ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/86765/

This is the author’s version of a work that was submitted to / accepted for publication.

Citation for final published version:

Baes, Michel, Oertel, Timm , Wagner, Christian and Weismantel, Robert 2013. Mirror-descent methods in mixed-integer convex optimization. Junger, Michael and Reinelt, Gerhard, eds. Facets of Combinatorial Optimization: Festschrift for Martin Grötschel, Springer, pp. 101-131. (10.1007/978-3-642-38189-8_5)

Publishers page: http://dx.doi.org/10.1007/978-3-642-38189-8_5

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Mirror-Descent Methods in Mixed-Integer Convex Optimization

Michel Baes, Timm Oertel, Christian Wagner, and Robert Weismantel

Abstract In this paper, we address the problem of minimizing a convex function f over a convex set, with the extra constraint that some variables must be integer. This problem, even when f is a piecewise linear function, is NP-hard. We study an algorithmic approach to this problem, postponing its hardness to the realization of an oracle. If this oracle can be realized in polynomial time, then the problem can be solved in polynomial time as well. For problems with two integer variables, we show with a novel geometric construction how to implemented the oracle efficiently, that is, in $\mathscr{O}(\ln(B))$ approximate minimizations of f over the continuous variables, where B is a known bound on the absolute value of the integer variables. Our algorithm can be adapted to find the second best point of a purely integer convex optimization problem in two dimensions, and more generally its k-th best point. This observation allows us to formulate a finite-time algorithm for mixed-integer convex optimization.

1 Introduction

One of the highlights in the list of publications of Martin Grötschel is his joint book with László Lovász and Alexander Schrijver on Geometric Algorithms and Combinatorial Optimization [GLS88]. This book develops a beautiful and general theory of optimization over (integer) points in convex sets. The generality comes from the fact that the convex sets under consideration are presented by oracles (membership, separation in different variations, optimization). The algorithms and their efficiency typically depend on the oracle presentation of the underlying convex set. This is precisely the theme of this paper as well: we present an algorithmic framework for solving mixed-integer convex optimization problems that is based on an oracle.

Michel Baes, Timm Oertel, Christian Wagner, Robert Weismantel
ETH Zurich, IFOR,
e-mail: {michel.baes, timm.oertel, christian.wagner, robert.weismantel}@ifor.math.ethz.ch

Whenever the oracle can be realized efficiently, then the overall running time of the optimization algorithm is efficient as well.

One of the results from the book [GLS88] that is perhaps closest to our results is the following. By $B(p,r)$ we denote a ball of radius r with center p.

Theorem 1. *[GLS88, Theorem 6.7.10] Let n be a fixed integer and $K \subseteq \mathbb{R}^n$ be any convex set given by a weak separation oracle and for which there exist $r,R > 0$ and $p \in K$ with $B(p,r) \subseteq K \subseteq B(0,R)$. There exists an oracle-polynomial algorithm that, for every fixed $\varepsilon > 0$, either finds an integral point in $K + B(0,\varepsilon)$ or concludes that $K \cap \mathbb{Z}^n = \emptyset$.*

The main distinction between results presented here and results from [GLS88] of such flavor as Theorem 1 is the way in which the statements are proven. Proofs of similar results in [GLS88] basically use a combination of the ellipsoid algorithm [Kha79] and a Lenstra-type algorithm [Len83]. Our proof techniques rather rely on methods from convex optimization.

Let us now make precise our assumptions. We study a general mixed-integer convex optimization problem of the kind

$$\min\{f(\hat{x},y) : (\hat{x},y) \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)\}, \tag{1}$$

where the function $f : \mathbb{R}^{n+d} \to \mathbb{R}_+ \cup \{+\infty\}$ is a nonnegative proper convex function, i.e., there is a point $z \in \mathbb{R}^{n+d}$ with $f(z) < +\infty$. Moreover, $S \subseteq \mathbb{R}^{n+d}$ is a convex set that is defined by a finite number of convex functional constraints, i.e., $S := \{(x,y) \in \mathbb{R}^{n+d} : g_i(x,y) \leq 0 \text{ for } 1 \leq i \leq m\}$. We denote by $\langle \cdot, \cdot \rangle$ a scalar product. The functions $g_i : \mathbb{R}^{n+d} \to \mathbb{R}$ are differentiable convex functions and encoded by a so-called *first-order oracle*. Given any point $(x_0,y_0) \in \mathbb{R}^{n+d}$, this oracle returns, for every $i \in \{1,\ldots,m\}$, the function value $g_i(x_0,y_0)$ together with a subgradient $g_i'(x_0,y_0)$, that is, a vector satisfying:

$$g_i(x,y) - g_i(x_0,y_0) \geq \langle g_i'(x_0,y_0), (x - x_0, y - y_0) \rangle$$

for all $(x,y) \in \mathbb{R}^{n+d}$.

In this general setting, very few algorithmic frameworks exist. The most commonly used one is “outer approximation”, originally proposed in [DG86] and later on refined in [VI90, FL94, BBC$^+$08]. This scheme is known to be finitely converging, yet there is no analysis regarding the number of iterations it takes to solve problem (1) up to a certain given accuracy.

In this paper we present oracle-polynomial algorithmic schemes that are (i) amenable to an analysis and (ii) finite for any mixed-integer convex optimization problem. Our schemes also give rise to the fastest algorithm so far for solving mixed-integer convex optimization problems in variable dimension with at most two integer variables.

2 An algorithm based on an “improvement oracle”

We study in this paper an algorithmic approach to solve (1), postponing its hardness to the realization of an improvement oracle defined below. If this oracle can be realized in polynomial time, then the problem can be solved in polynomial time as well. An oracle of this type has already been used in a number of algorithms in other contexts, such as in [AK07] for semidefinite problems.

Definition 1 (Improvement Oracle). Let $\alpha, \delta \geq 0$. For every $z \in S$, the oracle

a. returns $\hat{z} \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ such that $f(\hat{z}) \leq (1+\alpha)f(z) + \delta$, and/or
b. asserts correctly that there is no point $\hat{z} \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ for which $f(\hat{z}) \leq f(z)$.

We denote the query to this oracle at z by $\mathrm{O}_{\alpha,\delta}(z)$.

As stressed in the above definition, the oracle might content itself with a feasible point $\hat{z}$ satisfying the inequality in **a** without addressing the problem in **b**. However, we do not exclude the possibility of having an oracle that can occasionally report both facts. In that case, the point $\hat{z}$ that it outputs for the input point $z \in S$ must satisfy:

$$f(\hat{z}) - \hat{f}^* \leq \alpha f(z) + \delta + (f(z) - \hat{f}^*) \leq \alpha f(z) + \delta \leq \alpha \hat{f}^* + \delta,$$

where $\hat{f}^*$ is the optimal objective value of (1). Thus $f(\hat{z}) \leq (1+\alpha)f(z) + \delta$, and it is not possible to hope for a better point of S from the oracle. We can therefore interrupt the computations and output $\hat{z}$ as the final result of our method.

In the case where $\hat{f}^* > 0$ and $\delta = 0$, the improvement oracle might be realized by a relaxation of the problem of finding a suitable $\hat{z}$: in numerous cases, these relaxations come with a guaranteed value of α. In general, the realization of this oracle might need to solve a problem as difficult as the original mixed-integer convex instance, especially when $\alpha = \delta = 0$. Nevertheless, we will point out several situations where this oracle can actually be realized quite efficiently, even with $\alpha = 0$.

The domain of f, denoted by $\operatorname{dom} f$, is the set of all the points $z \in \mathbb{R}^{n+d}$ with $f(z) < +\infty$. For all $z \in \operatorname{dom} f$, we denote by $f'(z)$ an element of the subdifferential $\partial f(z)$ of f. We represent by $\hat{z}^* = (\hat{x}^*, y^*)$ a minimizer of (1), and set $\hat{f}^* := f(\hat{z}^*)$; more generally, we use a hat $(\hat{\cdot})$ to designate vectors that have their n first components integral by definition or by construction.

Let us describe an elementary method for solving *Lipschitz continuous* convex problems on S approximately. Lipschitz continuity of f on S, an assumption we make from now on, entails that, given a norm $||\cdot||$ on $\mathbb{R}^{n+d}$, there exists a constant $L > 0$ for which:

$$|f(z_1) - f(z_2)| \leq L||z_1 - z_2||$$

for every $z_1, z_2 \in S$. Equivalently, if $||\cdot||_*$ is the dual norm of $||\cdot||$, we have $||f'(z)||_* \leq L$ for every $f'(z) \in \partial f(z)$ and every $z \in \operatorname{dom} f$.

Our first algorithm is a variant of the well-known Mirror-Descent Method (see Chapter 3 of [NDY83]). It requires a termination procedure, which used alone constitutes our second algorithm as a minimization algorithm on its own. However, the second algorithm requires as input an information that is a priori not obvious to get: a point $z \in S$ for which $f(z)$ is a (strictly) positive lower bound of $\hat{f}^*$.

Let $V : \mathbb{R}^{n+d} \to \mathbb{R}_+$ be a differentiable σ-strongly convex function with respect to the norm $||\cdot||$, i.e., there exists a $\sigma > 0$ for which, for every $z_1, z_2 \in \mathbb{R}^{n+d}$, we have:

$$V(z_2) - V(z_1) - \langle V'(z_1), z_2 - z_1 \rangle \geq \frac{\sigma}{2} ||z_2 - z_1||^2.$$

We also use the conjugate V_* of V defined by $V_*(s) := \sup\{\langle s, z \rangle - V(z) : z \in \mathbb{R}^{n+d}\}$ for every $s \in \mathbb{R}^{n+d}$. We fix $z_0 \in S$ as the starting point of our algorithm and denote by M an upper bound of $V(\hat{z}^*)$. We assume that the solution of the problem $\sup\{\langle s, z \rangle - V(z) : z \in \mathbb{R}^{n+d}\}$ exists and can be computed easily, as well as the function $\rho(w) := \min\{||w - z|| : z \in S\}$ for every $w \in \mathbb{R}^{n+d}$, its subgradient, and the minimizer $\pi(w)$. In an alternative version of the algorithm we are about to describe, we can merely assume that the problem $\max\{\langle s, z \rangle - V(z) : z \in S\}$ can be solved efficiently.

A possible building block for constructing an algorithm to solve (1) is the continuous optimum of the problem, that is, the minimizer of (1) without the integrality constraints. The following algorithm is essentially a standard procedure meant to compute an approximation of this continuous minimizer, lined with our oracle that constructs simultaneously a sequence of mixed-integer feasible points following the decrease of f. Except in the rare case when we produce a provably suitable solution to our problem, this algorithm provides a point $z \in S$ such that $f(z)$ is a lower bound of $\hat{f}^*$. Would this lower bound be readily available, we can jump immediately to the termination procedure (see Algorithm 2).

The following proposition is an extension of the standard proof of convergence of Mirror-Descent Methods. We include it here for the sake of completeness.

Proposition 1. *Suppose that the oracle reports* **a** *for $k = 0, \ldots, N$ in Algorithm 1, that is, it delivers an output $\hat{z}_k$ for every iteration $k = 0, \ldots, N$. Then:*

$$\frac{1}{\sum_{k=0}^N h_k} \sum_{k=0}^N \frac{h_k f(\hat{z}_k)}{1 + \alpha_k} - f(\hat{z}^*) \leq \frac{M}{\sum_{k=0}^N h_k} + \frac{2L^2}{\sigma} \cdot \frac{\sum_{k=0}^N h_k^2}{\sum_{k=0}^N h_k} + \frac{1}{\sum_{k=0}^N h_k} \sum_{k=0}^N \frac{h_k \delta_k}{1 + \alpha_k}.$$

Proof. Since V is σ-strongly convex with respect to the norm $||\cdot||$, its conjugate V_* is differentiable and has a Lipschitz continuous gradient of constant $1/\sigma$ for the norm $||\cdot||_*$, i.e., $V_*(y) - V_*(x) \leq \langle V_*'(x), y - x \rangle + \frac{1}{2\sigma} \|y - x\|_*^2$ (see [HUL93, Chapter X]). Also $w_k = V_*'(s_k)$, in view of [Roc81, Theorem 23.5]. Finally, for every $z \in S$, we can write $\rho(w_k) + \langle \rho'(w_k), z - w_k \rangle \leq \rho(z) = 0$. Thus:

$$\langle \rho'(w_k), w_k - \hat{z}^* \rangle \geq \rho(w_k) = ||\pi(w_k) - w_k|| = ||z_k - w_k||. \tag{2}$$

Also, $||\rho'(w_k)||_* \leq 1$, because for every $z \in \mathbb{R}^{n+d}$:

$$\begin{aligned} \langle \rho'(w_k), z - w_k \rangle &\leq \rho(z) - \rho(w_k) = ||z - \pi(z)|| - ||w_k - \pi(w_k)|| \\ &\leq ||z - \pi(w_k)|| - ||w_k - \pi(w_k)|| \leq ||z - w_k||. \end{aligned} \tag{3}$$

By setting $\phi_k := V_*(s_k) - \langle s_k, \hat{z}^* \rangle$, we can write successively for all $k \geq 0$:

```
**Data**: $z_0 \in S$.
Set $\hat{z}_0 := z_0$, $w_0 := z_0$, $s_0 := 0$, and $\hat{f}_0 := f(\hat{z}_0)$.
Select sequences $\{h_k\}_{k\geq 0}$, $\{\alpha_k\}_{k\geq 0}$, $\{\delta_k\}_{k\geq 0}$.
for $k = 0, \ldots, N$ do
    Compute $f'(z_k) \in \partial f(z_k)$ and $\rho'(w_k) \in \partial \rho(w_k)$.
    Set $s_{k+1} := s_k - h_k f'(z_k) - h_k ||f'(z_k)||_* \rho'(w_k)$.
    Set $w_{k+1} := \arg\max\{\langle s_{k+1}, z\rangle - V(z) : z \in \mathbb{R}^{n+d}\}$.
    Set $z_{k+1} := \arg\min\{||w_{k+1} - z|| : z \in S\}$.
    Compute $f(z_{k+1})$.
    if $f(z_{k+1}) \geq \hat{f}_k$ then $\hat{z}_{k+1} := \hat{z}_k$, $\hat{f}_{k+1} := \hat{f}_k$.

    else
        Run $\mathrm{O}_{\alpha_{k+1},\delta_{k+1}}(z_{k+1})$.
        if the oracle reports a and b then
            Terminate the algorithm and return the oracle output from a.
        else if the oracle reports a but not b then
            Set $\hat{z}_{k+1}$ as the oracle output and $\hat{f}_{k+1} := \min\{f(\hat{z}_{k+1}), \hat{f}_k\}$.
        else
            Run the termination procedure with $z_0 := z_{k+1}$, $\hat{z}_0 := \hat{z}_{k+1}$,
            return its output, and terminate the algorithm.
        end
    end
end
```

Algorithm 1: Mirror-Descent Method.

```
Data: $z_0 \in S$ with $f(z_0) \leq \hat{f}^*$, $\hat{z}_0 \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$.
Set $l_0 := f(z_0)$, $u_0 := f(\hat{z}_0)$.
Choose $\alpha, \delta \geq 0$. Choose a subproblem accuracy $\varepsilon' > 0$.
for $k \geq 0$ do
    Compute using a bisection method a point $z_{k+1} = \lambda z_k + (1-\lambda)\hat{z}_k$
    for $0 \leq \lambda \leq 1$, for which $f(z_{k+1}) - (l_k(\alpha+1) + u_k)/(\alpha+2) \in [-\varepsilon', \varepsilon']$.
    Run $\mathrm{O}_{\alpha,\delta}(z_{k+1})$.
    if the oracle reports a and b then
        Terminate the algorithm and return the oracle output from a.
    else if the oracle reports a but not b then
        Set $\hat{z}_{k+1}$ as the oracle output, $l_{k+1} := l_k$, $u_{k+1} := \min\{f(\hat{z}_{k+1}), u_k\}$.
    else
        Set $\hat{z}_{k+1} := \hat{z}_k$, $l_{k+1} := f(z_{k+1})$, $u_{k+1} := u_k$.
    end
end
```

Algorithm 2: Termination procedure.

$$
\begin{aligned}
\phi_{k+1} &= V_*(s_{k+1}) - \langle s_{k+1}, \hat{z}^* \rangle \\
&\leq V_*(s_k) + \langle V_*'(s_k), s_{k+1} - s_k \rangle + \frac{1}{2\sigma} \|s_{k+1} - s_k\|_*^2 - \langle s_{k+1}, \hat{z}^* \rangle. \\
&= (V_*(s_k) - \langle s_k, \hat{z}^* \rangle) + \left\langle V_*'(s_k) - \hat{z}^*, s_{k+1} - s_k \right\rangle + \frac{1}{2\sigma} \|s_{k+1} - s_k\|_*^2 \\
&= \phi_k - h_k \left\langle w_k - z_k, f'(z_k) \right\rangle + h_k \left\langle \hat{z}^* - z_k, f'(z_k) \right\rangle
\end{aligned}
$$

$$-h_k||f'(z_k)||_* \left\langle w_k - \hat{z}^*, \rho'(w_k)\right\rangle + \frac{h_k^2\|f'(z_k)\|_*^2}{2\sigma}\left\|\frac{f'(z_k)}{\|f'(z_k)\|_*} + \rho'(w_k)\right\|_*^2,$$

where the inequality follows from the Lipschitz continuity of the gradient of V_*, and the last equality from the identities $V_*'(s_k) = w_k$, $s_{k+1} - s_k = -h_k f'(z_k) - h_k\|f'(z_k)\|_*\rho'(w_k)$, and $V_*(s_k) - \langle s_k, \hat{z}^*\rangle = \phi_k$. By the definition of the dual norm, it holds $-h_k\langle w_k - z_k, f'(z_k)\rangle \leq h_k\|f'(z_k)\|_*\|w_k - z_k\|$. Moreover, convexity of f implies $h_k\langle \hat{z}^* - z_k, f'(z_k)\rangle \leq f(\hat{z}^*) - f(z_k)$. Using this in the above expression we get:

$$\begin{aligned}\phi_{k+1} &\leq \phi_k + h_k||f'(z_k)||_* \left(||w_k - z_k|| - \left\langle w_k - \hat{z}^*, \rho'(w_k)\right\rangle\right)\\ &\quad + h_k(f(\hat{z}^*) - f(z_k)) + \frac{h_k^2||f'(z_k)||_*^2}{2\sigma}\left(\left\|\frac{f'(z_k)}{||f'(z_k)||_*}\right\|_* + \left\|\rho'(w_k)\right\|_*\right)^2\\ &\leq \phi_k + h_k(f(\hat{z}^*) - f(z_k)) + \frac{2h_k^2||f'(z_k)||_*^2}{\sigma}\\ &\leq \phi_k + h_k\left(f(\hat{z}^*) - \frac{f(\hat{z}_k) - \delta_k}{1+\alpha_k}\right) + \frac{2h_k^2||f'(z_k)||_*^2}{\sigma},\end{aligned}$$

where the second inequality follows from (2) and $\|\rho'(w_k)\|_* \leq 1$, and the third inequality from the fact that the oracle reports **a**. Summing up the above inequalities from $k := 0$ to $k := N$ and rearranging, it follows:

$$\frac{1}{\sum_{k=0}^N h_k}\sum_{k=0}^N \frac{h_k(f(\hat{z}_k) - \delta_k)}{1+\alpha_k} - f(\hat{z}^*) \leq \frac{\phi_0 - \phi_{N+1}}{\sum_{k=0}^N h_k} + \frac{2\sum_{k=0}^N h_k^2||f'(z_k)||_*^2}{\sigma\sum_{k=0}^N h_k}.$$

Note that $||f'(z_k)||_* \leq L$, $\phi_0 = \sup\{-V(z) : z \in \mathbb{R}^{n+d}\} \leq 0$, and $\phi_{N+1} \geq -V(\hat{z}^*) \geq -M$, yielding the desired result. □

In the special case when $\alpha_k = \alpha$ and $\delta_k = \delta$ for every $k \geq 0$, we can significantly simplify the above results. According to the previous proposition, we know that:

$$\begin{aligned}\left(\sum_{k=0}^N h_k\right)\left(\frac{\hat{f}_N - \delta}{1+\alpha} - \hat{f}^*\right) &= \left(\sum_{k=0}^N h_k\right)\left(\frac{\min_{1\leq i\leq N} f(\hat{z}_i) - \delta}{1+\alpha} - \hat{f}^*\right)\\ &\leq \sum_{k=0}^N \frac{h_k(f(\hat{z}_k) - \delta)}{1+\alpha} - \left(\sum_{k=0}^N h_k\right)\hat{f}^* \leq M + \frac{2L^2}{\sigma}\sum_{k=0}^N h_k^2. \qquad (4)\end{aligned}$$

We can divide both sides of the above inequality by $\sum_{k=0}^N h_k$, then determine the step-sizes $\{h_k : 0 \leq k \leq N\}$ for which the right-hand side is minimized. However, with this strategy, h_0 would depend on N, which is a priori unknown at the first iteration. Instead, as in [Nes04], we use a step-size of the form $h_k = c/\sqrt{k+1}$ for an appropriate constant $c > 0$, independent of N. Note that:

$$\sum_{k=0}^N \frac{1}{k+1} = \sum_{k=1}^{N+1}\frac{1}{k} \leq \int_1^{N+1}\frac{dt}{t} + 1 = \ln(N+1) + 1.$$

If we choose $c := \sqrt{\frac{\sigma M}{2L^2}}$, the right-hand side of (4) can be upper-bounded by $M\ln(N+1)+2M$. Finally, since

$$\frac{1}{c}\sum_{k=0}^{N} h_k = \sum_{k=0}^{N} \frac{1}{\sqrt{k+1}} = \sum_{k=1}^{N+1} \frac{1}{\sqrt{k}} \geq \int_{1}^{N+2} \frac{dt}{\sqrt{t}} = 2\sqrt{N+2}-2,$$

we can thereby conclude that:

$$\frac{\hat{f}_N - (1+\alpha)\hat{f}^* - \delta}{1+\alpha} \leq L\sqrt{\frac{M}{2\sigma}} \cdot \frac{\ln(N+1)+2}{\sqrt{N+2}-1}. \tag{5}$$

As the right-hand side converges to 0 when N goes to infinity, Algorithm 1 converges to an acceptable approximate solution or calls the termination procedure.

Let us now turn our attention to the termination procedure. We assume here that the oracle achieves a constant quality, that is, that there exists $\alpha, \delta \geq 0$ for which $\alpha_k = \alpha$ and $\delta_k = \delta$ for every $k \geq 0$.

Proposition 2. *Assume that $f(\hat{z}_0) \geq f(z_0) > 0$, and that there is no point $\hat{z} \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ for which $f(z_0) > f(\hat{z})$.*

(a) The termination procedure cannot guarantee an accuracy better than:

$$f(\hat{z}) \leq \hat{f}^* + (2+\alpha)\left(\alpha\hat{f}^* + (1+\alpha)\varepsilon' + \delta\right). \tag{6}$$

(b) For every $\varepsilon > 0$, the termination procedure finds a point $\hat{z} \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ satisfying:

$$f(\hat{z}) - \hat{f}^* \leq \varepsilon\hat{f}^* + (2+\alpha)\left(\alpha\hat{f}^* + (1+\alpha)\varepsilon' + \delta\right)$$

within

$$\max\left\{\left\lceil \ln\left(\frac{f(\hat{z}_0)-f(z_0)}{f(z_0)\varepsilon}\right) \Big/ \ln\left(\frac{2+\alpha}{1+\alpha}\right)\right\rceil, 0\right\}$$

iterations.

Proof. Part (a). At every iteration k, there is by construction no $\hat{z} \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ for which $l_k > f(\hat{z})$. Also, $f(\hat{z}_k) \geq u_k \geq \hat{f}^*$. For convenience, we denote $(1+\alpha)/(2+\alpha)$ by λ in this proof, and we set $\Delta_k := u_k - l_k$ for every $k \geq 0$.

Suppose first that the oracle finds a new point $\hat{z}_{k+1} \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ at iteration k. Then:

$$f(\hat{z}_{k+1}) \leq (1+\alpha)f(z_{k+1}) + \delta \leq (1+\alpha)\left(\lambda l_k + (1-\lambda)u_k + \varepsilon'\right) + \delta,$$

where the first inequality is due to the definition of our oracle and the second one comes from the accuracy by which our bisection procedure computes z_{k+1}. Observe that the oracle might return a point $\hat{z}_k$ such that $f(\hat{z}_k)$ is smaller than the above right-hand side. In this case, no progress is done. As $u_k \leq f(\hat{z}_k)$, this implies:

$$(\lambda + \lambda\alpha)l_k + (1+\alpha)\varepsilon' + \delta \geq (\lambda + \lambda\alpha - \alpha)f(\hat{z}_k). \tag{7}$$

Using that $\hat{f}^* \geq l_k$ we get an upper bound of the left-hand side. Rearranging the terms and replacing λ by its value, we get:

$$\hat{f}^* + (2+\alpha)(\alpha\hat{f}^* + (1+\alpha)\varepsilon' + \delta) \geq f(\hat{z}_k).$$

Since all the inequalities in the above derivation can be tight, a better accuracy cannot be guaranteed with our strategy. Thus, we can output $\hat{z}_k$.
Part (b). Note that we can assume $\frac{f(\hat{z}_0)-f(z_0)}{f(z_0)\varepsilon} > 1$, for otherwise the point $\hat{z}_0$ already satisfies our stopping criterion.

In order to assess the progress of the algorithm, we can assume that the stopping criterion (7) is not satisfied. As $l_{k+1} = l_k$ in our case where the oracle gives an output, we get:

$$\begin{aligned}
\Delta_{k+1} &= u_{k+1} - l_k \leq f(\hat{z}_{k+1}) - l_k \\
&\leq (1+\alpha)\left(\lambda l_k + (1-\lambda)u_k + \varepsilon'\right) + \delta - l_k \\
&= \frac{\alpha^2+\alpha-1}{2+\alpha} l_k + \frac{1+\alpha}{2+\alpha} u_k + (1+\alpha)\varepsilon' + \delta \\
&= \frac{1+\alpha}{2+\alpha}(u_k - l_k) + \alpha l_k + (1+\alpha)\varepsilon' + \delta \\
&\leq \frac{1+\alpha}{2+\alpha}\Delta_k + \alpha\hat{f}^* + (1+\alpha)\varepsilon' + \delta.
\end{aligned}$$

Suppose now that the oracle informs us that there is no mixed-integral point with a value smaller than $f(z_{k+1}) \geq \lambda l_k + (1-\lambda)u_k - \varepsilon'$. Then $\hat{z}_{k+1} = \hat{z}_k$ and $u_{k+1} = u_k$. We have:

$$\begin{aligned}
\Delta_{k+1} &= u_{k+1} - l_{k+1} = f(\hat{z}_k) - f(z_{k+1}) \\
&\leq u_k - \left(\lambda l_k + (1-\lambda)u_k - \varepsilon'\right) = \lambda\Delta_k + \varepsilon' \\
&\leq \frac{1+\alpha}{2+\alpha}\Delta_k + \alpha\hat{f}^* + (1+\alpha)\varepsilon' + \delta.
\end{aligned}$$

The above inequality is valid for every k that does not comply with the stopping criterion, whatever the oracle detects. Therefore, we get:

$$\Delta_N \leq \left(\frac{1+\alpha}{2+\alpha}\right)^N \Delta_0 + (2+\alpha)\left(\alpha\hat{f}^* + (1+\alpha)\varepsilon' + \delta\right),$$

and the proposition is proved because $f(\hat{z}_N) - \hat{f}^* \leq \Delta_N$. □

In the remainder of this paper, we elaborate on possible realizations of our hard oracle.

We proceed as follows. In Section 3, we focus on the special case when $n = 2$ and $d = 0$. We present a geometric construction that enables us to implement the improvement oracle in polynomial time. With the help of this oracle we then solve the problem (1) with $n = 2$ and $d = 0$ and obtain a “best point”, i.e., an optimal point. An adaptation of this construction can also be used to determine a second and, more generally, a "k-th best point". These results will be extended in Section 4 to the mixed-integer case with two integer variables and d continuous variables. The latter extensions are then used as a subroutine to solve the general problem (1) with arbitrary n and d in finite time.

3 Two-dimensional integer convex optimization

If $n = 1$ and $d = 0$, an improvement oracle can be trivially realized for $\alpha = \delta = 0$. Queried on a point $z \in \mathbb{R}$ the oracle returns $\hat{z} := \arg\min\{f(\lfloor z \rfloor), f(\lceil z \rceil)\}$ if one of these numbers is smaller or equal to $f(z)$, or returns $\mathbf{b}$ otherwise. The first non-trivial case arises when $n = 2$ and $d = 0$. This is the topic of this section.

3.1 Minimizing a convex function in two integer variables

In this section we discuss a new geometric construction that enables us to implement efficiently the oracle $\mathrm{O}_{\alpha,\delta}$ with $\alpha = \delta = 0$, provided that the feasible set is contained in a known finite box $[-B,B]^2$.

Theorem 2. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ and $g_i : \mathbb{R}^2 \to \mathbb{R}$ with $i = 1,\ldots,m$ be convex functions. Let $B \in \mathbb{N}$ and let $x \in [-B,B]^2$ such that $g_i(x) \leq 0$ for all $i = 1,\ldots,m$. Then, in a number of evaluations of f and $g_1,\ldots,g_m$ that is polynomial in $\ln(B)$, one can either*

(a) find an $\hat{x} \in [-B,B]^2 \cap \mathbb{Z}^2$ with $f(\hat{x}) \leq f(x)$ and $g_i(\hat{x}) \leq 0$ for all $i = 1,\ldots,m$ or
(b) show that there is no such point.

Note that we do not allow for the function f to take infinite values, in order to ensure that we can minimize f over the integers of any segment of $[-B,B]^2$ in $\mathcal{O}(\ln(B))$ evaluations of f using a bisection method. Indeed, if a convex function takes infinite values, it can cost up to $\mathcal{O}(B)$ evaluations of f to minimize it on a segment containing $\mathcal{O}(B)$ integer points, as there could be only one of those points on its domain.

The algorithm that achieves the performance claimed in Theorem 2 is described in the proof of the theorem. That proof requires two lemmata. We use the following notation. Let $Q \subset \mathbb{R}^2$. We denote by $\mathrm{vol}(Q)$ the volume of Q, i.e., its Lebesgue measure. By $\mathrm{aff}\{Q\}$ we denote the smallest affine space containing Q and by $\mathrm{conv}\{Q\}$ the convex hull of Q. The dimension $\dim(Q)$ of Q is the dimension of $\mathrm{aff}\{Q\}$. The scalar product we use in this section is exclusively the standard dot product.

Lemma 1. *Let $K \subset \mathbb{R}^2$ be a polytope with $\mathrm{vol}(K) < \frac{1}{2}$. Then $\dim(\mathrm{conv}(K \cap \mathbb{Z}^2)) \leq 1$.*

Proof. For the purpose of deriving a contradiction, assume that there exist three affinely independent points $\hat{x},\hat{y},\hat{z} \in K \cap \mathbb{Z}^2$. Then $\mathrm{vol}(K) \geq \mathrm{vol}(\mathrm{conv}(\{\hat{x},\hat{y},\hat{z}\})) = \frac{1}{2}|\det(\hat{x}-\hat{z},\hat{y}-\hat{z})| \geq \frac{1}{2}$. □

Lemma 2. *Let $u, v, w \in \mathbb{R}^2$ be affinely independent. If*

$$\big(\operatorname{conv}\{u, u+v, u+v+w\} \setminus (\operatorname{conv}\{u+v, u+v+w\} \cup \{u\})\big) \cap \mathbb{Z}^2 = \emptyset,$$

then the lattice points $\operatorname{conv}\{u, u+v, u+v-w\} \cap \mathbb{Z}^2$ *lie on at most three lines.*

Proof. We partition $\operatorname{conv}\{u, u+v, u+v-w\}$ into three regions. Then we show that in each region the integer points must lie on a single line using a lattice covering argument.

We define the parallelogram $P := \operatorname{conv}\{0, \frac{1}{2}v, \frac{1}{2}w, \frac{1}{2}v + \frac{1}{2}w\}$. Further, we set

$$A_1 := u - \frac{1}{2}w + P, \quad A_2 := u + \frac{1}{2}v - w + P, \quad \text{and} \quad A_3 := u + \frac{1}{2}v - \frac{1}{2}w + P.$$

Note that $\operatorname{conv}\{u, u+v, u+v-w\} \subset A_1 \cup A_2 \cup A_3$ (see Fig. 1). Our assumption implies that the set $u + \frac{1}{2}v + P$ does not contain any integer point except possibly on the segment $u + v + \operatorname{conv}\{0, w\}$. Therefore, for a sufficiently small $\varepsilon > 0$, the set $(u + \frac{1}{2}v - \varepsilon(v+w) + P) \cap \mathbb{Z}^2$ is empty.

Assume now that one of the three regions, say A_1, contains three affinely independent integer points $\hat{x}, \hat{y}, \hat{z}$. We show below that $A_1 + \mathbb{Z}^2 = \mathbb{R}^2$, i.e., that P defines a lattice covering, or equivalently that the set $t + P$ contains at least one integer point for every $t \in \mathbb{R}^2$. This fact will contradict that $(u + \frac{1}{2}v - \varepsilon(v+w) + P) \cap \mathbb{Z}^2 = \emptyset$ and thereby prove the lemma.

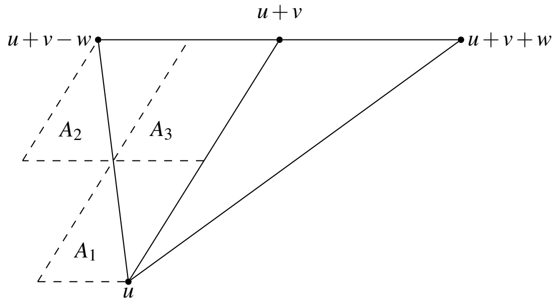

Fig. 1 Partitioning the triangle in regions.

Clearly, the parallelogram $Q := \operatorname{conv}\{\hat{x}, \hat{y}, \hat{z}, \hat{x} - \hat{y} + \hat{z}\}$ defines a lattice covering, as it is full-dimensional and its vertices are integral. We transform Q into a set $Q' \subseteq A_1$ for which $a \in Q'$ iff there exists $b \in Q$ such that $a - b \in \mathbb{Z}^2$. Specifically, we define a mapping T such that $Q' = T(Q) \subset A_1$ and $T(Q) + \mathbb{Z}^2 = \mathbb{R}^2$. Let $v^\perp := (-v_2, v_1)^\top$ and $w^\perp := (-w_2, w_1)^\top$, i.e., vectors orthogonal to v and w. Without loss of generality (up to a permutation of the names $\hat{x}, \hat{y}, \hat{z}$), we can assume that $\langle \hat{x}, w^\perp \rangle \leq \langle \hat{y}, w^\perp \rangle \leq \langle \hat{z}, w^\perp \rangle$. If $\hat{x} - \hat{y} + \hat{z} \in A_1$ there is nothing to show, so we suppose that $\hat{x} - \hat{y} + \hat{z} \notin A_1$.

Note that $\langle \hat{x}, w^\perp \rangle \leq \langle \hat{x} - \hat{y} + \hat{z}, w^\perp \rangle \leq \langle \hat{z}, w^\perp \rangle$. Assume first that $\langle \hat{x} - \hat{y} + \hat{z}, v^\perp \rangle < \langle \hat{z}, v^\perp \rangle \leq \langle \hat{x}, v^\perp \rangle, \langle \hat{y}, v^\perp \rangle$ — the strict inequality resulting from the fact that $\hat{x} - \hat{y} + \hat{z} \notin$

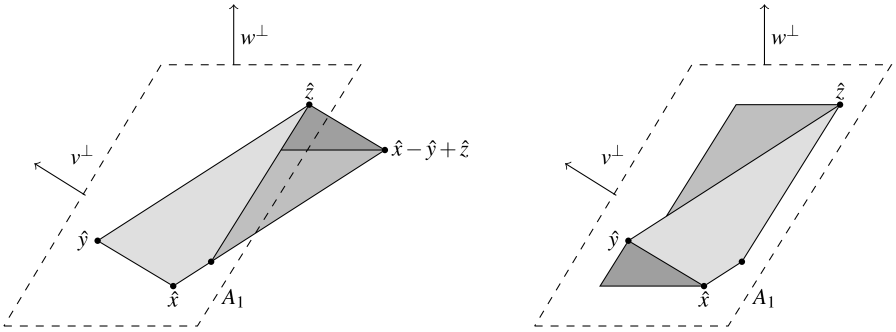

Fig. 2 Mapping T.

A_1. We define the mapping $T : Q \to A_1$ as follows,

$$T(l) = \begin{cases} l+\hat{y}-\hat{z}, & \text{if } \langle l, v^\perp\rangle < \langle \hat{z}, v^\perp\rangle \text{ and } \langle l, w^\perp\rangle > \langle \hat{x}-\hat{y}+\hat{z}, w^\perp\rangle, \\ l-\hat{x}+\hat{y}, & \text{if } \langle l, v^\perp\rangle < \langle \hat{z}, v^\perp\rangle \text{ and } \langle l, w^\perp\rangle \leq \langle \hat{x}-\hat{y}+\hat{z}, w^\perp\rangle, \\ l, & \text{otherwise} \end{cases}$$

(see Fig. 2). It is straightforward to show that $T(Q) \subset A_1$ and $T(Q) + \mathbb{Z}^2 = \mathbb{R}^2$. A similar construction can easily be defined for any possible ordering of $\langle \hat{x}-\hat{y}+\hat{z}, v^\perp\rangle$, $\langle \hat{z}, v^\perp\rangle$, $\langle \hat{x}, v^\perp\rangle$, and $\langle \hat{y}, v^\perp\rangle$. □

Remark 1. In each region A_i, the line containing $A_i \cap \mathbb{Z}^2$, if it exists, can be computed by the minimization of an arbitrary linear function $x \mapsto \langle c, x\rangle$ over $A_i \cap \mathbb{Z}^2$, with $c \neq 0$, and the maximization of the same function with the fast algorithm described in [EL05]. If these problems are feasible and yield two distinct solutions, the line we are looking for is the one joining these two solutions. If the two solutions coincide, that line is the one orthogonal to c passing through that point.

The algorithm in [EL05] is applicable to integer linear programs with two variables and m constraints. The data of the problem should be integral. This algorithm runs in $\mathscr{O}(m+\phi)$, where ϕ is the binary encoding length of the data. ◇

Proof (of Theorem 2). As described at the beginning of this subsection, a one-dimensional integer minimization problem can be solved polynomially with respect to the logarithm of the length of the segment that the function is optimized over. In the following we explain how to reduce the implementation of the two-dimensional oracle to the task of solving one-dimensional integer minimization problems. For notational convenience, we define $g(y) := \max_{i=1\ldots m} g_i(y)$ for $y \in \mathbb{R}^2$ which is again a convex function.

Let $F_1, \ldots, F_4$ be the facets of $[-B,B]^2$. Then $[-B,B]^2 = \bigcup_{j=1}^4 \operatorname{conv}\{x, F_j\}$. The procedure we are about to describe has to be applied to every facet $F_1, \ldots, F_4$ successively, until a suitable point $\hat{x}$ is found. Let us only consider one facet F. We define the triangle $T_0 := \operatorname{conv}\{x, F\}$, whose area is smaller than $2B^2$.

To find an improving point within T_0, we construct a sequence $T_0 \supset T_1 \supset T_2 \supset \dots$ of triangles that all have x as vertex, with $\mathrm{vol}(T_{k+1}) \leq \frac{2}{3}\mathrm{vol}(T_k)$, and such that $f(\hat{y}) > f(x)$ or $g(\hat{y}) > 0$ for all $\hat{y} \in (T_0 \setminus T_k) \cap \mathbb{Z}^2$. We stop our search if we have found an $\hat{x} \in [-B,B]^2 \cap \mathbb{Z}^2$ such that $f(\hat{x}) \leq f(x)$ and $g(\hat{x}) \leq 0$, or if the volume of one of the triangles T_k is smaller than $\frac{1}{2}$. The latter happens after at most $k = \lceil \ln(4B^2)/\ln(\frac{3}{2}) \rceil$ steps. Then, Lemma 1 ensures that the integral points of T_k are on a line, and we need at most $\mathscr{O}(\ln(B))$ iterations to solve the resulting one-dimensional problem.

The iterative construction is as follows. Let $T_k = \mathrm{conv}\{x, v_0, v_1\}$ be given. We write $v_\lambda := (1-\lambda)v_0 + \lambda v_1$ for $\lambda \in \mathbb{R}$ and we define the auxiliary triangle $\bar{T}_k := \mathrm{conv}\{x, v_{1/3}, v_{2/3}\}$. Consider the integer linear program

$$\min\{\langle h, \hat{y}\rangle : \hat{y} \in \bar{T}_k \cap \mathbb{Z}^2\} \tag{8}$$

where h is the normal vector to $\mathrm{conv}\{v_0, v_1\}$ such that $\langle h, x\rangle < \langle h, y\rangle$ for every $y \in F$. We distinguish two cases.

Case 1. The integer linear program (8) is infeasible. Then $\bar{T}_k \cap \mathbb{Z}^2 = \emptyset$. It remains to check for an improving point within $(T_k \setminus \bar{T}_k) \cap \mathbb{Z}^2$. By construction, we can apply Lemma 2 twice (with $(u, u+v-w, u+v+w)$ equal to $(x, v_0, v_{2/3})$ and $(x, v_{1/3}, v_1)$, respectively) to determine whether there exists an $\hat{x} \in (T_k \setminus \bar{T}_k) \cap \mathbb{Z}^2$ such that $f(\hat{x}) \leq f(x)$ and $g(\hat{x}) \leq 0$. This requires to solve at most six one-dimensional subproblems.

Case 2. The integer linear program (8) has an optimal solution $\hat{z}$. If $f(\hat{z}) \leq f(x)$ and $g(\hat{z}) \leq 0$, we are done. So we assume that $f(\hat{z}) > f(x)$ or $g(\hat{z}) > 0$. Define $H := \{y \in \mathbb{R}^2 \mid \langle h, y\rangle = \langle h, \hat{z}\rangle\}$, that is, the line containing $\hat{z}$ that is parallel to $\mathrm{conv}\{v_0, v_1\}$, and denote by H_+ the closed half-space with boundary H that contains x. By definition of $\hat{z}$, there is no integer point in $\bar{T}_k \cap \mathrm{int} H_+$. Further, let $L := \mathrm{aff}\{x, \hat{z}\}$.

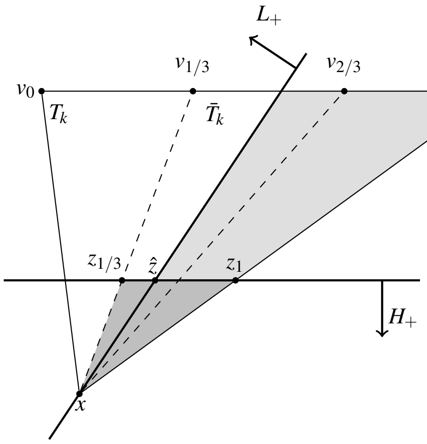

Fig. 3 Illustration of Case 2.

Due to the convexity of the set $\{y \in \mathbb{R}^2 \mid f(y) \leq f(x),\ g(y) \leq 0\}$ and the fact that $f(\hat{z}) > f(x)$ or $g(\hat{z}) > 0$, there exists a half-space L_+ with boundary L such that the possibly empty segment $\{y \in H \mid f(y) \leq f(x),\ g(y) \leq 0\}$ lies in L_+ (see Fig. 3). By convexity of f and g, the set $((T_k \setminus H_+) \setminus L_+)$ (the lightgray region in Fig. 3) contains no point y for which $f(y) \leq f(x)$ and $g(y) \leq 0$. It remains to check for an improving point within $((T_k \cap H_+) \setminus L_+) \cap \mathbb{Z}^2$. For that we apply again Lemma 2 on the triangle $\text{conv}\{z_{1/3}, z_1, x\}$ (the darkgray region in Fig. 3), with $z_{1/3} = H \cap \text{aff}\{x, v_{1/3}\}$ and $z_1 = H \cap \text{aff}\{x, v_1\}$. If none of the corresponding subproblems returns a suitable point $\hat{x} \in \mathbb{Z}^2$, we know that $T_k \setminus L_+$ contains no improving integer point. Defining $T_{k+1} := T_k \cap L_+$, we have by construction $f(\hat{y}) > f(x)$ or $g(\hat{y}) > 0$ for all $\hat{y} \in (T_k \setminus T_{k+1}) \cap \mathbb{Z}^2$ and $\text{vol}(T_{k+1}) \leq \frac{2}{3}\text{vol}(T_k)$.

It remains to determine the half-space L_+. If $g(\hat{z}) > 0$ we just need to find a point $y \in H$ such that $g(y) < g(\hat{z})$, or if $f(\hat{z}) > f(x)$, it suffices to find a point $y \in H$ such that $f(y) < f(\hat{z})$. Finally, if we cannot find such a point y in either case, convexity implies that there is no suitable point in $T_k \setminus H_+$; another application of Lemma 2 then suffices to determine whether there is a suitable $\hat{x}$ in $T_k \cap H_+ \cap \mathbb{Z}^2$. □

The algorithm presented in the proof of Theorem 2 can be adapted to output a minimizer $\hat{x}^*$ of f over $S \cap [-B,B]^2 \cap \mathbb{Z}^2$, provided that we know in advance that the input point x satisfies $f(x) \leq \hat{f}^*$: it suffices to store and update the best value of f on integer points found so far. In this case the termination procedure is not necessary.

Corollary 1. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ and $g_i : \mathbb{R}^2 \to \mathbb{R}$ with $i = 1, \ldots, m$ be convex functions. Let $B \in \mathbb{N}$ and let $x \in [-B,B]^2$ such that $g_i(x) \leq 0$ for all $i = 1, \ldots, m$. If $f(x) \leq \hat{f}^*$, then, in a number of evaluations of f and $g_1, \ldots, g_m$ that is polynomial in* $\ln(B)$*, one can either*

(a) find an $\hat{x} \in [-B,B]^2 \cap \mathbb{Z}^2$ with $f(\hat{x}) = \hat{f}^*$ and $g_i(\hat{x}) \leq 0$ for all $i = 1, \ldots, m$ or
(b) show that there is no such point.

Note that line 33 in Algorithm 3 requires the application of Lemma 1. Lines 11, 21 and 26 require the application of Lemma 2.

Remark 2 (Complexity). The following subroutines are used in Algorithm 3.

Line 9 and applications of Lemma 2. A two-dimensional integer linear program solver for problems having at most four constraints, such as the one described in [EL05]. The size of the data describing each of these constraints is in the order of the representation of the vector x as a rational number, which, in its standard truncated decimal representation, is in $\mathcal{O}(\ln(B))$.

Line 33 and applications of Lemma 2. A solver for one-dimensional integer convex optimization problems. At every iteration, we need to perform at most seven of them, for a cost of $\mathcal{O}(\ln(B))$ at each time.

Lines 19 and 20. Given a segment $[a,b]$ and one of its points z, we need a device to determine which of the two regions $[a,z]$ or $[z,b]$ intersects a level set defined by f and g that does not contain z. This procedure has a complexity of $\mathcal{O}(\ln(B))$ and only occurs in Case 2 above. ◇

```
Data: $x \in [-B,B]^2$ with $f(x) \le \hat{f}^*$ and $g_i(x) \le 0$ for all $i = 1,\dots,m$.
Let $F_1,\dots,F_4$ be the facets of $[-B,B]^2$.
Set $\hat{x}^* := 0$ and $\hat{f}^* := +\infty$.
for $t = 1,\dots,4$ do
    Set $F := F_t$ and define $v_0, v_1 \in \mathbb{R}^n$ such that $F := \mathrm{conv}\{v_0, v_1\}$.
    Write $h$ for the vector normal to $F$ pointing outwards $[-B,B]^2$.
    Set $T_0 := \mathrm{conv}\{x, F\}$ and $k := 0$.
    while $\mathrm{vol}(T_k) \ge \frac{1}{2}$ do
        Set $\bar{T}_k := \mathrm{conv}\{x, v_{1/3}, v_{2/3}\}$, with $v_\lambda := (1-\lambda)v_0 + \lambda v_1$.
        Solve $(\mathscr{P}) : \min\{\langle h, \hat{y}\rangle : \hat{y} \in \bar{T}_k \cap \mathbb{Z}^2\}$.
        if (Case 1) $(\mathscr{P})$ is infeasible, then
            Determine $\hat{x} := \arg\min\{f(\hat{y}) \mid \hat{y} \in (T_k \setminus \bar{T}_k) \cap \mathbb{Z}^2$ with $g(\hat{y}) \le 0\}$.
            if $\hat{x}$ exists and $f(\hat{x}) < \hat{f}^*$ then Set $\hat{x}^* := \hat{x}$ and $\hat{f}^* := f(\hat{x})$.

        else
            (Case 2) Let $\hat{z}$ be an optimal solution of $(\mathscr{P})$.
            Set $H_+ := \{y \in \mathbb{R}^2 : \langle h, y\rangle \le \langle h, \hat{z}\rangle\}$ and $H := \partial H_+$.
            Define the points $v := \mathrm{aff}\{x, \hat{z}\} \cap F$ and $z_i = H \cap \mathrm{conv}\{x, v_i\}$ for $i = 0, 1$.
            Denote $z_\lambda := (1-\lambda)z_0 + \lambda z_1$ for $\lambda \in (0,1)$.
            if $g(\hat{z}) \le 0$ and there is a $y \in \mathrm{conv}\{z_0, \hat{z}\}$ for which $f(y) < f(\hat{z})$ or
              $g(\hat{z}) > 0$ and there is a $y \in \mathrm{conv}\{z_0, \hat{z}\}$ for which $g(y) < g(\hat{z})$ then
                Determine $\hat{x} := \arg\min\{f(\hat{y}) \mid \hat{y} \in \mathrm{conv}\{x, z_{1/3}, z_1\} \cap \mathbb{Z}^2$ with $g(\hat{y}) \le 0\}$.
                if $\hat{x}$ exists and $f(\hat{x}) < \hat{f}^*$ then Set $\hat{x}^* := \hat{x}$ and $\hat{f}^* := f(\hat{x})$.

                Set $v_1 := v$, $T_{k+1} := \mathrm{conv}\{x, v_0, v\}$, and $k := k+1$.
            else
                Determine $\hat{x} := \arg\min\{f(\hat{y}) \mid \hat{y} \in \mathrm{conv}\{x, z_0, z_{2/3}\} \cap \mathbb{Z}^2$ with $g(\hat{y}) \le 0\}$.
                if $\hat{x}$ exists and $f(\hat{x}) < \hat{f}^*$ then Set $\hat{x}^* := \hat{x}$ and $\hat{f}^* := f(\hat{x})$.

                Set $v_0 := v$, $T_{k+1} := \mathrm{conv}\{x, v, v_1\}$, and $k := k+1$.
            end
        end
    end
    Determine $\hat{x} := \arg\min\{f(\hat{y}) \mid \hat{y} \in T_k \cap \mathbb{Z}^2$ with $g(\hat{y}) \le 0\}$.
    if $\hat{x}$ exists and $f(\hat{x}) < \hat{f}^*$ then Set $\hat{x}^* := \hat{x}$ and $\hat{f}^* := f(\hat{x})$.

end
if $\hat{f}^* < +\infty$ then Return $\hat{x}^*$.

else Return “the problem is unfeasible”.
```

Algorithm 3: Minimization algorithm for 2D problems.

3.2 Finding the k-th best point

In this subsection we want to show how to find the k-th best point, provided that the $k-1$ best points are known. A slight variant of this problem will be used in Subsection 4.3 as a subroutine for the general mixed-integer convex problem. In the following, we describe the necessary extensions of the previous Algorithm 3. Let $\hat{x}_1^* := \hat{x}^*$ and define for $k \geq 2$:

$$\hat{x}_k^* := \arg\min \left\{ f(\hat{x}) \mid \hat{x} \in (S \cap [-B,B]^2 \cap \mathbb{Z}^2) \setminus \{\hat{x}_1^*, \ldots, \hat{x}_{k-1}^*\} \right\}$$

to be the k-th best point. Observe that, due to the convexity of f and $g_1, \ldots, g_m$, we can always assume that $\operatorname{conv}\{\hat{x}_1^*, \ldots, \hat{x}_{k-1}^*\} \cap \mathbb{Z}^2 = \{\hat{x}_1^*, \ldots, \hat{x}_{k-1}^*\}$ for all $k \geq 2$. Although this observation appears plausible it is not completely trivial to achieve this algorithmically.

Lemma 3. *Let $\Pi_j := \{\hat{x}_1^*, \ldots, \hat{x}_j^*\}$ be the ordered j best points of our problem and P_j be the convex hull of Π_j. Suppose that, for a given $k \geq 2$, we have $P_{k-1} \cap \mathbb{Z}^2 = \Pi_{k-1}$. Let $\hat{x}_k^*$ be a k-th best point.*

(a) If $f(\hat{x}_k^*) > \hat{f}^*$, we can replace the point $\hat{x}_k^*$ by a feasible k-th best point $\hat{z}_k^*$ such that $\operatorname{conv}\{\Pi_{k-1}, \hat{z}_k^*\} \cap \mathbb{Z}^2 = \Pi_{k-1} \cup \{\hat{z}_k^*\}$ in $\mathscr{O}(1)$ operations.

(b) If $f(\hat{x}_k^*) = \hat{f}^*$, and if we have at our disposal the ν vertices of P_{k-1} ordered counterclockwise, we can construct such a point $\hat{z}_k^*$ in $\mathscr{O}(\nu \ln(B))$ operations.

Proof. Part (a). Suppose first that $f(\hat{x}_k^*) > \hat{f}^*$, and assume that we cannot set $\hat{z}_k^* := \hat{x}_k^*$, that is, that there exists $\hat{x} \in (P_k \cap \mathbb{Z}^2) \setminus \Pi_k$. Then $\hat{x} = \sum_{i=1}^k \lambda_i \hat{x}_i^*$ for some $\lambda_i \geq 0$ that sum up to 1. Note that $0 < \lambda_k < 1$, because $\hat{x} \notin P_{k-1} \cup \{\hat{x}_k^*\}$ by assumption, and that $f(\hat{x}) \geq f(\hat{x}_k^*)$. We deduce:

$$0 \leq f(\hat{x}) - f(\hat{x}_k^*) \leq \sum_{i=1}^{k} \lambda_i (f(\hat{x}_i^*) - f(\hat{x}_k^*)) \leq 0.$$

Thus $f(\hat{x}) = f(\hat{x}_k^*)$. Let $I := \{i : \lambda_i > 0\}$ and $Q_I := \operatorname{conv}\{\hat{x}_i^* : i \in I\}$, so that $\hat{x} \in \operatorname{relint} Q_I$. Observe that $|I| \geq 2$ and that f is constant on Q_I. Necessarily, Q_I is a segment. Indeed, if it were a two-dimensional set, we could consider the restriction of f on the line $\ell := \operatorname{aff}\{\hat{x}_1^*, \hat{x}\}$: it is constant on the open interval $\ell \cap \operatorname{int} Q_I$, but does not attain its minimum on it, contradicting the convexity of f. Let us now construct the point $\hat{z}_k^*$: it suffices to consider the closest point to $\hat{x}_k^*$ in $\operatorname{aff}\{Q_I\} \cap P_{k-1}$, say $\hat{x}_j^*$, and to take the integer point $\hat{z}_k^* \neq \hat{x}_j^*$ of $\operatorname{conv}\{\hat{x}_j^*, \hat{x}_k^*\}$ that is the closest to $\hat{x}_j^*$ (see Fig. 4).

Part (b). Suppose now that $f(\hat{x}_i^*) = \hat{f}^*$ for every $1 \leq i \leq k$, and define

$$\{\hat{y}_0^* \equiv \hat{y}_\nu^*, \hat{y}_1^*, \ldots, \hat{y}_{\nu-1}^*\} \subseteq \Pi_{k-1}$$

as the vertices of P_{k-1}, labeled counterclockwise. It is well-known that determining the convex hull of $P_{k-1} \cup \{\hat{x}_k^*\}$ costs $\mathscr{O}(\ln(\nu))$ operations. From these vertices, we

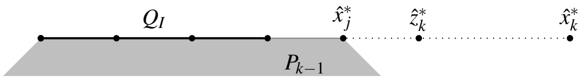

Fig. 4 Illustration of Part (a).

```
**Data**: $\hat{x}_k^*, \hat{y}_0^*, \hat{y}_1^*, \dots, \hat{y}_j^*$.
Set $i := 0$ and $x_k^*(0) := x_k^*$.
**while** $\det(\hat{x}_k^*(i) - \hat{y}_i^*, \hat{y}_{i+1}^* - \hat{y}_i^*) \geq 0$ **do**
    Set $\Delta_i := \mathrm{conv}\{\hat{x}_k^*(i), \hat{y}_i^*, \hat{y}_{i+1}^*\} \setminus \mathrm{aff}\{\hat{y}_i^*, \hat{y}_{i+1}^*\}$.
    Set $h_i$ a vector orthogonal to $\mathrm{aff}\{\hat{y}_i^*, \hat{y}_{i+1}^*\}$ such that $\langle h_i, \hat{x}_k^*(i) - \hat{y}_i^* \rangle > 0$.
    Set $\hat{x}_k^*(i+1) := \arg\min\{\langle h_i, \hat{x} \rangle : \hat{x} \in \Delta_i \cap \mathbb{Z}^2\}$.
    Set $i := i+1$.
**end**
Set $\hat{z}_k^* := \hat{x}_k^*(i)$.
```

Algorithm 4: Constructing a point $\hat{z}_k^*$ with $\mathrm{conv}\{\Pi_{k-1}, \hat{z}_k^*\} \cap \mathbb{Z}^2 = \Pi_{k-1} \cup \{\hat{z}_k^*\}$.

deduce the set $\{\hat{y}_i^* : i \in J\}$ of those points that are in the relative interior of that convex hull. Up to a renumbering of the $\hat{y}_l^*$'s, we have $J = \{1, 2, \dots, j-1\}$. We show below that Algorithm 4 constructs a satisfactory point $\hat{z}_k^*$.

We follow here the notation used in Algorithm 4. At every iteration i, the algorithm constructs from an integer point $\hat{x}_k^*(i)$ an integer point $\hat{x}_k^*(i+1)$, possibly identical to $\hat{x}_k^*(i)$. When the algorithm stops, after at most $j \leq \nu$ iterations, the point $\hat{z}_k^*$ we are looking for is, as we prove it below, the last $\hat{x}_k^*(i)$ we have constructed.

Define $T_l(i) := \mathrm{conv}\{\hat{x}_k^*(i), \hat{y}_l^*, \hat{y}_{l+1}^*\} \setminus P_{k-1}$ for $0 \leq l < j$ (see Fig. 5); the triangle Δ_i in Algorithm 4 corresponds to $T_i(i)$. Also, the vector h_i is orthogonal to the side $\mathrm{aff}\{\hat{y}_i^*, \hat{y}_{i+1}^*\}$ of the triangle $T_i(i)$.

At iteration i, the algorithm considers the triangle $T_i(i)$ if its signed area

$$\frac{1}{2}\det(\hat{x}_k^*(i) - \hat{y}_i^*, \hat{y}_{i+1}^* - \hat{y}_i^*)$$

is nonnegative, and finds a point $\hat{x}_k^*(i+1) \in T_i(i)$ such that $T_i(i+1)$ has only $\hat{x}_k^*(i+1)$ as integer point.

We prove by induction on $i \geq 1$ that $T_l(i)$ contains only $\hat{x}_k^*(i)$ as integer point whenever $l < i$. Consider the base case $i = 1$. By construction, the triangle $T_0(1)$ contains only $x_k^*(1)$ as integer point, for otherwise $x_k^*(1)$ would not minimize $\langle h_0, \hat{x} \rangle$ over $T_0(0) \cap \mathbb{Z}^2$.

Suppose now that the statement is true for i and let $l \leq i$. Let us verify that $\hat{x}_k^*(i+1)$ is the only integer in $T_l(i+1)$. We have:

$$\hat{x}_k^*(i+1) \in T_i(i) \subseteq \mathrm{conv}\{\hat{x}_k^*(i), \hat{y}_0^*, \dots, \hat{y}_{i+1}^*\} \setminus P_{k-1} = T_i(i) \cup \bigcup_{l=0}^{i-1} T_l(i).$$

This last equality represents a triangulation of the possibly non-convex polygon $\operatorname{conv}\{\hat{x}_k^*(i), \hat{y}_0^*, \ldots, \hat{y}_{i+1}^*\} \setminus P_{k-1}$. From the above inclusion, we deduce:

$$K := \operatorname{conv}\{\hat{x}_k^*(i+1), \hat{y}_0^*, \ldots, \hat{y}_{i+1}^*\} \setminus P_{k-1} \subseteq \operatorname{conv}\{\hat{x}_k^*(i), \hat{y}_0^*, \ldots, \hat{y}_{i+1}^*\} \setminus P_{k-1}.$$

As $T_l(i+1) \subseteq K$ for all $l \leq i$, the integers of $T_l(i+1)$ are either in $\bigcup_{l=0}^{i-1} T_l(i) \cap \mathbb{Z}^2$, which reduces to $\{\hat{x}_k^*(i)\}$ by induction hypothesis, or in $T_i(i)$. Since $\hat{x}_k^*(i) \in T_i(i)$, all the integers in $T_l(i+1)$ must be in $T_i(i)$. But $T_l(i+1) \cap T_i(i) \cap \mathbb{Z}^2 = \{\hat{x}_k^*(i+1)\}$ by construction of $\hat{x}_k^*(i+1)$. The induction step is proved.

It remains to take the largest value that i attains in the course of Algorithm 4 to finish the proof. We need to solve at most $\nu - 1$ two-dimensional integer linear problems over triangles to compute $\hat{x}_k^*$. As the data of these problems are integers bounded by B, the complexity of the minimization solver used to compute $x_k^*(i+1)$ at every step is bounded by $\mathscr{O}(\ln(B))$. The overall complexity of Algorithm 4 is thus bounded by $\mathscr{O}(\nu \ln(B))$. □

By Lemma 3, the k-th best point $\hat{x}_k^*$ can be assumed to be contained within $[-B,B]^2 \setminus \operatorname{conv}\{\hat{x}_1^*, \ldots, \hat{x}_{k-1}^*\}$. This property allows us to design a straightforward algorithm to compute this point. We first construct an inequality description of $\operatorname{conv}\{\hat{x}_1^*, \ldots, \hat{x}_{k-1}^*\}$, say $\langle a_i, x\rangle \leq b_i$ for $i \in I$ with $|I| < +\infty$. Then

$$[-B,B]^2 \setminus \operatorname{conv}\{\hat{x}_1^*, \ldots, \hat{x}_{k-1}^*\} = \bigcup_{i \in I} \{x \in [-B,B]^2 \mid \langle a_i, x\rangle > b_i\}.$$

As the feasible set is described as a union of simple convex sets, we could apply Algorithm 1 once for each of them. However, instead of choosing this straightforward approach one can do better: one can avoid treating each element of this disjunction separately by modifying Algorithm 3 appropriately.

Suppose first that $k = 2$. To find the second best point, we apply Algorithm 3 to the point $\hat{x}_1^*$ with the following minor modification: in Line 9, we replace $(\mathscr{P})$ with the integer linear problem $(\mathscr{P}') : \min\{\langle h, \hat{y}\rangle : \hat{y} \in \bar{T}_k \cap \mathbb{Z}^2,\ \langle h, \hat{y}\rangle \geq \langle h, \hat{x}_1\rangle + 1\}$, where $h \in \mathbb{Z}^2$ such that $\gcd(h_1, h_2) = 1$. This prevents the algorithm from returning $\hat{x}_1^*$ again.

Let $k \geq 3$. Let $\hat{y}_0^*, \ldots, \hat{y}_{\nu-1}^*, \hat{y}_\nu^* \equiv \hat{y}_0^*$ denote the vertices of P_{k-1}, ordered counterclockwise (they can be determined in $\mathscr{O}(k \ln(k))$ operations using the Graham Scan [Gra72]). Recall that the point we are looking for is not in P_{k-1}.

Let us call a triangle with a point $\hat{y}_i^*$ as vertex and with a segment of the boundary of $[-B,B]^2$ as opposite side a *search triangle* (see Fig. 7: every white triangle is a search triangle). The idea is to decompose $[-B,B]^2 \setminus P_{k-1}$ into search triangles, then to apply Algorithm 3 to these triangles instead of $(\operatorname{conv}\{x, F_t\})_{t=1}^4$.

For each $0 \leq i < \nu$, we define $H_i := \{y \in \mathbb{R}^2 : \det(y - \hat{y}_i^*, \hat{y}_{i+1}^* - \hat{y}_i^*) \geq 0\}$, so that $H_i \cap P_{k-1} = \operatorname{conv}\{\hat{y}_i^*, \hat{y}_{i+1}^*\}$. Consider the regions $R_i := ([-B,B]^2 \cap H_i) \setminus \operatorname{int} H_{i-1}$. Note that R_i contains only $\hat{y}_i^*$ and $\hat{y}_{i+1}^*$ as vertices of P_{k-1}. Also, at most four of the R_i's are no search triangles. If R_i is such, we triangulate it into (at least two) search triangles by inserting chords from $\hat{y}_i^*$ to the appropriate vertices of $[-B,B]^2$.

Note that a search triangle can contain two or more integer points of P_{k-1}. In order to prevent us from outputting one of those, we need to perturb the search

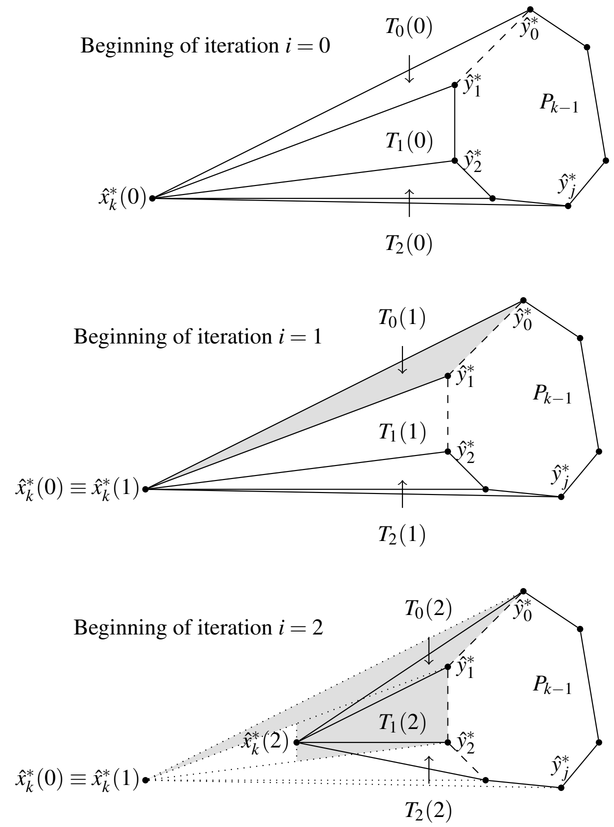

Fig. 5 Constructing P_k from P_{k-1}: first iterations of Algorithm 4. The point $\hat{x}_k^*(1)$ is the same as $\hat{x}_k^*(0)$ because $T_0(0)$ has no other integer point than $\hat{x}_k^*(0)$. The gray areas are, as the algorithm progresses, regions where we have established that they do not contain any integer point.

triangles slightly before using them in Algorithm 3. Let $T = \operatorname{conv}\{\hat{y}_i^*, b_1, b_2\}$ be one of the search triangles, with b_1, b_2 being points of the boundary of $[-B,B]^2$. The triangle T might contain $\hat{y}_{i+1}^*$, say $\hat{y}_{i+1}^* \in \operatorname{conv}\{\hat{y}_i^*, b_1\}$, a point we need to exclude from T. We modify b_1 slightly by replacing it with $(1-\varepsilon)b_1 + \varepsilon b_2$ for an appropriate positive $\varepsilon > 0$ whose encoding length is $\mathscr{O}(\ln(B))$.

So, we apply Algorithm 3 with all these modified search triangles instead of $\operatorname{conv}\{x, F_1\}, \ldots, \operatorname{conv}\{x, F_4\}$. A simple modification of Line 9 allows us to avoid the point $\hat{y}_i^*$ for $\hat{z}$: we just need to replace the linear integer problem $(\mathscr{P})$ with $\min\{\langle h, \hat{y}\rangle : \hat{y} \in \bar{T}_k \cap \mathbb{Z}^2,\ \langle h, \hat{y}\rangle \geq \langle h, \hat{y}_i^*\rangle + 1\}$, where $h \in \mathbb{Z}^2$ such that $\gcd(h_1, h_2) = 1$. Then, among the feasible integer points found, we return the point with smallest objective value.

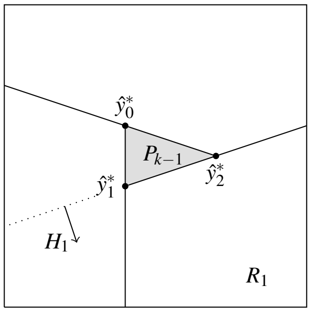

Fig. 6 Triangulation step 1.

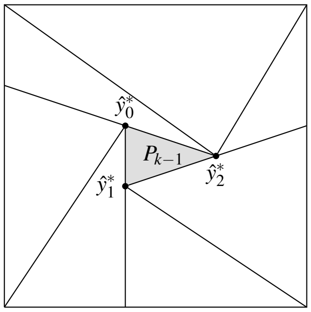

Fig. 7 Triangulation step 2.

Corollary 2. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ and $g_i : \mathbb{R}^2 \to \mathbb{R}$ with $i = 1, \dots, m$ be convex functions. Let $\hat{x}_1^*, \dots, \hat{x}_{k-1}^*$ be the $k-1$ best points for $\min\{f(\hat{x}) : \hat{x} \in S \cap [-B,B]^2 \cap \mathbb{Z}^2\}$. Then, in a number of evaluations of f and $g_1, \dots, g_m$ that is polynomial in $\ln(B)$ and in k, one can either find*

(a) a k-th best point, $\hat{x}_k^*$, or
(b) show that there is no such point.

4 Extensions and applications to the general setting

In this section, we extend our algorithm for solving two-dimensional integer convex optimization problems in order to solve more general mixed-integer convex problems. The first extension concerns mixed-integer convex problems with two integer variables and d continuous variables. For those, we first need results about problems with only one integer variable. We derive these results in Subsection 4.1 where we propose a variant of the well-known golden search method that deals with convex functions whose value is only known approximately. To the best of our knowledge, this variant is new.

In Subsection 4.2, we build an efficient method for solving mixed-integer convex problems with two integer and d continuous variables and propose an extension of Corollary 2. This result itself will be used as a subroutine to design a finite-time algorithm for mixed-integer convex problems in n integer and d continuous variables in Subsection 4.3.

In this section, the problem of interest is (1):

$$\min\{f(\hat{x},y) : g_i(\hat{x},y) \le 0 \text{ for } 1 \le i \le m, (\hat{x},y) \in \mathbb{Z}^n \times \mathbb{R}^d\}$$

with a few mild simplifying assumptions. We define the function

$$g : \mathbb{R}^n \to \mathbb{R}, \quad x \mapsto g(x) := \min_{y \in \mathbb{R}^d} \max_{1 \le i \le m} g_i(x,y).$$

We assume that this minimization in y has a solution for every $x \in \mathbb{R}^n$, so as to make the function g convex. Let $S := \{(x,y) \in \mathbb{R}^{n+d} : g_i(x,y) \leq 0 \text{ for } 1 \leq i \leq m\}$. We assume that the function f has a finite spread $\max\{f(x,y) - f(x',y') : (x,y),(x',y') \in S\}$ on S and that we know an upper bound V_f on that spread. Observe that, by Lipschitz continuity of f and the assumption that we optimize over $[-B,B]^n$, it follows $V_f \leq 2\sqrt{n}BL$. Finally, we assume that the partial minimization function:

$$\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, \quad x \mapsto \phi(x) := \min\{f(x,y) : (x,y) \in S\}$$

is convex. As for the function g, this property can be achieved e.g. if for every $x \in \mathbb{R}^n$ for which $g(x) \leq 0$ there exists a point y such that $(x,y) \in S$ and $\phi(x) = f(x,y)$.

Our approach is based on the following well-known identity:

$$\min\{f(\hat{x},y) : (\hat{x},y) \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)\} = \min\{\phi(\hat{x}) : g(\hat{x}) \leq 0,\ \hat{x} \in \mathbb{Z}^n\}.$$

For instance, when $n = 2$, we can use the techniques developed in the previous section on ϕ to implement the improvement oracle for f. However, we cannot presume to know exactly the value of ϕ, as it results from a minimization problem. We merely assume that, for a known accuracy $\gamma > 0$ and for every $x \in \operatorname{dom}\phi$ we can determine a point y_x such that $(x,y_x) \in S$ and $f(x,y_x) - \gamma \leq \phi(x) \leq f(x,y_x)$. Determining y_x can be, on its own, a non-trivial optimization problem. Nevertheless, it is a convex problem for which we can use the whole machinery of standard Convex Programming (see e.g. [**?**, **?**, Nes04] and references therein.).

Since we do not have access to exact values of ϕ, we cannot hope for an exact oracle for the function ϕ, let alone for f. The impact of the accuracy γ on the accuracy of the oracle is analyzed in the next subsections.

4.1 Mixed-integer convex problems with one integer variable

The Algorithm 3 uses as indispensable tools the bisection method for solving two types of problems: minimizing a convex function over the integers of an interval, and finding, in a given interval, a point that belongs to a level set of a convex function. In this subsection, we show how to adapt the bisection methods for mixed-integer problems. It is well-known that the bisection method is the fastest for minimizing univariate convex functions over a finite segment ([**?**, Chapter 1]).

Let $a,b \in \mathbb{R}$, $a < b$, and $\varphi : [a,b] \to \mathbb{R}$ be a convex function to minimize on $[a,b]$ and/or on the integers of $[a,b]$, such as the function ϕ in the preamble of this Section 4 when $n = 1$. Assume that, for every $t \in [a,b]$, we know a number $\tilde{\varphi}(t) \in [\varphi(t), \varphi(t) + \gamma]$. In order to simplify the notation, we scale the problem so that $[a,b] \equiv [0,1]$. The integers of $\operatorname{aff}\{a,b\}$ are scaled to a set of points of the form $t_0 + \tau\mathbb{Z}$ for a $\tau > 0$. Of course, the spread of the function φ does not change, but its Lipschitz constant does, and achieving the accuracy γ in its evaluation must be reinterpreted accordingly.

In the sequel of this section, we fix $0 \leq \lambda_0 < \lambda_1 \leq 1$.

Lemma 4. *Under our assumptions, the following statements hold.*

(a) If $\tilde{\varphi}(\lambda_0) \leq \tilde{\varphi}(\lambda_1) - \gamma$*, then* $\varphi(\lambda) \geq \tilde{\varphi}(\lambda_0)$ *for all* $\lambda \in [\lambda_1, 1]$.
(b) If $\tilde{\varphi}(\lambda_0) \geq \tilde{\varphi}(\lambda_1) + \gamma$*, then* $\varphi(\lambda) \geq \tilde{\varphi}(\lambda_1)$ *for all* $\lambda \in [0, \lambda_0]$.

Proof. We only prove Part (a) as the proof of Part (b) is symmetric. Thus, let us assume that $\tilde{\varphi}(\lambda_0) \leq \tilde{\varphi}(\lambda_1) - \gamma$. Then there exists $0 < \mu \leq 1$ for which $\lambda_1 = \mu\lambda + (1-\mu)\lambda_0$. Convexity of φ allows us to write:

$$\tilde{\varphi}(\lambda_0) \leq \tilde{\varphi}(\lambda_1) - \gamma \leq \varphi(\lambda_1) \leq \mu\varphi(\lambda) + (1-\mu)\varphi(\lambda_0) \leq \mu\varphi(\lambda) + (1-\mu)\tilde{\varphi}(\lambda_0),$$

implying $\tilde{\varphi}(\lambda_0) \leq \varphi(\lambda)$ as $\mu > 0$. Fig. 8 illustrates the proof graphically. □

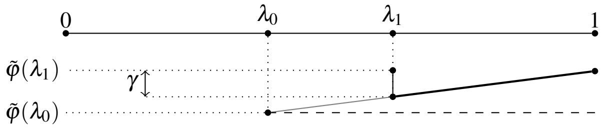

Fig. 8 Lemma 4: the bold line represents a lower bound on φ in Part (a).

If one of the conditions in Lemma 4 is satisfied, we can remove from the interval $[0,1]$ either $[0,\lambda_0[$ or $]\lambda_1,1]$. To have a symmetric effect of the algorithm in either case, we set $\lambda_1 := 1 - \lambda_0$, forcing λ_0 to be smaller than $\frac{1}{2}$. In order to recycle our work from iteration to iteration, we choose $\lambda_1 := \frac{1}{2}(\sqrt{5}-1)$, as in the golden search method: if we can eliminate, say, the interval $]\lambda_1, 1]$ from $[0,1]$, we will have to compute in the next iteration step an approximate value of the objective function at $\lambda_0\lambda_1$ and λ_1^2. The latter happens to equal λ_0 when $\lambda_1 = \frac{1}{2}(\sqrt{5}-1)$.

It remains to define a strategy when neither of the conditions in Lemma 4 is satisfied. In the lemma below, we use the values for λ_0, λ_1 chosen above.

Lemma 5. *Assume that* $\tilde{\varphi}(\lambda_1) - \gamma < \tilde{\varphi}(\lambda_0) < \tilde{\varphi}(\lambda_1) + \gamma$*. We define:*

$$\begin{aligned}\lambda_{0+} &:= (1-\lambda_0)\cdot\lambda_0 + \lambda_0\cdot\lambda_1 = 2\lambda_0\lambda_1,\\ \lambda_{1+} &:= (1-\lambda_1)\cdot\lambda_0 + \lambda_1\cdot\lambda_1 = 1 - 2\lambda_0\lambda_1.\end{aligned}$$

If $\min\{\tilde{\varphi}(\lambda_{0+}), \tilde{\varphi}(\lambda_{1+})\} \leq \min\{\tilde{\varphi}(\lambda_0) - \gamma, \tilde{\varphi}(\lambda_1) - \gamma\}$*, then* $\varphi(t) \geq \min\{\tilde{\varphi}(\lambda_{0+}), \tilde{\varphi}(\lambda_{1+})\}$ *for all* $t \in [0,1] \setminus [\lambda_0, \lambda_1]$*. Otherwise, it holds that* $\min\{\tilde{\varphi}(\lambda_0), \tilde{\varphi}(\lambda_1)\} \leq \min\{\varphi(t) : t \in [0,1]\} + (\kappa - 1)\gamma$*, where* $\kappa := \frac{2}{\lambda_0} \approx 5.236$.

Proof. The first conclusion follows immediately from Lemma 4. The second situation involves a tedious enumeration, summarized in Fig. 9. We assume, without loss of generality, that $\tilde{\varphi}(\lambda_0) \leq \tilde{\varphi}(\lambda_1)$. The bold lines in Fig. 9 represent a lower bound on the value of the function φ. We show below how this lower bound is constructed and determine its lowest point. In fact, this lower bound results from six applications

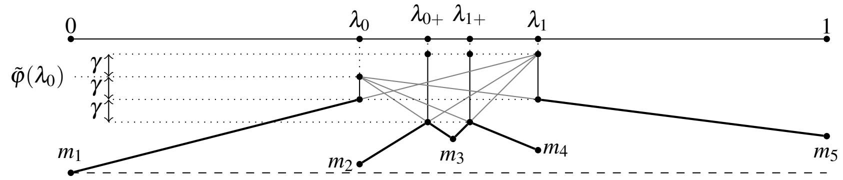

Fig. 9 Approximate bisection: bold lines represent a lower bound on φ in the termination case.

of a simple generic inequality (9) that we establish below, before showing how we can particularize it to different segments of the interval $[0,1]$.

Let $0 < t < 1$ and let $u, v \in \{\lambda_0, \lambda_{0+}, \lambda_{1+}, \lambda_1\}$. Suppose that we can write $v = \mu t + (1-\mu)u$ for a $\mu \in]\mu_0, 1]$ with $\mu_0 > 0$. If we can find constants $\gamma_-, \gamma_+ \geq 0$ that satisfy

$$\varphi(v) + \gamma_+ \geq \tilde{\varphi}(\lambda_0) \geq \varphi(u) - \gamma_-$$

then we can infer:

$$\mu\varphi(t) + (1-\mu)(\tilde{\varphi}(\lambda_0) + \gamma_-) \geq \mu\varphi(t) + (1-\mu)\varphi(u) \geq \varphi(v) \geq \tilde{\varphi}(\lambda_0) - \gamma_+,$$

and thus:

$$\varphi(t) - \tilde{\varphi}(\lambda_0) \geq \gamma_- - \frac{\gamma_+ + \gamma_-}{\mu} \geq \gamma_- - \frac{\gamma_+ + \gamma_-}{\mu_0}. \tag{9}$$

1. If $t \in]0, \lambda_0]$, we can take $u := \lambda_1$ and $v := \lambda_0$, giving $\mu_0 = 1 - \frac{\lambda_0}{\lambda_1} = \lambda_0$. Then $\gamma_- = \gamma_+ = \gamma$, and $\varphi(t) - \tilde{\varphi}(\lambda_0) \geq -\gamma(\frac{2}{\lambda_0} - 1)$.
2. If $t \in]\lambda_1, 1[$, we choose $u := \lambda_0$ and $v := \lambda_1$, and by symmetry with the previous case we obtain $\mu_0 = \lambda_0$. Now, $\gamma_- = 0$ and $\gamma_+ = \gamma$, yielding a higher bound than in the previous case.
3. Suppose $t \in]\lambda_0, \lambda_{0+}]$. Then with $u := \lambda_1$ and $v := \lambda_{0+}$, we get $\mu_0 = \frac{\lambda_1 - \lambda_{0+}}{\lambda_1 - \lambda_0} = \lambda_1$, $\gamma_- = \gamma$, $\gamma_+ = 2\gamma$, giving as lower bound $-\gamma(\frac{3}{\lambda_1} - 1)$, which is higher than the first one we have obtained.
4. Symmetrically, let us consider $t \in]\lambda_{1+}, \lambda_1]$. With $u := \lambda_0$ and $v := \lambda_{1+}$, we obtain also $\mu_0 = \lambda_1$. As $\gamma_- = 0$ and $\gamma_+ = 2\gamma$, the lower bound we get is larger than the one in the previous item.
5. Set $\lambda' := \frac{1}{5}(2\lambda_{0+} + 3\lambda_{1+})$. If $t \in]\lambda_{0+}, \lambda']$, we can use $u := \lambda_0$ and $v := \lambda_{0+}$, so that $\mu_0 = \frac{\lambda_{0+} - \lambda_0}{\lambda' - \lambda_0} = 5\lambda_0^2$, $\gamma_- = 0$, and $\gamma_+ = 2\gamma$. Thus, the lower bound is evaluated as $-\frac{2\gamma}{5\lambda_0^2}$, which is higher than any of the bounds we have obtained so far.
6. Finally, if $t \in]\lambda', \lambda_{1+}]$, we take $u := \lambda_1$ and $v := \lambda_{1+}$, so that $\gamma_- = \gamma$, $\gamma_+ = 2\gamma$, and $\mu_0 = \frac{\lambda_1 - \lambda_{1+}}{\lambda_1 - \lambda'} = \frac{5\lambda_0}{2+\lambda_0}$. Hence, we get $-\gamma(\frac{3(2+\lambda_0)}{5\lambda_0} - 1) = -\frac{2\gamma}{5\lambda_0^2}$ for the lower bound, just as in the previous item.

So, the lower bound for $\varphi(t) - \tilde{\varphi}(\lambda_0)$ on $[0,1]$ can be estimated as $-\gamma(\frac{2}{\lambda_0} - 1) \approx -4.236\gamma$. □

In the proof of the following proposition, we present an algorithm that returns a point $x \in [0,1]$ whose function value $\varphi(x)$ is close to $\min\{\varphi(t) : t \in [0,1]\}$.

Proposition 3. *There exists an algorithm that finds a point* $x \in [0,1]$ *for which* $\tilde{\varphi}(x) - (\kappa-1)\gamma \leq \min\{\varphi(t) : t \in [0,1]\} \leq \varphi(x)$ *in at most* $2 + \left\lceil \ln\left(\frac{(\kappa-1)\gamma}{V_\varphi}\right) / \ln(\lambda_1) \right\rceil$ *evaluations of* $\tilde{\varphi}$*, where* V_φ *is the spread of* φ *on* $[0,1]$.

Proof. We start with the interval $[0,1]$ and by evaluating $\tilde{\varphi}$ at λ_0 and λ_1. If one of the two conditions in Lemma 4 is satisfied, we can shrink the interval by a factor of $\lambda_0 \approx 38\%$ since it suffices to continue either with the interval $[0,\lambda_1]$ or with $[\lambda_0,1]$. If not, then Lemma 5 applies: if the first condition stated in Lemma 5 is met, then it suffices to continue with the interval $[\lambda_0,\lambda_1]$ so as to shrink the starting interval by a factor of $2\lambda_0 \approx 76\%$. Otherwise, any $x \in [\lambda_0,\lambda_1]$ satisfies the requirement of the lemma and we can stop the algorithm. Therefore, either the algorithm stops or we shrink the starting interval by a factor of at least λ_0. Iterating this procedure, it follows that — if the algorithm does not stop — at every step the length of the remaining interval is at most λ_1 times the length of the previous interval. Moreover, by the choice of λ_0, the function $\tilde{\varphi}$ is evaluated in two points at the first step, and in only one point as from the second step in the algorithm. So, at iteration k, we have performed at most $2+k$ evaluations of $\tilde{\varphi}$.

By construction, the minimum t^* of φ lies in the remaining interval I_k of iteration k. Also, the value of φ outside I_k is higher than the best value found so far, say $\tilde{\varphi}(\bar{t}_k)$. Finally, the size of I_k is bounded from above by λ_1^k. Consider now the segment $I(\lambda) := (1-\lambda)t^* + \lambda[0,1]$, of size λ. Observe that for every λ such that $1 \geq \lambda > \lambda_1^k$, the interval $I(\lambda)$ contains a point that is not in I_k. Therefore,

$$\begin{aligned}\tilde{\varphi}(\bar{t}_k) \leq \max\{\varphi(t) : t \in I(\lambda)\} &\leq (1-\lambda)\varphi(t^*) + \lambda \max\{\varphi(t) : t \in [0,1]\} \\ &\leq (1-\lambda)\varphi(t^*) + \lambda(V_\varphi + \varphi(t^*)).\end{aligned}$$

Hence $\tilde{\varphi}(\bar{t}_k) - \varphi(t^*) \leq \lambda V_\varphi$, and, by taking λ arbitrarily close to λ_1^k, we get $\tilde{\varphi}(\bar{t}_k) - \varphi(t^*) \leq \lambda_1^k V_\varphi$. If the algorithm does not end prematurely, we need at most $\left\lceil \ln\left(\frac{(\kappa-1)\gamma}{V_\varphi}\right) / \ln(\lambda_1) \right\rceil$ iterations to make $\lambda_1^k V_\varphi$ smaller than $(\kappa-1)\gamma$. □

Remark 3. If we content ourselves with a coarser precision $\eta \geq (\kappa-1)\gamma$, we merely need $\mathcal{O}(\ln(V_\varphi/\eta))$ evaluations of $\tilde{\varphi}$. ◇

It is now easy to extend this procedure to minimize a convex function approximately over the integers of an interval $[a,b]$, or, using our simplifying scaling, over $(t_0 + \tau\mathbb{Z}) \cap [0,1]$ for given $t_0 \in \mathbb{R}$ and $\tau > 0$.

Proposition 4. *There exists an algorithm that finds a point* $\hat{x} \in (t_0 + \tau\mathbb{Z}) \cap [0,1]$ *for which:*

$$\tilde{\varphi}(\hat{x}) - \kappa\gamma \leq \min\{\varphi(\hat{t}) : \hat{t} \in (t_0 + \tau\mathbb{Z}) \cap [0,1]\} \leq \varphi(\hat{x})$$

in less than

$$\min\left\{4 + \left\lceil \frac{\ln((\kappa-1)\gamma/V_\varphi)}{\ln(\lambda_1)} \right\rceil, 5 + \left\lceil \frac{\ln(\tau)}{\ln(\lambda_1)} \right\rceil \right\}$$

evaluations of $\tilde{\varphi}$, where V_φ is the spread of φ on $[0,1]$.

Proof. We denote in this proof the points in $(t_0+\tau\mathbb{Z})$ as *scaled integers*. To avoid a trivial situation, we assume that $[0,1]$ contains at least two such scaled integers.

Let us use the approximate bisection method described in the proof of Proposition 3 until the remaining interval has a size smaller than τ, so that it contains at most one scaled integer. Two possibilities arise: either the algorithm indeed finds such a small interval I_k, or it finishes prematurely, with a remaining interval I_k larger than τ.

In the first case, which requires at most $2+\lceil \ln(\tau)/\ln(\lambda_1)\rceil$ evaluations of $\tilde{\varphi}$, we know that I_k contains the continuous minimizer of φ. Hence, the actual minimizer of φ over $(t_0+\tau\mathbb{Z})\cap[0,1]$ is among at most three scaled integers, namely the possible scaled integer in I_k, and, at each side of I_k, the possible scaled integers that are the closest to I_k. By convexity of φ, the best of these three points, say $\hat{x}$, satisfies $\tilde{\varphi}(\hat{x})-\gamma\leq\varphi(\hat{x})=\min\{\varphi(\hat{t}):\hat{t}\in(t_0+\tau\mathbb{Z})\cap[0,1]\}$.

In the second case, we have an interval $I_k\subseteq[0,1]$ and a point $\bar{t}_k$ that fulfill $\tilde{\varphi}(\bar{t}_k)\leq\min\{\varphi(t):t\in[0,1]\}+(\kappa-1)\gamma$, which was determined within at most $2+\left\lceil\frac{\ln((\kappa-1)\gamma/V_\varphi)}{\ln(\lambda_1)}\right\rceil$ evaluations of $\tilde{\varphi}$. Consider the two scaled integers $\hat{t}_-$ and $\hat{t}_+$ that are the closest from $\bar{t}_k$. One of these two points constitutes an acceptable output for our algorithm. Indeed, suppose first that $\min\{\tilde{\varphi}(\hat{t}_-),\tilde{\varphi}(\hat{t}_+)\}\leq\tilde{\varphi}(\bar{t}_k)+\gamma$. Then:

$$\min\{\tilde{\varphi}(\hat{t}_-),\tilde{\varphi}(\hat{t}_+)\}\leq\tilde{\varphi}(\bar{t}_k)+\gamma\leq\min\{\varphi(t):t\in[0,1]\}+\kappa\gamma,$$

and we are done. Suppose that $\min\{\tilde{\varphi}(\hat{t}_-),\tilde{\varphi}(\hat{t}_+)\}>\tilde{\varphi}(\bar{t}_k)+\gamma$ and that there exists a scaled integer $\hat{t}$ with $\varphi(\hat{t})<\min\{\varphi(\hat{t}_-),\varphi(\hat{t}_+)\}$. Without loss of generality, let $\hat{t}_-\in\operatorname{conv}\{\hat{t},\bar{t}_k\}$, that is $\hat{t}_-=\lambda\hat{t}+(1-\lambda)\bar{t}_k$, with $0\leq\lambda<1$. We have by convexity of φ:

$$\varphi(\hat{t}_-)\leq\lambda\varphi(\hat{t})+(1-\lambda)\varphi(\bar{t}_k)<\lambda\varphi(\hat{t}_-)+(1-\lambda)(\tilde{\varphi}(\hat{t}_-)-\gamma),$$

which is a contradiction because $\lambda<1$ and $\tilde{\varphi}(\hat{t}_-)-\gamma\leq\varphi(\hat{t}_-)$. So, it follows that $\varphi(\hat{t})\geq\min\{\varphi(\hat{t}_-),\varphi(\hat{t}_+)\}$ for every $\hat{t}\in(t_0+\tau\mathbb{Z})\cap[0,1]$, proving the statement. □

In the the following we extend the above results to the problem $\min\{\varphi(t):t\in[0,1],\ g(t)\leq 0\}$, where $g:[0,1]\to\mathbb{R}$ is a convex function with a known spread V_g. In the case that we have access to exact values of g, an approach for attacking the problem would be the following: we first determine whether there exists an element $\bar{t}\in[0,1]$ with $g(\bar{t})\leq 0$. If $\bar{t}$ exists, we determine the exact bounds t_- and t_+ of the interval $\{t\in[0,1],g(t)\leq 0\}$. Then we minimize the function f over $[t_-,t_+]$.

The situation where we do not have access to exact values of g or where we cannot determine the feasible interval $[t_-,t_+]$ induces some technical complications. We shall not investigate them in this paper, except in the remaining of this subsection in order to appreciate the modification our method needs in that situation: let us assume, that we have only access to a value $\tilde{g}(t)\in[g(t),g(t)+\gamma]$. In order to ensure that the constraint g is well-posed we make an additional assumption: either $\{t\in[0,1]:|g(t)|\leq\gamma\}$ is empty, or the quantity $\min\{|g'(t)|:g'(t)\in\partial g(t),|g(t)|\leq\gamma\}$ is non-zero, and even reasonably large. This ensures that the (possibly empty) 0-level set of g is known with enough accuracy. We denote by $\theta > 0$ a lower bound on this minimum, and for simplicity assume that $\theta = 2^N \gamma$ for a suitable $N \in \mathbb{N}$.

Our strategy proceeds as follows. First we determine whether there exists a point $\bar{t} \in [0,1]$ for which $g(\bar{t}) < 0$ by applying the minimization procedure described in Proposition 3. If this procedure only returns nonnegative values, we can conclude after at most $2 + \lceil \ln((\kappa-1)\gamma/V_g)/\ln(\lambda_1) \rceil$ evaluations of $\tilde{g}$ that $g(t) \geq -(\kappa-1)\gamma$, in which case we declare that we could not locate any feasible point in $[0,1]$.

Otherwise, if we find a point $\bar{t} \in [0,1]$ with $\tilde{g}(\bar{t}) < 0$, we continue and compute approximate bounds t_- and t_+ of the interval $\{t \in [0,1], g(t) \leq 0\}$. For that, we assume $\tilde{g}(0), \tilde{g}(1) \geq 0$. By symmetry, we only describe how to construct t_- such that $\tilde{g}(t_-) \leq 0$ and $g(t_- - \eta) \geq 0$ for an $\eta > 0$ reasonably small. Note that $g(t) \leq 0$ on $[t_-, \bar{t}]$ by convexity of g.

In order to compute t_-, we adapt the standard bisection method for finding a root of a function. Note that the function $\tilde{g}$ might not have any root as it might not be continuous. Our adapted method constructs a decreasing sequence of intervals $[a_k, b_k]$ such that $\tilde{g}(a_k) > 0$, $\tilde{g}(b_k) \leq 0$, and $b_{k+1} - a_{k+1} = \frac{1}{2}(b_k - a_k)$. If $\tilde{g}(a_k) > \gamma$, we know that g is positive on $[0, a_k]$, and we know that there is a root of g on $[a_k, b_k]$. Otherwise, if $0 < \tilde{g}(a_k) \leq \gamma$ and that the interval $[a_k, b_k]$ has a length larger or equal to $\frac{\gamma}{\theta}$. Given the form of θ, we know that $k \leq N$. We claim that for every $0 \leq t \leq \min\{0, a_k - \frac{\gamma}{\theta}\}$ we have $g(t) \geq 0$, so that we can take $\eta := 2\frac{\gamma}{\theta}$ and $t_- := b_N$. Indeed, assume that $g'(a_k) \geq \theta$, then

$$\tilde{g}(b_k) \geq g(b_k) \geq g(a_k) + g'(a_k)(b_k - a_k) > -\gamma + \theta \cdot \frac{\gamma}{\theta} \geq 0$$

giving a contradiction, so we must have $g'(a_k) \leq -\theta$. We can exclude the case where t can only be 0. As claimed, we have

$$g(t) \geq g(a_k) + g'(a_k)(t - a_k) \geq -\gamma + \theta(a_k - t) \geq 0$$

as $\frac{\gamma}{\theta} \leq a_k - t$. This takes $\lceil \ln(\frac{\gamma}{\theta})/\ln(\frac{1}{2}) \rceil$ evaluations of $\tilde{g}$.

Summarizing this, we just sketched the proof of the following corollary.

Corollary 3. *There exists an algorithm that solves approximately* $\min\{\varphi(t) : t \in [0,1], g(t) \leq 0\}$*, in the sense that it finds, if they exist, three points* $0 \leq t_- \leq x \leq t_+ \leq 1$ *with:*

(a) $g(t) \leq \tilde{g}(t) \leq 0$ *for every* $t \in [t_-, t_+]$,
(b) if $t_- \geq 2\frac{\gamma}{\theta}$, *then* $g(t) \geq 0$ *for every* $t \in [0, t_- - 2\frac{\gamma}{\theta}]$,
(c) if $t_+ \leq 1 - 2\frac{\gamma}{\theta}$, *then* $g(t) \geq 0$ *for every* $t \in [t_+ + 2\frac{\gamma}{\theta}, 1]$,
(d) $\tilde{\varphi}(x) \leq \min\{\varphi(t) : t \in [t_-, t_+],\ g(t) \leq 0\} + (\kappa - 1)\gamma$

within at most $3 + \left\lceil \frac{\ln((\kappa-1)\gamma/V_g)}{\ln(\lambda_1)} \right\rceil + 2\left\lceil \frac{\ln(\gamma/\theta)}{\ln(1/2)} \right\rceil$ *evaluations of* $\tilde{g}$ *and at most* $2 + \left\lceil \frac{\ln((\kappa-1)\gamma/V_\varphi)}{\ln(\lambda_1)} \right\rceil$ *evaluations of* $\tilde{\varphi}$.

As stressed before above, we assume from now on that we can compute exactly the roots of the function g on a given interval, so that the segment $[t_-, t_+]$ in Corollary 3 is precisely our feasible set. This situation occurs e.g. in mixed-integer convex optimization with one integer variable when the feasible set $S \subset \mathbb{R} \times \mathbb{R}^d$ is a polytope.

Remark 4. In order to solve problem (1) with one integer variable, we can extend Proposition 4 to implement the improvement oracle $\mathrm{O}_{0,\kappa\gamma}$. We need three assumptions: first, $S \subseteq [a,b] \times \mathbb{R}^d$ with $a < b$; second, f has a finite spread on the feasible set; and third we can minimize $f(x,y)$ with $(x,y) \in S$ and x fixed up to an accuracy γ. That is, we have access to a value $\tilde{\varphi}(x) \in [\varphi(x), \varphi(x) + \gamma]$ with $\varphi(x) := \min\{f(x,y) : (x,y) \in S\}$ being convex.

Given a feasible query point $(x,y) \in [a,b] \times \mathbb{R}^d$, we can determine correctly that there is no point $(\hat{x}, \bar{y}) \in ((t_0 + \tau\mathbb{Z}) \cap [0,1]) \times \mathbb{R}^d$ for which $f(\hat{x}, \bar{y}) \leq f(x,y)$, provided that the output $\hat{x}$ of our approximate bisection method for integers given in Proposition 4 satisfies $\tilde{\varphi}(\hat{x}) - \kappa\gamma > f(x,y)$. Otherwise, we can determine a point $(\hat{x}, \bar{y})$ for which $f(\hat{x}, \bar{y}) \leq f(x,y) + \kappa\gamma$. Note that this oracle cannot report **a** and **b** simultaneously. ◇

4.2 Mixed-integer convex problems with two integer variables

We could use the Mirror-Descent Method in Algorithm 1 to solve the generic problem (1) when $n = 2$ with $z \mapsto \frac{1}{2}||z||_2^2$ as function V, so that $\sigma = 1$ and $M = \frac{1}{2}\mathrm{diam}(S)^2$, where $\mathrm{diam}(S) = \max\{||z - z'||_2 : z, z' \in S\}$. According to (5), the worst-case number of iterations is bounded by a multiple of $L\sqrt{M/\sigma} = \mathcal{O}(L\,\mathrm{diam}(S))$, where L is the Lipschitz constant of f. As $V_f \leq L\,\mathrm{diam}(S)$, the resulting algorithm would have a worst-case complexity of $\Omega(V_f)$.

We improve this straightforward approach with a variant of Algorithm 3, whose complexity is polynomial in $\ln(V_f)$. This variant takes into account the fact that we do not have access to exact values of the partial minimization function ϕ defined in the preamble of this section.

Proposition 5. *Suppose that we can determine, for every $x \in \mathbb{R}^n$ with $g(x) \leq 0$, a point $y_x \in \mathbb{R}^d$ satisfying $f(x,y_x) - \gamma \leq \min\{f(x,y) : (x,y) \in S\}$. Then we can implement the oracle $\mathrm{O}_{0,\kappa\gamma}$ such that for every $(x,y) \in S$ it takes a number of evaluations of f that is polynomial in $\ln(V_f/\gamma)$.*

Proof. We adapt the algorithm described in the proof of Theorem 2 for the function $\phi(x) := \min\{f(x,y) : (x,y) \in S\}$, which we only know approximately. Its available approximation is denoted by $\tilde{\phi}(x) := f(x,y_x) \in [\phi(x), \phi(x) + \gamma]$.

Let $(x,y) \in S$ be the query point and let us describe the changes that the algorithm in Theorem 2 requires. We borrow the notation from the proof of Theorem 2.

The one-dimensional integer minimization problems which arise in the course of the algorithm require the use of our approximate bisection method for integers in Proposition 4. This bisection procedure detects, if it exists, a point $\hat{x}$ on the line of interest for which $\tilde{\phi}(\hat{x}) = f(\hat{x}, y_{\hat{x}}) \leq f(x,y) + \kappa\gamma$ and we are done. Or it reports correctly that there is no integer $\hat{x}$ on the line of interest with $\phi(\hat{x}) \leq f(x,y)$.

In **Case 2**, we would need to check whether $\phi(\hat{z}) \leq f(x,y)$. In view of our accuracy requirement, we only need to check $\tilde{\phi}(\hat{z}) \leq f(x,y) + \kappa\gamma$.

We also need to verify whether the line H intersects the level set $\{x \in \mathbb{R}^2 \mid \phi(x) \leq f(x,y)\}$. We use the following approximate version:

"**check whether** there is a $v \in \operatorname{conv}\{z_0, \hat{z}\}$ for which $\tilde{\phi}(v) < f(x,y) + (\kappa - 1)\gamma$",

which can be verified using Proposition 3. If such a point v exists, the convexity of ϕ forbids any $w \in \operatorname{conv}\{\hat{z}, z_1\}$ to satisfy $\phi(w) \leq f(x,y)$, for otherwise:

$$\tilde{\phi}(\hat{z}) \leq \phi(\hat{z}) + \gamma \leq \max\{\phi(v), \phi(w)\} + \gamma \leq \max\{\tilde{\phi}(v), \tilde{\phi}(w)\} + \gamma < f(x,y) + \kappa\gamma,$$

a contradiction. Now, if such a point v does not exist, we perform the same test on $\operatorname{conv}\{\hat{z}, z_1\}$. We can thereby determine correctly which side of $\hat{z}$ on H has an empty intersection with the level set. □

Similarly as in Corollary 1, we can extend this oracle into an approximate minimization procedure, which solves our optimization problem up to an accuracy of $\kappa\gamma$, provided that we have at our disposal a point $(x,y) \in S$ such that $f(x,y)$ is a lower bound on the mixed-integer optimal value.

Let us now modify our method for finding the k-th best point for two-dimensional problems to problems with two integer and d continuous variables. Here, we aim at finding — at least approximately — the k-th best fiber $\hat{x}_k^* \in [-B,B]^2$, so that:

$$(\hat{x}_k^*, y_k^*) \in \arg\min\{f(x,y) : (x,y) \in S \cap ((\mathbb{Z}^2 \setminus \{\hat{x}_1^*, \ldots, \hat{x}_{k-1}^*\}) \times \mathbb{R}^d)\}$$

for a $y_k^* \in \mathbb{R}^d$. We set $\hat{f}_{[k]}^* := f(\hat{x}_k^*, y_k^*)$. The following proposition summarizes the necessary extensions of Subsection 3.2.

Proposition 6. *Let $k \geq 2$ and let $\Pi_{k-1} := \{\hat{z}_1^*, \ldots, \hat{z}_{k-1}^*\} \subseteq [-B,B]^2 \cap \mathbb{Z}^2$ be points for which $\phi(\hat{z}_i^*) \leq \hat{f}_i^* + i\kappa\gamma$, $g(\hat{z}_i^*) \leq 0$ when $1 \leq i < k$ and such that $\operatorname{conv}\{\Pi_{k-1}\} \cap \mathbb{Z}^2 = \Pi_{k-1}$. In a number of approximate evaluations of f and $g_1, \ldots, g_m$ that is polynomial in $\ln(V_f/\gamma)$ and k, one can either*

(a) find an integral point $\hat{z}_k^* \in [-B,B]^2$ for which $\phi(\hat{z}_k^*) \leq \hat{f}_{[k]}^* + k\kappa\gamma$, $g(\hat{z}_k^*) \leq 0$ and $\operatorname{conv}\{\Pi_{k-1}, \hat{z}_k^*\} \cap \mathbb{Z}^2 = \Pi_{k-1} \cup \{\hat{z}_k^*\}$, or
(b) show that there is no integral point $\hat{z}_k^* \in [-B,B]^2$ for which $g(\hat{z}_k^*) \leq 0$.

Proof. If $k = 2$, we run Algorithm 3 applied to $\hat{z}_1^*$ with Line 9 replaced by solving $\min\{\langle h, \hat{y}\rangle : \hat{y} \in \bar{T}_k \cap \mathbb{Z}^2,\ \langle h, \hat{y}\rangle \geq \langle h, \hat{z}_1^*\rangle + 1\}$, where $h \in \mathbb{Z}^2$ such that $\gcd(h_1, h_2) = 1$. We also need to use approximate bisection methods instead of exact ones. Following the proof of Proposition 5, the oracle finds, if it exists, a feasible point $\hat{z}_2^*$. Either $\tilde{\phi}(\hat{z}_2^*) \leq \tilde{\phi}(\hat{z}_1^*) + \kappa\gamma \leq \hat{f}_{[1]}^* + 2\kappa\gamma \leq \hat{f}_{[2]}^* + 2\kappa\gamma$, or $\tilde{\phi}(\hat{z}_2^*) > \tilde{\phi}(\hat{z}_1^*) + \kappa\gamma$, then $\phi(\hat{z}_2^*) \leq$

$\tilde{\phi}(\hat{z}_2^*) \leq \hat{f}_{[2]}^* + \kappa\gamma$. Note that, if $\phi(\hat{z}_2^*) > \phi(\hat{z}_1^*) + \kappa\gamma$, we can conclude a posteriori that z_1^* corresponds precisely to $f_{[1]}^*$.

For $k \geq 3$, we can define the same triangulation as in Figure 7. Replicating the observation sketched above, we generate indeed a feasible point $\hat{z}_k^*$ for which $\tilde{\phi}(\hat{z}_k^*) \leq \hat{f}_{[k]}^* + k\kappa\gamma$.

Lemma 3 is extended as follows. Suppose that there is an integer point $\hat{x}$ in $\mathrm{conv}\{\Pi_{k-1}, \hat{z}_k^*\} \setminus (\Pi_{k-1} \cup \{\hat{z}_k^*\})$. Since $\phi(x) \leq \tilde{\phi}(x) \leq \hat{f}_{[k]}^* + k\kappa\gamma$ and $g(x) \leq 0$ for every $x \in \Pi_{k-1} \cup \{\hat{z}_k^*\}$, we have $\phi(\hat{x}) \leq \hat{f}_{[k]}^* + k\kappa\gamma$ and $g(\hat{x}) \leq 0$ by convexity. Thus, we can apply Algorithm 4 to find a suitable point $\hat{z}_k^*$ in $\mathrm{conv}\{\Pi_{k-1}, \hat{z}_k^*\}$. □

4.3 *A finite-time algorithm for mixed-integer convex optimization*

In this subsection, we explain how to use the results of the previous subsection in order to realize the oracle $\mathrm{O}_{\alpha,\delta}$ for $\alpha \geq 0$, $\delta > 0$ in the general case, i.e., with $n \geq 3$ integer and d continuous variables as in (1).

Let $z \in S \subseteq [-B,B]^n \times \mathbb{R}^d$ be the query point of the oracle. The oracle needs to find a point $\hat{z} \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ for which $f(\hat{z}) \leq (1+\alpha) f(z) + \delta$ (so as to report **a**), or to certify that $f(z) < f(\hat{z})$ for every $\hat{z} \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ (so as to report **b**). To design such an oracle we have at our disposal a procedure to realize the oracle $\mathrm{O}_{\alpha,\delta}$ for any mixed-integer convex minimization problem of the kind (1) with $n = 2$. We propose a finite-time implementation of $\mathrm{O}_{\alpha,\delta}$ with $\alpha = 0$ and $\delta = \kappa\gamma$. The main idea is to solve the n-dimensional case iteratively through the fixing of integer variables. This works as follows. We start by solving approximately the relaxation:

$$\hat{f}_{12}^* := \min\{f(x,y) : (x,y) \in S \cap (\mathbb{Z}^2 \times \mathbb{R}^{(n-2)+d})\}$$

with the techniques developed in the previous subsection. If we can solve the partial minimization problems up to an accuracy of $\gamma \leq \delta/\kappa$, we obtain a point $(\hat{u}_1^*, \hat{u}_2^*, x_3^*, \ldots, x_n^*, y^*) \in S$ with $\hat{u}_1^*, \hat{u}_2^* \in \mathbb{Z}$ and for which:

$$\tilde{f}_{12}^* := f(\hat{u}_1^*, \hat{u}_2^*, x_3^*, \ldots, x_n^*, y^*) \leq \hat{f}_{12}^* + \kappa\gamma$$

As $\hat{f}_{12}^*$ is a lower bound on the mixed-integer optimal value $\hat{f}^*$, we can make our oracle output **b** if $\tilde{f}_{12}^* - \kappa\gamma > f(z)$. So, assume that $\tilde{f}_{12}^* - \kappa\gamma \leq f(z)$.

Then we fix $\hat{x}_i := \hat{u}_i^*$ for $i = 1, 2$ and solve (if $k \geq 4$; if $k = 3$, the necessary modifications are straightforward)

$$\hat{f}_{1234}^* := \min\{f(x,y) : (x,y) \in S \cap ((\hat{u}_1^*, \hat{u}_2^*) \times \mathbb{Z}^2 \times \mathbb{R}^{(n-4)+d})\}.$$

We obtain a point $(\hat{u}_1^*, \ldots, \hat{u}_4^*, x_5^*, \ldots, x_n^*, y^*) \in S$ with $\hat{u}_i^* \in \mathbb{Z}$ for $1 \leq i \leq 4$ and for which:

$$\tilde{f}_{1234}^* := f(\hat{u}_1^*, \ldots, \hat{u}_4^*, x_5^*, \ldots, x_n^*, y^*) \leq \hat{f}_{1234}^* + \kappa\gamma \leq \hat{f}^* + \kappa\gamma.$$

Now, if $\tilde{f}^*_{1234} - \kappa\gamma > f(z)$, we can make our oracle output **b**. Thus, we assume that $\tilde{f}^*_{1234} - \kappa\gamma \leq f(z)$ and fix $\hat{x}_i := \hat{u}^*_i$ for $1 \leq i \leq 4$. Iterating this procedure we arrive at the subproblem (again, the procedure can easily be modified if n is odd):

$$\min\{f(x,y) : (x,y) \in S \cap ((\hat{u}^*_1, \dots, \hat{u}^*_{n-2}) \times \mathbb{Z}^2 \times \mathbb{R}^d)\}.$$

Let $(\hat{u}^*_1, \dots, \hat{u}^*_n, y^*) \in \mathbb{Z}^n \times \mathbb{R}^d$ be an approximate optimal solution. If we cannot interrupt the algorithm, i.e., if $f(\hat{u}^*_1, \dots, \hat{u}^*_n, y^*) \not\leq (1+\alpha)f(z) + \kappa\gamma$, we replace $(\hat{u}^*_{n-3}, \hat{u}^*_{n-2})$ by the second best point for the corresponding mixed-integer convex minimization problem. In view of Proposition 6, the accuracy that we can guarantee on the solution is only $2\kappa\gamma$, so the criterion to output **b** must be adapted accordingly. Then we proceed with the computation of $(\hat{u}^*_{n-1}, \hat{u}^*_n)$ and so on.

It is straightforward to verify that this approach results in a finite-time algorithm for the general case. In the worst case the procedure forces us to visit all integral points in $[-B,B]^n$. However, in the course of this procedure we always have a feasible solution and a lower bound at our disposal. Once the lower bound exceeds the value of a feasible solution we can stop the procedure. It is precisely the availability of both, primal and dual information, that makes us believe that the entire algorithm is typically much faster than enumerating all the integer points in $[-B,B]^n$.

References

[AK07] S. Arora and S. Kale, *A combinatorial, primal-dual approach to semidefinite programs [extended abstract]*, STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, ACM, New York, 2007, pp. 227–236. MR 2402446

[BBC$^+$08] P. Bonami, L. Biegler, A. Conn, G. Cornuéjols, I. Grossmann, C. Laird, J. Lee, A. Lodi, F. Margot, N. Sawaya, and A. Wächter, *An algorithmic framework for convex mixed integer nonlinear programs*, Discrete Optimization **5** (2008), no. 2, 186–204. MR 2408416

[DG86] M. Duran and I. Grossmann, *An outer-approximation algorithm for a class of mixed-integer nonlinear programs*, Mathematical Programming **36** (1986), no. 3, 307–339. MR 88c:90100

[EL05] F. Eisenbrand and S. Laue, *A linear algorithm for integer programming in the plane*, Mathematical Programming **102** (2005), no. 2, Series A, 249–259. MR 2005k:90080

[FL94] R. Fletcher and S. Leyffer, *Solving mixed integer nonlinear programs by outer approximation*, Mathematical Programming **66** (1994), no. 3, Series A, 327–349. MR 95h:90083

[GLS88] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, Algorithms and Combinatorics: Study and Research Texts, vol. 2, Springer-Verlag, Berlin, 1988. MR 89m:90135

[Gra72] R.L. Graham, *An efficient algorithm for determining the convex hull of a finite planar set*, Information Processing Letters 1 (1972), 132–133.

[HUL93] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms. II*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 306, Springer-Verlag, Berlin, 1993, Advanced Theory and Bundle Methods. MR 95m:90002

[Kha79] L. Khachiyan, *A polynomial algorithm in linear programming*, Doklady Akademii Nauk SSSR **244** (1979), 1093–1096.

[Len83] H. Lenstra, Jr., *Integer programming with a fixed number of variables*, Mathematics of Operations Research **8** (1983), no. 4, 538–548. MR 86f:90106

[NDY83] A. Nemirovski and D. D. Yudin, *Problem Complexity and Method Efficiency in Optimization*, John Wiley, 1983.

[Nes04] Y. Nesterov, *Introductory Lectures on Convex Optimization*, Applied Optimization, vol. 87, Kluwer Academic Publishers, Boston, 2004. MR 2005k:90001

[Roc81] R. Rockafellar, *The Theory of Subgradients and its Applications to Problems of Optimization*, R & E, vol. 1, Heldermann Verlag, Berlin, 1981, Convex and Nonconvex Functions. MR 83b:90126

[VI90] J. Viswanathan and Grossmann I., *A combined penalty function and outer-approximation method for MINLP optimization*, Computers & Chemical Engineering **14** (1990), no. 7, 769–782.