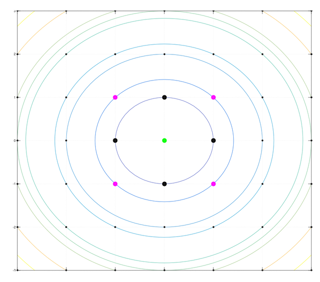

Mixed-Integer Derivative Free Optimization

Jeff Larson, Sven Leyffer, *Prashant Palkar, Stefan Wild

Argonne National Laboratory, *Indian Institute of Technology Bombay

Overview

We consider problem of the form,

$$\underset{x}{\text{minimize}}\ f(x), \quad \text{subject to } x \in \Omega \subset \mathbb{Z}^n \qquad \text{(P)}$$

Ω **bounded**, f **convex** on Ω, f **unevaluatable** if $x \notin \mathbb{Z}^n$.

- Evaluation of f is **expensive**, derivatives are unavailable
- Applications: optimal design of concentrating solar power plants, performance tuning codes for HPC etc.
- Any surrogate model of continuous relaxation can not be solved, prohibiting branch-and-bound.
- IP techniques like Bender's decomposition, Outer Approximation etc. can not be used due to unavailability of ∇f.
- DFO: global optimal not guaranteed (even for convex f).

We provide a method for solving (P) to global optimality under the stated assumptions on f and Ω.

The underestimation lemma

Given a set of points X satisfying $|X| \geq n+1$ and their function values, we define a **secant linear map**, using any $n+1$ affinely independent points, $\boldsymbol{i} \triangleq (i_1, \dots, i_{n+1})$, by solving

$$\begin{bmatrix} \bar{X}^{\boldsymbol{i}} & e \end{bmatrix} \begin{bmatrix} c^{\boldsymbol{i}} \\ b^{\boldsymbol{i}} \end{bmatrix} = f^{\boldsymbol{i}}, \qquad (1)$$

where

$$\bar{X}^{\boldsymbol{i}} \triangleq \begin{bmatrix} (x^{i_1})^T \\ (x^{i_2})^T \\ \vdots \\ (x^{i_{n+1}})^T \end{bmatrix}, \quad e \triangleq \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \text{and} \quad f^{\boldsymbol{i}} \triangleq \begin{bmatrix} f^{i_1} \\ f^{i_2} \\ \vdots \\ f^{i_{n+1}} \end{bmatrix}.$$

***Lemma 1**: Let f be convex and $X^{\boldsymbol{i}}$ be poised. The unique linear map, $m^{\boldsymbol{i}}(x) \triangleq (c^{\boldsymbol{i}})^T x + b^{\boldsymbol{i}}$ satisfying $m^{\boldsymbol{i}}(x^{i_j}) = f(x^{i_j}), \forall i_j \in \boldsymbol{i}$, satisfies, $f(x) \geq m^{\boldsymbol{i}}(x),\ \forall x \in U^{\boldsymbol{i}} \triangleq \bigcup_{j=1}^{n+1} \text{cone}\left(x^{i_j} - X^{\boldsymbol{i}}\right)$.*

***Corollary 1**: The linear mapping $m^{\boldsymbol{i}}(x)$ satisfies $f(x) \geq m^{\boldsymbol{i}}(x),\ \forall x \in \text{cone}\left(x^k - X^{\boldsymbol{i}} \setminus x^{i_j}\right)$, for x^k such that $f(x^k) \geq f(x^{i_j})$.*

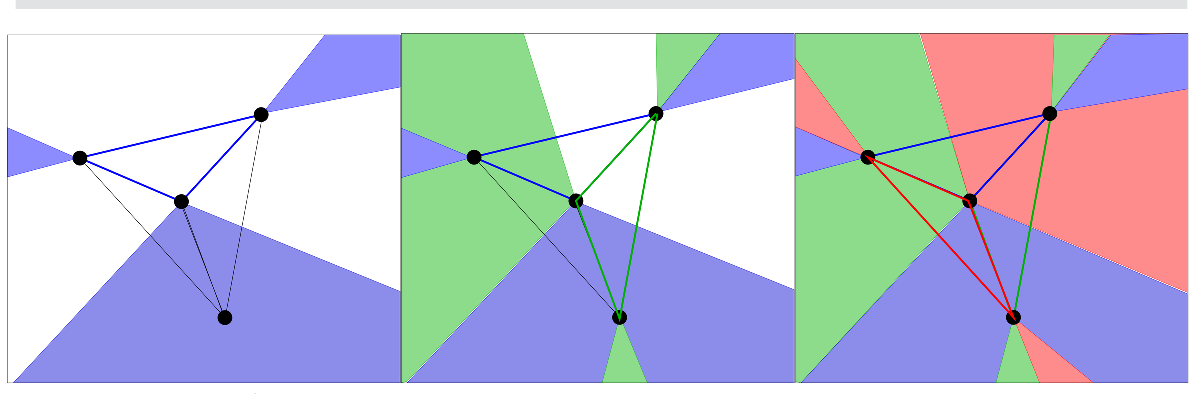

Figure 1 : $U^{\boldsymbol{i}}$: truncated polyhedral cones where secant cuts are valid.

***Proposition 1**: Given a poised set $X = \{x^1, \dots, x^{n+1}\}$ and $x^0 \in int(conv(X))$, let $W_0(X)$ be the set of all combination of $n+1$ points in $\{X \cup x^0\}$ that include x^0. Then,*

(i) the vectors $\{x^0 - x^j\},\ j = 1, \dots n+1$, constitute a positive spanning set.

(ii) $\bigcup_{\boldsymbol{i} \in W_0(X)} \text{cone}\left(x^0 - X^{\boldsymbol{i}}\right) = \mathbb{R}^n$.

Piecewise linear underestimator model

Let $W(X)$ denote the set of all multi-indices $\boldsymbol{i}$ corresponding to poised subsets $X^{\boldsymbol{i}}$ of X.

$$\begin{aligned} &\underset{x}{\text{minimize}} && \eta \\ &\text{subject to} && \eta \geq (c^{\boldsymbol{i}})^T x + b^{\boldsymbol{i}},\ \text{if } x \in U^{\boldsymbol{i}},\ \forall \boldsymbol{i} \in W(X) \qquad \text{(PLM)} \\ &&& x \in \Omega \end{aligned}$$

***Lemma 2**: If f is convex and $W(X)$ is nonempty, then the optimal value η^* of (PLM) satisfies $\eta^* \leq f(x), \forall x \in \Omega$. In fact, $\eta(x) \leq f(x), \forall x \in \Omega$.*

A globally convergent algorithm

Input A set of evaluated points $X^0 \subseteq \Omega$: $|W(X^0)| > 0$.

Find the best point, $\hat{x} \leftarrow \underset{x \in X^0}{\text{argmin}}\ f(x)$, and

set $u_f^1 \leftarrow f(\hat{x})$.

for $k = 1, 2, \dots$ **do**

  Update and solve master problem (PLM).

  Let x^k be an optimal solution to (PLM) and let l_f^k be the

  the value of (PLM) at x^k.

  if $l_f^k = u_f^k$ **then** STOP

  Evaluate $f(x^k)$ and set $X^k = x^k \cup X^{k-1}$.

  if $f(x^k) < f(\hat{x})$ **then**

    Set new incumbent, $\hat{x} = x^k$, and update $u_f^{k+1} \leftarrow f(x^k)$.

  else

    $u_f^{k+1} = u_f^k$

Algorithm 1: Derivative-free convex optimization over Ω

Pictorial demonstration

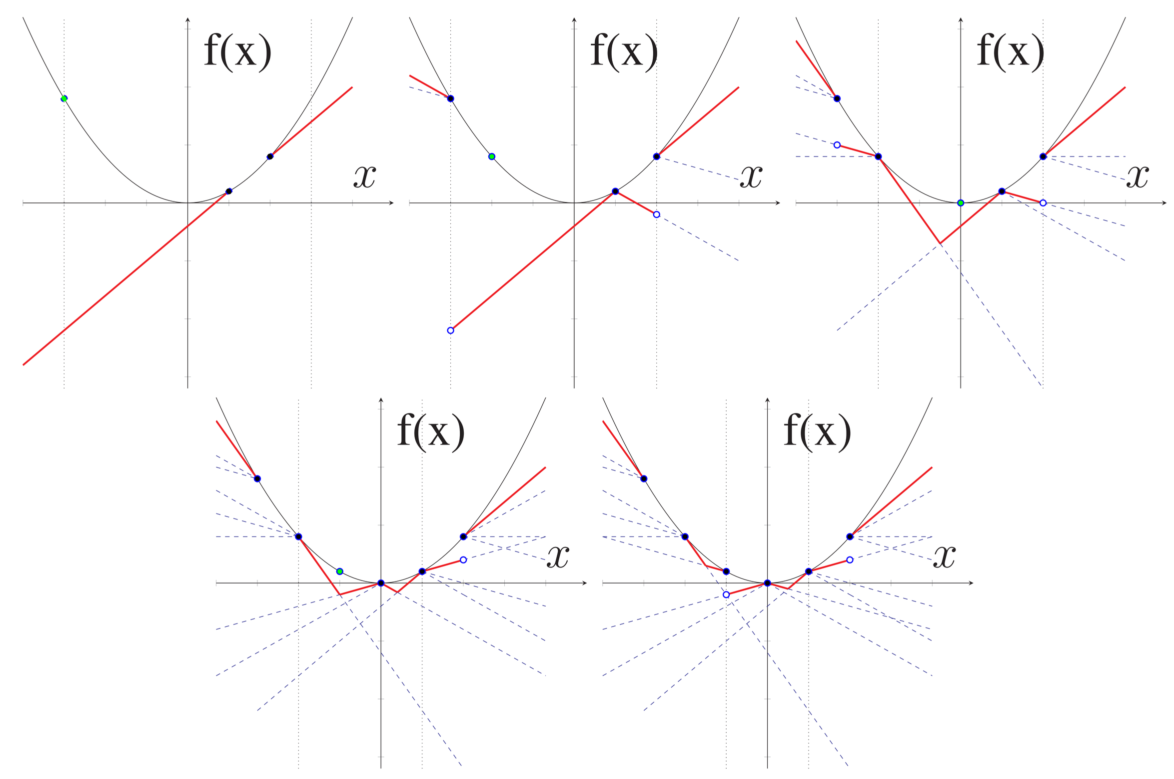

Figure 2 : Solving: minimize $f(x) = x^2$, subject to $x \in [-3, 3], x \in \mathbb{Z}$

***Theorem 1**: If f is convex and Ω is bounded, then Algorithm 1 terminates at a global optimum of (P) in finitely many iterations.*

An MILP formulation for (PLM)

- Define η := maximum of piecewise linear secant functions:

$$\eta \geq (c^{\boldsymbol{i}})^T x + b^{\boldsymbol{i}} - M_{\boldsymbol{i}}\Big(1 - \sum_{j=1}^{n+1} z^{i_j}\Big), \quad \forall \boldsymbol{i} \in W(X) \qquad (2)$$

- Define $z^{i_j} := 1 \Leftrightarrow x \in \text{cone}\left\{x^{i_j} - X^{\boldsymbol{i}}\right\}, i_j \in \boldsymbol{i}$:

$$x = x^{i_j} + \sum_{\substack{l=1,\\ l \neq j}}^{n+1} \lambda_l^{i_j}(x^{i_j} - x^{i_l}), \quad \forall \boldsymbol{i} \in W(X),\ \forall i_j \in \boldsymbol{i}. \qquad (3)$$

- For a given poised set $X^{\boldsymbol{i}}$, the $n+1$ cones are disjoint:

$$\sum_{j=1}^{n+1} z^{i_j} \leq 1,\ \forall \boldsymbol{i} \in W(X), \qquad (4)$$

- A lower bound on λ variables:

$$\lambda_l^{i_j} \geq -M_\lambda(1 - z^{i_j}), \quad \forall \boldsymbol{i} \in W(X),\ \forall i_j, i_l \in \boldsymbol{i}, j \neq l. \qquad (5)$$

- Define $w_l^{i_j} := 0 \Rightarrow \lambda_l^{i_j} < 0$:

$$\lambda_l^{i_j} \leq -\epsilon_\lambda + M_\lambda w_l^{i_j}, \quad \forall \boldsymbol{i} \in W(X),\ \forall i_j, i_l \in \boldsymbol{i}, j \neq l. \qquad (6)$$

- At least one of the $w_l^{i_j}$ be 0 if corresponding $z^{i_j} = 0$:

$$n z^{i_j} \leq \sum_{\substack{l=1,\\ l \neq j}}^{n+1} w_l^{i_j} \leq n - 1 + z^{i_j} \quad \forall \boldsymbol{i} \in W(X),\ \forall i_j \in \boldsymbol{i}. \qquad (7)$$

Finally, the full MIP model is given as:

$$\begin{aligned} &\underset{x,\lambda,z,w}{\text{minimize}} && \eta \\ &\text{subject to} && (2) - (7) \\ &&& w_l^{i_j}, z^{i_j} \in \{0,1\}, \forall l, j \in \{1, \dots, n+1\}, l \neq j; \forall \boldsymbol{i} \\ &&& x \in \Omega. \end{aligned}$$

Derivation of model parameters

- If l_f is a valid lower bound on f over Ω:

$$M_{\boldsymbol{i}} = \max_{x \in \Omega}\{(c^{\boldsymbol{i}})^T x + b^{\boldsymbol{i}}\} - l_f.$$

- False-termination: $f(x) = x^2$, $X = \{-1, 1\}$, $\epsilon_\lambda > 0.5$

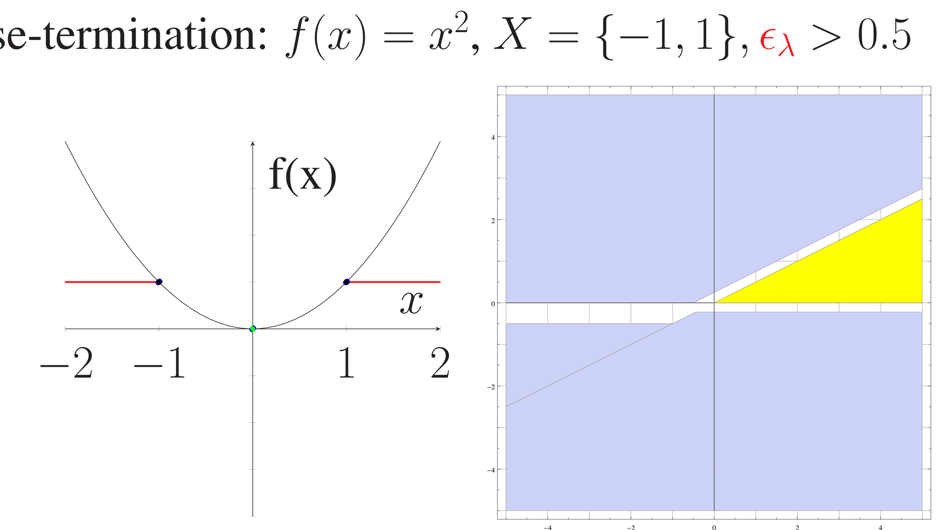

Figure 3 : Effect of insufficient ϵ_λ

- Use hyperplane through $X^i \setminus \{x^{i_j}\}, j = 1 \dots n+1$ to obtain bounds on ϵ_λ; denote by $c^{\boldsymbol{i}_{i_j-}}$, $b^{\boldsymbol{i}_{i_j-}}$ the solution of:

$$\begin{aligned} &\underset{x}{\text{minimize}} && \frac{|(c^{\boldsymbol{i}_{i_j-}})^T x + b^{\boldsymbol{i}_{i_j-}}|}{\|c^{\boldsymbol{i}_{i_j-}}\|_2} \\ &\text{subject to} && |(c^{\boldsymbol{i}_{i_j-}})^T x^{i_l} + b^{\boldsymbol{i}_{i_j-}}| \geq 1 \qquad (\text{P}-\epsilon) \\ &&& x \in \Omega. \end{aligned}$$

***Proposition 2**: The Euclidean distance between $S \triangleq \{x : c^T x + b = 0\}$ and an arbitrary point $\hat{x} \in \mathbb{Z}^n \setminus S$, where c and b are integral, is greater than or equal to $\frac{1}{\|c\|_2}$.*

- Use the proposition:

$$\epsilon_\lambda = \min_{\boldsymbol{i} \in W(X),\ i_j \in \boldsymbol{i}} \frac{1}{\|c^{\boldsymbol{i}_{i_j-}}\|_2}.$$

- Similarly, maximize the objective in (P$-\epsilon$) and set:

$$M_\lambda = \max_{\boldsymbol{i} \in W(X),\ i_j \in \boldsymbol{i}} M_\lambda^{\boldsymbol{i}_{i_j-}}.$$

Alternative formulation of (PLM)

- CPU time of MIP grows exponentially

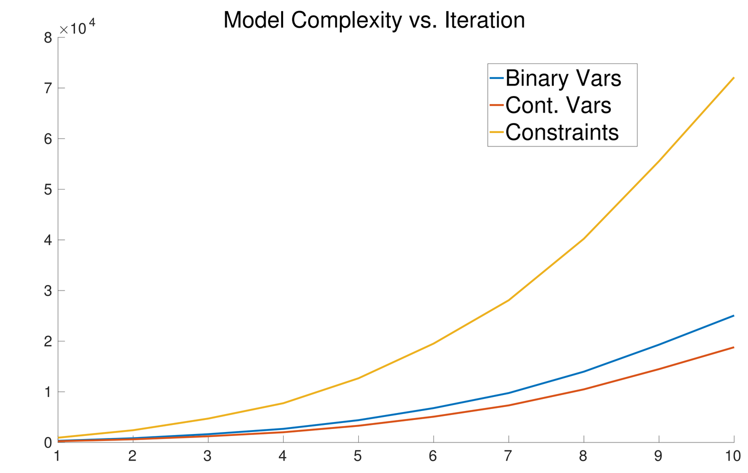

Figure 4 : Model size vs. iterations for Abhishek et al's function [1] (n=3)

- Instead, we update the η value of (PLM) at each grid point:

Input : A set of poised points $X^{\boldsymbol{i}}$

for *all* $i_j \in \boldsymbol{i}$ **do**

  for $l = 1, \dots, L^n$ **do**

    if $x_L^l \in \text{cone}\left(x^{i_j} - X^{\boldsymbol{i}}\right)$ **then**

      Set $\eta^l \leftarrow \max\left(\eta^l, (c^{\boldsymbol{i}})^T x_L^l + b^{\boldsymbol{i}}\right)$, see (2).

Algorithm 2: Routine for updating η values.

Few tricks to obtain a computationally efficient routine:

- Check if $X^{\boldsymbol{i}}$ is poised, and $x \in U^{\boldsymbol{i}}$ using QR factorization:

$$\begin{bmatrix} Q^{\boldsymbol{i}} & R^{\boldsymbol{i}} \end{bmatrix} = \begin{bmatrix} \bar{X}^{\boldsymbol{i}} & e \end{bmatrix}^T. \qquad (8)$$

- Update only $x \in U^{\boldsymbol{i}}$ and where $\eta_k < u_f^k$

Elementary results

Nonsmooth $f(x) := \max_{i \in \{1,\dots,n\}}\left\{\|x - c_i\|^2\right\}$ (see [2])

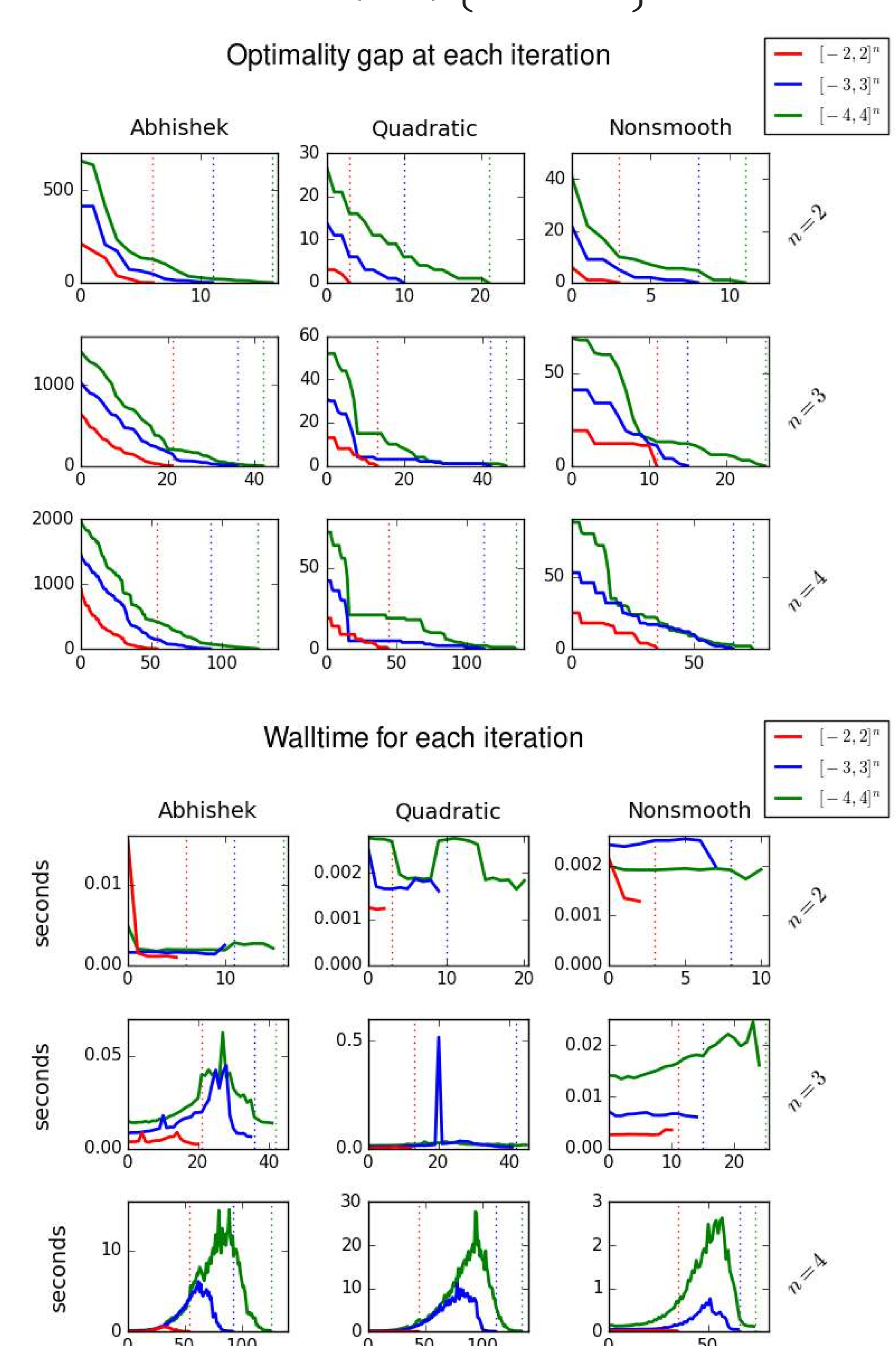

Figure 5 : Results using the alternate formulation of (PLM)

Ongoing interesting stuff

- Selecting 'useful' subsets of $W(X^k \cup x^{k+1})$ [bottleneck]
- Possible extension of ('cones of') underestimation Lemma
- Algorithmic refinements: surrogate models, trust region etc.
- Understanding '(minimal cardinality) optimal sets'

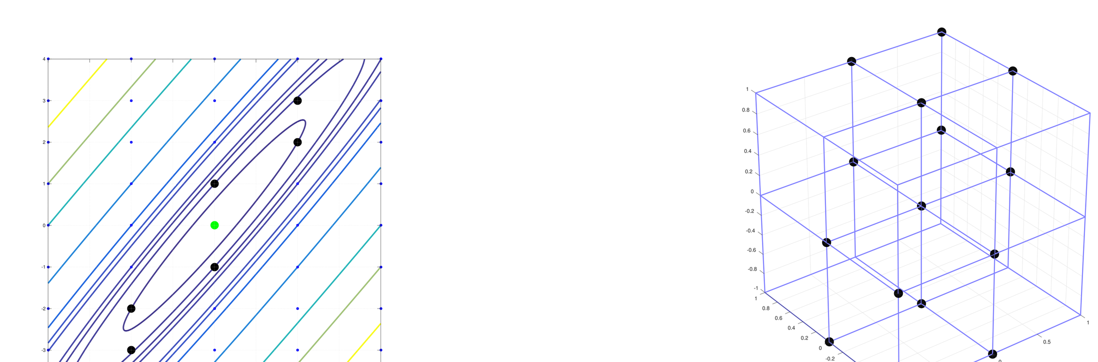

Figure 6 : Optimal sets: $f(x) =$ (i) $64(cx_1 - sx_2)^2 + (sx_1 - cx_2)^2$ [1] (ii) $x_1^2 + x_2^2 + x_3^2$

References

[1] K. Abhishek, S. Leyffer, and J. T. Linderoth, "Modeling without categorical variables: a mixed-integer nonlinear program for the optimization of thermal insulation systems," *Optimization and Engineering*, vol. 11, no. 2, pp. 185–212, 2010.

[2] T. G. Kolda, R. M. Lewis, and V. J. Torczon, "Optimization by direct search: New perspectives on some classical and modern methods," *SIAM Review*, vol. 45, no. 3, pp. 385–482, 2003.