DISS. ETH NO. 22288

Integer Convex Minimization in Low Dimensions

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

TIMM OERTEL

Diplom-Mathematiker, Otto-von-Guericke Universität Magdeburg

born on 11.05.1985

citizen of Germany

accepted on the recommnedation of

Prof. Dr. Robert Weismantel, examiner
Prof. Dr. Friedrich Eisenbrand, co-examiner

2014

Abstract

In this dissertation we discuss several approaches to solve integer and mixed-integer convex minimization problems. That is, we try to minimize a convex function over a convex set with the additional constraint that a small number variables must be integral.

The thesis consists of four parts.

In the first part we apply the Mirror-Descent Method from continuous convex optimization to the mixed-integer setting. The main feature of this method is that the number of iterations is independent of the dimension, however, this method relies on a strong oracle, the so called improvement oracle. We present an efficient realization of such an oracle for the case when only two variables are required to be integral.

The second part contains two alternative, short, and geometrically motivated proofs of the well known result that minimizing a convex function over the integral points of a bounded convex set is polynomial in fixed dimension. In particular, we present an oracle-polynomial algorithm that is based on a mixed-integer linear optimization oracle.

Then, in the third part, we extend the Method of Centers of Gravity to the integer and mixed-integer setting. The crucial step consists in replacing the points of center of gravity by more general center-points, allowing us to use measures other than the volume. We introduce the concepts of center-points and approximate center-points. For special instances we prove properties of the (approximate) center-points. In the integer setting and when the dimension is fixed, we present an algorithm to compute approximate center-points. Furthermore, we establish optimality certificates for (mixed-) integer minimization problems based on lattice free polyhedra and we present a algorithm based on center-points that terminates with such an optimality certificate.

In the last part we consider a special class of, not necessarily convex, optimization problems in variable dimension. We aim to optimize $f(Wx)$ over a set $P \cap \mathbb{Z}^n$, where f is a non-linear function, $P \subset \mathbb{R}^n$ is a polyhedron and $W \in \mathbb{Z}^{d \times n}$. The dimension n may vary, however, we assume that the dimension d is fixed. We obtain an efficient transformation from the latter class of problems to integer linear problems. The core result is a representation

theorem, characterizing the set $W(P \cap \mathbb{Z}^n)$, which can be seen as Frobenius type theorem for polyhedra.

Zusammenfassung

In der vorliegenden Dissertationsschrift betrachten wir ganzzahlige und gemischt-ganzzahlige Minimierungsprobleme. Das sind Probleme, bei denen eine konvexe Zielfunktion und eine konvexe Menge gegeben sind und versucht wird die Zielfunktion über die konvexe Menge zu minimieren, wobei wir fordern, dass eine, in unserem Fall kleine, Anzahl an Variablen ganzzahlig zu sein hat. Wir diskutieren mehrere algorithmische Ansätze um dies zu lösen.

Die Arbeit besteht aus vier Teilen.

Im ersten Teil betrachten wir einen Algorithmus aus der kontinuierlichen Optimierung, nämlich die *Mirrow-Descent Method*. Das Besondere dieser Methode ist, dass die Anzahl der Iterationen unabhängig von der Dimension ist. Jedoch beruht dieser Ansatz auf einem starken Orakel, einem sogenannten Verbesserungsorakel. Unser Kernbeitrag in diesem Teil ist eine effiziente Realisierung dieses Orakels für den Fall, dass höchstens zwei Variablen ganzzahlig sind.

Im zweiten Teil geben wir zwei alternative, kurze und geometrisch motivierte Beweise, dass das Minimieren einer konvexen Funktion über den ganzzahligen Punkten in einer konvexen Menge in polynomialer Zeit in der Eingabegröße realisierbar ist, vorausgesetzt, dass die Dimension der konvexen Menge fix ist. Speziell zeigen wir einen Orakel-polynomiellen Algorithmus der allein auf einem gemischt-ganzzahligen linearen Optimierungsorakel basiert.

Dann, im dritten Teil, betrachten wir die *Method of Centers of Gravity* für kontinuierliche konvexe Minimierung und adaptieren diese für den ganzzahligen und gemischt-ganzzahligen Fall. Dazu ersetzen wir den Massenmittelpunkt durch allgemeinere Mittelpunkte; dies erlaubt uns dann andere Maße als das Volumen zu nutzen. Als erstes definieren wir Mittelpunkte und approximative Mittelpunkte, dann beweisen wir für spezielle Fälle einige ihrer Eigenschaften und wir zeigen wie approximative Mittelpunkte gefunden werden können. Danach entwickeln wir Optimalitätskriterien die auf gitterpunktfreien Polyedern beruhen. Zum Schluss präsentieren wir noch einen Algorithmus der auf den zuvor genannten Mittelpunkten beruht und der mit den besagten Optimalitätskriterien terminiert.

Im letzten Teil untersuchen wir eine spezielle Klasse von, nicht notwendigerweise konvexen, Optimierungsproblemen in variabler Dimension. Wir wollen

$f(Wx)$ über $P \cap \mathbb{Z}^n$ optimieren, wobei f eine nichtlineare Funktion, $P \subset \mathbb{R}^n$ ein Polyeder und W eine Matrix aus $\mathbb{Z}^{d \times n}$ ist. Dabei nehmen wir an, dass n variabel, d jedoch fix ist. Wir erhalten eine effiziente Transformation die das zuvor genannte Problem auf ganzzahlige lineare Unterprobleme reduziert. Das zentrale Resultat ist hierbei eine kompakte Charakterisierung der Menge $W(P \cap \mathbb{Z}^n)$, dieses kann als eine Art von Frobenius-Theorem für Polyeder gesehen werde.

Contents

Chapter 1

Introduction

In this thesis we discuss ideas for solving general integer and mixed-integer convex minimization problems. Let $f : \mathbb{R}^{n+d} \mapsto \mathbb{R}$ be convex function and let $g : \mathbb{R}^{n+d} \mapsto \mathbb{R}^m$ be convex in each component $g(x)_i$, $i = 1, \ldots, m$. Our goal consists in solving problems of the following kind:

$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & g(x) \leq 0, \\
& x \in \mathbb{Z}^n \times \mathbb{R}^d.
\end{aligned}
\tag{1.1}
$$

Throughout this thesis we always assume that n is a small constant, but different from zero. Under this assumption it turns out that Problem (1.1) can be solved efficiently under mild assumptions about the underlying computational model. This was one of the landmark discoveries in the 1980's. Indeed, when f and all the g_i's are affine functions, Khachiyan proved in 1979 that, if all variables are continuous, the problem is solvable in polynomial time in the input size of the affine functions [Kha79]. The basic tool to prove this result is the ellipsoid method that itself has been the outcome of inventions by several famous mathematicians including Shor, Nemirovski and Yudin. Four years later, Lenstra proved that, if the number of integral variables is fixed and f and all the g_i's are affine functions, then the problem is still solvable in polynomial time in the input size [Len83]. This result is an application of the famous LLL-reduced bases theory [LLL82] in combination with the ellipsoid method. Grötschel, Lovasz and Schrijver adapted Lenstra's approach to solve general convex integer optimization problems [GLS88, Theorem 6.7.4]. In later years many researchers aimed at improving the running time of Lenstra's algorithm as well as the extensions to convex integer problems. This topic is not the focus of this thesis and hence, not further discussed here. Important references about this topic are [Kan87, KP00, Eis03, Hei05, EL05, HK12, Dad12].

We assume that a reader is familiar with linear, convex, and integer optimization. For detailed introductions to we refer to [Sch86], [NN94, Nes04, BV04b]

and [Sch86, GLS88] respectively. Further, we assume that the reader is familiar with the basic notations and concepts of convex and discrete geometry. Basic references on the latter topic are [Lek69, GL87, Gru07].

The following chapters are organized as followed:

Chapter 2 Let us consider the problem of minimizing a convex function f over a convex set, with the extra constraint that only some variables must be integer. We study an algorithmic approach to this problem, that is the *Mirror-Descent Method*, 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 implement the oracle efficiently, that is, in $\mathcal{O}(\ln(B))$ 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.

Chapter 3 We provide an alternative, short, and geometrically motivated proof of the result that minimizing a convex function over the integral points of a bounded convex set is polynomial in fixed dimension [GLS88, Theorem 6.7.4]. In particular, we present an oracle-polynomial algorithm that only utilizes a mixed integer linear optimization oracle.

Chapter 4 In this chapter we revisit the Method of Centers of Gravity [Nes04, Section 3.2.6.]. We generalize this approach by replacing the centers of gravity by general *center-points*. This allows us to utilize different measures, other than the volume, to analyze the progress of the algorithms.

We introduce the new term of center-points and their approximation. For several instances we prove properties of the center points and show how to compute them. As a further result we establish optimality certificates for (mixed-) integer minimization problems based on lattice free polyhedra. A natural algorithm based on center-points that terminates with such an optimality certificate will be presented.

Chapter 5 We prove a representation theorem about projections of sets of integer points by an integer matrix W. This can be seen as a polyhedral

analogue of several classical and recent results related to the *Frobenius problem.*

Our result is motivated by a large class of non-linear integer optimization problems in variable dimension. Concretely, we aim to optimize $f(Wx)$ over a set $\mathcal{F} = P \cap \mathbb{Z}^n$, where f is a non-linear function, $P \subset \mathbb{R}^n$ is a polyhedron and $W \in \mathbb{Z}^{d \times n}$. As a consequence of our representation theorem, we obtain a general efficient transformation from the latter class of problems to integer linear programming. Our bounds depend polynomially on various important parameters of the input data leading, among others, to first polynomial time algorithms for several classes of non-linear optimization problems.

1.1 The computational model, boundedness and hardness

In this section we first introduce the used notation and we introduce definitions that appear in all our models. Then we discuss one assumption about the tractability of Problem (1.1). It will turn out that the problem must be bounded, in order to be polynomial time solvable. Further, we give a lower bound on the complexity of integer, and therefore mixed-integer, convex minimization. Among other consequences it will show that the Problem (1.1) is in general NP-hard.

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a not necessarily differentiable, convex function. We define the *domain of f* as $dom(f) := \{x \in \mathbb{R}^n \mid f(x) < \infty\}$. Then, for every point $x \in \operatorname{int} \operatorname{dom}(f)$ we define the *subdifferential of f at x* as

$$\partial f(x) := \{h \in \mathbb{R}^n \mid f(y) \geq f(x) + h^\mathsf{T}(y - x) \text{ for all } y \in \mathbb{R}^n\},$$

i.e. the subdifferential is the set of all subgradients of f at a point x.

In our model we will assume that we have access to functions through *evaluation oracles* only. An analytic description of the functions describing Problem (1.1) is not requested in all our algorithms. Instead, we assume that the defining functions are given by *first-order evaluation oracles*.

Definition 1.1.1 (first-order evaluation oracle). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function presented by a first-order evaluation oracle. Then, queried on a point $\bar{x} \in \mathbb{R}^n$ the oracle returns $\bar{f} \in \mathbb{R}$ and $\bar{h} \in \mathbb{R}^n$ such that

$$\bar{f} = f(\bar{x}) \quad \text{and} \quad \bar{h} \in \partial f(\bar{x}).$$

Hence, we can evaluate function at a point x with a single operation. Compared to a zero-order evaluation oracle, we assume that we have, next to the function value, the additional information about the subgradient. On the other hand, for a twice differentiable function, a second-order evaluation oracle would deliver in addition the Hessian matrix.

In order to make integer convex minimization problems tractable one needs to assume that the problem is bounded, that is the optimal solution lies, say, in a box $[-B, B]^n$ where $B \in \mathbb{N}$. B is then considered to be part of the input explicitly or implicitly. Let us assume otherwise. Then, we can construct a family of 1-dimensional functions which cannot be distinguished if we only evaluate them on a fixed region $[-B, B]$. For $\alpha \in \mathbb{R}$ we define $f_\alpha : \mathbb{R} \mapsto \mathbb{R}$ as

$$f_\alpha(x) := \max\{x, \alpha - x\}.$$

For each α, the function f_α is minimized in the point $\frac{\alpha}{2}$ and the minimal function value is $f(\frac{\alpha}{2}) = \frac{\alpha}{2}$. We assume that we have access to the functions f_α through a first-order evaluation oracle. Hence, we assume that we do not have direct access to the function, but can only evaluate the function pointwise. In particular, we do not know the value α.

We aim in solving

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in \mathbb{R}, \end{aligned}$$

where $f \in \{f_\alpha \mid \alpha \in \mathbb{R}\}$. Suppose in the course of an algorithm we have evaluated the function on points $x_1, \ldots, x_l$, we can not recognize and distinguish the functions f_α with $|\alpha| \geq 2 \max\{|x_1|, \ldots, |x_l|\}$. Further, since α is not part of the input, but on the other side influences the optimal solution, it is, in general, not possible to output neither the optimal solution nor the optimal value in polynomial time, let alone find it.

Hence, in order to make general minimization problems tractable, we must either know a bound B explicitly or implicitly or content ourselves with the more restricted problem where we add additional bounding constraints.

In the next part we discuss the hardness of integer convex minimization. Already the case where f and all the g_i's are affine functions is well known to be NP complete [GJ79, Problem MP1]. We give a lower bound on the complexity of Problem (1.1): at least 2^n calls to the first-order evaluation oracle are requested to solve Problem (1.1).

We will define a class of convex functions with unique minimizers and such that they are hard to distinguish from the other functions. Let $\mathcal{F}$ denote this

class. We next consider the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in \{-1,1\}^n, \end{aligned}$$

where $f \in \mathcal{F}$. Again, we assume that we have access to the function only through a first-order evaluation oracle.

For our construction of $\mathcal{F}$, let $P := \text{conv}(\pm e_1, \ldots, \pm e_n)$ be the standard cross-polytope, where e_i denotes the i-th unit-vector $i = 1, \ldots, n$. We can describe P as follows

$$P = \{x \in \mathbb{R}^n \mid a^\mathsf{T} x \leq 1 \text{ for all } a \in \{-1,1\}^n\}.$$

Let $\epsilon > 0$. For every $b \in \{-1,1\}^n$ we define

$$\begin{aligned} P_b := \{x \in \mathbb{R}^n \mid & a^\mathsf{T} x \leq 1 \text{ for all } a \in \{-1,1\}^n \setminus b \\ & b^\mathsf{T} x \leq 1 + \epsilon\}. \end{aligned}$$

Then, we define for every $b \in \{-1,1\}^n$ the gauge-function $f_b : \mathbb{R}^n \mapsto \mathbb{R}$ by

$$f_b(x) := \min\{\lambda \geq 0 \mid x \in \lambda P_b\}.$$

Given $l < 2^n$ points $x_1, \ldots, x_l \in \mathbb{R}^n$ there always exist two distinct functions f_b and $f_{\bar{b}}$ in $\mathcal{F}$ such that $f_b(x_i) = f_{\bar{b}}(x_i)$ and $h \in \partial f_b(x_i) \cap \partial f_{\bar{b}}(x_i)$. Hence these two functions are not distinguishable. Let $\mathcal{F} := \{f_b \mid b \in \{-1,1\}^n\}$. Then the minimum of Problem (1.1) is always $\frac{n}{1+\epsilon}$ and it is attained in the point b corresponding to the function f_b. Note, however, that b is not part of the input. It follows that in general at least 2^n first-order evaluations oracle calls are necessary and sufficient to distinguish the functions and therefore to determine the optimum.

Many researchers believe that $\mathcal{O}(2^n)$ first-order evaluations oracle calls are also sufficient to solve (1.1). This conjecture and weakening of it are the content of several studies [KP00, Hei05, HK12, Dad12].

1.2 Lenstra Type Algorithms

In this section we consider the setting, where in Problem (1.1) we only have integral variables, i.e.

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0 \\ & x \in \mathbb{Z}^n. \end{aligned} \tag{1.2}$$

We assume that the problem is bounded, i.e. the optimal solution of (1.2) lies in $[-B, B]^n$ for some $B \in \mathbb{N}$. Moreover, n is assumed to be a constant. Several algorithms have been proposed in recent years to tackle the problem in time that is polynomial in the input size [Len83, GLS88, Kan87, KP00, Hei05, HK12, Dad12]. Although the analysis of the different algorithms is quite different, it turns out that they have a common scheme that goes back to Lenstra. In the following presentation we focus on these key ideas. We try to keep our exposition as simple as possible, omitting technical details whenever it is possible.

We start introducing tools and notation used. Let $K \subset \mathbb{R}^n$ be a centrally symmetric[1] and convex body. We denote with $\|\cdot\|_K$ the norm induced by K. For $x \in \mathbb{R}^n$ we define

$$\|x\|_K := \min\{\lambda \geq 0 \mid x \in \lambda K\}.$$

Let $B \in \mathbb{R}^{n \times n}$ be a non singular matrix and let $\Lambda = B\mathbb{Z}^n$ be a lattice. We call B the basis of Λ and we denote with $\det(\Lambda) = |\det(B)|$ the determinant of Λ. The basis assigned to a lattice Λ is not unique, any matrix $B' = BU$, where U is a unimodular matrix[2], is also a basis of Λ. However, there are ways to define a "nice" basis of Λ, so called *reduced basis*.

For that we first define the *Gram-Schmidt orthogonalization*, e.g. [Gru07, Section 28.1]. Let $b_1, \ldots, b_n$ denote the columns of B. Then the Gram-Schmidt orthogonalized basis is defined recursively as follows

$$\bar{b}_i := b_i - \sum_{j=1}^{i-1} \mu_{i,j} \bar{b}_j$$

where

$$\mu_{i,j} = \frac{b_i^\mathsf{T} \bar{b}_j}{\bar{b}_j^\mathsf{T} \bar{b}_j}, \ 1 \leq i \leq n, \ 1 \leq j < i.$$

The most famous reduced basis is the *LLL-reduced basis* [LLL82]. Let Λ be a lattice with basis B. Let $b_1, \ldots, b_n$ denote the columns of B and let $\bar{b}_1, \ldots, \bar{b}_n$ denote its Gram-Schmidt orthogonalized basis. Then B is called LLL-reduced if

$$\|\bar{b}_{i+1} + \mu_{i+1,i} \bar{b}_i\|^2 \geq \frac{3}{4} \|\bar{b}_i\|^2, \ 1 \leq i \leq n-1$$

---

[1] A set $K \subset \mathbb{R}^n$ is centrally symmetric if $K = -K$.
[2] U is a unimodular matrix if $U \in \mathbb{Z}^{n \times n}$ and $\det(U) = \pm 1$.

and

$$|\mu_{i,j}| \leq \frac{1}{2}, \ 1 \leq j < i \leq n.$$

The geometric interpretation of a LLL-reduced basis is that the vectors have a small norm and that they are nearly orthogonal to each other. This is also synthesized in the following property of a LLL-reduced basis, that is

$$\|b_1\| \cdot \ldots \cdot \|b_n\| \leq 2^{n(n-1)/4} \det(\Lambda). \tag{1.3}$$

For details on the existence of LLL-reduced basis, their computation and their properties we refer to [LLL82] and [Gru07, Chapter 28]. Another well known reduced basis that we will use occasionally is the *Korkine-Zolotareff-reduced basis* [KZ73]. This basis has slightly stronger properties, however it is harder to compute. For further details we refer to [KZ73, LLS90].

Let $\Lambda \subset \mathbb{R}^n$ be a lattice and and let $x \in \mathbb{R}^n$, Then, the *closest-lattice-point-problem* is defined as

$$\underset{z \in \Lambda}{\operatorname{argmin}} \|x - z\|_2.$$

Let B be a reduced basis of Λ. Then we can solve the latter problem approximately. Let $\lambda = U^{-1}x$ and let $\bar{\lambda} = \lfloor \lambda \rceil$ (i.e. component wise $-\frac{1}{2} < \lambda_i - \bar{\lambda}_i \leq \frac{1}{2}$). Then $U\bar{\lambda}$ solves the problem approximately, where the approximation only depends on n (and the choice of the reduced basis). Note that the closest-lattice-point-problem is equivalent to $\min_{z \in \mathbb{Z}^n} \|x - z\|_E$, where $E := \{x \in \mathbb{R}^n \mid x^\mathsf{T} B^{-\mathsf{T}} B^{-1} x \leq 1\}$.

Next, let $K \subset \mathbb{R}^n$ be a compact convex set. Let Λ be a lattice. With $\Lambda^\star$ we denote the polar lattice of Λ, i.e. $\Lambda^\star = \{x \in \mathbb{R}^n \mid x^\mathsf{T} z \in \mathbb{Z} \text{ for all } z \in \Lambda\}$. Note that if B is a basis of Λ then $B^{-\mathsf{T}}$ is a basis of $\Lambda^\star$. Let $v \in \mathbb{R}^n \setminus \{0\}$. We define the *width of K with respect to v* by

$$\omega(K, v) := \max\{v^\mathsf{T} x \mid x \in K\} - \min\{v^\mathsf{T} x \mid x \in K\}$$

and the *lattice width of K* by

$$\omega_\Lambda(K) := \min\{\omega(K, v) \mid v \in \Lambda^\star \setminus \{0\}\}.$$

If $\Lambda = \mathbb{Z}^n$ then we sometimes omit the subscript and write $\omega(K)$ instead of $\omega_{\mathbb{Z}^n}(K)$. A vector $v \in \Lambda \setminus \{0\}$ with $\omega_\Lambda(K) = \omega(K, v)$ is called *flatness direction for K*. We say that a vector $v \in \Lambda^\star \setminus \{0\}$ is an *approximate flatness direction for K* if $\omega(K, v)$ is bounded by $c \omega_\Lambda(K)$, where c is a constant that is only dependent on n. In fixed dimension, a flatness direction can be computed efficiently, whereas this problem is hard in variable dimension.

An approximate flatness direction in variable dimension can be recovered from an appropriate LLL-reduced basis. This task can be performed in polynomial time, even in variable dimension. Assume that we have an ellipsoidal approximation of K, that is we have an ellipsoid $E = \{x \in \mathbb{R}^n \mid x^\mathsf{T} D^{-\mathsf{T}} D^{-1} x \leq 1\}$, where $D \in \mathbb{R}^{n \times n}$ is a non singular matrix, and a center $e \in \mathbb{R}^n$ such that

$$e + E \subset K \subset e + nE.$$

Then $\omega(E, v) \leq \omega(K, v) \leq \omega(nE, v) = n\omega(K, v)$ for any $v \in \mathbb{R}^n \setminus \{0\}$. This implies that any approximate lattice width of E induces an approximate lattice width for K. Hence it suffices to compute an approximate flatness direction for E. For that we define $\phi : \mathbb{R}^n \mapsto \mathbb{R}^n$ by $\phi(x) := D^{-1} x$, i.e. ϕ maps E to the unit ball. Let $\Lambda := \phi(\mathbb{Z}^n)$, then it holds that $\omega(E) = \omega_\Lambda(\phi(E))$. Let B denote the LLL-reduced basis of Λ. With $b_1, \ldots, b_n$ we denote the columns of B and with $\bar{b}_1, \ldots, \bar{b}_n$ its Gram-Schmidt orthogonalized basis. Note that the distance between b_n and $\mathrm{lin}(b_1, \ldots, b_{n-1})$ is precisely $\|\bar{b}_n\|$ and

$$\phi(E) \cap \Lambda \subset \bigcup_{i=-k}^{k} i \cdot \bar{b}_n + \mathrm{lin}(b_1, \ldots, b_{n-1}),$$

where $k := \left\lfloor \frac{1}{\|\bar{b}_n\|} \right\rfloor$. The vector $v := \frac{\bar{b}_n}{b_n^\mathsf{T} b_n}$ is in $\Lambda^\star$, hence $\omega_\Lambda(\phi(K)) \leq \omega(\phi(K), v) = \frac{2}{\|\bar{b}_n\|}$. Since $\frac{1}{\omega(E)} \phi(E) \cap \Lambda = \{0\}$, it follows $\|b_n\| > \frac{1}{\omega(E)}$. From (1.3) and the two properties, $\|\bar{b}_i\| \leq \|b_i\|$ for $i = 1, \ldots, n$ and $\det(\Lambda) = \|\bar{b}_1\| \cdot \ldots \cdot \|\bar{b}_n\|$, it follows that

$$\|\bar{b}_n\| > \frac{1}{\omega(E) 2^{n(n-1)/4}}.$$

Hence $\omega(E, v) \leq 2^{n(n-1)/4} \omega(E)$.

Now we are ready to present a Lenstra type algorithm.

The idea is to construct a sequence of ellipsoids $E_0, E_1, \ldots$ with decreasing volume and which fulfill the invariant that each ellipsoid E_i contains the optimal solution(s). The algorithm starts with an initial ellipsoid $E_0 := \{x \in \mathbb{R}^n \mid (x - a_0)^\mathsf{T} A_0 (x - a_0) \leq \sqrt{n} B\}$, where A_0 is the identity matrix and $a_0 = 0$. It holds that $E_0 \supset [-B, B]^n$ which in turn implies that, by our assumption of boundedness, E_0 contains the optimal solution.

Then, using the LLL-reduced basis, the closest vector problem is solved approximately to compute an integral point x_0 close to a_0 with respect to the norm induced by $E_0 - E_0$. If $g_i(x_0) > 0$ for some i, then let h_0 be a subgradient of the subdifferential of g_i at x_0. Otherwise let h_0 be a subgradient of the subdifferential of f at x_0. Now h_0 and x_0 define a half-space

$H_0 := \{x \in \mathbb{R}^n \mid h_0^\mathsf{T} x \le h_0^\mathsf{T} x_0\}$. By the definition of the subgradient, the complement of H_0 either only contains infeasible points or points with objective value greater than or equal to $f(x_0)$. Hence, the optimal solution(s) of (1.2) must lie in $E_0 \cap H_0$ We compute a new ellipsoid E_1 containing $E_0 \cap H_0$. Then, clearly, also E_1 contains the optimal solution(s) of (1.2).

Let $\Omega \in (0, \frac{1}{n})$ be arbitrary but fixed. Let $\lambda_0 := \|x_0 - a_0\|_{E^0}$. It turns out that, if $\lambda_0 < \Omega$, one can compute an ellipsoid E^1 with

$$\mathrm{vol}(E^1) \le e^{-(1-n\Omega)^2/(5n)}\, \mathrm{vol}(E^0)$$

(see [GLS88, Theorem 3.3.9] and the lemmata within). This guarantees that the volume of the new ellipsoid is reduced by a factor that only depends on n. This procedure is repeated until, for some k, we obtain $\lambda_k \ge \Omega$. This implies that E_k is flat, ensuring the existence of a lattice width less than or equal to a constant only depending on Ω and n. This enables us to reduce the original problem to a small number of lower dimensional subproblems. For that we compute an approximate flatness direction $v \in \mathbb{Z}^n \setminus \{0\}$ such that $\omega(E^k, v) \le c(\Omega, n)$, where $c(\Omega, n)$ is again a constant only depending on Ω and n.

This allows us to enumerate $\lfloor \omega(E^k, v) \rfloor$ subproblems of lower dimension. For $i = \lceil \min_{x \in E^k} v^\mathsf{T} x \rceil, \ldots, \lfloor \max_{x \in E^k} v^\mathsf{T} x \rfloor$ we solve the subproblems

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \le 0 \\ & v^\mathsf{T} x = i \\ & x \in \mathbb{Z}^n. \end{aligned}$$

Finally, among all feasible solutions of the subproblems, if they exist, one determines the solution with smallest objective value. Since n is fixed, also the depth of this recursion is fixed. Hence, this procedure results in a polynomial time algorithm.

Chapter 2

Mirror-Descent Methods in Mixed-Integer Convex Optimization

This chapter is based on joint work with Michel Baes, Christian Wagner and Robert Weismantel [BOWW13]. We consider the Mirror-Descent Method and apply it to the Mixed-Integer setting. That is, we address the problem of minimizing a convex function f over a convex set, with the extra constraint that some variables must be integer. As we already mentioned in the first chapter, it is well known that if all variables are integer and if their total number is fixed that then this problem is polynomially solvable in the size of the input. This was first proven by Grötschel, Lovasz and Schrijver (see Theorem 6.7.10 in [GLS88]) and clearly this is also best possible, even when f is a piecewise linear function, as this problem is NP-hard (see [GJ79, Problem MP1]).

We will present a new approach, using the Mirror-Descent Method [NY79], which was first introduced by Nemirovski and Yudin. The most notable feature of the framework of the Mirror-Descent Method is that the dimension does not enter the number of iterations (see [NY83, Chapter 3]). This comes of course at a certain cost. The algorithm will rely on an improvement oracle, specified later, hiding many difficulties. The hardness will then lie within the realization of this oracle. However, if this oracle can be realized in polynomial time, then the problem can be solved in polynomial time as well.

We postpone the realization of the improvement oracle to later sections. In the first section we will introduce the general Mirror-Descent Method and give its analysis of the corresponding running-time. The (small) novelty will be here that we allow different errors in every iteration. The following sections will then deal with the realization of the oracle. In particular, for problems with two integer variables, we show with a novel geometric construction how to implement the oracle efficiently, that is, in $\mathcal{O}(\ln(B))$ approximate minimizations of f over the continuous variables, where B is a known bound on

the absolute value of the integer variables. Then we show how to adapt our algorithm 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.

The main distinction between results presented here and the result from [GLS88, Theorem 6.7.10] is the way in which the statements are proven. Proofs of similar results as in [GLS88] basically use a combination of the ellipsoid algorithm [Kha79] and a Lenstra-type algorithm [Len83]. Our proof techniques do not rely on these methods.

Another novelty for mixed integer optimization is that, on a high level, optimizing over the integral and continuous variables are not separated, i.e. our iteration involves integral and continuous variables at the same time (for a counter-example see Chapter 3). Further, we try to do pure continuous steps if possible, avoiding time expensive operations on the integral components.

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

$$\min\{f(z,y) : (z,y) \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)\}, \tag{2.1}$$

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 $x \in \mathbb{R}^{n+d}$ with $f(x) < +\infty$. 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 *first-order function oracles*. See Definition 1.1.1.

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 [VG90, 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 (2.1) up to a certain given accuracy.

In this chapter 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 to date for solving mixed-integer convex optimization problems in variable dimension with at most two integer variables.

2.1 An algorithm based on an "improvement oracle"

We study in this chapter an algorithmic approach to solve (2.1), 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 2.1.1 (Improvement Oracle). Let $\alpha, \delta \geq 0$. For every $x \in S$, the oracle

  a. returns $x' \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ such that $f(x') \leq (1 + \alpha)f(x) + \delta$, and/or

  b. asserts correctly that there is no point $x' \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ for which $f(x') \leq f(x)$.

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

As stressed in the above definition, the oracle might content itself with a feasible point x' 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 x' that it outputs for the input point $x \in S$ must satisfy:

$$f(x') - f^\star \leq \alpha f(x) + \delta + (f(x) - f^\star) \leq \alpha f(x) + \delta \leq \alpha f^\star + \delta,$$

where $f^\star$ is the optimal objective value of (2.1). Thus $f(x') \leq (1+\alpha)f(x)+\delta$, and it is not possible to hope for a better point of S from the oracle. We can therefore interrupt the computations and output x' as the final result of our method.

In the case where $f^\star > 0$ and $\delta = 0$, the improvement oracle might be realized by a relaxation of the problem of finding a suitable x': 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 $\mathrm{dom}\, f$, is the set of all the points $x \in \mathbb{R}^{n+d}$ with $f(x) < +\infty$. For all $x \in \mathrm{dom}\, f$, we denote by $f'(x)$ an element of the

subdifferential $\partial f(x)$ of f. We represent by $x^\star = (z^\star, y^\star)$ a minimizer of (2.1), and set $f^\star := f(x^\star)$. Further, we use a prime $(\cdot')$ to emphasise 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:

$$|f(x_1) - f(x_2)| \leq L||x_1 - x_2||$$

for every $x_1, x_2 \in S$. Equivalently, if $||\cdot||_*$ is the dual norm of $||\cdot||$, we have $||f'(x)||_* \leq L$ for every $f'(x) \in \partial f(x)$ and every $x \in \text{dom } f$.

Our first algorithm is a variant of the well-known Mirror-Descent Method (see Chapter 3 of [NY83]). 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 $x \in S$ for which $f(x)$ is a (strictly) positive lower bound of $f^\star$.

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 $x_1, x_2 \in \mathbb{R}^{n+d}$, we have:

$$V(x_2) - V(x_1) - \langle V'(x_1), x_2 - x_1 \rangle \geq \frac{\sigma}{2}||x_2 - x_1||^2.$$

We also use the conjugate V_* of V defined by $V_*(s) := \sup\{\langle s, x \rangle - V(x) : x \in \mathbb{R}^{n+d}\}$ for every $s \in \mathbb{R}^{n+d}$. We fix $x_0 \in S$ as the starting point of our algorithm and denote by M an upper bound of $V(x^\star)$. We assume that the solution of the problem $\sup\{\langle s, x \rangle - V(x) : x \in \mathbb{R}^{n+d}\}$ exists and can be computed easily, as well as the function $\rho(w) := \min\{||w - x|| : x \in S\}$ 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 $\max\{\langle s, x \rangle - V(x) : x \in S\}$ can be solved efficiently.

A possible building block for constructing an algorithm to solve (2.1) is the continuous optimum of the problem, that is, the minimizer of (2.1) 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 $x \in S$ such that $f(x)$ is a lower bound of $f^\star$. Would this

lower bound be readily available, we can jump immediately to the termination procedure (see Algorithm 2).

Data: $x_0 \in S$.
Set $x_0' := x_0$, $w_0 := x_0$, $s_0 := 0$, and $f_0 := f(x_0)$.
Select sequences $\{h_k\}_{k \geq 0}$, $\{\alpha_k\}_{k \geq 0}$, $\{\delta_k\}_{k \geq 0}$.
for $k = 0, \ldots, N$ **do**

    Compute $f'(x_k) \in \partial f(x_k)$ and $\rho'(w_k) \in \partial \rho(w_k)$.
    Set $s_{k+1} := s_k - h_k f'(x_k) - h_k \|f'(x_k)\|_* \rho'(w_k)$.
    Set $w_{k+1} := \arg\max\{\langle s_{k+1}, z \rangle - V(x) : x \in \mathbb{R}^{n+d}\}$.
    Set $x_{k+1} := \arg\min\{\|w_{k+1} - x\| : x \in S\}$.
    Compute $f(x_{k+1})$.
    if $f(x_{k+1}) \geq f_k$ **then** $x_{k+1}' := x_k'$, $f_{k+1} := f_k$.
    else

        Run $O_{\alpha_{k+1}, \delta_{k+1}}(x_{k+1})$.
        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 x_{k+1}' as the oracle output and $f_{k+1} := \min\{f(x_{k+1}'), f_k\}$.
        else
            Run the termination procedure with $x_0 := x_{k+1}$, $x_0' := x_{k+1}'$, return its output, and terminate the algorithm.
        end

    end

end

Algorithm 1: Mirror-Descent Method.

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 2.1.2. *Suppose that the oracle reports* **a** *for* $k = 0, \ldots, N$ *in Algorithm 1, that is, it delivers an output* x_k' *for every iteration* $k = 0, \ldots, N$. *Then:*

$$\frac{1}{\sum_{k=0}^{N} h_k} \sum_{k=0}^{N} \frac{h_k f(x_k')}{1 + \alpha_k} - f(x^\star)$$

$$\leq \frac{M}{\sum_{k=0}^{N} h_k} + \frac{2L^2}{\sigma} \cdot \frac{\sum_{k=0}^{N} h_k^2}{\sum_{k=0}^{N} h_k} + \frac{1}{\sum_{k=0}^{N} h_k} \sum_{k=0}^{N} \frac{h_k \delta_k}{1 + \alpha_k}.$$

Proof. Since V is σ-strongly convex with respect to the norm $\| \cdot \|$, its conju-

Data: $x_0 \in S$ with $f(x_0) \leq f^\star$, $x_0' \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$.
Set $l_0 := f(x_0)$, $u_0 := f(x_0')$.
Choose $\alpha, \delta \geq 0$. Choose a subproblem accuracy $\epsilon' > 0$.
for $k \geq 0$ **do**

> Compute using a bisection method a point $x_{k+1} = \lambda x_k + (1 - \lambda)x_k'$
> for $0 \leq \lambda \leq 1$, for which $f(x_{k+1}) - (l_k(\alpha + 1) + u_k)/(\alpha + 2) \in [-\epsilon', \epsilon']$.
> Run $O_{\alpha,\delta}(x_{k+1})$.
> **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 x_{k+1}' as the oracle output, $l_{k+1} := l_k$, $u_{k+1} := \min\{f(x_{k+1}'), u_k\}$.
>
> **else**
>> Set $x_{k+1}' := x_k'$, $l_{k+1} := f(x_{k+1})$, $u_{k+1} := u_k$.
>
> **end**

end

Algorithm 2: Termination procedure.

gate 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:

$$\langle \rho'(w_k), w_k - \hat{z}^* \rangle \geq \rho(w_k) = ||\pi(w_k) - w_k|| = ||z_k - w_k||. \qquad (2.2)$$

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

$$
\begin{aligned}
\langle \rho'(w_k), z - w_k \rangle &\leq \rho(z) - \rho(w_k) = ||z - \pi(z)|| - ||w_k - \pi(w_k)|| \\
&\leq ||z - \pi(w_k)|| - ||w_k - \pi(w_k)|| \leq ||z - w_k||. \quad (2.3)
\end{aligned}
$$

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

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

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

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

where the second inequality follows from (2.2) 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:

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

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

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:

$$
\left(\sum_{k=0}^N h_k \right) \left(\frac{f_N - \delta}{1 + \alpha} - x^\star \right) = \left(\sum_{k=0}^N h_k \right) \left(\frac{\min_{1 \leq i \leq N} f(x_i') - \delta}{1 + \alpha} - f^\star \right)
$$

$$
\leq \sum_{k=0}^N \frac{h_k (f(x_k') - \delta)}{1 + \alpha} - \left(\sum_{k=0}^N h_k \right) f^\star \leq M + \frac{2L^2}{\sigma} \sum_{k=0}^N h_k^2. \tag{2.4}
$$

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 (2.4) 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:

$$\frac{f_N - (1+\alpha)f^\star - \delta}{1+\alpha} \leq L\sqrt{\frac{M}{2\sigma}} \cdot \frac{\ln(N+1) + 2}{\sqrt{N+2} - 1}. \tag{2.5}$$

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.

Remark 2.1.3. Note that the Mirror-Decent Method is a generalization of the Gradient-Decent Method. For that consider the case that we have no constraints on our problem. (Except of course that the optimal solution is attained within a certain radius) Then, if we choose our norm to be the Euclidian, $s_k = W_k = x_k$. Hence, in each iteration we do a step in the direction of the gradient.

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 2.1.4. *Assume that $f(x_0') \geq f(x_0) > 0$, and that there is no point $x' \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ for which $f(x_0) > f(x')$.*

(a) The termination procedure cannot guarantee any $x' \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ an accuracy better than:

$$f(x') \leq f^\star + (2+\alpha)\left(\alpha f^\star + (1+\alpha)\epsilon' + \delta\right). \tag{2.6}$$

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

$$f(x') - f^\star \leq \epsilon f^\star + (2+\alpha)\left(\alpha f^\star + (1+\alpha)\epsilon' + \delta\right)$$

within

$$\max\left\{\left\lceil \ln\left(\frac{f(x_0')-f(x_0)}{f(x_0)\epsilon}\right)\Big/\ln\left(\frac{2+\alpha}{1+\alpha}\right)\right\rceil, 0\right\}$$

iterations.

Proof. Part (a). At every iteration k, there is by construction no $x' \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ for which $l_k > f(x')$. Also, $f(x_k') \geq u_k \geq \hat{f}^\star$. 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 $x_{k+1}' \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ at iteration k. Then:

$$f(x_{k+1}') \leq (1+\alpha)f(x_{k+1}) + \delta \leq (1+\alpha)(\lambda l_k + (1-\lambda)u_k + \epsilon') + \delta,$$

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 x_{k+1}. Observe that the oracle might return a point x_k' such that $f(x_k')$ is smaller than the above right-hand side. In this case, no progress is done. As $u_k \leq f(x_k')$, this implies:

$$(\lambda + \lambda\alpha)l_k + (1+\alpha)\epsilon' + \delta \geq (\lambda + \lambda\alpha - \alpha)f(x_k'). \tag{2.7}$$

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

$$f^\star + (2+\alpha)(\alpha f^\star + (1+\alpha)\epsilon' + \delta) \geq f(x_k').$$

Since all the inequalities in the above derivation can be tight, a better accuracy cannot be guaranteed with our strategy. Thus, we can output x_k'.

Part (b). Note that we can assume $\frac{f(x_0')-f(x_0)}{f(x_0)\epsilon} > 1$, for otherwise the point x_0' already satisfies our stopping criterion.

In order to assess the progress of the algorithm, we can assume that the stopping criterion (2.7) is not satisfied. As $l_{k+1} = l_k$ in our case where the oracle gives an output, we get:

$$
\begin{aligned}
\Delta_{k+1} &= u_{k+1} - l_k \leq f(x_{k+1}') - l_k \\
&\leq (1+\alpha)(\lambda l_k + (1-\lambda)u_k + \epsilon') + \delta - l_k \\
&= \frac{\alpha^2 + \alpha - 1}{2+\alpha}l_k + \frac{1+\alpha}{2+\alpha}u_k + (1+\alpha)\epsilon' + \delta \\
&= \frac{1+\alpha}{2+\alpha}(u_k - l_k) + \alpha l_k + (1+\alpha)\epsilon' + \delta \\
&\leq \frac{1+\alpha}{2+\alpha}\Delta_k + \alpha f^\star + (1+\alpha)\epsilon' + \delta.
\end{aligned}
$$

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

$$\begin{aligned} \Delta_{k+1} &= u_{k+1} - l_{k+1} = f(x'_k) - f(x_{k+1}) \\ &\leq u_k - (\lambda l_k + (1 - \lambda)u_k - \epsilon') = \lambda \Delta_k + \epsilon' \\ &\leq \frac{1+\alpha}{2+\alpha}\Delta_k + \alpha f^\star + (1 + \alpha)\epsilon' + \delta. \end{aligned}$$

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

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

and the proposition is proved because $f(x'_N) - f^\star \leq \Delta_N$. $\qquad\qquad\square$

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

We proceed as follows. In Section 2.2, 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 (2.1) 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 2.3 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 (2.1) with arbitrary n and d in finite time.

2.2 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 $x \in \mathbb{R}$ the oracle returns $x' := \arg\min\{f(\lfloor x \rfloor), f(\lceil x \rceil)\}$ if one of these numbers is smaller or equal to $f(x)$, or returns $\mathbf{b}$ otherwise. The first non-trivial case arises when $n = 2$ and $d = 0$. This is the topic of this section.

2.2.1 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 $O_{\alpha,\delta}$ with $\alpha = \delta = 0$, provided that the feasible set is contained in a known finite box $[-B, B]^2$.

Theorem 2.2.1. *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 $i = 1, \ldots, m$. 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 $z \in [-B, B]^2 \cap \mathbb{Z}^2$ with $f(z) \leq f(x)$ and $g_i(z) \leq 0$ 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 2.2.1 is described in the proof of the theorem. That proof requires two lemmata.

Lemma 2.2.2. *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 $x, y, z \in K \cap \mathbb{Z}^2$. Then $\mathrm{vol}(K) \geq \mathrm{vol}(\mathrm{conv}(\{x, y, z\})) = \frac{1}{2}|\det(x - z, y - z)| \geq \frac{1}{2}$. $\square$

Lemma 2.2.3. *Let $u, v, w \in \mathbb{R}^2$ be affinely independent. If*

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

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

Proof. We partition $\mathrm{conv}\{u, u+v, u+v-w\}$ 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 $P := \mathrm{conv}\{0, \frac{1}{2}v, \frac{1}{2}w, \frac{1}{2}v + \frac{1}{2}w\}$. 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 $\mathrm{conv}\{u, u+v, u+v-w\} \subset A_1 \cup A_2 \cup A_3$ (see Fig. 2.1). Our assumption implies that the set $u + \frac{1}{2}v + P$ does not contain any integer point except possibly on the segment $u + v + \mathrm{conv}\{0, w\}$. 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 x, y, 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.
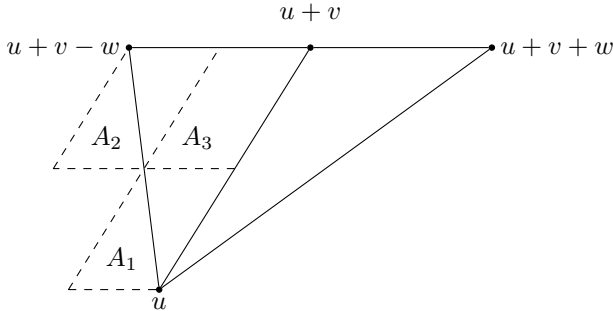
Figure 2.1: Partitioning the triangle in regions.

Clearly, the parallelogram $Q := \mathrm{conv}\{x, y, z, x - y + z\}$ 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 $a \in Q'$ if and only if 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 x, y, z), we can assume that $\langle x, w^\perp \rangle \leq \langle y, w^\perp \rangle \leq \langle z, w^\perp \rangle$. If $x - y + z \in A_1$ there is nothing to show, so we suppose that $x - y + z \notin A_1$.

Note that $\langle x, w^\perp \rangle \leq \langle x - y + z, w^\perp \rangle \leq \langle z, w^\perp \rangle$. Assume first that $\langle x - y + z, v^\perp \rangle < \langle z, v^\perp \rangle \leq \langle x, v^\perp \rangle, \langle y, v^\perp \rangle$ — the strict inequality resulting from the
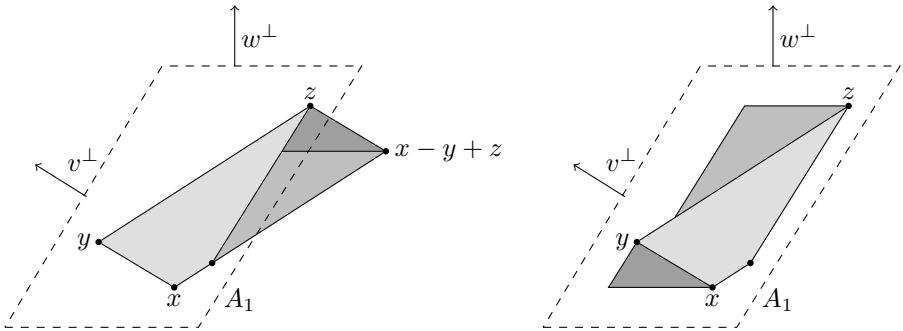
Figure 2.2: Mapping T.

fact that $x - y + z \notin A_1$. We define the mapping $T : Q \to A_1$ as follows,

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

(see Fig. 2.2). 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 $\langle x - y + z, v^\perp \rangle$, $\langle z, v^\perp \rangle$, $\langle x, v^\perp \rangle$, and $\langle y, v^\perp \rangle$. □

Remark 2.2.4. 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 $\mathcal{O}(m + \phi)$, where ϕ is the binary encoding length of the data.

Proof of Theorem 2.2.1. As described at the beginning of this section, 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 $g(y) :=$ $\max_{i=1...m} g_i(y)$ 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 $[-B, B]^2 = \bigcup_{j=1}^4 \text{conv}\{x, F_j\}$. The procedure we are about to describe has to be applied to every facet $F_1, \ldots, F_4$ successively, until a suitable point x' is found. Let us only consider one facet F. We define the triangle $T_0 := \text{conv}\{x, F\}$, whose area is smaller than $2B^2$.

To find an improving point within T_0, we construct a sequence $T_0 \supset T_1 \supset$ $T_2 \supset \ldots$ of triangles that all have x as vertex, with $\text{vol}(T_{k+1}) \leq \frac{2}{3} \text{vol}(T_k)$, and such that $f(z) > f(x)$ or $g(z) > 0$ for all $z \in (T_0 \setminus T_k) \cap \mathbb{Z}^2$. We stop our search if we have found an $x' \in [-B, B]^2 \cap \mathbb{Z}^2$ such that $f(x') \leq f(x)$ and $g(x') \leq 0$, 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 2.2.2 ensures that the integral points of T_k are on a line, and we need at most $\mathcal{O}(\ln(B))$ iterations to solve the resulting one-dimensional problem.

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

$$\min\{\langle h, z \rangle : z \in \bar{T}_k \cap \mathbb{Z}^2\} \tag{2.8}$$

where h is the normal vector to $\text{conv}\{v_0, v_1\}$ 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 (2.8) 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 2.2.3 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 $x' \in (T_k \setminus \bar{T}_k) \cap \mathbb{Z}^2$ such that $f(x') \leq f(x)$ and $g(x') \leq 0$. This requires to solve at most six one-dimensional subproblems.

Case 2. The integer linear program (2.8) has an optimal solution z. If $f(z) \leq f(x)$ and $g(z) \leq 0$, we return $x' = z$ and are done. So we assume that $f(z) > f(x)$ or $g(z) > 0$. Define $H := \{y \in \mathbb{R}^2 \mid \langle h, y \rangle = \langle h, z \rangle\}$, that is, the line containing z that is parallel to $\text{conv}\{v_0, v_1\}$, and denote by H_+ the closed half-space with boundary H that contains x. By definition of z, there is no integer point in $\bar{T}_k \cap \text{int } H_+$. Further, let $L := \text{aff}\{x, z\}$.

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 $f(z) > f(x)$ or $g(z) > 0$, 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
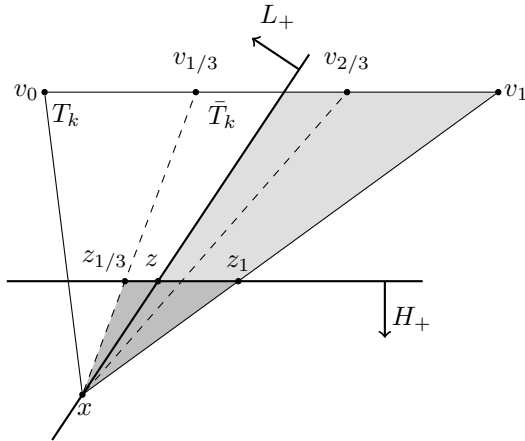
Figure 2.3: Illustration of Case 2.

in L_+ (see Fig. 2.3). By convexity of f and g, the set $((T_k \setminus H_+) \setminus L_+)$ (the lightgray region in Fig. 2.3) 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 2.2.3 on the triangle $\mathrm{conv}\{z_{1/3}, z_1, x\}$ (the darkgray region in Fig. 2.3), with $z_{1/3} = H \cap \mathrm{aff}\{x, v_{1/3}\}$ and $z_1 = H \cap \mathrm{aff}\{x, v_1\}$. If none of the corresponding subproblems returns a suitable point $x' \in \mathbb{Z}^2$, we know that $T_k \setminus L_+$ contains no improving integer point. Defining $T_{k+1} := T_k \cap L_+$, we have by construction $f(z) > f(x)$ or $g(z) > 0$ for all $z \in (T_k \setminus T_{k+1}) \cap \mathbb{Z}^2$ and $\mathrm{vol}(T_{k+1}) \leq \frac{2}{3} \mathrm{vol}(T_k)$.

It remains to determine the half-space L_+. If $g(z) > 0$ we just need to find a point $y \in H$ such that $g(y) < g(z)$, or if $f(z) > f(x)$, it suffices to find a point $y \in H$ such that $f(y) < f(z)$. 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 2.2.3 then suffices to determine whether there is a suitable x' in $T_k \cap H_+ \cap \mathbb{Z}^2$. $\qquad\square$

The algorithm presented in the proof of Theorem 2.2.1 can be adapted to output a minimizer $x^\star$ of f over $S \cap [-B, B]^2 \cap \mathbb{Z}^2$, provided that we know in advance that the input point x satisfies $f(x) \leq f^\star$: 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 2.2.5. *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 $f(x) \leq f^\star$, 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 $x^\star \in [-B, B]^2 \cap \mathbb{Z}^2$ with $f(x^\star) = \hat{f}^*$ and $g_i(x^\star) \leq 0$ for all $i = 1, \ldots, m$ or

(b) show that there is no such point.

Note that line 30 in Algorithm 3 requires the application of Lemma 2.2.2. Lines 11, 20 and 24 require the application of Lemma 2.2.3.

Remark 2.2.6 (Complexity). The following subroutines are used in Algorithm 3.

Line 9 and applications of Lemma 2.2.3. 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 30 and applications of Lemma 2.2.3. 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 18 and 19. 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.

2.2.2   Finding the k-th best point

In this section 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 Section 2.3.3 as a subroutine for the general mixed-integer convex problem. In the following, we describe the necessary extensions of the previous Algorithm 3. Let $x_1^\star := x^\star$ and define for $k \geq 2$:

$$x_k^\star := \arg\min \left\{ f(z) \mid z \in (S \cap [-B, B]^2 \cap \mathbb{Z}^2) \setminus \{x_1^\star, \ldots, x_{k-1}^\star\} \right\}$$

to be the k-th best point. Observe that, due to the convexity of the functions f and $g_1, \ldots, g_m$, we can always assume that $\text{conv}\{x_1^\star, \ldots, x_{k-1}^\star\} \cap \mathbb{Z}^2 =$

Data: $x \in [-B, B]^2$ with $f(x) \leq f^\star$ and $g_i(x) \leq 0$ for all $i = 1, \ldots, m$.

1 Let $F_1, \ldots, F_4$ be the facets of $[-B, B]^2$.

2 Set $x^\star := 0$ and $f^\star := +\infty$.

3 **for** $t = 1, \ldots, 4$ **do**

4   Set $F := F_t$ and define $v_0, v_1 \in \mathbb{R}^n$ such that $F := \text{conv}\{v_0, v_1\}$.

5   Write h for the vector normal to F pointing outwards $[-B, B]^2$.

6   Set $T_0 := \text{conv}\{x, F\}$ and $k := 0$.

7   **while** $\text{vol}(T_k) \geq \frac{1}{2}$ **do**

8   Set $\bar{T}_k := \text{conv}\{x, v_{1/3}, v_{2/3}\}$, with $v_\lambda := (1 - \lambda)v_0 + \lambda v_1$.

9   Solve $(\mathcal{P})$: $\min\{\langle h, z \rangle : z \in \bar{T}_k \cap \mathbb{Z}^2\}$.

10   **if** *(Case 1)* $(\mathcal{P})$ *is infeasible,* **then**

11   Determine $x' := \arg\min\{f(z) \mid z \in (T_k \setminus \bar{T}_k) \cap \mathbb{Z}^2 \text{ with } g(z) \leq 0\}$.

12   **if** x' *exists and* $f(x') < f^\star$ **then** Set $x^\star := x'$ and $f^\star := f(x')$.

13   **else**

14   *(Case 2)* Let z be an optimal solution of $(\mathcal{P})$.

15   Set $H_+ := \{y \in \mathbb{R}^2 : \langle h, y \rangle \leq \langle h, z' \rangle\}$ and $H := \partial H_+$.

16   Define $v := \text{aff}\{x, z\} \cap F$ and $z_i = H \cap \text{conv}\{x, v_i\}$ for $i = 0, 1$.

17   Denote $z_\lambda := (1 - \lambda)z_0 + \lambda z_1$ for $\lambda \in (0, 1)$.

18   **if** $g(z) \leq 0$ **and** *there is a* $y \in \text{conv}\{z_0, \hat{z}\}$ *s.t.* $f(y) < f(z)$ **or**

19              $g(z) > 0$ **and** *there is a* $y \in \text{conv}\{z_0, \hat{z}\}$ *s.t.* $g(y) < g(z)$ **then**

20   Determine

                     $x' := \arg\min\{f(y) \mid y \in \text{conv}\{x, z_{1/3}, z_1\} \cap \mathbb{Z}^2 \text{ with } g(y) \leq 0\}$.

21   **if** x' *exists and* $f(x') < f^\star$ **then** Set $x^\star := x'$ and $f^\star := f(x')$.

22   Set $v_1 := v$, $T_{k+1} := \text{conv}\{x, v_0, v\}$, and $k := k + 1$.

23   **else**

24   Determine

                     $x' := \arg\min\{f(y) \mid y \in \text{conv}\{x, z_0, z_{2/3}\} \cap \mathbb{Z}^2 \text{ with } g(y) \leq 0\}$.

25   **if** x' *exists and* $f(x') < f^\star$ **then** Set $x^\star := x'$ and $f^\star := f(x')$.

26   Set $v_0 := v$, $T_{k+1} := \text{conv}\{x, v, v_1\}$, and $k := k + 1$.

27   **end**

28   **end**

29   **end**

30   Determine $x' := \arg\min\{f(y) \mid y \in T_k \cap \mathbb{Z}^2 \text{ with } g(y) \leq 0\}$.

31   **if** x' *exists and* $f(x') < f^\star$ **then** Set $x^\star := x'$ and $f^\star := f(x')$.

32 **end**

33 **if** $f^\star < +\infty$ **then** Return $x^\star$.

34 **else** Return "the problem is unfeasible".

Algorithm 3: Minimization algorithm for 2D problems.

$\{x_1^\star, \ldots, x_{k-1}^\star\}$ for all $k \geq 2$. Although this observation appears plausible it is not completely trivial to achieve this algorithmically.

Lemma 2.2.7. *Let $\Pi_j := \{x_1^\star, \ldots, x_j^\star\}$ 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 $z_k^\star$ be a k-th best point.*

(a) *If $f(z_k^\star) > f^\star$, we can replace the point $z_k^\star$ by a feasible k-th best point $x_k^\star$ such that $\mathrm{conv}\{\Pi_{k-1}, x_k^\star\} \cap \mathbb{Z}^2 = \Pi_{k-1} \cup \{x_k^\star\}$ in $\mathcal{O}(1)$ operations.*

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

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

$$0 \leq f(z) - f(z_k^\star) \leq \sum_{i=1}^{k-1} \lambda_i (f(x_i^\star) - f(x_k^\star)) \leq 0.$$

Thus $f(z) = f(z_k^\star)$. Let $I := \{i : \lambda_i > 0\}$ and $Q_I := \mathrm{conv}\{z_k^\star\} \cup \{x_i^\star : i \in I\}$, so that $\hat{x} \in \mathrm{relint}\, Q_I$. 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 $\ell := \mathrm{aff}\{x_1^\star, z\}$: it is constant on the open interval $\ell \cap \mathrm{int}\, Q_I$, but does not attain its minimum on it, contradicting the convexity of f. Let us now construct the point $x_k^\star$: it suffices to consider the closest point to $x_k^\star$ in $\mathrm{aff}\{Q_I\} \cap P_{k-1}$, say $x_j^\star$, and to take the integer point $x_k^\star \neq x_j^\star$ of $\mathrm{conv}\{x_j^\star, x_k^\star\}$ that is the closest to $x_j^\star$ (see Fig. 2.4).
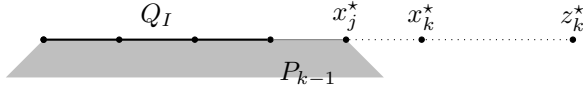
Figure 2.4: Illustration of Part (a).

Part (b). Suppose now that $f(x_i^\star) = f(z_k^\star) = f^\star$ for every $1 \leq i \leq k-1$, and define

$$\{v_0 \equiv v_\nu, v_1, \ldots, v_{\nu-1}\} \subseteq \Pi_{k-1}$$

Data: $z_k^\star, v_0, v_1, \ldots, v_j$.
Set $i := 0$ and $z(0) := z_k^\star$.
while $\det(z(i) - v_i, v_{i+1} - v_i) \geq 0$ **do**
> Set $\Delta_i := \text{conv}\{z(i), v_i, v_{i+1}\} \setminus \text{aff}\{v_i, v_{i+1}\}$.
> Set h_i a vector orthogonal to $\text{aff}\{v_i, v_{i+1}\}$ such that $\langle h_i, z(i) - v_i \rangle > 0$.
> Set $z(i+1) := \arg\min\{\langle h_i, z \rangle : z \in \Delta_i \cap \mathbb{Z}^2\}$.
> Set $i := i + 1$.

end
Set $x_k^\star := z(i)$.

Algorithm 4: Constructing a point $x_k^\star$ with $\text{conv}\{\Pi_{k-1}, x_k^\star\} \cap \mathbb{Z}^2 = \Pi_{k-1} \cup \{x_k^\star\}$.

as the vertices of P_{k-1}, labeled counterclockwise. It is well-known that determining the convex hull of $P_{k-1} \cup \{z_k^*\}$ costs $\mathcal{O}(\ln(\nu))$ operations. From these vertices, we deduce the set $\{v_i : i \in J\}$ of those points that are in the relative interior of that convex hull. Up to a renumbering of the v_l's, we have $J = \{1, 2, \ldots, j-1\}$. We show below that Algorithm 4 constructs a satisfactory point $x_k^\star$.

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

Define $T_l(i) := \text{conv}\{z(i), v_l, v_{l+1}\} \setminus P_{k-1}$ for $0 \leq l < j$ (see Fig. 2.5); the triangle Δ_i in Algorithm 4 corresponds to $T_i(i)$. Also, the vector h_i is orthogonal to the side $\text{aff}\{v_i, v_{i+1}\}$ of the triangle $T_i(i)$.

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

$$\frac{1}{2} \det(z(i) - v_i, v_{i+1} - v_i)$$

is nonnegative, and finds a point $z(i+1) \in T_i(i)$ such that $T_i(i+1)$ has only $z(i+1)$ as integer point.

We prove by induction on $i \geq 1$ that $T_l(i)$ contains only $z(i)$ as integer point whenever $l < i$. Consider the base case $i = 1$. By construction, the triangle $T_0(1)$ contains only $z(1)$ as integer point, for otherwise $z(1)$ would not minimize $\langle h_0, z \rangle$ 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

$z(i+1)$ is the only integer in $T_l(i+1)$. We have:

$$z(i+1) \in T_i(i) \subseteq \mathrm{conv}\{z(i), v_0, \ldots, v_{i+1}\} \setminus P_{k-1} = T_i(i) \cup \bigcup_{l=0}^{i-1} T_l(i).$$

This last equality represents a triangulation of the possibly non-convex polygon $\mathrm{conv}\{z(i), v_0, \ldots, v_{i+1}\} \setminus P_{k-1}$. From the above inclusion, we deduce:

$$K := \mathrm{conv}\{z(i+1), v_0, \ldots, v_{i+1}\} \setminus P_{k-1} \subseteq \mathrm{conv}\{z(i), v_0, \ldots, v_{i+1}\} \setminus P_{k-1}.$$

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 $\{z(i)\}$ by induction hypothesis, or in $T_i(i)$. Since $z(i) \in T_i(i)$, all the integers in $T_l(i+1)$ must be in $T_i(i)$. But $T_l(i+1) \cap T_i(i) \cap \mathbb{Z}^2 = \{z(i+1)\}$ by construction of $z(i+1)$. 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 $x_k^\star$. As the data of these problems are integers bounded by B, the complexity of the minimization solver used to compute $z(i+1)$ at every step is bounded by $\mathcal{O}(\ln(B))$. The overall complexity of Algorithm 4 is thus bounded by $\mathcal{O}(\nu \ln(B))$. $\square$

By Lemma 2.2.7, the k-th best point $x_k^\star$ can be assumed to be contained within $[-B, B]^2 \setminus \mathrm{conv}\{x_1^\star, \ldots, x_{k-1}^\star\}$. This property allows us to design a straightforward algorithm to compute this point. We first construct an inequality description of $\mathrm{conv}\{x_1^\star, \ldots, x_{k-1}^\star\}$, say $\langle a_i, x \rangle \leq b_i$ for $i \in I$ with $|I| < +\infty$. Then

$$[-B, B]^2 \setminus \mathrm{conv}\{x_1^\star, \ldots, x_{k-1}^\star\} = \bigcup_{i \in I}\{x \in [-B, B]^2 \mid \langle a_i, x \rangle > b_i\}.$$

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 $x_1^\star$ with the following minor modification: in Line 9, we replace $(\mathcal{P})$ with the integer linear problem $(\mathcal{P}') : \min\{\langle h, z \rangle : z \in \bar{T}_k \cap \mathbb{Z}^2, \langle h, z \rangle \geq \langle h, x_1^\star \rangle + 1\}$, where $h \in \mathbb{Z}^2$ such that $\gcd(h_1, h_2) = 1$. This prevents the algorithm from returning $x_1^\star$ again.

Let $k \geq 3$. Let $v_0, \ldots, v_{\nu-1}, v_\nu \equiv v_0$ denote the vertices of P_{k-1}, ordered counterclockwise (they can be determined in $\mathcal{O}(k \ln(k))$ operations using the
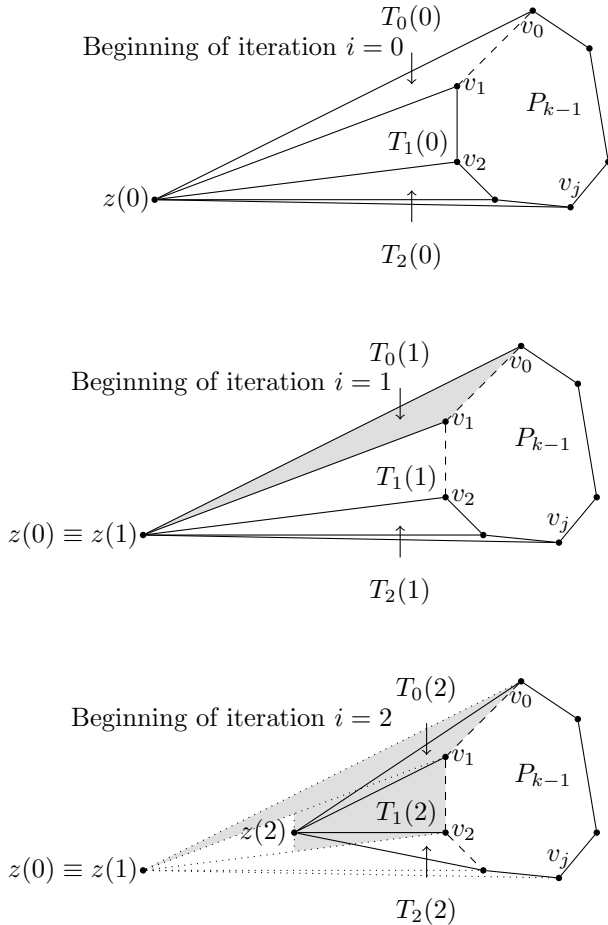
Figure 2.5: Constructing P_k from P_{k-1}: first iterations of Algorithm 4. The point $z(1)$ is the same as $z(0)$ because $T_0(0)$ has no other integer point than $z(0)$. The gray areas are, as the algorithm progresses, regions where we have established that they do not contain any integer point.

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 v_i as vertex and with a segment of the boundary of $[-B, B]^2$ as opposite side a *search triangle* (see Fig. 2.7: 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 $(\text{conv}\{x, F_t\})_{t=1}^4$.

For each $0 \leq i < \nu$, we define $H_i := \{y \in \mathbb{R}^2 : \det(y - v_i, v_{i+1} - v_i) \geq 0\}$, so that $H_i \cap P_{k-1} = \text{conv}\{v_i, v_{i+1}\}$. Consider the regions $R_i := ([-B, B]^2 \cap H_i) \setminus \text{int}\, H_{i-1}$. Note that R_i contains only v_i and v_{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 v_i to the appropriate vertices of $[-B, B]^2$.
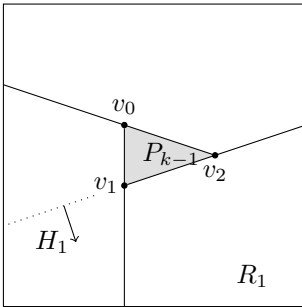
Figure 2.6: Triangulation step 1.   **Figure 2.7:** Triangulation step 2.
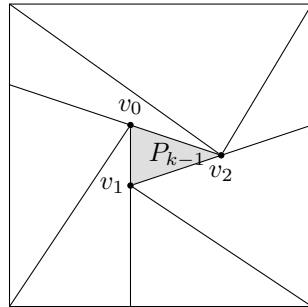
Note that a search triangle can contain two or more integer points of P_{k-1}. In order to prevent us from outputting one of those, we need to perturb the search triangles slightly before using them in Algorithm 3. Let $T = \text{conv}\{v_i, b_1, b_2\}$ be one of the search triangles, with b_1, b_2 being points of the boundary of $[-B, B]^2$. The triangle T might contain v_{i+1}, say $v_{i+1} \in \text{conv}\{v_i, b_1\}$, 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 $\mathcal{O}(\ln(B))$.

So, we apply Algorithm 3 with all these modified search triangles instead of $\text{conv}\{x, F_1\}, \ldots, \text{conv}\{x, F_4\}$. A simple modification of Line 9 allows us to avoid the point v_i for z: we just need to replace the linear integer problem $(\mathcal{P})$ with $\min\{\langle h, z \rangle : z \in \bar{T}_k \cap \mathbb{Z}^2, \langle h, z \rangle \geq \langle h, v_i \rangle + 1\}$, 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.

Corollary 2.2.8. *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 $x_1^\star, \ldots, x_{k-1}^\star$ be the $k-1$ best points for $\min\{f(z) : z \in S \cap [-B, B]^2 \cap \mathbb{Z}^2\}$. Then, in a number of evaluations of f and $g_1, \ldots, g_m$ that is polynomial in $\ln(B)$ and in k, one can either find*

(a) a k-th best point, $x_k^\star$, or

(b) show that there is no such point.

2.3 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 Section 2.3.1 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 Section 2.3.2, we build an efficient method for solving mixed-integer convex problems with two integer and d continuous variables and propose an extension of Corollary 2.2.8. 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 Section 2.3.3.

In this section, the problem of interest is (2.1):

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

with a few mild simplifying assumptions. We define the function

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

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:

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

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 $x \in \text{dom} \, \phi$ 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. [NN94, CGT00, 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 sections.

2.3.1   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 section, 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 ([Nem94, 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 2.3 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 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 2.3.1. *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. 2.8 illustrates the proof graphically. $\square$
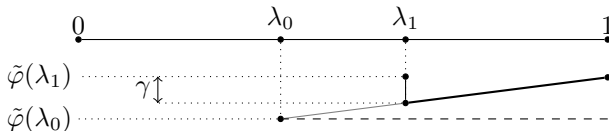
Figure 2.8: Lemma 2.3.1: the bold line represents a lower bound on φ in Part (a).

If one of the conditions in Lemma 2.3.1 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 2.3.1 is satisfied. In the lemma below, we use the values for λ_0, λ_1 chosen above.

Lemma 2.3.2. *Assume that $\tilde{\varphi}(\lambda_1) - \gamma < \tilde{\varphi}(\lambda_0) < \tilde{\varphi}(\lambda_1) + \gamma$. We define:*

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

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* $t \in [0,1] \backslash [\lambda_0, \lambda_1]$. *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 2.3.1. The second situation involves a tedious enumeration, summarized in Fig. 2.9. We assume, without loss of generality, that $\tilde{\varphi}(\lambda_0) \leq \tilde{\varphi}(\lambda_1)$. The bold lines in Fig. 2.9 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 of a simple generic inequality (2.9) that we establish below, before showing how we can particularize it to different segments of the interval $[0, 1]$.
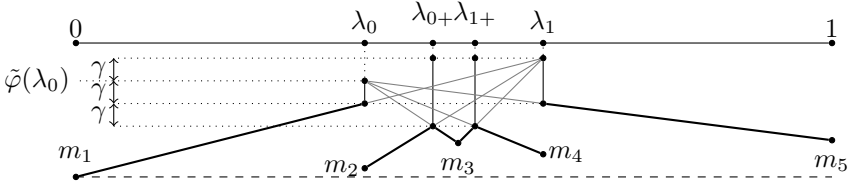
Figure 2.9: Approximate bisection: bold lines represent a lower bound on φ in the termination case.

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:

$$\varphi(t) - \tilde{\varphi}(\lambda_0) \geq \gamma_- - \frac{\gamma_+ + \gamma_-}{\mu} \geq \gamma_- - \frac{\gamma_+ + \gamma_-}{\mu_0}. \tag{2.9}$$

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 $-\gamma(\frac{2}{\lambda_0} - 1) \approx -4.236\gamma$. $\qquad\square$

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 2.3.3. *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 2.3.1 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 2.3.2 applies: if the first condition stated in Lemma 2.3.2 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 $t^\star$ 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 $I(\lambda) := (1 - \lambda)t^\star + \lambda[0, 1]$, 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,

$$\tilde{\varphi}(\bar{t}_k) \leq \max\{\varphi(t) : t \in I(\lambda)\} \quad \leq \quad (1 - \lambda)\varphi(t^\star) + \lambda \max\{\varphi(t) : t \in [0, 1]\}$$
$$\leq \quad (1 - \lambda)\varphi(t^\star) + \lambda(V_\varphi + \varphi(t^\star)).$$

Hence $\tilde{\varphi}(\bar{t}_k) - \varphi(t^\star) \leq \lambda V_\varphi$, and, by taking λ arbitrarily close to λ_1^k, we get $\tilde{\varphi}(\bar{t}_k) - \varphi(t^\star) \leq \lambda_1^k V_\varphi$. 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 $(\kappa-1)\gamma$. $\square$

Remark 2.3.4. If we content ourselves with a coarser precision $\eta \geq (\kappa - 1)\gamma$, we merely need $\mathcal{O}(\ln(V_\varphi/\eta))$ evaluations of $\tilde{\varphi}$. $\diamond$

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 2.3.5. *There exists an algorithm that finds a point $z^\star \in (t_0 + \tau\mathbb{Z}) \cap [0, 1]$ for which:*

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

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 2.3.3 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 $z^\star$, satisfies $\tilde{\varphi}(z^\star) - \gamma \leq \varphi(z^\star) = \min\{\varphi(z) : z \in (t_0 + \tau\mathbb{Z}) \cap [0,1]\}$.

In the second case, we have an interval $I_k \subseteq [0,1]$ and a point $\bar{t}_k$ that fulfill $\tilde{\varphi}(\bar{t}_k) \leq \min\{\varphi(z) : z \in [0,1]\} + (\kappa - 1)\gamma$, 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 z_- and z_+ that are the closest from $\bar{t}_k$. One of these two points constitutes an acceptable output for our algorithm. Indeed, suppose first that $\min\{\tilde{\varphi}(z_-), \tilde{\varphi}(z_+)\} \leq \tilde{\varphi}(\bar{t}_k) + \gamma$. Then:

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

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

$$\varphi(z_-) \leq \lambda\varphi(z) + (1 - \lambda)\varphi(z_k) < \lambda\varphi(z_-) + (1 - \lambda)(\tilde{\varphi}(z_-) - \gamma),$$

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

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 chapter, except in the remaining of this section 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 2.3.3. 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 2.3.6. *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 2.3.6 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 2.3.7. In order to solve problem (2.1) with one integer variable, we can extend Proposition 2.3.5 to implement the improvement oracle $O_{0,\kappa\gamma}$. 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 $(x',y') \in ((t_0 + \tau\mathbb{Z}) \cap [0,1]) \times \mathbb{R}^d$ for which $f(x',y') \leq f(x,y)$, provided that the output x' of our approximate bisection method for integers given in Proposition 2.3.5 satisfies $\tilde{\varphi}(x') - \kappa\gamma > f(x,y)$. Otherwise, we can determine a point (x',y') for which $f(x',y') \leq f(x,y) + \kappa\gamma$. Note that this oracle cannot report **a** and **b** simultaneously.

2.3.2   Mixed-integer convex problems with two integer variables

We could use the Mirror-Descent Method in Algorithm 1 to solve the generic problem (2.1) 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 (2.5), 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 2.3.8. *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 2.2.1 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 2.2.1 requires. We borrow the notation from the proof of Theorem 2.2.1.

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 2.3.5. This bisection procedure detects, if it exists, a point z on the line of interest for which $\tilde{\phi}(z) = f(z, y_z) \leq f(x, y) + \kappa\gamma$ and we are done. Or it reports correctly that there is no integer z on the line of interest with $\phi(z) \leq f(x, y)$.

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

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 $v \in \text{conv}\{z_0, z\}$ for which
>
> $$\tilde{\phi}(v) < f(x,y) + (\kappa - 1)\gamma",$$

which can be verified using Proposition 2.3.3. If such a point v exists, the convexity of ϕ forbids any $w \in \text{conv}\{z, z_1\}$ to satisfy $\phi(w) \leq f(x,y)$, for otherwise:

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

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

Similarly as in Corollary 2.2.5, 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 $x_k^\star \in [-B, B]^2$, so that:

$$(x_k^\star, y_k^\star) \in \arg\min\{f(x,y) : (x,y) \in S \cap ((\mathbb{Z}^2 \setminus \{x_1^\star, \ldots, x_{k-1}^\star\}) \times \mathbb{R}^d)\}$$

for a $y_k^\star \in \mathbb{R}^d$. We set $f_{[k]}^\star := f(x_k^\star, y_k^\star)$. The following proposition summarizes the necessary extensions of Section 2.2.2.

Proposition 2.3.9. *Let $k \geq 2$ and let $\Pi_{k-1} := \{z_1^\star, \ldots, z_{k-1}^\star\} \subseteq [-B, B]^2 \cap \mathbb{Z}^2$ be points for which $\phi(z_i^\star) \leq \hat{f}_{[i]}^\star + i\kappa\gamma$, $g(z_i^\star) \leq 0$ when $1 \leq i < k$ and such that $\text{conv}\{\Pi_{k-1}\} \cap \mathbb{Z}^2 = \Pi_{k-1}$. 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 $z_k^\star \in [-B, B]^2$ for which $\phi(z_k^\star) \leq \hat{f}_{[k]}^\star + k\kappa\gamma$, $g(z_k^\star) \leq 0$ and $\text{conv}\{\Pi_{k-1}, z_k^\star\} \cap \mathbb{Z}^2 = \Pi_{k-1} \cup \{z_k^\star\}$, or*

(b) *show that there is no integral point $z_k^\star \in [-B, B]^2$ for which $g(z_k^\star) \leq 0$.*

Proof. If $k = 2$, we run Algorithm 3 applied to $z_1^\star$ with Line 9 replaced by solving $\min\{\langle h, z \rangle : z \in \bar{T}_k \cap \mathbb{Z}^2, \langle h, z \rangle \geq \langle h, z_1^\star \rangle + 1\}$, 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 2.3.8, the oracle finds, if it exists, a feasible point $z_2^\star$. Either $\tilde{\phi}(z_2^\star) \leq \tilde{\phi}(z_1^\star) + \kappa\gamma \leq f_{[1]}^\star + 2\kappa\gamma \leq f_{[2]}^\star + 2\kappa\gamma$, or $\tilde{\phi}(z_2^\star) > \tilde{\phi}(z_1^\star) + \kappa\gamma$, then $\phi(z_2^\star) \leq \tilde{\phi}(z_2^\star) \leq f_{[2]}^\star + \kappa\gamma$. Note that, if $\phi(z_2^\star) > \phi(z_1^\star) + \kappa\gamma$, we can conclude a posteriori that $z_1^\star$ corresponds precisely to $f_{[1]}^\star$.

For $k \geq 3$, we can define the same triangulation as in Figure 2.7. Replicating the observation sketched above, we generate indeed a feasible point $z_k^\star$ for which $\tilde{\phi}(z_k^\star) \leq f_{[k]}^\star + k\kappa\gamma$.

Lemma 2.2.7 is extended as follows. Suppose that there is an integer point z in $\mathrm{conv}\{\Pi_{k-1}, z_k^\star\} \setminus (\Pi_{k-1} \cup \{z_k^\star\})$. Since $\phi(x) \leq \tilde{\phi}(x) \leq f_{[k]}^\star + k\kappa\gamma$ and $g(x) \leq 0$ for every $x \in \Pi_{k-1} \cup \{z_k^\star\}$, we have $\phi(z) \leq f_{[k]}^\star + k\kappa\gamma$ and $g(z) \leq 0$ by convexity. Thus, we can apply Algorithm 4 to find a suitable point $z_k^\star$ in $\mathrm{conv}\{\Pi_{k-1}, z_k^\star\}$. $\qquad\square$

2.3.3 A finite-time algorithm for mixed-integer convex optimization

In this section, we explain how to use the results of the previous section in order to realize the oracle $O_{\alpha,\delta}$ for $\alpha \geq 0$, $\delta > 0$ in the general case, i.e., with $n \geq 3$ integer and d continuous variables as in (2.1).

Let $x \in S \subseteq [-B, B]^n \times \mathbb{R}^d$ be the query point of the oracle. The oracle needs to find a point $x' \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ for which $f(x') \leq (1 + \alpha)f(x) + \delta$ (so as to report **a**), or to certify that $f(x) < f(x')$ for every $x' \in S \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ (so as to report **b**). To design such an oracle we have at our disposal a procedure to realize the oracle $O_{\alpha,\delta}$ for any mixed-integer convex minimization problem of the kind (2.1) with $n = 2$. We propose a finite-time implementation of $O_{\alpha,\delta}$ 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:

$$f_{12}^\star := \min\{f(z, y) : (z, y) \in S \cap (\mathbb{Z}^2 \times \mathbb{R}^{(n-2)+d})\}$$

with the techniques developed in the previous section. If we can solve the partial minimization problems up to an accuracy of $\gamma \leq \delta/\kappa$, we obtain a point $(u_1^\star, u_2^\star, y_3^\star, \ldots, y_{n+d}^\star) \in S$ with $u_1^\star, u_2^\star \in \mathbb{Z}$ and for which:

$$\tilde{f}_{12}^\star := f(u_1^\star, u_2^\star, y_3^\star, \ldots, y_{n+d}^\star) \leq f_{12}^\star + \kappa\gamma$$

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

Then we fix $u_1^\star$ and $u_2^\star$ and solve (if $k \geq 4$; if $k = 3$, the necessary modifications are straightforward)

$$f_{1234}^\star := \min\{f(u_1^\star, u_2^\star, z, y) : (u_1^\star, u_2^\star, z, y) \in S \cap ((u_1^\star, u_2^\star) \times \mathbb{Z}^2 \times \mathbb{R}^{(n-4)+d})\}.$$

We obtain a point $(u_1^\star, \ldots, u_4^\star, y_5^\star, \ldots, y_{n+d}^\star) \in S$ with $u_i^\star \in \mathbb{Z}$ for $1 \leq i \leq 4$ and for which:

$$\tilde{f}_{1234}^\star := f(u_1^\star, \ldots, u_4^\star, y_5^\star, \ldots, y_{n+d}^\star) \leq f_{1234}^\star + \kappa\gamma \leq f^\star + \kappa\gamma.$$

Now, if $\tilde{f}_{1234}^\star - \kappa\gamma > f(x)$, we can make our oracle output **b**. Thus, we assume that $\tilde{f}_{1234}^\star - \kappa\gamma \leq f(x)$ and fix $u_i^\star$ 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):

$$\min\{f(u_1^\star, \ldots, u_{n-2}^\star, z, y) : (u_1^\star, \ldots, u_{n-2}^\star, z, y) \in S \cap ((u_1^\star, \ldots, u_{n-2}^\star) \times \mathbb{Z}^2 \times \mathbb{R}^d)\}.$$

Let $(u_1^\star, \ldots, u_n^\star, y^\star) \in \mathbb{Z}^n \times \mathbb{R}^d$ be an approximate optimal solution. If we cannot interrupt the algorithm, i.e., if $f(u_1^\star, \ldots, u_n^\star, y^\star) \not\leq (1 + \alpha)f(x) + \kappa\gamma$, we replace $(u_{n-3}^\star, u_{n-2}^\star)$ by the second best point for the corresponding mixed-integer convex minimization problem. In view of Proposition 2.3.9, 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 $(u_{n-1}^\star, u_n^\star)$ 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 $[-B, B]^n$.

Chapter 3

Cutting Plane Methods

This chapter is based on a joint paper with Christian Wagner and Robert Weismantel [OWW14]. The core of this chapter are two alternative, short, and geometrically motivated algorithms proving that integer convex minimization in fixed dimension can be polynomially reduced to mixed integer linear optimization. That is, provided that the dimension is fixed, we can reduce the latter problem to a polynomial number of mixed integer linear optimization problems and to a polynomial number of function evaluations, plus some polynomial work.

We aim at solving

$$\min\{f(x) \mid x \in \mathbb{Z}^n \text{ and } g(x) \leq 0\}, \tag{3.1}$$

where $f, g : \mathbb{R}^n \to \mathbb{R}$ are convex functions. We assume that the functions are given by a *first-order evaluation oracle*. Queried on a specific point such oracles return a function value and a subgradient of the subdifferential at this point. We further assume that (i) some $B \in \mathbb{N}$ is known satisfying $\{x \in \mathbb{Z}^n \mid g(x) \leq 0\} \subset [0, B]^n$, and (ii) the output of the first order evaluation oracles is of sufficient precision.

In order to quantify what we consider sufficient precision. Let ϵ and δ be given nonnegative constants. Throughout this chapter we consider them to be fixed. Note that, in Definition 1.1.1 the output of the first-order evaluation oracle is exact. For this chapter we define our *first-order evaluation oracle* as follows:

Definition 3.0.10 (first-order evaluation oracle)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex function presented by a first-order evaluation oracle. Then, queried on $\bar{x}$ the oracle returns $\bar{f}$ and $\bar{h}$ such that

$$|f(\bar{x}) - \bar{f}| \leq \epsilon$$

and

$$\bar{h} = 0 \qquad \qquad \text{if} \quad 0 \in \partial f(\bar{x})$$

$$\text{or} \quad \left\| \frac{h}{\|h\|_\infty} - \frac{\bar{h}}{\|\bar{h}\|_\infty} \right\|_\infty \leq \delta \qquad \text{for some } h \in \partial f(\bar{x}) \setminus \{0\},$$

where $\partial f(\bar{x})$ denotes the subdifferential of f at $\bar{x}$.

Since we are only interested in the separating property of the subgradients, we assume that, whenever a subgradient is nonzero, then it is normalized, i.e. $\|\bar{h}\|_\infty = 1$. That is, we assume to have an error on the function values such as on the subgradients. Most of the Section 3.1 will be dedicated to this error.

Further we assume to have excess to the following *mixed integer linear optimization oracle*.

Definition 3.0.11 (mixed integer linear optimization oracle)**.** The oracle returns an optimal solution when queried with a mixed integer linear optimization problem (MILP) with a fixed number of integer variables.

Our motivation for using such an oracle lies in the significant progress in developing efficient solution techniques for MILP's that has been achieved over the last decades. Today, one can solve MILP's that were considered out of reach twenty years ago. Moreover, if one intends to solve ICP's, it is natural to assume the existence of an oracle capable of solving easier optimization problems. Thus, it is plausible to postulate that the linear case can be solved.

Also, on the theoretical side, Lenstra proved in [Len83] that, given a system of linear inequalities with rational coefficients in fixed dimension, in polynomial time in the size of the encoding length of the input data one can either find an integral solution to the system, or show that all integral points of the solution set of the system lie on few parallel hyperplanes. Further, Lenstra showed that, given a system of linear inequalities with rational coefficientsn and the additional constraint that a fixed number of variables are required to be integral, in polynomial time in the size of the encoding length of the input data one can either find an mixed-integral solution to the system, or show that no such point exists.

Combined with binary search this yields a polynomial algorithm, realising a *mixed integer linear optimization oracle* as in Definition 3.0.11

The main result that is shown in this chapter is stated below.

Theorem 3.0.12. *Let $n \in \mathbb{N}$ be fixed. Let $B \in \mathbb{N}$, and $\delta, \epsilon \geq 0$ with $\delta \in \mathcal{O}(B^{-n})$ be given and satisfying the assumptions (i) and (ii). Assume to have at hand first-order evaluation oracles for f and g, and a mixed integer linear optimization oracle able to solve MILP's with at most n integer variables. Then problem (3.1) can be solved within a number of oracle calls bounded by a polynomial in the binary encoding of B. That is, we either find a point $\bar{x} \in \mathbb{Z}^n$ with $g(\bar{x}) \leq 2\epsilon$ and*

$$f(\bar{x}) \leq \min\{f(x) \mid x \in \mathbb{Z}^n \text{ and } g(x) \leq 0\} + 2\epsilon,$$

or show that $g(x) > 0$ for all $x \in \mathbb{Z}^n$.

We acknowledge that this result, in a slightly different setting, was first proven by Grötschel, Lovasz and Schrijver in [GLS88, Theorem 6.7.10]. There, they generalized the approach of Lenstra.

We point out that the accuracy of 2ϵ in Theorem 3.0.12 comes from the fact that the first-order evaluation oracles for f and g return the function value only with a precision of ϵ. When fed with a point $\bar{x} \in \mathbb{R}^n$, the first-order evaluation oracle for g returns a value $\bar{g}$ such that $|g(\bar{x}) - \bar{g}| \leq \epsilon$. Hence, if $\bar{g} \leq \epsilon$, then $g(\bar{x}) \leq 2\epsilon$, and if $\bar{g} > \epsilon$ then $g(\bar{x}) > 0$. Moreover, the evaluation oracle for f is, in general, not able to distinguish the function values of two points $\bar{x}, \bar{y} \in \mathbb{R}^n$ with $|f(\bar{x}) - f(\bar{y})| \leq 2\epsilon$. Thus, the accuracy in Theorem 3.0.12 is best possible, assuming evaluation oracles as given. We note that even enumerating all points in $[0, B]^n \cap \mathbb{Z}^n$ would not lead to a better accuracy.

The standard approach to solve (3.1) is to set $F_\gamma := f - \gamma$ and to solve the feasibility problem for the level-set $\{x \in \mathbb{Z}^n : F_\gamma(x) \leq 0 \text{ and } g(x) \leq 0\}$ while applying binary search on γ. In particular, this procedure is used in the original proof that (3.1) is solvable in oracle-polynomial time in [GLS88]. The drawback of this approach is that the underlying minimization problem (3.1) can only be solved up to a certain accuracy, even if the evaluation oracles provide exact output, i.e. $\epsilon = 0$ (and $\delta \in \mathcal{O}(B^{-n})$). Our methods solve (3.1) without binary search on the objective function value by only making use of the local information from oracle outputs. The virtue of our methods is that they solve (3.1) exactly when $\epsilon = 0$. Nevertheless, if $\epsilon > 0$, then all the approaches can approximate the optimal solution up to 2ϵ in polynomial time. Though, in the algorithm described in [GLS88], ϵ enters the run-time while the run-time of our approach is independent of ϵ.

To the best of our knowledge it has never been stated directly that problems of type (3.1) are oracle-polynomially solvable in fixed dimension. However, this result is derivable from the work of Lenstra [Len83] and Grötschel et al. [GLS88].

After the preliminaries we present in section 3.2 and 3.3 two constructive proofs for Theorem 3.0.12. Both use cutting plane methods. The first proof culminates in an algorithm that uses centroids – whose computation is time-consuming. The second proof results in an algorithm that avoids the computation of centroids at the expense of more iteration steps. The special feature of the second algorithm is that it only needs to solve MILP's as subproblems. This is of practical relevance as the computation of centroids is theoretically doable in fixed dimension, but intractable in practice. Moreover, to the best of our knowledge, it is not known how to accelerate the computation of centroids with an MILP oracle at hand.

Following the general cutting plane schemes presented in [Nes04, Section 3.2.6], we see that there are parallels between the development of solution techniques for continuous convex optimization and the introduced integral techniques. The Ellipsoid Method (see [Nes04, p. 154]) bears resemblance with the algorithms of [GLS88, KP00, Hei05, HK12] to solve (3.1), using ellipsoidal approximations. Furthermore, the Method of Centers of Gravity (see [Nes04, p. 152], or for a randomized version see [BV04a]) exhibits many similarities to our centroid algorithm in Section 3.2. Finally, the Kelly Method [Nes04, Section 3.3.2] and the Level Method [Nes04, Section 3.3.3] can be seen as analogues to our MILP algorithm in Section 3.3 in the sense that linear techniques are applied to solve non-linear problems.

3.1 Preliminaries

In this section, we present auxiliary lemmata and observations that are needed for the proofs in Sections 3.2 and 3.3.

Recall the notion of the lattice width and the flatness direction. See Section 1.2.

called *flatness direction for K*.

Observation 3.1.1. *Let $n \in \mathbb{N}$ be fixed and $P = \{x \in \mathbb{R}^n : Ax \leq b\} \subset \mathbb{R}^n$ be a rational polytope. Given a mixed integer linear optimization oracle in n integer variables, we can compute a flatness direction for P in polynomial time.*

Proof. W.l.o.g. let $\text{int}(P) \neq \emptyset$ and let $a_i^\top$ denote the i-th row vector of A. Further, we want to assume that $Ax \leq b$ has no redundant inequalities. By scaling the rows, we may assume that $b_i - \min_{x \in P} a_i x = 1$. Then $P_0 :=$

$P - P = \{x \in \mathbb{R}^n \mid -\mathbb{1} \leq Ax \leq \mathbb{1}\}$, where $\mathbb{1}$ denotes the all-one vector. Observe that $2\omega(P) = \omega(P_0)$. Let $P_0^\star := \{x \in \mathbb{R}^n \mid x^\top y \leq 1 \text{ for all } y \in P_0\}$ denote the *polar set* of P_0 and let $\|x\|_{P_0^\star} := \min\{\gamma \geq 0 \mid x \in \gamma P_0^\star\}$ denote the norm induced by $P_0^\star$. Then, we can reformulate the lattice width as follows

$$\omega(P_0) = 2 \min_{x \in \mathbb{Z}^n \setminus \{0\}} \max_{y \in P_0} x^\top y = 2 \min_{x \in \mathbb{Z}^n \setminus \{0\}} \|x\|_{P_0^\star}.$$

The latter minimization problem can be solved using MILP's. For that, note $\gamma P_0^\star = \gamma \operatorname{conv}\{\pm a_1, \ldots, \pm a_m\} = \{x \in \mathbb{R}^n \mid x = A^\top \lambda, \ \|\lambda\|_1 \leq \gamma\}$. For $i = 1, \ldots, n$ we solve the following MILP's that we call F^i.

$$
\begin{array}{lll}
\min & \gamma \\
\text{s.t.} & x = A^\top \lambda, & \|\lambda\|_1 \leq \gamma, \\
& x \in \mathbb{Z}^n, & x_i \geq 1, \\
& \lambda \in \mathbb{R}^m, & \gamma \in \mathbb{R}.
\end{array}
\qquad (F^i)
$$

Let $(\bar{\gamma}^i, \bar{x}^i, \bar{\lambda}^i)$ be an optimal solution of F^i and let $\bar{\gamma}^j := \min_{i=1,\ldots,n} \bar{\gamma}^i$. Then $\omega(P) = \bar{\gamma}^j$ and $\bar{x}^j$ is a flatness direction for P. □

The following lemma is similar to known results in literature, see [BLPS99] for instance. It states that a convex set is flat whenever its volume is sufficiently small. It thus defines the threshold at which to switch from adding cutting planes to enumerating lower-dimensional subproblems. As it is stated here we are not aware of a reference. This is why we outline a short proof.

Lemma 3.1.2. *Let $K \subset \mathbb{R}^n$ be a bounded convex set. If $\operatorname{vol}(K) < 1$ then*

$$\omega(K) \leq cn^{\frac{3}{2}}$$

for a universal constant c.

Proof. We show that K has a lattice-free translate, i.e. there exists a point $x \in \mathbb{R}^n$ such that $(x+K) \cap \mathbb{Z}^n = \emptyset$. Then $\omega(K) \leq cn^{\frac{3}{2}}$ for a universal constant c (see [BLPS99]).

Let χ_K denote the characteristic function of K and for a set $S \subset \mathbb{Z}^n$, $|S|$ denotes the cardinality of S. Assume that for all $x \in \mathbb{R}^n$ it holds that $|(x +$

$K) \cap \mathbb{Z}^n| \geq 1$. Then it follows that

$$1 > \mathrm{vol}(K) = \int_{\mathbb{R}^n} \chi_K(x) dx = \sum_{z \in \mathbb{Z}^n} \int_{[0,1)^n} \chi_K(x+z) dx$$

$$= \int_{[0,1)^n} \sum_{z \in \mathbb{Z}^n} \chi_K(x+z) dx$$

$$= \int_{[0,1)^n} \sum_{z \in \mathbb{Z}^n} \chi_{K-x}(z) dx$$

$$= \int_{[0,1)^n} |(K-x) \cap \mathbb{Z}^n| dx \geq 1,$$

a contradiction. $\qquad\square$

Given a point $\bar{x} \in [0,B]^n$ the first-order evaluation oracle provides us with a vector $\bar{h} \in \mathbb{R}^n$. Either $\bar{h} = 0$ or $\|\bar{h}\|_\infty = 1$ and there exists a $h \in \partial f(\bar{x})$ (resp. $\in \partial g(\bar{x})$) such that $\|\frac{h}{\|h\|_\infty} - \bar{h}\|_\infty \leq \delta$. In the following we want to discuss the error of the oracle, i.e. the value δ, of the second possible outcome. Let us assume that $\bar{h} \neq 0$. For our algorithms in Sections 3.2 and 3.3 we need to investigate how the the error of the oracle affects the volume, i.e. the ratio between $\mathrm{vol}(\{x \in P \mid h^\mathsf{T} x \leq h^\mathsf{T} \bar{x}\})$ and $\mathrm{vol}(\{x \in P \mid \bar{h}^\mathsf{T} x \leq \bar{h}^\mathsf{T} \bar{x}\})$ for a polyhedron $P \subset [0,B]^n$. For that, we need the following observation.

Observation 3.1.3. $M_1 := \{x \in [0,B]^n \mid h^\mathsf{T} x \leq h^\mathsf{T} \bar{x}\} \subset \{x \in [0,B]^n \mid \bar{h}^\mathsf{T} x \leq \bar{h}^\mathsf{T} \bar{x} + nB\delta\} =: M_2$.

Proof. Suppose that $x \in M_1 \setminus M_2$. Then $0 \leq -h^\mathsf{T}(x-\bar{x})$ and $nB\delta < \bar{h}^\mathsf{T}(x-\bar{x})$. Adding these inequalities yields $nB\delta < (x-\bar{x})^\mathsf{T}(\bar{h}-h) \leq \sum_{i=1}^n |x_i - \bar{x}_i| \cdot |\bar{h}_i - h_i| \leq nB\delta$, a contradiction. $\qquad\square$

The following lemma states a lower bound for the ratio between the volumes of $K \cap H'$ and K, provided that a lower bound for the ratio between the volumes of $K \cap H$ and K is known, where $K \subset [0,B]^n$ is a convex set and H and H' are half-spaces whose boundary hyperplanes are translates.

Lemma 3.1.4. *Let $K \subset [0,B]^n$ be a bounded convex set with $\mathrm{vol}(K) \geq 1$. Let $H := \{x \in \mathbb{R}^n \mid \alpha^\mathsf{T} x \leq \beta\}$ and $H' := \{x \in \mathbb{R}^n \mid \alpha^\mathsf{T} x \leq \beta - \kappa\}$ with $\|\alpha\|_\infty = 1$, $\beta \in \mathbb{R}$ and $\kappa \geq 0$. Moreover, let $\mathrm{vol}(K \cap H) \geq C\,\mathrm{vol}(K)$ for a constant $C > 0$. Let σ denote the volume of the five-dimensional unit ball. If $\kappa \leq \frac{C}{2\sigma}(\frac{2}{\sqrt{n}B})^{n-1}$, then*

$$\mathrm{vol}(K \cap H') \geq \frac{C}{2}\,\mathrm{vol}(K).$$

Proof. Note that the volume of an n-dimensional unit ball is maximal for $n = 5$. It holds that $\operatorname{vol}(K \cap H') = \operatorname{vol}(K \cap H) - (\operatorname{vol}(K \cap H) - \operatorname{vol}(K \cap H')) \geq C \operatorname{vol}(K) - (\operatorname{vol}(K \cap H) - \operatorname{vol}(K \cap H'))$. Let $S := \{x \in \mathbb{R}^n \mid \|x - \frac{B}{2}\mathbb{1}\|_2 \leq \frac{\sqrt{n}B}{2}\} \supset [0, B]^n$. Then

$$
\begin{aligned}
&\operatorname{vol}(K \cap H') - C \operatorname{vol}(K) \\
&\geq -\operatorname{vol}(\{x \in K \mid \beta - \kappa \leq \alpha^\mathsf{T} x \leq \beta\}) \\
&\geq -\operatorname{vol}(\{x \in [0, B]^n \mid \beta - \kappa \leq \alpha^\mathsf{T} x \leq \beta\}) \\
&\geq -\operatorname{vol}\left(\left\{x \in S \;\middle|\; \frac{\beta - \kappa}{\|\alpha\|_2} \leq \frac{\alpha^\mathsf{T} x}{\|\alpha\|_2} \leq \frac{\beta}{\|\alpha\|_2}\right\}\right) \\
&= -\int_{-\frac{\kappa}{\|\alpha\|_2}}^{0} \operatorname{vol}_{n-1}\left(\left\{x \in S \;\middle|\; \frac{\alpha^\mathsf{T} x}{\|\alpha\|_2} = \frac{\beta}{\|\alpha\|_2} - y\right\}\right) dy \\
&\geq -\frac{\kappa}{\|\alpha\|_2} \operatorname{vol}_{n-1}\left(\left\{x \in S \;\middle|\; \alpha^\mathsf{T} x = \alpha^\mathsf{T} \frac{B}{2}\mathbb{1}\right\}\right) \\
&\geq -\kappa \operatorname{vol}_{n-1}\left(\left\{x \in \mathbb{R}^{n-1} \;\middle|\; \|x\|_2 \leq \frac{\sqrt{n}B}{2}\right\}\right) \\
&= -\kappa \left(\frac{\sqrt{n}B}{2}\right)^{n-1} \operatorname{vol}_{n-1}\left(\{x \in \mathbb{R}^{n-1} \mid \|x\|_2 \leq 1\}\right) \\
&\geq -\kappa \left(\frac{\sqrt{n}B}{2}\right)^{n-1} \sigma \geq -\frac{C}{2}.
\end{aligned}
$$

The second and third inequality follow from the fact that $K \subset [0, B]^n \subset S$. For the first equation we apply Cavalieri's principle. Then, in the fourth inequality we use that $\operatorname{vol}_{n-1}\left(\{x \in S \mid \alpha^\mathsf{T} x = y\}\right)$ is maximal for $y = \alpha^\mathsf{T} \frac{B}{2}\mathbb{1}$. In the fifth inequality we exploit that $\|\alpha\|_2 \geq 1$ and that the $(n-1)$-dimensional ball

$$
\left\{x \in S \;\middle|\; \alpha^\mathsf{T} x = \alpha^\mathsf{T} \frac{B}{2}\mathbb{1}\right\}
$$

is equivalent, up-to rotation and translation, to

$$
\left\{x \in \mathbb{R}^{n-1} \mid \|x\|_2 \leq \frac{\sqrt{n}B}{2}\right\}.
$$

Finally, in the last inequality we use that the n-dimensional volume of an n-dimensional unit ball (i.e. $\pi^{n/2}/\Gamma(n/2 + 1)$) is maximal for $n = 5$.

Hence $\operatorname{vol}(K \cap H') \geq C(\operatorname{vol}(K) - \frac{1}{2}) \geq \frac{C}{2} \operatorname{vol}(K)$. $\square$

3.2 Cutting plane scheme based on centroids

In this section, we present our first algorithm to solve (3.1).

Let $K \subset \mathbb{R}^n$ be a compact convex set. The centroid of K is defined to be the point $c_K := \mathrm{vol}(K)^{-1} \int_K x dx$. In the case where K is a polytope, one possible way of computing the centroid is to triangulate K into simplices $S_1, \ldots, S_r$ and to compute the centroids $c_{S_1}, \ldots, c_{S_r}$ of the simplices. In turn, the centroid of a simplex S with vertices $v_0, \ldots, v_n$ is $c_S = \frac{1}{n+1} \sum_{i=0}^{n} v_i$. Finally, $c_K = \mathrm{vol}(K)^{-1} \sum_{i=1}^{r} c_{S_i} \cdot \mathrm{vol}(S_i)$. We note that the computation of a triangulation of the polytope K can be done in polynomial time in the number of vertices of K (see, for instance, [For97] or [DLRS10, Chapter 8]).

For a given compact convex body K and a $0 \leq \lambda \leq 1$ we define

$$K_\lambda := \lambda(K - c_K) + c_K, \tag{3.2}$$

i.e. the scaling of K by the factor λ with respect to its centroid. Note that $K_1 = K$ and $K_0 = \{c_K\}$. Again, in the case where K is a polytope and is given by linear inequalities, i.e. $K = \{x \in \mathbb{R}^n \mid Ax \leq b\}$, then $K_\lambda = \{x \in \mathbb{R}^n \mid Ax \leq \lambda b + (1 - \lambda)Ac_K\}$.

In the following lemma we give a straightforward generalization of a theorem of Grünbaum [Grü60, Theorem 2].

Lemma 3.2.1. *Let $0 \leq \lambda \leq 1$. Let $K \subset \mathbb{R}^n$ be a closed convex set, and let $H \subset \mathbb{R}^n$ be a half-space. If $K_\lambda \cap H \neq \emptyset$, then*

$$\mathrm{vol}(K \cap H) \geq (1 - \lambda)^n \cdot \left(\frac{n}{n+1}\right)^n \mathrm{vol}(K).$$

Proof. In [Grü60], Grünbaum defined the set

$$S := \{x \in \mathbb{R}^n \mid \text{for all half-spaces } G \text{ with } x \in G$$

$$\text{holds } \mathrm{vol}(K \cap G) \geq \left(\frac{n}{n+1}\right)^n \mathrm{vol}(K)\}.$$

In the proof of [Grü60, Theorem 2] it is shown that $c_K \in S$. This implies that if $c_K \in H$, then $\mathrm{vol}(K \cap H) \geq (\frac{n}{n+1})^n \mathrm{vol}(K)$. Note that $K = K_\lambda + K_{1-\lambda} - c_K$. Let $x \in K_\lambda \cap H$ and let $K^x := x + K_{1-\lambda} - c_K$. Then $K^x \subset K$ and $\mathrm{vol}(K \cap H) \geq \mathrm{vol}(K^x \cap H)$. Since $c_{K^x} = x$ we have $c_{K^x} \in H$. Hence $\mathrm{vol}(K^x \cap H) \geq (\frac{n}{n+1})^n \mathrm{vol}(K^x)$. We can rewrite $\mathrm{vol}(K^x)$ in terms of $\mathrm{vol}(K)$, namely $\mathrm{vol}(K^x) = (1 - \lambda)^n \mathrm{vol}(K)$. Then the lemma follows. $\square$

Observation 3.2.2. *If* $\text{int}(K_\lambda) \cap \mathbb{Z}^n = \emptyset$, *then*

$$\omega(K) = \frac{1}{\lambda}\omega(K_\lambda) \leq \frac{1}{\lambda}cn\frac{3}{2}$$

for a universal constant c (see [BLPS99]).

We are now ready to give a first algorithmic proof of Theorem 3.0.12.

Proof of Theorem 3.0.12. We follow the idea of the Method of Centers of Gravity in [Nes04, p. 152]. Our proof uses induction on n. We fix $\Lambda \in (0, 1)$. Further we assume that

$$\delta \leq \frac{1}{4\sigma\sqrt{n}}\left(\frac{2n(1-\Lambda)}{(n+1)\sqrt{n}B}\right)^n,$$

where σ denotes the volume of the five-dimensional unit ball. We define P^0 to be the box $[0, B]^n$, to which we add cutting planes until we can reduce the original problem to a small number of lower-dimensional subproblems. Among all points visited in the course of the algorithm, we keep record of the feasible point with smallest objective function value.

In the following, we construct a sequence of polytopes $P^0 \supset P^1 \supset P^2 \ldots$, such that P^{i+1} arises from P^i by intersecting P^i with a half-space $H = \{x \in \mathbb{R}^n \mid \bar{h}^\mathsf{T}x \leq \bar{h}^\mathsf{T}\bar{x} + nB\delta\}$. Here, $\bar{x}$ is an integral point of P^i and $\bar{h}$ is a vector provided by the first-order evaluation oracles. Note that, in order to avoid cutting off an optimal integral point – if any feasible integral point exists – we correct the oracle error δ by increasing the right hand side from $\bar{h}^\mathsf{T}\bar{x}$ to $\bar{h}^\mathsf{T}\bar{x} + nB\delta$ (see Observation 3.1.3)). Also, note that $P^i \cap \mathbb{Z}^n \neq \emptyset$ for all i.

The construction works as follows. Let $P^i = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be given, where $A \in \mathbb{R}^{m\times n}$ with rows $a_i^\mathsf{T} \in \mathbb{R}^n$ and $\|a_i\|_\infty = 1$ for all $i = 1, \ldots, m$.

We solve the mixed integer linear minimization problem

$$\begin{aligned}
\min \quad & \lambda \\
\text{s.t.} \quad & Ax + (Ac_{P^i} - b)\lambda \leq Ac_{P^i} \\
& \lambda \in \mathbb{R}_+ \\
& x \in \mathbb{Z}^n.
\end{aligned} \qquad \text{(MILP-1)}$$

Let (λ^*, x^*) be an optimal solution. Note that (MILP-1) always has a solution. Further, observe that $x^* \in P^i_{\lambda^*} = \{x \in \mathbb{R}^n \mid Ax + (Ac_{P^i} - b)\lambda^* \leq Ac_{P^i}\}$ and that $P^i_{\lambda^*}$ is lattice-free.

We distinguish two cases.

Case 1 If $\lambda^* > \Lambda$, then we compute a flatness direction $v \in \mathbb{Z}^n \setminus \{0\}$ for P^i (see Observation 3.1.1). Further, we compute $s := \left\lceil \min_{x \in P_i} v^\mathsf{T} x \right\rceil$ and $k := \left\lfloor \max_{x \in P_i} v^\mathsf{T} x \right\rfloor - s$. Let $H_j = \{x \in \mathbb{R}^n \mid v^\mathsf{T} x = s + j\}$, for $j = 1, \ldots, k$. It holds that $P^i \cap \mathbb{Z}^n \subset H_1 \cup \ldots \cup H_k$ and, by Observation 3.2.2,

$$k \leq \left\lfloor \frac{cn^{\frac{3}{2}}}{\lambda^*} \right\rfloor + 1 \leq \left\lfloor \frac{cn^{\frac{3}{2}}}{\Lambda} \right\rfloor + 1 =: \psi.$$

So we need to solve at most ψ subproblems of dimension $n - 1$. For all $j = 1, \ldots, k$, we solve the lower-dimensional problems $\min\{f(x) \mid x \in H_j \cap \mathbb{Z}^n \text{ and } g(x) \leq 0\}$. Among all feasible points, if they exist, we return the point with smallest objective function value. We have reduced our initial problem to a polynomial number of $(n-1)$-dimensional subproblems. By the induction hypothesis, all these subproblems can be solved in polynomial time.

Case 2 Let $\lambda^* \leq \Lambda$. Let $\bar{g}$ be the output of the first-order evaluation oracle of g at x^*. Depending on the value of $\bar{g}$, we use either the first-order evaluation oracle for f or the one for g to get $\bar{h}$: if $\bar{g} \leq \epsilon$ we use the first-order evaluation oracle for f or otherwise, if $\bar{g} > \epsilon$, we use the first-order evaluation oracle for g. Let $\bar{h}$ be the output of the first-order evaluation oracle at x^*. If $\bar{h} = 0$, then either x^* is the optimum solution or Problem 3.1 is infeasible. In both cases are be done; we either return x^* or state that $g(x) > 0$ for all $x \in \mathbb{Z}^n$. Hence let us assume that $\bar{h} \neq 0$. Then we define $P^{i+1} := P^i \cap \{x \in \mathbb{R}^n \mid \bar{h}^\mathsf{T} x \leq \bar{h}^\mathsf{T} x^* + nB\delta\}$. With $C = (1-\Lambda)^n \left(\frac{n}{n+1}\right)^n$ as in Lemma 3.2.1 and with $\kappa = nB\delta$, it follows from Lemma 3.1.4 that

$$\mathrm{vol}(P^{i+1}) \leq \left(1 - \frac{1}{2}(1 - \Lambda)^n \left(\frac{n}{n+1}\right)^n\right) \mathrm{vol}(P^i). \tag{3.3}$$

In particular, (3.3) guarantees that after at most

$$\Omega := \left\lfloor \frac{-\log(B^n)}{\log\left(1 - \frac{1}{2}(1 - \Lambda)^n \left(\frac{n}{n+1}\right)^n\right)} \right\rfloor + 1 \tag{3.4}$$

iterations we obtain a polytope P^l with $\mathrm{vol}(P^l) < 1$. We compute a flatness direction $v \in \mathbb{Z}^n \setminus \{0\}$ for P^l (see Observation 3.1.1). Next, we can construct as in Case 1 the value $s \in \mathbb{Z}$ and, by Lemma 3.1.2, a

$$k \leq \left\lfloor cn^{\frac{3}{2}} \right\rfloor + 1 =: \phi$$

and parallel hyperplanes $H_j = \{x \in \mathbb{R}^n \mid v^\mathsf{T} x = s + j\}$, $j = 1, \ldots, k$, such that $P^l \cap \mathbb{Z}^n \subset H_1 \cup \ldots \cup H_k$. So we need to solve at most ϕ subproblems of

dimension $n-1$. For all $j = 1, \ldots, k$, we solve the lower-dimensional problems $\min\{f(x) \mid x \in H_j \cap \mathbb{Z}^n \text{ and } g(x) \leq 0\}$. Among all feasible points, if they exist, we return the point with smallest objective function value. We have reduced our initial problem to a polynomial number of $(n-1)$-dimensional subproblems. By the induction hypothesis, all these subproblems can be solved in polynomial time.

Each iteration needs a constant number of oracle calls and the number of iterations is polynomial in $\log(B)$. $\qquad\square$

3.3 Cutting plane scheme based on mixed integer linear programs

In this section, we propose an alternative algorithm that avoids the computation of centroids. For that, we sacrifice on the fraction of volume decrease of our polytopes in every iteration. Let $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ be a full-dimensional polytope with $P \cap \mathbb{Z}^n \neq \emptyset$. We assume that $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ is a matrix with rows $a_i^\top$ and $\|a_i\|_\infty = 1$ for $i = 1, \ldots, m$. Further, we assume that $Ax \leq b$ has no redundant inequalities.

For $i = 1, \ldots, m$ let $l_i := \min_{x \in P} a_i^\top x$, and let $l := (l_1, \ldots, l_m)^\top$. Since P is bounded and $\text{int}(P) \neq \emptyset$, l_i exists and $l_i < b_i$ for all i. Consider the following problem in variables $x = (x_1, \ldots, x_n)^\top$ and λ.

$$
\begin{aligned}
\max \quad & \lambda \\
\text{s.t.} \quad & Ax + (b - l)\lambda \leq b \\
& \lambda \in \mathbb{R}_+ \\
& x \in \mathbb{Z}^n.
\end{aligned}
\qquad \text{(MILP-2)}
$$

Since $P \cap \mathbb{Z}^n \neq \emptyset$, (MILP-2) has an optimal solution. We will use (MILP-2) to replace (MILP-1).

The following observation relates feasible points of (MILP-2) with the difference body of P.

Observation 3.3.1. $\{(\lambda, x) \in \mathbb{R}_+ \times \mathbb{R}^n \mid Ax + (b - l)\lambda \leq b\} = \{(\lambda, x) \in \mathbb{R}_+ \times \mathbb{R}^n \mid x + \lambda(P - P) \subset P\}$.

Proof. See Figure 3.1 for an illustration.

Let $(\lambda, x) \in \mathbb{R}_+ \times \mathbb{R}^n$.

Assume that $Ax + (b - l)\lambda \leq b$. For any $z \in P - P$ there exist $x_1, x_2 \in P$ such that $z = x_1 - x_2$ and $l \leq Ax_1 \leq b$, $l \leq Ax_2 \leq b$. It follows that $A(x + \lambda z) = Ax + \lambda A(x_1 - x_2) \leq Ax + \lambda(b - l) \leq b$.

Assume that $x + \lambda(P - P) \subset P$. Then for each $i = 1, \ldots, m$ there exists a pair $x_1, x_2 \in P$ such that $a_i^\mathsf{T} x_1 = b_i$ and $a_i^\mathsf{T} x_2 = l_i$ (we assume that there are no redundant inequalities). Hence, $a_i^\mathsf{T} x + (b_i - l_i)\lambda = a_i^\mathsf{T} x + (a_i^\mathsf{T} x_1 - a_i^\mathsf{T} x_2)\lambda = a_i^\mathsf{T}(x + \lambda(x_1 - x_2)) \leq b_i$. $\qquad\square$

Let (λ, x) be a feasible point with respect to MILP-2. Observe that $Ax \leq b - \lambda(b - m) = (1-\lambda)b + \lambda m \leq (1-\lambda)b + \lambda Ac_P$. Hence, $x \in P_{1-\lambda}$, where $P_{1-\lambda}$ is defined as in (3.2). It is now straightforward to adapt Lemma 3.2.1. Let $H \subset \mathbb{R}^n$ be a half-space. If $x \in H$, then $\mathrm{vol}(P \cap H) \geq \lambda^n \cdot \left(\frac{n}{n+1}\right)^n \mathrm{vol}(P)$. However, we can do better. The next lemma is a suprior analogue to Lemma 3.2.1 for the new algorithm in this section.

Lemma 3.3.2. *Let $(\lambda, x) \in \mathbb{R}_+ \times \mathbb{Z}^n$ be a feasible point of* (MILP-2)*, and let $H \subset \mathbb{R}^n$ be a half-space. If $x \in H$, then $\mathrm{vol}(P \cap H) \geq 2^{n-1}\lambda^n \mathrm{vol}(P)$.*

Proof. By Observation 3.3.1, we have $x + \lambda(P - P) \subset P$. Furthermore, due to the central symmetry of the difference body $P - P$, we have

$$
\begin{aligned}
\mathrm{vol}(P \cap H) &\geq \mathrm{vol}\Big(\big(x + \lambda(P - P)\big) \cap H\Big) \\
&\geq \frac{1}{2}\,\mathrm{vol}\big(\lambda(P - P)\big) \\
&= \frac{\lambda^n}{2}\,\mathrm{vol}(P - P) \geq 2^{n-1}\lambda^n\,\mathrm{vol}(P).
\end{aligned}
$$

See Figure 3.1.

The last inequality follows from the Brunn-Minkowski inequality (see, for instance, Gruber [Gru07, Theorem 8.5]), stating that $2^n \mathrm{vol}(P) \leq \mathrm{vol}(P - P)$. $\qquad\square$

The following lemma is an analogue to Observation 3.2.2.

Lemma 3.3.3. *Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ and let $P' = \{x \in \mathbb{R}^n \mid Ax + (b - l)\lambda \leq b\}$ for some $\lambda \in [0, \frac{1}{2n})$. If $\mathrm{int}(P') \cap \mathbb{Z}^n = \emptyset$ then*

$$
\omega(P) \leq \frac{cn^{\frac{5}{2}}}{1 - 2\lambda n}
$$

for a universal constant c.
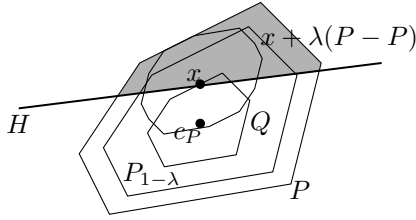
Figure 3.1: This figure illustrates the correlation between P, $Q = \{y \in \mathbb{R}^n \mid Ay \leq b - \lambda(b-m)\}$, $\lambda(P_P)$, $P_{1-\lambda}$ and the centroid c_P. Further, it shows the main proof idea of Lemma 3.3.2, i.e. that a half space containing $x \in Q$ contains half of $x + \lambda(P - P)$.

Proof. Since $\text{int}(P') \cap \mathbb{Z}^n = \emptyset$ it holds that $\omega(P') \leq cn^{\frac{3}{2}}$ for a universal constant c (see [BLPS99]).

By Observation 3.3.1, $P' = \{x \in \mathbb{R}^n \mid x + \lambda(P - P) \subset P\}$. By John's characterization of inscribed ellipsoids of maximal volume (see John [Joh48] and Ball [Bal92]), there exists an ellipsoid E centered at the origin, and a point q such that $q + E \subset P \subset q + nE$. By the definition of the difference body $P - P$, it follows that $2E = E - E \subset P - P \subset nE - nE = 2nE$. This implies $\lambda(P-P) \subset 2\lambda nE$ and thus $(1 - 2\lambda n)E + \lambda(P-P) \subset E \subset P - q$. Hence, $q + (1 - 2\lambda n)E + \lambda(P - P) \subset P$. This implies that $q + (1 - 2\lambda n)E \subset P'$. Thus, we obtain $P \subset q + nE \subset q + \frac{n}{1-2\lambda n}(P' - q)$. Hence $\omega(P) \leq \frac{n}{1-2\lambda n}\omega(P')$. $\square$

We now give an alternative proof of Theorem 3.0.12.

Proof of Theorem 3.0.12. The main structure remains equivalent to the proof in Section 3.2. This time we set the threshold value $\Lambda \in (0, \frac{1}{2n})$. Further we assume that

$$\delta \leq \frac{1}{8\sigma\sqrt{n}} \left(\frac{4\Lambda}{\sqrt{n}B} \right)^n,$$

where σ denotes the volume of the five-dimensional unit ball. We replace (MILP-1) by (MILP-2). Let $(\lambda^\star, x^\star)$ be an optimal solution of (MILP-2). We define $P_{\lambda^\star}^i := \{x \in \mathbb{R}^n \mid Ax + (b-l)\lambda^\star \leq b\}$. Again, we distinguish two cases.

In Case 1, if $\lambda^\star \leq \Lambda$, we apply Lemma 3.3.3. It follows that we have to solve at most ψ subproblems, where

$$\psi := \left\lfloor \frac{cn^{\frac{5}{2}}}{1 - 2\lambda^\star n} \right\rfloor + 1 \leq \left\lfloor \frac{cn^{\frac{5}{2}}}{1 - 2\Lambda n} \right\rfloor + 1.$$

In Case 2, if $\lambda^\star > \Lambda$, we set $C = 2^{n-1}\Lambda^n$ as in Lemma 3.3.2 and $\kappa = nB\delta$. Then we apply Lemma 3.1.4. Thus, we ensure to reduce the volume of P^i by a constant factor of $1 - 2^{n-2}\Lambda^n$. This guarantees that after at most

$$\Omega := \left\lfloor \frac{-\log(B^n)}{\log(1 - (1 - 2^{n-2}\Lambda^n))} \right\rfloor + 1$$

iterations we obtain a polytope P^k with $\mathrm{vol}(P^k) < 1$. Again, each iteration needs a constant number of oracle calls and the number of iterations is polynomial in $\log(B)$. $\hfill\square$

3.4 Extension to the mixed-integer setting

Let $F, G : \mathbb{Z}^n \times \mathbb{R}^d \to \mathbb{R}$ be convex functions given by first-order evaluation oracles. Then the general mixed-integer convex minimization problem is

$$\min\{F(x,y) \mid (x,y) \in \mathbb{Z}^n \times \mathbb{R}^d \text{ and } G(x,y) \leq 0\}.$$

We assume that the problem is bounded, i.e. $\{(x,y) \in \mathbb{Z}^n \times \mathbb{R}^d \mid G(x,y) \leq 0\} \subset [0,B]^{n+d}$. We define for $x \in \mathbb{Z}^n$ the function

$$g(x) := \min\{G(x,y) \mid y \in [0,B]^d\} \tag{3.5}$$

and for every $x \in \mathbb{Z}^n$ with $g(x) \leq 0$ we define

$$f(x) := \min\{F(x,y) \mid y \in [0,B]^d \text{ with } G(x,y) \leq 0\}. \tag{3.6}$$

In order to solve the mixed integer problem with a pure integer algorithm we emulate the first-order evaluation oracles of f and g with a *continuous convex minimization oracle*, as outlined below. For that we omit the errors ϵ and δ.

Function g. Let $\bar{x} \in \mathbb{Z}^n$ and let $y^\star$ be an optimal solution of (3.5). We assume that the oracle returns a $\bar{G} = g(\bar{x}) = G(\bar{x}, y^\star)$ and a $\bar{H} \in \mathbb{R}^{n+d}$. Either $\bar{H} = 0$ if $0 \in \partial G$ or, otherwise, $\|\bar{H}\|_\infty = 1$ such that there exists some H in the subdifferential of G at $(\bar{x}, y^\star)$, (i.e. $H \in \partial G(\bar{x}, y^\star)$) with $H_i = 0$ for $i = n+1, \ldots, n+d$, and $\bar{H} = \frac{1}{\|H\|_\infty} H$.

Function f. Let $\bar{x} \in \mathbb{Z}^n$ be feasible and let $y^\star$ be an optimal solution of (3.6). We assume that the oracle returns a $\bar{F} = f(\bar{x}) = F(\bar{x}, y^\star)$ and a $\bar{H} \in \mathbb{R}^{n+d}$. We distinguish two cases:
If $G(\bar{x}, y^\star) < 0$ then $\bar{H}$ is equal to zero if $0 \in \partial F$ or, otherwise, $\|\bar{H}\|_\infty = 1$

such that there exists $H \in \partial F(\bar{x}, y^\star)$, with $H_i = 0$ for $i = n + 1, \ldots, n + d$, and $\bar{H} = \frac{1}{\|H\|_\infty} H$.

If $G(\bar{x}, y^\star) = 0$ then there exists a $H_G \in \partial G(\bar{x}, y^\star)$, $H_F \in \partial F(\bar{x}, y^\star)$ and an $\alpha \geq 0$ such that $(H_F + \alpha H_G)_i = 0$ for $i = n + 1, \ldots, n + d$, and

$$\bar{H} = \begin{cases} 0, & \text{if } H_F + \alpha H_G = 0, \\ \frac{1}{\|H_F + \alpha H_G\|_\infty} (H_F + \alpha H_G) & \text{otherwise.} \end{cases}$$

Note that, the link between H_F and H_G arises from the Karush-Kuhn-Tucker conditions.

Realizations of such a continuous convex minimization oracle exists provided that we do not omit the errors ϵ and δ. Examples for such realizations can be found, for instance, in [Nes04].

Chapter 4

Center-Points

This chapter is partially based on joint work with Kalle Klimkewitz and
Robert Weismantel. Some parts are published in Kalle Klimkewitz's Master's Thesis [Kli14]. Again, our aim is to minimize a convex function f over
a vector-field $\mathbb{E}^n$, i.e.

$$\min_{x \in \mathbb{E}^n} f(x).$$

It is also possibly to add convex constraints $g(x) \leq 0$.

Our main goal in this chapter is to revisit the Method of Center of Gravity
and generalize it. See [Nes04, Section 3.2.6.], or for a randomized version see
[BV04a]. The main difference between Section 3.2 and the material developed
here is that we introduce a general measure function. With respect to integer
convex minimization we will use the number of integer points in polytopes as
a measure. We construct a series of polytopes always containing the optimal
solution and such that the measure decreases in each step by a constant fraction. One advantage of this new approach is that we can avoid to enumerate
lower dimensional sub-problems.

We begin our discussions by introduceing the general terms of center-points
and approximate center-points of sets. Those are points that guarantee that
any half-space containing them *cuts* off a certain part of the set. This approach
of generating iteratively center-points will allow us to develop a general theory
for convex continuous, integer, and mixed integer optimization. Whereas the
continuous case has been well studied and leads to the Method of Center of
Gravity, the other two approaches seem to be novel. Common features of the
new approaches are (i) that they avoid any kind of enumeration (at least on
a high level) and (ii) that they lead to an optimal algorithm in the sense that
the worst case number of calls of a first-order evaluation oracle is minimized.
Further, in the mixed integer setting another novelty is that we can treat
integer and continuous variables simultaneously.

The remaining Chapter is organized as follows. We start with the first sec-

tion, introducing center-points and approximate center-points, proving their existence and proving lower bounds for several instances of $\mathbb{E}^n$. In the second section we introduce two algorithms to compute approximate center-points in $\mathbb{Z}^n$. In Section 4.3 we introduce optimality certificates for convex (mixed-) integer optimization problems. Then, in the last section, we discuss an optimization scheme that terminates with, or at leasts converges to, the latter optimality certificates.

4.1   Center-points and their existence

In this section we extend results of Grünbaum [Grü60] which itself are extensions of results of Neumann [Neu45] and Eggleston [Egg53, Egg57, Egg58] for the two dimensional case. For that we introduce the definition of a *k-hull*, which is a generalization of a definition first introduced by Cole et al. in [CSY87].

Let $\mathbb{E}^n$ denote a vector-space equipped with an inner product $\langle \cdot, \cdot \rangle$. This allows us to define half-spaces, that is, given an element of the vector-space $a \in \mathbb{E}^n$ and a scalar $b \in \mathbb{E}$ we can define the half-space $H := \{x \in \mathbb{E}^n \mid \langle a, x \rangle \leq b\}$.

Definition 4.1.1 (*k*-Hull)**.** Let μ be a finite measure on $\mathbb{E}^n$ and let $S \subset \mathbb{E}^n$ be a compact subset with a non-zero measure. For a $k \in [0,1]$ we define the *k-hull* of S as

$$\mathrm{hull}_k^\mu(S) = \left\{ x \in \mathbb{E}^n \ \middle| \ \forall \text{ half-spaces } H \ni x \text{ it holds } \frac{\mu(H \cap S)}{\mu(S)} \geq k \right\}.$$

In other words, the k-hull defines the subset of the vector-space such that any half-space containing any point of this set cuts of a k-fraction of S with respect to the measure μ. Depending on our vector-space $\mathbb{E}^n$ we will work with different measures. We will consider the Lebesgue measure for $\mathbb{R}^n$, i.e. the volume, the counting measure for $\mathbb{Z}^n$ and the d-dimensional Lebesgue measure for $\mathbb{Z}^n \times \mathbb{R}^d$.

We are interested in the optimum of the following maximization problem:

$$\begin{aligned}
\max \quad & k \\
\text{s.t.} \quad & x \in \mathrm{hull}_k^\mu(S) \\
& x \in \mathbb{E}^n \\
& k \in [0,1].
\end{aligned} \tag{4.1}$$

An optimal solution $(k^\star, x^\star)$ of (4.1) minimizes the worst case, that is $x^\star$ is a solution of

$$\max_{x \in \mathbb{E}^n} \min_{\substack{\text{half-space } H \\ x \in H}} \frac{\mu(H \cap S)}{\mu(S)}.$$

That is, a half-space containing $x^\star$ automatically contains a *largest possible* fraction of S with respect to the measure. In view of optimization, center-points guarantee a priori that we can significantly reduce the region in which the optimal solution lies. We synthesize this in the following definition of *center-points*.

Definition 4.1.2 (Center-Point)**.** Let μ be a finite measure on $\mathbb{E}^n$ and let $S \subset \mathbb{E}^n$ be a compact subset with positive measure. If $(k^\star, x^\star)$ is optimal with respect to (4.1), then we call $x^\star$ a *center-point* of S.

Given a $k \in [0, 1]$, trying to solve the feasibility problem, amounts to finding a point in $\text{hull}_k^\mu(S)$ or determining that $\text{hull}_k^\mu(S)$ is empty. This problem is already hard on its own. On the one hand we will address the question of determining and characterizing $k^\star$. On the other hand we will present two algorithms that can solve the feasibility problem in $\mathbb{Z}^n$ for special values of k.

Notation 4.1.3. A point $x \in \text{hull}_k^\mu(S)$ we call a *k-approximate center-point*.

In the remainder of this section we consider the three cases where the vector space equals to $\mathbb{R}^n$, $\mathbb{Z}^n$ or $\mathbb{Z}^n \times \mathbb{R}^d$.

4.1.1 Existence of center-points in $\mathbb{R}^n$

For the special setting where S is a bounded convex set and where μ is the Lebesgue measure, the centroid gives us a particularly interesting center-point. Recall the definition. The *centroid* of a compact convex set $K \subset \mathbb{R}^n$, is defined as

$$c_K := \text{vol}(K)^{-1} \int_K x dx.$$

Grünbaum proved the following theorem.

Theorem 4.1.4. (Grünbaum, [Grü60, Theorem 2]) *Let μ be the Lebesgue measure on $\mathbb{R}^n$ and let $K \subset \mathbb{R}^n$ be a compact convex set. Then*

$$c_K \in \text{hull}_{\left(\frac{n}{n+1}\right)^n}^\mu(K).$$

Or in other terms:
*Let $K \subset \mathbb{R}^n$ be a compact convex set, and let $H \subset \mathbb{R}^n$ be a half-space. If c_K,
the centroid of K, is in H, then*

$$\mathrm{vol}(K \cap H) \geq \left(\frac{n}{n+1} \right)^n \mathrm{vol}(K).$$

Note that $\lim_{n \to \infty} \left(\frac{n}{n+1} \right)^n = e^{-1}$, i.e. the bound is essentially independent
of the dimension n. This bound is best possible when K is a simplex.

Corollary 4.1.5. *Let $K \subset \mathbb{R}^n$ be a closed convex set and let $f : K \mapsto \mathbb{R}_+$ be
a concave, nonnegative and Riemann measurable function. Then there exists
a point $x \in K$ such that for every half-space H with $x \in H$, it holds that*

$$\int_{K \cap H} f(x)dx \geq \left(\frac{n+1}{n+2} \right)^{n+1} \int_K f(x)dx.$$

Proof. We define $K' \subset \mathbb{R}^{n+1}$ as follows,

$$K' := \{(x,y) \in \mathbb{R}^{n+1} \mid x \in K \text{ and } 0 \leq y \leq f(x)\}.$$

K' is a closed convex set. From Theorem 4.1.4 it follows that there exists a
point $(x^*, y^*) \in \mathbb{R}^n \times \mathbb{R}$ such that $\mathrm{vol}(K' \cap H') \geq (\frac{n+1}{n+2})^{n+1} \mathrm{vol}(K')$ for any
half-space $H' \in \mathbb{R}^{n+1}$ containing (x^*, y^*). In particular this holds for any
half-space $H = \{x \in \mathbb{R}^{n+1} \mid u^\mathsf{T} x \leq w_u\}$ containing x^* and with $u_{n+1} = 0$.
Observing that $\mathrm{vol}(K') = \int_K f(x)dx$ and $\mathrm{vol}(K' \cap H') = \int_{K \cap H} f(x)dx$ the
Corollary follows. $\square$

Grünbaum also proved the following more general Theorem.

Theorem 4.1.6. (Grünbaum, [Grü60, Theorem 1]) *Let μ be a finite measure
on $\mathbb{R}^n$ and let $K \subset \mathbb{R}^n$ be compact. Then*

$$\mathrm{hull}^\mu_{\frac{1}{n+1}}(K) \neq \emptyset.$$

4.1.2   Existence of center-points in $\mathbb{Z}^n$

We now consider the case when $\mathbb{E}^n = \mathbb{Z}^n$. To illustrate that there is no direct
connection between center-points in $\mathbb{R}^n$ and $\mathbb{Z}^n$ consider $K = \mathrm{conv}((-10,1)^\mathsf{T},
(0,1)^\mathsf{T}, (0,0)^\mathsf{T}, (10,0)^\mathsf{T})$. For $\mathbb{R}^2$ consider the Lebesgue measure and for $\mathbb{Z}^2$

the counting measure, i.e. the number of integers points in a set. Then, in case of $\mathbb{R}^2$ together with the Lebesgue measure, the center-point of K is uniquely determinate by the point $\{(0, 0.5)^\mathsf{T}\}$, however, if we chose $\mathbb{Z}^2$ and the counting measure, then the center-points of $K \cap \mathbb{Z}^2$ are $(5, 0)^\mathsf{T}$ and $(-5, 1)^\mathsf{T}$ (see Figure 4.1).

The core argument in the proof of Theorem 4.1.6 is Helly's Theorem (e.g. [Gru07, Theorem 3.2.]). To achieve a similar result, we need the following Helly-type theorem for $\mathbb{Z}^n$.

Theorem 4.1.7. (Doignon, [Doi73, Proposition 4.2]) *Let* $K_1, \ldots, K_m \subset \mathbb{R}^n$ *be a family of closed convex sets. Suppose for every subfamily* $K_{i_1}, \ldots, K_{i_l}$ *with* $l \leq 2^n$

$$\bigcap_{j=1}^{l} (K_{i_j} \cap \mathbb{Z}^n) \neq \emptyset.$$

Then

$$\bigcap_{j=1}^{m} (K_j \cap \mathbb{Z}^n) \neq \emptyset.$$

With this we can now prove an analogue of Theorem 4.1.4.

Theorem 4.1.8. *Let* μ *be a finite measure on* $\mathbb{Z}^n$ *and let* $K \subset \mathbb{Z}^n$ *be a finite set. Then*

$$\text{hull}^{\mu}_{\frac{1}{2^n}} (K) \neq \emptyset.$$

Proof. We assume without loss of generality that $\mu(K) = 1$. Let $S^{n-1} := \{u \in \mathbb{R}^n \mid u^\mathsf{T} u = 1\}$ denote the unit sphere. For $u \in S^{n-1}$ and $w \in \mathbb{R}$ we denote by $H(u, w)$ the half-space $\{x \in \mathbb{Z}^n \mid u^\mathsf{T} x \leq w\}$. We define $g : S^{n-1} \times \mathbb{R} \mapsto [0, 1]$ as

$$g(u, w) := \mu(K \cap H(u, w)).$$

Note that g is right-continuous in w. Next, for each $u \in S^{n-1}$ we choose $w_u \in \mathbb{R}$ such that $g(u, w_u) > 1 - \frac{1}{2^n}$ and such that for any $\varepsilon > 0$ it holds that
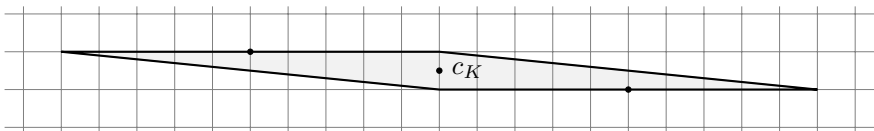
Figure 4.1: An example where the center-points can vary significantly when one changes the vector spaces and the measures.

$g(u, w_u - \varepsilon) \leq 1 - \frac{1}{2^n}$. Assume there exists a $x \in \bigcap_{u \in S^{n-1}} H(u, w_u)$. Then for any half-space H containing x we know that $\mu(K \cap H) \geq \frac{1}{2^n}$. Hence, proving Theorem 4.1.8 is equivalent to proving

$$\bigcap_{u \in S^{n-1}} H(u, w_u) \neq \emptyset. \tag{4.2}$$

The set K is a finite set of integers. The number of different partitions of K into two parts is upper bounded by $2^{|K|}$. We are interested in all partitions $K = K_1 \cup K_2$ such that $K_1 \subset H(u, w_u)$ and $K_2 \subset H(-u, -w_u)$. For each such partition we select one representative $u \in S^{n-1}$. Let us denote them by $u_1, u_2, ..., u_m$. Then

$$\bigcap_{u \in S^{n-1}} H(u, w_u) = \bigcap_{i=1}^{m} H(u_i, w_{u_i}). \tag{4.3}$$

In order to apply Theorem 4.1.7, consider any subset of $\{1, \ldots, m\}$ of cardinality less than or equal to 2^n. Without loss of generality, $\{1, \ldots, l\}$ with $l \leq 2^n$. Then, by the definition of $H(u_1, w_{u_1})$, we know that $\mu(K \cap H(u_1, w_{u_1})) > 1 - \frac{1}{2^n}$. By the additivity of measures we have that $\mu(K \cap H(u_1, w_{u_1}) \cap H(u_2, w_{u_2})) > 1 - \frac{2}{2^n}$. Hence,

$$\mu \left(K \bigcap_{i=1}^{l} H(u_i, w_{u_i}) \right) > 1 - \frac{l}{2^n} \geq 0.$$

In particular, this implies that $\bigcap_{i=1}^{l} H(u_i, w_{u_i}) \neq \emptyset$. Applying (4.3) and Theorem 4.1.7 we obtain (4.2). $\qquad \square$

Choosing μ to be the counting measure, i.e. for $S \subset \mathbb{Z}^n$

$$\mu(S) = \begin{cases} |S|, & \text{if } S \text{ is finite}, \\ \infty, & \text{otherwise}, \end{cases}$$

we obtain the following particularly interesting case.

Corollary 4.1.9. *Let $K \subset \mathbb{R}^n$ be a compact convex set. Then there exists a point $x \in K \cap \mathbb{Z}^n$, such that for any half-space $H \ni x$*

$$|K \cap H \cap \mathbb{Z}^n| \geq \frac{1}{2^n} |K \cap \mathbb{Z}^n|.$$

Remark 4.1.10. The bound in Corollary 4.1.9 is tight for $K = [0, 1]^n$. Furthermore it is possible to give examples of convex bodies containing an arbitrary number of lattice points such that the bound is tight. For that, consider the following example. Let $n \geq 2$ and let $K = [0, 1]^{n-1} \times [0, 2m-1]$ with $m \in \mathbb{N}$. Then, for any point $x \in K \cap \mathbb{Z}^n$ there exists a half-space H such that $|K \cap H \cap \mathbb{Z}^n| \leq \frac{1}{2^n} |K \cap \mathbb{Z}^n| = m$.

There is a natural connection between the volume of a convex set and the number of integer points contained within. That is, for convex bodies with a large lattice width, the volume and the number of integer points contained in the body is approximately the same. This allows us to adapt Theorem 4.1.4 to the counting measure on $\mathbb{Z}^n$. We quantify this in Lemma 4.1.13 and in Lemma 4.1.16.

For Lemma 4.1.13 we first need the following auxiliary lemma.

Lemma 4.1.11. *Let $K \subset \mathbb{R}^n$ be a closed convex set. If there exists an affine unimodular transformation U such that $U(c[0,1]^n) \subset K$, where $c \in \mathbb{N}$ and $c > 0$, then*

$$\left(1 - \frac{1}{c}\right)^n \leq \frac{|K \cap \mathbb{Z}^n|}{\mathrm{vol}(K)} \leq \left(1 + \frac{1}{c}\right)^n$$

Proof. Without loss of generality assume that $c[0,1]^n \subset K$, i.e. U is the identity matrix. Let $b := \frac{1}{2c}(1, \ldots, 1)^T$ and let $B := \frac{1}{2}[-1, 1]^n$. Further we define $\bar{K} := (K \cap \mathbb{Z}^n) + B$ and $K_\lambda := (1 + \lambda)(K - b) + b$ where $\lambda \in \mathbb{R}$ and $\lambda > -1$. Then $|K \cap \mathbb{Z}^n| = \mathrm{vol}(\bar{K}) \leq \mathrm{vol}(K + B)$. It remains to observe that $K + B \subset K_{1/c}$. Hence $|K \cap \mathbb{Z}^n| = \mathrm{vol}(\bar{K}) \leq \mathrm{vol}(K_{1/c}) = (1 + 1/c)^n \, \mathrm{vol}(K)$. In order to prove the lower bound assume that there exists an $x \in K_{-1/c} \setminus \bar{K}$. We define $z := \lfloor x_i \rceil_U$.[1] Then, since $\|x - z\|_K \leq 1/c$,[2] the point z must be in $K \cap \mathbb{Z}^n$. This contradicts $x \notin z + B \subset \bar{K}$. Hence $|K \cap \mathbb{Z}^n| = \mathrm{vol}(\bar{K}) \geq \mathrm{vol}(K_{-1/c}) = (1 - 1/c)^n \, \mathrm{vol}(K)$. □

Remark 4.1.12. This bound is tight. For example, let $K = c[0,1]^n$.

Lemma 4.1.13. *Let $K \subset \mathbb{R}^n$ be a compact convex set. If there exists a $c \in \mathbb{N}$, $c > 0$ and a unimodular matrix U such that $\lfloor cK \rceil_U + \left(c + \frac{1}{2}\right)[-1, 1]^n \subset K$, then, for any half-space $H \ni \lfloor cK \rceil_U$*

$$|K \cap H \cap \mathbb{Z}^n| \geq \left(\frac{n}{n+1}\right)^n \left(\frac{4c^2 - 4c}{4c^2 + 4c + 1}\right)^n |K \cap \mathbb{Z}^n|.$$

---

[1] Let $x \in \mathbb{R}^n$ and let U be a unimodular matrix. With $\lfloor x \rceil$ we denote the point $z \in \mathbb{Z}^n$ such that for each component $-\frac{1}{2} < z_i - x_i \leq \frac{1}{2}$. With $\lfloor x \rceil_U$ we then denote $U \cdot \lfloor U^{-1}x \rceil$.

[2] With $\|x\|_K$ we denote the norm induced by $\frac{1}{2}(K - K)$, i.e. $\|x\|_K = \min\left\{\lambda \geq 0 \mid x \in \frac{\lambda}{2}(K - K)\right\}$

Proof. Without loss of generality we assume that U is equal to the identity matrix. Hence $\lfloor c_K \rfloor_U = \lfloor c_K \rfloor$. Let $H \subset \mathbb{R}^n$ be an arbitrary half-space containing $\lfloor c_K \rfloor$. Further, let $\bar{K} := \lfloor c_K \rfloor + \frac{2c}{2c+1}(K - c_K)$. It holds that $\bar{K} \subset K$ and $c_{\bar{K}} = \lfloor c_K \rfloor$. Thus, together with Theorem 4.1.4, it follows that

$$\text{vol}(K \cap H) \geq \text{vol}(\bar{K} \cap H) \geq \left(\frac{n}{n+1} \right)^n \text{vol}(\bar{K}) = \left(\frac{n}{n+1} \right)^n \left(\frac{2c}{2c+1} \right)^n \text{vol}(K).$$
(4.4)

Recall that K contains a translation of $2c[0,1]^n$ and note that $\bar{K} \cap H$, and in particular $K \cap H$, contains a translation of $cn[0,1]^n$. Then, due to Lemma 4.1.11, we can rewrite (4.4) as

$$\left(\frac{c}{c-1} \right)^n |K \cap H \cap \mathbb{Z}^n| \geq \left(\frac{n}{n+1} \right)^n \left(\frac{2c}{2c+1} \right)^n \left(\frac{2c}{2c+1} \right)^n |K \cap \mathbb{Z}^n|.$$

Hence,

$$|K \cap H \cap \mathbb{Z}^n| \geq \left(\frac{n}{n+1} \right)^n \left(\frac{4c^2 - 4c}{4c^2 + 4c + 1} \right)^n |K \cap \mathbb{Z}^n|.$$

$\square$

Note that

$$\lim_{c \to \infty} \left(\frac{4c^2 - 4c}{4c^2 + 4c + 1} \right)^n = 1.$$

Hence, for large values of c Lemma 4.1.13 and Theorem 4.1.4 are essentially the same. Recall the definition of the lattice width of a compact convex set K. See Section 1.2. We can weaken the conditions in Lemma 4.1.13 and give a more precise dependence of the number $k^\star$ from the lattice width of the underlying polytope. First we prove two auxiliary lemmata.

Lemma 4.1.14. *Let $K \subset \mathbb{R}^n$ be a bounded convex set with nonempty interior. There exists an ellipsoid E such that*

$$c_K + E \subset K \subset c_K + n^{3/2}E.$$

Proof. Without loss of generality we assume that $c_K = 0$. Let E' be the maximum volume ellipsoid contained in $K - K$. Then, by John's characterization of inscribed ellipsoids of maximum volume for centrally symmetric convex bodies (see John [Joh48] and Ball [Bal92] or [Gru07, Section 11.1]) it holds that

$$E' \subset K - K \subset \sqrt{n}E'.$$

Without loss of generality we assume that $E' = B^n := \{x \in \mathbb{R}^n \mid \|x\|_2 \leq 1\}$.

A well known property regarding the centroid is that for any $u \in S^{n-1}$ it holds

$$\frac{1}{n} \leq \left| \frac{\max_{x \in K} u^{\mathsf{T}} x}{\min_{x \in K} u^{\mathsf{T}} x} \right| \leq n \tag{4.5}$$

(e.g. [Grü60]). To prove this property, we assume that (4.5) does not hold. Without loss of generality we also assume that $\min_{x \in K} u^{\mathsf{T}} x = -1$ and $\max_{x \in K} u^{\mathsf{T}} x =: a > n$ where $u = e_1$, i.e. u is equal to the first unit vector. Let $z := \operatorname{argmax}_{x \in K} u^{\mathsf{T}} x$. We define for every $t \in \mathbb{R}$ the set $K_t := K \cap \{x \in \mathbb{R}^n \mid u^{\mathsf{T}} x = t\}$. Further, we define $C := z + \operatorname{cone}(K_0 - z)$, $S_1 := \{x \in \mathbb{R}^n \mid -1 \leq u^{\mathsf{T}} x \leq 0\}$ and $S_2 = \{x \in \mathbb{R}^n \mid 0 \leq u^{\mathsf{T}} x \leq a\}$. Then

$$K \cap S_1 \subset C \cap S_1 \text{ and } K \cap S_2 \supset C \cap S_2.$$

See Figure 4.2 for an illustration. It follows

$$- \int_{-1}^{0} t \operatorname{vol}_{n-1}(K_t) dt$$

$$\leq - \int_{-1}^{0} t \operatorname{vol}_{n-1}\left(\frac{a-t}{a} K_0\right)$$

$$< \int_{0}^{a} t \operatorname{vol}\left(\frac{a-t}{a} K_0\right)$$

$$\leq \int_{0}^{a} t \operatorname{vol}_{n-1}(K_t) dt.$$

But our assumption that $c_K = 0$ implies that

$$- \int_{-1}^{0} t \operatorname{vol}_{n-1}(K_t) dt = \int_{0}^{a} t \operatorname{vol}_{n-1}(K_t) dt,$$
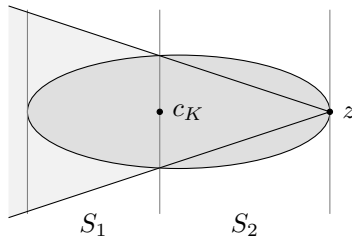
Figure 4.2: An illustration of the sets K, C, $K \cap S_1$, $C \cap S_1$, $K \cap S_2$ and $C \cap S_2$.

which is a contradiction. Hence, this proves (4.5).

Now, it follows on the one hand that

$$K \subset \frac{n}{n+1}\sqrt{n}E'$$

and on the other hand it follows that

$$\frac{1}{n+1}E' \subset K$$

The lemma follows by choosing $E = \frac{1}{n+1}E'$. $\square$

Lemma 4.1.15. *Let $K \subset \mathbb{R}^n$ be a bounded convex set with lattice-width $\omega(K)$. Then there exists a universal constant α and a unimodular matrix $U \in \mathbb{Z}^{n \times n}$ such that*

$$c_K + \frac{\omega(K)}{2\alpha n^{n-1}}U[-1,1]^n \subset K$$

Proof. By Lemma 4.1.14, we know that there exists an ellipsoid E such that $c_K + E \subset K \subset c_K + n^{3/2}E$. Let $E = \{x \in \mathbb{R}^n \mid x^\mathsf{T} D^{-2}x \leq 1\}$ where $D \in \mathbb{S}^n_{++}$ is a positive definite matrix. We want to map $c_K + E$ to the n-dimensional unit ball B^n. For that we define $\phi : \mathbb{R}^n \mapsto \mathbb{R}^n$ by

$$\phi(x) := D^{-1}x.$$

Let $\Lambda := \phi(\mathbb{Z}^n)$ and $K' := \phi(K)$. Note that $c_{K'} = \phi(c_K)$. Further, note that $\omega_\Lambda(K') = \omega_{\mathbb{Z}^n}(K)$ and that $c_K + B^n \subset K \subset c_K + n^{3/2}B^n$.

Let B denote a Korkine-Zolotareff basis of Λ [KZ73]. Then, a well known property is that

$$\|B_{\star,1}\|_2 \cdots \|B_{\star,n}\|_2 \leq \alpha n^n \det(\Lambda) \tag{4.6}$$

(see [LLS90, Theorem 2.3]), where α is a universal constant and $B_{\star,i}$ denotes the i-th column of B for $i = 1, \ldots, n$.

Let C denote the cross-polytope $\mathrm{conv}(\{\pm e_1, \ldots, \pm e_n\})$, where e_i denotes the i-th unit vector for $i = 1, \ldots, n$. Next we show that $c_{K'} + \sqrt{n}\frac{\omega(K)}{2\alpha n^{n-1}}BC \subset K'$. This then implies that

$$c_K + \frac{\omega(K)}{2\alpha n^{n-1}}DB[-1,1]^n \subset c_K + \sqrt{n}\frac{\omega(K)}{2\alpha n^{n-1}}DBC \subset K.$$

To derive a contradiction let us assume that $\sqrt{n}\frac{\omega(K)}{2\alpha n^{n-1}}B_{\star,n} \notin K'$. Then

$$\|B_{\star,n}\|_2 > \frac{1}{\sqrt{n}\frac{\omega(K)}{2\alpha n^{n-1}}}. \tag{4.7}$$

Let $\tilde{B}_{\star,1}, \ldots, \tilde{B}_{\star,n}$ denote the Gram-Schmidt orthogonalization of $B_{\star,1}, \ldots, B_{\star,n}$ (see for example [Gru07, Chapter 28]). It holds that $\|\tilde{B}_{\star,i}\| \leq \|B_{\star,i}\|_2$ for all $i = 1, \ldots, n$ and it holds that $\det(\Lambda) = \det(B) = \prod_{i=1}^{n} \|\tilde{B}_{\star,i}\|_2$. Together with (4.6) and (4.7) this implies

$$\|\tilde{B}_{\star,n}\|_2 > \frac{1}{\sqrt{n}\frac{\omega(K)}{2\alpha n^{n-1}}} \frac{1}{\alpha n^n} = \frac{2}{n^{3/2}\omega(K)}.$$

This, in turn, implies that $\omega(K) < 2\frac{n^{3/2}\omega(K)}{2}\frac{1}{n^{3/2}} = \omega(K)$, which isa contradiction. Then, setting $U := DB$ proves the lemma. $\qquad \square$

Now we are ready to prove the alternative version of Lemma 4.1.13 depending on the lattice-width.

Lemma 4.1.16. *Let $K \subset \mathbb{R}^n$ be a compact convex set with lattice-width $\omega(K) \geq 3\alpha n^{n-1}$ (where α is defined as in Lemma 4.1.15). There exists a point $z \in \mathbb{Z}^n$ such that for any half-space $H \ni z$, we have*

$$|K \cap H \cap \mathbb{Z}^n| \geq c \left(\frac{n}{n+1}\right)^n |K \cap \mathbb{Z}^n|,$$

where c is a constant only depending on $\omega(K)$ and n.

Proof. From Lemma 4.1.15 we know that

$$c_K + \frac{\omega(K)}{2\alpha n^{n-1}}U[-1,1]^n \subset K,$$

where U is defined and constructed as in the proof of Lemma 4.1.15. Let

$$c' := \frac{\omega(K)}{2\alpha n^{n-1}} - \frac{1}{2} \geq 1. \tag{4.8}$$

Define $z := \lfloor c_K \rceil_U$. (For the notation $\lfloor x \rceil_U$, see the proof of Lemma 4.1.13 and its corresponding footnote.) First, we define the unimodular mapping $\psi : \mathbb{R}^n \mapsto \mathbb{R}^n$ as

$$\psi(x) := U^{-1}x.$$

Then

$$c_{\psi(K)} + (c' + \frac{1}{2})[-1,1]^n \subset \psi(K).$$

Then the lemma follows from Lemma 4.1.15, with

$$c := \left(\frac{4c'^2 - 4c'}{4c'^2 + 4c' + 1}\right)^n. \tag{4.9}$$

$\qquad \square$

Remark 4.1.17. We want to point out that the proof of Lemma 4.1.16 is constructive. In particular, if K is a polytope and the dimension n is fixed, then the point z can be computed in polynomial time in the input-size of the polytope.

4.1.3 Existence of center-points in $\mathbb{Z}^n \times \mathbb{R}^d$

In this section we consider the mixed-integer setting. In order to derive an analogue result to Theorem 4.1.6 and Theorem 4.1.8, we need a Helly-type theorem for the mixed-integer setting $\mathbb{Z}^n \times \mathbb{R}^d$.

Theorem 4.1.18. (Averkov and Weismantel, [AW12]) *Let* $K_1, \dots, K_m \subset \mathbb{R}^{n+d}$ *be a family of closed convex sets. If for every subfamily* $K_{i_1}, \dots, K_{i_l}$ *with* $l \leq 2^n(d+1)$

$$\bigcap_{j=1}^l (K_{i_j} \cap \mathbb{Z}^n \times \mathbb{R}^d) \neq \emptyset,$$

then

$$\bigcap_{j=1}^m (K_j \cap \mathbb{Z}^n \times \mathbb{R}^d) \neq \emptyset.$$

Theorem 4.1.19. *Let* μ *be a finite measure on* $\mathbb{Z}^n \times \mathbb{R}^d$ *and let* $K \subset \mathbb{Z}^n \times \mathbb{R}^d$ *be a compact set. Then*

$$\mathrm{hull}^\mu_{\frac{1}{2^n}\frac{1}{d+1}}(K) \neq \emptyset.$$

Proof. The proof follows precisely the lines in the proof of Theorem 4.1.8. Instead of Doignon's theorem, Theorem 4.1.7, we use Theorem 4.1.18. $\square$

If we choose μ to be the d-dimensional Lebesgue measure, a very interesting open question is whether one can improve the constant k. In terms of volume, we conjecture the following:

Conjecture 4.1.20. *Let* $K \subset \mathbb{R}^{n \times d}$ *be a closed convex set. There exists a point* $x \in \mathbb{Z}^n \times \mathbb{R}^d$ *such that any half-space* H *containing* x *satisfies*

$$\mathrm{vol}_d(K \cap H \cap (\mathbb{Z}^n \times \mathbb{R}^d)) \geq \frac{1}{2^n}\left(\frac{d}{d+1}\right)^d \mathrm{vol}_d(K \cap (\mathbb{Z}^n \times \mathbb{R}^d)).$$

This bound is tight, in the sense that if we choose $K = \text{conv}\{\{0,1\}^n \times S\}$, where $S \subset \mathbb{R}^d$ is a d-dimensional simplex, there exists a hyperplane that only truncates one corner of one of the copies of the simplices. The interesting part would be that the bound is independent of d in the sense that $\lim_{d \to \infty} \frac{1}{2^n} \left(\frac{d}{d+1} \right)^d = \frac{1}{2^n} e^{-1}$.

4.2 Computation of approximate center-points

In this section we present two algorithms which compute approximate center-points for the pure integer case with respect to polytopes P and the counting measure μ. The first algorithm deals with the special case where $n = 2$, which yields a point in $\text{hull}^{\mu}_{\frac{1}{4}}(P)$. The second algorithm will handle any fixed dimension n, however it will only compute a point in $\text{hull}^{\mu}_{\frac{1}{2^{n^2}}}(P)$.

4.2.1 The 2-dimensional case

The main result in this section is the following one:

Theorem 4.2.1. *Let $P = \{x \in \mathbb{R}^2 \mid Ax \leq b\}$, where $A \in \mathbb{Z}^{m \times 2}$ be a bounded polytope and $b \in \mathbb{Z}^m$, such that $P \cap \mathbb{Z}^2 \neq \emptyset$. Then, in polynomial time in the input-size of A and b, one can compute a point*

$$z \in \text{hull}^{\mu}_{\frac{1}{4}}(P),$$

where μ denotes the counting measure.

Proof. First we compute the lattice-width $\omega(P)$ and the flatness-direction v of P. See Observation 3.1.1. Then, we distinguish between the two cases where P either has a large or a small lattice-width. Let $c(\omega)$ be defined as in (4.9) and (4.8). Then, let $c^\star$ be sufficiently large, such that $c(\omega) \left(\frac{4}{9} \right) \geq \frac{1}{4}$ for $n = 2$ and all $\omega \geq c^\star$.

If $\omega(P) > c^\star$, then, by Lemma 4.1.16, we can compute $z \in \text{hull}^{\mu}_{\frac{1}{4}}(P)$. See Remark 4.1.17 on how to compute z.

Otherwise, if $\omega(K) \leq c^\star$ we determine the center-point of P (and not an approximation). For that we define the function $f : S^1 \times \mathbb{Z}^2 \mapsto \mathbb{N}$ as

$$f(u, y) := |P \cap \{x \in \mathbb{Z}^2 \mid u^\mathsf{T} x \leq u^\mathsf{T} y\}|.$$

That is, the function f returns the number of integer points contained in P intersected with the half-space defined by u and y. For any fixed $y \in \mathbb{Z}^2$ we can optimize $\min_{u \in S^1} f(u, y)$. We will prove this in a separate claim below. With f, we can now express the center-point of P as the following optimization problem

$$\underset{y \in K \cap \mathbb{Z}^2}{\mathrm{argmax}} \min_{u \in S^1} f(u, y).$$

Since P is flat, we can divide the latter problem into $\lfloor \omega(K) + 1 \rfloor$ subproblems. We consider all levels $L_i = \{x \in \mathbb{R}^2 \mid v^\mathsf{T} x = i\}$ such that $L_i \cap P \neq \emptyset$. Note that f is quasi-convex in y. Therefore, with binary search on $L_i \cap \mathbb{Z}^2$, we can solve the subproblems

$$\underset{y \in L_i \cap \mathbb{Z}^2}{\mathrm{argmax}} \min_{u \in S^1} f(u, y).$$

Among all the subproblems we pick the optimal solution.

It remains to prove the following claim.

Claim: Given a point $y \in \mathbb{Z}^2$, then, in polynomial time in $\omega(P)$ and the binary encoding length of A and b one can solve

$$\min_{u \in S^1} f(u, y).$$

We assume that the flatness-direction $v = (0, 1)^\mathsf{T}$, $0 = \lceil \min_{x \in K} v^\mathsf{T} x \rceil$ and that $l = \lfloor \max_{x \in K} v^\mathsf{T} x \rfloor = \lfloor \omega(K) + 1 \rfloor$. That is $K \cap \mathbb{Z}^2 \subset \bigcup_{i=0}^l L_i$. Let $w := (1, 0)^\mathsf{T}$. For $i = 0, \ldots, l$ we define

$$w_i^- := \underset{x \in K \cap L_i}{\mathrm{argmin}} \, w^\mathsf{T} x \quad \text{and} \quad w_i^+ := \underset{x \in K \cap L_i}{\mathrm{argmax}} \, w^\mathsf{T} x.$$

For $y \neq w_i^-, w_i^+$ we define

$$u_i^- = \frac{w_i^- - y}{\|w_i^- - y\|_2} \quad \text{and} \quad u_i^+ = \frac{w_i^+ - y}{\|w_i^+ - y\|_2}.$$

There are three main observations to make: (i) f is piece-wise constant in u (ii) given a $u \in S^1$ one can compute $f(u, y)$ (one only needs to count integer points on at most $l + 1$ line segments) and (iii) one can divide S^1 into $2d + 2$ intervals (defined by the u_i^- and u_i^+) such that f is monotone in each of them. See Figure 4.3.

It remains to evaluate $f(u, y)$ for a constant number of points u, i.e. $u \in U := \{u_0^- \pm \epsilon, u_0^+ \pm \epsilon, u_1^- \pm \epsilon, \ldots\}$ where ϵ is a sufficiently small perturbation. It holds that

$$min_{u \in S^1} f(u, y) = min_{u \in U} f(u, y).$$
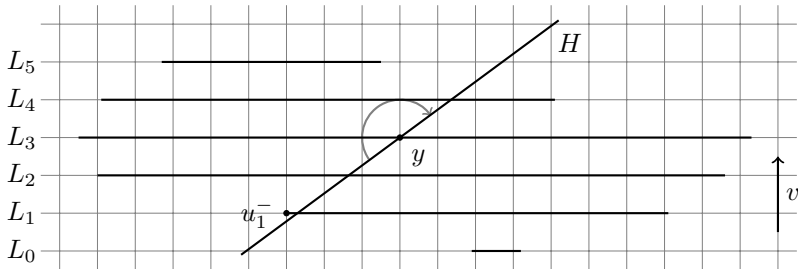
Figure 4.3: Let $H = \{x \in \mathbb{R}^2 \mid u^{\mathsf{T}}x = u^{\mathsf{T}}y\}$. This figure illustrates the characteristics of f. In particular the monotonicity of f only changes around critical points such as u_1^-.

$\square$

Remark 4.2.2. With the same approach as in Theorem 4.2.1 one can solve for any fixed $\bar{k} < \frac{4}{9}$

$$
\begin{aligned}
(k^\star, x^\star) := \quad &\underset{}{\mathrm{argmax}} \quad k \\
&\text{s.t.} \quad x \in \mathrm{hull}_k^\mu(S) \\
&\phantom{\text{s.t.}} \quad x \in \mathbb{E}^2 \\
&\phantom{\text{s.t.}} \quad k \in [0, \bar{k}].
\end{aligned}
$$

For that, one needs to adapt the threshold for when one uses Lemma 4.1.16 or not.

4.2.2 An approximation algorithm for fixed dimension n

Analogous to the previous section we could define a function $f : S^{n-1} \times \mathbb{Z}^n \mapsto \mathbb{N}$ with $f(u, y) := |P \cap \{x \in \mathbb{Z}^n \mid u^{\mathsf{T}}x \leq u^{\mathsf{T}}y\}|$. Then, if we could evaluate for a fixed $y \in \mathbb{Z}^n$ the subproblem $\min_{u \in S^{n-1}} f(u, y)$, we could proceed in a similar way as before. However, up to this point it is not clear to us how to achieve this.

As an alternative, we present the following weaker result, which is based on induction.

Theorem 4.2.3. *Let n be fixed, let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be a polytope, where $A \in \mathbb{Z}^{m \times n}$ and $b \in \mathbb{Z}^m$, such that $P \cap \mathbb{Z}^n \neq \emptyset$. Then, in polynomial time in the input-size of A and b, one can find a point*

$$
z \in \mathrm{hull}_{\frac{1}{2^{n^2}}}^\mu(P),
$$

where μ denotes the counting measure.

Proof. Clearly, the theorem holds true for $n \leq 2$ by Theorem 4.2.1. For the purpose of induction, let us assume that the theorem holds for $n-1$.

As in the proof of Theorem 4.2.1 we distinguish two cases. For that let $c(\omega)$ be defined as in (4.9) and (4.8) depending on ω. Further, let $c^\star$ be sufficiently large, such that $c(\omega) \left(\frac{n}{n+1} \right)^n \geq \frac{1}{2^{n^2}}$ for all $\omega \geq c^\star$. Note that $c^\star$ is a constant only depending on n.

If $\omega(P) > c^\star$, we apply Lemma 4.1.16 and compute a point z as in Remark 4.1.17. By the choice of $c^\star$ we have that $z \in \text{hull}^\mu_{\frac{1}{2^{n^2}}}(P)$.

Otherwise, if $\omega(P) \leq c^\star$, we can partition $P \cap \mathbb{Z}^n$ into $\omega(K)$ sets. Let $v \in \mathbb{Z}^n$ be the flatness direction corresponding to the lattice-width $\omega(P)$. Without loss of generality we assume that $v = (0, \ldots, 0, 1)^\mathsf{T}$, $0 = \lceil \min_{x \in K} v^\mathsf{T} x \rceil$ and that $l = \lfloor \max_{x \in K} v^\mathsf{T} x \rfloor = \lfloor \omega(K) + 1 \rfloor$. We define $L_i = \{x \in \mathbb{R}^n \mid v^\mathsf{T} x = i\}$ and $P_i = P \cap L_i$ for $i = 0, \ldots, l$. Hence $P \cap \mathbb{Z}^n = \bigcup_{i=0}^l (P_i \cap (\mathbb{Z}^{n-1} \times \{i\}))$.

Each P_i, with $i \in \{0, \ldots, l\}$, is of the dimension $n-1$ or smaller. By the inductive hypothesis we can compute the points

$$z_i \in \text{hull}^\mu_{\frac{1}{2^{(n-1)^2}}}(P_i)$$

for $i = 0, \ldots, l$. We define the following measure $\bar{\mu}$ on $\mathbb{Z}^n$. For each $z \in \mathbb{Z}^n$ set

$$\bar{\mu}(z) := \begin{cases} |P_i \cap \mathbb{Z}^n| & \text{if } z = z_i, \\ 0 & \text{otherwise.} \end{cases}$$

Note that with Barvinok's algorithm [Bar94] we can compute the number $|P_i \cap \mathbb{Z}^n|$ in polynomial time in the input size. Further note that $\bar{\mu}(\mathbb{Z}^n) = \bar{\mu}(P \cap \mathbb{Z}^n) = |P \cap \mathbb{Z}^n|$. By Theorem 4.1.8 we know that there exists a point

$$z \in \text{hull}^{\bar{\mu}}_{\frac{1}{2^n}}(P).$$

There is a constant number of possible partitions of $\{z_0, \ldots, z_l\}$. Hence, this point z can be computed by brute force.

Let H denote a half-space containing z. If $\text{bd}(H)$, the boundary of H, is parallel to L_0, then $|H \cap P \cap \mathbb{Z}^n| \geq \frac{1}{2^n} |P \cap \mathbb{Z}^n|$. Otherwise, H defines $(n-1)$-dimensional half-spaces H_i with respect to the affine sub-spaces L_i for $i = 0, \ldots, l$, i.e. $H_i := H \cap L_i$ for $i = 0, \ldots, l$. We have that

$$\frac{\bar{\mu}(H \cap \mathbb{Z}^n)}{\bar{\mu}(\mathbb{Z}^n)} = \frac{\sum_{i=0}^l \bar{\mu}(H_i \cap \mathbb{Z}^n)}{\bar{\mu}(\mathbb{Z}^n)} \geq \frac{1}{2^n}.$$

If $z_i \in H_i$ then $\bar{\mu}(H_i \cap \mathbb{Z}^n) \neq 0$. By construction of the z_i's we have that $|H \cap P_i \cap \mathbb{Z}^n| \geq \frac{1}{2^{(n-1)^2}} |P_i \cap \mathbb{Z}^n| = \frac{1}{2^{(n-1)^2}} \bar{\mu}(H_i \cap \mathbb{Z}^n)$. Hence,

$$\sum_{i=0}^{l} |H \cap P_i \cap \mathbb{Z}^n| \geq \frac{1}{2^{(n-1)^2}} \frac{1}{2^n} |P \cap \mathbb{Z}^n|.$$

This finally proves that $|H \cap P \cap \mathbb{Z}^n| \geq \frac{1}{2^{n^2}} |P \cap \mathbb{Z}^n|$ for any half-space H containing z. $\qquad\square$

4.3 Mixed-integer optimality conditions for convex functions

Given a convex function $f : \mathbb{R}^n \mapsto \mathbb{R}$. Assume that f has a, not necessarily unique, minimizer $x^\star$. Then a necessary and sufficient certificate for $x^\star$ being a minimizer of f is that $0 \in \partial f(x^\star)$, i.e. the zero-function is in the subdifferential of f at $x^\star$. Hence

$$x^\star = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x) \iff 0 \in \partial f(x^\star).$$

We can add convex constraints $g(x) \leq 0$, where $g : \mathbb{R}^n \mapsto \mathbb{R}^m$ and consider the following problem

$$x^\star = \underset{\substack{x \in \mathbb{R}^n, \\ g(x) \leq 0}}{\operatorname{argmin}} f(x).$$

Assume that there exists a point y fulfilling the Slater condition, that is, $g(y) < 0$. In this case the Karush-Kuhn-Tucker conditions (e.g. [Kar39, KT51]) provide necessary and, in our setting, sufficient optimality conditions. Namely, let $h_f \in \partial f(x^\star)$ and $h_{g_i} \in \partial g_i(x^\star)$, $i = 1, \ldots, m$. Then there exist non-negative λ_i, $i = 1, \ldots, m$, such that

$$-h_f = \sum_{i=1}^{m} \lambda_i h_{g_i}.$$

Note that it suffices only to consider those $g_i(x^\star)$ that are active, i.e. $\lambda_i \neq 0$, and linearly independent.

It is natural to ask, whether it is possible to formulate optimality conditions for the integer setting

$$x^\star = \underset{\substack{x \in \mathbb{Z}^n, \\ g(x) \leq 0}}{\operatorname{argmin}} f(x) \qquad (4.10)$$

and for the mixed-integer setting

$$x^\star = \underset{\substack{x \in \mathbb{Z}^n \times \mathbb{R}^d, \\ g(x) \leq 0}}{\operatorname{argmin}} f(x). \tag{4.11}$$

One key element in the proof of the Karush-Kuhn-Tucker conditions is the Carathéodory theorem (e.g. [Gru07, Theorem 3.1.]). In order to obtain an integer or mixed-integer analogue we will have to replace Carathéodory's theorem by an integer and mixed-integer Helly-type theorem, i.e. Theorem 4.1.7 and Theorem 4.1.18 proven in [Doi73] and [AW12] respectively.

For an easier exposition of this material we start with the unconstrained case.

Theorem 4.3.1. *A point $x^\star \in \mathbb{Z}^n \times \mathbb{R}^d$ is optimal with respect to*

$$\min_{x \in \mathbb{Z}^n \times \mathbb{R}^d} f(x)$$

if and only if there exist $k \leq 2^n$ points $x_1 = x^\star, x_2, \ldots, x_k$ with corresponding $h_{x_i} \in \partial f(x_i)$ such that $f(x_i) \geq f(x^\star)$ for $i = 1, \ldots, k$ and

$$\{x \in \mathbb{R}^{n+d} \mid h_{x_i}^\mathsf{T} x < h_{x_i}^\mathsf{T} x_i \text{ for all } i = 1, \ldots, k\} \cap (\mathbb{Z}^n \times \mathbb{R}^d) = \emptyset.$$

Proof. One key observation is that given a point $x \in \mathbb{Z}^n \times \mathbb{R}^d$ and a subgradient $h \in \partial f(x)$ we have that $x = \operatorname{argmin}\{f(y) \mid h^\mathsf{T} y \geq h^\mathsf{T} x, \ y \in \mathbb{Z}^n \times \mathbb{R}^d\}$, i.e. x minimizes the function f with respect to the half-space defined by x and its subgradient h, If $h = 0$, then x minimizes f over $\mathbb{Z}^n \times \mathbb{R}^d$.

We assume that $x^\star$ is optimal. Let $X^\star$ denote the set of all optimal solutions. We may assume that $X^\star \neq \emptyset$, otherwise there is nothing to prove. If there exists a point $x \in X^\star$ with $0 \in \partial f(x)$, then the theorem follows directly from the purely continuous case, described at the beginning of this section. Next, assume there exists an $x \in X^\star \cap \operatorname{int}(\operatorname{conv}(X^\star))$ and $h_x \in \partial f(x)$ such that $h_x \neq 0$. Then, since x is in the interior of $\operatorname{conv}(X^\star)$, there exists a $x' \in X^\star$ such that $f(x') - f(x) \geq h_x^\mathsf{T}(x' - x) > 0$, which contradicts that x' is optimal. This implies that if $X^\star \cap \operatorname{int}(\operatorname{conv}(X^\star)) \neq \emptyset$ that then $0 \in \partial f(x)$ for all $x \in X^\star$. Hence, let us assume that $X^\star \cap \operatorname{int}(\operatorname{conv}(X^\star)) = \emptyset$ and that $0 \notin \partial f(x)$ for all $x \in X^\star$.

Let the function $F : \mathbb{R}^n \mapsto \mathbb{R}$ be defined as

$$F(z) := \min_{y \in \mathbb{R}^d} f(z, y).$$

Note that F is convex again. For each $z \in \mathbb{R}^n$ let $y_z := \operatorname{argmin}_{y \in \mathbb{R}^d} f(z, y)$. Since y_z is an optimal solution of a continuous problem there exists a $(h_z, h_{y_z}) \in$

$\partial f((z, y_z))$ such that (h_z, h_{y_z}) is orthogonal to $\{z\} \times \mathbb{R}^d$ (i.e. $h_{y_z} = 0$). Further, $h_z \in \partial F(z)$.

Let $L := \{y \in \mathbb{R}^n \mid h_z^\mathsf{T} y < h_z^\mathsf{T} z, \ z \in \mathbb{Z}^n, \ h_z \in \partial F(z)\}$. Clearly $L \cap \mathbb{Z}^n = \emptyset$, i.e. L is lattice-point free. It follows from [Doi73] that a sub-selection of $k \leq 2^n$ inequalities $h_{z_i}^\mathsf{T} y \leq h_{z_i}^\mathsf{T} z_i, \ i = 1, \ldots, k$ (we can choose z_1 such that $(z_1, y_{z_1}) = x^\star$) suffice, such that the relaxation $\bar{L} := \{y \in \mathbb{R}^n \mid h_{z_i}^\mathsf{T} y < h_{z_i}^\mathsf{T} z_i, \ i = 1, \ldots, k\}$ remains lattice-point free. Then the one direction follows by choosing $x_1 = x^\star, x_2 = (z_2, y_{z_2}), \ldots, x_k = (z_k, y_{z_k})$.

To prove the other direction, let

$$P := \{x \in \mathbb{R}^{n+d} \mid h_{x_i}^\mathsf{T} x < h_{x_i}^\mathsf{T} x_i \text{ for all } i = 1, \ldots, k\}$$

and assume that $P \cap (\mathbb{Z}^n \times \mathbb{R}^d) = \emptyset$. Let $\bar{x} \in \mathbb{Z}^n \times \mathbb{R}^d$. Then $\bar{x}$ must violate at least one inequality of P, say the i-th inequality. But then we know from our observation in the beginning that x_i minimizes f over the half-space $\{x \in \mathbb{Z}^n \times \mathbb{R}^d \mid h_{x_i}^\mathsf{T} x \geq h_{x_i}^\mathsf{T} x_i\}$. Hence, $f(\bar{x}) \geq f(x_i) \geq f(x^\star)$. Therefore $x^\star$ must be optimal. $\qquad\square$

Now we prove the following theorem, providing if and only if conditions for an optimal point with respect to (4.11) (which includes (4.10)).

Theorem 4.3.2. *Assume that there exists a point $s \in \mathbb{R}^{n+d}$ such that $g(s) < 0$. A point $x^\star \in \mathbb{Z}^n \times \mathbb{R}^d$ is optimal with respect to (4.11) if and only if $g(x^\star) \leq 0$ and there exist $k + l \leq 2^n(d+1)$ points $x_1 = x^\star, x_2, \ldots, x_k$ and $y_1, \ldots, y_l$ in $\mathbb{R}^{n+d}$ with corresponding $h_{x_i} \in \partial f(x_i)$, and $h_{y_i} \in \partial g_j(y_i)$ such that $f(x_i) \geq f(x^\star)$ for $i = 1, \ldots, k$, $g(y_i) \geq 0$ for $i = 1, \ldots, l$ and such that*

$$\{x \in \mathbb{R}^{n+d} \mid h_{x_i}^\mathsf{T} x < h_{x_i}^\mathsf{T} x_i \text{ for all } i = 1, \ldots, k,$$
$$h_{y_i}^\mathsf{T} x \leq h_{y_i}^\mathsf{T} y_i \text{ for all } i = 1, \ldots, l\} \cap (\mathbb{Z}^n \times \mathbb{R}^d) = \emptyset.$$

Proof. Let $X^\star$ denote the set of all optimal solutions. It follows from the arguments in the proof of Theorem 4.3.1 that we may assume that $X^\star \cap \text{int}(\text{conv}(X^\star)) = \emptyset$ and that $0 \notin \partial f(x)$ for all $x \in X^\star$.

Let $x^\star \in X^\star$, $L := \{x \in \mathbb{R}^{n+d} \mid f(x) \leq f(x^\star)\}$ and $S := \{x \in \mathbb{R}^{n+d} \mid g(x) \leq 0\}$. We can describe $X^\star$ as the intersection of the level-set L, the feasible region S and mixed-integer lattice. In turn, we can describe L and S as the intersection of half-spaces defined by the boundary points and their corresponding subdifferentials

$$L = \bigcap_{\substack{z \in \text{bd}(L), \\ h \in \partial f(z)}} \{x \in \mathbb{R}^{n+d} \mid h^\mathsf{T} x \leq h^\mathsf{T} z\}$$

and

$$S = \bigcap_{\substack{z \in \mathrm{bd}(S), \\ h \in \partial g(z)}} \{x \in \mathbb{R}^{n+d} \mid h^\mathsf{T} x \leq h^\mathsf{T} z\}.$$

Assume that $x + B_\epsilon \subset L \cap S$, where $x \in X^\star$ and B_ϵ denotes a ball of radius $\epsilon > 0$. Then, $f(x) = f(x^\star)$ and $f(y) \leq f(x^\star)$ for all $y \in x + B_\epsilon$. This then, implies that $0 \in \partial f(x)$, which then in turn contradicts our assumption. Hence, $X^\star \subset \mathrm{bd}(L \cap S)$. By our assumption $\mathrm{int}(\mathrm{conv}(X^\star)) \cap (\mathbb{Z}^n \times \mathbb{R}^d) = \emptyset$, i.e. $\mathrm{conv}(X^\star)$ is mixed-integer lattice free, it holds that $\mathrm{int}(L \cap S) \cap (\mathbb{Z}^n \times \mathbb{R}^d) = \emptyset$. It follows from [AW12] that a subset of $2^n(d+1)$ half-spaces suffice in order to guarantee that the corresponding intersection remains mixed-integer lattice free.

To prove the other direction, let

$$P := \{x \in \mathbb{R}^{n+d} \mid h_{x_i}^\mathsf{T} x < h_{x_i}^\mathsf{T} x_i \text{ for all } i = 1, \ldots, k,$$
$$h_{y_i}^\mathsf{T} x \leq h_{y_i}^\mathsf{T} y_i \text{ for all } i = 1, \ldots, l\}.$$

Since $P \cap (\mathbb{Z}^n \times \mathbb{R}^d) = \emptyset$, P cannot contain any feasible point $x \in \mathbb{Z}^n \times \mathbb{R}^d$ with a smaller objective value. Any other point $x \in \mathbb{Z}^n \times \mathbb{R}^d \setminus P$ must violate at least one inequality, say either $h_{x_1}^\mathsf{T} x \geq h_{x_1}^\mathsf{T} x_1$ or $h_{y_1}^\mathsf{T} x > h_{y_1}^\mathsf{T} y_1$. By the definition of subgradients, x must either have objective value greater than or equal to $f(x^\star)$ or it is infeasible. Hence, no feasible mixed-integer lattice point exists with better objective value than $x^\star$. □

In the following section, we present an algorithm that will terminate with precisely the optimality certificate established here.

4.4   Algorithmic implications

We consider (mixed)-integer convex minimization problems, that is

$$\min_{\substack{x \in \mathbb{Z}^n \times \mathbb{R}^d, \\ g(x) \leq 0}} f(x),$$

where $f, g : \mathbb{R}^n \times \mathbb{R}^d \mapsto \mathbb{R}$ are convex functions. In this section we want to present a general cutting plane method based on center-points. This can be interpreted as a direct extension of the well-known Method of Centers of Gravity or the more general cutting plane methods (e.g. [Nes04, Section 3.2.6.]). Algorithms such as ellipsoid method, outer approximation, Kelly's

cutting plane method and level method also fit into this framework. As input we consider two functions f and g given by first-order evaluation oracles. See Definition 1.1.1.

Further, we assume that we have access to approximate center-points of polytopes. For that we define the following oracle.

Definition 4.4.1 (α-central-point-oracle)**.** For a polytope P, the oracle returns a point

$$z \in \text{hull}_\alpha^\mu(P).$$

This way we hide the problem of computing central points with its complexity in the oracle and keep the following discussion as general as possible. However, for several instances the oracle can be realized. For example in the continuous case (i.e. $n = 0$) and the Lesbesgue measure we have the centroid as an approximate center-point or another alternative is the center of the maximum inscribed ellipsoid of the polytope. For the case $d = 0$, we have seen two possible choices of approximate center-points in sections 4.2.1 4.2.2.

The general algorithmic framework is as follows. We start with the bounding box, say $P^0 := [0, B]^n$. Then we construct iteratively a sequence of polytopes $P^1, P^2, \ldots$ by intersecting P^k with the half-space defined by its (approximate) center-point and the corresponding subgradient arising from either f or g. That is, let $x_k \in \text{hull}_\alpha^\mu(P_k)$ and let $\bar{h} \in \partial f(x_k)$ (or $\bar{h} \in \partial g(x_k)$ if x_k is not feasible). Then we define $P^{k+1} := P^k \cap \{x \in \mathbb{R}^n \mid \bar{h}^\mathsf{T} x < \bar{h}^\mathsf{T} x_k\}$. It follows that

$$P^k \supset P^{k+1} \supset \underset{\substack{x \in \mathbb{Z}^n \times \mathbb{R}^d, \\ g(x) \leq 0}}{\text{argmin}} f(x)$$

for all $k \in \mathbb{N}$. Further, by the choice of x_k, the measure of P^k decreases in each iteration by a fraction of at least $1 - \alpha$. Among all (approximate) center-points x_k that we encounter, we keep track of the feasible point with smallest objective value. We denote that point by $x^\star$ and its objective value by $f^\star = f(x^\star)$. This discussion is summarized in Algorithm 5.

Let us first recall the classical continuous case without the constraint $g(x) \leq 0$. We consider the method of centers of gravity, i.e. we choose the centroid as the approximate-central-point with $\alpha = e^{-1}$. Let $\hat{x} := \text{argmin}_{x \in [0,B]^n} f(x)$ and $\hat{f} := f(\hat{x})$. With L we denote the maximal Lipschitz constant of the function f. On the one hand we know, by Theorem 4.1.4, that in each iteration we reduce the volume by a factor $1 - e^{-1}$, i.e. $\text{vol}(P_k) \leq (1 - e^{-1}) \text{vol}(P_{k-1})$. Hence, after k iterations $\text{vol}(P_k) \leq (1 - e^{-1})^k \text{vol}([0, B]^n) = (1 - e^{-1})^k B^n$. On

Data: Functions f and g given by first-order evaluation oracles and $B \in \mathbb{N}$
      such that $\{x \in \mathbb{R}^{n+d} \mid g(x) \leq 0\} \subset [0, B]^{n+d}$.
Set $P^0 := [0, B]^{n+d}$ and $f^\star = \infty$.
for $k = 0, 1, \dots$ **do**
    | Let $x_k \in \text{hull}_\alpha^\mu(P_k)$ and let $\bar{g} = g(x_k)$
    | **if** $\bar{g} \leq 0$ **then**
    | | Let $\bar{f} = f(x_k)$
    | | **if** $\bar{f} < f^\star$ **then**
    | | | $f^\star = \bar{f}$ and $z^\star = x_k$.
    | | **end**
    | | Let $\bar{h} \in \partial f(x_k)$
    | **else**
    | | Let $\bar{h} \in \partial g(x_k)$
    | **end**
    | $P^{k+1} := P^k \cap \{x \in \mathbb{R}^{n+d} \mid \bar{h}^\mathsf{T} x < \bar{h}^\mathsf{T} x_k\}$
end
return $z^\star$ and $f^\star$.

Algorithm 5: The general cutting-plane method

the other hand, following from the Lipschitz constant, we have that $f^\star - \hat{f} \leq L\|\hat{x} - x\|_2$ for all $x \in \mathbb{R}^n$ with $f(x) = f^\star$. Hence,

$$P_k \supset \{x \in \mathbb{R}^n \mid f(x) \leq f^\star\}$$

$$\supset \left\{ x \in \mathbb{R}^n \;\middle|\; \|\hat{x} - x\|_2 \leq \frac{f^\star - \hat{f}}{L} \right\}$$

It follows,

$$\text{vol}(P_k) \geq \text{vol}\left(\left\{ x \in \mathbb{R}^n \;\middle|\; \|\hat{x} - x\|_2 \leq \frac{f^\star - \hat{f}}{L} \right\} \right)$$

$$= \left(\frac{\hat{f} - f^\star}{L} \right)^n \kappa_n,$$

where κ_n denotes the volume of the n-dimensional unit ball, i.e. $\kappa_n = \frac{\pi^{n/2}}{\Gamma(n/2+1)}$. This gives us then an upper bound on $\hat{f} - f^\star$ with respect to the number of iteration k. Namely, it holds that

$$\hat{f} - f^\star \leq LB(1 - e^{-1})^{k/n} \kappa_n^{-n}.$$

Given a precision $\epsilon > 0$ this implies that, after at most

$$k \leq \frac{n \ln(\epsilon/LB)}{\ln(1 - e^{-1})}$$

iterations the algorithm returns a solution $x^\star$, $f^\star$ such that $f^\star - \hat{f} \leq \epsilon$.

The main argument for the quality of the output of the algorithm was based on volumes. However, when we consider integer or mixed-integer settings the volume tends to be a bad measure. Only when the convex sets P_k have a big lattice width, there is a strong correlation between the volume and the number of integer points. (See Lemma 4.1.11.) When the lattice width is small there is basically no connection. For the pure integer case it appears that the number of integer points in P_k is a better measure. In the mixed-integer case (i.e. $\mathbb{Z}^n \times \mathbb{R}^d$) it would be natural to consider the d-dimensional volume, that is $\text{vol}_d(P_k \cap \mathbb{Z}^n \times \mathbb{R}^d) = \sum_{z \in \mathbb{Z}^n} \text{vol}_d(\{x \in \mathbb{R}^d \mid (z, x) \in P_k\})$.

In the following we want discuss these two measures for the integer and the mixed integer case.

4.4.1 Integer and mixed-integer convex minimization

Let us begin our discussion with the problem

$$\min_{\substack{x \in \mathbb{Z}^n, \\ g(x) \leq 0}} f(x),$$

where $f, g : \mathbb{R}^n \mapsto \mathbb{R}$ are convex functions. Here, we have again the interesting fact (as already stressed in Chapter 3) that we can compute the optimal point rather than an ϵ approximation. If we use the counting measure instead of the volume to evaluate our progress, this follows from the fact that the optimal solution(s), if they exist, have a measure greater than or equal to one.

Again we need the assumption that the problem is bounded, i.e. the optimum lies within a box $[0, B]^n$. In the beginning we have $(B + 1)^n \approx B^n$ integer points to consider. In each iteration we reduce this number by a fraction $1 - \alpha$. Hence after at most

$$k \leq \frac{n \ln(B)}{\ln(1 - \alpha)} + 1$$

iterations we are left with no integer points in the interior of P^k. Hence we have found the optimal solution and furthermore, up to redundant inequalities we also have a certificate for its optimality in form of P^k.

Let us now consider the mixed-integer problem

$$\min_{x \in \mathbb{Z}^n \times \mathbb{R}^d} f(x),$$

where $f : \mathbb{R}^n \times \mathbb{R}^d \mapsto \mathbb{R}$ is a convex function. For simplicity of exposition we discuss the unconstrained problem. Let L denote the Lipschitz constant of f and let $\epsilon > 0$ be a given precision. Again, in order to make the problem tractable, let us assume that the problem is bounded, i.e. the optimum lies within a box, say $[0, B]^{n+d}$.

Now again we can employ the same analysis as for the continuous case. Let μ denote the d-dimensional volume on $\mathbb{Z}^n \times \mathbb{R}^d$ and let

$$(\hat{x}, \hat{y}) := \operatorname*{argmin}_{(x,y) \in \mathbb{Z}^n \times \mathbb{R}^d} f((x,y))$$

and $\hat{f} := f((\hat{x}, \hat{y}))$. We have $\mu(\{0, \ldots, B\}^n \times [0, B]^d) \approx B^{n+d}$. This time we bound $\mu(P_k)$ from below as follows

$$\mu(P_k) \geq \mu(\{(x,y) \in \mathbb{Z}^n \times \mathbb{R}^d \mid f((x,y)) - \hat{f} \leq f^\star - \hat{f}\})$$

$$\geq \mu\left(\left\{(x,y) \in \mathbb{Z}^n \times \mathbb{R}^d \,\middle|\, \|(\hat{x}, \hat{y}) - (x,y)\|_2 \leq \frac{f^\star - \hat{f}}{L}\right\}\right)$$

$$\geq \mu\left(\left\{(\hat{x}, y) \in \{\hat{x}\} \times \mathbb{R}^d \,\middle|\, \|(\hat{x}, \hat{y}) - (\hat{x}, y)\|_2 \leq \frac{f^\star - \hat{f}}{L}\right\}\right)$$

$$= \left(\frac{\hat{f} - f^\star}{L}\right)^d \kappa_d.$$

Then, it follows that after at most

$$k \leq \frac{d \ln\left(\frac{\epsilon}{LB}\right) + n \ln(B)}{\ln(1 - \alpha)}$$

iterations we have that $f(x^\star, y^\star) - f(\hat{x}, \hat{y}) \leq \epsilon$.

We close this chapter with a remark on the constrained case. By adding constraints, we might encounter the problem that we do not find any feasible point, i.e. $g(x_i) > 0$ for all $i = 1, \ldots, k$. However, if g is also Lipschitz continuous, with Lipschitz constant L, then we can conclude that $\{x \in \mathbb{Z}^n \times \mathbb{R}^d \mid g(x) \geq -\epsilon\} = \emptyset$. One can avoid this problem if one knows a priori that the value $\mu(\{x \in \mathbb{Z}^n \times \mathbb{R}^d \mid g(x) \geq 0\})$ is of sufficient size.

Chapter 5

A Polyhedral Frobenius Theorem with Applications to Integer Optimization

This chapter is based on a joint work with David Adjiashvili and Robert Weismantel [AOW14].

Non-linear integer programming is concerned with optimizing a non-linear function over the integer points in a polyhedron. Significant effort has been made in recent years to extend the well-established theory of *linear* integer programming to the non-linear case. Along these line, polynomial algorithms for various classes of nonlinear functions were developed, including convex functions [GLS88], bounded-degree polynomials [DPW14, DPHWZ14] and more. Apart from very few exceptions [HS90, DLHOW08], however, all results in this vein were proved for the *fixed-dimension* case, namely for the case where the total number of variables is a fixed constant. The latter fact makes these methods less practical, limiting their potential domain of applications.

There are, of course, good reasons why positive algorithmic results in non-linear variable-dimension integer programming are harder to come by. Firstly, this class of problems trivially generalized linear integer programming, which is NP-hard in almost every variable-dimension setup. Secondly, non-linear variable-dimension integer problems often become hard already in the fixed-dimensional case. Finally, if the non-linear function acts directly on the variable-dimensional space, even stronger hardness results can be proven. For example, in the function oracle model one can prove simple information-theoretic exponential lower bounds on the complexity of any algorithm approximating the minimum of a convex function over the hypercube. If the function class is further restricted to be convex quadratic polynomials and stronger oracles are assumed, the latter problem becomes "merely" NP-hard to solve exactly. Worse still, the latter example shows the large increase in

complexity when a linear objective function is replaced with a non-linear one.
This means that any algorithm reducing the latter problems to integer linear
programming will most likely need to replace the well-structured feasible set,
namely the hypercube, with a much more complicated one.

Still, it is meaningful to ask: *Which non-linear variable-dimension integer
programming problems can be reduced to the linear case, maintaining the
structure of the problem class?* In this chapter we study one such class of
problems. Our class of problems contains an additional component, namely
that of a *projection* into a low-dimensional space. The previous discussion
suggests that this is, to a large extent, unavoidable when efficient reductions
of the latter type are sought. Formally, we are interested in studying problems
of the form

$$\min\{f(Wx) \mid x \in \mathbb{Z}^n \cap P\}, \tag{5.1}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is a function from our function class, $P := \{x \in \mathbb{R}^n \mid Ax \leq b\}$ is a polyhedron in n-dimensional space, (with $A \in \mathbb{Z}^{m \times n}, b \in \mathbb{Z}^m$) and
W is a $n \times d$ integer matrix. We discuss minimization here, but our results
also hold for maximization problems. The set $\mathcal{F} := \{x \in \mathbb{Z}^n \mid x \in P\}$ is
called the *feasible set*, and points $x \in \mathcal{F}$ are called *feasible*. Although not
necessary for our main result, we think of n as being large (variable) and of
d as being small (fixed). We note that this class of problems includes linear
integer programming already for $d = 1$ and f the identity.

In this chapter we give a first general-purpose efficient reduction from the
latter class of problems to integer programming. The efficiency of our re-
duction depends on various input parameters. We elaborate on this exact
dependence later. As a result, we obtain the first polynomial algorithms
for several classes variable-dimension non-linear integer problems. For other
problem classes, our method provides a polynomial time reduction from the
non-linear problem to linear integer programming, maintaining the structure
of the feasible set.

We assume black-box access to two oracles, namely a *fiber oracle* and an *d-
dimensional non-linear optimization oracle* (or simply, *optimization oracle*),
defined as follows.

Definition 5.0.2 (fiber oracle, optimization oracle)**.**

- The fiber oracle accepts as input a point $y \in \mathbb{Z}^d$ and either returns a
  feasible point (a point $x \in \mathcal{F}$) such that $Wx = y$, or asserts that no
  such point exists.

- The optimization oracle accepts descriptions of a polyhedron $R \subset \mathbb{R}^d$ and a affine sub-lattice $\Lambda \subset \mathbb{Z}^d$ of the integer lattice, and returns a point y^* in

$$\arg\min\{f(y) \mid y \in \Lambda \cap R\},$$

if one exists, or asserts that the latter set is empty.

We note that both oracles can be implemented in polynomial time for various classes of input parameters. We defer a detailed discussion on this topic to a later stage.

Our algorithmic results follow from a careful analysis of the set

$$\mathcal{R} = W\mathcal{F} := \{y = Wx \mid x \in \mathcal{F}\},$$

namely the projection of the feasible set with respect to the matrix W. Let us first explain why understanding this set can have important algorithmic consequences. Assume, for example, that $\mathcal{R} = Q \cap \mathbb{Z}^d$ holds, where

$$Q := WP = \{Wx \mid x \in P\}.$$

In this case we can solve Problem (5.1) with two oracle calls as follows. First, use the optimization oracle to obtain $y^* \in \arg\min\{f(y) \mid y \in Q \cap \mathbb{Z}^d\}$. This is possible since $\mathbb{Z}^d$ is clearly a lattice, thus the latter problem has the required form. Then, use the fiber oracle to obtain $x^* \in \mathcal{F}$ with $Wx^* = y^*$. The oracle is guaranteed to return a point $x^* \in \mathcal{F}$ since we assumed that $\mathcal{R} = Q \cap \mathbb{Z}^d$, namely that *every* integer point in Q has a feasible pre-image under the projection with W. The obtained x^* is clearly an optimal solution. In the following remark we give a concrete example of a class of matrices with this property.

Remark 5.0.3. One important case in which $\mathcal{R} = Q \cap \mathbb{Z}^d$ is the case of a totally unimodular matrix $\binom{W}{A}$ (see e.g. [Sch86, Theorem 19.1]). In this case one can show the inclusion $Q \cap \mathbb{Z}^d \subset \mathcal{R}$ as follows. Let $y \in Q \cap \mathbb{Z}^d$. Since $y \in Q$ there exists $x \in \mathbb{R}^n$ such that $Ax \leq b$ and $Wx = y$. Since $\binom{W}{A}$ is totally unimodular and $y \in \mathbb{Z}^d$, the solution set to the latter system is an integral polyhedron. Thus, there exists an integral point $\bar{x} \in \mathbb{Z}^n$ with $A\bar{x} \leq b$ and $W\bar{x} = y$. This implies that $y \in \mathcal{R}$.

It is, unfortunately, rarely the case that $\mathcal{R} = Q \cap \mathbb{Z}^d$, as typically one has

$$\mathcal{R} \subsetneq Q \cap \mathbb{Z}^d,$$

namely, the set $Q \cap \mathbb{Z}^d$ contains *holes*, i.e., points without pre-images in $\mathcal{F}$. We illustrate this with the following simple example.

Example 5.0.4. *Let $n = 3, d = 2, P = \{x \in \mathbb{R}^3 \mid 0 \leq x_i \leq 3, \ i = 1, 2, 3\}$, and consider the matrix*

$$W = \begin{pmatrix} 1 & 2 & 1 \\ -2 & 0 & 1 \end{pmatrix}.$$

Figure 5.1 illustrates the polyhedron Q and the set $\mathcal{R}$, which corresponds to the thick points. All other points are holes.

In this more common situation, the latter simple strategy can not be directly applied. One can still hope, however, to decompose the problem into sub-problems, each solvable in this way. Ideally, this decomposition should have the form

$$\mathcal{R} = \bigcup_{i \in I} Q^i \cap \Lambda^i,$$

where $Q^i \subset \mathbb{R}^d$ are polyhedra and $\Lambda^i \subset \mathbb{Z}^d$ are affine sub-lattices, the description of which can be efficiently computed from the input data. Then, $k = 2|I|$ oracle calls are sufficient to solve the problem, by simply repeating our simple procedure for every sub-problem $i \in I$, defined over $Q^i \cap \Lambda^i$, and taking the best solution, among the $|I|$ resulting candidates. It is hence of particular interest to study such efficient representations, trying to minimize k, while maintaining the property that both Q^i and Λ^i are efficiently computable.

Our main contribution provides such a decomposition. Concretely, we show strong existential bounds on some important parameters of such decompositions. These bounds, in turn, lead to strong bounds on efficiently computable decompositions, which are later exploited to obtain efficient algorithms.

It is now evident that we deal with a problem of representability of sets of integer vectors. Indeed, what we seek in our decomposition is a way to cover
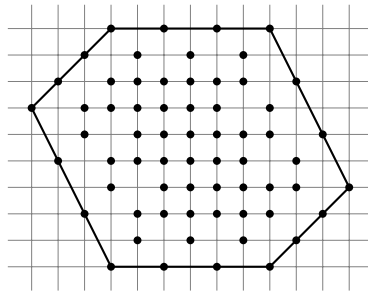
Figure 5.1: An illustration of the set $\mathcal{R}$ and the notion of holes.

all points in $\mathcal{R}$ by "simple" sets, with the property that none of these sets contains a hole, namely a point in $(Q \cap \mathbb{Z}^d) \setminus \mathcal{R}$. One can almost equivalently ask: *How complicated can the set of holes be?*

Our result can hence be seen as a polyhedral variant of the *Frobenius problem* [1]. Given a set $S = \{a_1, \cdots, a_m\} \subset \mathbb{Z}_+$ of positive integers with $\gcd(S) = 1$, the Frobenius problem asks to find the largest integer $k \in \mathbb{Z}_+$ that can not be represented as a positive integer combination of numbers in S. The Frobenius problem is known to be NP-hard [RA96] in all but a few special cases [Kan92, Kan89]. Several bounds on the Frobenius number were also proven [EG72, BS62, Bra42].

For a positive integer s, let $[s] = \{1, \cdots, s\}$. In higher dimensions one can define the following generalization of the Frobenius problem, called the *diagonal Frobenius problem* [2] [AH10]. Given an $d \times n$ integer matrix M with the property that $\text{cone}(M) = \{M\lambda \mid \lambda \in \mathbb{R}_+^n\}$ forms a full-dimensional pointed cone, and such that $M\mathbb{Z}^n = \mathbb{Z}^d$, find the smallest $t \in \mathbb{Z}$ with the property that

$$(tv + \text{cone}(M)) \cap \mathbb{Z}^d \subset \{Mx \mid x \in \mathbb{Z}_+^m\},$$

where $v = \sum_{i \in [m]} M_{.i}$ is the sum of the columns of M. Intuitively, the diagonal Frobenius number is the smallest factor by which one needs to shift the cone $\text{cone}(M)$ inwards (in the direction $v \in \text{cone}(M)$), so that every integer point in it be expressible as a positive integer combination of the columns of M (For an illustration see the second row in Figure 5.2). The following result of Aliev and Henk [AH10] proves a strong bound on the diagonal Frobenius number.

Theorem 5.0.5 (Aliev and Henk 2010). *Let $M \in \mathbb{Z}^{d \times n}$, such that $M\mathbb{Z}^n = \mathbb{Z}^d$ with $\text{cone}(M)$ pointed. Then the diagonal Frobenius number of M is at most*

$$\mathbf{c}(M) = \frac{(n-d)\sqrt{n}}{2}\sqrt{\det(MM^\mathsf{T})}.$$

Theorem 5.0.5 guarantees that the set $M\mathbb{Z}_+^n$ becomes very regular in the cone $\text{cone}(M)$ shifted by the vector $\mathbf{c}(M)v$. For our purposes we need a similar result for arbitrary polyhedra, instead of cones. To this end we define a notion of regularity, suitable for our needs.

Definition 5.0.6 (Δ-regular set)**.** We call a set $\mathcal{S} \subset \mathbb{Z}^d$ Δ-regular, with respect to a region $B \subset \mathbb{R}^d$, if there exists a family of full-dimensional affine

---

[1] also known as the *coin problem.*

[2] We note that there are several ways to define such a generalization.

sub-lattices $\Lambda_1, \cdots, \Lambda_k$ of $\mathbb{Z}^d$ with determinants $\det(\Lambda_i) \leq \Delta$ such that

$$\mathcal{S} \cap B = \bigcup_i \Lambda_i \cap B. \qquad (5.2)$$

Theorem 5.0.5 can be restated in terms of our new definition as follows. For a matrix M satisfying the conditions of Theorem 5.0.5, the set $\mathbb{Z}^d$ is 1-regular with respect to $B = \mathbf{c}(M)v + \mathrm{cone}(M)$. Furthermore, only the lattice $\Lambda = \mathbb{Z}^d$ is needed to certify this fact.

Our main result proves a similar statement for a much more general setup. Firstly, the matrix M is replaced with the arbitrary matrix W. Secondly, the admissible set of positive combinations is no longer the convenient set $\mathbb{Z}_+^n$, but rather the set $\mathcal{F}$. Finally, we prove regularity with respect to a polyhedron $Q' \subset Q$. Since Q can be bounded, Q' can no longer be a translate of Q. We use instead the notion of α-*inscribed polyhedron* defined as follows. Let $R \subset \mathbb{R}^d$ be a polyhedron, and let $B(\alpha) = \{x \in \mathbb{R}^d \mid \|x\|_\infty \leq \alpha\}$ denote the ℓ_∞ ball with radius α. Then the α-inscribed polyhedron of R is the polyhedron

$$R_\alpha := \{x \in R \mid x + B(\alpha) \subset R\}.$$

We are now ready to state our main result. We henceforth fix the notations $P, A, b, W, \mathcal{F}, \mathcal{R}, Q, d$ and n to represent the input to our problem. We denote by Δ and ω the maximum absolute sub-determinant of A, and the largest absolute-value of an entry in W, respectively.

Theorem 5.0.7. *$\mathcal{R}$ is $\boldsymbol{\delta}$-regular with respect to the $\boldsymbol{\gamma}$-inscribed polyhedron $Q_{\boldsymbol{\gamma}}$ of Q, where $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ are bounded polynomially in Δ, ω and n.*

Remark 5.0.8. We remark that one can also define a clean notion of a *polyhedral Frobenius number* as follows. Given two matrices $A \in \mathbb{Z}^{n \times d}$ and $W \in \mathbb{Z}^{d \times m}$ let the polyhedral Frobenius number of A and W be

$$F(A, W) = \min \left\{ \|\boldsymbol{\gamma}, \boldsymbol{\delta}\|_\infty \mid \mathcal{R} \text{ is } \boldsymbol{\delta}\text{-regular with respect to } Q_{\boldsymbol{\gamma}} \ \forall b \in \mathbb{Z}^m \right\},$$

where Q, P and $\mathcal{R}$ are defined from A, W and b, as before. We stress that one can define a polyhedral Frobenius number in various alternative ways. With the latter definition, however, Theorem 5.0.7 can be restated as follows. The polyhedral Frobenius number $F(A, W)$ is polynomially bounded in Δ, ω and n, and exponentially by d. For an illustration of the conection between the classical Frobenius problem, the diagonal Frobenius problem and the polyhedral Frobenius problem see Figure 5.2.

The remainder of the chapter is organized as follows. In Section 5.1 we prove Theorem 5.0.7. In Section 5.2 we use Theorem 5.0.7 to prove algorithmic results for various classes of non-linear integer optimization problems.
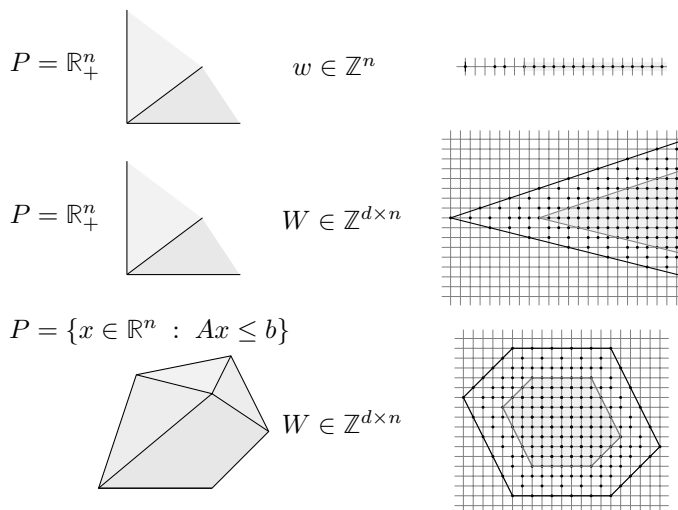
Figure 5.2: Given a polytope P and a matrix W we illustrates the the relationship between $W(P \cap \mathbb{Z}^n)$ and $WP \cap \mathbb{Z}^n$. From top to bottom, the first row shows the classical Frobenius setting, the second row the diagonal Frobenius problem and the last rows shows the polyhedral Frobenius setting (see Remark 5.0.8).

5.1   A proof of Theorem 5.0.7

In this section we prove Theorem 5.0.7. For that we first introduce notation and we prove an auxiliary lemma that adapts Theorem 5.0.5 to our needs. Then we prove our main theorem.

We start with some notation. Let $B, C \subset \mathbb{R}^d$ and let $D \in \mathbb{R}^{m \times d}$ be a matrix. With $B + C$ we denote the Minkowski sum $\{x \in \mathbb{R}^d \mid x = b + c \text{ with } b \in B \text{ and } c \in C\}$. With DB we denote the set $\{x \in \mathbb{R}^m \mid x = Db \text{ with } b \in B\}$. Further, we denote with $D_{i,\star}$ the i-th row of D and with $D_{\star,i}$ the i-th column. The operator $\lfloor \cdot \rfloor$ maps component wise every entry to the largest integer smaller than or equal to the corresponding entry. Finally, let $\|D\|_{max} := \max_{i,j} |D_{i,j}|$ denote the maximum absolute value of an entry of D.

Lemma 5.1.1. Let $M \in \mathbb{Z}^{d \times n}$, such that $\mathrm{cone}(M) = \mathbb{R}^n$ and let $\Lambda = M\mathbb{Z}^n$. Let $z \in \Lambda$ and let $\alpha := \|z\|_\infty$. Then z can be expressed as $z = M\lambda$ such that $\lambda \in \mathbb{Z}^n_+$ and $\|\lambda\|_\infty \leq \mathbf{p}(\alpha, n, \omega)$, where $\mathbf{p}(\alpha, n, \omega) \in \mathbb{R}_+[x_1, x_2, x_3]$ is a polynomial in α, n and $\omega := \|M\|_{max}$.

Proof. To start with, we show why we may assume that $\Lambda = \mathbb{Z}^d$. Let B be the Korkin-Zolotariv basis of Λ [KZ73]. Then, a well known property is that the following inequality holds $\|B_{\star,1}\|_2 \cdots \|B_{\star,d}\|_2 \leq a d^d \det(\Lambda)$ (see [LLS90, Theorem 2.3]), where a is a universal constant. It follows that the entries of B are bounded polynomially in $\det \Lambda$. Therefore, there is also a polynomial bound for the entries of its inverse matrix B^{-1}. We can hence transform M, z and Λ by B^{-1} to arrive at a matrix $M' = B^{-1}M$, a vector $z' = B^{-1}z$ and a lattice $\Lambda' = B^{-1}\Lambda$. The entries in M' and z' are polynomially bounded in the entries in M and z, respectively. Furthermore, Λ' becomes the standard lattice, that is $\Lambda' = \mathbb{Z}^d$. We hence assume hereafter that $\Lambda = \mathbb{Z}^d$.

Case 1. Assume that $\{M_{\star,j} \mid j = 1, \ldots, n\} = \{-M_{\star,j} \mid j = 1, \ldots, n\}$, i.e., the negative of every column in M is also a column in M. Let $p = (2^{d-1}\omega^{d-1}, \ldots, 2^0\omega^0)^\mathsf{T}$. It holds that $p^\mathsf{T} M_{\star,j} \neq 0$ for all $j = 1, \ldots, n$. Without loss of generality, we assume that n is even, that $p^\mathsf{T} M_{\star,j} > 0$ for all $j = 1, \ldots, n/2$ and that $M_{\star,j} = -M_{\star,n/2+j}$ for $j = 1, \ldots, n/2$. This implies that $\mathrm{cone}(M_{\star,1}, \ldots, M_{\star,n/2})$ is pointed.

By Caratheodory's Theorem (see e.g. [Gru07, Theorem 3.1]), we can express z as a positive combination of at most d linearly independent columns of M, say $z = \sum_{j=1}^d \gamma_{i_j} M_{\star,i_j}$ with $\boldsymbol{\gamma}_{\mathbf{i_j}} \in \mathbb{R}_+$. Using Cramer's rule, the Lagrange expansion of determinants and Hadamard's inequality, we can compute the

bound

$$\gamma_{i_j} \le d\alpha\omega^{d-1}(d-1)^{(d-1)/2} =: \boldsymbol{\rho_1}.$$

We set $\gamma_{i_j} = 0$ for $j = d + 1, \ldots, n$.

Let $\mathbf{c} = \mathbf{c}(M)$, defined as in Theorem 5.0.5. Note that $\mathbf{c}$ is polynomially bounded by ω and n. Next, define $\bar{\gamma}_i := \max\{0, \gamma_i - \mathbf{c}\}$ for $i = 1, \ldots, n/2$ and $\bar{\gamma}_i := \min\{\gamma_i, \gamma_{i-n/2} - \mathbf{c}\}$ for $i = n/2 + 1, \ldots, n$. Let $\bar{z} := M\lfloor\bar{\gamma}\rfloor$. Then $z \in \bar{z} + \mathbf{c}\sum_{j=1}^{n/2} M_{\star,j} + \text{cone}(M_{\star,1}, \ldots, M_{\star,n/2})$. It follows from Theorem 5.0.5 that $z - \bar{z}$ can be expressed as a positive integer combination of $M_{\star,1}, \ldots, M_{\star,n/2}$. Let $z - \bar{z} = \sum_{j=1}^{n/2} \mu_j M_{\star,j}$, with $\mu_j \in \mathbb{Z}_+$ be such a combination.

Next, we exploit the fact

$$1 \le p^\mathsf{T} M_{\star,j} \le d2^d\omega^d \text{ for all } j \in \{1, \ldots, n/2\}.$$

It holds that $p^\mathsf{T}(z - \bar{z}) \le p^\mathsf{T}((\mathbf{c} + 1)\sum_{j=1}^{n/2} M_{\star,j}) \le (\mathbf{c} + 1)n/2d2^d\omega^d$. In particular, this implies that

$$\mu_j \le (\mathbf{c} + 1)n/2d2^d\omega^d =: \boldsymbol{\rho_2}.$$

Finally, let $\lambda := \lfloor\bar{\gamma}\rfloor + \mu$. It holds that $\lambda \in \mathbb{Z}_+^n$, $z = M\lambda$ and $\|\lambda\|_\infty \le \mathbf{p}(\alpha, n, \omega) := \boldsymbol{\rho_1} + \boldsymbol{\rho_2}$. This completes the proof for Case 1.

Case 2. In the general case $\{M_{\star,j} \mid j = 1, \ldots, n\} \ne \{-M_{\star,j} \mid j = 1, \ldots, n\}$. Without loss of generality we assume that $-M_{\star,1} \notin \{M_{\star,j} \mid j = 1, \ldots, n\}$.

Since $\text{cone}(M) = \mathbb{R}^d$ there exists, by Caratheodory's Theorem, a selection of at most d linearly independent columns of M, such that $-M_{\star,1} = \sum_{j=1}^d \xi_j M_{\star,i_j}$ with $\xi_j \in \mathbb{R}_+$. With $\delta_1 := \det(M_{\star,i_1}, \ldots, M_{\star,i_d})$ and $\delta_j = \delta_1\xi_j \in \mathbb{Z}_+$ it follows that

$$-M_{\star,1} = (\delta_1 - 1)M_{\star,1} + \sum_{j=1}^d \delta_j M_{\star,i_j} \text{ and } \delta_j \le \omega^d d^{d/2} =: \boldsymbol{\rho_3} \qquad (5.3)$$

for all $j = 1, \ldots, d$. From this it follows that we can insert $-M_{\star,1}$ to the set $\{M_{\star,j} \mid j = 1, \ldots, n\}$ with the slight modification that whenever a multiplier β for column $-M_{\star,1}$ is used in a representation, we replace it by β times its expression for (5.3). By performing the latter replacement for all n columns independently, we obtain the general bound

$$\|\lambda\|_\infty \le \mathbf{p}(\alpha, n, \omega) := \boldsymbol{\rho_1} + \boldsymbol{\rho_2} + n\boldsymbol{\rho_2}\boldsymbol{\rho_3}.$$

$\square$

We are now ready to prove Theorem 5.0.7

Proof of Theorem 5.0.7. In order to distinguish between vectors in fixed dimension d from those in variable dimension n, we denote elements in the n-dimensional space with capital letters and elements in the d-dimensional space with small letters. Without loss of generality we assume that P is given in the form $\{x \in \mathbb{R}_+^n \mid Ax = b\}$. We can do so by introducing m slack-variables and decomposing any vector into the difference of two positive vectors of the same dimension, i.e. $Ax - Ay + Iz = b$ with $x, y \in \mathbb{R}_+^n$ and $z \in \mathbb{R}_+^m$. Note that the dimension only grows linearly and that the maximum absolute sub-determinant remains the same. For each orthant $\mathcal{O}_i$ we define $\mathcal{H}_i$ to be the Hilbert basis of the cone $\{x \in \mathcal{O}_i \mid Ax = 0\}$. By Cramer's rule we can conclude that we can express $\{x \in \mathcal{O}_i \mid Ax = 0\}$ as $\text{cone}(G_1, \ldots, G_k)$ with $G_j \in \mathcal{O}_i \cap \{-\Delta, \ldots \Delta\}^d$ for $j = 1, \ldots, k$. Further, it is well known that $\mathcal{H}_i \subset \{x \in \mathcal{O}_i \cap \mathbb{Z}^n \mid \|x\|_\infty \leq \boldsymbol{\eta_1}\}$ with

$$\boldsymbol{\eta_1} := d\Delta.$$

This implies that for each $H \in \mathcal{H}_i$ it holds that

$$\|WH\|_\infty \leq n\omega d\Delta =: \boldsymbol{\eta_2}.$$

For an introduction to Hilbert bases see, e.g. [Sch86, Section 16.4].

Let $v_1, v_2, \ldots, v_l$ denote the vertices of Q. For each $j = 1, \ldots, l$ there exists a vertex V_j of P such that $WV_j = v_j$. Assuming that $\mathcal{F}$ is non empty we can conclude that for each $j = 1, \ldots, l$ there exist a $Y_j \in \mathcal{F}$ such that

$$\|V_j - Y_j\|_\infty \leq n\Delta$$

(see [Sch86, Theorem 17.3]). In words, there exists a feasible integral point Y_j close to each vertex V_j for $j = 1, \ldots, l$. Then, with $y_i := WY_i$ it follows that

$$\|v_i - y_i\|_\infty \leq n^2\omega\Delta =: \boldsymbol{\eta_3}.$$

For an illustration of the points in d-dimensional space see Figure 5.3.

We will proceed from here as follows. For suitable polynomials $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$, we will consider the pre-image Z of an arbitrary point $z \in Q_{\boldsymbol{\gamma}} \cap \mathcal{R}$, construct an affine lattice Λ induced by the Hilbert basis representation of $Y_1 - Z, \ldots, Y_l - Z$ containing z, such that $\det(\Lambda) \leq \boldsymbol{\delta}$, and then prove that this lattice intersected with $Q_{\boldsymbol{\gamma}}$ is contained in $\mathcal{R}$. This will then prove our theorem.

Let $z \in Q_{\boldsymbol{\gamma}} \cap \mathcal{R}$ and $Z \in \mathcal{F}$ such that $WZ = z$. We first exhibit our construction.

For each index $j \in \{1, \ldots, l\}$ we consider the vector $Y_j - Z$. Let us say that, $Y_j - Z$ is contained in the orthant $\mathcal{O}_{i_j}$, with $i_j \in \{1, \ldots, 2^n\}$. In view of [Seb90], we can express $Y_j - Z$ as the positive integer combination of at most $2n - 2$ elements of the Hilbert basis $\mathcal{H}_{i_j}$, i.e.

$$Y_j - Z = \sum_{k=1}^{2n-2} \lambda_j^k H_j^k,$$

with $\lambda_j^k \in \mathbb{Z}_+$ and $H_j^k \in \mathcal{H}_{i_j}$. Note that all points in $\{Z + \sum_{k=1}^{2n-2} \gamma_j^k H_j^k \mid \gamma_j^k \in \mathbb{Z}_+,$ and $\gamma_j^k \leq \lambda_j^k\}$ are feasible, i.e. they are a subset of $\mathcal{F}$. This follows from the fact that $H_j^k \in \{x \in \mathcal{O}_{i_j} \mid Ax = 0\}$ for every $k = 1, \ldots, 2n - 2$.

Let

$$\mathbf{p} := \mathbf{p}((2n-2)l\boldsymbol{\eta_1}, (2n-2)l, \boldsymbol{\eta_2})$$

be the polynomial defined in Lemma 5.1.1. Letting $\boldsymbol{\eta_4} := (2n-2)l(\mathbf{p}+1)\boldsymbol{\eta_1}$, we define

$$\bar{\lambda}_j^k := \max\{0, \lambda_j^k - \boldsymbol{\eta_4}\}$$

and

$$\bar{Y}_j := Z + \sum_{k=1}^{2n-2} \bar{\lambda}_j^k H_j^k. \tag{5.4}$$

Notice that $\bar{y}_j := W\bar{Y}_j$ remains close to its corresponding vertex v_j. That is

$$\|v_j - \bar{y}_j\|_\infty \leq \boldsymbol{\eta_3} + (2n-2)\boldsymbol{\eta_2}\boldsymbol{\eta_4} =: \boldsymbol{\eta_5}.$$

Choosing $\boldsymbol{\gamma} \geq \boldsymbol{\eta_5}$ we ensure that z is sufficiently far from each vertex v_j so that at least one λ_j^k must be greater or equal than $\boldsymbol{\eta_4}$ for each j.

For simplicity, we assume that $\bar{\lambda}_j^k > 0$ for all j and k. This can be assumed without loss of generality, as if $\lambda_j^k = 0$ we can simply modify Y_j and consider a representation of it with one Hilbert basis element less. Let $h_j^k := WH_j^k$ for every $j = 1, \ldots, 2n - 2$ and $k = 1, \ldots, l$. We define the affine lattice

$$\Lambda = \{x \in \mathbb{Z}^d \mid x = z + \sum_{j=1}^{l} \sum_{k=1}^{2n-2} \gamma_j^k h_j^k, \ \gamma_j^k \in \mathbb{Z}^{(2n-2)\times l}\}.$$

We can bound the determinant of Λ by any sub-lattice induced by d linearly independent h_j^i-s. Hence, by Hadarmad's inequality and since $\|h_j^k\|_\infty \leq \boldsymbol{\eta_2}$, it holds that $\det(\Lambda) \leq \boldsymbol{\delta} := \boldsymbol{\eta_2^d} d^{d/2}$. Next, we define the matrix

$$M := [h_1^1, \ldots, h_1^{(2n-2)}, \ldots, h_l^1, \ldots, h_l^{(2n-2)}].$$

It holds that $\|M\|_{max} \leq \boldsymbol{\eta_2}$. In order to apply Lemma 5.1.1 let us first verify Claim 1.

Claim 1. $\mathrm{cone}(M) = \mathbb{R}^d$.

Proof of Claim 1. Assume that $\mathrm{cone}(M) \neq \mathbb{R}^d$. Then there exists a $u \in \mathbb{R}^d$ with $\|u\|_2 = 1$ defining a half-space $\{x \in \mathbb{R}^d \mid u^\mathsf{T} x \leq 0\}$ such that $\mathrm{cone}(M) \subset \{x \in \mathbb{R}^d \mid u^\mathsf{T} x \leq 0\}$. Since $z \in Q_{\boldsymbol{\gamma}}$ it holds that $z + B(\boldsymbol{\gamma}) \subset Q$. This implies that there exists a vertex v_i such that $u^\mathsf{T} v_i - u^\mathsf{T} z \geq \boldsymbol{\gamma}$. On the one hand, it holds that

$$\boldsymbol{\gamma} \leq u^\mathsf{T}(v_i - y_i + y_i - z) \leq \boldsymbol{\eta_3} + \sum_{k=1}^{2n-2} \lambda_i^k u^\mathsf{T} h_i^k.$$

On the other hand, for each $k = 1, \ldots, 2n-2$ it holds that $u^\mathsf{T} h_i^k \leq \|h_i^k\|_\infty \leq \boldsymbol{\eta_2}$. It follows that for some $j \in \{1, \ldots, 2n-2\}$ we have that $u^\mathsf{T} h_i^j > 0$ and $\lambda_i^j \geq \boldsymbol{\eta_4}$. By construction, this implies that h_i^j is a column of M, contradicting that $\mathrm{cone}(M) \subset \{x \in \mathbb{R}^d \mid u^\mathsf{T} x \leq 0\}$.   □

It remains to show that $\mathcal{R}$ is $\boldsymbol{\delta}$-regular with respect to $Q_{\boldsymbol{\gamma}}$. We split the proof into Claim 2 and 3. In Claim 2 we show that for every $j = 1, \ldots, l$, all lattice points sufficiently close to $\bar{y}_j$ have a feasible pre-image.

Claim 2. For every $\gamma \in \mathbb{Z}_+^{(2n-2)l}$ with $\|\gamma\|_\infty \leq \mathbf{p} + 1$ it holds that $\bar{y}_j + M\gamma$ has a feasible pre-image, i.e. it is not a hole.

Proof of Claim 2. We prove the claim by showing that $\bar{Y}_j + (2n - 2)l(\mathbf{p} + 1)H_i^k \in P$ for every i, j and k. This will then imply the slightly stronger result, that

$$\bar{y}_j + M\gamma$$

is feasible for any $\gamma \in \mathbb{Z}_+^{(2n-2)l}$ with $\|\gamma\|_1 \leq (2n - 2)l(\mathbf{p} + 1)$.

In order do derive a contradiction, let us assume that the latter does not hold for $j = 1$ and $i = 2$, i.e. that $Y_1 + (2n - 2)l(\mathbf{p} + 1)H_2^1 \notin P$. The only constraints defining P that can be violated by this vector are the non-negativity constraints, thus some component of this vector must be strictly negative. Let us assume that the first component is negative. We have an upper and a lower bound for this component, namely

$$-(2n - 2)l(\mathbf{p} + 1)\boldsymbol{\eta_1} \leq (Y_1 + (2n - 2)l(\mathbf{p} + 1)H_2^1)_1 < 0.$$

Since $Z + (2n-2)l(\mathbf{p}+1)H_2^1$ is feasible, it must hold that $(-H_1^k)_1 \geq 0$ for all $k = 1, \ldots, 2n - 2$. In particular, there must be at least one index $k \in \{1, \ldots, 2n - 2\}$ such that $(-H_1^k)_1 \geq 1$. Hence,

$$0 \leq \left(Y_1 + (2n-2)l(\mathbf{p}+1)H_2^1 - \boldsymbol{\eta_4} \sum_{k=1}^{2n-2} H_1^k\right)_1 = \left(\bar{Y}_1 + (2n-2)l(\mathbf{p}+1)H_2^1\right)_1 < 0.$$

We obtained a contradiction. □

We now use Claim 2 to show that all points in $\Lambda \cap Q_{\boldsymbol{\gamma}}$ have pre-images, i.e. they are not holes.

Claim 3. $\Lambda \cap Q_{\boldsymbol{\gamma}} \subset \mathcal{R}$.

Proof of Claim 3. Let $\bar{z} \in \Lambda \cap Q_{\boldsymbol{\gamma}}$. We prove that there exists a $\bar{Z} \in P \cap \mathbb{Z}^d$ such that $\bar{z} = W\bar{Z}$. Note that $Q_{\boldsymbol{\gamma}} \subset \text{conv}(\bar{y}_1, \ldots, \bar{y}_l)$. By Caratheodory's theorem there exist $i_1, \ldots, i_d \in \{1, \ldots, l\}$, such that $\bar{z} \in \text{conv}(z, \bar{y}_{i_1}, \ldots, \bar{y}_{i_d})$. Without loss of generality we may assume that $i_1 = 1, \ldots, i_d = d$. Let $\alpha \in [0,1]$ and $\alpha_j \in \mathbb{R}_+$ for $j = 1, \ldots, d$, such that $\bar{z} = \alpha z + (1-\alpha) \sum_{j=1}^d \alpha_j \bar{y}_j$ and $\sum_{j=1}^d \alpha_j = 1$. Hence, using (5.4), $\bar{z}$ is the image under W of a, not necessarily integral point

$$Z + (1-\alpha) \sum_{j=1}^d \alpha_j \sum_{k=1}^{2n-2} \bar{\lambda}_j^k H_j^k,$$

which is included in P. We can approximate this point by

$$\hat{Z} = Z + \sum_{j=1}^d \sum_{k=1}^{2n-2} \lceil (1-\alpha)\alpha_j \bar{\lambda}_j^k \rceil H_j^k.$$

Clearly, $\hat{z} = W\hat{Z} \in \Lambda$. Let $L := \sum_{j=1}^d \sum_{k=1}^{2n-2} (\lceil (1-\alpha)\alpha_j \bar{\lambda}_j^k \rceil - (1-\alpha)\alpha_j \lambda_j^k)H_j^k$. Since $Z + L \in P$ and $Y_j + L \in P$ holds for all $j = 1, \ldots, d$ (see Claim 2), $\hat{Z}$ must be feasible. From Claim 2 it follows again that $\hat{z} + M\gamma$ is feasible for any $\gamma \in \mathbb{Z}_+^{2nl-2l}$ with $\|\gamma\|_\infty \leq \mathbf{p}$. It holds that $\|\bar{z} - \hat{z}\|_\infty \leq (2n-2)l\boldsymbol{\eta_1}$. Finally, we can apply Lemma 5.1.1 to guarantee the existence of $\bar{Z} \in \mathcal{F}$ such that $\bar{z} = W\bar{Z}$. □

This completes the proof of the theorem. □
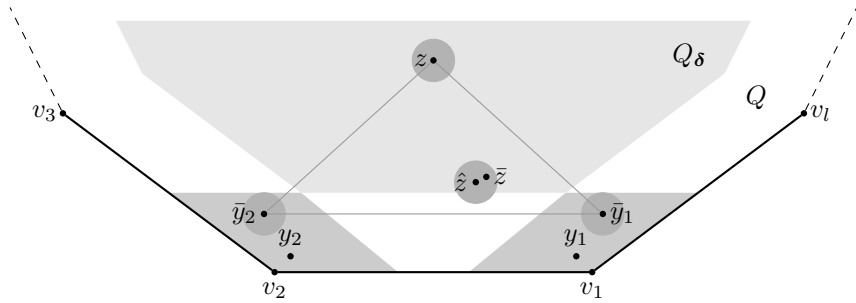
Figure 5.3: An illustration of the notation in the proof of Theorem 5.0.7.

5.2   Applications to non-linear integer optimization

We describe next a general algorithmic framework that allows us to apply Theorem 5.0.7 to solve variable-dimension non-linear integer optimization problems. More precisely, we show a general purpose algorithm that solves Problem (5.1) with a number of oracle calls that is polynomial in the input size, Δ, n and ω. For brevity, we will henceforth say "in polynomial time" to imply a running time of the latter type. We recall that the oracles available to our algorithm are an optimization oracle and a fiber oracle (see Definition 5.0.2). We stress that the dependence in d can be exponential. We assume in this section that d is an arbitrary fixed constant. We later mention a number of concrete examples of problem classes, for which, using polynomial-time implementations of the oracles, our algorithm runs in polynomial time in the encoding length of the input.

Our main goal in this section is to prove the following theorem.

Theorem 5.2.1. *Let d be any fixed constant. There is an algorithm that solves the non-linear optimization problem*

$$\operatorname{opt}\left\{f(Wx) \mid Ax \leq b,\ x \in \mathbb{Z}^n\right\},$$

with input $A \in \mathbb{Z}^{m \times n}$, $W \in \mathbb{Z}^{d \times n}$, $b \in \mathbb{Z}^m$ and $f : \mathbb{R}^d \to \mathbb{R}$. The number of oracle calls it performs (to the optimization and fiber oracles) is polynomial in n, the maximum sub-determinant Δ of A and the unary encoding length of W.

To simplify notation, we will henceforth restrict our attention to minimization problems. We stress that Theorem 5.2.1 works also for maximization problems.

Our algorithm works with an inequality description of the polyhedra Q and Q_γ. Since the input only provides implicit representations of these polyhedra, we need the following lemma, which also states a useful connection between the two descriptions.

Lemma 5.2.2. *One can compute in polynomial time a matrix $F \in \mathbb{Z}^{q \times d}$ and vectors $g, g' \in \mathbb{Z}^q$ such that $Q = \{x \in \mathbb{R}^d \mid Fx \le g\}$ and $Q_\gamma = \{x \in \mathbb{R}^d \mid Fx \le g'\}$, with $\|F\|_{max} \le (n\omega\Delta)^{d-1}(d-1)^{(d-1)/2}$ and*

$$|g_i - g'_i| \le \gamma \|F_{i,\star}\|_\infty$$

for every $i \in [q]$.

Proof. We start with some notation. Let v_1 and v_2 denote two adjacent vertices of Q. Together they define the edge $\mathrm{conv}(v_1, v_2)$ of Q. In the following we call a vector e an *edge-direction* of an edge $\mathrm{conv}(v_1, v_2)$, if $e \in \mathrm{lin}(v_2 - v_1)$.

To prove the lemma, we exploit that each edge-direction of Q is the image (under the linear mapping W) of an edge-direction of P. An edge-direction E of P, which corresponds to an edge-direction of Q, can be expressed as the intersection of $n - 1$ linearly independent facets. Let us assume without loss of generality that these are $A_{1,\star}, \ldots, A_{n-1,\star}$. Applying Cramer's rule we know that there exists a non-trivial solution $E \in \mathbb{Z}^n$ such that $A_{i,\star}E = 0$ for all $i \in \{1, \ldots, n-1\}$ and $\|E\|_\infty \le \Delta$. Let $e := WE$. It follows that, $\|e\|_\infty \le n\omega\Delta$.

A facet of Q is defined by $d - 1$ linear independent edge-directions, say $e_1, \ldots, e_{d-1}$. Then, a facet defining vector $F_{i,\star}$ is defined by a non-trivial solution to $e_i^\mathsf{T} c = 0$. Using Cramer's rule and Hadarmad's inequality we can choose $F_{i,\star} \in \mathbb{Z}^d$ such that $\|F_{i,\star}^\mathsf{T}\|_\infty \le (n\omega\Delta)^{d-1}(d-1)^{(d-1)/2}$. It remains to note that there is only a polynomial number of possible $F_{i,\star}$. Hence, one can compute F and g by brute force with linear programming [Sch86].

We can now find an inequality description of Q_γ as follows. First, by normalizing the inequalities defining Q, i.e., by setting $\bar{F}_{i,\star} = \frac{1}{\|F_{i,\star}\|_\infty} F_{i,\star}$ and $\bar{g}_i = \frac{1}{\|F_{i,\star}\|_\infty} g_i$, one easily verifies that

$$Q_\gamma = \{x \in \mathbb{R}^d \mid \bar{F}_{i,\star}^\mathsf{T} x \le \bar{g}_i - \gamma \;\; \forall i \in [q]\}.$$

A description with integral coefficients is hence given by

$$Q_{\boldsymbol{\gamma}} = \{x \in \mathbb{R}^d \mid F_{i,\star}^{\mathsf{T}} x \leq g_i - \boldsymbol{\gamma}\|F_{i,\star}\|_{\infty} \;\; \forall i \in [q]\},$$

so we can set $g_i' = g_i - \boldsymbol{\gamma}\|F_{i,\star}\|_{\infty}$ for all $i \in [q]$. The bound $|g_i - g_i'| \leq \boldsymbol{\gamma}\|F_{i,\star}\|_{\infty}$ immediately follows. $\qquad\square$

As was discussed in the introduction, our algorithmic approach relies on a decomposition of the problem into "sufficiently regular" sub-problems. Each sub-problem corresponds to a projected feasible set $R \cap \Lambda \subset \mathbb{Z}^d$ containing no holes, where R is a polyhedron and Λ is a lattice. Then, the optimization oracle is invoked to obtain a point $y^* \in R \cap \Lambda$ attaining

$$\min\left\{f(y) \mid y \in R \cap \Lambda\right\},$$

and a point $x^* \in \mathcal{F}$ is computed with $Wx^* = y^*$ using the fiber oracle. The best solution across all sub-problems is then an optimal solution.

We distinguish between two types of sub-problems. The first type is concerned with the polyhedron $Q_{\boldsymbol{\gamma}}$, i.e., such sub-problems optimize over the restricted feasible region

$$\mathcal{F}' := \{x \in \mathcal{F} \mid Wx \in Q_{\boldsymbol{\gamma}}\}.$$

In the following lemma we prove, using Theorem 5.0.7, that the optimal point in this region can be found efficiently.

Lemma 5.2.3. *The problem*

$$\min\left\{f(Wx) \mid x \in \mathcal{F}'\right\}$$

can be solved with a polynomial number of calls to the optimization and fiber oracles.

Proof. As guaranteed by Theorem 5.0.7, for every point $x \in \mathcal{F}'$ there is a lattice Λ_x with determinant at most $\boldsymbol{\delta}$ such that $x \in \Lambda_x$ and $\Lambda_x \cap Q_{\boldsymbol{\gamma}} \subset \mathcal{R}$, i.e., $\Lambda_x \cap Q_{\boldsymbol{\gamma}}$ contains no holes. Consider an optimal solution y^* to the problem

$$\min\{f(y) \mid y \in \Lambda_x \cap Q_{\boldsymbol{\gamma}}\},$$

obtainable by a single oracle call to the optimization oracle. Since $\Lambda_x \cap Q_{\boldsymbol{\gamma}}$ contains no holes, one can obtain, using a call to the fiber oracle, a pre-image $x^* \in \mathcal{F}'$ of y^*. Furthermore, due to $x \in \Lambda_x \cap Q_{\boldsymbol{\gamma}}$ we also know that $f(x^*) \leq f(x)$. Consequently, to minimize over $\mathcal{F}'$ it suffices to consider the problem

$$\min\{f(y) \mid y \in \Lambda \cap Q_{\boldsymbol{\gamma}}\},$$

for every affine lattice Λ with determinant bounded by δ. Next, we bound the number of such lattices.

An affine lattice can be represented by a basis $B \subset \mathbb{Z}^{d \times d}$ and a translation vector $t \in \{B\lambda \mid \lambda \in [0,1)^d\} \cap \mathbb{Z}^d$ as

$$\Lambda = \{t + v \mid \exists z \in \mathbb{Z}^d \ \ v = Bz\}.$$

We can assume that B comprises the columns of a matrix in Hermite Normal Form. The bound on the determinant of the lattice now translates to a bound on the maximum absolute value of an entry in B. We can thus roughly estimate the number of affine lattices by $\delta^{d^2+d} d^d$.

It follows that, by considering every bounded-determinant lattice, as described before, one can obtain the best solution $x \in \mathcal{F}'$ with at most $2\delta^{d^2+d} d^d$ oracle calls. $\qquad\square$

To treat the region $\mathcal{F} \setminus \mathcal{F}'$ we use a recursive decomposition into lower-dimensional problems. To control the number of such problems we use the fact that all points $y = Wx$ for points $x \in \mathcal{F} \setminus \mathcal{F}'$ fall close to the boundary of Q. This fact is used in the following lemma to prove a bound on the number of hyperplanes needed to cover all integer points in $Q \setminus Q_\gamma$.

Lemma 5.2.4. *There is a polynomial time procedure that computes a set $\mathcal{H}$ of hyperplanes parallel to the facets of Q, with the property that all integer points in $Q \setminus Q_\gamma$ lie on at least one hyperplane in $\mathcal{H}$, i.e.,*

$$(Q \setminus Q_\gamma) \cap \mathbb{Z}^d \subset \bigcup_{H \in \mathcal{H}} H$$

In particular, $\mathcal{H}$ has polynomial size.

Proof. Lemma 5.2.2 asserts that Q and Q_γ admit inequality descriptions $Q = \{x \in \mathbb{R}^d \mid Fx \leq g\}$ and

$$Q_\gamma = \{x \in \mathbb{R}^d \mid F_{i,\star}^\mathsf{T} x \leq g_i - \gamma \|F_{i,\star}\|_\infty \ \ \forall i \in [q]\},$$

with $\|F_{i,\star}\|$ polynomially bounded for all $i \in [q]$. We can now use this description to cover all integer points in $Q \setminus Q_\gamma$ with a polynomial number of hyperplanes parallel to the facets of Q. More precisely, for $i \in [q]$ let $\mathcal{H}_i$ denote the set of hyperplanes of the form

$$H_i(s) = \{x \in \mathbb{R}^d \mid F_{i,\star}^\mathsf{T} x = s\},$$

where $s \in \{g_i - \boldsymbol{\gamma}\|F_{i,\star}\|_\infty, g_i - \boldsymbol{\gamma}\|F_{i,\star}\|_\infty + 1, \cdots, g_i\}$ ranges over all integer right-hand sides between $g_i - \boldsymbol{\gamma}\|F_{i,\star}\|_\infty$ and g_i. Note that, by Lemma 5.2.2, the number of such hyperplanes is indeed polynomially bounded. We can now take the union over all facets of Q of the sets $\mathcal{H}_i$, i.e.

$$\mathcal{H} = \bigcup_{i \in [q]} \mathcal{H}_i$$

to arrive at the desired set of hyperplanes. Since the number of facets of Q is polynomially bounded, the lemma is proved.

$\square$

We now have almost all ingredients for the proof of Theorem 5.2.1. The following remark states that the constraint matrix of sub-problems arising by restricting the feasible set to the pre-image of an arbitrary face of Q has a determinant that is polynomially bounded.

Remark 5.2.5. Let $I \subset \mathbb{R}^d$ denote an i-face of Q. Let $F_{i_1,\star}, \ldots, F_{i_{d-i},\star} \in \mathbb{Z}^d$ be the corresponding face defining facets, i.e. $I = \{x \in Q \mid F_{i_j,\star}^\mathsf{T} x = g_{i_j} \;\; \forall j \in [d-i]\}$. Then $W^{-1}I$, the pre-image of I under W, can be expressed as $\{x \in \mathbb{R}^n \mid \bar{A}x \leq \bar{b}\}$ with

$$\bar{A} := [A^\mathsf{T}, (F_{i_1,\star}W)^\mathsf{T}, -(F_{i_1,\star}W)^\mathsf{T}, \ldots, (F_{i_{d-i},\star}W)^\mathsf{T}, -(F_{i_{d-i},\star}W)^\mathsf{T}]^\mathsf{T}$$

and

$$\bar{b} := [b^\mathsf{T}, g_{i_1}, -g_{i_1}, \ldots, g_{i_{d-i}}, -g_{i_{d-i}}]^\mathsf{T}.$$

In particular, note that the maximum absolute sub-determinant of $\bar{A}$ is polynomially bounded.

We are now ready to prove Theorem 5.2.1.

Proof of Theorem 5.2.1. The algorithm starts by computing the inequality descriptions of Q and $Q_{\boldsymbol{\gamma}}$, as in Lemma 5.2.2. Then, the algorithm proceeds by solving the problem

$$\min \{f(Wx) \mid x \in \mathcal{F}'\}$$

by invoking the procedure in Lemma 5.2.3. To treat the region $\mathcal{F} \setminus \mathcal{F}'$, the algorithm obtains first the polynomially-bounded set of hyperplanes $\mathcal{H}$, using the procedure in Lemma 5.2.4. For every hyperplane $H \in \mathcal{H}$, the algorithm recursively solves the $(d-1)$-dimensional problem

$$\min \{f(Wx) \mid x \in \mathcal{F}(H)\},$$

where $\mathcal{F}(H) := \{x \in \mathcal{F} \mid Wx \in H \cap Q\}$. Each such sub-problem admits the same form as the original one. Additionally, Remark 5.2.5 implies that the matrix corresponding to the inequality description of $P' := \{x \in \mathbb{R}^d \mid Wx \in H \cap Q\}$ has determinant that is polynomially bounded, as well. See Figure 5.4 for an illustration of the algorithm.

Finally, since d is fixed, so is the depth of the recursion, implying that the algorithm performs, in total, a polynomial number of oracle calls, and additional polynomial work. □

Theorem 5.2.1 achieves our main algorithmic goal, namely a general-purpose efficient reduction from non-linear integer programming to linear integer programming. The linear integer programs that arise corresponds to the fiber problem, implying that their feasible set is defined from the matrices given in the input data. This property is desirable, since our algorithm does not require solving linear integer programs with a feasible set, whose structure dramatically differs from that of the original non-linear problem. Consequently, our reduction makes it possible to solve a large class of non-linear integer problems using well-known techniques for linear integer programming, such as cutting planes methods etc.

We conclude the chapter by mentioning some concrete class of problems solved by our algorithm. Unless stated otherwise, no polynomial time algorithms were known for these problems. To arrive at the desired polynomial algorithms we need to present polynomial implementations of the optimization and fiber oracles. Let us first list a number classes of non-linear functions for which the optimization oracle can be implemented in polynomial time. We stress that the latter results hold in fixed dimension, i.e., whenever d is an arbitrary, but fixed constant. In all cases the feasible set comprises an arbitrary intersection of a polyhedron and an affine lattice, whose descriptions are provided in the input, and the functions are presented with evaluation oracles.

- **Minimization of convex functions.** Grötschel, Lovász and Schrijer [GLS88] presented an algorithm for the minimization of a convex function.

- **Minimization of bounded degree polynomials.** Del Pia and Weismantel [DPW14] presented an algorithm for minimizing arbitrary degree-two polynomials with integer coefficients in the plane. This result was recently extended by Del Pia, Hildebrand, Weismantel and Zemmer [DPHWZ14] to cubic polynomials in two variables, in the case of a
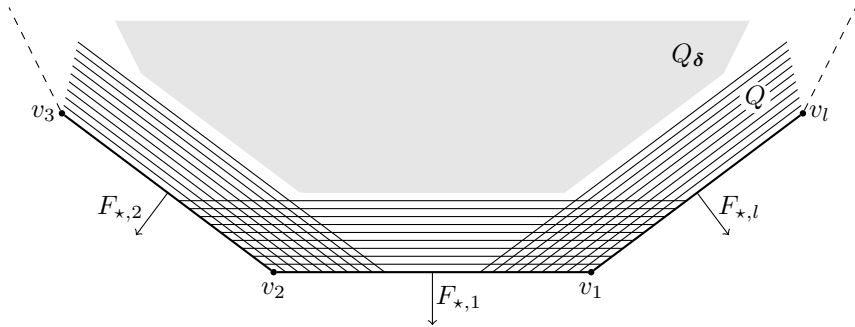
Figure 5.4: An illustration of the algorithm.

bounded polyhedron. With the same restriction on the feasible set, the authors also present a polynomial algorithm for minimizing a homogeneous polynomial with two variables and an arbitrary fixed degree.

- **Approximate maximization of non-negative polynomials.** De Loera, Hemmecke, Köppe and Weismantel [DLHKW06] showed that a polynomial in fixed dimension can be approximately maximized in polynomial time over the integer points in a polyherdon, provided that the polynomial is non-negative over the polyhedron. Concretely, the authors show a fully polynomial-time approximation scheme (FPTAS) for the problem.

We stress that the latter list gives a few prominent examples of classes for which the optimization oracle can be implemented efficiently, but it is far from being a complete list. We remark that in order to obtain an approximate solution to Problem (5.1) it suffices to employ an approximate implementation of the optimization oracle.

We turn to implementations of the fiber oracle. Recall that the fiber oracle is required to provide a point in $\{x \in \mathbb{Z}^n \ : \ Ax \le b, \ Wx = y\}$ for an arbitrary $y \in \mathbb{Z}^d$, if one exists, or correctly report that the latter set is empty.

- **A Constant number of constraints.** Eisenbrand, Vempala and Weismantel [EVW14] recently showed that an integer program with a fixed number of rows can be solved in time polynomial in the dimension and the maximum sub-determinant of the constraint matrix, and *independent* of the right-hand side. This result implies that when $\binom{W}{A}$ has

a constant number of rows, and the entries in this matrix are polynomially bounded in the input length, the fiber oracle can be implemented in polynomial time.

- **N-fold systems.** It is well-known that if A is an N-fold matrix then the matrix $\binom{W}{A}$ can be transformed to an equivalent N-fold matrix, provided that all entries in W form a set $K \subset \mathbb{Z}$ of fixed size. As was shown by De Loera, Hemmecke, Köppe and Weismantel [DLHOW08], integer programs with an N-fold constraint matrix admit polynomial-time algorithms.

We note that there are several other interesting classes of matrices that admit polynomial algorithms. One obvious example is when $\binom{W}{A}$ is totally unimodular. In such cases, however, one has $\mathcal{R} = Q \cap \mathbb{Z}^d$, so the Problem (5.1) can in these cases be solved with two oracle calls (see Remark 5.0.3).

Bibliography

[AH10]   I. Aliev and M. Henk. Feasibility of integer knapsacks. *SIAM J. Optim.*, 20(6):2978–2993, 2010.

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

[AOW14]   D Adjiashvili, T. Oertel, and R. Weismantel. A polyhedral frobenius theorem with applications to integer opti- mization. Manuscript submitted for publication, 2014.

[AW12]   G. Averkov and R. Weismantel. Transversal numbers over subsets of linear spaces. *Adv. Geom.*, 12(1):19–28, 2012.

[Bal92]   K. Ball. Ellipsoids of maximal volume in convex bodies. *Geometriae Dedicata*, 41(2):241–250, 1992.

[Bar94]   A.I. Barvinok. A polynomial time algorithm for counting integral points in polyhedra when the dimension is fixed. *Mathematics of Operations Research*, 19(4):769–779, 1994.

[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(2):186–204, 2008.

[BLPS99]   W. Banaszczyk, A.E. Litvak, A. Pajor, and S.J. Szarek. The flatness theorem for nonsymmetric convex bodies via the local theory of Banach spaces. *Mathematics of Operations Research*, 24(3):728–750, 1999.

[BOWW13]   M. Baes, T. Oertel, C. Wagner, and R. Weismantel. Mirror-descent methods in mixed-integer convex optimization. In *Facets of combinatorial optimization*, pages 101–131. Springer, Heidelberg, 2013.

[Bra42]   A. Brauer. On a problem of partitions. *American Journal of Mathematics*, 64(1):299–312, 1942.

[BS62]   A. Brauer and J.E. Shockley. On a problem of frobenius. *J. reine angew. Math*, 211:215–220, 1962.

[BV04a]   D. Bertsimas and S. Vempala. Solving convex programs by random walks. *J. ACM*, 51(4):540–556, July 2004.

[BV04b]   S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.

[CGT00]   A.R. Conn, N.I.M. Gould, and P.L. Toint. *Trust-region methods*. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[CSY87]   R. Cole, M. Sharir, and C.-K. Yap. On k-hulls and related problems. *SIAM J. Comput.*, 16(1):61–77, 1987.

[Dad12]   D.N. Dadush. *Integer programming, lattice algorithms, and deterministic volume estimation*. ProQuest LLC, Ann Arbor, MI, 2012. Thesis (Ph.D.)–Georgia Institute of Technology.

[DG86]   M. Duran and I. Grossmann. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming*, 36(3):307–339, 1986.

[DLHKW06]   J.A. De Loera, R. Hemmecke, M. Köppe, and R. Weismantel. Integer polynomial optimization in fixed dimension. *Mathematics of Operations Research*, 31:147–153, 2006.

[DLHOW08]   J.A. De Loera, R. Hemmecke, S. Onn, and R. Weismantel. N-fold integer programming. *Discrete Optimization*, 5(2):231–241, 2008.

[DLRS10]   J.A. De Loera, J. Rambau, and F. Santos. *Triangulations*, volume 25 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, 2010. Structures for Algorithms and Applications.

[Doi73]   J.-P. Doignon. Convexity in cristallographical lattices. *J. Geometry*, 3:71–85, 1973.

[DPHWZ14]   A. Del Pia, R. Hildebrand, R. Weismantel, and K. Zemmer. Minimizing cubic and homogeneous polynomials over integers in the plane. Manuscript, 2014.

[DPW14]   A. Del Pia and R. Weismantel. Integer quadratic programming in the plane. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 840–846, 2014.

[EG72]   P. Erdös and R.L. Graham. On a linear diphantine problem of Frobenius. *Acta Arithmetica*, 21:399–408, 1972.

[Egg53]   H.G. Eggleston. Some properties of triangles as extremal convex curves. *J. London Math. Soc.*, 28:32–36, 1953.

[Egg57]   H.G. Eggleston. *Problems in Euclidean space: Applications of convexity.* International Series of Monographs on Pure and Applied Mathematics, Vol. V. Pergamon Press, New York-London-Paris-Los Angeles, 1957.

[Egg58]   H.G. Eggleston. *Convexity.* Cambridge Tracts in Mathematics and Mathematical Physics, No. 47. Cambridge University Press, New York, 1958.

[Eis03]   F. Eisenbrand. Fast integer programming in fixed dimension. In *Algorithms—ESA 2003*, volume 2832 of *Lecture Notes in Comput. Sci.*, pages 196–207. Springer, Berlin, 2003.

[EL05]   F. Eisenbrand and S. Laue. A linear algorithm for integer programming in the plane. *Mathematical Programming*, 102(2, Series A):249–259, 2005.

[EVW14]   F. Eisenbrand, S. Vempala, and R. Weismantel. Revisiting dynamic programming for integer optimization. Private communication, 2014.

[FL94]   R. Fletcher and S. Leyffer. Solving mixed integer nonlinear programs by outer approximation. *Mathematical Programming*, 66(3, Series A):327–349, 1994.

[For97]   S. Fortune. Voronoi diagrams and Delaunay triangulations. In *Handbook of Discrete and Computational Geometry*, CRC Press Ser. Discrete Math. Appl., pages 377–388. CRC, Boca Raton, FL, 1997.

[GJ79]   M.R. Garey and D.S. Johnson. *Computers and intractability.* W. H. Freeman and Co., San Francisco, Calif., 1979. A guide to the theory of NP-completeness, A Series of Books in the Mathematical Sciences.

[GL87]   P. M. Gruber and C. G. Lekkerkerker. *Geometry of numbers*, volume 37 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition, 1987.

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

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

[Grü60]   B. Grünbaum. Partitions of mass-distributions and of convex bodies by hyperplanes. *Pacific J. Math.*, 10:1257–1261, 1960.

[Gru07]   P.M. Gruber. *Convex and Discrete Geometry*, volume 336 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2007.

[Hei05]   S. Heinz. Complexity of integer quasiconvex polynomial optimization. *Journal of Complexity*, 21(4):543–556, 2005.

[HK12]   R. Hildebrand and M. Köppe. A new Lenstra-type algorithm for quasiconvex polynomial integer minimization with complexity $2^{O(n log n)}$. Manuscript, http://arxiv.org/abs/1006.4661, 2012.

[HS90]   D.S. Hochbaum and J.G. Shanthikumar. Convex separable optimization is not much harder than linear optimization. *J. ACM*, 37(4):843–862, 1990.

[HUL93]   J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms II*, volume 306 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

[Joh48]   F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, pages 187–204. Interscience Publishers, Inc., New York, 1948.

[Kan87]   R. Kannan. Minkowski's convex body theorem and integer programming. *Math. Oper. Res.*, 12(3):415–440, 1987.

[Kan89]   R. Kannan. Solution of the frobenius problem and its general-
            ization. Manuscript, 1989.

[Kan92]   R. Kannan. Lattice translates of a polytope and the Frobenius
            problem. *Combinatorica*, 12(2):161–177, 1992.

[Kar39]   W. Karush. Minima of functions of several variables with in-
            equalities as side conditions. Dissertation, Department of Math-
            ematics, University of Chicago, 1939.

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

[Kli14]   K. Klimkewitz. Integer convex optimization in fixed dimension
            using integer approximations to centerpoints. Master's thesis,
            ETH Zurich, 2014.

[KP00]   L. Khachiyan and L. Porkolab. Integer optimization on con-
            vex semialgebraic sets. *Discrete & Computational Geometry*,
            23(2):207–224, 2000.

[KT51]   H.W. Kuhn and A.W. Tucker. Nonlinear programming. In
            *Proceedings of the Second Berkeley Symposium on Mathematical
            Statistics and Probability, 1950*, pages 481–492. University of
            California Press, Berkeley and Los Angeles, 1951.

[KZ73]   A. Korkine and G. Zolotareff. Sur les formes quadratiques.
            Math. Ann., 6(3):366–389, 1873.

[Lek69]   C. G. Lekkerkerker. *Geometry of numbers.* Bibliotheca Math-
            ematica, Vol. VIII. Wolters-Noordhoff Publishing, Groningen;
            North-Holland Publishing Co., Amsterdam-London, 1969.

[Len83]   H.W. Lenstra, Jr. Integer programming with a fixed number of
            variables. *Mathematics of Operations Research*, 8(4):538–548,
            1983.

[LLL82]   A.K. Lenstra, H.W. Lenstra, Jr., and L. Lovász. Factoring
            polynomials with rational coefficients. *Mathematische Annalen*,
            261(4):515–534, 1982.

[LLS90]   J.C. Lagarias, H.W. Lenstra, Jr., and C.-P. Schnorr. Korkin-
            Zolotarev bases and successive minima of a lattice and its re-
            ciprocal lattice. *Combinatorica*, 10(4):333–348, 1990.

[Nem94]   A. Nemirovski.   Efficient   methods   in   con-
          vex   programming.   Available   online   at
          http://www2.isye.gatech.edu/∼nemirovs/Lect_EMCO.pdf,
          1994.

[Nes04]   Y. Nesterov. *Introductory Lectures on Convex Optimization*,
          volume 87 of *Applied Optimization*. Kluwer Academic Publish-
          ers, Boston, 2004.

[Neu45]   B.H. Neumann. On an invariant of plane regions and mass
          distributions. *J. London Math. Soc.*, 20:226–237, 1945.

[NN94]   Y. Nesterov and A. Nemirovskii. *Interior-point polynomial al-
          gorithms in convex programming*, volume 13 of *SIAM Studies in
          Applied Mathematics*. Society for Industrial and Applied Math-
          ematics (SIAM), Philadelphia, PA, 1994.

[NY79]   A. Nemirovski and D. Yudin.   Efficient methods of solv-
          ing convex-programming problems of high dimensionality.
          Ekonomika i matem. metody (in Russian), 1979.

[NY83]   A. Nemirovski and D. Yudin. *Problem Complexity and Method
          Efficiency in Optimization*. John Wiley, 1983.

[OWW14]   Timm Oertel, Christian Wagner, and Robert Weismantel. Inte-
          ger convex minimization by mixed integer linear optimization.
          Operations Research Letters, 42(6‚Äì7):424 – 428, 2014.

[RA96]   J.L. Ramírez-Alfonsín. Complexity of the Frobenius problem.
          Combinatorica, 16(1):143–147, 1996.

[Roc81]   R. Rockafellar. *The Theory of Subgradients and its Applications
          to Problems of Optimization*, volume 1 of *R & E*. Heldermann
          Verlag, Berlin, 1981.

[Sch86]   A. Schrijver. *Theory of linear and integer programming*. Wiley-
          Interscience Series in Discrete Mathematics. John Wiley & Sons
          Ltd., Chichester, 1986. A Wiley-Interscience Publication.

[Seb90]   A. Sebö. Hilbert bases, caratheodory's theorem and combina-
          torial optimization. In *Proceedings of the 1st Integer Program-
          ming and Combinatorial Optimization Conference*, pages 431–
          455. University of Waterloo Press, 1990.

[VG90]   J. Viswanathan and I. Grossmann. A combined penalty func-
           tion and outer-approximation method for MINLP optimization.
           Computers & Chemical Engineering, 14(7):769–782, 1990.

Short Curriculum Vitae

Timm Oertel

Born on May 11, 1985

citizen of Germany

Education

2010-2014	**Ph.D., Mathematics**
	ETH Zürich
2005-2010	**Dipl., Mathematics**
	Otto-von-Guericke Universität, Magdeburg
1998-2004	**High school**
	Hegel Gymnasium, Magdeburg

Work Experience

2010-2014	**Teaching Assistant**
	ETH Zürich
2008-2010	**Student Research Assistant**
	Otto-von-Guericke Universität, Magdeburg

Academic Visists

2008	**Internship**
	Zuse-Institut, Berlin