A reduction of cardinality to comlementarity in sparse optimization

Oleg Burdakov

Linköping University, Sweden

Joint work with:
Christian Kanzow (University of Würzburg, Germany) and
Alexandra Schwartz (TU Darmstadt, Germany)

Cardinality constrained optimization

$$\text{(CardCP)} \qquad \min_x \ f(x) \quad \text{s.t.} \quad x \in X, \ \|x\|_0 \leq \kappa,$$

where $\|x\|_0 = \text{card}(x)$ denotes the number of nonzero elements in $x \in R^n$, and $X \subseteq \mathbb{R}^n$ is a subset that contains any further constraints on x.

Applications

- signal processing
- portfolio selection
- subset selection
- uncapacitated facility location
- etc.

Xiaoling Sun, Xiaojin Zheng and Duan Li
Recent Advances in Mathematical Programming with Semi-continuous Variables and Cardinality Constraint, JORC (2013) 1:55–77

Applications

- signal processing
- portfolio selection
- subset selection
- uncapacitated facility location
- etc.

Xiaoling Sun, Xiaojin Zheng and Duan Li
Recent Advances in Mathematical Programming with Semi-continuous Variables and Cardinality Constraint, JORC (2013) 1:55–77

Applications

- signal processing
- portfolio selection
- subset selection
- uncapacitated facility location
- etc.

Xiaoling Sun, Xiaojin Zheng and Duan Li
Recent Advances in Mathematical Programming with Semi-continuous Variables and Cardinality Constraint, JORC (2013) 1:55–77

Applications

- signal processing
- portfolio selection
- subset selection
- uncapacitated facility location
- etc.

Xiaoling Sun, Xiaojin Zheng and Duan Li
Recent Advances in Mathematical Programming with Semi-continuous Variables and Cardinality Constraint, JORC (2013) 1:55–77

Applications

- signal processing
- portfolio selection
- subset selection
- uncapacitated facility location
- etc.

Xiaoling Sun, Xiaojin Zheng and Duan Li
Recent Advances in Mathematical Programming with Semi-continuous Variables and Cardinality Constraint, JORC (2013) 1:55–77

Applications

- signal processing
- portfolio selection
- subset selection
- uncapacitated facility location
- etc.

Xiaoling Sun, Xiaojin Zheng and Duan Li
Recent Advances in Mathematical Programming with Semi-continuous Variables and Cardinality Constraint, JORC (2013) 1:55–77

Approaches

- greedy algorithms
- approximation of l_0-norm by l_p-norms ($0 < p \leq 1$), l_1-norm $\Rightarrow$ convex relaxation
- parametric approximation of l_0-norm, e.g. by a parametric family of exponential functions (O. Mangasarian, 1999)
- DC Approximation of l_0-norm (X.J. Zheng, X.L. Sun, D. Li and J. Sun, 2012)
- branch-and-bound and branch-and-cut algorithms
- heuristic procedures based on metaheuristics (e.g. genetic algorithms, tabu search, simulated annealing)
- etc.
- reduction of cardinality to comlementarity

Approaches

- greedy algorithms
- approximation of l_0-norm by l_p-norms $(0 < p \leq 1)$,
  l_1-norm $\Rightarrow$ convex relaxation
- parametric approximation of l_0-norm, e.g. by a parametric family of exponential functions (O. Mangasarian, 1999)
- DC Approximation of l_0-norm (X.J. Zheng, X.L. Sun, D. Li and J. Sun, 2012)
- branch-and-bound and branch-and-cut algorithms
- heuristic procedures based on metaheuristics (e.g. genetic algorithms, tabu search, simulated annealing)
- etc.
- reduction of cardinality to comlementarity

Approaches

- greedy algorithms
- approximation of l_0-norm by l_p-norms $(0 < p \leq 1)$,
  l_1-norm $\Rightarrow$ convex relaxation
- parametric approximation of l_0-norm, e.g. by a parametric family of exponential functions (O. Mangasarian, 1999)
- DC Approximation of l_0-norm (X.J. Zheng, X.L. Sun, D. Li and J. Sun, 2012)
- branch-and-bound and branch-and-cut algorithms
- heuristic procedures based on metaheuristics (e.g. genetic algorithms, tabu search, simulated annealing)
- etc.
- reduction of cardinality to comlementarity

Approaches

- greedy algorithms
- approximation of l_0-norm by l_p-norms $(0 < p \leq 1)$,
  l_1-norm $\Rightarrow$ convex relaxation
- parametric approximation of l_0-norm, e.g. by a parametric family of exponential functions (O. Mangasarian, 1999)
- DC Approximation of l_0-norm (X.J. Zheng, X.L. Sun, D. Li and J. Sun, 2012)
- branch-and-bound and branch-and-cut algorithms
- heuristic procedures based on metaheuristics (e.g. genetic algorithms, tabu search, simulated annealing)
- etc.
- reduction of cardinality to comlementarity

Approaches

- greedy algorithms
- approximation of l_0-norm by l_p-norms ($0 < p \leq 1$),
  l_1-norm $\Rightarrow$ convex relaxation
- parametric approximation of l_0-norm, e.g. by a parametric family of exponential functions (O. Mangasarian, 1999)
- DC Approximation of l_0-norm (X.J. Zheng, X.L. Sun, D. Li and J. Sun, 2012)
- branch-and-bound and branch-and-cut algorithms
- heuristic procedures based on metaheuristics (e.g. genetic algorithms, tabu search, simulated annealing)
- etc.
- reduction of cardinality to comlementarity

Approaches

- greedy algorithms
- approximation of l_0-norm by l_p-norms ($0 < p \leq 1$),
  l_1-norm $\Rightarrow$ convex relaxation
- parametric approximation of l_0-norm, e.g. by a parametric family of exponential functions (O. Mangasarian, 1999)
- DC Approximation of l_0-norm (X.J. Zheng, X.L. Sun, D. Li and J. Sun, 2012)
- branch-and-bound and branch-and-cut algorithms
- heuristic procedures based on metaheuristics (e.g. genetic algorithms, tabu search, simulated annealing)
- etc.
- reduction of cardinality to comlementarity

Approaches

- greedy algorithms
- approximation of l_0-norm by l_p-norms $(0 < p \leq 1)$,
  l_1-norm $\Rightarrow$ convex relaxation
- parametric approximation of l_0-norm, e.g. by a parametric family of exponential functions (O. Mangasarian, 1999)
- DC Approximation of l_0-norm (X.J. Zheng, X.L. Sun, D. Li and J. Sun, 2012)
- branch-and-bound and branch-and-cut algorithms
- heuristic procedures based on metaheuristics (e.g. genetic algorithms, tabu search, simulated annealing)
- etc.
- reduction of cardinality to comlementarity

Approaches

- greedy algorithms
- approximation of l_0-norm by l_p-norms $(0 < p \leq 1)$,
  l_1-norm $\Rightarrow$ convex relaxation
- parametric approximation of l_0-norm, e.g. by a parametric family of exponential functions (O. Mangasarian, 1999)
- DC Approximation of l_0-norm (X.J. Zheng, X.L. Sun, D. Li and J. Sun, 2012)
- branch-and-bound and branch-and-cut algorithms
- heuristic procedures based on metaheuristics (e.g. genetic algorithms, tabu search, simulated annealing)
- etc.
- reduction of cardinality to comlementarity

Approaches

- greedy algorithms
- approximation of l_0-norm by l_p-norms $(0 < p \leq 1)$,
  l_1-norm $\Rightarrow$ convex relaxation
- parametric approximation of l_0-norm, e.g. by a parametric family of exponential functions (O. Mangasarian, 1999)
- DC Approximation of l_0-norm (X.J. Zheng, X.L. Sun, D. Li and J. Sun, 2012)
- branch-and-bound and branch-and-cut algorithms
- heuristic procedures based on metaheuristics (e.g. genetic algorithms, tabu search, simulated annealing)
- etc.
- reduction of cardinality to comlementarity

Approaches

- greedy algorithms
- approximation of l_0-norm by l_p-norms $(0 < p \leq 1)$,
  l_1-norm $\Rightarrow$ convex relaxation
- parametric approximation of l_0-norm, e.g. by a parametric family of exponential functions (O. Mangasarian, 1999)
- DC Approximation of l_0-norm (X.J. Zheng, X.L. Sun, D. Li and J. Sun, 2012)
- branch-and-bound and branch-and-cut algorithms
- heuristic procedures based on metaheuristics (e.g. genetic algorithms, tabu search, simulated annealing)
- etc.
- reduction of cardinality to comlementarity

Example 1: Compressed sensing

$$\min_{x \in R^n} \|b - Ax\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \leq \kappa,$$

where $A \in R^{m \times n}$ (typically, $m << n$)

Example 1: Compressed sensing

$$\min_{x \in R^n} \|b - Ax\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \leq \kappa,$$

where $A \in R^{m \times n}$ (typically, $m << n$)

Example 2: Sparse multilinear least squares minimization

$$\min_{x \in R^n} \|b - (A_1 x_1) \circ (A_2 x_2) \circ \ldots \circ (A_L x_L)\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \leq \kappa,$$

where $x = (x_1, x_2, \ldots, x_L)$, $x_i \in R^{n_i}$, $n_1 + \ldots + n_L = n$,
$u \circ v$ is the component-wise product of u and v.

M. Andersson, OB, H. Knutsson, S. Zikrin
Sparsity Optimization in Design of Multidimensional Filter Networks,
Optimization and Engineering (2015), 16(2): 259–277

Example 3: Mean-variance portfolio selection with cardinality and minimum buy-in threshold

Given n risky assets with the vector of their expected returns μ and the covariance matrix Q. Let x_i be the fraction of the available capital to be invested into asset $i = 1, \ldots, n$ with the minimum buy-in threshold a_i, and let ρ be the desired expected return. Then $\mu^T x$ is the expected portfolio return and $x^T Q x$ is the variance of the portfolio (a measure of the risk).

Portfolio selection model

$$\begin{array}{ll} \min & x^T Q x \\ \text{s.t.} & \mu^T x \geq \rho, \quad e^T x = 1, \quad \|x\|_0 \leq \kappa, \\ & x_i \in \{0\} \cup [a_i, b_i], \quad i = 1, \ldots, n \end{array}$$

Remark: $x_1, \ldots, x_n$ are semi-continuous variables.

Example 3: Mean-variance portfolio selection with cardinality and minimum buy-in threshold

Given n risky assets with the vector of their expected returns μ and the covariance matrix Q. Let x_i be the fraction of the available capital to be invested into asset $i = 1, \ldots, n$ with the minimum buy-in threshold a_i, and let ρ be the desired expected return. Then $\mu^T x$ is the expected portfolio return and $x^T Q x$ is the variance of the portfolio (a measure of the risk).

Portfolio selection model

$$\begin{array}{ll} \min & x^T Q x \\ \text{s.t.} & \mu^T x \geq \rho, \quad e^T x = 1, \quad \|x\|_0 \leq \kappa, \\ & x_i \in \{0\} \cup [a_i, b_i], \quad i = 1, \ldots, n \end{array}$$

Remark: $x_1, \ldots, x_n$ are semi-continuous variables.

Example 3: Mean-variance portfolio selection with cardinality and minimum buy-in threshold

Given n risky assets with the vector of their expected returns μ and the covariance matrix Q. Let x_i be the fraction of the available capital to be invested into asset $i = 1, \ldots, n$ with the minimum buy-in threshold a_i, and let ρ be the desired expected return. Then $\mu^T x$ is the expected portfolio return and $x^T Q x$ is the variance of the portfolio (a measure of the risk).

Portfolio selection model

$$\begin{array}{ll} \min & x^T Q x \\ \text{s.t.} & \mu^T x \geq \rho, \quad e^T x = 1, \quad \|x\|_0 \leq \kappa, \\ & x_i \in \{0\} \cup [a_i, b_i], \quad i = 1, \ldots, n \end{array}$$

Remark: $x_1, \ldots, x_n$ are semi-continuous variables.

Example 3: Mean-variance portfolio selection with cardinality and minimum buy-in threshold

Given n risky assets with the vector of their expected returns μ and the covariance matrix Q. Let x_i be the fraction of the available capital to be invested into asset $i = 1, \ldots, n$ with the minimum buy-in threshold a_i, and let ρ be the desired expected return. Then $\mu^T x$ is the expected portfolio return and $x^T Qx$ is the variance of the portfolio (a measure of the risk).

Portfolio selection model

$$\begin{aligned} \min \quad & x^T Q x \\ \text{s.t.} \quad & \mu^T x \geq \rho, \quad e^T x = 1, \quad \|x\|_0 \leq \kappa, \\ & x_i \in \{0\} \cup [a_i, b_i], \quad i = 1, \ldots, n \end{aligned}$$

Remark: $x_1, \ldots, x_n$ are semi-continuous variables.

Example 3: Mean-variance portfolio selection with cardinality and minimum buy-in threshold

Given n risky assets with the vector of their expected returns μ and the covariance matrix Q. Let x_i be the fraction of the available capital to be invested into asset $i = 1, \ldots, n$ with the minimum buy-in threshold a_i, and let ρ be the desired expected return. Then $\mu^T x$ is the expected portfolio return and $x^T Qx$ is the variance of the portfolio (a measure of the risk).

Portfolio selection model

$$\begin{array}{ll} \min & x^T Q x \\ \text{s.t.} & \mu^T x \geq \rho, \quad e^T x = 1, \quad \|x\|_0 \leq \kappa, \\ & x_i \in \{0\} \cup [a_i, b_i], \quad i = 1, \ldots, n \end{array}$$

Remark: $x_1, \ldots, x_n$ are semi-continuous variables.

Cardinality constraints in disjunctive form

Lemma

For any $\kappa \in N$, *where* $N = \{1, 2, \ldots, n\}$, *the set identity*

$$\{x \in R^n : \ \|x\|_0 \leq \kappa\} = X_\kappa$$

holds, where

$$X_\kappa = \bigcup_{J \subseteq N: \ |J| = \kappa} \{x \in R^n : \ x_i = 0, \ \forall i \in N \setminus J\}.$$

Corollary

(CardCP) is equivalent to the following disjunctive program

$$\min_{x \in X \cap X_\kappa} f(x).$$

Cardinality constraints in disjunctive form

Lemma

For any $\kappa \in N$, *where* $N = \{1, 2, \dots, n\}$, *the set identity*

$$\{x \in R^n : \ \|x\|_0 \leq \kappa\} = X_\kappa$$

holds, where

$$X_\kappa = \bigcup_{J \subseteq N: \ |J| = \kappa} \{x \in R^n : \ x_i = 0, \ \forall i \in N \setminus J\}.$$

Corollary

(CardCP) is equivalent to the following disjunctive program

$$\min_{x \in X \cap X_\kappa} \ f(x).$$

Local feasible variations in CardCPs

For any x^* such that $\|x^*\|_0 \leq \kappa$, denote

$$Z(x^*) = \begin{cases} \{x \in R^n : \; x_i = 0, \forall i \in I_0(x^*)\}, & \text{if } \|x^*\|_0 = \kappa, \\ \bigcup\limits_{I \subseteq I_0(x^*): \; |I| = n-\kappa} \{x \in R^n : \; x_i = 0, \forall i \in I\}, & \text{if } \|x^*\|_0 < \kappa. \end{cases}$$

where $I_0(x) = \{i \in N : \; x_i = 0\}$.

Proposition

For any feasible point x^* in (CardCP), there exists a radius $r > 0$ such that

$$\{x \in X : \; \|x\|_0 \leq \kappa\} \cap B_r(x^*) = X \cap Z(x^*) \cap B_r(x^*).$$

$\Rightarrow$ The disjunctive nature of (CardCP) affects the optimality conditions only when the cardinality constraint is not active.

Local feasible variations in CardCPs

For any x^* such that $\|x^*\|_0 \le \kappa$, denote

$$Z(x^*) = \begin{cases} \{x \in R^n : \; x_i = 0, \forall i \in I_0(x^*)\}, & \text{if } \|x^*\|_0 = \kappa, \\ \bigcup\limits_{I \subseteq I_0(x^*) :\; |I| = n-\kappa} \{x \in R^n : \; x_i = 0, \forall i \in I\}, & \text{if } \|x^*\|_0 < \kappa. \end{cases}$$

where $I_0(x) = \{i \in N : \; x_i = 0\}$.

Proposition

For any feasible point x^* in (CardCP), there exists a radius $r > 0$ such that

$$\{x \in X : \; \|x\|_0 \le \kappa\} \cap B_r(x^*) = X \cap Z(x^*) \cap B_r(x^*).$$

$\Rightarrow$ The disjunctive nature of (CardCP) affects the optimality conditions only when the cardinality constraint is not active.

Local feasible variations in CardCPs

For any x^* such that $\|x^*\|_0 \le \kappa$, denote

$$Z(x^*) = \begin{cases} \{x \in R^n : \ x_i = 0, \forall i \in I_0(x^*)\}, & \text{if } \|x^*\|_0 = \kappa, \\ \bigcup\limits_{I \subseteq I_0(x^*):\ |I| = n-\kappa} \{x \in R^n : \ x_i = 0, \forall i \in I\}, & \text{if } \|x^*\|_0 < \kappa. \end{cases}$$

where $I_0(x) = \{i \in N : \ x_i = 0\}$.

Proposition

For any feasible point x^* in (CardCP), there exists a radius $r > 0$ such that

$$\{x \in X : \ \|x\|_0 \le \kappa\} \cap B_r(x^*) = X \cap Z(x^*) \cap B_r(x^*).$$

⇒ The disjunctive nature of (CardCP) affects the optimality conditions only when the cardinality constraint is not active.

Local optimality conditions for CardCPs

$$\text{(CardCP)} \qquad \min_x \; f(x) \quad \text{s.t.} \quad x \in X, \; \|x\|_0 \leq \kappa.$$

- For $X = R^n$, the necessary optimality conditions are:

$$\nabla_{x_i} f(x^*) = 0, \quad \forall i : \; x_i^* \neq 0.$$

Amir Beck and Yonina Eldar
Sparsity constrained nonlinear optimisation: optimality conditions and algorithms, SIAM J. Optim. (2013) 23:1480-1509

- For $X = \{x \in R^n : \; g(x) \leq 0, h(x) = 0\}$, the necessary optimality conditions are:

$$\nabla_{x_i} L(x^*, \lambda^*, \mu^*) = 0, \quad \forall i : \; x_i^* \neq 0.$$

Local optimality conditions for CardCPs

$$\text{(CardCP)} \qquad \min_x \; f(x) \quad \text{s.t.} \quad x \in X, \; \|x\|_0 \le \kappa.$$

- For $X = R^n$, the necessary optimality conditions are:

$$\nabla_{x_i} f(x^*) = 0, \quad \forall i : \; x_i^* \neq 0.$$

Amir Beck and Yonina Eldar
Sparsity constrained nonlinear optimisation: optimality conditions and algorithms, SIAM J. Optim. (2013) 23:1480-1509

- For $X = \{x \in R^n : \; g(x) \le 0, h(x) = 0\}$, the necessary optimality conditions are:

$$\nabla_{x_i} L(x^*, \lambda^*, \mu^*) = 0, \quad \forall i : \; x_i^* \neq 0.$$

Standard mixed integer reformulation

- Consider the case

$$0 \le x \le u, \quad \|x\|_0 \le \kappa.$$

- Introduce auxiliary variables counting the nonzeros of x:

$$z_i \in \{0, 1\}, \quad 0 \le x_i \le u_i z_i, \quad e^T z \le \kappa.$$

- Solve a mixed integer problem.
- The relaxation $z_i \in [0, 1]$ preserves possible linearity / convexity, but it doesn't preserve the required cardinality.

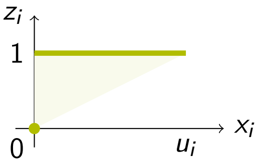
Standard mixed integer reformulation

- Consider the case

$$0 \leq x \leq u, \quad \|x\|_0 \leq \kappa.$$

- Introduce auxiliary variables counting the nonzeros of x:

$$z_i \in \{0,1\}, \quad 0 \leq x_i \leq u_i z_i, \quad e^T z \leq \kappa.$$



- Solve a mixed integer problem.
- The relaxation $z_i \in [0,1]$ preserves possible linearity / convexity, but it doesn't preserve the required cardinality.

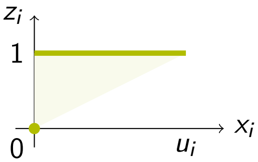
Standard mixed integer reformulation

- Consider the case

$$0 \le x \le u, \quad \|x\|_0 \le \kappa.$$

- Introduce auxiliary variables counting the nonzeros of x:

$$z_i \in \{0, 1\}, \quad 0 \le x_i \le u_i z_i, \quad e^T z \le \kappa.$$



- Solve a mixed integer problem.
- The relaxation $z_i \in [0, 1]$ preserves possible linearity / convexity, but it doesn't preserve the required cardinality.

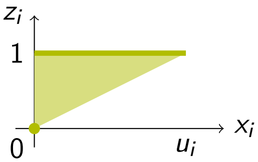
Standard mixed integer reformulation

- Consider the case

$$0 \le x \le u, \quad \|x\|_0 \le \kappa.$$

- Introduce auxiliary variables counting the nonzeros of x:

$$z_i \in \{0,1\}, \quad 0 \le x_i \le u_i z_i, \quad e^T z \le \kappa.$$



- Solve a mixed integer problem.
- The relaxation $z_i \in [0,1]$ preserves possible linearity / convexity, but it doesn't preserve the required cardinality.

Alternative reformulation

- Cardinality constraint: $\|x\|_0 \leq \kappa$
- Mixed-integer reformulation:

$$
\begin{aligned}
& y_i \in \{0,1\}, \quad i = 1, \dots, n, \\
& x_i y_i = 0, \quad i = 1, \dots, n, \\
& e^T y \geq n - \kappa.
\end{aligned}
$$

- Continuous relaxation:

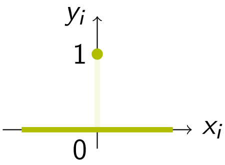
$$
\begin{aligned}
& y_i \in [0,1], \quad i = 1, \dots, n, \\
& x_i y_i = 0, \quad i = 1, \dots, n, \\
& e^T y \geq n - \kappa.
\end{aligned}
$$

- The relaxation preserves the cardinality, but it introduces a nonlinearity (the complementarity condition).

Alternative reformulation

- Cardinality constraint: $\|x\|_0 \leq \kappa$
- Mixed-integer reformulation:

$$
\begin{aligned}
& y_i \in \{0,1\}, \quad i = 1, \dots, n, \\
& x_i y_i = 0, \qquad i = 1, \dots, n, \\
& e^T y \geq n - \kappa.
\end{aligned}
$$



- Continuous relaxation:

$$
\begin{aligned}
& y_i \in [0,1], \quad i = 1, \dots, n, \\
& x_i y_i = 0, \qquad i = 1, \dots, n, \\
& e^T y \geq n - \kappa.
\end{aligned}
$$

- The relaxation preserves the cardinality, but it introduces a nonlinearity (the complementarity condition).

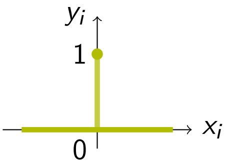
Alternative reformulation

- Cardinality constraint: $\|x\|_0 \leq \kappa$
- Mixed-integer reformulation:

$$\begin{aligned} & y_i \in \{0,1\}, \quad i = 1,\dots,n, \\ & x_i y_i = 0, \quad i = 1,\dots,n, \\ & e^T y \geq n - \kappa. \end{aligned}$$

- Continuous relaxation:

$$\begin{aligned} & y_i \in [0,1], \quad i = 1,\dots,n, \\ & x_i y_i = 0, \quad i = 1,\dots,n, \\ & e^T y \geq n - \kappa. \end{aligned}$$

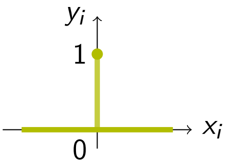


- The relaxation preserves the cardinality, but it introduces a nonlinearity (the complementarity condition).

Alternative reformulation

- Cardinality constraint: $\|x\|_0 \le \kappa$
- Mixed-integer reformulation:

$$\begin{aligned} & y_i \in \{0,1\}, \quad i = 1,\dots,n, \\ & x_i y_i = 0, \quad i = 1,\dots,n, \\ & e^T y \ge n - \kappa. \end{aligned}$$



- Continuous relaxation:

$$\begin{aligned} & y_i \in [0,1], \quad i = 1,\dots,n, \\ & x_i y_i = 0, \quad i = 1,\dots,n, \\ & e^T y \ge n - \kappa. \end{aligned}$$

- The relaxation preserves the cardinality,
  but it introduces a nonlinearity (the complementarity condition).

Mixed integer reformulation of CardCPs with reduction of cardinality to comlementarity

$$
\min_{x,y} \; f(x) \quad \text{s.t.} \quad
\begin{array}{ll}
x \in X & \\
e^T y = n - \kappa, & \\
x_i y_i = 0, & i = 1, \ldots, n, \\
y_i \in \{0, 1\}, & i = 1, \ldots, n.
\end{array}
$$

Relaxed problem

Theorem

A vector $x^* \in R^n$ is a solution to problem (CardCP) if and only if there exists a vector $y^* \in R^n$ such that the pair (x^*, y^*) is a solution to the relaxed problem

$$\begin{aligned} \min_{x,y} \ f(x) \quad s.t. \quad & x \in X \\ & e^T y = n - \kappa, \\ & x_i y_i = 0, \qquad i = 1, \ldots, n, \\ & y_i \in [0, 1], \quad i = 1, \ldots, n. \end{aligned}$$

M. Feng, J.E. Mitchell, J.S. Pang, X. Shen, A. Wächter:
Complementarity formulations of l_0-norm optimization problems (2013)

Relaxed problem

Theorem

A vector $x^* \in R^n$ is a solution to problem (CardCP) if and only if there exists a vector $y^* \in R^n$ such that the pair (x^*, y^*) is a solution to the relaxed problem

$$\begin{aligned} \min_{x,y} \; f(x) \quad s.t. \quad & x \in X \\ & e^T y = n - \kappa, \\ & x_i y_i = 0, \qquad i = 1, \ldots, n, \\ & y_i \in [0, 1], \quad i = 1, \ldots, n. \end{aligned}$$

M. Feng, J.E. Mitchell, J.S. Pang, X. Shen, A. Wächter:
Complementarity formulations of l_0-norm optimization problems (2013)

Local minimizers of CardCP and relaxed problem

Theorem

i) *Let $x^* \in R^n$ be a local minimizer of (CardCP). Then there exists a vector $y^* \in R^n$ such that the pair (x^*, y^*) is also a local minimum of the relaxed problem.*

ii) *Let (x^*, y^*) be a local minimizer of the relaxed problem. Then $\|x^*\|_0 = \kappa$ holds if and only if y^* is unique. In this case, the components of y^* are binary.*

iii) *Let (x^*, y^*) be a local minimizer of the relaxed problem satisfying $\|x^*\|_0 = \kappa$. Then x^* is a local minimum of (CardCP).*

No conclusions, in general, in the case when $\|x^*\|_0 < \kappa$.

Local minimizers of CardCP and relaxed problem

Theorem

i) *Let $x^* \in R^n$ be a local minimizer of (CardCP). Then there exists a vector $y^* \in R^n$ such that the pair (x^*, y^*) is also a local minimum of the relaxed problem.*

ii) *Let (x^*, y^*) be a local minimizer of the relaxed problem. Then $\|x^*\|_0 = \kappa$ holds if and only if y^* is unique. In this case, the components of y^* are binary.*

iii) *Let (x^*, y^*) be a local minimizer of the relaxed problem satisfying $\|x^*\|_0 = \kappa$. Then x^* is a local minimum of (CardCP).*

No conclusions, in general, in the case when $\|x^*\|_0 < \kappa$.

Local minimizers of CardCP and relaxed problem

Theorem

i) *Let $x^* \in R^n$ be a local minimizer of (CardCP). Then there exists a vector $y^* \in R^n$ such that the pair (x^*, y^*) is also a local minimum of the relaxed problem.*

ii) *Let (x^*, y^*) be a local minimizer of the relaxed problem. Then $\|x^*\|_0 = \kappa$ holds if and only if y^* is unique. In this case, the components of y^* are binary.*

iii) *Let (x^*, y^*) be a local minimizer of the relaxed problem satisfying $\|x^*\|_0 = \kappa$. Then x^* is a local minimum of (CardCP).*

No conclusions, in general, in the case when $\|x^*\|_0 < \kappa$.

Local minimizers of CardCP and relaxed problem

Theorem

i) *Let $x^* \in R^n$ be a local minimizer of (CardCP). Then there exists a vector $y^* \in R^n$ such that the pair (x^*, y^*) is also a local minimum of the relaxed problem.*

ii) *Let (x^*, y^*) be a local minimizer of the relaxed problem. Then $\|x^*\|_0 = \kappa$ holds if and only if y^* is unique. In this case, the components of y^* are binary.*

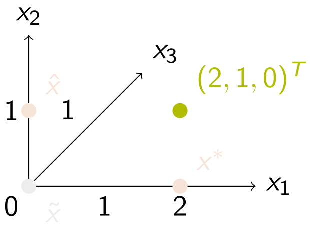
iii) *Let (x^*, y^*) be a local minimizer of the relaxed problem satisfying $\|x^*\|_0 = \kappa$. Then x^* is a local minimum of (CardCP).*

No conclusions, in general, in the case when $\|x^*\|_0 < \kappa$.

Example for spurious local solutions

Consider the 3-dimensional cardinality constrained problem

$$\min_x \|x - (2,1,0)^T\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \le 1.$$



The global minimizer is $x^* = (2,0,0)^T$ and the local minimizer $\hat{x} = (0,1,0)^T$.

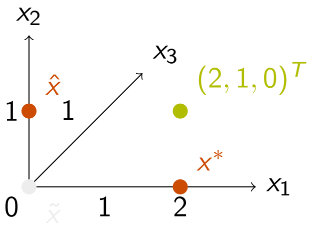
However, the continuous relaxation has additional local minimizer such as $(\tilde{x}, \tilde{y})$ with $\tilde{x} = (0,0,0)^T$ and $\tilde{y} = (1,1,0)$.

Example for spurious local solutions

Consider the 3-dimensional cardinality constrained problem

$$\min_x \|x - (2,1,0)^T\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \leq 1.$$

The global minimizer is $x^* = (2,0,0)^T$ and the local minimizer $\hat{x} = (0,1,0)^T$.



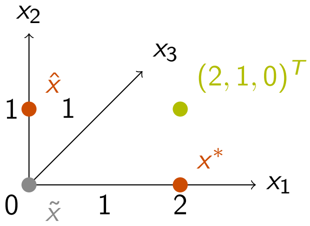
However, the continuous relaxation has additional local minimizer such as $(\tilde{x}, \tilde{y})$ with $\tilde{x} = (0,0,0)^T$ and $\tilde{y} = (1,1,0)$.

Example for spurious local solutions

Consider the 3-dimensional cardinality constrained problem

$$\min_x \|x - (2,1,0)^T\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \le 1.$$

The global minimizer is $x^* = (2,0,0)^T$ and the local minimizer $\hat{x} = (0,1,0)^T$.



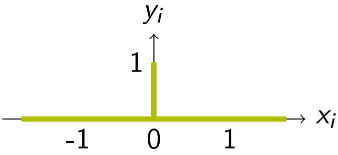
However, the continuous relaxation has additional local minimizer such as $(\tilde{x}, \tilde{y})$ with $\tilde{x} = (0,0,0)^T$ and $\tilde{y} = (1,1,0)$.

Optimality and stationarity for the relaxed problem

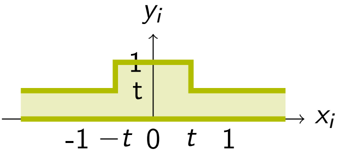
- Constraint qualifications.
- Optimality conditions.
- Though the relaxed problem looks similar to mathematical programs with complementarity constraints (MPCC), it is an easier object to study.

Geometric idea of regularization

Idea: Replace the orthogonality conditions $x_i y_i = 0$ and $0 \leq y_i \leq 1$ by an enlarged regularized set, which approaches the original one as $t \downarrow 0$



original set



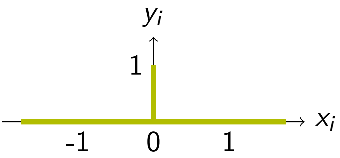
regularized set, $t > 0$

regularized set, $t = 0$

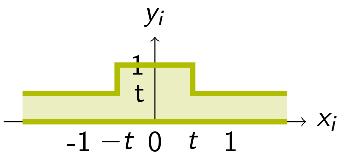
regularized set, $t \geq 1$

Geometric idea of regularization

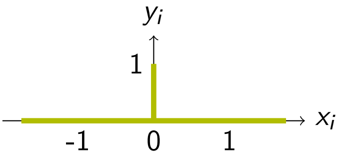
Idea: Replace the orthogonality conditions $x_i y_i = 0$ and $0 \le y_i \le 1$ by an enlarged regularized set, which approaches the original one as $t \downarrow 0$



original set



regularized set, $t > 0$

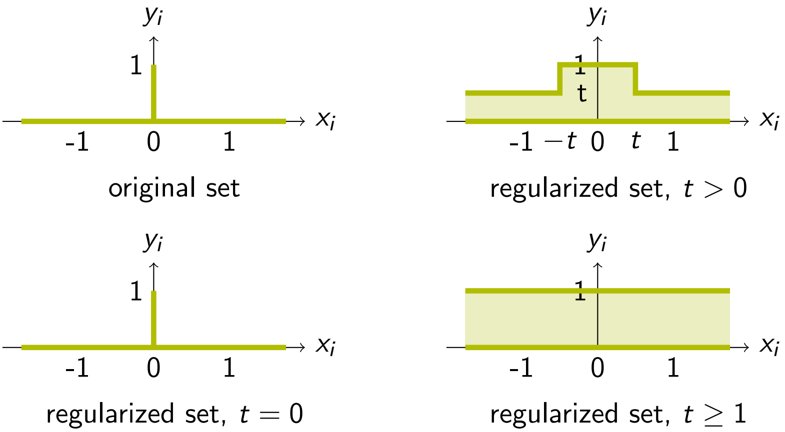


regularized set, $t = 0$

regularized set, $t \ge 1$

Geometric idea of regularization

Idea: Replace the orthogonality conditions $x_i y_i = 0$ and $0 \le y_i \le 1$ by an enlarged regularized set, which approaches the original one as $t \downarrow 0$



Algebraic formulation of the regularization

- Define the functions

$$\varphi^+(a,b;t) \;:=\; \begin{cases} (a-t)(b-t) & \text{if } a+b \geq 2t, \\ -\frac{1}{2}[(a-t)^2+(b-t)^2] & \text{if } a+b < 2t \end{cases},$$

$$\varphi^-(a,b;t) \;:=\; \varphi^+(-a,b;t).$$

- φ^+ and φ^- are continuously differentiable.
- For $t \geq 0$ the function φ^+ has the property

$$\varphi^+(a,b;t) \leq 0 \quad \Longleftrightarrow \quad \min\{a,b\} \leq t.$$

- Similarly, for $t \geq 0$

$$\varphi^-(a,b;t) \leq 0 \quad \Longleftrightarrow \quad \min\{-a,b\} \leq t.$$

Algebraic formulation of the regularization

- Define the functions

$$\varphi^+(a,b;t) \;:=\; \begin{cases} (a-t)(b-t) & \text{if } a+b \geq 2t, \\ -\frac{1}{2}[(a-t)^2+(b-t)^2] & \text{if } a+b < 2t \end{cases},$$

$$\varphi^-(a,b;t) \;:=\; \varphi^+(-a,b;t).$$

- φ^+ and φ^- are continuously differentiable.

- For $t \geq 0$ the function φ^+ has the property

$$\varphi^+(a,b;t) \leq 0 \quad \Longleftrightarrow \quad \min\{a,b\} \leq t.$$

- Similarly, for $t \geq 0$

$$\varphi^-(a,b;t) \leq 0 \quad \Longleftrightarrow \quad \min\{-a,b\} \leq t.$$

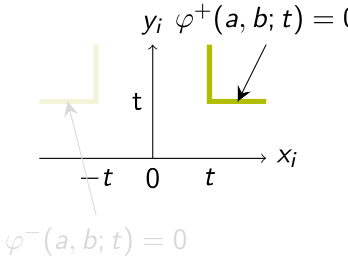
- Define the functions

$$\varphi^{+}(a,b;t) := \begin{cases} (a-t)(b-t) & \text{if } a+b \geq 2t, \\ -\frac{1}{2}[(a-t)^2+(b-t)^2] & \text{if } a+b < 2t \end{cases},$$

$$\varphi^{-}(a,b;t) := \varphi^{+}(-a,b;t).$$

- φ^{+} and φ^{-} are continuously differentiable.
- For $t \geq 0$ the function φ^{+} has the property

$$\varphi^{+}(a,b;t) \leq 0 \iff \min\{a,b\} \leq t.$$



- Similarly, for $t \geq 0$

$$\varphi^{-}(a,b;t) \leq 0 \iff \min\{-a,b\} \leq t.$$

Algebraic formulation of the regularization

- Define the functions

$$\varphi^+(a,b;t) := \begin{cases} (a-t)(b-t) & \text{if } a+b \geq 2t, \\ -\frac{1}{2}[(a-t)^2+(b-t)^2] & \text{if } a+b < 2t \end{cases},$$

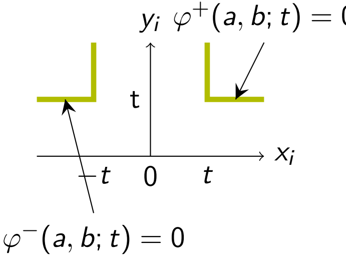
$$\varphi^-(a,b;t) := \varphi^+(-a,b;t).$$

- φ^+ and φ^- are continuously differentiable.
- For $t \geq 0$ the function φ^+ has the property

$$\varphi^+(a,b;t) \leq 0 \iff \min\{a,b\} \leq t.$$

- Similarly, for $t \geq 0$

$$\varphi^-(a,b;t) \leq 0 \iff \min\{-a,b\} \leq t.$$



Regularization approach

- For $t > 0$ replace the complementarity constraints

$$0 \le y_i, \; x_i y_i = 0$$

by the inequalities

$$0 \le y_i, \; \varphi^+(x_i, y_i; t) \le 0, \; \varphi^-(x_i, y_i; t) \le 0$$

- For $t \downarrow 0$ solve a sequence of regularized problems:

$$\begin{aligned} \min_{x,y} f(x) \quad \text{s.t.} \quad & x \in X, \\ & e^T y \ge n - \kappa, \\ & 0 \le y \le e, \\ & \varphi^+(x, y; t) \le 0, \varphi^-(x, y; t) \le 0 \quad \text{component-wise.} \end{aligned}$$

Regularization approach

- For $t > 0$ replace the complementarity constraints

$$0 \le y_i,\ x_i y_i = 0$$

by the inequalities

$$0 \le y_i,\ \varphi^+(x_i, y_i; t) \le 0,\ \varphi^-(x_i, y_i; t) \le 0$$

- For $t \downarrow 0$ solve a sequence of regularized problems:

$$\begin{aligned}
\min_{x,y} f(x) \quad \text{s.t.} \quad & x \in X, \\
& e^T y \ge n - \kappa, \\
& 0 \le y \le e, \\
& \varphi^+(x, y; t) \le 0, \varphi^-(x, y; t) \le 0 \quad \text{component-wise.}
\end{aligned}$$

Test problems

- We used randomly generated portfolio selection problems

$$\min_x x^T Q x \quad \text{s.t.} \quad \mu^T x \geq \rho, \; e^T x \leq 1,$$
$$0 \leq x \leq u,$$
$$\|x\|_0 \leq \kappa.$$

- The data Q, μ, ρ was taken from Frangioni and Gentile (2007) and is available at their website

  `http://www.dt.unipi.it/optimize/Data/MV.html`

- We used 30 instances for each for the dimensions $n = 200, 300, 400$. Each instance was run with $\kappa = 5, 10, 20$, which results in 270 test problems altogether.

Solution alternatives

We compared the following three solution alternatives:

- GUROBI: A mixed-integer solver applied to

$$\min_{x,z} x^T Qx \quad \text{s.t.} \quad \mu^T x \geq \rho,\ e^T x \leq 1,\ e^T z \leq \kappa,$$
$$z \in \{0,1\}^n,\ 0 \leq x \leq u \circ z.$$

- SNOPT: A nonlinear solver applied directly to

$$\min_{x,y} x^T Qx \quad \text{s.t.} \quad \mu^T x \geq \rho,\ e^T x \leq 1,\ e^T y \geq n - \kappa,$$
$$0 \leq y \leq e,\ 0 \leq x \leq u,\ x \circ y = 0.$$

- Regularization : Solve a sequence of regularized problems with SNOPT

$$\min_{x,y} x^T Qx \quad \text{s.t.} \quad \mu^T x \geq \rho,\ e^T x \leq 1,\ e^T y \geq n - \kappa,$$
$$0 \leq y \leq e,\ 0 \leq x \leq u,\ \varphi^+(x, y; t_k) \leq 0.$$

Solution alternatives

We compared the following three solution alternatives:

- GUROBI: A mixed-integer solver applied to

$$\min_{x,z} x^T Q x \quad \text{s.t.} \quad \mu^T x \geq \rho,\ e^T x \leq 1,\ e^T z \leq \kappa,$$
$$z \in \{0,1\}^n,\ 0 \leq x \leq u \circ z.$$

- SNOPT: A nonlinear solver applied directly to

$$\min_{x,y} x^T Q x \quad \text{s.t.} \quad \mu^T x \geq \rho,\ e^T x \leq 1,\ e^T y \geq n - \kappa,$$
$$0 \leq y \leq e,\ 0 \leq x \leq u,\ x \circ y = 0.$$

- Regularization : Solve a sequence of regularized problems with SNOPT

$$\min_{x,y} x^T Q x \quad \text{s.t.} \quad \mu^T x \geq \rho,\ e^T x \leq 1,\ e^T y \geq n - \kappa,$$
$$0 \leq y \leq e,\ 0 \leq x \leq u,\ \varphi^+(x, y; t_k) \leq 0.$$

Solution alternatives

We compared the following three solution alternatives:

- GUROBI: A mixed-integer solver applied to

$$\min_{x,z} x^T Q x \quad \text{s.t.} \quad \mu^T x \geq \rho,\; e^T x \leq 1,\; e^T z \leq \kappa,$$
$$z \in \{0,1\}^n,\; 0 \leq x \leq u \circ z.$$

- SNOPT: A nonlinear solver applied directly to

$$\min_{x,y} x^T Q x \quad \text{s.t.} \quad \mu^T x \geq \rho,\; e^T x \leq 1,\; e^T y \geq n - \kappa,$$
$$0 \leq y \leq e,\; 0 \leq x \leq u,\; x \circ y = 0.$$

- Regularization : Solve a sequence of regularized problems with SNOPT

$$\min_{x,y} x^T Q x \quad \text{s.t.} \quad \mu^T x \geq \rho,\; e^T x \leq 1,\; e^T y \geq n - \kappa,$$
$$0 \leq y \leq e,\; 0 \leq x \leq u,\; \varphi^+(x, y; t_k) \leq 0.$$

Implementation details

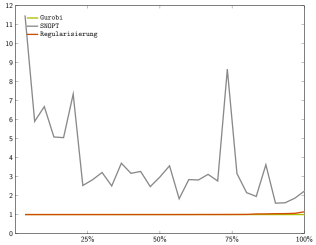
- For all three approaches, only one initial point was used:

$$x^0 = (0, \dots, 0)^T \quad \text{and} \quad y^0 = (1, \dots, 1)^T.$$

- The time limit for running GUROBI was 600 seconds, which corresponds to approximately 2 hours of computation time.
- The initial regularization parameter was $t_0 = 1$, and the updating fromula $t_{k+1} = 0.01 t_k$ was used.
- The regularization method was terminated when either $t_k < 10^{-8}$ or $\max_{i=1,\dots,n} |x_i y_i| \leq 10^{-6}$.

Test Results

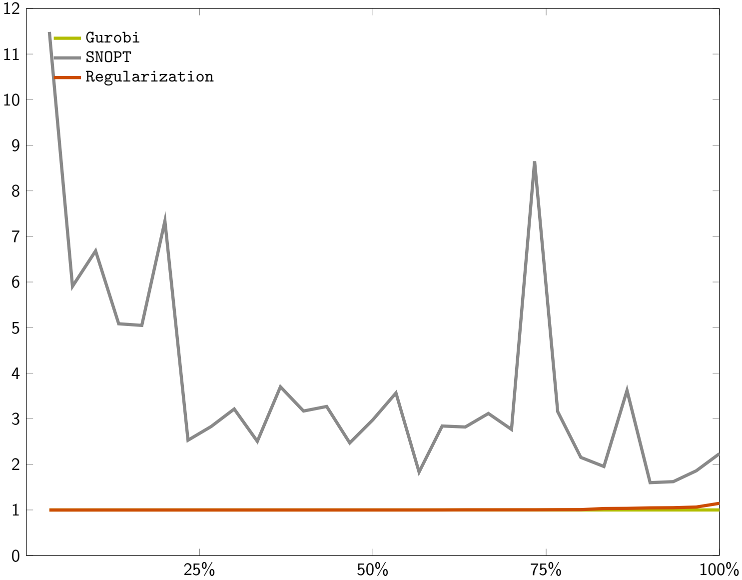
- SNOPT needs in average 0.145 sec, the regularization in average 16.796 sec and GUROBI by definition 600 sec.
- Solutions found by GUROBI and SNOPT are feasible. Solutions found by the regularization satisfy $|x_i| \leq 10^{-6}$ for all but the κ largest components in all but 1 out of 270 examples.
- GUROBI produces the smallest objective function values. SNOPT gets stuck in bad (local?) solutions quite often.



Normalized function values for $n = 300$, $\kappa = 5$

The regularization finds solutions with at most 1% larger function value than GUROBI in 71.5% of the examples.

Normalized objective function values



Continuous reformulation of portfolio optimization problem with semi-continuous variables and cardinality constraint

Theorem

A vector $x^* \in R^n$ is a solution to the portfolio optimization problem

$$\begin{aligned} \min \quad & x^T Q x \\ s.t. \quad & \mu^T x \geq \rho, \quad e^T x = 1, \quad \|x\|_0 \leq \kappa, \\ & x_i \in \{0\} \cup [a_i, b_i], \quad i = 1, \ldots, n \end{aligned}$$

if and only if there exists a vector $y^* \in R^n$ such that the pair (x^*, y^*) is a solution to the problem

$$\begin{aligned} \min \quad & x^T Q x \\ s.t. \quad & \mu^T x \geq \rho, \quad e^T x = 1, \quad e^T y \geq n - \kappa, \\ & x_i y_i = 0, \qquad i = 1, \ldots, n, \\ & 0 \leq y_i \leq 1, \qquad i = 1, \ldots, n, \\ & a_i(1 - y_i) \leq x_i \leq b_i(1 - y_i), \quad i = 1, \ldots, n. \end{aligned}$$