Sino-Japan Optimization Meeting, September 26-29, 2011, Beijing

Modeling and Algorithmic Challenges from Financial Optimization

Xiaoling Sun

School of Management
Fudan University

Outline

- Classical optimization models of portfolio selection
- Cardinality and minimum buy-in threshold constraints
- Local methods for cardinality constrained QP
  - p-norm approximation
  - Piecewise linear D.C. approximation
- Probabilistic constraints
- Factor-risk constraints
- Conclusions
- Refernces

Portfolio selection

- *Portfolio selection is to seek a best allocation of wealth among a basket of securities*.
- Markowitz (1952) developed a mean-variance (MV) model for portfolio selection which was the first return-risk optimization framework of investment theory.

Henry. M. Markowitz

Mean-variance model

- In MV model, the expected value of portfolio is measured by the mean of the portfolio and the risk is measured by the variance of the portfolio.
- Let ξ be the random vector of expected returns of n risky assets. Suppose ξ has the following mean vectors:

$$\mu = (\mu_1, \ldots, \mu_n)^T, \quad \mu_i = E(\xi_i), \; i = 1, \ldots, n,$$

  and co-variance matrix:

$$\Sigma = E[(\xi - E(\mu))(\xi - E(\mu))^T].$$

- The variance of the portfolio x is

$$\sigma^2(\xi^T x) = x^T Q x$$

- The mean-variance optimization model is

$$(MV) \qquad \begin{aligned} \min\ & x^T Qx \\ \text{s.t. } & \mu^T x \geq \rho, \\ & x \in X, \end{aligned}$$

  where ρ is a prescribed return level, and X is a set of constraints representing real-world trading conditions such as no shorting, bounds on exposure to groups of assets, sector allocation and regulation conditions. These constraints usually can be expressed as linear equality or inequality constraints.
- The classical MV model is a *convex quadratic program* which is polynomially solvable.
- Various extensions and improvement of MV model have been proposed since the pioneering work of Markowitz.

Extensions and improvement

- Various alternative risk measures:
  - Absolute deviation;
  - Downside risk measures such as semi-variance and lower partial moment;
  - Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR);
- Index tracking (passive portfolio management)
- Robust portfolio selection
- Multi-period portfolio selection
- Continuous time portfolio selection models

Portfolio selection models with hard constraints

- In this talk, we focus on portfolio selection models with hard constraints which arise from real-world financial optimization modeling.
- We consider the following three types of constraints in the mean-variance framework:
  - Cardinality and minimum buy-in threshold constraints;
  - Probabilistic constraint (VaR constraint);
  - Factor-risk constraints (marginal risk).

Cardinality and minimum buy-in threshold constraints

Let $x = (x_1, \ldots, x_n)$ be the vector of portfolio weights investing on n securities.

- Cardinality constraint: the number of assets in the optimal portfolio should be limited:

$$|\mathrm{supp}(x)| \leq K,$$

  where $\mathrm{supp}(x) = \{i \mid x_i \neq 0\}$, $1 \leq K << n$.
- The need to account for this limit is due to the transaction cost and managerial concerns.
- $|\mathrm{supp}(x)| = \|x\|_0$ is also called zero norm of x. A portfolio x with few nonzero elements is called sparse solution, or limited diversified portfolio.

- Minimum buy-in threshold:

$$x_i \geq \alpha_i, \;\; i \in \mathrm{supp}(x),$$

  or

$$x_i \in \{0\} \cup \{\alpha_i, 1\}.$$

  So, x_i is a semi-continuous variable.
- Cardinality constraint arises in portfolio selection models using both active and passive investment strategies.

MV models with cardinality and minimum buy-in threshold

- Cardinality constrained QP:

$$
\begin{aligned}
(\mathrm{CCQP}_1) \qquad & \min \ \frac{1}{2}x^T Qx + c^T x \\
& \text{s.t. } x \in X, \\
& \qquad |\mathrm{supp}(x)| \le K, \qquad (\text{cardinality constraint}) \\
& \qquad x_i \ge \alpha_i, \ \forall i \in \mathrm{supp}(x), \qquad (\text{threshold constraint}) \\
& \qquad 0 \le x_i \le u_i, \ i = 1, \ldots, n,
\end{aligned}
$$

where $\mathrm{supp}(x) = \{i \mid x_i > 0\}$, $\alpha_i > 0$, $0 < K < n$.

- Difficulty: testing the feasibility of (CCQP_1) is already NP-complete when A has three rows (Bienstock (1996)).

- Construct a portfolio with a few assets to track the performance of some market index:

$$\text{tracking error} = (x - x_0)^T \Sigma (x - x_0),$$

where x is the trading vector with small number of nonzero variables and x_0 is the weight vector of the benchmark index (S&P 500, FTSE 100, N225).

- Portfolio selection with cardinality and tracking error control:

$$
\begin{aligned}
(\mathrm{CCQP}_2) \qquad & \min \ x^T \Sigma x, \\
& \text{s.t. } (x - x_B)^T \Sigma (x - x_B) \le \sigma_0, \\
& \qquad \mu^T x \ge \rho, \ e^T x = 1, \\
& \qquad |\mathrm{supp}(x)| \le K, \\
& \qquad 0 \le x \le u, \ x_i \ge a_i, \ \forall i \in \mathrm{supp}(x).
\end{aligned}
$$

- The cardinality constraint can be represented by

$$e^T y \leq K, \; 0 \leq x_i \leq u_i y_i, \; y \in \{0,1\}^n.$$

- The minimum buy-in threshold $x_i \in \{0\} \cup [\alpha_i, u_i]$ can be expressed as

$$\alpha_i y_i \leq x_i \leq u_i y_i, \; y \in \{0,1\}^n.$$

- So that the cardinality constrained QP can be reformulated as a mixed-integer quadratic program (MIQP):

$$\begin{aligned} \min \; & \frac{1}{2} x^T Q x + c^T x \\ \text{s.t. } & x \in X, \\ & e^T y \leq K, \; y \in \{0,1\}^n, \\ & \alpha_i y_i \leq x_i \leq u_i y_i, \; i = 1, \ldots, n. \end{aligned}$$

Existing solution methods for cardinality constrained QP

- Branch-and-bound methods (based on continuous relaxation), e.g., MIQP solvers in CPLEX 12.1, Gurobi and Zimpl (?). Only small-size problems ($n \leq 50$) can be solved to global optimality.
- New MIQP reformulation using Lagrangian decomposition or perspective reformulation techniques (Frangioni and Gentile (2006), Zheng, Sun, Li (2010)).
- Branch-and-cut methods using cutting plane derived from the cardinality constraints.

Local Methods for Cardinality Constrained Problems

- Consider a general cardinality constrained QP:

$$
\begin{aligned}
(\mathrm{P}) \qquad \min\ & \frac{1}{2}x^T Qx + c^T x \\
\text{s.t. } & x \in X, \\
& \|x\|_0 \le K.
\end{aligned}
$$

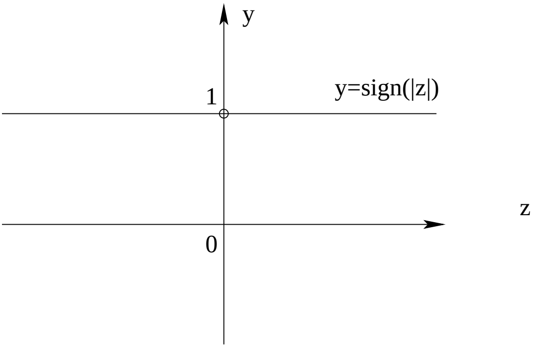
  This problem is still NP-hard even without the minimum buy-in threshold constraints.

- Note that

$$
\|x\|_0 = \sum_{i=1}^{n} \mathrm{sign}(|x_i|),
$$

  where $y = \mathrm{sign}(|z|)$ is discontinuous at 0.

- The function $y = \mathrm{sign}(|z|)$:



- Linear or nonlinear approximations (smooth or nonsmooth) to $y = \mathrm{sign}(|z|)$ can be considered. For example:
  - convex approximation (relaxation), e.g., ℓ_1-norm relaxation
  - p-norm approximation ($0 < p < 1$)
  - piecewise smooth approximation
  - piecewise linear approximation

p-norm approximation

- p-norm approximation:

$$\lim_{p\to 0} \|x\|_p^p = \lim_{p\to 0} \sum_{i=1}^{n} |x_i|^p = \|x\|_0 \quad (\textit{not uniformly convergent}).$$

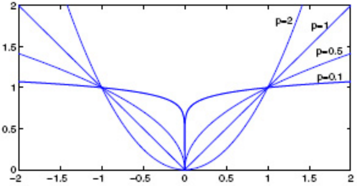


Figure: the behavior of p-norm function

- p-norm approximation to cardinality constraint:

$$\begin{aligned} \min \ & \frac{1}{2}x^T Qx + c^T x \\ \text{s.t. } & x \in X, \\ & \|x\|_p^p \leq K. \end{aligned}$$

- p-norm approximation and penalized problem:

$$\begin{aligned} \min \ & \frac{1}{2}x^T Qx + c^T x + \lambda\|x\|_p^p \\ \text{s.t. } & x \in X, \end{aligned}$$

ℓ_1 norm approximation

- If $p = 1$, then we have ℓ_1-norm approximation approximation:

$$\begin{aligned} \min\ & \frac{1}{2}x^T Qx + c^T x \\ \text{s.t. } & x \in X,\ x \in [-1, 1]^n, \\ & \|x\|_1 \le K, \end{aligned}$$

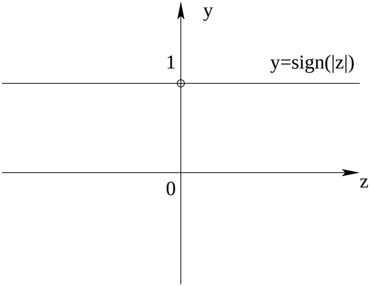
  where we have included the box constraint $x \in [-1, 1]^n$.

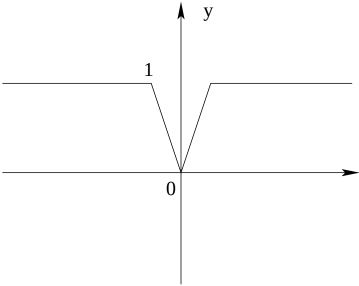
- Interestingly, the ℓ_1-norm approximation is equivalent to the continuous relaxation of (MIQP):

$$\begin{aligned} \min\ & \frac{1}{2}x^T Qx + c^T x \\ \text{s.t. } & x \in X, \\ & e^T y \le K,\ y \in [0, 1]^n, \\ & -y_i \le x_i \le y_i,\ i = 1, \ldots, n. \end{aligned}$$

Piecewise linear D.C. approximation

- Consider a piecewise linear approximation to the step function $y = |\mathrm{sign}(z)|$:





- This piecewise linear function can be expressed as a D.C. function:

$$\varphi(z,t) = \min\{\frac{1}{t}\|x\|_1, 1\} = \frac{1}{t}|z| - \frac{1}{t}\left[(z-t)^+ + (-z-t)^+\right].$$

- It is an underestimation: $\varphi(z,t) \le |\mathrm{sign}(z)|$, $\forall z \in \Re$.
- Let

$$\phi(x,t) = \sum_{i=1}^{n} \varphi(x_i,t) = \frac{1}{t}|x|_1 - \frac{1}{t}\sum_{i=1}^{n}\left[(x_i-t)^+ + (-x_i-t)^+\right].$$

Then, for any $x \in \Re^n$,

$$\lim_{t\to 0^+} \phi(x,t) = \|x\|_0.$$

(Not uniformly convergent)

- Consider the piecewise linear approximation to cardinality constraint:

$$
\begin{aligned}
(\mathrm{P_t}) \qquad\qquad & \min \ \frac{1}{2}x^T Qx + c^T x \\
& \text{s.t. } x \in X, \\
& \quad \frac{1}{t}\|x\|_1 - g(x,t) \le K
\end{aligned}
$$

where $g(x,t) = \sum_{i=1}^{n}[(x_i - t)^+ + (-x_i - t)^+$. This problem can be also expressed as

$$
\begin{aligned}
& \min \ \frac{1}{2}x^T Qx + c^T x \\
& \text{s.t. } x \in X, \\
& \quad \frac{1}{t}e^T z - g(x,t) \le K, \\
& \quad -x_i \le z_i \le x_i, \qquad i = 1, \ldots, n,
\end{aligned}
$$

Successive Linearization Algorithm

- Step 0: Find an initial feasible solution x^0 of (P) (via ℓ_1-norm relaxation). Choose $\xi^0 \in \partial g(x^0, t)$, set $k = 0$.
- Step 1: Solve the linearization subproblem (a convex QP):

$$
\begin{aligned}
&\min\ f(x) = x^T Qx + c^T x \\
&\text{s.t. } x \in X, \\
&\qquad \frac{1}{t} e^T z - \frac{1}{t}[g(x^k, t) + (\xi^k)^T (x - x^k)] \le K, \\
&\qquad -x_i \le z_i \le x_i, \qquad i = 1, \dots, n
\end{aligned}
$$

  to obtain an optimal solution (x^{k+1}, z^{k+1}).
- Step 2: If $x_{k+1} = x_k$, stop (KKT solution).
- Step 3: Choose $\xi^{k+1} \in \partial g(x^{k+1}, t)$. Set $k := k + 1$ and go to Step 1.

Questions

- How to analyze the relation between the solution of the approximation problem and the solution of (P)?
- How to design *efficient* algorithms for the approximation problems?
- How to recover a feasible solution of (P) from the approximation solution x^*? (e.g., setting some $x_i^* = 0$ if $|x_i^*| \leq \epsilon$ and resolve the QP)
- What is the quality of the recovered feasible solution from the KKT point of (P_t)?

Probabilistic Constraints

- General form of quadratic program with a probabilistic (or chance) constraint:

$$
\begin{aligned}
(P) \qquad & \min \; x^T Qx + c^T x \\
& \text{s.t. } x \in X, \\
& \qquad \mathbb{P}(\xi^T Bx \geq R) \geq 1 - \epsilon,
\end{aligned}
$$

where

$$X = \{x \mid Ex \leq f,\ 0 \leq x \leq u,\ x^T A_i x + b_i^T x + d_i \leq 0,\ i = 1, \ldots, r\},$$

Q and A_i, $i = 1, \ldots, r$ are positive semidefinite $n \times n$ symmetric matrices, $c \in \Re^n$, B is an $m \times n$ matrix, ξ is a random vector in $\Re^m$, $\mathbb{P}$ denotes the probability, $0 < \epsilon < 1$.

VaR constrained portfolio selection

- The Value-at-Risk (VaR) constraint can be expressed as

$$\mathbb{P}(\xi^T x \geq R) \geq 1 - \epsilon,$$

where $\xi^T x$ represents the random return of the portfolio with weight vector x, R is the prescribed minimal level of return, and ϵ is usually a small number, $\epsilon = 0.05$, for example.

- The VaR-constrained mean-variance portfolio selection model:

$$\begin{aligned} \min \ & x^T \Sigma x - \sigma \mu^T x \\ \text{s.t. } & \mathbb{P}(x^T \xi \geq R) \geq 1 - \epsilon, \\ & x \in X. \end{aligned}$$

Existing solution methods

- Extensive study for LP with a special probabilistic constraint: $\mathbb{P}(Ax \geq \xi) \geq 1 - \epsilon$, where ξ is a random vector, Prékopa (2003), Ruszczynski (2002), Luedtke, Ahmed and Nemhauser (2010) ...
- If the random vector ξ has a known (continuous) distribution, then safe (conservative) approximation technique can be used to obtain a convex approximation, e.g., CVaR approximation, Nemirovski and Shapiro (2006).
- Scenario approximation is another way of constructing tractable convex approximations to probabilistic constraint. Lower bounds of sample size to ensure the feasibility of the solution generated from scenario approximations are derived in Calafiore and Campi (2005, 2006) and Nemirovski and Shapiro (2009).

- Suppose that ξ has a *finite discrete distribution*: ξ takes finite number of values $\xi^1, \ldots, \xi^N \in \Re^m$ with equal probability, called scenarios.
- Let α_i be the minimum value of $\xi_i^T Bx$ for all possible scenarios, i.e, $(\xi^i)^T Bx \geq \alpha_i$, $i = 1, \ldots, N$. Let $K = \lfloor N\epsilon \rfloor$.
- Then, (P) can be reformulated as a mixed integer QP program (standard MIQP):

$$
(\mathrm{MIQP}_0) \qquad \begin{aligned} \min\ & x^T Qx + c^T x \\ \text{s.t. } & (\xi^i)^T Bx \geq R + y_i(\alpha_i - R),\ i = 1, \ldots, N, \\ & \sum_{i=1}^{T} y_i \leq K, \\ & x \in X,\ y_i \in \{0, 1\},\ i = 1, \ldots, N. \end{aligned}
$$

A new reformulation Lagrangian decomposition

- Define

$$\begin{aligned} \alpha_i &= \min_{x \in X} \ (\xi^i)^T Bx, \ i = 1, \dots, N \\ \beta_i &= \max_{x \in X} \ (\xi^i)^T Bx, \ i = 1, \dots, N \\ \Theta &= \{\theta \in \Re^N \mid Q - \sum_{i=1}^{N} \theta_i B^T \xi^i (\xi^i)^T B \succeq 0\}. \end{aligned}$$

- For any $\theta \in \Theta$, problem (P) can be written as

$$\begin{aligned} (P_\theta) \qquad \min \ & x^T(Q - \sum_{i=1}^{N} \theta_i B^T \xi^i (\xi^i)^T B)x + c^T x + \sum_{i=1}^{N} \theta_i v_i^2 \\ \text{s.t. } & v_i = (\xi^i)^T Bx, \quad i = 1, \dots, N, \quad \text{(link constraint)} \\ & v_i \geq R + y_i(\alpha_i - R), \ i = 1, \dots, N, \\ & e^T y \leq K \\ & x \in X, \ \alpha \leq v \leq \beta, \ y \in \{0, 1\}^N. \end{aligned}$$

- Associating a multiplier λ_i to the link constraint $v_i = (\xi^i)^T Bx$, we have the following Lagrangian relaxation problem of (P):

$$d(\lambda) = d_1(\lambda) + d_2(\lambda),$$

where

$$\begin{aligned} d_1(\lambda) &= \min\ x^T(Q - \sum_{i=1}^N \theta_i B^T \xi^i (\xi^i)^T B)x + (c - \sum_{i=1}^N \lambda_i B^T \xi^i)^T x \\ &\quad \text{s.t. } x \in X \\ d_2(\lambda) &= \min\ \sum_{i=1}^N \theta_i v_i^2 + \lambda_i v_i \\ &\quad \text{s.t. } v_i \geq R + y_i(\alpha_i - R),\ y_i \in \{0,1\},\ i = 1,\dots,N, \\ &\qquad\quad e^T y \leq K,\ \alpha \leq v \leq \beta. \end{aligned}$$

- $d_1(\lambda)$ and $d_2(\lambda)$ are two easy subproblems!

- The Lagrangian dual of (P_θ) is

$$(D_\theta) \quad \max_{\lambda} \; d(\lambda)$$

- The best θ can be found via the following program

$$(D) \quad \max_{\theta \in \Theta} \; v(D_\theta)$$

- We can show that
  - (D_θ) can be reduced to an SOCP problem (for fixed $\theta \in \Theta$).
  - (D_θ) is tighter than (or at least as tight as) the continuous relaxation of (MIQP_0) for any fixed $\theta \in \Theta$ and $\theta \geq 0$.
  - (D_θ) is equivalent to the continuous relaxation of a new reformulation of (P).
  - (D) can be reduced to an SDP problem (best bound for all admissible θ).

Challenging problems

- Computational difficulty arises when the number of scenarios (N) is large which leads to a large-scale (number of constraints) MIQP.
- One of the open questions for the MIQP reformulation of probabilistically constrained QP is how to reduce the number of scenario constraints in MIQP using polyhedral properties of the constraints: valid inequalities, cutting planes, scenarios aggregation, scenario clustering? ...
- How to construct approximate or heuristic methods to large-scale QP with probabilistic constraints?

Factor-risk constrained MV model

- We assume that the random return R_i is driven by a group of factors:

$$R_i = \alpha_i + \beta_i^T f + \epsilon_i,$$

where $f \in \Re^m$ is the vector of random factors, α_i is the intercept representing the the alpha value of the asset and $\beta_i \in \Re^m$ is the factor loading sensitivities, and ϵ_i is a random scalar representing the asset-specific return.

- The variance of the portfolio x is

$$\sigma^2(x) = \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_i \beta_j \sigma_{ij} + \sum_{i=1}^{n} x_i^2 \sigma_{\varepsilon_i},$$

where $\beta_j = \sum_{k=1}^{n} \beta_{kj} x_k$, $\sigma_{ij} = \mathrm{Cov}(f_i, f_j)$, $\sigma_{\varepsilon_i}^2 = \mathrm{Var}(\varepsilon_i)$.

- The systematic risk is

$$\sigma_{\mathrm{sys}}^2(\mathrm{x}) = \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_i \beta_j \sigma_{ij} = \sum_{j=1}^{m} \beta_j^2 \sigma_{jj} + \sum_{1 \le i < j \le n}^{m} 2\beta_i \beta_j \sigma_{ij}.$$

- The cross term can be decomposed as

$$2\beta_i\beta_j\sigma_{ij} = \eta_{ij}(2\beta_i\beta_j\sigma_{ij}) + \eta_{ji}(2\beta_i\beta_j\sigma_{ij}).$$

  where

$$\eta_{ij} = \frac{\sigma_{jj}}{\sigma_{ii} + \sigma_{jj}}, \quad \eta_{ji} = \frac{\sigma_{ii}}{\sigma_{ii} + \sigma_{jj}}.$$

- The risk associated with factor f_j as

$$\sigma_{f_j}^2(x) = \beta_j^2\sigma_{jj} + 2\sum_{i=1, i\neq j}^{m} \eta_{ij}\beta_i\beta_j\sigma_{ij}.$$

- The relative risk associated with factor f_j is then given by

$$\frac{\sigma_{f_j}^2(x)}{\sigma_{\mathrm{sys}}^2(\mathrm{x})} = \frac{\beta_j^2\sigma_{jj} + 2\sum_{i=1, i\neq j}^{m} \eta_{ij}\beta_i\beta_j\sigma_{ij}}{\sum_{i=1}^{m}\sum_{k=1}^{m}\beta_i\beta_k\sigma_{ik}}.$$

- Factor-risk control:

$$\frac{\sigma_{f_j}(x)}{\sigma_{sys}(x)} \le \psi_j,\ j \in J \subseteq \{1, \ldots, m\},$$

where $\psi_j \in (0, 1)$ is a given control parameter, which is equivalent to

$$\beta_j^2 \sigma_{jj} + 2 \sum_{i=1, i \neq j}^{m} \eta_{ij} \beta_i \beta_j \sigma_{ij} - \psi_j \sum_{i=1}^{m} \sum_{k=1}^{m} \beta_i \beta_k \sigma_{ik} \le 0,$$

where $\beta_j = \sum_{k=1}^{n} \beta_{kj} x_k$. This is a *nonconvex* quadratic constraint.

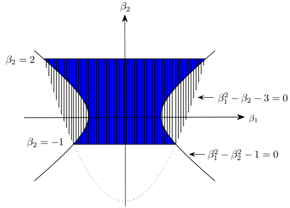
- MV model with factor-risk constraints:

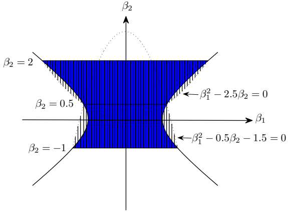
$$
\begin{aligned}
(\mathrm{MV_F}) \quad \min \quad & f(x,\beta) = \sum_{i=1}^{m}\sum_{j=1}^{m} \beta_i\beta_j\sigma_{ij} + \sum_{i=1}^{n} x_i^2\sigma_{\varepsilon_i} \\
\text{s.t.} \quad & \sum_{i=1}^{n} \mu_i x_i \ge \rho, \\
& \beta_j = \sum_{i=1}^{n} x_i\beta_{ij}, \quad j = 1,\dots,m, \\
& \beta_j^2\sigma_{jj} + 2\sum_{i=1, i\neq j}^{m} \eta_{ij}\beta_i\beta_j\sigma_{ij} - \psi_j \sum_{i=1}^{m}\sum_{k=1}^{m} \beta_i\beta_k\sigma_{ik} \le 0, \quad j \in J, \\
& x \in \mathcal{X}, \ \beta \in \mathcal{R}^m.
\end{aligned}
$$

- This is a quadratic program with *nonconvex quadratic constraints*.

Convex outer approximation to nonconvex quadratic constraint

- Example: $\beta_1^2 - \beta_2^2 \leq 1$.
- Convex outer approximation:





Conclusions

- Many challenging modeling and algorithmic problems arising from Financial optimization:
  - Cardinality constraint (sparse solution, zero-norm problem)
  - Probabilistic constraints (VaR constraints)
  - Factor-risk constraints
- Discrete, combinatorial and nonconvex nature
- Solution methods:
  - p-norm approximation, D.C. approximation
  - Lagrangian decomposition, SDP and SOCP approximation
  - Large-scale problem with scenario approximation
  - Outer and inner approximation to nonconvex quadratic constraints.
- Open questions: global solution (tight reformulations, cutting planes, ...)? or local solution (quality guarantee, efficient heuristics, ...)

References

1. X. J. Zheng, X. L. Sun, D. Li, Convex relaxations and mixed-integer quadratic reformulations for cardinality constrained quadratic programs, working paper, 2010.
2. X. T. Cui, X. J. Zheng and S. S. Zhu, Convex relaxations and MIQCQP reformulations for a class of cardinality constrained portfolio selection problems, *J. Global Optim.*, 2011 (accepted).
3. X. J. Zheng, X. L. Sun, D. Li, Lagrangian decomposition and mixed-integer quadratic programming reformulations for probabilistically constrained quadratic programs, working paper, 2011.
4. S. S. Zhu, D. Li, X. L. Sun, Portfolio selection with marginal risk control, *J. Comput. Finance*, Vol.14(1), 3-28, 2010
5. S.S. Zhu, X. T. Cui, X. L. Sun, D. Li, Factor-risk constrained mean-variance portfolio selection: formulation and global optimization solution approach, *Journal of Risk*, 2011 (accepted)