Robust Optimization

Mustafa Ç. Pınar
Bilkent University

September 2005

What are we dealing with?

Consider the optimization problem

$$\begin{array}{rrcl} \min & & f(x,\xi) & \text{(P1)} \\ \text{subject to} & & & \\ & g_i(x,\xi) & \leq & 0 \\ & x & \in & X \end{array}$$

- x is the vector of variables
- ξ is the vector of data (uncertain)
- f and g_i are convex functions
- X is a possibly non-convex (e.g., the set of nonnegative integers, or binary integers) set

Our Typical Optimization Problems

Linear Programming

$$\begin{array}{rcl} \min & c^T x & \\ \text{subject to} & & \\ Ax & \geq & b \end{array} \tag{P2}$$

- A, b, and c could be plagued with uncertainty, or could be just estimates from a simulation or discretization process.

Our Typical Optimization Problems II

Linear Programming with Integers

$$\begin{array}{rcl} \min & c^T x & \\ \text{subject to} & & \\ Ax & \geq & b \\ x & \in & X \end{array} \tag{P3}$$

- X is the set of integers, or binary integers,
- A, b, and c could be plagued with uncertainty, or could be just estimates from a simulation or discretization process.

Our Typical Optimization Problems III

Quadratic Programming

$$\begin{array}{lcr} \min & \frac{1}{2}x^TQx + c^Tx & \text{(P4)} \\ \text{subject to} & & \\ & Ax \geq b & \end{array}$$

- Q is a symmetric positive (semi-)definite matrix
- Q, A, b, c could be plagued with uncertainty, or could be just estimates from a simulation or discretization process.

Our Typical Optimization Problems IV

Second-order Cone Programming

$$\begin{array}{lcr} \min & \frac{1}{2}c^Tx & \text{(P5)} \\ \text{subject to} & & \\ & \|Ax - b\|_2 \leq f^tx + g & \end{array}$$

- A is a $m \times n$ matrix, the remaining quantities have conformable dimensions
- SOCP is usually obtained as a result of incorporating robustness into LP
- It can be encountered as a model in engineering applications, with uncertainty in the parameters (things get harder..)

How is Uncertainty Treated in Optimization?

- **Sensitivity Analysis**: Solve the problem with fixed values of the parameters (perhaps the most likely values) and see how the optimal solution is affected by small perturbations.
- Easy to perform for Linear Programming
- Uses Duality Theory
- But, this is a post-mortem tool.

Another tool for treating uncertainty in optimization

- **Stochastic Programming**: Under an assumed probability distribution of the uncertain parameters, the objective function becomes a collection of random variables. Choose the "best" in this collection according to some criterion, e.g., the expected value or some other suitable utility function.
- It is not a post-mortem tool, is pro-active
- It is easy to model recourse issues (more on this later)
- But, this approach results in huge optimization problems with heavy data requirements.

Robust Optimization (RO): what is it?

- A complementary methodology to stochastic programming and sensitivity analysis
- Seeks a solution that will have an "acceptable" performance under most realizations of the uncertain inputs
- Usually, no distribution assumption is made on uncertain parameters (if such information is available, it can be utilized beneficially)
- Usually, it is a conservative (worst-case oriented) methodology.

Robust Optimization is useful if

- some parameters come from an estimation process and may be contaminated with estimation errors
- there are "hard" constraints that must be satisfied no matter what
- the objective function value/optimal solutions are highly sensitive to perturbations
- the modeler/designer cannot afford low probability high-magnitude risks (typical example: designing a bridge).

Main Contributors to Robust Optimization

- Charnes, Cooper et al. (probabilistic constraints)
- Soyster (column uncertainty in LP)
- Ben-Tal and Nemirovski (ellipsoidal uncertainty)
- Kouvelis and Yu (minimax regret)
- El-Ghaoui et al.

Some more contributors..

- Zenios and co.
- Mulvey and co.
- Tansel and co.
- Yaman, Karaşan and P.
- Bertsimas and Sim
- Ordoñez
- Tütüncü

Two major problems:

- How to represent uncertainty?
- How to compute a robust solution?
- What is a robust solution anyway?
- **Definition**: An optimal (feasible) solution is robust if it stays optimal (feasible) under any realization of the data
- Definition too restrictive. Most unlikely that such a solution exists.

Another definition:

An optimal solution is **robust** if it minimizes **maximum relative regret**.

What is maximum relative regret?

- **Example**: Consider the shortest path problem on a directed graph where arc costs are subject to uncertainty. Let us assume arc (i, j) can have as cost value any value in the interval $[\underline{c}_{ij}, \bar{c}_{ij}]$.
- Define the maximum relative regret associated with any path: Put all arc costs on the path to their upper bounds and all other arc costs to their lower bounds. Find the shortest path in this realization of arc costs.
- **Maximum Relative Regret** = the cost of the path (at lower bounds) − that of the shortest path.
- Compute the solution which minimizes maximum regret. **NP-hard**!

Another example

(Yaman, Karaşan, P. (2001))

- Consider the minimum spanning tree problem on an undirected graph where edge costs are subject to uncertainty. Let us assume edge (i, j) can have as cost value any value in the interval $[\underline{c}_{ij}, \bar{c}_{ij}]$.
- Define the **maximum relative regret** associated with any tree: Put all edge costs on the path to their upper bounds and all other edge costs to their lower bounds. Find the minimum spanning tree in this realization of arc costs.
- Approach too **conservative** from a modeling perspective. We are acting as if all random data will assume their worst possible values simultaneously.

Soyster's Approach to Robust Optimization

- Consider the linear program

$$\begin{aligned} \max \quad & c^T x \qquad \text{(P6)} \\ \text{subject to} \quad & \\ & a_i^T x - b_i \leq 0, \; \forall i = 1, \ldots, m \end{aligned}$$

- Consider a particular row i, a_i and let J_i represent the set of coefficients in row i that are subject to uncertainty
- Each entry $a_{ij}, j \in J_i$ is modeled as a symmetric and bounded random variable $\tilde{a}_{ij}$ taking values in $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$.

Soyster's Robust Formulation

$$\begin{array}{rll} \max & c^T x & \text{(P7)} \\ \text{subject to} & & \\ \sum_j a_{ij}x_j + \sum_{j \in J_i} \hat{a}_{ij} y_j & \leq & b_i, \ \forall i = 1, \ldots, m \\ -y_j \leq x_j & \leq & y_j \ \forall j \\ y & \geq & 0 \end{array}$$

- Let x^* be the optimal solution vector. At optimality, we have $y_j = |x_j^*|$.
- Thus the constraint is actually equivalent to

$$\sum_j a_{ij}x_j + \sum_{j \in J_i} \hat{a}_{ij}|x_j^*| - b_i \leq 0, \ \forall i = 1, \ldots, m$$

- For every possible realization $\tilde{a}_{ij}$, the solution remains feasible.

Ben-Tal and Nemirovski Approach to Robust Optimization

- Consider the linear program

$$\begin{array}{rll} \min & c^T x & \text{(P8)} \\ \text{subject to} & & \\ Ax & \geq & b \end{array}$$

- Assume each row a_i of A is uncertain but known to lie in **ellipsoids**

$$E_i = \{a_i : a_i = \bar{a}_i + P_i u_i, \|u_i\|_2 \leq 1\}$$

where P_i is symmetric positive (semi)definite matrix for all i.

- Assume rows a_i assume values independently of one another.

What is a robust solution in the world of Ben-Tal and Nemirovski?

- We want to make sure the constraints $Ax \geq b$ are satisfied for **all realizations** of the data.
- We do not tolerate a violation of the constraints for any values of the uncertain parameters in the uncertainty set
- Among such solutions, pick one that minimizes the objective function value.

Robust Counterpart to Linear Programs

- Consider

$$\begin{aligned} \min \quad & c^T x \\ \text{subject to} \quad & \\ & a_i^T x - b_i \geq 0, \ \forall\, a_i \in E_i, \forall i = 1, \ldots, m \end{aligned} \tag{P9}$$

- But we can rewrite this

$$\begin{aligned} \min \quad & c^T x \\ \text{subject to} \quad & \\ & \bar{a}_i^T x + x^T P_i u_i - b_i \geq 0, \ \forall\, \|u_i\|_2 \leq 1, i = 1, \ldots, m \end{aligned} \tag{P10}$$

Derivation of Robust Counterpart cont'd

- The previous problem is equivalent to

$$\min \quad c^T x \qquad \text{(P11)}$$

subject to

$$\bar{a}_i^T x - b_i + \min_{u_i \,|\, \|u_i\|_2 \le 1} x^T P_i u_i \quad \ge \quad 0, i = 1, \ldots, m$$

Derivation of Robust Counterpart cont'd

- Now use the fact that

$$\min_{u_i \,|\, \|u_i\|_2 \le 1} x^T P_i u_i = -\|P_i^T x\|_2$$

- Therefore, we obtain the robust counterpart

$$\min \quad c^T x \qquad \text{(P12)}$$

subject to

$$\bar{a}_i^T x - b_i - \|P_i x\|_2 \quad \ge \quad 0, i = 1, \ldots, m$$

- This is a **Second-order Cone Programming** problem!

Comparison with Soyster's Approach to Robust Optimization

- Remember the linear program

$$\begin{array}{rcl} \max & c^T x & \text{(P13)} \\ \text{subject to} & & \\ a_i^T x - b_i & \leq & 0,\ \forall i = 1, \ldots, m \end{array}$$

- Consider a particular row i, a_i and let J_i represent the set of coefficients in row i that are subject to uncertainty
- Each entry $a_{ij}, j \in J_i$ is modeled as a symmetric and bounded random variable $\tilde{a}_{ij}$ taking values in $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$.
- What is the BTN robust counterpart here?

The BTN Robust Counterpart

is

$$\begin{array}{rcl} \max & & c^T x \text{ (P14)} \\ \text{subject to} & & \\ \sum_j a_{ij} x_j + \sum_{j \in J_i} \hat{a}_{ij} y_j + \Omega_i \sqrt{\sum_{j \in J_i} \hat{a}_{ij}^2 z_{ij}^2} & \leq & b_i,\ \forall i \\ -y_j \leq x_j - z_{ij} & \leq & y_j\ \forall j \\ y & \geq & 0 \end{array}$$

- BTN showed that the probability that the ith constraint is violated is at most $\exp -\Omega_i^2/2$.
- Less conservative than the Soyster's problem (why?)

Engineering Interpretation

- This is like replacing a random quantity by its expectation minus a constant (2 or 3) times its standard deviation.
- Engineers use a similar approach for safety.
- A similar problem is obtained when we deal with **probabilistic chance constraints** where the rows of A are assumed to have Gaussian distribution with known mean and variance, c.f. Boyd and Vandenberghe.
- Have to use the engineering safety approach (RC) with caution, c.f., example later..

Another Example of Robust Counterpart

- Assume now c is subject to ellipsoidal uncertainty:

$$E = \{c : c = \bar{c} + Pu, \|u\|_2 \leq 1\}$$

- We can address this problem in two equivalent ways: The first one is by passing to the epigraph form of the problem:

$$\begin{array}{rrcl} \min & & t & \\ \text{subject to} & & & \\ & c^T x & \leq & t \; \forall \; c \in E \\ & a_i^T x - b_i & \geq & 0, \forall i = 1, \ldots, m \end{array} \tag{P15}$$

Another example of Robust Counterpart

- The second one is by passing to a worst-case min-max formulation:

$$\begin{aligned} \min \quad & \max_{c \in E} c^T x \\ \text{subject to} \quad & \\ & a_i^T x - b_i \geq 0, \forall i = 1, \ldots, m \end{aligned} \tag{P16}$$

- In both cases, the end result is the same:

$$\begin{aligned} \min \quad & \bar{c}^T x - \|Px\|_2 \\ \text{subject to} \quad & \\ & a_i^T x - b_i \geq 0, \forall i = 1, \ldots, m \end{aligned} \tag{P17}$$

A second-order cone program!

S-Lemma and S-Procedure: Every day bread in Control Theory

Theorem S-Lemma. Let $A, B \in S^n$ and some $\bar{x}$ such that $\bar{x}^t A \bar{x} > 0$. Then $x^t A x \geq 0 \;\Rightarrow\; x^t B x \geq 0$ if and only if there exists $\lambda \geq 0$ such that $B - \lambda A \succeq 0$

Lemma. S-Procedure. Let $F_0, F_1, \ldots, F_p$ be quadratic functions of the variable $\zeta \in \mathbb{R}^m$:

$$F_i(\zeta) = \zeta^T T_i \zeta + 2u_i^T \zeta + v_i, i = 0, 1, \ldots, p,$$

where $T_i = T_i^T$. The following condition on $F_0, F_1, \ldots, F_p$:

$$F_0(\zeta) \geq 0 \text{ for all } \zeta \text{ such that } F_i(\zeta) \geq 0, i = 1, \ldots, p$$

holds if there exist $\tau_1 \geq 0, \ldots, \tau_p \geq 0$ such that

$$\begin{bmatrix} T_0 & u_0 \\ u_0^T & v_0 \end{bmatrix} - \sum_{i=1}^{p} \tau_i \begin{bmatrix} T_i & u_i \\ u_i^T & v_i \end{bmatrix} \geq 0$$

When $p = 1$, the converse holds (S-Lemma) provided there is ζ_0 such that $F_1(\zeta_0) > 0$.

- Very useful for finding robust counterparts of quadratic programming problems
- See Ben-Tal and Nemirovski, or El-Ghaoui and Lebret

Example: Structured Robust Least Squares Approximation

- Given $A_0, A_1, \ldots, A_p \in \mathbb{R}^{n \times m}$ and $b_0, b_1, \ldots, b_p \in \mathbb{R}^n$, define for every $\delta \in \mathbb{R}^p$,

$$A(\delta) = A_0 + \sum_{i=1}^{p} \delta_i A_i, b(\delta) = b_0 + \sum_{i=1}^{p} \delta_i b_i$$

- Consider now the robust problem:

$$\min_x \max_{\|\delta\|_2 \leq \rho} \|A(\delta)x - b(\delta)\|.$$

- Using the notation

$$M(x) = [A_1 x - b \ldots A_p x - b]$$

and $F = M(x)^T M(x)$, $g = M(x)^T (A_0 x - b)$, $h = \|A_0 x - b_0\|_2^2$

the inner max problem can be posed

$$\max_{\delta^T\delta \le 1} [1 \ \delta^T] \begin{bmatrix} h & g^T \\ g & F \end{bmatrix} \begin{bmatrix} 1 \\ \delta \end{bmatrix}$$

where we assumed $\rho = 1$ for simplicity.

- Homework: Can you complete the transformation of the inner max problem for tomorrow?

Pros and Cons I

- Can interpret $\bar{a}_i$ as **expected value** and P_i as the square root of the **variance/covariance** matrix.
- We can explicitly give the robust counterpart problem to an uncertain LP where uncertainty affects the rows independently in ellipsoidal form.
- All random data will not assume their worst possible values simultaneously.
- Therefore, approach **less conservative** compared to minimax regret.
- Can handle equally easily uncertainty in c and b.

Pros and Cons II

- The big issue is to find the appropriate uncertainty set. This is an application dependent question.
- The robust counterpart is not an LP. It is a second-order cone program.
- Second-order cone programs are not much harder than LPs theoretically. They can be solved to any desired accuracy by polynomial interior point methods.
- However, in practice they are **not as nice** as LPs. The current limit is a few thousand constraints/variables.
- Furthermore, things get much more complicated when we move to e.g., **convex quadratic programs**, and other types of ellipsoidal uncertainty.

- When x is a **discrete** variable, how are we to solve the robust counterpart under ellipsoidal uncertainty?

Bertsimas and Sim Approach to Robust Optimization

Consider the linear program

$$\begin{aligned} \min \quad & c^T x \\ \text{subject to} \quad & \\ x &\in X \end{aligned} \tag{P18}$$

- Assume uncertainty is in the objective function where the coefficients c_j take values in the interval $[\bar{c}_j, \bar{c}_j + d_j]$, **independently** of one another.
- The set X represents non-negativity, linear constraints, binary/integer requirements.

- What is the biggest damage nature/adversary can inflict?
- Assuming you have chosen x a priori: the **maximum damage** one can incur
$$\sum_{j=1}^{n} d_j x_j$$
- But, most likely, nature/adversary will not push all c_j's to their upper bounds simultaneously
- So, how can I avoid over-protecting myself?

Restricted Adversary

- **Restricted damage**: What if I hedge myself against a restricted adversary: assume only a **subset** of the c_j's will be pushed to their upper bounds.
- Choose a Γ between 1 and n. And consider the **maximum restricted damage**:

$$\max_{\{S|S\subseteq\{1,\ldots,n\},|S|=\Gamma\}} \sum_{j\in S} d_j x_j$$

- Therefore, I want to solve the robust counterpart

$$\min_{x\in X} c^T x + \max_{\{S|S\subseteq\{1,\ldots,n\},|S|=\Gamma\}} \sum_{j\in S} d_j x_j$$

- Γ will control the **conservatism** in the solution, c.f. Theorem later.

What can I say about the inner max problem?

$$\begin{aligned} \max \quad & \sum_{j=1}^{n} x_j d_j s_j && \text{(P19)} \\ \text{subject to} \quad & \\ & \textstyle\sum_{j=1}^{n} s_j = \Gamma \\ & s_j \in \{0,1\}, j = 1,\ldots,n \end{aligned}$$

- Relax the binary requirements to

$$0 \le s_j \le 1, j = 1,\ldots,n$$

- **Easy to see fact**: The relaxed problem always has a binary optimal solution
- Now, use **LP duality** to obtain the following robust

counterpart

$$\min_{x \in X, y} \bar{c}^T x - \Gamma y + \sum_{j=1}^{n} \max(0, d_j x_j + y)$$

- Can **linearize** the max terms by introducing extra nonnegative variables.
- This is a problem of the same nature as the problem of departure. I.e., if it was an LP to begin with, it stays an LP.
- If it was a **polynomially solvable** $0-1$ **LP** it remains so, e.g., shortest path, spanning tree, weighted non-bipartite matching etc..
- In the $0-1$ case, further simplification results into the solution of **at most** n **problems of identical structure** to obtain the robust solution. Does **not** carry over to the general integer case.

Simplification in the $0-1$ case..

Recall the robust problem where we impose now $X \subset \{0,1\}^n$:

$$\min_{x \in X, y} -\Gamma y + \sum_{j=1}^{n} \bar{c}_j x_j + \max(0, d_j x_j + y)$$

Observe that because x_j's are binary variables the above is equivalent to

$$\min_{x \in X, y} -\Gamma y + \sum_{j=1}^{n} (\max(0, d_j + y) + \bar{c}_j) x_j$$

Result

Assume the c_j's are **iid** random variables, symmetrically distributed.

Rewrite the uncertain problem as

$$\begin{array}{rcl} \min & t & \\ \text{subject to} & & \\ c^T x & \le & t \\ x & \in & X \end{array} \qquad \text{(P20)}$$

Optimal Bound

Thm.[**BS**] Let x^* denote a robust solution for some Γ. Then

$$\text{Prob}(\sum_{j=1}^{n} c_j x_j^* > t^*) \le \frac{1}{2^n}\{(1-\mu)\sum_{l=\lfloor \nu \rfloor}^{n} \binom{n}{l} + \mu \sum_{l=\lfloor \nu \rfloor + 1}^{n} \binom{n}{l}\}$$

where $\nu = \frac{\Gamma + n}{2}$ and $\mu = \nu - \lfloor \nu \rfloor$.

How to interpret the result?

- For a given Γ let x_{rob} denote an optimal solution to the robust problem. Let z_{rob} denote the **optimal value** of the robust problem. Then, **most of the time** for random c, the value $c^T x_{rob}$ will remain smaller or equal to z_{rob}.
- Notice that for $\Gamma = n$ the right hand reduces to $\frac{1}{2^n}$.
- **Immediate Corollary.** Let z_{rand} denote the optimal value of the uncertain problem for a random c. Since $c^T x_{rob}$ is an upper bound on z_{rand}, then with probability

$$1 - \frac{1}{2^n}\{(1-\mu)\sum_{l=\lfloor\nu\rfloor}^{n}\binom{n}{l} + \mu\sum_{l=\lfloor\nu\rfloor+1}^{n}\binom{n}{l}\}$$

we will have $z_{rand} \leq z_{rob}$.

Another bound [P. 2002]

Thm. Assume that the cost coefficients randomly take values in the following fashion: For positive ϵ and $\bar{c}_j > 0$, $j = 1, \ldots, n$, let

$$c_j = \bar{c}_j(1 + \epsilon\xi_j), \tag{21}$$

where ξ_j are iid, symmetrically distributed in $[-1, 1]$, and assume w.l.o.g. $\sum_{j=1}^{n} \epsilon^2 \bar{c}_j^2 = 1$. Let x^{rob}, y^{rob} denote a robust solution for some $\Gamma < n$, and t^{rob} the corresponding optimal value. If $y^{rob} > 0$ then, we have

$$\Pr\{c^T x^{rob} > t^{rob}\} \leq e^{-\frac{\Gamma^2\epsilon^2(\min_i \bar{c}_i)^2}{2}}. \tag{22}$$

For reflection and discussion..

Are these results entirely satisfactory?

- The result does not say anything about the relative performance of a robust solution x_{rob} with respect to a **nominal** optimal solution obtained by ignoring uncertainty. e.g., by taking an expected (average) value for c and solving the resulting problem.
- How does the BS robust solution fare compared to the minimax regret solution in the case of e.g., minimum spanning tree, shortest path problems etc..?

Adjustable Robustness

- Consider a multi-period optimization problem with uncertain parameters where uncertainty is revealed progressively through periods.
- A subset of the decision variables can be chosen after observing realizations of some uncertain parameters.
- This allows to correct the earlier decision made under a smaller information set.
- In stochastic programming, this is called "recourse".
- In robust optimization it is called "Adjustable Robust Optimization" (ARO).
- Due to Ben-Tal, Guslitzer and Nemirovski

Adjustable Robustness: what is it?

- Consider the two-stage linear optimization problem

$$\min_{x^1,x^2}\{c^T x^1 : A^1 x^1 + A^2 x^2 \le b\}$$

  where x^1 is the first stage variables, and x^2 are the second-stage variables, A^1, A^2, b are uncertain parameters.

- Assume that A^1 is revealed to the modeler after choosing x^1. So, at the moment of choosing x^2 the modeler knows the value of A^1.
- Could we make robust choices of x^2 taking this sequential nature of the decision process into account?

Adjustable Robustness (cont'd)

- Let $\mathcal{U}$ denote the *uncertainty* set for parameters A^1, A^2, b. The robust counterpart is (RC)

$$\min_{x^1}\{c^T x^1 : \exists x^2 \forall (A^1, A^2, b) \in \mathcal{U} : A^1 x^1 + A^2 x^2 \le b\}$$

- Here notice that the choice of x^2 is independent of the realized values of the uncertain parameters.
- This ignores the multi-stage nature of the decision process.

Adjustable Robustness: what is it?

- The ARO formulation is (ARC):

  $$\min_{x^1}\{c^T x^1 : \forall (A^1, A^2, b) \in \mathcal{U} \exists x^2 = x^2(A^1, A^2, b) : A^1 x^1 + A^2 x^2 \le b\}$$

- The feasible set of the second problem (ARO) is larger than the feasible set of the RC.
- ARO is therefore less conservative compared to RC
- However, ARO is harder to formulate explicitly as an easy optimization problem e.g., linear programs.
- Because we do not know the functional form of the dependency.
- Only very simple uncertainty sets allow "nice" ARCs. Otherwise, we have to assume a simple functional form for dependencies, e.g., affine dependency

An Example from Network Design

- Ordoñez and Zhao considered the following network capacity expansion problem subject to a budget constraint:

  $$\min_{x \ge 0, y \ge 0}\{c^T x : Nx = b, x \le u + y, d^T y \le I\}$$

  where the vector $y \in \mathbf{R}^m$ denotes capacity expansion decisions, $x \in \mathbf{R}^m$ the flow variables
- The constraints $Nx = b$ represent the flow balance equations, c represents the transportation cost coefficients, and d the cost of incremental unit capacity.
- Assume c and b are uncertain, i.e., $c \in \mathcal{U}_c$ and $b \in \mathcal{U}_b$, where $\mathcal{U}_c$ and $\mathcal{U}_b$ are suitable (closed, bounded, convex) uncertainty sets.

- The Adjustable Robust Counterpart is

$$z_{ARC} = \min_{y,\gamma} \gamma$$

subject to

$$d^T y \leq I, y \geq 0,$$

$$\forall c \in \mathcal{U}_c,\ b \in \mathcal{U}_b\ \exists x : \begin{cases} Nx = b \\ 0 \leq x \leq u + y \\ c^T x \leq \gamma \end{cases}$$

- Ordoñez and Zhao proved :

$$z_{ARC} = \min_{\substack{y \geq 0 \\ d^T y \leq I}} \max_{\substack{c \in \mathcal{U}_c \\ b \in \mathcal{U}_b}} \min_{\substack{Nx = b \\ 0 \leq x \leq u+y}} c^T x$$

Takeda, Taguchi and Tütüncü (T^3) prove essentially the same result, but in more general form. T^3 also prove other results giving conditions for tractability of the ARC.

Negative result

- Unfortunately, the problem

$$z_{ARC} = \min_{\substack{y \geq 0 \\ d^T y \leq I}} \max_{\substack{c \in \mathcal{U}_c \\ b \in \mathcal{U}_b}} \min_{\substack{Nx = b \\ 0 \leq x \leq u+y}} c^T x$$

is intractable

- because the inner problem

$$\max_{\substack{c \in \mathcal{U}_c \\ b \in \mathcal{U}_b}} \min_{\substack{Nx = b \\ 0 \leq x \leq u+y}} c^T x$$

is equivalent to

- $$\max_{\substack{c\in\mathcal{U}_c, b\in\mathcal{U}_b \\ \lambda,\pi}} \{b^T\lambda - (u+y)^T\pi : N^T\lambda - \pi \le c, \pi \ge 0\}$$

  with the bilinear term $b^T\lambda$ in b and λ, therefore, not concave!
- Unless $\mathcal{U}_b$ is a singleton, i.e., b is not uncertain.
- But, in that case, $z_{ARC} = z_{RC}$.

Another Negative Result

- Assume that

  $$\mathcal{U}_c = \{\bar{c} + P_c u : \|u\|_2 \le 1\}$$

  and

  $$\mathcal{U}_b = \{\bar{b} + P_b u : \|u\|_2 \le 1\}$$

  are ellipsoidal uncertainty sets
- Now, consider the inner problem rewritten as

  $$\max_{\lambda,\pi} \max_{c\in\mathcal{U}_c} \max_{b\in\mathcal{U}_b} \{b^T\lambda - (u+y)^T\pi : N^T\lambda - \pi \le c, \pi \ge 0\}$$
- Then replace the inner max over b with min:

  $$\max_{\lambda,\pi} \max_{c\in\mathcal{U}_c} \min_{b\in\mathcal{U}_b} \{b^T\lambda - (u+y)^T\pi : N^T\lambda - \pi \le c, \pi \ge 0\}$$

  This is a relaxation

- which is equivalent to

$$\max_{\lambda,\pi} \max_{c \in \mathcal{U}_c} \{\bar{b}^T \lambda - \|P_b \lambda\|_2 - (u+y)^T \pi : N^T \lambda - \pi \leq c, \pi \geq 0\}$$

- Take now the dual of the resulting problem to obtain the relaxation:

$$\min_{x \geq 0, y \geq 0, \|v\|_2 \leq 1} \{\bar{c}^T x + \|P_c x\|_2 : Nx = \bar{b} + P_b v, x \leq u + y, d^T y \leq I\}$$

- but this is nothing other than the RC applied to the problem of capacity expansion.

A Most Recent Approach: Comprehensive Robust Counterpart

Back to the origins of RC methodology:

$$\begin{array}{rcl} \min & & F_0(\chi, \zeta) \\ \text{subject to} & & \\ F_i(\chi, \zeta) & \geq & 0 \end{array} \qquad \text{(P23)}$$

- χ is the vector of variables
- ζ is the vector of data (uncertain)
- Let $\mathcal{U}$ represent the uncertainty set for ζ.

Comprehensive Robust Counterpart

- Recall the RC:

$$\min_{\chi}\left\{\sup_{\zeta\in\mathcal{U}} F_0(\chi,\zeta) : F_i(\chi,\zeta)\le 0, \forall i, \forall \zeta\in\mathcal{U}\right\}$$

- Recall that the ARC that allowed the decision variables (at least a portion of them) to depend on the realizations of the data:

$$\min_{\chi(\cdot)}\left\{\sup_{\zeta\in\mathcal{U}} F_0(\chi(\zeta),\zeta) : F_i(\chi(\zeta),\zeta)\le 0, \forall i, \forall \zeta\in\mathcal{U}\right\}$$

  Recall that this is intractable in general, unless $\chi(\zeta)$ is assumed to be e.g., affine.

Comprehensive Robust Counterpart

- Apart from intractability, the RC and ARC have another drawback.
- They only protect for the events represented in the uncertainty set $\mathcal{U}$. They do not, except some special cases, provide any performance guarantees for those events not represented in the uncertainty set $\mathcal{U}$.
- The CRC tries to address this final issue (but assumes affine functional dependency)
- Treat the uncertainty set $\mathcal{U}$ as the "normal range" of the data

Comprehensive Robust Counterpart

An affinely adjustable solution $\chi(.)$ is acceptable in CRC if

- When $\zeta \in \mathcal{U}$ the solution must satisfy the constraints $F_i(\chi(\zeta), \zeta) \leq 0, \forall i$,
- For all data ζ, the violations of the constraints should not exceed a prescribed multiple of the deviation of the data from its normal range:

  $$\forall \zeta \in \mathbb{R}^{n_\zeta} : \text{dist}(F_i(\chi(\zeta), \zeta), \mathbb{R}_-) \leq \alpha_i \text{dist}(\zeta, \mathcal{U}), \forall i$$

  where dist is given by $\text{dist}(a, A) = \inf_{z \in A} \|a - z\|$.
- Ben-Tal, Boyd, and Nemirovski, develop this fully, and apply it to an optimal control problem in linear dynamical systems.

Challenges in Comprehensive Robust Counterpart

- How can we apply this methodology to other problems?
- We have to understand the details.
- A natural application area is multi-period financial optimization problems
- Any multi-period planning problem under uncertainty may be the subject of CRC.

Applications of Robust Optimization

Applications Covered

- Robust Optimization in Finance
- Robust Optimization in Macroeconomics
- Robust Optimization in Telecommunications
- Robust Optimization in Engineering
- Robust Optimization in Supply Chain/Inventory Management

Robust Optimization Applied to Finance

First, a piece of wisdom..

> Si j'avais été riche, je n'aurais jamais consacré ma vie aux mathématiques.
>
> Joseph-Louis Lagrange (1736-1813)

Robust Optimization Applied to Finance

- Before we begin to review contributions in this area, we have to warn the potential user of the dangers that lie ahead!
- Applying the Robust Counterpart approach is an art.
- It is important to formulate the problem in a way suitable for the application of Robust Counterpart technique
- Otherwise, the RC may be useless (if not intractable)

Extremely Simple Example from Finance

- Single period investment problem: the investor has 1 USD in cash and no other assets
- Distribute part of this cash between n assets in order to maximize the value of the resulting portfolio at the end of the period
- In other words, solve the problem

$$\max_x \{y | y \leq \sum_{i=1}^n r_i x_i, \sum_{i=1}^n x_i \leq 1, x \geq 0\}$$

where the r_i denotes the one period return of asset i.

Example (cont'd)

- Assume that the r_i are log-normal independent random variables with expectations ρ_i and standard deviations σ_i, all these quantities being of the same order of magnitude:

$$\rho_i, \sigma_i \in [1/\kappa, \kappa]$$

where $\kappa > 1$ is fixed.

- The RC applied to this problem yields

$$\max_x \{\sum_{i=1}^n \rho_i x_i - \theta \sqrt{\sum_{i=1}^n \sigma_i^2 x_i^2} | x \geq 0, \sum_{i=1}^n x_i \leq 1\}$$

Example (cont'd)

- For fixed small $\alpha > 0$ one can choose $\theta = \theta(\alpha)$ in such a way that, uniformly in n and in x the probability of the event

$$\{\sum_{i=1}^{n} r_i x_i < \sum_{i=1}^{n} \rho_i x_i - \theta \sqrt{\sum_{i=1}^{n} \sigma_i^2 x_i^2}\}$$

will be less than α.

- What does this mean?

Example (cont'd)

- It means "the actual portfolio value yielded by the optimal robust portfolio will be less than the optimal value of the RC with small probability (less than α)
- It can be shown that the optimal value of RC for large n is at least

$$(1 - o(1))\frac{1}{n}\sum_{i=1}^{n} \rho_i$$

- Therefore, for large n, the RC yields with probability at least $1 - \alpha$ a final value of $O(1)$.

Example (cont'd)

- But we could have equally formulated the initial problem as

$$\max_x \{\sum_{i=1}^n y_i | 0 \le y_i \le r_i x_i, \sum_{i=1}^n x_i \le 1, x \ge 0\}$$

- and applied the RC to obtain

$$\max_x \{\sum_{i=1}^n y_i | 0 \le y_i \le (\rho_i - \theta\sigma_i) x_i, \sum_{i=1}^n x_i \le 1, x \ge 0\}$$

- But, for $\theta \ge \kappa^2$, the optimal solution here is the portfolio $x_i = 0, i = 1, \ldots, n$!
- This means no investment to assets is made, the 1 USD cash is kept.
- Totally useless!

Single Period Robust Portfolio Selection

- Consider an investor who wants to allocate his/her funds among a set of asset classes $S_1, S_2, \ldots, S_n$.
- The returns $r_j, j = 1, \ldots, n$ from these asset classes at the end of an investment period are random with expected value μ_j and standard deviation σ_j.
- Let x_j denote the proportion of the investor's money to be allocated to asset class j.
- The values of x_j will be determined from an optimization problem.
- First, let us find the expressions for the expected return and variance of the portfolio return for a portfolio $x = (x_1, \ldots, x_n)$.

Robust Portfolio Selection II

- The expected return of the portfolio $x = (x_1, \ldots, x_n)$ is

$$\mathbb{E}(x) = \sum_j \mu_j x_j = \mu^T x.$$

- The variance of the return of the portfolio is

$$Var(x) = \sum_j \sigma_j^2 x_j^2 + 2 \sum_{i<j} \sigma_{ij} x_i x_j = x^T \Sigma x$$

  where $\Sigma_{ij} = \sigma_{ij}$ and $\Sigma_{jj} = \sigma_j^2$.

- We have to deal with competing objectives: it is desirable to find the portfolio that minimizes the variance of the return and maximize the expected return at the same time.

- Represent the constraints defining the portfolio by $x \in \mathcal{X}$
- we are dealing now with the Markowitz problem that can be presented in various forms:

$$\begin{array}{rcl} \max & \mu^T x & \\ \text{subject to} & & \\ x & \in & \mathcal{X} \\ x^T \Sigma x & \leq & \sigma^2 \end{array} \qquad \text{(P24)}$$

  where σ^2 is a limit on tolerable variability of return

- or,

$$\begin{array}{rcl} \max & x^T \Sigma x & \\ \text{subject to} & & \\ x & \in & \mathcal{X} \\ \mu^T x & \geq & R \end{array} \qquad \text{(P25)}$$

where R is minimum target return

- or,

$$\begin{array}{rcl} \max & & \mu^T x - \lambda x^T \Sigma x \\ \text{subject to} & & \\ & x \in \mathcal{X} & \end{array} \qquad \text{(P26)}$$

where λ is a non-negative scalar representing the risk aversion level of the investor.

- All three models are equivalent under mild conditions (see PhD thesis by F. Lutgens).
- However, usually we never know μ and Σ exactly. We must estimate them from historical data. But, experience shows that the composition of optimal portfolio is very sensitive to perturbations in μ (and in Σ, but to a lesser extent)

Robust Portfolio Selection Models

- Define an appropriate uncertainty set $\mathcal{U}$ and try to formulate a robust counterpart to the portfolio selection problem under parameter uncertainty
- I.e., have to deal with the problem

$$\max_{x \in \mathcal{X}} \min_{(\mu,\Sigma) \in \mathcal{U}} \mu^T x - \lambda x^T \Sigma x$$

- Typical question of robust optimization: couple "representative of reality" uncertainty set with solvable (an explicit convex optimization problem) robust counterpart

How to deal with the last question?

There is some literature:

1. Rüstem, Settergren, Gülpınar etc..
2. Tütüncü and Koenig
3. Goldfarb and Iyengar
4. Ceria and Stubbs
5. Nice overview in PhD thesis by F. Lutgens (available on the Web)

Scenario-based Models (Rüstem et al.)

- We assume given a finite set of possible scenarios $\mu^1, \ldots, \mu^l$ and $\Sigma^1, \ldots, \Sigma^J$.
- In the return maximizing model the robust problem is:

$$\begin{array}{rcl} \max & & \min_i (\mu^i)^T x \\ \text{subject to} & & \\ x & \in & \mathcal{X} \\ \max_j x^T \Sigma^j x & \leq & \sigma^2 \end{array} \qquad \text{(P27)}$$

- It is easy to transform this into an equivalent convex optimization problem

Scenario-based Models cont'd

- The result is

$$
\begin{array}{rrcl}
\max & & t & \\
\text{subject to} & & & \\
& t & \leq & (\mu^i)^T x, \; i = 1, \ldots, l \\
& x & \in & \mathcal{X} \\
& x^T \Sigma^j x & \leq & \sigma^2, \; j = 1, \ldots, J
\end{array}
\tag{P28}
$$

- Critical question: How to generate the scenarios?

Interval-based Models

- Tütüncü and Koenig (2004) use the following uncertainty sets:

$$\mathcal{U} = \{(\mu, \Sigma) : \mu^L \leq \mu \leq \mu^U, \Sigma^L \leq \Sigma \leq \Sigma^U, \Sigma \succeq 0\}.$$

- These sets can be obtained as confidence intervals from historical data
- The robust optimization model in this case:

$$\max_{x \in \mathcal{X}} \min_{(\mu,\Sigma) \in \mathcal{U}} \mu^T x - \lambda x^T \Sigma x$$

- No further simplification in the formulation
- The model is solved using an interior point algorithm

Interval-based Models: Observations

- Robust efficient portfolios have stable behavior under unfavorable market conditions (as expected)
- Slight deterioration in performance for average market conditions. Under favorable market less conservative efficient portfolios might be preferable
- The model is computationally intensive. Too many constraints. Limited to small data sets.

Models based on Ellipsoidal Uncertainty

- Goldfarb and Iyengar (MOR, 2003) consider a factor model of returns

$$r = \mu + V^T f + \varepsilon$$

where

- μ is the mean return
- f is the random vector representing market factors
- V is a factor loading matrix
- ε represents residuals

Models based on Ellipsoidal Uncertainty II

- Motivation: if these parameters are obtained from time series data via linear regression confidence regions around the least squares estimate have the following structures:

$$S_v = \{V : V = V_0 + W : \|W_i\|_g \leq \rho_i, i = 1, \ldots, n\},$$

  where $\|u\|_g = \sqrt{u^T G u}$.
- These are ellipsoidal uncertainty sets
- The robust portfolio selection problem is transformed into a convex Second-order Cone optimization problem
- Details are involved, but not hard to follow.

Multi-Period Portfolio Selection

Let us look at a typical model:

- Assume, for simplicity, a time horizon of 2 periods
- Initial budget: $1, no short positions.
- $r^1(\omega), r^2(\omega) \in \mathbf{R}_+^{m+1}$: vectors of random variables representing returns for periods 1 and 2, respectively.
- $x^0 \in \mathbf{R}_+^{m+1}$: portfolio vector at the beginning of period 1 x_i^0 dollar value of asset i in the initial portfolio, x_{m+1}^0 is the cash.
- $x^1 \in \mathbf{R}_+^{m+1}$: portfolio vector at the beginning of period 2

Multi-Period Formulation (cont'd)

- Assume the return of risk-less asset known with certainty.
- Include proportional transaction costs for buying and selling, μ and ν.
- Buy-variables: $y^1 \in \mathbf{R}^m_+$
- Sell-variables: $z^1 \in \mathbf{R}^m_+$

Asset Dynamics

- For all risky assets $i = 1, \ldots, m$:

$$x^1_i = r^1_i x^0_i + y^1_i - z^1_i,$$

- For the risk-less asset (cash) :

$$x^1_{m+1} = r^1_{m+1} x^0_{m+1} - \sum_{i=1}^{m} (1 + \mu^1_i) y^1_i + \sum_{i=1}^{m} (1 - \nu^1_i) z^1_i.$$

- The initial budget equation is :

$$\sum_{i=1}^{m} (1 + \mu^0_i) x^0_i = 1 - x^0_{m+1}.$$

- Objective: maximize final wealth

$$\max (r^2)^T x^1.$$

Robust Formulation

- The above formulations are mathematically non-sense because r^1 and r^2 are random vectors.
- One can pass to a Robust Counterpart formulation after obtaining first and second moment information
- But this will not be really useful, because in each constraint we have a single uncertain coefficient
- Furthermore, the RC approach ignores the sequential nature of the decision process: decide on x^1 after observing r^1.
- Ben-Tal, Margalit and Nemirovski pass to an equivalent formulation where each constraint has more than one uncertain coefficient and apply RC to that formulation. They obtain a second-order cone program.

Robust Formulation: Open Question

- The problem is the kind of problem where ARC is needed.
- Unfortunately, ARC is intractable even for polytopic uncertainty sets.
- Can you try to formulate ARC for the above problem?

Robust Risk Management

- Financial institutions are required, by law, to measure and manage their risks.
- The best known and widely accepted risk measure in the financial industry is Value-at-Risk (VaR)
- VaR is a measure obtained using the percentiles of loss distributions
- VaR represents the predicted maximum loss with a specified probability level (e.g., 95%) over a certain period of time
- VaR is not sub-additive. It does not yield convex optimization problems.

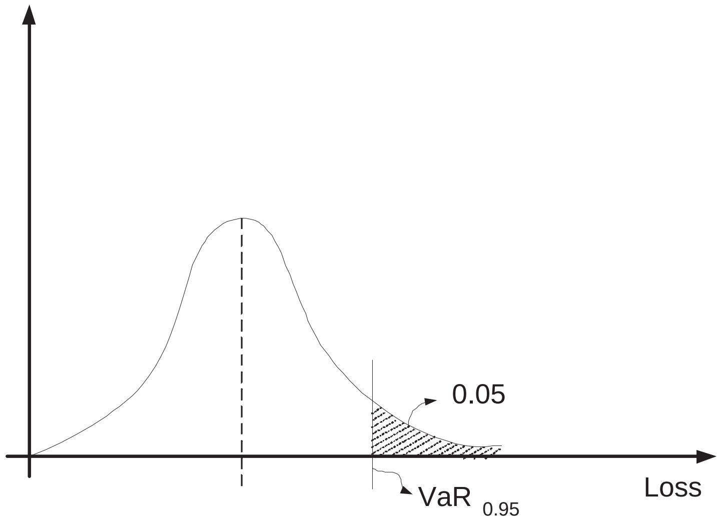
Value-at-Risk

- Consider a random variable X representing loss from an investment portfolio over a fixed period of time. A negative value indicates gains.
- Given a probability (confidence) level α, VaR_α of X is given by

$$VaR_\alpha(X) := \min\{\gamma : P(X \leq \gamma) \geq \alpha\}$$

- For fixed α you want the probability of the loss exceeding some γ to be at most $1 - \alpha$.
- Among all such γ, find the smallest one.

Value-at-Risk illustrated



Robust Value-at-Risk

- We rarely know exactly the distribution of losses
- We may have moment estimates, i.e., means and covariances
- Robust VaR: Among all distributions for loss with a fixed mean and covariance, which one gives the worst VaR?
- Therefore, the worst-case VaR at level α:

$$\min\{\gamma : \inf_q P_q(X \leq \gamma) \geq \alpha\}$$

  where the inf is taken over all distributions q with fixed mean and covariance
- El-Ghaoui *et al.* (O.R. 2003) formulate and solve the problem using Lagrange duality on the space of probability distributions and a well-known result from Control Theory, S-Lemma.

Robust VaR

- The resulting problem is convex optimization problem (in fact a linear optimization problem in the space of symmetric positive semidefinite matrices).
- They also extend the result to cases where the mean and covariances are not necessarily fixed but lie in an uncertainty set (in a polytopic set, or ellipsoidal set, in a factor model type uncertainty set)

Robust Optimization in Macroeconomics

- Based on the work of M. Giannoni (2000).
- Problem: Monetary policy design under parameter uncertainty
- A simple forward-looking macro-economic model
- The model is composed of a monetary policy rule and two structural equations: an inter-temporal IS equation and and aggregate supply equation.
- Important property of the model: the policymaker faces a trade-off between the stabilization of inflation and the output gap on one hand, and the nominal interest rate on the other hand.

Monetary Policy Design Model

- Two structural equations:
- The first, the inter-temporal IS equation, which relates spending decisions to the interest rate:

$$x_t = E_t x_{t+1} - \frac{1}{\sigma}(i_t - E_t \pi_{t+1} - r_t^n)$$

and the second aggregate supply equation

$$\pi_t = \kappa x_t + \beta E_t \pi_{t+1},$$

where x_t is the output gap defined as the deviation of output from its natural level, i.e., the equilibrium level of output under flexible prices, π_t is the inflation rate, i_t is the deviation of the short-term interest rate from its steady-state value.

Monetary Policy Design Model (cont'd)

- The composite exogenous disturbance r_t^n represents "natural interest rate", i.e., the real interest rate that equates output to its natural level, or alternatively the interest rate that would prevail in equilibrium under flexible prices.
- κ, σ, β are positive parameters.
- σ is the inverse of inter-temporal elasticity of substitution
- κ which is the slope of the short run aggregate supply curve, can be seen as the speed of price adjustment
- $\beta \in (0, 1)$ is a time discount factor of price-setters (assumed to be equal to that of a representative household).

Monetary Policy Design Model (cont'd)

Recall the first equation:

$$x_t = E_t x_{t+1} - \frac{1}{\sigma}(i_t - E_t \pi_{t+1} - r_t^n)$$

- $i_t - E_t \pi_{t+1}$ represents the real interest rate
- Therefore, $(i_t - E_t \pi_{t+1}) - r_t^n$ is an interest rate gap between real rate and the natural rate
- Perturbations to the natural rate of interest represent all non-monetary disturbances affecting inflation and output gap
- E.g., a temporary increase in r_t^n could reflect a temporary exogenous increase in aggregate demand, or alternatively a temporary decrease in the natural level of output

Monetary Policy Design Model (cont'd)

- Moreover, non-monetary perturbations affect inflation and output gap only if the interest rate controlled by the Central Bank is such that the interest rate gap is non-zero.
- The first equation can be viewed a log-linear approximation to the representative Household's Euler equation for optimal timing of consumption in the presence of complete financial markets
- The second equation can be viewed as a log-linear approximation to the first-order condition for the supplier's optimal price-setting decision.
- The model is known to provide an accurate description of the actual behavior of inflation, output, and the average Federal funds rates.

Optimal Monetary Policy

- Question: what is the objective of monetary policy?
- Traditionally, policymakers seek to minimize a weighted average of some measure of variability of inflation and of the output gap.
- In the model above, set $i_t = r_t^n$ in every period. This perfectly stabilizes inflation and output gap, so that interest rate perfectly tracks the exogenous fluctuations in the natural rate of interest.
- However, it may be undesirable to vary the interest rate as much as the natural rate of interest (Friedman (1969)).
- Desirable to reduce both the level and the variability of nominal interest rates

Optimal Monetary Policy

- Use the loss criterion

$$L_0 = E_0\{(1-\beta)\sum_{t=0}^{\infty}\beta^t[\pi_t^2 - \lambda_x x_t^2 + \lambda_i i_t^2]\}$$

  where λ_x, λ_i are weights placed on the stabilization of the output gap and the the nominal interest rate.
- This can be viewed as a second-order Taylor approximation to the expected utility of the representative household in the model
- An implication of this loss criterion is that an equilibrium with complete stabilization of inflation and the output gap is not fully efficient.

Optimal Monetary Policy (cont'd)

- What is monetary policy?
- Interest-rate feedback rules.
- Assume the policymaker commits credibly at the beginning of period 0 to a policy rule of the form:

$$i_t = \mathcal{P}_t(\pi_t, \pi_{t-1}, \ldots, x_t, x_{t-1}, i_{t-1}, i_{t-2}, \ldots, r_t^n, r_{t-1}^n, \ldots)$$

  for each date $t \geq 0$.
- Policymaker's problem: determine the functions $\mathcal{P}_t(.), t = 0, 1, 2, \ldots$ to minimize the loss $E[L_0]$ subject to the structural equations.

A Dynamic Optimization Problem

- As the objective is quadratic and constraints are linear in all variables, we restrict our attention to linear functions $\mathcal{P}_t(.)$
- Denote by ϕ the vector of coefficients that characterize $\{\mathcal{P}_t(.)]\}_{t=0}^{\infty}$
- ϕ is called a "policy rule".
- Policy rules are drawn from a polyhedral set $\tilde{\Phi} \subseteq \mathbb{R}^n$
- Let $\theta = [\theta_1, \theta_2, \ldots, \theta_m]^T$ denote the vector of structural parameters of the model, taking values in a set $\Theta \subseteq \mathbb{R}^m$.
- Let $q_t = [\pi_t, x_t, i_t]'$ denote the vector of endogenous variables at date t, and q for the stochastic process $\{q_t\}_{t=0}^{\infty}$ specifying q_t at each date as a function of the history of exogenous shocks until that date.

A Dynamic Optimization Problem (cont'd)

- To be feasible the stochastic process q needs to satisfy the structural equations at all dates t.
- Re-write them compactly as

$$\tilde{\mathbf{S}}(q, \theta) = \mathbf{0}.$$

- The restrictions imposed by the policy rule commitment at all dates $t \geq 0$ are written

$$\tilde{\mathbf{P}}(q, \phi) = \mathbf{0}.$$

- Assume both $\tilde{\mathbf{S}}, \tilde{\mathbf{P}}$ are linear in q.
- The loss function can be denoted now $L_0(q, \theta)$

A Dynamic Optimization Problem (cont'd)

- A rational expectations equilibrium is a stochastic process $q(\phi, \theta)$ satisfying the above equations
- **Definition** In the case of known structural parameters θ, let $\Phi \subseteq \tilde{\Phi}$ be a set of policy rules such that there is a unique bounded equilibrium. Then an **optimal monetary policy rule** is a vector that solves

$$\min_{\phi \in \Phi} E[L_0(q(\phi, \theta), \theta)]$$

  where the conditional expectation is taken over all possible histories of the disturbances $\{r_t^n\}$.
- Potential problem with this approach: parameters are supposed to be constant and known by all economic agents. The only uncertainty in the model is due to exogenous perturbations r_t^n.

Monetary Policy with Model Uncertainty

- The model is completely unrealistic with respect to the previous remark
- Assume the vector θ of structural parameters lies in a given known compact set $\Theta \subset \mathbb{R}^m$, and that the distribution of θ is unknown.
- **Definition** Let Φ be a set of policy rules such that there is a unique bounded equilibrium process $q(\phi, \theta)$ for all $\phi \in \Phi, \theta \in \Theta$. Then an **robust optimal monetary policy rule** is a vector that solves
$$\min_{\phi \in \Phi} \left\{ \max_{\theta \in \Theta} E[L_0(q(\phi, \theta), \theta)] \right\}$$
where the conditional expectation is taken over all possible histories of the disturbances $\{r_t^n\}$.

Monetary Policy with Model Uncertainty

- This is our usual min-max framework.
- The rest of the paper is quite technical, but accessible.
- See the journal Macro-economic Dynamics, special issue on Robustness in Economics (2001).

Robust Optimization Applied to Telecommunications

- Design of Networks under Aggregate Traffic Uncertainty
- Problem motivated by the advent of Virtual Private Networks (VPN). A multi-million dollar industry...
- Joint work with Amaldi, Belotti and Altın
- Originated from the work of Gupta et al., with contributions from Oriolo, Erlebach, Rüegg, Yener, Kleinberg, Rastogi, Roughgarden etc..

VPN Design under Uncertainty

- Network design problem
- Reserve capacity (bandwidth) in a network so as to support a given set of pairwise traffic demands
- Reserving bandwidth is costly, therefore
- Seek the least costly bandwidth reservation.
- Multicommodity network design
- Previous work forms a vast body of literature..

Network Capacity Allocation Problem

Given:

- a graph $G = (V, E)$ with per-unit reservation cost c_e for each $e \in E$;
- a subset $Q \subseteq V$ of terminals
- a traffic matrix D between terminals, i.e., for each (ordered) pair of terminals $u, v \in Q$, the value (bandwidth) d_{uv} of the expected demand

Find:

A minimum cost reservation vector x such that the network G, equipped with capacity x_e for all edges $e \in E$, is able to support the traffic matrix D.

Observations

- Problem is easy if traffic matrix D is known in advance.
- Several applications where communication patterns may change over time render this assumption questionable at best.
- Therefore, we have to consider situations where we are not given a known traffic matrix D.
- Duffield et al. proposed a model of uncertainty representation for traffic, the *hose model*.
- In the hose model, we are given bounds on the cumulative amount of traffic each terminal can send and receive,
- and we must design the network so as to support any traffic matrix respecting these bounds.

Hose Network Capacity Allocation Problem

Asymmetric Case

Given:

- a graph $G = (V, E)$ with per-unit reservation cost c_e for each $e \in E$;
- a subset $Q \subseteq V$ of terminals
- for each terminal $v \in Q$, a couple of upper bounds b_v^+ and b_v^- such that

$$\sum_{u \in Q, u \neq v} d_{uv} \leq b_v^-$$

and

$$\sum_{u \in Q, u \neq v} d_{vu} \leq b_v^+$$

Find:

A minimum cost reservation vector x such that the network G, equipped with capacity x_e for all edges $e \in E$, is able to support any valid traffic matrix D respecting the above cumulative bounds.

Hose Network Capacity Allocation Problem

Symmetric Case A single upper bound on the cumulative amount of traffic a terminal can send or receive.

Given:

- a graph $G = (V, E)$ with per-unit reservation cost c_e for each $e \in E$;
- a subset $Q \subseteq V$ of terminals
- for each terminal $v \in Q$, upper bound b_v such that

$$\sum_{u \in Q, u \neq v} d_{vu} \leq b_v$$

Find:

A minimum cost reservation vector x such that the network G, equipped with capacity x_e for all edges $e \in E$, is able to support any valid traffic matrix D respecting the above cumulative bounds.

- We gave the first compact, solvable numerically, formulations of this problem.

Formulation

Variables:

- Nonnegative variable x_e representing the amount of bandwidth reserved on edge $e = (i, j)$.
-
$$y_{ij}^{st} = \begin{cases} 1 & \text{if demand pair (s,t) uses edge (i,j)} \\ 0 & \text{otherwise} \end{cases}$$

Known Parameters:

- Nonnegative upper bound b_s for all $s \in Q$
- Positive bandwidth costs c_e for each edge $e \in E$.

Uncertain parameter:

d_{st}: traffic demand between terminals s and t such that

$$\sum_{t \in Q, s \neq t} d_{st} \leq b_s.$$

How would we formulate the problem if we assumed d_{st}'s are known for the moment?

$$\text{Minimize} \sum_{ij} c_{ij} x_{ij}$$

subject to

(path constraints for all $i \in V$ and ordered terminal pair (s,t) with $s < t$, $s,t \in Q$):

$$\sum_j y_{ij}^{st} - \sum_j y_{ji}^{st} = \begin{cases} 1 & i = s \\ -1 & i = t \\ 0 & \text{o.w.} \end{cases}$$

(capacity constraints for all edges (i,j)):

$$\sum_{(s,t)} d_{st} y_{ij}^{st} \leq x_{ij}$$

- Fix (i,j) and consider the capacity constraint:

$$\sum_{(s,t)} d_{st} y_{ij}^{st} \leq x_{ij}$$

- Assume we have fixed y_{ij}^{st} variables for the moment and consider the following linear program over the variables d_{st}

$$\text{Maximize} \sum_{(s,t)} d_{st} y_{ij}^{st}$$

subject to

$$\sum_{t \in Q, s \neq t} d_{st} \leq b_s \forall s \in Q$$

$$d_{st} \geq 0, \forall (s,t)$$

- The dual of this problem over variables w_s^{ij} is:

$$\text{Minimize} \sum_s b_s w_s^{ij}$$

subject to

$$w_s^{ij} + w_t^{ij} \geq y_{ij}^{st} + y_{ji}^{st} \forall (s,t)$$
$$w_s^{ij} \geq 0, \forall s \in Q$$

- Now, replace each "uncertain" capacity constraint

$$\sum_{(s,t)} d_{st} y_{ij}^{st} \leq x_{ij}$$

by the constraints

- $$\sum_s b_s w_s^{ij} \leq x_{ij}$$
$$w_s^{ij} + w_t^{ij} \geq y_{ij}^{st} + y_{ji}^{st} \ \forall (s,t)$$
$$w_s^{ij} \geq 0, \quad \forall s \in Q$$

- Notice that we got rid of the "min" since we are minimizing the weighted sum of bandwidths.
- Therefore, we obtain the following mixed-integer linear programming formulation for *SymG*.

$$\text{Minimize} \sum_{ij} c_{ij} x_{ij}$$

subject to

(path constraints for all $i \in V$ and ordered terminal pair (s,t) with $s < t$, $s,t \in Q$):

$$\sum_j y_{ij}^{st} - \sum_j y_{ji}^{st} = \begin{cases} 1 & i = s \\ -1 & i = t \\ 0 & \text{o.w.} \end{cases}$$

$$\sum_{s \in Q} b_s w_s^{ij} \leq x_{ij} \ \forall (i,j)$$

$$w_s^{ij} + w_t^{ij} \geq y_{ij}^{st} + y_{ji}^{st} \ \forall (s,t)$$

$$w_s^{ij} \geq 0, \quad \forall s \in Q$$

and the binary requirements on the y's and non-negativity on x's.

Observations

- Both the symmetric and asymmetric cases are solved easily by off-the-shelf MIP solvers (CPLEX) for small to medium instances
- The asymmetric problem is harder to solve in general
- For large instances, a Branch-and-price and cutting plane (BPC) algorithm works very well
- The Bertsimas-Sim type uncertainty leads to NP-hard problem
- The corresponding robust formulation is solved very efficiently by the BPC algorithm while it creates difficulties for CPLEX.

Application of Robust Optimization to Least Squares Problems

- For an inconsistent linear system $Ax = b$, the least squares solution x_{LS} solves

$$
\begin{aligned}
\min_x \|Ax - b\|_2 \quad &\Leftrightarrow \quad \min_x (Ax-b)^T(Ax-b) \\
&\Leftrightarrow \quad \begin{array}{ll} \min_{x,\Delta b} & \|\Delta b\|_2 \\ \text{s.t.} & Ax = b + \Delta b \end{array} \\
&= \quad r(A,b)
\end{aligned}
$$

- Interpretation: Approximate the vector b by a linear combination of the columns of A, also called the regression problem

Other interpretation

- **Estimation:** Estimate a parameter vector on the basis of an imperfect linear vector measurement
- Consider a linear measurement model:

$$y = Ax + v$$

  where $y \in \mathbb{R}^m$ is a vector measurement, $x \in \mathbb{R}^n$ is a vector of parameters to be estimated, and $v \in \mathbb{R}^m$ is some unknown measurement error but, presumed to be small
- . Problem: find the most plausible guess x, given y
- The most plausible guess for x may be found by

$$\min_z \|Az - y\|_2$$

Typical Application in Engineering

- Optimal input design: Consider a dynamical system with scalar input sequence $u(0), u(1), \ldots, u(N)$, and scalar output sequence $y(0), y(1), \ldots, y(N)$ related by convolution:

$$y(t) = \sum_{\tau=0}^{t} h(\tau)u(t-\tau), t = 0, 1, \ldots, N.$$

- The sequence $h(0), h(1), \ldots, h(N)$ is called the *convolution kernel* or *impulse response* of the system
- **Output tracking**: The output y should track, or follow, a desired target or reference signal y_{des}. Therefore, solve

$$\min_h \frac{1}{N+1} \sum_{t=0}^{N} (y(t) - y_{\text{des}})^2$$

Errors-in-Variables

- The approach above assumed the errors are confined to observations
- If the errors are also present in the model data A, the following problem is solved in Engineering applications
- The total least squares solution, x_{TLS} solves

$$\begin{aligned} r_T(A, b) = \quad & \min_{x,\Delta A,\Delta b} & \|(\Delta A, \Delta b)\|_F \\ & \text{s.t.} & (A + \Delta A)x = b + \Delta b \end{aligned}$$

- Solved by an application of SVD (Singular Value Decomposition)

Robust Least Squares

- In the robust setting we assume that the data (A, b) is uncertain, and $(A, b) \in \mathcal{U}$.
- Write down the worst residual for a given solution x, and the robust least squares problem as to find the solution with best worst residual:

$$\min_x \max_{A,b \in \mathcal{U}} \|Ax - b\|_2$$

- Consider the uncertainty set $\mathcal{U}(A, b, \rho) = \{(A + \Delta A, b + \Delta b) \mid \|(\Delta A, \Delta b)\|_F \leq \rho\}$.
- Wlog you can consider only the robust counterpart for $\mathcal{U}(A, b, 1)$.

The robust counterpart:

$$\min_x \max_{\|\Delta A, \Delta b\|_F \leq \rho} \|(A + \Delta)x - (b + \Delta b)\|_2$$

- **Theorem** The robust counterpart problem for the unstructured total least squares problem is

$$\begin{array}{ll} \min & \lambda \\ \text{s.t.} & \|Ax - b\|_2 \leq \lambda - \tau \\ & \|(x, 1)\|_2 \leq \tau \end{array}$$

- Analytic solution for this SOCP:

$$X_{RLS} = \begin{cases} (\mu I + A^T A)^{-1} A^T b & \text{if } \mu = \frac{\lambda - \tau}{\tau} > 0 \\ \text{GINV}(A) b & \text{if } \mu = 0 \end{cases}$$

Other Versions I

- The ∞-norm version of the problem is much harder
- I.e., the problem:

$$\min_x \max_{\|(A,b\|_\infty \le \rho} \|Ax - b\|_2$$

- The inner max problem is as hard as an optimization problem can be!
- Because it as hard as solving the MAX-CUT problem, or, maximizing a convex quadratic (homogeneous) function over box constraints.

Other Versions II

- Typically, in engineering applications a solution can be only implemented in some limited accuracy, *bit quantization*
- This is addressed by introducing uncertainty into the variables x
- We have to solve

$$\min_x \max_{\|\Delta A, \Delta b\|_F \le \rho, \|\Delta x\|_2 \le \mu} \|(A + \Delta)(x + \Delta x) - (b + \Delta b)\|_2$$

- We cannot usually find an equivalent explicit convex programming problem; have to use approximations

Contributors

- El-Ghaoui and Lebret
- Chandrasekaran, Golub, Gu and Sayed
- Ben-Tal and Nemirovski
- Lewis
- Watson
- Arikan and P.

Application to Supply Chain Management

- Bertsimas and Thiele (IPCO 2004). Full paper available on the web.
- Apply the robust optimization methodology of Bertsimas and Sim to a problem of ordering items at one or more installations, under stochastic demand over a finite discrete horizon of T periods.
- We will review the uncapacitated single station case
- Define, for $k = 0, \ldots, K$
- x_k the stock available at the beginning of the kth period
- u_k the stock ordered at the beginning of the kth period
- w_k the demand during the kth period

The Model I

- The evolution of stock over time:

$$x_{k+1} = x_k + u_k - w_k,\ k = 0, 1, \dots, T-1$$

or

$$x_{k+1} = x_0 + \sum_{i=0}^{k}(u_i - w_i),\ k = 0, 1, \dots, T-1$$

- The cost incurred at period k has two components: the purchasing cost $C(u_k)$ and and a holding/shortage cost resulting from this order $R(x_k + u_k - w_k)$
- The purchasing cost is of the form:

$$C(u) = \begin{cases} K + cu & \text{if } u > 0 \\ 0 & \text{if} u = 0 \end{cases}$$

The Model II

- The holding/shortage cost represents the cost associated with having either excess inventory (positive stock) or unfilled demand (negative stock).
- Consider a convex piecewise linear holding/shortage cost:

$$R(x) = \max(hx, -px)$$

where h and p are non-negative. Assume $p > c$ so that ordering stock remains a possibility until the last period.

The Model III

- Assume no lead time in the deliveries
- Ignoring stochasticity for the moment the model is:

$$\min \sum_{k=0}^{T-1} (cu_k + Kv_k + y_k)$$

subject to

$$y_k \geq h(x_0 + \sum_{i=0}^{k}(u_i - w_i)),\ k = 0, 1, \ldots, T-1$$

$$y_k \geq -p(x_0 + \sum_{i=0}^{k}(u_i - w_i)),\ k = 0, 1, \ldots, T-1$$

$$0 \leq u_k \leq Mv_k, v_k \in \{0, 1\}\ k = 0, 1, \ldots, T-1$$

where M is a large (positive) number.

Incorporating Robustness into the Model

- Model demand uncertainty as

$$w_i = \bar{w}_i + \hat{w}_i z_i$$

such that

$$z \in \mathcal{P} = \{|z_i| \leq 1\ \forall i \geq 0,\ \sum_{i=0}^{k} |z_i| \leq \Gamma_k, \forall k \geq 0\}.$$

- The choice of Γ_k is discussed in Bertsimas and Thiele. It is an increasing parameter with k because uncertainty increases with the number of periods considered.

The Robust Model

$$\min \sum_{k=0}^{T-1} (cu_k + Kv_k + y_k)$$

subject to

$$y_k \geq h(x_0 + \sum_{i=0}^{k} (u_i - w_i) + q_k \Gamma_k + \sum_{i=0}^{k} r_{ik}), \forall k$$

$$y_k \geq p(-x_0 - \sum_{i=0}^{k} (u_i - w_i) + q_k \Gamma_k + \sum_{i=0}^{k} r_{ik}), \forall k$$

$$q_k + r_{ik} \geq \hat{w}_i, \ \forall k, \ \forall i \leq k$$

$$q_k \geq 0, r_{ik} \geq 0, \ \forall k, \ \forall i \leq k$$

$$0 \leq u_k \leq M v_k, v_k \in \{0, 1\} \ k = 0, 1, \ldots, T-1$$

The Robust Optimal Policy

- **Definition**: The optimal policy of a discrete-horizon inventory problem is said to be (s, S) or base-stock, if there exists a threshold sequence (s_k, S_k) such that at each time period k, it is optimal to order $S_k - x_k$ if $x_k < s_k$ and 0 otherwise, with $s_k \leq S_k$. If there is no fixed ordering cost $(K = 0)$, $s_k = S_k$.
- The nominal problem where all w_i are replaced by $\bar{w}_i$ the optimal policy of the s, S type with closed-form expressions for s_j, S_j.
- The optimal robust policy for is the optimal policy for the nominal problem with a modified demand.

Other Applications (not covered here)

- Robust Truss Topology Design (Ben-Tal and Nemirovski)
- Robot Arm Design under Uncertainty (Ben-Tal and Nemirovski)
- Robust One-Period Option Modeling (Lutgens, Sturm)
- Robustness in Control Theory and Dynamical Systems (Ben-Tal, Nemirovski, Boyd)
- Robustness in Markov Chains (El-Ghaoui, Nilim)
- Robust location analysis (Thiele)