-anticipativity constraints



- Replace x with $x_1, x_2, \ldots, x_K$
- Non-anticipativity: $(x_1, x_2, \ldots, x_K) \in L$ (a subspace) - the H constraints

Computational methods exploit the separability of these constraints, essentially by dualization of the non-anticipativity constraints.

- Primal and dual decompositions (Lagrangian relaxation, progressive hedging, etc)
- L shaped method (Benders decomposition applied to det. equiv.)
- Trust region methods and/or regularized decomposition

Complications

- Multistage problems
  - recursive application of above, scenario trees
  - dynamic programming approaches - see Judd talk
  - reinforcement learning, neuro-dynamic programming
  - real options
- Stochastic integer programming
- Stochastic variational inequalities / complementarity problems
- Nonlinear (convex or otherwise) recourse models

Sampling methods

But what if the number of scenarios is too big (or the probability distribution is not discrete)? use sample average approximation (SAA)

- Take sample $\xi_1, \ldots, \xi_N$ of N realizations of random vector ξ
  - viewed as historical data of N observations of ξ, or
  - generated via Monte Carlo sampling
- for any $x \in X$ estimate $f(x)$ by averaging values $F(x, \xi_j)$

$$\text{(SAA):} \ \min_{x \in X} \left\{ \hat{f}_N(x) := \frac{1}{N} \sum_{j=1}^{N} F(x, \xi_j) \right\}$$

- Nice theoretical asymptotic properties
- Can use standard optimization tools to solve the SAA problem
- Implementation uses common random numbers, distributed computation
- Monte Carlo Sampling (Quasi-Monte Carlo Sampling)

Variance reduction

Choose $\xi_1, \xi_2, \ldots, \xi_N$ carefully (essentially exploting properties of numerical integration) to reduce the variance of $\hat{f}_N(x)$

- Latin Hypercube Sampling
- Importance Sampling, Likelihood ratio methods

Significantly improves performance of optimization codes

Example: Robust Linear Programming

Data in LP not known with certainty:

$$\min c^T x \text{ s.t. } a_i^T x \leq b_i, i = 1, 2, \ldots, m$$

Suppose the vectors a_i are known to be lie in the ellipsoids (no distribution)

$$a_i \in \varepsilon_i := \{\bar{a}_i + P_i u : \|u\|_2 \leq 1\}$$

where $P_i \in \mathbb{R}^{n \times n}$ (and could be singular, or even 0).
Conservative approach: robust linear program

$$\min c^T x \text{ s.t. } a_i^T x \leq b_i, \text{ for all } a_i \in \varepsilon_i, i = 1, 2, \ldots, m$$

Robust Linear Programming as SOCP

The constraints can be rewritten as:

$$\begin{aligned} b_i &\geq \sup \left\{ a_i^T x : a_i \in \varepsilon_i \right\} \\ &= \bar{a}_i^T x + \sup \left\{ u^T P_i^T x : \|u\|_2 \leq 1 \right\} = \bar{a}_i^T x + \left\| P_i^T x \right\|_2 \end{aligned}$$

Thus the robust linear program can be written as

$$\min c^T x \text{ s.t. } \bar{a}_i^T x + \left\| P_i^T x \right\|_2 \leq b_i, i = 1, 2, \ldots, m$$

$$\min c^T x \text{ s.t. } (b_i - \bar{a}_i^T x, P_i^T x) \in C$$

where C represents the second-order cone. Solution (as SOCP) by Mosek or Sedumi, CVX, etc

Example: Simulation Optimization

- Computer simulations are used as substitutes to understand or predict the behavior of a complex system when exposed to a variety of realistic, stochastic input scenarios
- Simulations are widely applied in epidemiology, engineering design, manufacturing, supply chain management, medical treatment and many other fields
- Optimization applications: calibration, parameter tuning, inverse optimization, pde-constrained optimization

$$\min_{x \in X} f(x) = \mathbb{E}[F(x, \xi)],$$

- The sample response function $F(x, \xi)$
  - typically does not have a closed form, thus cannot provide gradient or Hessian information
  - is normally computationally expensive
  - is affected by uncertain factors in simulation

Bayesian approach

- The underlying objective function $f(x)$ still has to be estimated.
- Denote the mean of the simulation output for each system as $\mu_i = f(x_i) = \mathbb{E}[F(x_i, \xi)]$
- In a Bayesian perspective, the means are considered as Gaussian random variables whose posterior distributions can be estimated as

$$\mu_i | X \sim N(\bar{\mu}_i, \hat{\sigma}_i^2 / N_i),$$

  where $\bar{\mu}_i$ is sample mean and $\hat{\sigma}_i^2$ is sample variance. The above formulation is one type of posterior distribution.
- Instrument existing optimization codes to use this derived distribution information
  - Derivative free optimization, surrogate optimization
  - Response surface methodology
  - Evolutionary methods

Example: Chance Constrained Problems

$$\min_{x \in X} f(x) \text{ s.t. } Prob(C(x,\xi) > 0) \le \alpha$$

α is some threshold parameter, C is vector valued

- joint probabilistic constraint: all constraints satisfied simultaneously - possible dependence between random variables in different rows
- extensive literature
- linear programs with probabilistic constraints are still largely intractable (except for a few very special cases)
  - for a given $x \in X$, the quantity $Prob(C(x,\xi) > 0)$ requires multi-dimensional integration
  - the feasible region defined by a probabilistic constraint is not convex
- Recent work by Ahmed, Leudtke, Nemhauser and Shapiro

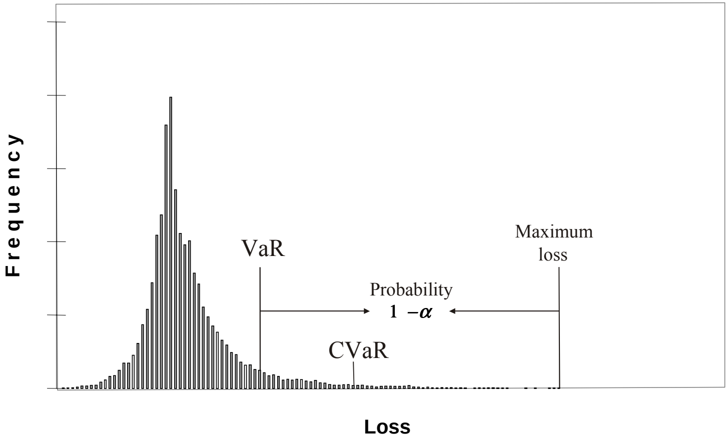
Example: Risk Measures

- Classical: utility/disutility function $u(\cdot)$:

$$\min_{x \in X} f(x) = \mathbb{E}[u(F(x, \xi))],$$

- Modern approach to modeling risk aversion uses concept of risk measures
  - mean-risk
  - semi-deviations
  - mean deviations from quantiles, VaR, CVaR
  - Römish, Schultz, Rockafellar, Urasyev (in Math Prog literature)
  - Much more in mathematical economics and finance literature
  - Optimization approaches still valid, different objectives

CVaR constraints: mean excess dose (radiotherapy)



Move mean of tail to the left!

So what's my point?

- Modeling and optimization model building is key!
- Economic models versus scientific engineering models: philosophical differences in usage
- Many different optimization approaches to treat (model) uncertainties
- How much do I know about distribution of data?
- Specific models needed for these applications
- Modeling systems (GAMS, AMPL, AIMMS) have had limited impact due to no common input/model format
- Stochastic model implementation and interfaces to these tools are needed