Minimizing convex functions over integer points

Dr. Michel Baes,

IFOR / ETH Zürich

joint work with A. Del Pia, Y. Nesterov,

S. Onn, C. Wagner, R. Weismantel

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Playing with switches to approximate a function

Let $f : [0, T] \to \mathbb{R}$ be a given function *(signal)*
$g : \mathbb{R} \to \mathbb{R}$ be another given function *(filter)*
with $g(t) = 0$ for $t \leq 0$
$\delta := T/n$, where $n \in \mathbb{Z}_{++}$ *(time-step)*

We want to approximate f as well as possible by

$$\hat{f}_x(t) := \sum_{i=1}^{n} x_i g(t - (i-1)\delta), \text{ with } x_i \in \{0, 1\}.$$

Playing with switches to approximate a function

Let $f : [0, T] \to \mathbb{R}$ be a given function *(signal)*
$g : \mathbb{R} \to \mathbb{R}$ be another given function *(filter)*
with $g(t) = 0$ for $t \leq 0$
$\delta := T/n$, where $n \in \mathbb{Z}_{++}$ *(time-step)*

We want to approximate f as well as possible by

$$\widehat{f}_x(t) := \sum_{i=1}^{n} x_i g(t - (i-1)\delta), \text{ with } x_i \in \{0, 1\}.$$

- $||f - \widehat{f}_x||_2^2 = \int_0^T (f(t) - \widehat{f}_x(t))^2 dt \to \min$

Playing with switches to approximate a function

Let $f : [0, T] \to \mathbb{R}$ be a given function *(signal)*
$g : \mathbb{R} \to \mathbb{R}$ be another given function *(filter)*
with $g(t) = 0$ for $t \leq 0$
$\delta := T/n$, where $n \in \mathbb{Z}_{++}$ *(time-step)*

We want to approximate f as well as possible by

$$\hat{f}_x(t) := \sum_{i=1}^{n} x_i g(t - (i-1)\delta), \text{ with } x_i \in \{0, 1\}.$$

- $\|f - \hat{f}_x\|_2^2 = \int_0^T (f(t) - \hat{f}_x(t))^2 dt \to \min$

  $\Leftrightarrow \min\{x^T Q x + 2c^T x + d : x \in \{0, 1\}^n\}$

$Q_{ij} = \int_0^T g(t - (i-1)\delta) g(t - (j-1)\delta) dt$, $c_i = \int_0^T g(t - (i-1)\delta) f(t) dt$

Playing with switches to approximate a function

Let $f : [0, T] \to \mathbb{R}$ be a given function *(signal)*
$g : \mathbb{R} \to \mathbb{R}$ be another given function *(filter)*
with $g(t) = 0$ for $t \leq 0$
$\delta := T/n$, where $n \in \mathbb{Z}_{++}$ *(time-step)*

We want to approximate f as well as possible by

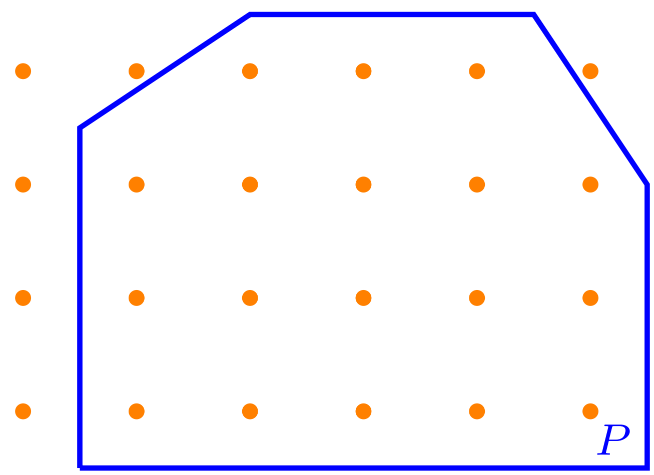
$$\hat{f}_x(t) := \sum_{i=1}^{n} x_i g(t - (i-1)\delta), \text{ with } x_i \in \{0, 1\}.$$

- $||f - \hat{f}_x||_2^2 = \int_0^T (f(t) - \hat{f}_x(t))^2 dt \to \min$
  $\Leftrightarrow \min\{x^T Q x + 2c^T x + d : x \in \{0,1\}^n\}$
- We have a convex objective with binary variables.

Our problem is very hard

$$f(x^*) := \min_{x\in\mathcal{F}} f(x)$$

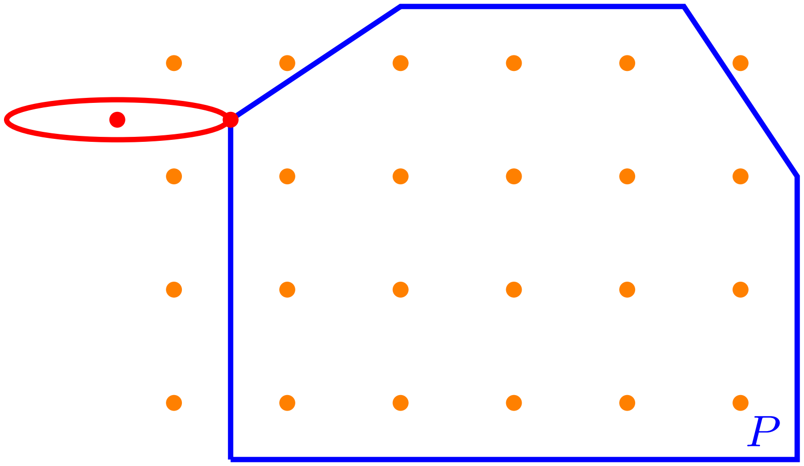
- $\star$ $\mathcal{F} = P \cap \mathbb{Z}^n$, where P is a polytope;
- $\star$ $f : \mathbb{R}^n \to \mathbb{R}$ is convex.



Our problem is very hard

$$f(x^*) := \min_{x \in \mathcal{F}} f(x)$$

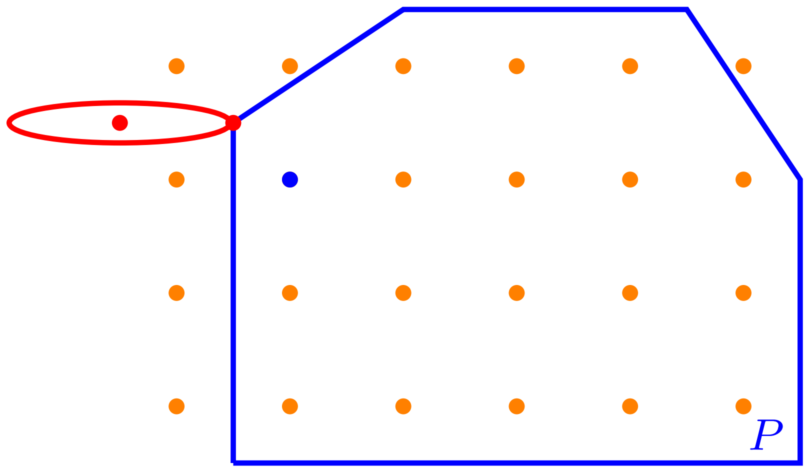
- $\star$ $\mathcal{F} = P \cap \mathbb{Z}^n$, where P is a polytope;
- $\star$ $f : \mathbb{R}^n \to \mathbb{R}$ is convex.



Our problem is very hard

$$f(x^*) := \text{min}_{x \in \mathcal{F}} f(x)$$

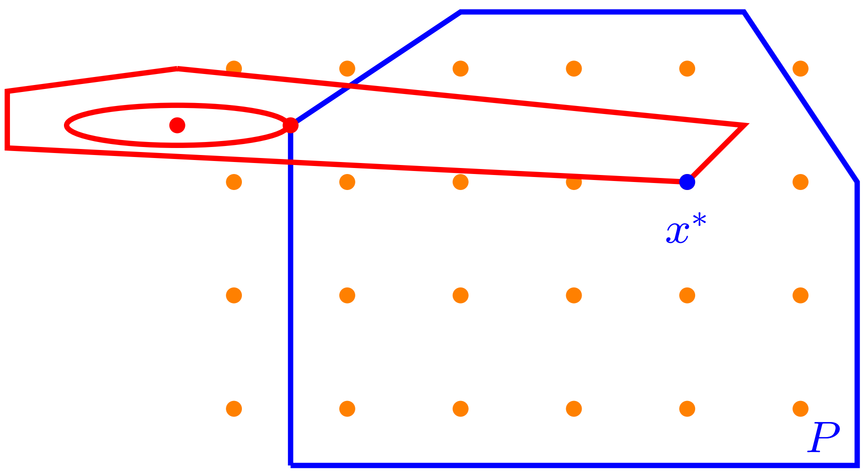
- $\star$ $\mathcal{F} = P \cap \mathbb{Z}^n$,
  where P is a polytope;
- $\star$ $f : \mathbb{R}^n \to \mathbb{R}$ is convex.



Our problem is very hard

$$f(x^*) := \min_{x \in \mathcal{F}} f(x)$$

- ⋆ $\mathcal{F} = P \cap \mathbb{Z}^n$, where P is a polytope;
- ⋆ $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex.



x^* can be arbitrarily far from the continuous minimum

A way to measure how hard our problem is

$$f(x^*) := \min_{x \in \mathcal{F}} f(x)$$

- ⋆ $\mathcal{F} = P \cap \mathbb{Z}^n$,
  where P is a polytope;
- ⋆ $f : \mathbb{R}^n \to \mathbb{R}$ is convex.

We assume that f is strongly convex and L-smooth:
for every $x, y \in \mathbb{R}^n$:

$$\frac{l}{2}||x-y||_2^2 \leq f(y) - f(x) - f'(x)^T(y-x) \leq \frac{L}{2}||x-y||_2^2.$$

A way to measure how hard our problem is

$$f(x^*) := \min_{x \in \mathcal{F}} f(x)$$

⋆ $\mathcal{F} = P \cap \mathbb{Z}^n$,
where P is a polytope;

⋆ $f : \mathbb{R}^n \to \mathbb{R}$ is convex.

We assume that f is strongly convex and L-smooth:
for every $x, y \in \mathbb{R}^n$:

$$\frac{l}{2}||x - y||_2^2 \leq f(y) - f(x) - f'(x)^T(y - x) \leq \frac{L}{2}||x - y||_2^2.$$

- L/l bounds the asphericity of level sets.
  A trivial enumeration algorithm would take
  $\mathcal{O}(\min\{(L/l)^n, \text{diameter}(\mathcal{F})^n\})$ evaluations of f.
- In fact, the complexity of our methods depends on $L - l$.

Our problem is very hard

$$f(x^*) := \min_{x \in \mathcal{F}} f(x)$$

- ⋆ $\mathcal{F} = P \cap \mathbb{Z}^n$, where P is a polytope;
- ⋆ $f : Q \to \mathbb{R}$ is convex.

- ▶ Even with f linear, the problem is **NP-Hard**
- ▶ Even with f quadratic and $\mathcal{F} \subseteq \{0, 1\}^n$, the problem is **hard**, independently of the NP-Conjecture.

Our problem is very hard

$$f(x^*) := \min_{x\in\mathcal{F}} f(x)$$

- $\star$ $\mathcal{F} = P \cap \mathbb{Z}^n$,
  where P is a polytope;
- $\star$ $f : Q \to \mathbb{R}$ is convex.

- Even with f linear, the problem is **NP-Hard**
- Even with f quadratic and $\mathcal{F} \subseteq \{0,1\}^n$,
  the problem is **hard**, independently of the NP-Conjecture.
  Consider $n = 4m$, $c \in \{2,3\}$, $\gamma = 5m - 1$,
  $\mathcal{F} \subseteq \{0,1\}^n$ be an independence system, and

$$\min\{f(x) = n^2(c^T x - \gamma)^2 + \mathbf{1}^T x : x \in \mathcal{F}\}.$$

  $\mathcal{F}$ is presented by an LP solver:
  taking as input $v \in \mathbb{R}^n$, returns $\arg\max_{x\in\mathcal{F}} v^T x$.

Note: We have $L/l \leq 9n^3 + 1$.

Our problem is very hard

$$f(x^*) := \min_{x \in \mathcal{F}} f(x)$$

⋆ $\mathcal{F} = P \cap \mathbb{Z}^n$, where P is a polytope;

⋆ $f : Q \to \mathbb{R}$ is convex.

- Even with f linear, the problem is **NP-Hard**
- Even with f quadratic and $\mathcal{F} \subseteq \{0,1\}^n$, the problem is **hard**, independently of the NP-Conjecture. Consider $n = 4m$, $c \in \{2,3\}$, $\gamma = 5m - 1$, $\mathcal{F} \subseteq \{0,1\}^n$ be an independence system, and

$$\min\{f(x) = n^2(c^T x - \gamma)^2 + 1^T x : x \in \mathcal{F}\}.$$

Theorem 1 *Any method that solves less than* $\binom{2m}{m+1} \geq 2^m$ *LPs on* $\mathcal{F}$ *fails to find an* $\hat{x}$ *for which* $f(\hat{x}) - f(x^*) \leq n^2 - n$.

Very few methods exist for our problem

$$f(x^*) := \min_{x \in \mathcal{F}} f(x)$$

- $\star$ $\mathcal{F} = P \cap \mathbb{Z}^n$,
  where P is a polytope;
- $\star$ $f : \mathbb{R}^n \to \mathbb{R}$ is convex.

1. **Branch and bound approach** *(Gupta, Ravindran, Leyffer)*
   Branching by fixing some variables.
   Lower bounds from continuous relaxation
   of all the other ones, with possibly extra cuts

2. **Outer Approximation** *(Duran, Grossman)*
   (assumes that we can solve the problem with f linear)
   f is replaced by a piecewise linear model,
   which is minimized in $\mathcal{F}$, then enriched.

Our methods use simple ideas from Convex Optimization

$$f(x^*) := \min_{x \in Q} f(x)$$

- $\star$ $Q \subseteq \mathbb{R}^n$ is convex and closed;
- $\star$ $f : \mathbb{R}^n \to \mathbb{R}$ is convex.

Subgradient method:

Select $x_0 \in Q$

for $k = 0, 1, 2, \ldots$

Choose $h_k > 0$, compute $f'(x_k) \in \partial f(x_k)$

Set $x_{k+1} := \pi_Q(x_k - h_k f'(x_k))$

$$= \arg\min_{y \in Q} \left\{ f'(x_k)(y - x_k) + \frac{1}{2h_k} ||y - x_k||_2^2 \right\}.$$

- Guaranteed decrease with $h_k = 1/L$, theoretically the best choice.

A substitute for the subgradient step

$$f(x^*) := \min_{x \in \mathcal{F}} f(x)$$

- $\star$ $\mathcal{F} = P \cap \mathbb{Z}^n$, where P is a polytope;
- $\star$ $f : \mathbb{R}^n \to \mathbb{R}$ is convex.

Gradient method:

$$x_{k+1} := \arg\min_{y \in Q} \left\{ f'(x_k)(y - x_k) + \frac{L}{2} ||y - x_k||_2^2 \right\}.$$

Our strategy:

$$x_{k+1} := \arg\min_{y \in \mathcal{F}} \left\{ f'(x_k)(y - x_k) + \frac{L}{2} ||y - x_k||_2^2 \right\}.$$

A substitute for the subgradient step

$$f(x^*) := \min_{x \in \mathcal{F}} f(x)$$

- ⋆ $\mathcal{F} = P \cap \mathbb{Z}^n$, where P is a polytope;
- ⋆ $f : \mathbb{R}^n \to \mathbb{R}$ is convex.

Gradient method:

$$x_{k+1} := \arg\min_{y \in Q} \left\{ f'(x_k)(y - x_k) + \frac{L}{2}||y - x_k||_2^2 \right\}.$$

Our strategy:

$$x_{k+1} := \arg\min_{y \in \mathcal{F}} \left\{ f'(x_k)(y - x_k) + \frac{1}{2\tau}||y - x_k||_2^2 \right\}.$$

▶ We allow for a step-size τ larger than $1/L$.

A substitute for the subgradient step

$$f(x^*) := \min_{x \in \mathcal{F}} f(x)$$

- $\star$ $\mathcal{F} = P \cap \mathbb{Z}^n$, where P is a polytope;
- $\star$ $f : \mathbb{R}^n \to \mathbb{R}$ is convex.

Gradient method:

$$x_{k+1} := \arg\min_{y \in Q} \left\{ f'(x_k)(y - x_k) + \frac{L}{2}||y - x_k||_2^2 \right\}.$$

Our strategy:

$$x_{k+1} \approx \arg\min_{y \in \mathcal{F}} \left\{ f'(x_k)(y - x_k) + \frac{1}{2\tau}||y - x_k||_2^2 \right\}.$$

- We allow for a step-size τ larger than $1/L$.
- This quadratic problem can be very hard to solve exactly
- $x_{k+1} \in \mathcal{F}$ and $g_{x_k}(x_{k+1}) \leq (1 - \alpha) \min_{y \in \mathcal{F}} g_{x_k}(y) < 0$,

  with $g_{x_k}(y) := f'(x_k)(y - x_k) + \frac{1}{2\tau}||y - x_k||_2^2$.

A substitute for the subgradient step

$$f(x^*) := \min_{x \in \mathcal{F}} f(x)$$

⋆ $\mathcal{F} = P \cap \mathbb{Z}^n$,
where P is a polytope;
⋆ $f : \mathbb{R}^n \to \mathbb{R}$ is convex.

Gradient method:

$$x_{k+1} := \arg\min_{y \in Q} \left\{ f'(x_k)(y - x_k) + \frac{L}{2} ||y - x_k||_2^2 \right\}.$$

Our strategy:

$$x_{k+1} \approx \arg\min_{y \in \mathcal{F}} \left\{ f'(x_k)(y - x_k) + \frac{1}{2\tau} ||y - x_k||_2^2 \right\}.$$

- We allow for a step-size τ larger than $1/L$.
- This quadratic problem can be very hard to solve exactly
- $x_{k+1} \in \mathcal{F}$ and $g_{x_k}(x_{k+1}) \leq (1 - \alpha) \min_{y \in \mathcal{F}} g_{x_k}(y) < 0$,

  with $g_{x_k}(y) := f'(x_k)(y - x_k) + \frac{1}{2\tau} ||y - x_k||_2^2$.

If we don't move anymore, we are done

Select $x_0 \in \mathcal{F}$, $0 \leq \alpha < 1$, $l \leq \tau \leq L$, $N \in \mathbb{Z}_+$
for $k = 0, 1, \ldots, N-1$,
  Compute x_{k+1} such that
    $g_{x_k}(x_{k+1}) \leq (1-\alpha) \arg\min_{y \in \mathcal{F}} g_{x_k}(y)$
  if $x_{k+1} \in \{x_0, \ldots, x_k\}$
    STOP
return the best $\widehat{x} \in \{x_0, \ldots, x_N\}$.

Theorem 2 *If $\tau = 1/l$ and $x_{k+1} = x_k$, then $f(x_k) = f(x^*)$.*

Our method can solve the problem approximately

Select $x_0 \in \mathcal{F}$, $0 \leq \alpha < 1$, $1/L \leq \tau \leq 1/l$, $N \in \mathbb{Z}_+$
for $k = 0, 1, \ldots, N-1$,
  Compute x_{k+1} such that
    $g_{x_k}(x_{k+1}) \leq (1-\alpha) \arg\min_{y \in \mathcal{F}} g_{x_k}(y)$
  if $x_{k+1} \in \{x_0, \ldots, x_k\}$
    STOP
return the best $\widehat{x} \in \{x_0, \ldots, x_N\}$.

Theorem 3 *If* $\alpha > 0$, $\eta > 0$, $\delta_{\mathcal{F}} =$ *diameter*$(\mathcal{F})$ *and*

$$N := \left\lceil \frac{1}{\ln(1/\alpha)} \ln\left(\max\left\{1, \frac{f(x_0) - f(x^*)}{\eta}\right\}\right)\right\rceil,$$

then $f(\widehat{x}) - f(x^*) \leq \dfrac{L-l}{2(1-\alpha)}\delta_{\mathcal{F}}^2 + \eta.$

The subproblem can sometimes be solved

We know: quadratic problems on $\mathcal{F}$ can be **hard**

Theorem 4 (Heinz) *We can solve the subproblem exactly ($\alpha = 0$) in $s^{\mathcal{O}(1)} 2^{\mathcal{O}(n)}$ arithmetic operations, where s is its binary encoding length.*

The subproblem can sometimes be solved

We know: quadratic problems on $\mathcal{F}$ can be **hard**

Theorem 4 (Heinz) *We can solve the subproblem exactly ($\alpha = 0$) in $s^{\mathcal{O}(1)} 2^{\mathcal{O}(n)}$ arithmetic operations, where s is its binary encoding length.*

- If $\mathcal{F} \subseteq \{0, 1\}^n$, the subproblem is linear (as $x_i^2 = x_i$). We can solve it efficiently for classes of problems where IP is easy (e.g. matroid problems, matchings)
- Many quadratic problems admit a fast approximate method with guaranteed α. (e.g. binary knapsack problems)

A delicate issue: avoiding cycling

Guarantee: $f(\widehat{x}) - f(x^*) \le \frac{L-l}{1-\alpha}\delta_{\mathcal{F}}^2$
after a known number of steps.

Let's keep it running a bit more, hoping for the best.

A delicate issue: avoiding cycling

Guarantee: $f(\widehat{x}) - f(x^*) \leq \frac{L-l}{1-\alpha}\delta_{\mathcal{F}}^2$
after a known number of steps.

Let's keep it running a bit more, hoping for the best.

However, we can get a previously visited point *(cycling)*

A delicate issue: avoiding cycling

Guarantee: $f(\widehat{x}) - f(x^*) \leq \frac{L-l}{1-\alpha}\delta_{\mathcal{F}}^2$
after a known number of steps.

Let's keep it running a bit more, hoping for the best.

However, we can get a previously visited point *(cycling)*
Remedy: Assume $f(x) \in \mathbb{Z}$ for $x \in \mathcal{F}$ and set $\tau = 1/l$.
Instead of computing $\min\{g_{x_k}(x) : x \in \mathcal{F}\}$, do:

$$\min\{g_{x_k}(x) : g_{x_i}(x) \leq -1 \text{ for } 0 \leq i \leq k,\ x \in \mathcal{F}\}$$

A delicate issue: avoiding cycling

Guarantee: $f(\widehat{x}) - f(x^*) \leq \frac{L-l}{1-\alpha}\delta_{\mathcal{F}}^2$
after a known number of steps.

Let's keep it running a bit more, hoping for the best.

However, we can get a previously visited point *(cycling)*
Remedy: Assume $f(x) \in \mathbb{Z}$ for $x \in \mathcal{F}$ and set $\tau = 1/l$.
Instead of computing $\min\{g_{x_k}(x) : x \in \mathcal{F}\}$, do:

$$\min\{g_{x_k}(x) : g_{x_i}(x) \leq -1 \text{ for } 0 \leq i \leq k,\ x \in \mathcal{F}\}$$

- Never generates twice the same point

A delicate issue: avoiding cycling

Guarantee: $f(\widehat{x}) - f(x^*) \leq \frac{L-l}{1-\alpha}\delta_{\mathcal{F}}^2$
after a known number of steps.

Let's keep it running a bit more, hoping for the best.

However, we can get a previously visited point *(cycling)*
Remedy: Assume $f(x) \in \mathbb{Z}$ for $x \in \mathcal{F}$ and set $\tau = 1/l$.
Instead of computing $\min\{g_{x_k}(x) : x \in \mathcal{F}\}$, do:

$$\min\{g_{x_k}(x) : g_{x_i}(x) \leq -1 \text{ for } 0 \leq i \leq k,\ x \in \mathcal{F}\}$$

- Never generates twice the same point
- Computes x^* in $\mathcal{O}(c_f(L-l)\delta_{\mathcal{F}}^2 s^{\mathcal{O}(1)})$,

  where $c_f = \max_{\alpha \in \mathbb{Z}} |\mathcal{F} \cap \{x : f(x) = \alpha\}|$.

A delicate issue: avoiding cycling

Guarantee: $f(\widehat{x}) - f(x^*) \leq \frac{L-l}{1-\alpha}\delta_{\mathcal{F}}^2$
after a known number of steps.

Let's keep it running a bit more, hoping for the best.

However, we can get a previously visited point *(cycling)*
Remedy: Assume $f(x) \in \mathbb{Z}$ for $x \in \mathcal{F}$ and set $\tau = 1/l$.
Instead of computing $\min\{g_{x_k}(x) : x \in \mathcal{F}\}$, do:

$$\min\{g_{x_k}(x) : g_{x_i}(x) \leq -1 \text{ for } 0 \leq i \leq k,\ x \in \mathcal{F}\}$$

- Never generates twice the same point
- Computes x^* in $\mathcal{O}(c_f(L-l)\delta_{\mathcal{F}}^{2} s^{\mathcal{O}(1)})$,

  where $c_f = \max_{\alpha \in \mathbb{Z}} |\mathcal{F} \cap \{x : f(x) = \alpha\}|$.
- Modest, but can be better than an enumeration ($\mathcal{O}(\delta_{\mathcal{F}}^{n})$).

A delicate issue: avoiding cycling

Guarantee: $f(\widehat{x}) - f(x^*) \leq \frac{L-l}{1-\alpha}\delta^2_{\mathcal{F}}$
after a known number of steps.

Let's keep it running a bit more, hoping for the best.

However, we can get a previously visited point *(cycling)*
Remedy: Assume $f(x) \in \mathbb{Z}$ for $x \in \mathcal{F}$ and set $\tau = 1/l$.
Instead of computing $\min\{g_{x_k}(x) : x \in \mathcal{F}\}$, do:

$$\min\{g_{x_k}(x) : g_{x_i}(x) \leq -1 \text{ for } 0 \leq i \leq k,\ x \in \mathcal{F}\}$$

- Never generates twice the same point
- Subproblems polynomially solvable
  for some well-structured binary problems
  (e.g. when $\mathcal{F} \subseteq \{0,1\}^n$ is a vectorial matroid)

Conclusion and outlook 1

- The field is almost new.
- Convex integer problems are intrinsically hard.
- Our methods provide polynomial algorithms for instances where quadratic problems can be easily minimized on their feasible set.
- We have complexity and accuracy guarantees when quadratic functions can only be solved approximately.
- The mixed-integer case remains largely unaddressed

Conclusion and outlook 2

- We are developing new methods based on the existence of a *level set oracle*:

  Fix $\alpha \geq 0$, $\delta \geq 0$. For every $x \in \mathbb{R}^n$,

  the oracle finds $\widehat{x} \in \mathbb{Z}^n$

  such that $f(\widehat{x}) \leq (1+\alpha)f(x) + \delta$,

  or declares it does not exists.
- They can be extended to general lattices.
- They allow us to solve 2 dim. problems polynomially in $\ln(L/l)$.
- They seem promising to attack mixed-integer convex problems.

Thank you