Optimization Models and Methods with Applications in Finance

Javier Nogales
www.est.uc3m.es/Nogales
@fjnogales

BCAM & UPV/EHU Courses
Bilbao, February, 2013

Universidad Carlos III de Madrid

Some references

- J. Nocedal and S.J. Wright. Numerical Optimization. Springer-Verlag, 2006
- S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2004
- G. Cornuejols and R. Tütüncü: Optimization Methods in Finance. Cambridge University Press, 2007
- W. T. Ziemba and R. G. Vickson (Ed.): Stochastic optimization models in finance. World Scientific, 2006
- A. Ruszczynski, A. Shapiro (Ed.): Stochastic Programming. Elsevier , 2003
- S. W. Wallace (Ed.) Applications of Stochastic programming. Book Data Limited, UK, 2005

Structure of the course

1. Introduction and modeling
2. Unconstrained Optimization and Applications
3. Constrained Optimization and Applications
4. Optimization under Uncertainty with Applications

Universidad Carlos III de Madrid

Introduction

Optimization:

An important tool in decision sciences and the analysis of physical systems

Several definitions:

- The discipline of applying advanced analytical methods to help make better decisions

- Narrowing your choices to the very best when there are virtually innumerable feasible options and comparing them is difficult

- Making the best possible choice for a vector in $\mathbb{R}^n$ from a set of possible choices

Introduction

Some examples:

- Airlines companies need to schedule their crew and planes in order to minimize their costs; investors build portfolios to maximize their returns (given a level of risk); industry companies try to maximize their efficiency in design and operations of their production planning; etc.

- Several applications in: Economy, Finance, Engineering, Medicine, Government, etc.

Framework

- Formulating the problem: Modeling
- Solving the problem
  - Studying the properties of problems and solutions
  - Designing efficient algorithms to compute these solutions
  - Applying the algorithms to obtain a solution
  - Validating the solution - conducting sensitivity analysis
- Applying these solutions in practice

Formulating the problem

Three basic elements of an optimization problem (mathematical program):

- Objective to be optimized: profit, time, energy, costs,. . .
- Variables (decisions): timetable for airplanes taking-off, amount of money to invest in each asset, . . .
- Constraints: some decisions are not allowed: airplanes taking-off is constrained for air security, risk must be controlled,. . .

Formulating the problem

- General framework:

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{X}, \end{array}$$

  where f is the objective function of n variables and $\mathcal{X}$ is a subset of $\mathbb{R}^n$ containing the constraints of the variables. This set is called feasible region

- The objective can be minimized or maximized. It is equivalent noting that

$$\max f(x) = -\min -f(x)$$

Solving the problem

- Both commercial and open source
- Specific for different classes of problems (CPLEX and Gurobi for LP, KNITRO and SNOPT for NLP, SeDuMi and SDPT3 for SDP, ...)
- Also available in general packages (Solver in Excel, Optimization toolbox and CVX in Matlab, ...)
- Open source software: COIN-OR, http://www.coin-or.org

Basic Definitions

- If $\mathcal{X} = \mathbb{R}^n$, then Unconstrained Optimization
- If f is linear and $\mathcal{X}$ is a polyhedron, then Linear Programming. Otherwise, Nonlinear Programming
- If f and $\mathcal{X}$ are convex, then Convex Optimization
- If $\mathcal{X}$ contains discrete variables, then Discrete Optimization or Integer Programming

In this course, focus on Nonlinear Optimization:

$$\mathcal{X} = \{x : \ c_i(x) = 0, i \in \mathcal{E}, \ c_i(x) \geq 0, i \in \mathcal{I}\},$$

where f and c_i are sufficiently differentiable. In this case, we can obtain good local information in an efficient way

Global vs Local Solutions

- For the general optimization framework:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in \mathcal{X}, \end{aligned} \tag{P}$$

- A point (decision) x^* is a local solution of (P) if there are no better points in a neighborhood of the solution, that is,

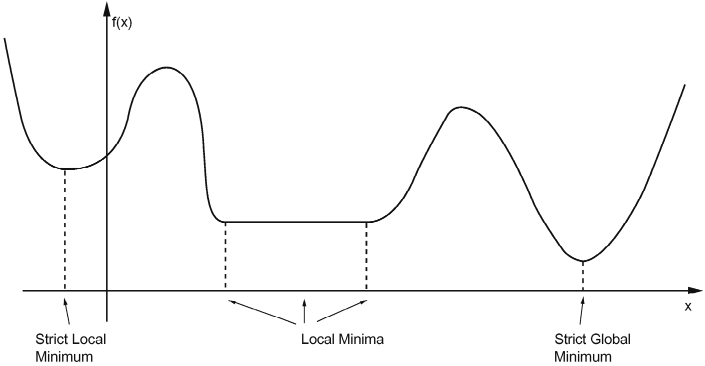
$$\exists \epsilon > 0 \text{ t.q. } \quad f(x^*) \leq f(x) \; \forall x \in \mathcal{X} \text{ s.t. } \quad \|x - x^*\| < \epsilon$$

- A point (decision) x^* is a global solution of (P) if there are no better points in all the feasible region, that is,

$$f(x^*) \leq f(x) \; \forall x \in \mathcal{X}$$

Global vs Local Solutions

- If previous inequalities are strict, solutions are locally strict
- Global optima are in general difficult to identify and to locate. One exception: convex optimization (global $\equiv$ local)
- Local Optimization: improved solutions and easy to find them
  Global Optimization: best overall solutions but difficult to find them



Existence of solutions

- Before trying to find a solution, does a solution exist for a given problem?

- We can assure the existence of at least one minimizer if:
  - f is a continuous function and $\mathcal{X}$ is a compact subset (Weierstrass Theorem)
  - In practice, the feasible region will be closed but not necessarily bounded. We need a stronger condition:
  - f is a continuous function, $\mathcal{X}$ is a closed subset and f is coercive in $\mathcal{X}$, that is,

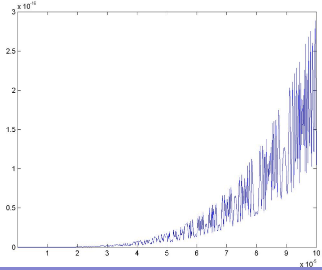
$$f(x) \to \infty \text{ when } \|x\| \to \infty$$

Existence of solutions

- There are difficulties if there exist infinite solutions. Hence, find isolated solutions:
- A local solution x^* is isolated if there exist a neighborhood of x^* such that there are no other local solutions
- Isolated is stronger than strict. Example:

$$f(x) = x^4(2 + \cos(1/x)), \; f(0) = 0$$

has an strict local minimum at $x = 0$ but there exists an infinity number of local minimizers near $x = 0$



Modeling: Some Examples

- Portfolio Optimization and Asset Allocation

- Developed by Harry Markowitz in the 1950s: formalize the diversification principle in portfolio selection

- 1990 winner of the Nobel Memorial Prize for Economics

Portfolio Optimization

- An investor has n available assets where she can allocate her money (buy and hold). The variable R_i represents the (random) return of asset i

  The problem is:

$$\begin{aligned} \text{maximize}_x \quad & \sum_i R_i x_i \\ \text{subject to} \quad & \sum x_i = 1, \end{aligned}$$

  but, is this problem well-defined?

Portfolio Optimization

- An investor has n available assets where she can allocate her money (buy and hold). The variable R_i represents the (random) return of asset i

  The problem is:

$$\begin{array}{ll} \text{maximize}_x & \sum_i R_i x_i \\ \text{subject to} & \sum x_i = 1, \end{array}$$

  but, is this problem well-defined?

- Well-defined problem:

$$\begin{array}{ll} \text{maximize}_x & \sum_i \mu_i x_i \\ \text{subject to} & \sum x_i = 1, \end{array}$$

  where $\mu_i = E(R_i)$. But, is its solution reasonable?

Portfolio Optimization

- Markowitz: Risk measured through the variance of the portfolio return

$$\begin{aligned} \text{maximize}_x \quad & \mu^T x - \frac{1}{2}\,\gamma\, x^T \Sigma x \\ \text{subject to} \quad & \sum x_i = 1, \end{aligned}$$

where $\Sigma = \mathrm{Var}(R)$ and γ is the risk-aversion coefficient

- This model allows to build the so-called efficient frontier (those portfolios that, for a given return, have a minimum risk-level)

Portfolio Optimization

- In the classical model, μ and Σ are estimated through historical data: past returns, R_{it}
- Nowadays, there exist much better estimations (factor analysis, Black-Litterman, shrinked estimators, etc.)
- If short-selling is not allowed, then $x \geq 0$ must be a constraint
- This is a static model: one decision in time (not a sequence of decisions)
- It is a quadratic (convex) problem

Portfolio Optimization

- Other objective functions and constraints are possible

- Based on utilities:

$$\begin{array}{ll} \text{maximize}_x & E_R\big(U(R_1x_1 + \cdots + R_nx_n)\big) \\ \text{subject to} & \sum x_i = 1, \end{array}$$

  where U is a given utility function (logarithmic, power utility, etc.).
  But, how can we solve it?

- Nonlinear Programming and Optimization under Uncertainty

Portfolio Optimization

- Based on Risk Management:

$$\begin{array}{ll} \text{minimize}_x & \text{VaR}_\beta\left(-(R_1x_1 + \cdots + R_nx_n)\right) \\ \text{subject to} & \sum x_i = 1, \end{array}$$

where VaR_β is the Value-at-Risk for a given $0 \leq \beta \leq 1$

- This problem is nonlinear and nonconvex (local solutions)
- Conservative solutions: now required by regulators
- Other risk measures to be optimized: Conditional Value-at-Risk (CVaR) , Maximum Drawdown, etc.
- But, how can we solve it?
- Nonlinear Programming and Optimization under Uncertainty

Portfolio Optimization

Extensions (following days):

- Dynamic formulation: decisions made today affect future decisions
- Dynamic risk-measures (volatiities and correlations)
- Transaction costs
- Robust portfolios
- See http://estimationrisk.blogspot.com for real performance

Structure of the course

1. Introduction and modeling
2. Unconstrained Optimization and Applications
3. Constrained Optimization and Applications
4. Optimization under Uncertainty with Applications

Universidad Carlos III de Madrid

Introduction

- Once we know how to model a problem: how to solve it?

- Simplest case: unconstrained optimization

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

where $x \in \mathbb{R}^n$ and $f: \mathbb{R}^n \to \mathbb{R}$

Introduction

- Once we know how to model a problem: how to solve it?
- Simplest case: unconstrained optimization

$$\min_x f(x)$$

where $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$

- Structure:
    - Optimality characterization
    - Applications and Examples
    - Convex Optimization

Definition and Assumptions

- The function f must satisfy:
  - It has continuous (Lipschitz) second-order derivatives
  - It is bounded below
  - It goes to infinity when $x \to \infty$

Definition and Assumptions

- The function f must satisfy:
  - It has continuous (Lipschitz) second-order derivatives
  - It is bounded below
  - It goes to infinity when $x \to \infty$
- Main interest: find one local solution
- This solution satisfies necessary and sufficient conditions (optimality characterization)

Optimality Conditions

- If f is sufficiently differentiable, we can use Taylor's expansions to approximate the objective function in a neighborhood:

$$f(x^* + \Delta x) - f(x^*) = \nabla f(x^*)^T \Delta x + O(\|\Delta x\|^2)$$

$$f(x^* + \Delta x) - f(x^*) = \nabla f(x^*)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x^*) \Delta x + O(\|\Delta x\|^3)$$

- The deviations respect to the optimum:
  $\nabla f(x^*)^T \Delta x$ or $\nabla f(x^*)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x^*) \Delta x$
  help us to infer the behavior near the solution: necessary conditions

Optimality Conditions

- First-order necessary conditions:
  If x^* is a local minimizer, then it is easy to guess that the slope at this point must be zero, that is,

$$\nabla f(x^*) = 0 \tag{NC1}$$

  A point x^* satisfying (NC1) is called stationary point

- Second order necessary conditions:
  If x^* is a local minimizer, then it is stationary and the curvature at this point must be nonnegative, that is,

$$\nabla^2 f(x^*) \succeq 0 \tag{NC2}$$

Optimality Conditions

- If the problem is nonconvex, we need conditions with second derivatives to guarantee local optimizers, that is, sufficient conditions

- Second-order sufficient conditions:
  If x^* is a point satisfying:

$$\nabla f(x^*) = 0 \tag{SC1}$$

$$\nabla^2 f(x^*) \succ 0 \tag{SC2}$$

  then x^* is a strict local minimizer of f

Optimality Conditions

- Example:

$$f(x_1, x_2) = \frac{1}{3}x_1^3 + \frac{1}{2}x_1^2 + 2x_1x_2 + \frac{1}{2}x_2^2 - x_2 + 9$$

Necessary conditions: $\nabla f(x) = \begin{pmatrix} x_1^2 + x_1 + 2x_2 \\ 2x_1 + x_2 - 1 \end{pmatrix} = 0$

Therefore, there are two stationary points: $x_a = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ y $x_b = \begin{pmatrix} 2 \\ -3 \end{pmatrix}$

Optimality Conditions

- Sufficient conditions: $\nabla^2 f(x) = \begin{pmatrix} 2x_1 + 1 & 2 \\ 2 & 1 \end{pmatrix} \quad \Rightarrow$
  $\nabla^2 f(x_a) = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$ y $\nabla^2 f(x_b) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$

- Therefore:
  $\nabla^2 f(x_b)$ is positive definite $\Rightarrow x_b$ is a local minimizer
  $\nabla^2 f(x_a)$ is indefinite and x_a is neither a local minimizer nor a maximizer

- This function has neither a global minimum nor a maximum, because f is not bounded as $x_1 \to \pm\infty$

Optimality Conditions

- Local minimizers not satisfying the sufficient conditions are called singular points

- These points are very difficult to deal with: (i) their optimality cannot be assured through sufficient conditions with low computational time (in the absence of convexity); (ii) near these points, almost all the algorithms are slow and erratic

Optimality conditions

- Example:

$$f_1(x) = x^3,\ f_2(x) = x^4,\ f_3(x) = -x^4$$

- All of these functions satisfy $\nabla f(0) = \nabla^2 f(0) = 0 \quad \Rightarrow$
  $x = 0$ is a candidate to be a local minimizer for all of these functions, but:

- while f_2 has a local minimizer at $x = 0$,
  f_1 has a saddle point
  f_3 has an local maximizer

Optimality Conditions

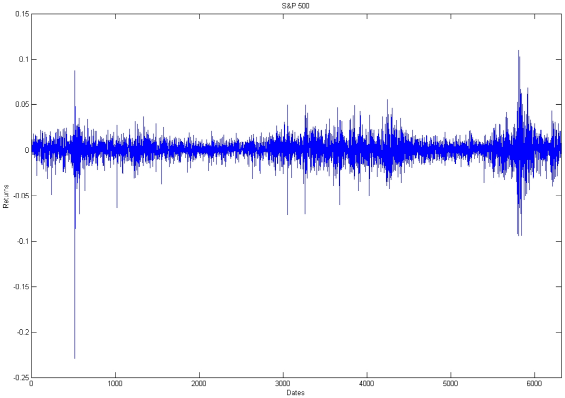
- As a summary, for stationary points:

$$
\begin{aligned}
\nabla^2 f(x^*) \succ 0 &\Rightarrow \text{ minimizer} \\
\nabla^2 f(x^*) \prec 0 &\Rightarrow \text{ maximizer} \\
\nabla^2 f(x^*) \text{ indefinite} &\Rightarrow \text{ saddle point} \\
\nabla^2 f(x^*) \text{ singular} &\Rightarrow \text{ several performances}
\end{aligned}
$$

Applications: GARCH models

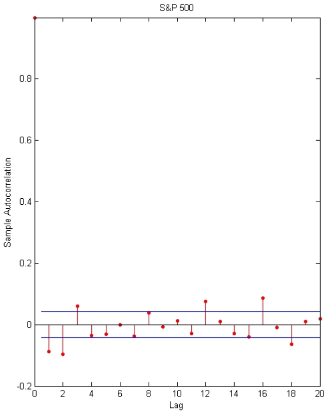
- GARCH model: generalization of the autoregressive conditional heteroscedastic (ARCH) model
- Econometric term developed in 1982 by Robert F. Engle, 2003 winner of the Nobel Memorial Prize for Economics
- Describe an approach to estimate dynamic volatility in financial markets
- The GARCH model is widely used in Finance: more real-world context than other forms when trying to predict prices and rates of financial instruments

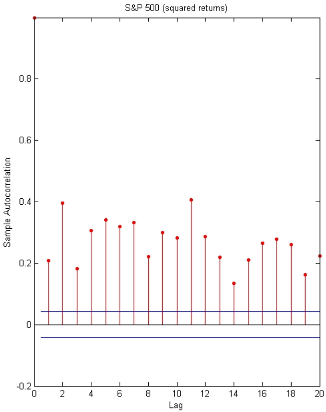
GARCH model: daily S&P 500 returns



Dynamic volatility and volatility clusters

GARCH model: daily S&P 500 correlations





Stock returns are uncorrelated, but dependent

Classical Estimation

- Model for daily returns:

$$r_t = \sigma_t \epsilon_t$$

where ϵ_t is a vector of disturbances and σ_t represents the volatilities

- Classical model. Assume i.i.d. normal returns: $\epsilon_t \sim \mathcal{N}(0,1)$
Then, $\sigma_t = \sigma$ for all t

- The log-likelihood to be maximized over σ is

$$L = -\frac{1}{2} \sum_{t=1}^{T} (N \log(2\pi) + \log \sigma^2 + \frac{r_t^2}{\sigma^2})$$

Apply the optimality conditions and obtain: $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} r_t^2$

- This is the sample variance

GARCH(1,1) model

- GARCH(1,1) model: Volatilities are dynamic

- The log-likelihood to be maximized over (α, β) is

$$L = -\frac{1}{2}\sum_{t=1}^{T}(N\log(2\pi) + \log\sigma_t^2 + \frac{r_t^2}{\sigma_t^2})$$

with

$$\sigma_t^2 = \sigma^2 \times (1 - \alpha - \beta) + \alpha r_{t-1}^2 + \beta\sigma_{t-1}^2$$

- In this case, the solution is not explicit: nonconvex problem

- Apply optimization algorithms

Convex Optimization

- When f is convex, then local and global optimizers are equivalent and easy to identify

- A function f is convex if

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \ \forall 0 \leq \alpha \leq 1$$

- If f is differentiable ($f \in C^1$), then it is convex if and only if

$$f(y) \geq f(x) + \nabla f(x)^T(y-x) \quad \forall x, y$$

- If $f \in C^2$, then it is convex if and only if

$$\nabla^2 f(x) \succeq 0 \quad \forall x$$

Differentiable Convex Optimization

- If f is differentiable, there are two reasons why convex problems are important:
  - In this case, there is no difference between a local and a global minimizer, that is, every local minimizer is a global solution too. Moreover, the set of global minimizers is convex
    If f is strictly convex, then the global minimizer is unique
  - The necessary condition $\nabla f(x^*) = 0$ becomes a sufficient one

Nondifferentiable Convex Optimization

- What happens if f is not differentiable (but convex)?

- We say g_x is a subgradient of f at x if

$$f(y) \geq f(x) + g_x^T(y - x) \quad \forall y$$

- The subdifferential of f at x is the set of all the subgradients:

$$\partial f(x) = \{g_x : \; f(y) \geq f(x) + g_x^T(y - x) \quad \forall y\}$$

- The set $\partial f(x)$ is a closed convex set (maybe empty) even if f is not convex

Convex Optimization

- If f is not differentiable, there are two reasons why convex problems are important:
    - Again, there is no difference between a local and a global minimizer, that is, every local minimizer is a global solution too
    - x^* is a minimizer of $f(x)$ if and only if $0 \in \partial f(x^*)$

Examples

- $f(x) = ||x||_1$

- $\partial f(0) = \{x : \ -e \leq x \leq e\} = \{x : \ ||x||_\infty \leq 1\}$

- $\partial f(-2e) = -e$

- In general, $\partial f(x) = \{u : \ u^T x = ||x||_1 \ \text{ and } ||u||_\infty \leq 1\}$

Examples

- $f(x) = ||x||_2$

- If $x \neq 0$, then $\partial f(x) = \frac{1}{||x||_2} x$

- If $x = 0$, then $\partial f(0) = \{x : \; ||x||_2 \leq 1\}$

Applications

- Large regression problems. For the model, $y = X\beta + u$, the popular LASSO estimator is:

$$\min_{\beta} ||y - X\beta||_2^2 + \rho||\beta||_1$$

Widely used in Data Mining and Big Data analysis

Applications

- Transaction costs in Finance. For the portfolio optimization problem:

$$\max_x \mu^T x - \frac{\gamma}{2} x^T \Sigma x - \kappa ||x - \bar{x}||_1$$

where κ represents the proportional transaction costs (around 20 bps) and $\bar{x}$ denotes previous portfolio allocation before rebalancing

Widely used in Portfolio Management

Applications

- Jensen inequality. If f is convex and X is a random variable, then $f(E(X)) \leq E(f(X))$

  Widely used in Probability and Statistics

Structure of the course

Universidad Carlos III de Madrid

Constrained Optimization

- Now add constraints

- For instance, in the GARCH(1,1) objective function:

$$L = -\frac{1}{2}\sum_{t=1}^{T}(N\log(2\pi) + \log\sigma_t^2 + \frac{r_t^2}{\sigma_t^2})$$

we can add the constraints $\sigma_t^2 = \sigma^2 \times (1 - \alpha - \beta) + \alpha r_{t-1}^2 + \beta\sigma_{t-1}^2$, and $\alpha + \beta < 1$ to assure stationarity

Constrained Optimization

- Now add constraints

- For instance, in the GARCH(1,1) objective function:

$$L = -\frac{1}{2}\sum_{t=1}^{T}(N\log(2\pi) + \log\sigma_t^2 + \frac{r_t^2}{\sigma_t^2})$$

  we can add the constraints $\sigma_t^2 = \sigma^2 \times (1-\alpha-\beta) + \alpha r_{t-1}^2 + \beta\sigma_{t-1}^2$, and $\alpha + \beta < 1$ to assure stationarity

- For instance, in the portfolio selection problem:

$$\mu^T x - \frac{1}{2}\,\gamma\, x^T \Sigma x$$

  we can add the constraints $e^T x = 1$ (initial wealth), $x \geq 0$ (no short-selling), $Ax = b$ diversification by sectors, etc.

General Framework

- Constrained Nonlinear Optimization Problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & c_{\mathcal{E}}(x) = 0 \\ & c_{\mathcal{I}}(x) \geq 0 \end{aligned} \qquad \text{(NLP)}$$

where f and c_i have continuous (Lipschitz) second-order derivatives

- Goal: compute a local solution satisfying some optimality conditions

Optimality conditions

- The terms KKT point and KKT conditions (Karush-Kuhn-Tucker) are used to characterize stationary points when there are constraints

- We say that the first-order KKT conditions for problem (NLP) are satisfied at a point x^* if there exists a vector λ^* (Lagrange multipliers) such that:

$$\begin{aligned}
\nabla_x f(x^*) - \nabla_x c(x^*)\lambda^* &= 0 && \text{stationarity} && \text{(KKT1)}\\
c_{\mathcal{I}}(x^*) \geq 0 \text{ and } c_{\mathcal{E}}(x^*) &= 0 && \text{feasibility} && \text{(KKT2)}\\
c_{\mathcal{I}}(x^*) \circ \lambda^*_{\mathcal{I}} &= 0 && \text{complementarity} && \text{(KKT3)}\\
\lambda^*_{\mathcal{I}} &\geq 0 && \text{multipliers sign} && \text{(KKT4)}
\end{aligned}$$

- A point x^* satisfying these conditions is called a KKT point

Constrained Optimization

- Example:

$$\begin{aligned} \text{minimize}_x \quad & f(x) = (x_1 - 3/2)^2 + (x_2 - 1/4)^2 \\ \text{subject to} \quad & c(x) = \begin{pmatrix} 1 - x_1 - x_2 \\ 1 - x_1 + x_2 \\ 1 + x_1 - x_2 \\ 1 + x_1 + x_2 \end{pmatrix} \geq 0. \end{aligned}$$

The first-order KKT conditions are satisfied at point $x^* = (1, 0)$

Constrained Optimization

- We can check it:

$$\begin{pmatrix} -1 \\ -0.5 \end{pmatrix} - \begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1^* \\ \lambda_2^* \\ \lambda_3^* \\ \lambda_4^* \end{pmatrix} = 0 \qquad \text{(KKT1)}$$

$$\begin{pmatrix} 1-1-0 \\ 1-1+0 \\ 1+1-0 \\ 1+1+0 \end{pmatrix} \geq 0 \qquad \text{(KKT2)}$$

$$\begin{pmatrix} 0 \\ 0 \\ 2 \\ 2 \end{pmatrix} \circ \begin{pmatrix} \lambda_1^* \\ \lambda_2^* \\ \lambda_3^* \\ \lambda_4^* \end{pmatrix} = 0 \qquad \text{(KKT3)}$$

$$\lambda^* \geq 0, \qquad \text{(KKT4)}$$

where $\lambda^* = (3/4, 1/4, 0, 0)^T$

Constrained Optimization

- The stationarity condition (KKT1) can be written as

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0,$$

where

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T c(x)$$

is the Lagrangian function

- We need to take care with the multipliers sign: if we are maximizing, they change; if constraints are $\leq$, they change

- Notation: $g(x) = \nabla_x f(x), \quad J(x) = \nabla_x c(x)^T$

- Active constraints: $\mathcal{A}(\bar{x})$, those satisfied with equality at a point $\bar{x}$

Constraint Qualifications

- To obtain first-order necessary conditions, first ensure that local linear approximations are good representations

- LICQ (linear independence constraint qualification): This condition is satisfied at a feasible point $\bar{x}$ if $J_{\mathcal{A}}(\bar{x})$ (the gradients of the active constraints at $\bar{x}$) is a full row rank matrix (linearly-independent gradients)

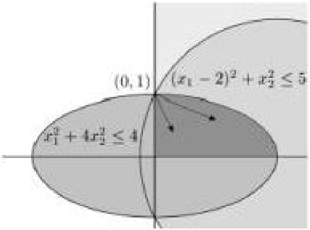
- MFCQ (Mangasarian-Fromovitz constraint qualification): This condition is satisfied at a feasible point $\bar{x}$ if the gradients of the equality constraints at $\bar{x}$ are linearly independent (full row rank jacobian of the equality constraints) and if there exists $p \neq 0$ such that $J_{\mathcal{A}}(\bar{x})p > 0$ for all $i \in \mathcal{A}(\bar{x}) \cap \mathcal{I}$ and $J_{\mathcal{A}}(\bar{x})p = 0$ for all $i \in \mathcal{E}$

- The MFCQ is a weaker condition than the LICQ

- Example

  Feasible region:

$$4 - x_1^2 - 4x_2^2 \geq 0$$
$$5 - (x_1 - 2)^2 - x_2^2 \geq 0$$
$$x_1, x_2 \geq 0$$



- At $x = (0, 1)$, the matrix $J_{\mathcal{A}}(x)$ has not full rank, hence, LICQ does not hold.

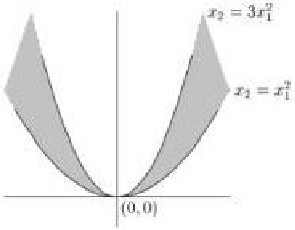
- But the MFCQ holds because $J_{\mathcal{A}}(x)p > 0$ if $p = (1, -1)$

Constraint Qualifications

- Example

  Feasible region:

$$-x_1^2 + x_2 \geq 0$$
$$3x_1^2 - x_2 \geq 0$$



- At $x = (0,0)$, the matrix $J_{\mathcal{A}}(x)$ has not full rank, hence, the LICQ does not hold

- In this case, the MFCQ does not hold because $J_{\mathcal{A}}(x)p > 0$ would imply $p_2 > 0$ and $-p_2 > 0$. Note there does not exist a linear path with strict interior to reach the point (although there exist interior nonlinear paths)

Constraint Qualifications

- LICQ is the most demanding, but it is very easy to check (check the rank of a matrix). It implies that the multiplier vector is unique

- MFCQ is satisfied for a very large class of problems, but it is more difficult to check (solve a linear program)

  Given a point x, if the solution of

$$\begin{aligned} \text{minimize}_{p,\theta} \quad & -\theta \\ \text{subject to} \quad & J_{\mathcal{A}}(x)p - \theta e \geq 0, \\ & 0 \leq \theta \leq 1, \end{aligned}$$

  is obtained at $\theta^* = 1$, then the MFCQ is satisfied at x

  With MFCQ there may be an infinite number of multiplier vectors. But these multiplier vectors are bounded

First-order necessary conditions

- If a constraint qualification holds, then the KKT conditions are necessary for a local minimizer x^* of the constrained problem

$$
\begin{aligned}
g(x^*) - J(x^*)^T \lambda^* &= 0 && \text{stationarity} && \text{(KKT1)} \\
c_{\mathcal{I}}(x^*) \geq 0 \text{ and } c_{\mathcal{E}}(x^*) &= 0 && \text{feasibility} && \text{(KKT2)} \\
c_{\mathcal{I}}(x^*) \circ \lambda^*_{\mathcal{I}} &= 0 && \text{complementarity} && \text{(KKT3)} \\
\lambda^*_{\mathcal{I}} &\geq 0 && \text{multipliers sign} && \text{(KKT4)}
\end{aligned}
$$

Second-order necessary conditions

- The Lagrangian hessian is denoted by

$$H(x, \lambda) = \nabla^2_{xx}\mathcal{L}(x, \lambda) = \nabla^2_{xx} f(x) - \sum_i \lambda_i \nabla^2_{xx} c_i(x)$$

- When a local solution x^* satisfies the LICQ, then it is a KKT point and

$$p^T H(x^*, \lambda^*) p \geq 0 \ \forall p \ \text{ tal que } J^*_{\mathcal{A}} p = 0 \qquad \text{(SONC)}$$

- Condition (SONC) is equivalent to

$$Z^T H(x^*, \lambda^*) Z \succeq 0,$$

where Z is matrix whose columns for a basis for the null space of $J^*_{\mathcal{A}}$:

$$J_{\mathcal{A}}(x^*)^T = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R \\ 0 \end{pmatrix} = QR,$$

Q_1 dimension $n \times m$, $\boxed{Z = Q_2}$ dimension $n \times (n - m)$, R dimension $m \times m$

Second-order necessary conditions

- Example

$$\begin{array}{ll} \text{minimize}_x & f(x) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3 \\ \text{subject to} & x_1 - x_2 + 2x_3 = 0. \end{array}$$

The point $x^* = (2.5, -1.5, -1)^T$ satisfies the KKT conditions with $\lambda^* = 3$

$$H(x^*, \lambda^*) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{pmatrix},\ Z = \begin{pmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix},\ Z^T H(x^*, \lambda^*) Z = \begin{pmatrix} 4 & -4 \\ -4 & 6 \end{pmatrix} \succ 0.$$

Hence, conditions (SONC) are satisfied at $x^* = (2.5, -1.5, -1)^T$

Second-order sufficient conditions

- Given a KKT point x^*, it satisfies the strict complementarity condition (SCS) if there exists a Lagrange multiplier λ^* such that $\lambda_i^* > 0$ for all $i \in \mathcal{A} \cap \mathcal{I}$

  That is, all active constraints have multipliers different from zero

- A point x^* is an isolated solution to (NLP) if: (i) x^* is a KKT point, (ii) the LICQ and the SCS are satisfied at x^* and (iii) for the unique $\lambda^* \in \mathcal{M}_\lambda(x^*)$ and for all $p \neq 0$ such that ${g^*}^T p = 0$, $J_{\mathcal{E}}^* p = 0$ and $J_{\mathcal{I}\cap\mathcal{A}}^* p \geq 0$, there exists $\omega > 0$ such that $p^T H(x^*, \lambda) p \geq \omega \|p\|^2$

  Condition (iii) is equivalent to

$$Z^T H(x^*, \lambda^*) Z \succ 0$$

Mean-variance portfolios

- Markowitz: Investor who cares only about mean and variance of static portfolio returns should hold the following portfolio:

$$\begin{aligned} \max_{x} \quad & \mu^T x - \frac{\gamma}{2} x^T \Sigma x \\ \text{s.t.} \quad & x^T e = 1, \end{aligned}$$

where $\mu = E(R)$, $\Sigma = \text{Var}(R)$ and $\gamma \equiv$ risk aversion parameter

- From KKT conditions, the solution is:

$$x = \frac{1}{\gamma} \Sigma^{-1} \mu - \frac{\lambda_\gamma}{\gamma} \Sigma^{-1} e,$$

where $\lambda_\gamma = \frac{\mu^T \Sigma^{-1} e - \gamma}{e^T \Sigma^{-1} e}$

- $H(x^*, \lambda^*) = -\gamma \Sigma$ which is negative definite (we are maximizing)

Convex Optimization

- Consider the (not necessarily convex) problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & c_{\mathcal{E}}(x) = 0 \\ & c_{\mathcal{I}}(x) \geq 0 \end{aligned} \qquad \text{(Primal)}$$

- Let $\mathcal{L}(x, \lambda) = f(x) - \lambda^T c(x)$ be the Lagrangian function

- The dual function is $g(\lambda) = \inf_x \mathcal{L}(x, \lambda)$

- The dual problem is

$$\begin{aligned} \text{maximize} \quad & g(\lambda) \\ \text{subject to} \quad & \lambda_{\mathcal{I}} \geq 0 \end{aligned} \qquad \text{(Dual)}$$

Convex Optimization

- Let f^* be the optimal value function of (Primal)

- If $\lambda_{\mathcal{I}} \geq 0$, then $g(\lambda) \leq f^*$
  This is the weak duality (lower bound for f^*)

- Hence, the dual problem provides the best bound $g^* \leq f^*$ always

- Problem (Dual) is convex even if problem (Primal) is not

- If $g^* = f^*$, then strong duality

Convex Optimization

- Slater condition (sufficient condition for strong duality):
  If (Primal) is convex and there exists x s.t. $c_{\mathcal{I}}(x) > 0$, then $g^* = f^*$ (duality gap is 0)

- If the problem (Primal) is convex, that is

  f convex, linear equality constraints, concave inequality constraints

  then

  - There is no difference between a local and a global minimizer, that is, every local minimizer is a global solution too
  - Under the Slater condition, x^* is a minimizer if and only if the KKT conditions are satisfied
    (replace gradients in KKT by subdifferentials)

Structure of the course

1. Introduction and modeling
2. Unconstrained Optimization and Applications
3. Constrained Optimization and Applications
4. Optimization under Uncertainty with Applications

Universidad Carlos III de Madrid

Optimization under Uncertainty with Applications

- Up to now, we made the implicit assumption that the data of the problem are all known

- Often, the problem parameters will only be known in the future, or cannot be known exactly before the problem is solved

Optimization under Uncertainty with Applications

- Up to now, we made the implicit assumption that the data of the problem are all known

- Often, the problem parameters will only be known in the future, or cannot be known exactly before the problem is solved

- Two fundamentally different approaches that address optimization under uncertainty:

  - Stochastic programming: data uncertainty is random and can be explained by some probability distribution (average view)

  - Robust optimization: data uncertainty is deterministic, but unknown. Solution behaves well under all possible realizations of the data (worst-case view)

Introduction

- Optimization under uncertainty is a decision-making framework to deal with model parameters (data) that are unknown or stochastic

- Some applications:

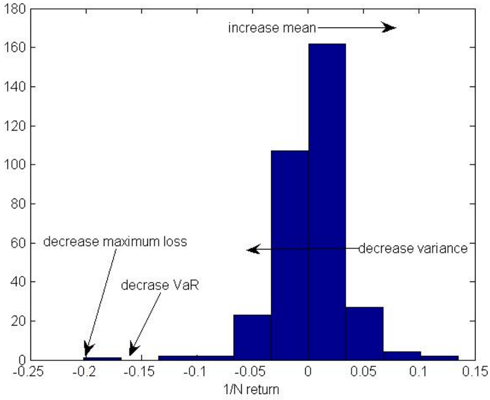
  - Transport and logistic (unknown demand)
  - Electricity markets (prices and demands are stochastic)
  - Finance (asset prices and demands are unknown)
  - Nuclear engineering (accident probability)
  - Environment (probability of non-satisfying regulations, i.e. Kyoto)
  - Airline companies (passenger demands, no shows)

Example

- Portfolio Optimization:

  For a given vector of portfolio weights, x

  The (empirical) distribution of the portfolio return is:



Optimize, over x, some characteristic of this random variable

Portfolio Optimization

- General Framework:

  Given N risky assets, select the vector of portfolio weights, x, that solves:

$$\begin{array}{rl} \text{max or min} & c(R^T x) \\ \text{subject to} & x^T e = 1, \end{array}$$

  where R is the random (next-period) return vector and $c(\cdot)$ measures a characteristic of the r. v. (mean-var, expected utility, risk measure, etc.)

- This is an Optimization Problem with Uncertainty

Portfolio Optimization

- General Framework:

  Given N risky assets, select the vector of portfolio weights, x, that solves:

$$
\begin{aligned}
\text{max or min} \quad & c(R^T \mathrm{x}) \\
\text{subject to} \quad & \mathrm{x}^T e = 1,
\end{aligned}
$$

  where R is the random (next-period) return vector and $c(\cdot)$ measures a characteristic of the r. v. (mean-var, expected utility, risk measure, etc.)

- This is an Optimization Problem with Uncertainty It can be solved:

- By Stochastic Programming techniques

- By Robust Optimization techniques

Example

- To feed his cattle, a farmer can plant his land with either wheat, corn and rice

  Excess or shortfall production can be sold or bought from a wholesaler

Available land for planting: 500 acres			
	Wheat	Corn	Rice
Yield (Ton/acre)	2.5	3	20
Planting cost ($/acre)	150	230	260
Selling price ($/Ton)	170	150	36 (under 6000 Ton) 10 (above 6000 Ton)
Purchase price ($/Ton)	238	210	-
Minimum requirement for cattle (Ton)	200	240	-

Example

- $x_{W,C,R}$: Acres of land devoted to wheat, corn and rice
- $w_{W,C,R,r}$: Tons of wheat, corn and rice (favorable/lower price) sold
- $y_{T,M}$: Tons of wheat and corn purchased

$$
\begin{aligned}
\text{minimize}_{x,w,y} \quad & +150x_W + 230x_C + 260x_R - 170w_W - 150w_C - 36w_R - 10w_r \\
& \quad + 238y_W + 210y_C \\
\text{subject to} \quad & x_W + x_C + x_R \leq 500 \\
& 2.5x_W + y_W - w_W \geq 200 \\
& 3x_C + y_C - w_C \geq 240 \\
& 20x_R + y_R - w_R - w_r \geq 0 \\
& w_R \leq 6000 \\
& x, y, w \geq 0
\end{aligned}
$$

Example

- Solution:

Cost: -$118600			
	Wheat	Corn	Rice
Plant (acres)	120	80	300
Production (Ton)	300	240	6000
Sales (Ton)	100	-	6000
Purchase (Ton)	-	-	-

  - It is profitable to have an excess of 100 wheat tons for selling
  - Corn is little profitable, plant only necessary to meet minimum requirements
  - Rice is only profitable if sold at favorable price

- But, is it a reasonable solution (model)?

Approaches

The previous solution is based on expected estimations of:

- Yield for each plant (it is weather-dependent)
- Sale and purchase prices (one year ahead)
- Minimum requirement (one year in advance)

But these quantities are unknown at the time to make the decision

So, how to deal with this uncertainty?

Approaches

Alternatives:

- Sensitivity analysis: changes in the solution and objective respect to changes in parameters. In general, this is a local approach (only one parameter can be changed at the same time) and it is only effective in small problems

- Scenario analysis or wait-and-see: The problem is solved for each parameter scenario and the obtained solutions are studied. The different optimal solutions and objectives are aggregated and the final optimal solution is obtained by an heuristic

- Stochastic programming or here-and-now: All parameter scenarios are taken into account at the same time

Example: Scenario Analysis

- Consider uncertainty only in the weather. Farmer considers two more scenarios: if weather is better then yield is 20% more than that of expected weather; if weather is worse then yield is 20% less than that of expected weather.

- Wait and see solutions:

Scenario	Best scenario			Worst scenario		
Cost	-$167667			-$59950		
	Wheat	Corn	Rice	Wheat	Corn	Rice
Plant (acres)	183.33	66.67	250	100	25	375
Production (Ton)	550	240	6000	200	60	6000
Sales (Ton)	350	-	6000	-	-	6000
Purchases (Ton)	-	-	-	-	180	-

Stochastic Programming

Note that:

- the solution is quite dependent on the weather

- impossible to make an optimal decision here and now: i.e., if he plants 300 acres of rice and weather is good then he will have to sell 1200 Ton at unfavorable price

The best alternative is to minimize the expected cost considering at the same time ALL the possible scenarios: Stochastic Programming

In our example, we have two types of decisions:

- first-stage decisions (now): amount of wheat, corn and rice to plant

- second-stage decisions (once the uncertainty is known): sales and purchases

Stochastic Programming

- Suppose the three different scenarios ($s = 1$ if better weather, $s = 2$ if expected weather, $s = 3$ if worse weather) occur with the same probability

- Then the stochastic programming model is the following:

$$
\begin{aligned}
\text{minimize}_{x,w,y} \quad & +150x_W + 230x_C + 260x_R + \frac{1}{|S|}\sum_{s\in S}(-170w_{Ws} - 150w_{Cs} \\
& \quad - 36w_{Rs} - 10w_{rs} + 238y_{Ws} + 240y_{Cs}) \\
\text{subject to} \quad & x_W + x_C + x_R \leq 500 \\
& d_{Ws}x_W + y_{Ws} - w_{Ws} \geq 200 \quad \forall s \in S \\
& d_{Cs}x_C + y_{Cs} - w_{Cs} \geq 240 \quad \forall s \in S \\
& d_{Rs}x_R + y_{Rs} - w_{Rs} - w_{rs} \geq 0 \quad \forall s \in S \\
& w_{Rs} \leq 6000 \quad \forall s \in S \\
& x, y, w \geq 0
\end{aligned}
$$

Stochastic Programming

- Solution here-and-now:

Expected cost: -$108390				
		Wheat	Corn	Rice
First stage	Land (acres)	170	80	250
s = 1	Production (Ton)	510	288	6000
Good	Sales (Ton)	310	48	6000
	Purchases (Ton)	-	-	-
s = 2	Production (Ton)	425	240	5000
Average	Sales (Ton)	225	-	5000
	Purchases (Ton)	-	-	-
s = 3	Production (Ton)	340	192	4000
Bad	Sales (Ton)	140	-	4000
	Purchases (Ton)	-	48	-

Stochastic Programming

- Summary:

Expected profit with expected weather	\$118600
Expected profit with good weather	\$167667
Expected profit with bad weather	\$59950
Expected profit with stochastic solution	\$108390

With stochastic solution:

- Rice is never sold at unfavorable price.
- Corn is only purchased if weather is bad.
- Most important: the stochastic solution is not optimal under EACH scenario but it is hedged under the risk associated to ALL the scenarios

Stochastic Programming: General Framework

- Given a probability space for the parameters with uncertainty, $(\Omega, \mathcal{A}, P)$, an stochastic optimization problem can be written as follows:

$$\begin{array}{ll} \text{minimize}_x & f(x) = E_\omega(F(x,\omega)) \\ \text{subject to} & x \in \mathcal{X}. \end{array}$$

- The probability distribution F is supposed to be known (or estimated through available data)

- This is the best approach although the size of the problem can be increased considerably

- Let x^* and $z^* = f(x^*)$ be the stochastic solution and the corresponding optimal value

Wait-and-see approach

- For each possible scenario ω, the following problem is solved:

$$\begin{aligned} \text{minimize} \quad & F(x, \omega) \\ \text{subject to} \quad & x \in \mathcal{X} \end{aligned}$$

  and solutions x_ω and $z_\omega = F(x_\omega, \omega)$ are obtained. Then, we can compute the expected cost under perfect information:

$$z_{\text{ws}} = \sum_{\omega} z_\omega p(\omega),$$

  where $p(\omega)$ denotes the probability of scenario ω

- In our example, $z_{\text{ws}} = \frac{-118600-167667-59950}{3} = -115406$

Wait-and-see approach: EVPI

- With this approach, we can define the expected value of perfect information (EVPI):

$$\text{EVPI} = z^* - z_{\text{ws}},$$

that measures how much farmer would be willing to pay for this perfect information: or for a good forecasting method

- An small value of EVPI indicates that more accurate forecasts do not guarantee significant profits. On the other hand, a large value of EVPI indicates that bad forecasts will have associated a great cost

- In our example, EVPI $= -108390 + 115406 = \$7016$. This is the amount of money that farmer would have to pay to obtain perfect information (via a fortune teller?)

Predictive approach

- The parameter values ω are estimated (predicted) by $\hat{\omega}$ and the following deterministic optimization problem is solved:

$$\begin{aligned} \text{minimize} \quad & F(x, \hat{\omega}) \\ \text{subject to} \quad & x \in \mathcal{X} \end{aligned}$$

- We obtain solutions $x_{\hat{\omega}}$ and $\hat{z}_{\text{d}} = F(x_{\hat{\omega}}, \hat{\omega})$

- Given the solution $x_{\hat{\omega}}$, the expected cost of this approach is denoted by z_{d}:

$$z_{\text{d}} = E_{\omega}(F(x_{\hat{\omega}}, \omega))$$

- In our example, because we have two-stage decisions, z_{d} is obtained as follows: first, the original problem is solved using $\hat{\omega}$ instead of ω. Then, the first-stage decisions are fixed and a second-stage problem is solved for each scenario. Finally, the average objective function for these problems is computed: $z_{\text{d}} = \frac{-118600-148000-55120}{3} = -107240$

Predictive approach: VSS

- With this approach, the value of the stochastic solution (VSS) can be defined:

$$\text{VSS} = z_{\text{d}} - z^*,$$

that measures the value of including uncertainty in the optimization problem

- In other words, it measures our profit if we use the stochastic programming approach instead of the predictive one
In our example, $\text{VSS} = -107240 + 108390 = \$1150$ is the gaining from using the stochastic solution rather than the predictive solution

- As a summary, if $F(x_{\hat{\omega}}, \omega)$ is convex, then:

$$z_{\text{ws}} \leq z^* \leq z_{\text{d}}$$

Two-stage Stochastic Programs

- The most used stochastic program is the following:

$$\begin{aligned} \text{minimize}_{x,y} \quad & c^T x + E_\xi \left[q(\omega)^T y(\omega) \right] \\ \text{subject to} \quad & Ax = b, \\ & T(\omega)x + Wy(\omega) = h(\omega), \\ & x, y(\omega) \geq 0, \end{aligned}$$

  where $T(\omega)$ is the technology matrix, W is the recourse matrix, x denotes the first-stage decisions, and y denotes the second-stage decisions

- If matrix W does not have uncertainty, then the problem has fixed recourse

Two-stage Stochastic Programs

- The deterministic equivalent program is the following:

$$\begin{aligned} \text{minimize}_{x,y} \quad & c^T x + \mathcal{Q}(x) \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0, \end{aligned}$$

where

$$\mathcal{Q}(x) = E_\xi \left[Q(x, \xi(\omega)) \right]$$

is the expected value of the objective function at the second stage and

$$Q(x, \xi(\omega)) = \min_y \{ q(\omega)^T y(\omega) | \;\; Wy(\omega) = h(\omega) - T(\omega)x, y \geq 0 \}$$

is the value for the function at the second stage given a scenario ω and a first-stage decision x

Properties

- The deterministic equivalent is a deterministic nonlinear problem
- The computational cost in these problems is very expensive: evaluation of $\mathcal{Q}(x)$
- Properties of $\mathcal{Q}(x)$ are very important when designing efficient solution methods:
  - It is continuous, but non linear. But if the probability distributions are discrete, then it is piecewise linear
  - It is convex and differentiable (if the probability distributions are continuous)

Financial Application

- Dynamic Portfolio Selection (Asset Allocation)
  Given an initial budget w_0, decide (for $t = 0, \ldots, T-1$) $x_{it} \equiv$ the amount of money to invest in each available asset i, $i = 1, \ldots, n$ to maximize the final profit (en $t = T$)

- The decisions depend on the asset returns, R_{it}, $t = 1, \ldots, T$, and they will be known progressively. Hence, the decisions are made recursively:

  decision $x_0 \rightsquigarrow R_1$ is known $\rightsquigarrow$ decision $x_1 \rightsquigarrow R_2$ is known $\cdots \rightsquigarrow$ decision x_{T-1}

Dynamic Portfolio Selection

- The dynamic stochastic problem is the following:

$$
\begin{aligned}
\text{maximize}_{x,w} \quad & E(U(w_T)) \\
\text{subject to} \quad & w_t = \sum_{i=1}^{n} (1 + R_{it}) x_{i,t-1}, \quad t = 1, \ldots, T \\
& \sum_{i=1}^{n} x_{it} = w_t, \quad t = 0, \ldots, T-1 \\
& w_t \geq 0, \quad t = 0, \ldots, T-1
\end{aligned}
$$

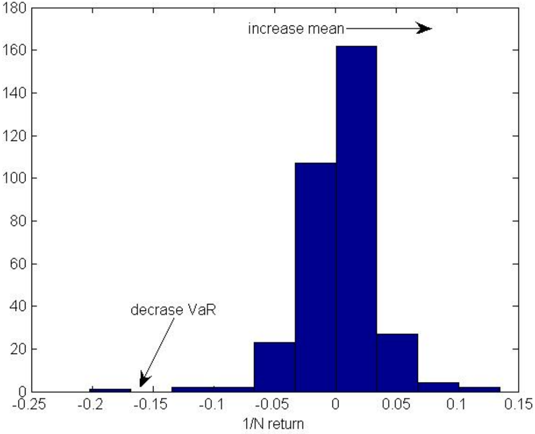
- The variables (states) w_t are random, except w_0
- The variables (controls) x_t are random, except x_0 (unknown but deterministic)

Chance Constraints

- Portfolio Optimization and Risk Management

  Now, risk is not modeled through the variance (or volatility)

  Now, focus on losses (instead of profits)



Chance Constraints

- Up to now, decisions feasible for (almost) all the uncertainty

- Relax this assumption to improve the objective function
  Example: allow the constraints to be satisfied with certain probability (probability of losing more than 5% of invested money less than 1%)

Chance Constraints

- Up to now, decisions feasible for (almost) all the uncertainty
- Relax this assumption to improve the objective function
  Example: allow the constraints to be satisfied with certain probability (probability of losing more than 5% of invested money less than 1%)
- Portfolio Optimization and Risk Management

$$
\begin{aligned}
\text{maximize}_x \quad & \mu^T x \\
\text{subject to} \quad & P(R^T x < -b) \le \alpha \\
& x^T e = 1
\end{aligned}
$$

- With this constraint relaxation, we can achieve a larger profit but assuming a higher loss risk

Chance Constraints

- Constraint $P(R^T x < -b) \leq \alpha$ is equivalent to $\text{VaR}_{1-\alpha}(-R^T x) \leq b$, where
$$\text{VaR}_\alpha(Z) = \min\{\gamma : \; P(Z \leq \gamma) \geq \alpha\},$$
represents the $100\alpha\%$ percentile of the variable Z

- VaR is the most-used risk measure in Finance and Insurance

- But in general, the optimization problem is non-differentiable and nonconvex. Moreover, it is not coherent

Coherent Risk Measures

- A risk measure, $\rho(Z)$ (for a loss Z), is coherent if:
  1. it is convex: $\rho(\lambda Z_1 + (1-\lambda)Z_2) \leq \lambda\rho(Z_1) + (1-\lambda)\rho(Z_2)$
  2. it is monotone: If $Z_2 \geq Z_1$, then $\rho(Z_2) \geq \rho(Z_1)$
  3. it is positive homogeneous: $\rho(\lambda Z) = \lambda\rho(Z)$ si $\lambda > 0$
  4. it is equivariant to traslations: $\rho(Z + a) = \rho(Z) + a$

- In addition, coherent risk measures have very good optimization properties:

$$\text{If } f(\cdot,\omega) \text{ is convex } \forall\omega, \text{ then } \rho(f(x,\omega)) \text{ is convex}$$

Coherent risk measures

- Conditional Value-at-risk:

$$\text{CVaR}_\alpha(Z) = E(Z|\ Z \geq \text{VaR}_\alpha(Z))$$

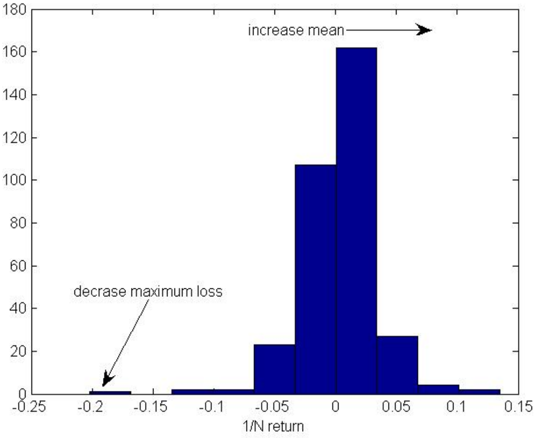
CVaR is more conservative than VaR: $\text{CVaR}_\alpha \geq \text{VaR}_\alpha$

- But it is a coherent risk-measure. Hence, the associated Risk Management problem is convex. Indeed, it can be reformulated as a LP !

- But it is not widely used in Finance (at least compared to VaR)

Robust Optimization

Now, focus is on the worst-case scenario (under a bounded set)



Robust Optimization

Stochastic Programming:

$$\begin{array}{ll} \text{minimize} & f(x) = E_\omega(F(x,\omega)) \\ \text{subject to} & x \in \mathcal{X} \end{array}$$

Robust Optimization:

$$\begin{array}{ll} \text{minimize} & f(x) = \sup_{\omega\in\Omega} F(x,\omega) \\ \text{subject to} & x \in \mathcal{X} \end{array}$$

- SP: uncertainty is described through a probability function. It provides good solutions in practice, but it suffers from the curse of dimensionality
- RO: uncertainty is not stochastic (no random variables), it is deterministic and based on bounded and convex sets. It is a worst-case scheme but the associated problems can be solved in an efficient way: they are tractable

Robust Portfolio Optimization

- Portfolio Selection:

$$\begin{aligned} \text{maximize}_x \quad & \mu^T x - \lambda x^T \Sigma x \\ \text{subject to} \quad & x^T e = 1 \end{aligned}$$

Now suppose μ and Σ are unknown, but not stochastic

For instance, we assume the expected return of a given asset will be in the interval $15\% \pm 7\%$ and the associated volatility inside $12\% \pm 2\%$

- Suppose for a moment that Σ is known, and consider the following ellipsoidal uncertainty set for the unknown μ:

$$\Omega = \{\mu : \ (\mu - \bar{\mu})^T \Sigma^{-1} (\mu - \bar{\mu}) \leq k^2\}$$

Robust Portfolio Optimization

- The robust approach is:

$$\begin{aligned} \text{maximize}_x \quad & \min_{\mu \in \Omega} \; \mu^T x - \lambda x^T \Sigma x \\ \text{subject to} \quad & x^T e = 1 \end{aligned}$$

or equivalently

$$\begin{aligned} \text{maximize}_x \quad & t - \lambda x^T \Sigma x \\ \text{subject to} \quad & \min_{\mu \in \Omega} \; \mu^T x \geq t \\ & (\mu - \bar{\mu})^T \Sigma^{-1} (\mu - \bar{\mu}) \leq k^2 \\ & x^T e = 1 \end{aligned}$$

- This is a problem with an infinite number of constraints
  We would like to deal with a tractable reformulation

Robust Portfolio Optimization

- The KKT conditions of the inner problem

$$\begin{array}{ll} \text{minimize}_\mu & \mu^T x \\ \text{subject to} & (\mu - \bar{\mu})^T \Sigma^{-1} (\mu - \bar{\mu}) \leq k^2 \end{array}$$

are

$$x - 2\lambda \Sigma^{-1} (\mu - \bar{\mu}) = 0$$
$$(\mu - \bar{\mu})^T \Sigma^{-1} (\mu - \bar{\mu}) \leq k^2$$

- Then, $\mu = \bar{\mu} + \frac{1}{\lambda} \Sigma x$ and (assuming the constraint is active) $\lambda = -\frac{1}{k} \sqrt{x^T \Sigma x}$

Robust Portfolio Optimization

- Hence, the optimal objective value for

$$\begin{array}{ll} \text{minimize}_{\mu} & \mu^T x \\ \text{subject to} & (\mu - \bar{\mu})^T \Sigma^{-1} (\mu - \bar{\mu}) \leq k^2 \end{array}$$

is $\bar{\mu}^T x - k \|\Sigma^{1/2} x\|$

- Now, we write this expression in our original robust formulation:

$$\begin{array}{ll} \text{maximize}_{x} & \min_{\mu \in \Omega} \mu^T x - \lambda x^T \Sigma x \\ \text{subject to} & x^T e = 1 \end{array}$$

Robust Portfolio Optimization

- Finally, the robust scheme is equivalent to the following tractable framework:

$$\begin{aligned} \text{maximize}_x \quad & \bar{\mu}^T x - \lambda x^T \Sigma x - k\|\Sigma^{1/2} x\| \\ \text{subject to} \quad & x^T e = 1 \end{aligned}$$

  This is a convex (non-quadratic) program:
  Second Order Cone Program≡ SOCP

- Now, compare with Markowitz approach:

$$\begin{aligned} \text{maximize}_x \quad & \bar{\mu}^T x - \lambda x^T \Sigma x \\ \text{subject to} \quad & x^T e = 1 \end{aligned}$$

  This is a convex QP

- The robust formulation is more conservative
  (new term penalizing estimation risk in μ)

Robust Portfolio Optimization

- Now what happens if Σ is unknown too?

- The robust formulation is:

$$\begin{array}{ll} \text{maximize}_x & \{\min_{(\mu,\Sigma)\in\Omega} \mu^T x - \lambda x^T \Sigma x\} \\ \text{subject to} & x^T e = 1, \end{array}$$

  where now Ω represents the uncertainty in (μ, Σ)

- For Σ, the following set leads to a tractable formulation (SOCP):

$$\Omega = \{\Sigma \succeq 0 : \; ||\hat{\Sigma}_{target}^{-1/2}(\Sigma - \hat{\Sigma})\hat{\Sigma}_{target}^{-1/2}||^2 \leq \delta^2\},$$

  where $||\cdot||$ denotes the 2- or Frobenius norm and $\hat{\Sigma}_{target}$ is a target matrix (identity, or 1-factor matrix, or even $\hat{\Sigma}$)

Robust Portfolio Optimization

- In this case, the corresponding tractable formulation is equivalent to the following norm-constrained portfolio formulation:

$$\begin{array}{ll} \max\limits_{x} & \hat{\mu}^T x - k\|\hat{\Sigma}^{1/2} x\| - \frac{\gamma}{2} x^T \hat{\Sigma} x - \frac{\delta}{2} x^T \hat{\Sigma}_{target} x \\ \text{subject to} & x^\top e = 1 \end{array}$$

- Solution even more conservative: another new term penalizing estimation risk in Σ

Robust Optimization

- In general, robust optimization problems are typically solved with second-order cone programs (SOCP) (convex problems that are solved efficiently)

- Advantages: no need of probability distribution for uncertainty. Problems solved in an efficient way

- Main limitation: conservative approach (worst-case view)

- Applications: finance (robust portfolio optimization), statistics (least squares with noise in explanatory variables), engineering (robust design), telecommunications (robust networks), signal processing, supply chain management, air traffic control, dynamic systems, etc.

- Active research topic: especially in portfolio optimization and (robust) optimal control

Robust Optimization

- Robust optimization is more than a minimax approach
  The solution methods are efficient!

- Chance constraints and robust optimization are related:

- Remember the chance constraints view: $P(c_i(x, \omega_i) \leq 0) \geq 1 - \alpha$
  Here, we look for feasible solutions for all the uncertainty and optimal in some sense. Hence, we look for robust solutions against uncertainty

- Robust optimization view: $c_i(x, \omega_i) \leq 0, \ \forall \omega_i \in \Omega_i$

Robust Optimization

- General formulation:

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & c_i(x, \omega_i) \leq 0, \ \forall \omega_i \in \Omega_i, \ i = 1, \dots, m. \end{array}$$

- Note in this formulation there is no probability

- Uncertainty is modelled through the uncertainty sets Ω_i (closed, bounded and typically convex)

- But the general formulation has an infinite number of constraints
  We need efficient solution methods $\equiv$ tractable formulations

Robust Optimization

- Objective of RO: obtain an equivalent reformulation that is tractable
  We can achieve this objective if we restrict the form of the sets Ω_i: typically ellipsoids (tractable convex sets)

- The objective function may depend on unknown parameters ω_0, too. But introducing an auxiliary variable, t, it fits with the general formulation:

$$\begin{array}{rl} \text{minimize}_{x,t} & t \\ \text{subject to} & c_i(x, \omega_i) \le 0,\ \forall \omega_i \in \Omega_i,\ i = 1, \ldots, m \\ & \max\limits_{\omega_0 \in \Omega_0} f(x, \omega_0) \le t \end{array}$$

Robust Linear Programming

- General formulation:

$$\begin{array}{rl} \text{minimize} & c^T x \\ \text{subject to} & Ax \le b, \ \forall a_i^T \in \Omega_i, \ i = 1, \ldots, m \end{array}$$

- It is equivalent to

$$\begin{array}{rl} \text{minimize} & c^T x \\ \text{subject to} & \max\limits_{a_i \in \Omega_i} a_i^T x \le b_i, \ i = 1, \ldots, m \end{array}$$

But this formulation is not tractable for general Ω

Robust Optimization

- Ellipsoidal uncertainty: if Ω is restricted to be

$$\Omega = \{A : \; a_i = a_i^0 + \Delta_i u_i, \; i = 1, \ldots, m, \; \|u\|_2 \leq \rho\},$$

then, the equivalent robust counterpart is

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & (a_i^0)^T x + \rho \|\Delta_i x\|_2 \leq b_i, \; i = 1, \ldots, m \end{array}$$

And this is a SOCP (tractable formulation)

Robust Optimization

- Polyhedral uncertainty: if Ω is restricted to be a polyhedra, then the equivalent robust counterpart is a LP
- More general cases:
  - For certain quadratic problems, the equivalent robust counterpart is a SDP
  - For certain SOCP problems, the equivalent robust counterpart is a SDP
  - But for SDP problems, the equivalent robust counterpart is in general a NP-hard problem

That's all

Thanks for your attention!!

Javier Nogales
www.est.uc3m.es/Nogales
@fjnogales