Analysis of a Class of Heaviside Composite Minimization Problems

Shaoning Han

Department of Mathematics
National University of Singapore

2025 December

Collaborators

Ying Cui
IEOR, UC Berkeley

Jong-Shi Pang
ISE, USC

Introduction

Heaviside Composite Optimization Problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \sum_{j=1}^{k} f_{0j}(x) \underbrace{\mathbb{I}_{\mathbb{R}_{++}}(g_{0j}(x))}_{\text{composite indicator}} \\ \text{s.t.} \ & x \in X_{\text{HSC}} \overset{\text{def}}{=} \left\{ x \in P \ \middle| \ \sum_{j=1}^{k} f_{ij}(x) \, \mathbb{I}_{\mathbb{R}_{++}}(g_{ij}(x)) \le b_i, \ i \in [m] \right\} \end{aligned} \tag{HSC}$$

where

- $P \subseteq \mathbb{R}^n$ is a polyhedron
- $f_{ij}, g_{ij} : \mathbb{R}^n \to \mathbb{R}$ are tractable continuous functions

Introduction

Heaviside Composite Optimization Problem

$$\min_{x \in \mathbb{R}^n} \sum_{j=1}^{k} f_{0j}(x) \underbrace{\mathbb{I}_{\mathbb{R}_{++}}(g_{0j}(x))}_{\text{composite indicator}}$$

$$\text{s.t. } x \in X_{\text{HSC}} \stackrel{\text{def}}{=} \left\{ x \in P \;\middle|\; \sum_{j=1}^{k} f_{ij}(x)\, \mathbb{I}_{\mathbb{R}_{++}}(g_{ij}(x)) \leq b_i, \ i \in [m] \right\} \qquad \text{(HSC)}$$

where

- $P \subseteq \mathbb{R}^n$ is a polyhedron
- $f_{ij}, g_{ij} : \mathbb{R}^n \to \mathbb{R}$ are tractable continuous functions
- $\mathbb{I}_{\mathbb{R}_{++}}(\bullet)$ is the Heaviside function defined by

$$\mathbb{I}_{\mathbb{R}_{++}}(t) = \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{if } t \leq 0. \end{cases}$$

- $\mathbb{I}_{\mathbb{R}_{++}}(g_{ij}(x))$ captures discrete structures or logical conditions

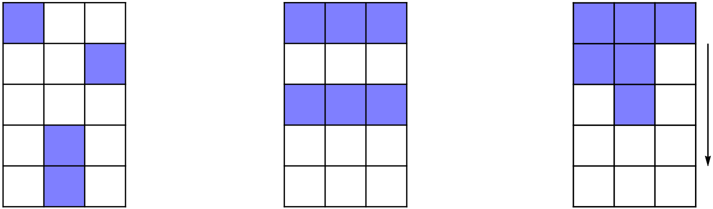
Some problem sources

$|\cdot|_0$-Optimization Given observations (a^i, y_i), consider regression problem

$$\begin{aligned} \min \ & \sum_{i=1}^{m} (\langle a^i, x\rangle - y_i)^2 \\ \text{s.t.} \ & \big(|x_1|_0, \ldots, |x_n|_0\big) \in X, \end{aligned}$$

where

- $|t|_0 = \mathbb{I}_{\mathbb{R}_{++}}(|t|) = \begin{cases} 0 & \text{if } t = 0 \\ 1 & \text{if } t \neq 0. \end{cases}$



(a) Sparsity $\sum_i |x_i|_0 \leq k$

(b) Group sparsity $|x_i|_0 = |x_j|_0$ if $i, j \in \mathcal{G}_k$

(c) Hierarchy structure $|x_i|_0 \leq |x_j|_0$

Some problem sources - stochastic optimization

Chance constraint

$$\mathbb{P}[g_1(x,\xi) > 0] \leq \mathbb{P}[g_2(x,\xi) > 0] = \mathbb{E}[\mathbb{I}_{\mathbb{R}_{++}}(g_2(x,\xi))]$$

Sample average approximation (SAA) $\Rightarrow$

$$\frac{1}{N}\sum_{s=1}^{N} \mathbb{I}_{\mathbb{R}_{++}}(g_1(x,\xi^s)) - \frac{1}{N}\sum_{s=1}^{N} \mathbb{I}_{\mathbb{R}_{++}}(g_2(x,\xi^s)) \leq 0$$

Conditional expectation

$$b \geq \mathbb{E}[f(x,\xi) | g(x,\xi) > 0] = \frac{\mathbb{E}[f(x,\xi)\mathbb{I}_{\mathbb{R}_{++}}(g(x,\xi))]}{\mathbb{P}[g(x,\xi)]}$$

Sample average approximation (SAA) $\Rightarrow$

$$\frac{1}{N}\sum_{s=1}^{N} f(x,\xi^s)\mathbb{I}_{\mathbb{R}_{++}}(g(x,\xi^s)) - \frac{b}{N}\sum_{s=1}^{N} \mathbb{I}_{\mathbb{R}_{++}}(g(x,\xi^s)) \leq 0$$

Outline

In the rest of this talk, we present some elementary analysis

- Closedness
- Optimality conditions
- Reformulation via lifting

Closedness

How hard is it to solve

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} \sum_{j=1}^{k} f_{0j}(x)\mathbb{I}_{\mathbb{R}_{++}}(g_{0j}(x)) \\ &\text{s.t. } x \in X_{\text{HSC}} \stackrel{\text{def}}{=} \left\{ x \in P \;\middle|\; \sum_{j=1}^{k} f_{ij}(x)\mathbb{I}_{\mathbb{R}_{++}}(g_{ij}(x)) \le b_i,\ i \in [m] \right\} \end{aligned} \qquad \text{(HSC)}$$

Closedness

How hard is it to solve

$$\begin{aligned}
&\min_{x \in \mathbb{R}^n} \sum_{j=1}^{k} f_{0j}(x)\mathbb{I}_{\mathbb{R}_{++}}(g_{0j}(x)) \\
&\text{s.t. } x \in X_{\mathrm{HSC}} \stackrel{\text{def}}{=} \left\{ x \in P \;\middle|\; \sum_{j=1}^{k} f_{ij}(x)\mathbb{I}_{\mathbb{R}_{++}}(g_{ij}(x)) \le b_i,\ i \in [m] \right\}
\end{aligned} \qquad \text{(HSC)}$$

To simplify, consider

$$X_{\mathrm{ASC}} = \left\{ x \in P \;\middle|\; \sum_{j=1}^{n} a_{ij}|x_j|_0 \le b_i,\ i = 1, \ldots, m \right\}.$$

- If $a_{ij} \ge 0\ \forall i, j$, then X_{ASC} is closed.

Closedness

How hard is it to solve

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} \sum_{j=1}^{k} f_{0j}(x) \mathbb{I}_{\mathbb{R}_{++}}(g_{0j}(x)) \\ &\text{s.t. } x \in X_{\text{HSC}} \overset{\text{def}}{=} \left\{ x \in P \;\middle|\; \sum_{j=1}^{k} f_{ij}(x) \mathbb{I}_{\mathbb{R}_{++}}(g_{ij}(x)) \le b_i, \; i \in [m] \right\} \end{aligned} \qquad \text{(HSC)}$$

To simplify, consider

$$X_{\text{ASC}} = \left\{ x \in P \;\middle|\; \sum_{j=1}^{n} a_{ij} |x_j|_0 \le b_i, \; i = 1, \ldots, m \right\}.$$

- If $a_{ij} \ge 0 \; \forall i, j$, then X_{ASC} is closed.
- X_{ASC} may **NOT** be a closed set in general! e.g., $\{x \mid |x_1|_0 \le |x_2|_0\}$

Closedness

If we take the closure of X_{ASC}...

- The resulting solution can be infeasible for the original problem
- The resulting set could be nonsense, e.g.,

$$\mathrm{cl}\{x \mid |x_1|_0 \leq |x_2|_0\} = \mathbb{R}^2$$

Closedness

If we take the closure of X_{ASC}...

- The resulting solution can be infeasible for the original problem
- The resulting set could be nonsense, e.g.,

$$\text{cl}\{x \mid |x_1|_0 \leq |x_2|_0\} = \mathbb{R}^2$$

- Computing $\text{cl}(X_{\text{ASC}})$ can be difficult

$$X = \bigcap_{i=1}^{m} X_i \not\Rightarrow \text{cl}(X) = \bigcap_{i=1}^{m} \text{cl}(X_i).$$

Example Consider $X_1 = \{x \mid |x_1|_0 \leq |x_2|_0\}$ and $X_2 = \{x \mid x_2 = 0\}$. Then

$$\text{cl}(X) = \text{cl}(X_1 \cap X_2) = (0,0) \neq \text{cl}(X_1) \cap \text{cl}(X_2) = \mathbb{R} \times \{0\}.$$

Closedness

For

$$X_{\mathrm{ASC}} = \{x \in P \mid A|x|_0 \leq b_i,\ i = 1, \ldots, m\},$$

where $|x|_0 \in \mathbb{R}^n$ is defined by $(|x|_0)_i = |x_i|_0$.

Proposition

There exists a matrix $\tilde{A} \geq 0$ and a $\{0,1\}$-vector $\tilde{b}$ such that

$$\mathrm{cl}(X_{\mathrm{ASC}}) = \left\{x \in P \,\middle|\, \tilde{A}|x|_0 \leq \tilde{b}\right\}$$

- A point $z \in S$ is called a maximal element if $z' \geq z \Rightarrow z' = z \,\forall z' \in S$.
- $(\tilde{A}, \tilde{b})$ merely depends on the maximal elements in the support set $\{z \in \{0,1\}^n |\ z = |x|_0,\ x \in X_{\mathrm{ASC}}\}$
- It is unclear how to compute $(\tilde{A}, \tilde{b})$.

Optimality condition for HSC

Given a "stationary" solution $\bar{x}$ to

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} \sum_{j=1}^{k} f_{0j}(x)\,\mathbb{I}_{\mathbb{R}_{++}}(g_{0j}(x)) \\ &\text{s.t. } x \in X_{\text{HSC}} \overset{\text{def}}{=} \left\{ x \in P \;\middle|\; \sum_{j=1}^{k} f_{ij}(x)\,\mathbb{I}_{\mathbb{R}_{++}}(g_{ij}(x)) \le b_i,\ i \in [m] \right\} \end{aligned} \qquad \text{(HSC)}$$

Question: under what conditions, $\bar{x}$ is a local minimizer of (HSC)?

A detour – generalization of convex functions

Consider a convex optimization problem

$$\min_x f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a closed convex function. A significant feature of convex programs is

$$0 \in \partial f(x) \Rightarrow x \in \arg\min_x f(x).$$

A detour – generalization of convex functions

Consider a convex optimization problem

$$\min_x f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a closed convex function. A significant feature of convex programs is

$$0 \in \partial f(x) \Rightarrow x \in \arg\min_x f(x).$$

Extension

- Quasi-convex function: every local minimizer is a global minimizer.
- Pseudo-convex function: every stationary solution is a global minimizer.

A detour – generalization of convex functions

Consider a convex optimization problem

$$\min_x f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a closed convex function. A significant feature of convex programs is

$$0 \in \partial f(x) \Rightarrow x \in \arg\min_x f(x).$$

Extension

- Quasi-convex function: every local minimizer is a global minimizer.
- Pseudo-convex function: every stationary solution is a global minimizer.

Aim to generalize convexity in a local manner...

Locally convex-like property

Define the directional derivative function of f at $\bar{x}$

$$f'(\bar{x}; x - \bar{x}) = \lim_{t \to 0} \frac{f(\bar{x} + t(x - \bar{x})) - f(\bar{x})}{t}.$$

A convex function f satisfies

- Global relaxation:

$$f(x) \geq \ell_{\bar{x}}(x) \overset{\text{def}}{=} f(\bar{x}) + f'(\bar{x}; x - \bar{x}) \; \forall x \in \mathbb{R}^n$$

- Touching property: $f(\bar{x}) = \ell_{\bar{x}}(\bar{x})$.

Then

$\bar{x}$ is a stationary solution $\Leftrightarrow \bar{x} \in \underset{x \in \mathbb{R}^n}{\arg\min}\, \ell_{\bar{x}}(x) \Rightarrow \bar{x} \in \underset{x \in \mathbb{R}^n}{\arg\min}\, f(x)$

From convexity to locally convex-like property

Define the directional derivative function of f at $\bar{x}$

$$f'(\bar{x}; x - \bar{x}) = \lim_{t \to 0} \frac{f(\bar{x} + t(x - \bar{x})) - f(\bar{x})}{t}.$$

If the function f satisfies

- Local relaxation:

$$f(x) \geq \ell_{\bar{x}}(x) \overset{\text{def}}{=} f(\bar{x}) + f'(\bar{x}; x - \bar{x}) \ \forall x \in \mathcal{B}_r(\bar{x}) \overset{\text{def}}{=} \{x \mid \|x - \bar{x}\| \leq r\}$$

- Touching property: $f(\bar{x}) = \ell_{\bar{x}}(\bar{x})$.

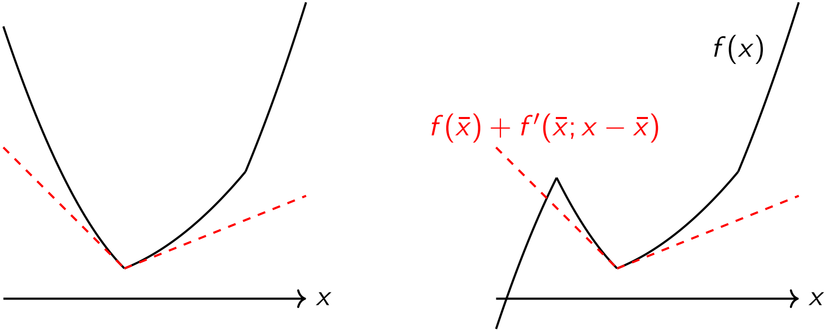
Then

$\bar{x}$ is a stationary solution $\Leftrightarrow \bar{x} \in \underset{x \in \mathcal{B}_r(\bar{x})}{\arg\min}\, \ell_{\bar{x}}(x) \Rightarrow \bar{x} \in \underset{x \in \mathcal{B}_r(\bar{x})}{\arg\min}\, f(x)$

Locally convex-like functions

Locally convex-like functions A function $f : \mathbb{R}^n \to \mathbb{R}$ is called (locally) convex like at $\bar{x}$ if there exists a neighborhood $\mathcal{B}_r(\bar{x})$ of $\bar{x}$ such that

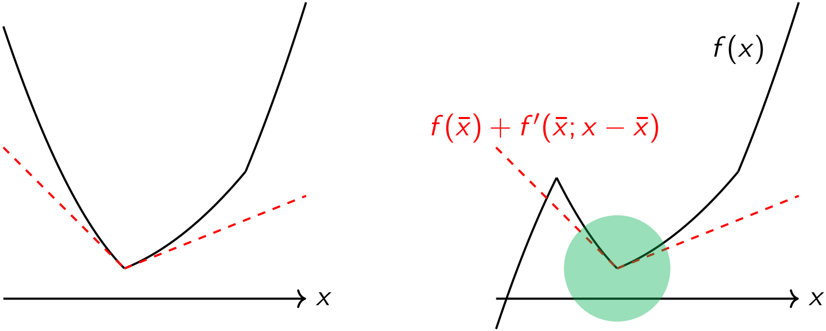
$$f(x) \geq f(\bar{x}) + f'(\bar{x}; x - \bar{x}),\ \forall x \in \mathcal{B}_r(\bar{x})$$



Locally convex-like functions

Locally convex-like functions A function $f : \mathbb{R}^n \to \mathbb{R}$ is called <u>(locally) convex like</u> at $\bar{x}$ if there exists a neighborhood $\mathcal{B}_r(\bar{x})$ of $\bar{x}$ such that

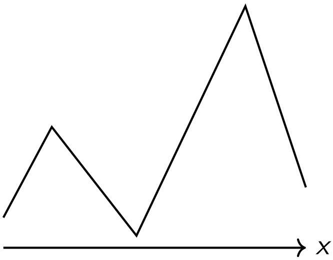
$$f(x) \geq f(\bar{x}) + f'(\bar{x}; x - \bar{x}),\ \forall x \in \mathcal{B}_r(\bar{x})$$



Locally convex-like functions

Examples

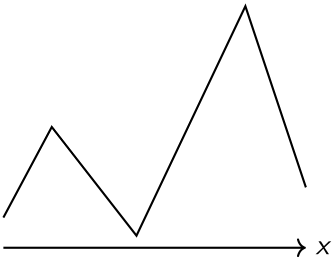
- Convex functions are locally convex-like
- Piecewise affine functions are locally convex-like

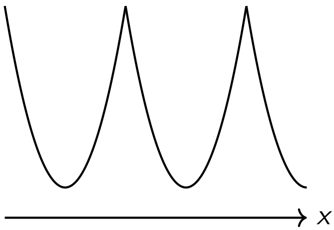


Locally convex-like functions

Examples

- Convex functions are locally convex-like
- Piecewise affine functions are locally convex-like
- The composition $f \circ g \circ h$ is locally convex-like if f is isotone and piecewise affine, g is convex, and h is piecewise affine





Locally convex-like functions

Locally convex-like functions A function $f : \mathbb{R}^n \to \mathbb{R}$ is called (locally) convex like at $\bar{x}$ if there exists a neighborhood $\mathcal{B}_r(\bar{x})$ of $\bar{x}$ such that

$$f(x) \geq f(\bar{x}) + f'(\bar{x}; x - \bar{x}), \ \forall x \in \mathcal{B}_r(\bar{x})$$

The gap between everywhere local convex-like property and the global convexity is Clarke regularity.

Proposition

Assume $f : \mathbb{R}^n \to \mathbb{R}$ is locally convex-like everywhere. Then

$$f \text{ is Clarke regular everywhere} \Leftrightarrow f \text{ is convex}$$

Locally convex-like sets

Define the tangent cone of a given set S at $\bar{x} \in S$ as

$$\mathcal{T}(\bar{x}; S) \stackrel{\text{def}}{=} \left\{ v \in \mathbb{R}^n \;\middle|\; \exists \{t_k\} \downarrow 0 \text{ and } \{x^k\} \subset S \text{ s.t. } v = \lim_{k \to \infty} \frac{x^k - \bar{x}}{t_k} \right\}$$

Locally convex-like sets A set S is called locally convex-like at $\bar{x} \in S$ if there exists a neighborhood $\mathcal{B}_r(\bar{x})$ such that the relaxation property holds

$$S \cap \mathcal{B}_r(\bar{x}) \subseteq \bar{x} + \mathcal{T}(\bar{x}; S)$$

- f is locally convex like at $\bar{x} \Leftrightarrow \text{epi}(f)$ is locally convex like at $(\bar{x}, f(\bar{x}))$
- Convex sets and open sets are always locally convex like
- Cartesian product/union of finitely many locally convex sets is locally convex

Locally convex-like sets

Unless suitable constraint qualification holds, the intersection of locally convex sets is generally not locally convex. In particular, the level set of a locally convex functions is not necessarily locally convex.

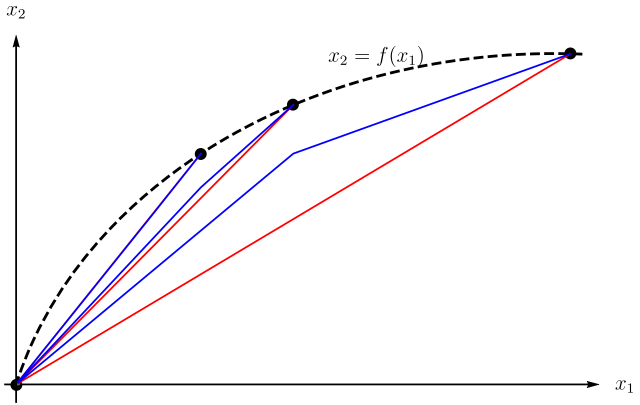


Figure: Intersection of two locally convex sets. X_1 and X_2 consists of the red and blue line segments, respectively; their intersection is represented by black points.

Epistationarity

Optimality condition for linear optimization over S

$$\bar{x} \in \arg\min_{x \in S} c^\top x \Rightarrow c^\top v \geq 0,\ \forall v \in \mathcal{T}(\bar{x}; S)$$

Note $\min\{f(x) \,|\, x \in S\} \Leftrightarrow \min\{t \,|\, (x, t) \in \text{epi}(f), x \in S\}$.

Epistationary solution Point $\bar{x}$ is called an epistationary solution of $\min\{f(x) : x \in S\}$ if $(\bar{x}, f(\bar{x}))$ satisfies the optimality condition for the lifted linear program over $\hat{S} \stackrel{\text{def}}{=} \{(x, t) \in \text{epi}(f) \,|\, x \in S\}$.

Proposition

If $\bar{x}$ is an epistationary solution and $\hat{S}$ is locally convex-like at $(\bar{x}, f(\bar{x}))$, then $\bar{x}$ is a local minimizer of $\min\{f(x) \,|\, x \in S\}$.

Sufficient optimality condition for HSC

Back to the Heaviside composite optimization problem,

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \sum_{j=1}^{k} f_{0j}(x)\, \mathbb{I}_{\mathbb{R}_{++}}(g_{0j}(x)) \\ \text{s.t.} \ & x \in X_{\text{HSC}} \overset{\text{def}}{=} \left\{ x \in P \;\middle|\; \sum_{j=1}^{k} f_{ij}(x)\, \mathbb{I}_{\mathbb{R}_{++}}(g_{ij}(x)) \leq b_i,\ i \in [m] \right\} \end{aligned} \qquad \text{(HSC)}$$

To establish sufficient optimality conditions, one turns to study the local convex-like property of the lifted set

$$\hat{X}_{\text{HSC}} \overset{\text{def}}{=} \left\{ (x, t) \;\middle|\; \sum_{j=1}^{k} f_{0j}(x)\, \mathbb{I}_{\mathbb{R}_{++}}(g_{0j}(x)) \leq t,\ x \in X_{\text{HSC}} \right\}$$

which itself is HSC.

Locally convex-like ASC constraints

Recall

$$X_{\mathrm{ASC}} = \left\{ x \in P \;\middle|\; \sum_{j=1}^{n} a_{ij} |x_j|_0 \leq b_i,\; i = 1, \ldots, m \right\}.$$

Tangent cone of X_{ASC}

$$\mathcal{T}(\bar{x}; X_{\mathrm{ASC}}) = \mathrm{cl} \left\{ v \in \mathcal{T}(\bar{x}; P) \;\middle|\; \sum_{i \notin \bar{\beta}} a_{ij} |v_j|_0 \leq b_i - \sum_{j \in \beta} a_{ij} \right\},$$

where $\beta = \mathrm{supp}(\bar{x}) \overset{\mathrm{def}}{=} \{ i \mid \bar{x}_i \neq 0 \}$.

Proposition

The set X_{ASC} is locally convex like at every $\bar{x} \in X_{\mathrm{ASC}}$.

Locally convex-like HSC constraints

In general, the $\mathcal{T}(x; X_{\text{HSC}})$ does not admit a clean description

$$X_{\text{HSC}} \stackrel{\text{def}}{=} \left\{ x \in P \;\middle|\; \sum_{j=1}^{k} f_{ij}(x)\, \mathbb{I}_{\mathbb{R}_{++}}(g_{ij}(x)) \le b_i,\ i \in [m] \right\}$$

However, if each f_{ij}, g_{ij} is either convex or piecewise affine, we have

Proposition

The set X_{HSC} is locally convex-like everywhere if the assumptions given by any entry of the following table is true

f_{ij} \ g_{ij}	convex	piecewise affine
convex	$f_{ij} \ge 0$	free
piecewise affine	$f_{ij} \ge 0$	free

Computation via lifting

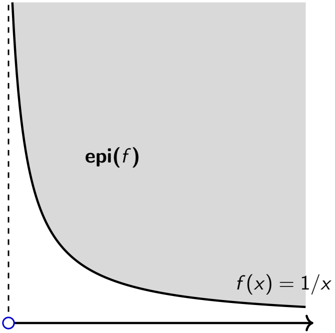
Consider

$$X_{\text{ASC}} = \left\{ x \in P \;\middle|\; \sum_{j=1}^{n} a_{ij} |x_j|_0 \leq b_i,\; i = 1, \ldots, m \right\}.$$

Lifting A non-closed set can be closed in a lifted space by introducing additional variables.

Example

$$\operatorname*{proj}_{x} \{(t, x) \in \mathbb{R}^2_+ : tx \geq 1\} = (0, \infty)$$



Computation via lifting

Lifting A non-closed set can be closed in an extended space of variables. Consider

$$X_{\text{ASC}} = \left\{ x \in P \;\middle|\; \sum_{j=1}^{n} a_{ij} |x_j|_0 \le b_i,\; i = 1, \dots, m \right\}.$$

Define

$$\tilde{X}_{\text{ASC}} = \left\{ (x, t, s, y) \;\middle|\; \begin{array}{l} \sum_{j=1}^{n} (a_{ij}^{+} + \epsilon) s_j \le \sum_{j=1}^{n} (a_{ij}^{-}) t_j + b_i,\; i = 1, \dots, m \\ t_j \le |x_j| y_j,\; 0 \le t_j \le 1,\; y_j \ge 0,\; j = 1, \dots, n \\ x_j (1 - s_j) = 0,\; 0 \le s_j \le 1,\; j = 1, \dots, n \end{array} \right\}$$

Then

$$\text{proj}_x(\tilde{X}_{\text{ASC}}) = X_{\text{ASC}}$$

and

$$\min\{ f(x) \,|\, x \in X_{\text{ASC}} \} \Leftrightarrow \min\{ f(x) \,|\, (x, t, s, y) \in \tilde{X}_{\text{ASC}} \}.$$

Computation via lifting

Comments

- Similar reformulation techniques are applicable to get $\tilde{X}_{\mathrm{HSC}}$ for X_{HSC}.
- The lifting process is not unique and can be performed in multiple ways.
- Lifting offers a valid solution method that leverages on existing NLP algorithms.

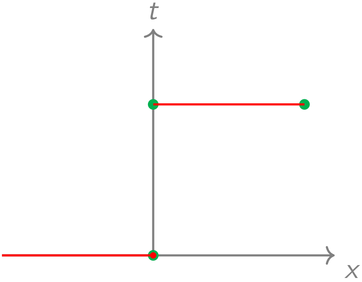
Computation via lifting

Comments

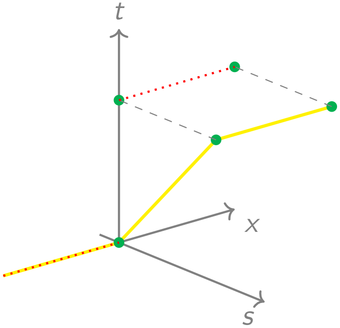
- Similar reformulation techniques are applicable to get $\tilde{X}_{\mathrm{HSC}}$ for X_{HSC}.
- The lifting process is not unique and can be performed in multiple ways.
- Lifting offers a valid solution method that leverages on existing NLP algorithms.
- In the worst case, the stationary solution obtained in the lifted space can be weak in the original space of variables.

Computation via lifting

Graph of $t = \mathbb{I}_{\mathbb{R}_{++}}(x)$

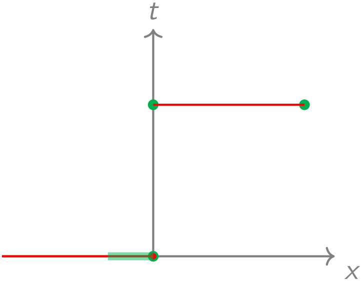


Graph of $t = \mathbb{I}_{\mathbb{R}_{++}}(x)$ in lifted space

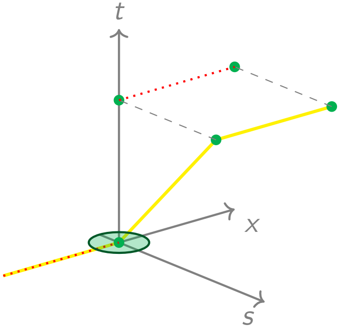


Computation via lifting

Graph of $t = \mathbb{I}_{\mathbb{R}_{++}}(x)$

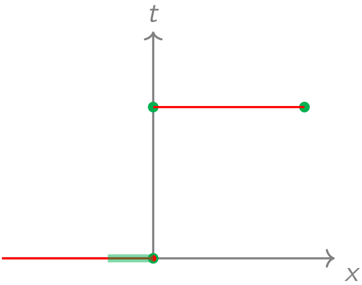


Graph of $t = \mathbb{I}_{\mathbb{R}_{++}}(x)$ in lifted space

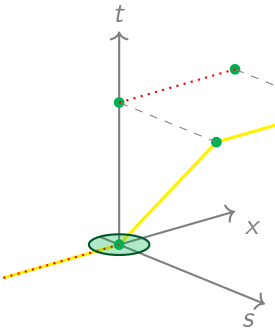


Computation via lifting

Graph of $t = \mathbb{I}_{\mathbb{R}_{++}}(x)$



Graph of $t = \mathbb{I}_{\mathbb{R}_{++}}(x)$ in lifted space



Proposition (Informal)

If $(\bar{x}, \bar{t}, \bar{y}, \bar{s})$ is an epistationary solution of f on $\tilde{X}_{\mathrm{HSC}}$, then $\bar{x}$ is a <u>pseudostationary</u> solution of f on X_{HSC}. Under more restrictive conditions, $\bar{x}$ is an epistationary solution of f on X_{HSC}.

Take Home Message

Summary

- Heaviside composite optimization can lack lower semicontinuity and is challenging to solve
- Locally convex-like property + epi-stationarity $\Rightarrow$ local optimality
- Possibility of solving Heaviside via existing NLP solution methods in a lifted space

Take Home Message

Summary

- Heaviside composite optimization can lack lower semicontinuity and is challenging to solve
- Locally convex-like property + epi-stationarity $\Rightarrow$ local optimality
- Possibility of solving Heaviside via existing NLP solution methods in a lifted space

Thanks for your listening!