

Hesse's Redemption: Efficient Convex Polynomial Programming

Lucas Slot

lucas.slot@inf.ethz.ch

ETH Zurich

David Steurer

david.steurer@inf.ethz.ch

ETH Zurich

Manuel Wiedmer

manuel.wiedmer@inf.ethz.ch

ETH Zurich

November 5, 2025

Abstract

Efficient algorithms for convex optimization, such as the ellipsoid method, require an *a priori* bound on the radius of a ball around the origin guaranteed to contain an optimal solution if one exists. For linear and convex quadratic programming, such solution bounds follow from classical characterizations of optimal solutions by systems of *linear* equations. For other programs, e.g., semidefinite ones, examples due to Khachiyan show that optimal solutions may require huge coefficients with an exponential number of bits, even if we allow approximations. Correspondingly, semidefinite programming is not even known to be in NP.

The unconstrained minimization of convex polynomials of degree four and higher has remained a fundamental open problem between these two extremes: its optimal solutions do not admit a linear characterization and, at the same time, Khachiyan-type examples do not apply. We resolve this problem by developing new techniques to prove solution bounds when no linear characterizations are available. Even for programs minimizing a convex polynomial (of arbitrary degree) over a polyhedron, we prove that the existence of an optimal solution implies that an approximately optimal one with polynomial bit length also exists. These solution bounds, combined with the ellipsoid method, yield the first polynomial-time algorithm for convex polynomial programming, settling a question posed by Nesterov (Math. Program., 2019). Before, no polynomial-time algorithm was known even for unconstrained minimization of a convex polynomial of degree four.

Our results rely on a structural decomposition of any convex polynomial into a sum of a linear function and a polynomial on a linear subspace that admits a strongly convex lower bound, where the logarithm of the strong convexity parameter is polynomially bounded in the input size. A key component of our proof is a strong local-to-global property for convex polynomials: if at every point *some* directional second derivative vanishes, then a *single* directional second derivative must vanish everywhere. While Hesse erroneously claimed that this property holds for general polynomials (J. Reine Angew. Math., 1851), we show that it holds for convex ones.

Contents

1	Introduction	1
1.1	Minimizers of (convex) polynomials	1
1.2	Main contributions	2
1.3	Complexity landscape of polynomial programming	3
1.4	Related work	5
2	Technical overview	7
2.1	Structure theorem for convex polynomials	7
2.2	Applications in convex polynomial programming	11
3	Preliminaries	13
3.1	Notations	13
3.2	The ellipsoid method	13
4	Structure theorem for convex polynomials	14
4.1	A strongly convex lower bound: Proof of Proposition 4.1	15
4.2	A Hessian decomposition theorem: Proof of Proposition 4.2	16
5	Applications in convex polynomial programming	18
5.1	Deciding unboundedness: Proof of Lemma 5.1	20
5.2	Bounding the norm of minimizers in a subspace: Proof of Lemma 5.2	22
5.3	Finding a minimizer of small norm: Proof of Lemma 5.3	25
	Acknowledgments	26
A	Linear algebra in the bit model	26
B	A polynomial with identically vanishing hessian	27
C	Witnesses for polynomial programs	28
D	Ellipsoid method	29
D.1	Additional preliminaries on the ellipsoid method	30
D.2	Proof of Proposition 3.2	31

1 Introduction

A central result in convex programming is the algorithmic equivalence of optimization and separation, emerging from a body of work around the ellipsoid method [YN76, Kha79, GLS81, KP80, PR81, GLS93]. However, this equivalence crucially requires *effective solution bounds*, even when we allow approximations: if optimal solutions exist, then (approximately) optimal solutions must exist such that the logarithm of their norm is polynomially bounded in the input size. The ellipsoid method requires such solution bounds when determining the radius of its initial ellipsoid and the running time depends polynomially on the logarithm of this radius.¹ Effective solution bounds are necessary in order to be able to output (approximately) optimal solutions in (worst-case) polynomial time. Without such bounds, the corresponding decision problem is typically not known to be in **NP**, as its natural witnesses may have size superpolynomial in the input size.

Linear programs are prominent examples of optimization problems that enjoy effective solution bounds. Their optimal solutions arise as solutions to systems of (consistent and independent) linear equations [Dan63]. By Cramer’s rule, such solutions have polynomial bit length [Edm67], and thus the logarithm of their norms is polynomially bounded. This effective solution bound plays a crucial role for the seminal result that linear programming has polynomial time algorithms [Kha79]. For quadratic objective functions, optimal solutions continue to have a linear characterization (because the gradient of a quadratic is an affine linear map). Correspondingly, convex quadratic programs, where one minimizes a convex polynomial of degree two over a polyhedron, have effective solution bounds and polynomial-time algorithms like linear programs [KTK80]. On the other hand, for other natural generalizations of linear programs, like semidefinite programs, the situation is drastically different and no effective solution bound is possible: There are feasible semidefinite programs such that all feasible and even approximately feasible solutions have huge coefficients and the logarithm of their norms is exponential in the input size.² Therefore, the ellipsoid method does not yield a polynomial-time algorithm for semidefinite programming (even approximately), even though the corresponding separation problem has polynomial-time algorithms. The lack of effective solution bounds is also the reason why semidefinite programming is not known to be in **NP** [PK97, Ram97].

The unconstrained minimization of convex polynomials of degree four and higher is a fundamental optimization problem whose optimal solutions have no linear characterization (because critical points are defined by equations of degree three and higher). At the same time, previous “bad” examples (like for semidefinite programs) do not apply. In this paper, we develop new techniques to show effective solution bounds for this problem. These techniques extend to *convex polynomial programming*, where one minimizes a convex polynomial (of arbitrary degree) over a polyhedron. Our solution bounds, combined with the ellipsoid method, yield the first efficient algorithm for solving convex polynomial programs at degree four and above (to an arbitrary degree of accuracy). Previously, no efficient algorithm for computing the approximate minimum of a convex polynomial of degree four was known (even without constraints).

1.1 Minimizers of (convex) polynomials

If p is a polynomial of degree 2, its critical points are solutions to the *linear* system $\nabla p(x) = 0$. This permits us to control the norm of its minimizers, similarly to the case of linear programs.

¹This radius is sometimes called “big R ”. The ellipsoid method has another parameter, often referred to as “little r ”, relating to the existence of a small ball inside the feasible region. However, the need for “little r ” can usually be avoided by accepting approximate solutions that may violate optimality and the feasibility constraints by a tiny amount.

²These examples are commonly attributed to Khachiyan; see [Ram97, O’D17, PT24].

For polynomials of larger degree, this approach does not work. In fact, already for degree 4, *nonconvex* polynomials may not admit an effectively bounded minimizer at all. Consider, e.g.,

$$p(x) := (x_1^2 - x_2)^2 + \dots + (x_{n-1}^2 - x_n)^2 + (x_n - 2)^2. \quad (1)$$

The minimum of p over \mathbb{R}^n is 0, and it is attained if and only if $x_n = 2$, and $x_{i-1} = x_i^2$ for all $1 \leq i < n$. Thus, the unique minimizer x^* of p satisfies $x_1^* = 2^{2^{n-1}}$, meaning its norm is *doubly-exponential* in n .

The upshot is that, in order to establish effective solution bounds for *convex* polynomials, it is imperative to exploit their convexity. On the other hand, it is NP-hard to decide whether a given polynomial is convex [AOPT11]. This means we cannot rely on an algorithmic characterization of convexity for polynomials. Furthermore, convex polynomials do not (obviously) satisfy the type of quantitative properties typically used in the analysis of optimization algorithms. For example, even strictly convex polynomials are not *strongly convex* in general (consider $f(x) = x^4$).

The Hessian determinant. Instead, we will work with a purely *qualitative* characterization: f is convex if and only if its matrix $\nabla^2 f(x)$ of second derivatives is positive semidefinite for all $x \in \mathbb{R}^n$. To translate from this qualitative description to the quantitative bounds we are after, the key is to consider the *Hessian determinant* $h_f(x) := \det(\nabla^2 f(x))$ of f , which describes its (local) curvature. Note that, by convexity, we have $h_f(x) \geq 0$ for all $x \in \mathbb{R}^n$. The Hessian determinant captures the local behavior of f around a point $a \in \mathbb{R}^n$ via the following dichotomy:

- If $h_f(a) > 0$, then f is *locally μ -strongly convex* around a , meaning there are $\mu, r > 0$ such that

$$f(x) \geq f(a) + \langle \nabla f(a), x - a \rangle + \mu \cdot \|x - a\|^2 \quad \forall x \in B(a, r).$$

- If $h_f(a) = 0$, then f has a *local direction of linearity* at a , meaning there is a $v \in \mathbb{R}^n$ such that

$$\frac{\partial^2}{\partial t^2} f(a + tv) = 0.$$

Local to global. The core of our analysis is to show that both sides of the dichotomy above exhibit *local-to-global* behavior. First, we show that if the Hessian determinant is positive at even a single point $a \in \mathbb{R}^n$, this immediately implies a *global μ -strongly convex lower bound* on f . Moreover, the parameter μ can be controlled in terms of the binary encoding length of f independently of a . Second, we show that if f has a *local* direction of linearity *everywhere*, then it has a *global* direction of linearity. That is, if $h_f(x) = 0$ for all $x \in \mathbb{R}^n$, there is $v \in \mathbb{R}^n$ (independent of x) such that

$$\frac{\partial^2}{\partial t^2} f(x + tv) = 0 \quad (\forall x \in \mathbb{R}^n),$$

which implies that f is linear along v . Projecting onto the complement of this direction allows us to reduce to a problem in fewer variables. This fact has an interesting connection to work of O. Hesse himself [Hes51], who (mistakenly) claimed it holds for general (nonconvex) polynomials. Together, these observations will allow us to write any convex polynomial as the sum of a linear function and a polynomial (of fewer variables) that admits a strongly convex lower bound (Theorem 1.3).

1.2 Main contributions

As our main contribution, we show a singly-exponential upper bound on the norm of a global minimizer of a convex polynomial f on a convex polyhedron $P = \{Ax \leq b\}$. To state our precise

results, we must first specify our input model. We write $\text{bl}(q)$ for the bit length of $q \in \mathbb{Q}$. The bit length of a vector or matrix is the sum of the bit lengths of its entries. We write $\text{enc}(P) := \text{bl}(A) + \text{bl}(b)$ for the (total) encoding length of P . We encode a polynomial $f(x) = \sum_{\alpha} f_{\alpha} x^{\alpha}$ using a *binary* representation for its (nonzero) coefficients $f_{\alpha} \in \mathbb{Q}$, and a *unary* representation for the multi-indices $\alpha \in \mathbb{N}^n$. Thus, its encoding length is $\text{enc}(f) = \Theta(n + \deg(f) + \sum_{\alpha: f_{\alpha} \neq 0} \text{bl}(f_{\alpha}))$.³

Theorem 1.1. *Let $f \in \mathbb{Q}[x]$ be a convex polynomial. Let P be a (nonempty) convex polyhedron. Then, either f is unbounded from below on P , or it attains its minimum on P at a point x^* of norm*

$$\log \|x^*\| \leq \text{poly}(\text{enc}(f), \text{enc}(P)).$$

As a consequence, we can minimize f over P in polynomial time using the ellipsoid method.

Corollary 1.2. *Let $f \in \mathbb{Q}[x]$ be a convex polynomial. Let P be a (nonempty) convex polyhedron. Let $\varepsilon > 0$. We can decide in time $\text{poly}(\text{enc}(f), \text{enc}(P), \log(1/\varepsilon))$ whether f is unbounded from below on P , and, if not, output a point $\tilde{x} \in P$ with $f(\tilde{x}) \leq \min_{x \in P} f(x) + \varepsilon$.*

The main technical tool in our proof of [Theorem 1.1](#) is a structure theorem for convex polynomials, which we believe to be of independent interest. It shows that any convex polynomial can be written as the sum of a linear function and a polynomial (of fewer variables) that has a μ -strongly convex *lower bound*, where μ is at least inversely exponential in $\text{enc}(f)$.

Theorem 1.3. *Let $f \in \mathbb{Q}[x]$ be a convex polynomial. Then, there is a subspace $\mathcal{U} \subseteq \mathbb{R}^n$, and a rational vector $w \in \mathcal{U}^\perp$ (possibly $w = 0$), such that, writing $x_{\mathcal{U}}$ for the projection of x onto \mathcal{U} ,*

$$f(x) = f(x_{\mathcal{U}}) - \langle w, x \rangle,$$

where $f(x_{\mathcal{U}})$ has a μ -strongly convex lower bound on \mathcal{U} . That is, there is a matrix $U \in \mathbb{Q}^{k \times n}$ whose rows are orthogonal and span \mathcal{U} , and a k -variate, μ -strongly convex, quadratic polynomial $q \in \mathbb{Q}[y_1, \dots, y_k]$ such that $f(x_{\mathcal{U}}) \geq q(Ux)$ for all $x \in \mathbb{R}^n$. Moreover, the matrix U and the vector w can be computed in polynomial time in $\text{enc}(f)$, the lower bound q satisfies $\text{enc}(q) \leq \text{poly}(\text{enc}(f))$, and $\mu \geq 2^{-\text{poly}(\text{enc}(f))}$.

[Theorem 1.3](#) allows us to give a quantitative description of the minimizers of f on P . For example, in the unconstrained case (where $P = \mathbb{R}^n$), we have $f_{\min} = -\infty$ if and only if $w \neq 0$. If $w = 0$, the μ -strongly convex lower bound on f straightforwardly shows that its global minimizer is attained at a point of norm at most exponential in $\text{poly}(\text{enc}(f))$. For general P , the situation is more complicated. There, we show that f is unbounded from below if and only if a certain linear system of inequalities (involving U and w) has a solution. If not, Farkas' lemma gives us a witness for this fact, which we use to show that any minimizer of f on P has small norm when projected onto the subspace $\mathcal{U} \oplus \text{span}\{w\}$. Finally, by a lifting argument, we show that there exists a minimizer of small norm in the full space.

1.3 Complexity landscape of polynomial programming

A *polynomial program* asks to minimize a rational polynomial f over a convex polyhedron P described by rational linear inequalities $Ax \leq b$; that is, to output a binary representation of

$$x^* = \operatorname{argmin}_{x \in \mathbb{R}^n} \{f(x) : x \in P\}. \quad (\text{POP})$$

³This encoding is more efficient than the “naive” encoding of f as a (dense) vector of coefficients in the monomial basis (whose size scales exponentially in $\deg(f)$). Our main results [Theorem 1.1](#) and [Corollary 1.2](#) are novel even in this dense model; in fact already for fixed degree $\deg(f) = 4$.

In this section, we discuss the complexity of this problem (in the Turing model) in the convex and general setting, and for different values of $d = \deg(f)$. Our discussion is summarized in [Table 1](#). For ease of exposition, we assume that x^* is well-defined, i.e., that the minimum above is attained and unique. We are interested in algorithms whose runtime is bounded polynomially in n and the number of bits required to specify f and P . Consider the associated *decision problem*:

$$\text{Given } (f, P), \text{ output } \begin{cases} \text{YES} & \text{if } \exists x \in P : f(x) \leq 0, \\ \text{NO} & \text{if } \forall x \in P : f(x) > 0. \end{cases} \quad (\text{D})$$

Compact witnesses. A *witness* for [\(D\)](#) is a point $w \in P$ satisfying $f(w) \leq 0$. Clearly, the answer to [\(D\)](#) is YES if and only if there exists a witness. Furthermore, if there exists a witness $w \in \mathbb{Q}^n$ of polynomial bit length (in the size of the input), then this YES-answer can be verified efficiently by evaluating $f(w)$ and the inequalities $Aw \leq b$. We call such a witness *compact*. Thus, the existence of a compact witness for every YES-instance of [\(D\)](#) implies that the problem lies in **NP**. We have already seen that this is the case for linear programs ($d = 1$) and for quadratic programs ($d = 2$) [[Vav90](#)]. In both cases, the “canonical” witness $w = x^*$ happens to be compact.⁴ As a direct result, linear programming and *convex* quadratic programming lie in **P**, as x^* can be identified efficiently by the ellipsoid method [[Kha79](#), [KTK80](#)]. On the other hand, (nonconvex) quadratic programming is **NP**-complete (while x^* is compact, it might be hard to find [[Sah74](#)]).

For $d \geq 4$, the minimizer x^* is typically not compact. In fact, there might not exist any compact witness for [\(D\)](#). For example, the global minimum of $f(x) = (x^2 - 2)^2$ over $P = \mathbb{R}$ is 0, but the only witnesses of this fact are $\pm\sqrt{2} \notin \mathbb{Q}$. Even if we assume f is convex, there are examples of degree 4 where $x^* \notin \mathbb{Q}^n$, and examples of degree 6 where no rational witnesses exist at all (see [Section C](#)).

Effective solution bounds. If the canonical witness x^* is not rational (compact), this means that the polynomial program [\(POP\)](#) cannot be solved (efficiently), as it is not possibly to output a binary representation of x^* (in polynomial time). This appears a somewhat artificial limitation. Instead, one could ask to output an *approximate* minimizer $\tilde{x} \in P$, satisfying $f(\tilde{x}) \leq f(x^*) + \varepsilon$, where $\varepsilon > 0$ is part of the input. This corresponds to the following *promise problem*:

$$\text{Given } (f, P, \varepsilon), \text{ output } \begin{cases} \text{YES} & \text{if } \exists x \in P : f(x) \leq 0, \\ \text{NO} & \text{if } \forall x \in P : f(x) > \varepsilon. \end{cases} \quad (\text{P})$$

A witness for [\(P\)](#) is a point $\tilde{w} \in P$ with $f(\tilde{w}) \leq \varepsilon$. Existence of such a \tilde{w} is equivalent to YES being a valid answer to [\(P\)](#). As before, the existence of a compact witness for each YES-instance implies that [\(P\)](#) is in **NP**. The upshot is that any witness w for [\(D\)](#) whose *norm* is exponentially bounded in (a polynomial of) the size of the input can be rounded to a *compact* witness \tilde{w} for [\(P\)](#). Thus, to show that [\(P\)](#) is in **NP**, it suffices to show that x^* has exponentially bounded norm.

Notably, polynomial programs generally do not admit an exponentially bounded solution, already when $d = 4$. Indeed, the quartic polynomial defined in [\(1\)](#) has a unique global minimizer of norm $2^{2^{n-1}}$ on $P = \mathbb{R}^n$. Our main contribution ([Theorem 1.1](#)) is to show that *convex* polynomials (of any degree) *do* admit an exponentially bounded minimizer. Contrary to the degree-2 case, the existence of compact (approximate) witnesses for $d \geq 4$ thus depends on convexity. Using the ellipsoid method, this shows that (approximate) convex polynomial programming is in **P** ([Corollary 1.2](#)).

⁴For linear programs, x^* is a vertex of P , and thus the solution to a subsystem of $Ax = b$ [[Dan63](#), [Edm67](#)]. For unconstrained quadratic programs ($P = \mathbb{R}^n$), x^* is a critical point of f , and thus a solution to the linear system $\nabla f(x) = 0$. In the general setting, it turns out x^* is still the solution to a linear system, involving KKT(-like) conditions [[KTK80](#), [Vav90](#)].

objective	compact wit. (D)	compact wit. (P)	complexity (D)	complexity (P)
linear	yes ^a [Dan63, Edm67]	—	P [Kha79]	—
convex quadratic	yes ^a [KTK80]	—	P [KTK80]	—
quadratic	yes ^a [Vav90]	—	—	NP-hard [Sah74]
convex quartic	unknown ^b	yes (Thm. 1.1)	unknown	P (Cor. 1.2)
quartic	—	no (Eq. (1))	—	NP-hard ^c
convex $d \geq 6$	no (App. C)	yes (Thm. 1.1)	unknown	P (Cor. 1.2)

Table 1: Complexity and availability of compact witnesses for problems (D) and (P), corresponding to exact and approximate polynomial programming, respectively. Omitted entries (“—”) follow directly from adjacent ones (e.g., a compact witness for (D) is also a compact witness for (P)).

^athe canonical witness $w = x^*$ is compact.

^bthe canonical witness $w = x^*$ is not compact in general, but a compact witness might exist. See [Section C](#).

^cthe problem is NP-hard even in the unconstrained setting ($P = \mathbb{R}^n$) [Lau08].

Local polynomial optimization. Even *local* optimization tasks involving polynomials are typically hard: Determining whether a polynomial of degree 4 has a local minimum or a critical point is hard; it is even hard to decide if a given (critical) point is a local minimum [AZ22a]. Determining an (approximate) local minimizer of a quadratic on a polytope is also hard [AZ22b].

Real models of computation. In real models of computation (e.g., in the *Blum-Schub-Smale* model [BSS89] or in the *real RAM* model), the decision problem (D) is always in NP (as evaluating $f(x)$ and $Ax \leq b$ at any point $x \in \mathbb{R}^n$ is easy there). However, in these models, it is not even known whether *linear* programs can be solved in polynomial time [Sma00] (as the runtime of an algorithm is no longer allowed to depend on the size of the entries of A and b).

1.4 Related work

Higher-order Newton methods. The global minimization of convex polynomials comes up naturally in Newton’s method for optimization. This method attempts to find the minimizer of a smooth function $g : \mathbb{R}^n \rightarrow \mathbb{R}$ by successively approximating it by its second-order Taylor expansion around an iterate $x^k \in \mathbb{R}^n$, setting x^{k+1} to be a global minimizer of this approximation. If the Hessian matrix $\nabla^2 g(x^k) \succeq 0$ is positive semidefinite at x^k , then the Taylor expansion around x^k is a convex quadratic, and its critical points are global minimizers. Newton’s method converges quadratically within a small basin around local minimizers where the Hessian of g is positive definite and locally Lipschitz. It has been extensively studied whether faster convergence can be achieved by considering Taylor expansions of g of higher order, see e.g., [Nes19, ACZ24]. A difficulty is that the degree- d expansion of g around x^k is typically nonconvex when $d > 2$ (even if $\nabla^2 g(x^k) \succeq 0$), and it is not clear how to compute its global minimizer (if it even exists). To avoid this problem, the authors of [Nes19] and [ACZ24] add a regularizing term that ensures convexity. Still, efficient optimization of the resulting convex polynomial then relies on additional properties inherited from the specifics of the regularization scheme: in [Nes19], it satisfies a *relative smoothness condition*; in [ACZ24] it is *sum-of-squares-convex* (see below). This raises the question whether such additional properties are necessary.⁵ Our work shows that convexity alone is enough.

⁵In fact, this question is raised already by the author in [Nes19], who states that they were unable to locate any method in the literature aimed at solving the general problem of minimizing a convex, multivariate polynomial.

Complexity of SDP. Semidefinite programming (SDP) is a natural extension of linear programming with important applications in (combinatorial) optimization. This includes the Goemans-Williamson approximation for max-cut [GW95], and an efficient algorithm to compute the independence number of a perfect graph [GLS81]. Under additional assumptions, SDPs can be solved in polynomial time using the ellipsoid method [GLS81] or interior-point methods [dKV16]. However, the computational complexity of general SDP is not known. The reason is that SDPs do not admit effective solution bounds [PK97, PT24]; consider the following example due to Khachiyan:

$$\exists x \in \mathbb{R}^n : \begin{bmatrix} x_1 & x_2 \\ x_2 & 1 \end{bmatrix} \succeq 0, \dots, \begin{bmatrix} x_{n-1} & x_n \\ x_n & 1 \end{bmatrix} \succeq 0, x_n \geq 2? \quad (2)$$

Any solution to this problem must satisfy $x_1 \geq 2^{2^{n-1}}$, meaning it has doubly-exponential norm. The best available results are that, in the Turing model, SDP is either in $\mathbf{NP} \cap \mathbf{coNP}$ or not in $\mathbf{NP} \cup \mathbf{coNP}$ [Ram97]; in the real (Blum-Shub-Smale) model, it is in $\mathbf{NP} \cap \mathbf{coNP}$ [Ram97].

Effective solution bounds for sum-of-squares relaxations. There is a well-developed literature on convex relaxations that produce tractable bounds on the (global) minimum p_{\min} of a polynomial p . These are based on the observation that $p_{\min} = \max_{\lambda \in \mathbb{R}} \{\lambda : p(x) - \lambda \geq 0 \forall x \in \mathbb{R}^n\}$, which transforms a (nonconvex) polynomial optimization problem into a (convex) problem over the cone \mathcal{P} of nonnegative polynomials. Testing membership in \mathcal{P} is \mathbf{NP} -hard, but one may obtain a tractable *lower bound* on p_{\min} by optimizing over a subcone of \mathcal{P} that admits an efficient separation oracle. The most famous example is the cone $\Sigma \subseteq \mathcal{P}$ of *sums of squares* (*sos*), i.e., polynomials of the form $\sigma(x) = s_1(x)^2 + \dots + s_\ell(x)^2$, $s_i \in \mathbb{R}[x]$. It is possible to optimize over Σ via a semidefinite program. This approach can be generalized to constrained problems, leading to a sequence of lower bounds called the *sum-of-squares hierarchy* [Par00, Las01]. Algorithms based on the sos hierarchy have had a significant impact in theoretical computer science [BS14]. However, its precise computational complexity is not known: Pathological examples similar to (2) can be realized as the sos relaxation of a (constrained) polynomial optimization problem [O'D17]. Recent work shows that polynomial-time computability of the sos hierarchy can be guaranteed under additional assumptions [RW17, GPS23, BMPV25, BMV25]. Essentially, these works show that certain special classes of semidefinite programs admit effectively bounded solutions.

Sum-of-squares-convexity. The interplay between convex polynomials and sums of squares has also been studied. A polynomial $p \in \mathbb{R}[x]$ is called *sum-of-squares-convex* if the $2n$ -variate polynomial $\langle y, \nabla^2 p(x) \cdot y \rangle$ is a sum of squares of polynomials in $\mathbb{R}[x, y]$. This implies that $\nabla^2 p(x) \succeq 0$ for all $x \in \mathbb{R}^n$. Thus, sos-convex polynomials are convex in the regular sense, and their convexity has an algebraic proof. (But, not all convex polynomials are sos-convex [AP13].) Contrary to regular convexity (which is \mathbf{NP} -hard to detect [AOPT11]), sos-convexity of a polynomial can be determined by testing feasibility of a semidefinite program [ACZ24]. Moreover, the (global) minimum of an sos-convex polynomial can be determined by solving an SDP. Indeed, the first level of the sum-of-squares hierarchy is exact for sos-convex polynomials (even under sos-convex constraints) [Las08]. In light of the discussion on sum-of-squares relaxations above, we remark that (to the best of the authors' knowledge) it has never been examined whether these semidefinite programs can be solved in polynomial time.

2 Technical overview

At a high level, we are concerned with understanding the quantitative behavior of the minimizers of a convex polynomial $f \in \mathbb{Q}[x]$ on a convex polyhedron $P = \{Ax \leq b\} \subseteq \mathbb{R}^n$. In particular, we will show that f is either unbounded from below on P , in which case this can be detected efficiently, or f attains its minimum on P at a point $x^* \in P$ whose norm is at most exponential in the encoding lengths of f and P . In the latter case, the ellipsoid method can (approximately) identify this point in polynomial time. Together, this yields our main results [Theorem 1.1](#) and [Corollary 1.2](#).

The main technical tool for establishing this dichotomy is a structure theorem for convex polynomials ([Theorem 1.3](#)), which allows us to write f as the sum of a linear function and a polynomial (of fewer variables) which is bounded from below by a μ -strongly convex quadratic. Namely, we show that any convex polynomial f decomposes as

$$f(x) = f(x_{\mathcal{U}}) - \langle w, x \rangle \geq q(Ux) - \langle w, x \rangle \quad (\forall x \in \mathbb{R}^n), \quad (3)$$

where $U \in \mathbb{Q}^{k \times n}$ is a matrix with orthogonal rows spanning a subspace $\mathcal{U} \subseteq \mathbb{R}^n$, $w \in \mathcal{U}^\perp$ is a rational vector (possibly $w = 0$), and $q \in \mathbb{Q}[y]$ is a k -variate, μ -strongly convex quadratic. Moreover, we show that U, w can be computed in polynomial time in $\text{enc}(f)$, and that $\mu \geq 2^{-\text{poly}(\text{enc}(f))}$.

In the unconstrained case (when $P = \mathbb{R}^n$), this immediately gives us the desired norm bound on the minimizer of f . Indeed, we then have $f_{\min} = -\infty$ if and only if $w \neq 0$, in which case $\lim_{\lambda \rightarrow \infty} f(\lambda w) = -\infty$. On the other hand, if $w = 0$, the strongly convex lower bound (3) on f implies that it has a global minimizer of small norm (in terms of μ). For general P , rather than checking whether $w = 0$, we will show that f is unbounded from below on P if and only if a certain linear system of inequalities (involving U and w) is feasible. If not, Farkas' lemma provides an explicit witness of this fact, which we use to show that f attains its minimum on P at a point whose norm is at most singly-exponential in $\text{enc}(f)$ and $\text{enc}(P)$. We explain this in more detail in [Section 2.2](#) below. First, we outline our proof of [Theorem 1.3](#), i.e., of the decomposition (3).

2.1 Structure theorem for convex polynomials

The key to our proof of [Theorem 1.3](#) is to consider the *Hessian determinant* (or just *hessian*) of f , which is given by $h_f(x) := \det(\nabla^2 f(x))$. By convexity, $h_f(x) \geq 0$ for all $x \in \mathbb{R}^n$. Intuitively, the hessian captures whether the graph of f is locally ‘curved’ around a point $x \in \mathbb{R}^n$ ($h_f(x) > 0$), or ‘straight’ in at least one direction ($h_f(x) = 0$), corresponding to a zero eigenvector of $\nabla^2 f(x)$. For our proof, we exhibit two complimentary local-to-global phenomena involving the hessian of f :

1. If $h_f \not\equiv 0$, i.e., if $\nabla^2 f(a) > 0$ at even a single point $a \in \mathbb{R}^n$, then f admits a *global*, μ -strongly convex lower bound of degree two. Moreover, we have $\mu \geq 2^{-\text{poly}(\text{enc}(f))}$ independently of a .
2. If $h_f \equiv 0$, i.e., if $\nabla^2 f(x)$ is singular for all $x \in \mathbb{R}^n$, then f admits a *global* direction of linearity. That is, there is a nonzero $v \in \mathbb{R}^n$ such that the directional derivative $\nabla_v f$ is constant on \mathbb{R}^n . In other words, we have $f(x) = f(x_{\mathcal{V}^\perp}) + c\langle x, v \rangle$ for all $x \in \mathbb{R}^n$, where $c \in \mathbb{R}$ and $\mathcal{V} = \text{span}\{v\}$.

At a high level, these observations allow us to prove [Theorem 1.3](#) as follows: if $h_f \not\equiv 0$ is not identically zero, we are done immediately (setting $\mathcal{U} = \mathbb{R}^n$ and $w = 0$). If $h_f \equiv 0$, we may conclude that f has a direction of linearity. After projecting onto the complement of that direction, we are left with a polynomial in one fewer variable which differs from f only by a linear term. Iterative application of this idea allows us to “eliminate” variables one at a time until the “remaining” polynomial has a nonzero hessian (and thus admits a strongly convex lower bound). For technical reasons, it turns out that it is preferable to eliminate all variables “in one shot”, and we will show

how to do so. In what follows, we explain in more detail how to prove the two observations above separately, and how to combine them to obtain the decomposition (3).

Nonzero hessian implies a strongly convex lower bound. We show that any convex polynomial $f \in \mathbb{Q}[x]$ with $h_f \not\equiv 0$ admits a μ -strongly convex, quadratic lower bound, where μ is at least inverse exponential in the encoding size of f . Our starting point is the following lemma.

Lemma 2.1 ([ACZ24, Lemma 5]). *Let $p \in \mathbb{R}[x]$ be a convex polynomial. For any $a \in \mathbb{R}^n$, we have*

$$p(x) \geq p(a) + \langle \nabla p(a), x - a \rangle + \frac{\langle x - a, \nabla^2 p(a) \cdot (x - a) \rangle}{4 \deg(p)^2} \quad (\forall x \in \mathbb{R}^n). \quad (4)$$

This lemma can be seen as an extension of the fact that a convex function is globally lower bounded by its tangents: The RHS of (4) is equal to the second-order Taylor expansion of p around a , up to a rescaling of the quadratic term by a factor $\approx 1/\deg(p)^2$. Its proof relies on a clever combination of basic facts on integration of low-degree (matrix) polynomials: Write the first-order error term of the Taylor series as an integral involving $\nabla^2 p$; apply a classical quadrature rule [CC60] to reduce the integral to a finite, weighted sum $\sum_i w_i \cdot \nabla^2 p(y^{(i)})$ with weights $\approx 1/\deg(p)^2$ [Imh63]; finally, use the fact that $\nabla p(x) \geq 0$ for all x to lower bound the sum by a single term $w_0 \cdot \nabla^2 p(a)$.

From (4), we get a strongly convex lower bound on f if $h_f \not\equiv 0$. Namely, if $\nabla^2 f(a) > 0$, then

$$f(x) \geq q(x) := f(a) + \langle \nabla f(a), x - a \rangle + \frac{\lambda_{\min}(\nabla^2 f(a))}{4 \deg(f)^2} \cdot \|x - a\|^2 \quad (\forall x \in \mathbb{R}^n). \quad (5)$$

This lower bound was used in [ACZ24] to establish that convex polynomials with nonzero hessian are *coercive*, and therefore have a unique global minimizer. For us, this is not enough; we need an explicit bound on the norm of that minimizer. Note that (5) actually gives us a μ -strongly convex lower bound on f , where

$$\mu = \mu(a) = \frac{\lambda_{\min}(\nabla^2 f(a))}{2 \deg(f)^2} > 0. \quad (6)$$

Unfortunately, this μ could be arbitrarily close to 0, as there is no way to control $\lambda_{\min}(\nabla^2 f(a))$. We would need to find an $a \in \mathbb{R}^n$ such that $\mu(a)$ in (6) is at least inversely exponential in $\text{enc}(f)$. Importantly, we need to do so without making any further assumptions on f . We show that this is possible by establishing the following two facts:

- If $h_f \not\equiv 0$, then there exists an $a \in \mathbb{Q}^n$ such that $\nabla^2 f(a) > 0$ with $\text{bl}(a) \leq \text{poly}(n, \deg(f))$. That is, if $h_f(a) \neq 0$ for any $a \in \mathbb{R}^n$, then in fact $h_f(a) \neq 0$ for an $a \in \mathbb{Q}^n$ of small bit length.
- For any $a \in \mathbb{Q}^n$ with $\nabla^2 f(a) > 0$, we have $\lambda_{\min}(\nabla^2 f(a)) \geq 2^{-\text{poly}(\text{enc}(f), \text{bl}(a))}$. That is, for any a of small bit length, the lower bound (5) is μ -strongly convex with $\mu > 0$ not too small.

Combined, these facts show that

$$h_f \not\equiv 0 \implies f \text{ admits a } \mu\text{-strongly convex, quadratic lower bound, where } \mu \geq 2^{-\text{poly}(\text{enc}(f))}. \quad (7)$$

To prove the first fact, note that the set of (real) zeroes of a nonzero polynomial has measure zero in \mathbb{R}^n . Thus, if $h_f \not\equiv 0$, we intuitively expect virtually any point $a \in \mathbb{Q}^n$ to satisfy $\nabla^2 f(a) > 0$. This intuition can be made precise using the Schwartz-Zippel lemma (Lemma 4.5), which gives an explicit set of points $B \subseteq \mathbb{Z}^n$, each of small bit length, such that $h_f(a) \neq 0$ for at least one $a \in B$.

For the the second fact, note that $\nabla^2 f(a)$ is a positive definite, rational matrix whose bit length is bounded by $\text{poly}(\text{enc}(f), \text{bl}(a))$. The eigenvalues of such matrices can be bounded away from 0 using Cramer's rule (see Lemma 4.7 and Corollary 4.8).

Identically zero hessian implies direction of linearity. In light of the above, we investigate when the hessian h_f of a convex polynomial f is identically zero, i.e., when $\nabla^2 f$ is *nowhere* definite. We show that this happens if, *and only if*, f has a (global) direction of linearity. That is to say,

$$h_f \equiv 0 \iff \exists v \in \mathbb{R}^n \setminus \{0\}, c \in \mathbb{R} : \nabla_v f \equiv c. \quad (8)$$

Note that the backward implication of (8) is immediate: if there exist (fixed) $v \in \mathbb{R}^n$, $c \in \mathbb{R}$ so that $\nabla_v f(x) = c$ for all $x \in \mathbb{R}^n$, then v is a kernel vector of $\nabla^2 f(x)$ for all $x \in \mathbb{R}^n$. The forward implication tells us that we may reverse the quantifiers: If f has a *local* direction of linearity for every $x \in \mathbb{R}^n$ (i.e., $\nabla^2 f(x)$ has a kernel vector $v = v(x)$ for all $x \in \mathbb{R}^n$), then in fact it has a *global* direction of linearity. Its validity is more subtle; in particular it *does not* hold for nonconvex polynomials. Interestingly, O. Hesse himself worked on this question, and (mistakenly) claimed the equivalence holds in the general case. We discuss this further below.

To prove the forward implication of (8), assume that $h_f \equiv 0$. We view the gradient ∇f of f as the smooth function $\mathbb{R}^n \rightarrow \mathbb{R}^n$ given by $x \mapsto \nabla f(x)$, and consider its range

$$\nabla f(\mathbb{R}^n) := \{\nabla f(x) : x \in \mathbb{R}^n\} \subseteq \mathbb{R}^n.$$

We can deduce the following two properties of this set:

- The Hessian $\nabla^2 f$ may be seen as the Jacobian of ∇f . The fact that $\det(\nabla^2 f) \equiv 0$ means that every $x \in \mathbb{R}^n$ is a critical point of ∇f . Thus, Sard's theorem (Theorem 4.11) tells us that $\nabla f(\mathbb{R}^n)$ has Lebesgue measure zero in \mathbb{R}^n .
- The range of the gradient of a (sufficiently smooth) convex function is known to be an *almost convex* set, meaning $\nabla f(\mathbb{R}^n)$ contains the interior of its convex hull.⁶

It follows that $\nabla f(\mathbb{R}^n)$ is contained in an affine hyperplane: If not, its convex hull would contain a full-dimensional simplex, and its interior would have positive measure. But that means that there is a $v \in \mathbb{R}^n$ and a $c \in \mathbb{R}$ such that $\nabla_v f(x) := \langle v, \nabla f(x) \rangle = c$ for all $x \in \mathbb{R}^n$, meaning f has a (global) direction of linearity.

One-shot variable elimination: proof of the structure theorem. Next, we explain how to combine the above to prove our structure theorem. Recall that we wish to write a convex polynomial $f \in \mathbb{Q}[x]$ as the sum of a linear function and a polynomial (in fewer variables) that is lower bounded by a strongly convex quadratic. If the hessian $h_f \not\equiv 0$ is not identically zero, we are done immediately by (7). The idea is that this assumption is essentially without loss of generality in light of (8), which tells us that if $h_f \equiv 0$, we can identify a direction in which f is linear. Projecting onto the complement of this direction allows us to reduce to a problem in one fewer variable. Successive application of this procedure yields a decomposition of f into a linear function and a polynomial \hat{f} whose hessian is not identically zero (to which we may then apply (7)). In what follows, we formalize this idea. In fact, we will show how to eliminate all variables “in one shot”.

First, we identify all directions of linearity of f , which make up a subspace $\mathcal{L} \subseteq \mathbb{R}^n$. For this, note that the partial derivatives $\partial f / \partial x_i$ of f are polynomials (in x). Thus, we can adopt a new way of viewing the gradient ∇f , namely as the linear operator $\mathbb{R}^n \rightarrow \mathbb{R}[x]$ given by

$$\nabla f : v \mapsto \nabla_v f = \sum_{i=1}^n v_i \cdot \frac{\partial f}{\partial x_i} \in \mathbb{R}[x] \quad (v \in \mathbb{R}^n).$$

⁶This fact appears to be rather classical, going back at least to [Min64, Corollary 2] (which proves a more general statement). For completeness, we include a short proof using convex conjugates (Lemma 4.10).

Then, $\mathcal{L} \subseteq \mathbb{R}^n$ is just the subspace of vectors mapped to a constant polynomial by ∇f , i.e.,

$$\mathcal{L} := \{v \in \mathbb{R}^n : \nabla_v f \equiv c \text{ for some } c \in \mathbb{R}\}.$$

Clearly, $\mathcal{L} \supseteq \ker(\nabla f)$. In fact, since \mathcal{L} is the inverse image of a subspace of dimension 1, we find that $\mathcal{L} = \ker(\nabla f) \oplus \text{span}\{w\}$, where $w \in \ker(\nabla f)^\perp$ is either zero or satisfies $\nabla_w f \equiv -1$. Now, set $\mathcal{U} = \mathcal{L}^\perp$. Construct a matrix $U \in \mathbb{R}^{k \times n}$ whose rows U_1, \dots, U_k form an orthonormal basis of \mathcal{U} . Let $\hat{f} \in \mathbb{R}[y_1, \dots, y_k]$ be the polynomial defined by $\hat{f}(y) := f(\sum_{i=1}^k y_i U_i)$, so that $\hat{f}(Ux) = f(x_{\mathcal{U}})$. By construction,

$$f(x) = f(x_{\mathcal{U}} + x_{\mathcal{L}}) = f(x_{\mathcal{U}}) - \langle w, x \rangle = \hat{f}(Ux) - \langle w, x \rangle \quad (\forall x \in \mathbb{R}^n). \quad (9)$$

We claim that the hessian of \hat{f} is not identically zero. By (8), it suffices to show that \hat{f} has no directions of linearity. For this, note that the partial derivatives of \hat{f} are linear combinations of the partial derivatives of f along directions in \mathcal{U} . Thus, if \hat{f} had a direction of linearity, this would imply that f has a direction of linearity in \mathcal{U} , which contradicts the fact that $\mathcal{U} = \mathcal{L}^\perp$.

Bit complexity. We have shown how to construct a matrix $U \in \mathbb{Q}^{k \times n}$, a polynomial \hat{f} with nonzero hessian, and a vector w with $Uw = 0$ so that $f(x) = \hat{f}(Ux) - \langle w, x \rangle$. As we noted, we may then apply (7) to \hat{f} to obtain the desired decomposition of f into a linear function and a polynomial with a μ -strongly convex, quadratic lower bound. A technical but important detail that we have ignored here is that in order to fully prove [Theorem 1.3](#), we need to make sure that this μ is sufficiently large, namely at least inverse exponential in $\text{enc}(f)$. For that, it would suffice to show that $\text{enc}(\hat{f})$ is at most polynomial in $\text{enc}(f)$. Moreover, we need to compute the matrix $U \in \mathbb{Q}^{k \times n}$ and the vector $w \in \mathbb{Q}^n$ in polynomial time in $\text{enc}(f)$. In short, we need to show that the linear algebra used to construct U, w and \hat{f} above can be performed in polynomial time (in the bit model).

The important observation is that, for $f \in \mathbb{Q}[x]$, the linear operator $\nabla f : v \mapsto \nabla_v f$ can be expressed as a rational matrix $[\nabla f]$, whose columns represent the (monomial) expansion of the partial derivatives $\partial f / \partial x_i \in \mathbb{Q}[x]$. This matrix has bit length at most $\text{poly}(\text{enc}(f))$. Let $\mathbb{1}$ denote the expansion of the constant polynomial $x \mapsto 1$. It suffices to perform the following operations:

1. Find a $w \in \ker([\nabla f])^\perp$ with $[\nabla f] \cdot w = -\mathbb{1}$, or set $w = 0$ if no solution exists;
2. Compute an orthogonal basis for $\mathcal{U} = \mathcal{L}^\perp$, where $\mathcal{L} := \ker([\nabla f]) \oplus \text{span}\{w\}$.

Both can be achieved in polynomial time via standard results on linear system solving and Gram-Schmidt orthogonalization in the bit model, see [Section A](#). A detail here is that orthonormalization of a set of vectors is typically not possible in polynomial time (the resulting vectors might not be rational). For this reason, the matrix U appearing in [Theorem 1.3](#) has merely orthogonal rows.

Some remarks on scope: Hesse's redemption. One might wonder if the decomposition (9) of a polynomial into a linear part and a part with nonzero hessian holds for *nonconvex* polynomials as well. We only use convexity at one point in its proof, namely to show that $h_f \equiv 0$ if and only if f has a direction of linearity (8). Intuitively, it seems plausible that this equivalence holds in general. It turns out that it does not, and the question has an interesting history. In particular, O. Hesse himself (mistakenly) made this claim [[Hes51](#)], but was later corrected [[GN76](#)]; see also [[GR09](#)]. Without convexity, we are only able to conclude from $h_f \equiv 0$ that the partial derivatives of f are *algebraically dependent*, i.e., there exists a nonzero polynomial π such that $\pi(\partial f / \partial x_1, \dots, \partial f / \partial x_n) \equiv 0$. There are examples where this π has to be of degree at least 2 [[GN76](#)]. For completeness, we include such

an example in [Section B](#). On the other hand, Hesse's claim finds some redemption in this work: it holds for convex polynomials, and this is a crucial ingredient of our proof.

A further subtlety here is that our proof of (8) does not actually make use of the fact that f is a polynomial (although, we do use this fact to *compute* the directions of linearity). Indeed, any (sufficiently smooth) convex function can be written as the sum of a linear function and a function whose Hessian matrix is not everywhere singular. On the other hand, our primary motivation for deriving this decomposition was to then apply (7) to obtain a strongly convex, quadratic lower bound on the nonlinear part. Such a bound cannot be obtained in general. For example, $t \mapsto \exp(t)$ is convex, with *everywhere* definite Hessian, but it is not even coercive.

2.2 Applications in convex polynomial programming

In this section, we show how to apply our structure theorem ([Theorem 1.3](#)) to convex polynomial programming. Let $f \in \mathbb{Q}[x]$ be a convex polynomial, and let $P = \{Ax \leq b\}$ be a (nonempty) convex polyhedron. Our goal is to show that f is either unbounded from below on P (in which case we show that this can be determined efficiently) or it has a minimizer x^* on P of norm at most $2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$ (which can then be identified efficiently by the ellipsoid method). Together, this yields [Theorem 1.1](#) and [Corollary 1.2](#). Recall that the structure theorem gives us a decomposition

$$f(x) = f(x_{\mathcal{U}}) - \langle w, x \rangle \quad (\forall x \in \mathbb{R}^n), \quad (10)$$

where $\mathcal{U} \subseteq \mathbb{R}^n$ is a subspace spanned by the (orthogonal) rows of a matrix $U \in \mathbb{Q}^{k \times n}$, $w \in \mathcal{U}^\perp$, and $f(x_{\mathcal{U}}) \geq q(Ux)$ for some k -variate, μ -strongly convex quadratic, polynomial $q \in \mathbb{Q}[y]$. Furthermore, the bit lengths of U , w and q are all polynomial in $\text{enc}(f)$.

Detecting unboundedness. We first prove that f is unbounded from below if and only if there exists a solution to a certain linear system of inequalities, which will allow us to detect unboundedness efficiently. Recall that in the unconstrained case (i.e., when $P = \mathbb{R}^n$), f is unbounded from below if and only if $w \neq 0$. In the constrained case, we prove that unboundedness is equivalent to the feasibility of the following linear system of inequalities:

$$Ax^0 \leq 0, \quad Ux^0 = 0, \quad \langle w, x^0 \rangle = 1. \quad (11)$$

If $P = \mathbb{R}^n$, (11) reduces to $\exists x^0 : Ux^0 = 0, \langle w, x^0 \rangle = 1$ or in other words $w \neq 0$. Recall that, if $P = \mathbb{R}^n$ and $w \neq 0$, f is unbounded from below since $\lim_{\lambda \rightarrow \infty} f(\lambda w) = -\infty$. For general P , the points λw need not be in P . However, for any solution x^0 to (11) and any $x \in P$, by (10),

$$x + \lambda x^0 \in P \quad (\lambda \geq 0) \quad \text{and} \quad f(x + \lambda x^0) = f(x_{\mathcal{U}}) - \langle w, x + \lambda x^0 \rangle = f(x) - \lambda \rightarrow -\infty \quad (\lambda \rightarrow \infty).$$

This shows condition (11) is sufficient. To show it is also necessary, assume that f is unbounded from below. This means there is a sequence $(x^\ell)_{\ell \geq 1} \subseteq P$ of points with $\lim_{\ell \rightarrow \infty} f(x^\ell) = -\infty$. Using a compactness argument, a subsequence of the normalized points $x^\ell / \|x^\ell\|$ converges to a unit vector x^0 . If $\mathcal{U} \oplus \text{span}\{w\} = \mathbb{R}^n$, using (10), we are able to show that this x^0 satisfies (11). If not, we need to do a more careful limit argument to ensure that $x^0 \notin (\mathcal{U} \oplus \text{span}\{w\})^\perp$; see [Section 5.1](#).

Bounding the norm in a subspace. From now on, we assume that f is bounded from below on P . Recall that our goal is to show that f attains its minimum at a point of small norm. The structure theorem (cf. (10)) shows that the value of $f(x)$ only depends on the component of x in the subspace $\mathcal{U} \oplus \text{span}\{w\}$. Thus, we cannot bound the norm of all minimizers. Instead, we first

show that any minimizer has small norm in the subspace $\mathcal{U} \oplus \text{span}\{w\}$, and then show we can lift to the full space without increasing the norm too much. Using the structure theorem, after writing q as its degree-2 Taylor expansion around 0 and using μ -strong convexity of q , we get

$$f(x) \geq q(0) - \|\nabla q(0)\| \cdot \|Ux\| + \frac{\mu}{2} \|Ux\|^2 - \langle w, x \rangle \quad (\forall x \in \mathbb{R}^n). \quad (12)$$

We want to show that if x^* is a minimizer of f on P , then both $\|Ux^*\|$ and $|\langle w, x^* \rangle|$ are small. To do so, we will show that for any $x \in P$, we have that $|\langle w, x \rangle|$ is large only if $\|Ux\|$ is large as well. Then, since $\|Ux\|$ appears quadratically in (12), we can conclude that both quantities must be small for minimizers.

Formally, since f is bounded from below on P , there does not exist an x^0 as in (11). Farkas' lemma gives us a witness for this fact; namely a vector $\lambda \geq 0$ and a vector $z \in \mathbb{R}^n$ such that $A^\top \lambda + U^\top z - w = 0$. We thus get $\langle w, x \rangle = \langle z, Ux \rangle + \langle \lambda, Ax \rangle$ for any $x \in \mathbb{R}^n$. For $x \in P$, we have $Ax \leq b$. Together with $\lambda \geq 0$, this implies that $\langle \lambda, Ax \rangle \leq \lambda^\top b$ is uniformly bounded for all $x \in P$. Thus, we have

$$|\langle w, x \rangle| \leq \|z\| \cdot \|Ux\| + \|\lambda\| \cdot \|b\| \quad (\forall x \in P). \quad (13)$$

Using this in (12) we get

$$f(x) \geq q(0) - \|\nabla q(0)\| \cdot \|Ux\| - \|z\| \cdot \|Ux\| - \|\lambda\| \cdot \|b\| + \frac{\mu}{2} \|Ux\|^2 \quad (\forall x \in P). \quad (14)$$

As the right-hand side of (14) is a quadratic polynomial in $\|Ux\|$ with positive leading coefficient, $\|Ux^*\|$ needs to be small for any minimizer x^* of f on P . Combining this with (13), this implies that $|\langle w, x^* \rangle|$ is small. To finish the argument, we need to relate these bounds to the bit length of the input. Before we do so, we first show how to lift this bound to the full space.

Lifting to the full space. So far, we have shown that for any minimizer x^* of f on P , the norm in the subspace $\mathcal{U} \oplus \text{span}\{w\}$ needs to be small. The norm of x^* in the full space might still be large and the projection to $\mathcal{U} \oplus \text{span}\{w\}$ might not be feasible. Thus, it remains to lift this projection to a feasible solution in the full space without increasing the norm too much. This lifted point will automatically be a minimizer of f on P since adding any vector in $(\mathcal{U} \oplus \text{span}\{w\})^\perp$ does not change the value of f . Formally, let x^* be any minimizer of f on P . We want to find

$$x' \in P : x'_{\mathcal{U} \oplus \text{span}\{w\}} = x^*_{\mathcal{U} \oplus \text{span}\{w\}} \quad (15)$$

with x' having small norm. The condition (15) defines a linear system. It is feasible since $x^* \in P$. If this system had small bit length, we could conclude immediately that it has a solution x' of small bit length. However, the right-hand side of (15) need not even be rational (we only know its norm is small). We show that the matrix of this linear system has small bit length and the vector of the system has small norm, which we use to show there is a solution of small norm; see [Section 5.3](#).

Bit complexity. It remains to relate the norm bounds on x^* derived above to the bit length of the input. For this, we need to bound all the terms that occur in (13) and (14). By the structure theorem we have that $\text{bl}(q) \leq \text{poly}(\text{enc}(f))$ and $\mu \geq 2^{-\text{poly}(\text{enc}(f))}$ and thus the terms involving q and μ can be bounded appropriately. Using a quantitative version of Farkas' lemma we also get that the terms involving z and λ can be bounded in terms of the bit length of the input. This allows us to conclude that the norm of any minimizer x^* in the subspace $\mathcal{U} \oplus \text{span}\{w\}$ is at most $2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$. The lifting argument then also shows that there exists a minimizer x^* that has norm at most $2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$ in the full space, which completes the proof.

3 Preliminaries

3.1 Notations

Linear algebra. A symmetric matrix $M \in \mathbb{R}^{N \times N}$ is positive (semi)definite if all of its eigenvalues are greater than (or equal to) 0. For a positive semidefinite matrix, we denote its smallest eigenvalue by $\lambda_{\min}(M)$, and its smallest nonzero eigenvalue by $\lambda_{\min}^+(M)$. For a subspace $\mathcal{V} \subseteq \mathbb{R}^N$, and a vector $x \in \mathbb{R}^N$, we write $x_{\mathcal{V}} \in \mathcal{V}$ for the projection of x onto \mathcal{V} . For a subset $S \subseteq \mathbb{R}^N$, we define the affine hull $\text{aff}(S)$ of S as the set of all affine combinations of elements of S , i.e., of points $\sum_{i=1}^k \alpha_i x^i$ where $k \geq 1$, $x^i \in S$ for all $1 \leq i \leq k$ and $\sum_{i=1}^k \alpha_i = 1$. We call points $x^1, \dots, x^k \in \mathbb{R}^N$ affinely independent if no point is an affine combination of the other points.

Smooth (convex) functions. For a sufficiently smooth $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$, we write $\nabla \varphi = (\partial \varphi / \partial x_i)_{1 \leq i \leq n}$ for its gradient, and $\nabla^2 \varphi = (\partial^2 \varphi / \partial x_i \partial x_j)_{1 \leq i, j \leq n}$ for its Hessian. For a vector $v \in \mathbb{R}^n$, we write $\nabla_v \varphi(x) = \langle v, \nabla \varphi(x) \rangle$ for the unnormalized directional derivative. For $c \in \mathbb{R}$, we write $\varphi \equiv c$ if $\varphi(x) = c$ for all $x \in \mathbb{R}^n$. We say $v \in \mathbb{R}^n$ is a *direction of linearity* of φ if $\nabla_v \varphi \equiv c$ for some $c \in \mathbb{R}$, i.e., the directional derivative is constant. For $\mu > 0$, we say that φ is μ -strongly convex if $x \mapsto \varphi(x) - (\mu/2) \cdot \|x\|^2$ is convex, or equivalently if $\lambda_{\min}(\nabla^2 \varphi(x)) \geq \mu$ for all $x \in \mathbb{R}^n$. If there exists any such $\mu > 0$, we say φ is strongly convex.

Polynomials. We write $\mathbb{R}[x]$ (resp. $\mathbb{Q}[x]$) for the space of real (resp. rational) polynomials in n variables $x = (x_1, x_2, \dots, x_n)$, with (monomial) basis $\{x^\alpha : \alpha \in \mathbb{N}^n\}$. For $p(x) = \sum_\alpha p_\alpha x^\alpha$, we write $\text{supp}(p) \subseteq \mathbb{N}^n$ for the set of α with $p_\alpha \neq 0$. We write $h_p(x) := \det(\nabla^2 p(x))$ for the Hessian determinant (or hessian) of p , which is itself a polynomial of degree at most $n \cdot \deg(p)$.

Bit length. The bit length $\text{bl}(k)$ of an integer k is $\max\{1, \lceil \log_2 |k| \rceil\}$, which is the smallest natural number such that $|k| \leq 2^{\text{bl}(k)}$. The bit length of a maximally simplified fraction $r = p/q \in \mathbb{Q}$ is $\text{bl}(p) + \text{bl}(q)$. The bit length of a rational vector or matrix is the sum of the bit lengths of its entries. The bit length of a rational polynomial $f \in \mathbb{Q}[x]$ is the sum of the bit lengths of its (nonzero) coefficients in the monomial basis. We write $\text{enc}(f) = \Theta(n + \deg(f) + \text{bl}(f))$ for the total encoding length of f . For a convex polyhedron $P = \{Ax \leq b\}$ given by a matrix $A \in \mathbb{Q}^{m \times n}$ and vector $b \in \mathbb{Q}^m$, we similarly write $\text{enc}(P) = \text{bl}(A) + \text{bl}(b)$. When $P = \mathbb{R}^n$, we assume that $\text{enc}(P) \geq n$.

3.2 The ellipsoid method

We need the following two statements about the ellipsoid method.

Proposition 3.1 ([Sch86, Theorem 14.1]). *Let $A \in \mathbb{Q}^{M \times N}$ and $b \in \mathbb{Q}^M$. Consider the linear program $P = \{Ax \leq b\}$. In polynomial time in $\text{enc}(P)$ we can check whether the program is feasible and if so, compute a feasible solution. In particular, this feasible solution has bit size polynomial in $\text{enc}(P)$.*

Proposition 3.2. *Let $R > 0$, $\varepsilon > 0$. Let $f \in \mathbb{Q}[x]$ be a polynomial. Let $P = \{Ax \leq b\}$ be a (nonempty) polyhedron. Then, there exists an algorithm that outputs $\tilde{x} \in P$ such that $f(\tilde{x}) \leq \min_{x \in P \cap B_R(0)} f(x) + \varepsilon$ in time $\text{poly}(\text{enc}(f), \text{enc}(P), \log(R), \log(1/\varepsilon))$.*

All the ideas needed to prove Proposition 3.2 are standard results about the ellipsoid method. Similar statements can for example be found in [Sch86, GLS93, Vis21]. However, as we were unable to find the exact statement needed for our application (i.e., Proposition 3.2) we include a proof of

this statement in [Section D](#) for completeness. Our proof mainly relies on the following standard results about the ellipsoid method:

- It is known how to get [Proposition 3.2](#) if P is full-dimensional and in a real model of computation where we are allowed to take square roots (see e.g. [[Vis21](#), Chapter 13]).
- The ellipsoid method can be implemented in the bit model (see e.g. [[GLS93](#), Chapter 3]).
- One can, again using the ellipsoid method, find the affine hull of P , which allows us to reduce to the full-dimensional case (see e.g. [[GLS93](#), Chapters 5 and 6]).

In [Section D](#) we combine these ideas and give a complete proof of [Proposition 3.2](#) (in the bit model). The main idea is to first find the affine hull of P , then project to this affine hull (which does increase the bit complexity by at most a polynomial factor) and finally apply the ellipsoid method on the projected polyhedron, which is now full-dimensional.

4 Structure theorem for convex polynomials

In this section, we give the full proof of our structure theorem for convex polynomials, which we restate for convenience.

Theorem 1.3. *Let $f \in \mathbb{Q}[x]$ be a convex polynomial. Then, there is a subspace $\mathcal{U} \subseteq \mathbb{R}^n$, and a rational vector $w \in \mathcal{U}^\perp$ (possibly $w = 0$), such that, writing $x_{\mathcal{U}}$ for the projection of x onto \mathcal{U} ,*

$$f(x) = f(x_{\mathcal{U}}) - \langle w, x \rangle,$$

where $f(x_{\mathcal{U}})$ has a μ -strongly convex lower bound on \mathcal{U} . That is, there is a matrix $U \in \mathbb{Q}^{k \times n}$ whose rows are orthogonal and span \mathcal{U} , and a k -variate, μ -strongly convex, quadratic polynomial $q \in \mathbb{Q}[y_1, \dots, y_k]$ such that $f(x_{\mathcal{U}}) \geq q(Ux)$ for all $x \in \mathbb{R}^n$. Moreover, the matrix U and the vector w can be computed in polynomial time in $\text{enc}(f)$, the lower bound q satisfies $\text{enc}(q) \leq \text{poly}(\text{enc}(f))$, and $\mu \geq 2^{-\text{poly}(\text{enc}(f))}$.

Our proof relies on two key technical tools. The first provides a μ -strongly convex lower bound on a convex polynomial $f \in \mathbb{Q}[x]$ whose hessian $h_f(x) := \det(\nabla^2 f(x))$ is not identically zero.

Proposition 4.1. *Let $f \in \mathbb{Q}[x]$ be a convex polynomial with $h_f \not\equiv 0$. Then, there is a μ -strongly convex, quadratic $q \in \mathbb{Q}[x]$ with $\text{bl}(q) \leq \text{poly}(\text{enc}(f))$ such that $\hat{f}(x) \geq q(x)$ for all $x \in \mathbb{R}^n$. Moreover, $\mu \geq 2^{-\text{poly}(\text{enc}(f))}$.*

The second shows that any convex polynomial $f \in \mathbb{Q}[x]$ can be written as the sum of a linear function and a polynomial \hat{f} (of fewer variables) with nonzero hessian $h_{\hat{f}} \not\equiv 0$.

Proposition 4.2. *Let $f \in \mathbb{Q}[x]$ be a convex polynomial. Then, there is a subspace $\mathcal{U} \subseteq \mathbb{R}^n$, a matrix $U \in \mathbb{Q}^{k \times n}$ whose orthogonal rows span \mathcal{U} , and a rational vector $w \in \mathcal{U}^\perp$ (possibly $w = 0$), such that*

$$f(x) = \hat{f}(Ux) - \langle w, x \rangle \quad (\forall x \in \mathbb{R}^n), \tag{16}$$

where the polynomial $\hat{f} : \mathbb{R}^k \rightarrow \mathbb{R}$ defined by $\hat{f}(Ux) = f(x_{\mathcal{U}})$ satisfies $h_{\hat{f}} \not\equiv 0$. Moreover, we can compute U, w and \hat{f} in polynomial time in $\text{enc}(f)$. In particular, U, w, \hat{f} have bit length $\text{poly}(\text{enc}(f))$.

Together, these statements immediately imply [Theorem 1.3](#). Indeed, one first decomposes f according to (16), and then applies [Proposition 4.1](#) to \hat{f} . In the remainder of this section, we prove [Proposition 4.1](#) and [Proposition 4.2](#) separately.

4.1 A strongly convex lower bound: Proof of Proposition 4.1

Let $f \in \mathbb{Q}[x]$ be a convex polynomial of degree d with $h_f \not\equiv 0$. For any $a \in \mathbb{R}^n$, we know from [ACZ24, Lemma 5] that

$$f(x) \geq f(a) + \langle \nabla f(a), x - a \rangle + \frac{\langle x - a, \nabla^2 f(a) \cdot (x - a) \rangle}{4d^2} \quad (\forall x \in \mathbb{R}^n).$$

If $\nabla^2 f(a) > 0$, this gives us a μ -strongly convex, quadratic lower bound on f , namely,

$$f(x) \geq q(x) := f(a) + \langle \nabla f(a), x - a \rangle + \mu \cdot \|x - a\|^2 \quad (\forall x \in \mathbb{R}^n),$$

where

$$\mu = \mu(a) = \frac{\lambda_{\min}(\nabla^2 f(a))}{2d^2} > 0.$$

Note that, if $a \in \mathbb{Q}^n$, then $\text{bl}(q) \leq \text{poly}(\text{enc}(f), \text{bl}(a))$. Thus, it suffices to find an $a \in \mathbb{Q}^n$ with $\text{bl}(a) \leq \text{poly}(\text{enc}(f))$ such that $\mu(a) \geq 2^{-\text{poly}(\text{enc}(f))}$. To do so, we establish the following two facts:

Fact 4.3. *If $h_f \not\equiv 0$, then there is an $a \in \mathbb{Q}^n$ with $\text{bl}(a) \leq \text{poly}(n, d)$ such that $h_f(a) > 0$.*

Fact 4.4. *For any $a \in \mathbb{Q}^n$ with $\text{bl}(a) \leq \text{poly}(\text{enc}(f))$ and $h_f(a) > 0$, we have $\mu(a) \geq 2^{-\text{poly}(\text{enc}(f))}$.*

4.1.1 Polynomial identity testing: proof of Fact 4.3

The Schwartz-Zippel lemma is a standard tool in (probabilistic) polynomial identity testing. It states that, for any polynomial $p \not\equiv 0$, and $S \subseteq \mathbb{R}$ finite,

$$\mathbb{P}_{x \sim \text{Unif}(S^n)}[p(x) = 0] \leq \frac{\deg(p)}{|S|},$$

where $\text{Unif}(S^n)$ is the uniform distribution on $S^n \subseteq \mathbb{R}^n$. The following is a deterministic version:

Lemma 4.5 (Schwartz-Zippel, see [Sch80, Corollary 1]). *Let $p \in \mathbb{R}[x]$ be a nonzero polynomial of degree d . Let $S = \{0, 1, \dots, d\}$. Then, there is an $a \in S^n$ with $p(a) \neq 0$.*

Applying the deterministic Schwartz-Zippel lemma to the Hessian determinant $h_f \not\equiv 0$ yields the desired point $a \in \mathbb{Q}^n$ of small bit size for which $h_f(a) > 0$ (recall that $h_f(x) \geq 0$ for all $x \in \mathbb{R}^n$).

Corollary 4.6. *Let p be a polynomial in n variables of degree d and assume that $h_p \not\equiv 0$. Then, there is an $a \in \mathbb{Z}^n$ with $\text{bl}(a) \leq O(n \log(dn))$ such that $h_p(a) = \det(\nabla^2 p(a)) \neq 0$.*

Proof. By assumption, the Hessian determinant h_p of p is a nonzero polynomial in n variables of degree at most dn . By Lemma 4.5, there is thus an $a \in \{0, 1, \dots, dn\}^n$ with $h_p(a) \neq 0$. This a has bit size at most $O(n \log(dn))$. \square

4.1.2 Spectral bounds for rational matrices: proof of Fact 4.4

Now, let $a \in \mathbb{Q}^n$ with $\text{bl}(a) \leq \text{enc}(f)$ and $h_f(a) > 0$. We aim to show that $\mu(a) \geq 2^{-\text{poly}(\text{enc}(f))}$, for which it suffices to show that $\lambda_{\min}(\nabla^2 f(a)) \geq 2^{-\text{poly}(\text{enc}(f))}$. For this, note that $\nabla^2 f(a) > 0$ is a rational matrix of bit size at most $\text{poly}(\text{enc}(f))$. Indeed, the entries of $\nabla^2 f(a)$ each have bit length at most $\text{poly}(\text{enc}(f), \text{bl}(a))$. The eigenvalues of such matrices can be bounded away from 0 as follows.

Lemma 4.7 ([RW17, Lemma 3.1]). *Let $M \in \mathbb{Z}^{N \times N}$ be a positive semidefinite matrix with $|M_{ij}| \leq B$ for all $1 \leq i, j \leq N$. Then $\lambda_{\min}^+(M) \geq (BN)^{-N}$.*

Corollary 4.8. *Let $M \in \mathbb{Q}^{N \times N}$ be a positive definite matrix. Then, $\lambda_{\min}(M) \geq 2^{-\text{poly}(\text{bl}(M))}$.*

Proof. Let $C \in \mathbb{Z}$ be the least common multiple of the denominators of the entries of M . Note that $\text{bl}(C) \leq \text{poly}(\text{bl}(M))$, and so $C \cdot M$ is an integer matrix whose entries are bounded in magnitude by $2^{\text{poly}(\text{bl}(M))}$. Applying Lemma 4.7 yields

$$\lambda_{\min}(M) = \frac{1}{C} \cdot \lambda_{\min}(C \cdot M) \geq \frac{1}{2^{\text{poly}(\text{bl}(M))}} \cdot (2^{\text{poly}(\text{bl}(M))} \cdot N)^{-N} \geq 2^{-\text{poly}(\text{bl}(M))}. \quad \square$$

4.2 A Hessian decomposition theorem: Proof of Proposition 4.2

The primary tool in our proof of Proposition 4.2 is the following characterization of convex functions whose Hessian determinant is identically zero.

Proposition 4.9. *Let $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ be twice continuously differentiable and convex. Then, φ has a direction of linearity if and only if its Hessian is nowhere definite:*

$$\det(\nabla^2 \varphi) \equiv 0 \iff \exists v \in \mathbb{R}^n \setminus \{0\}, c \in \mathbb{R} : \nabla_v \varphi \equiv c.$$

To establish this proposition, we consider $\nabla \varphi$ as function $\mathbb{R}^n \rightarrow \mathbb{R}^n$, and investigate its range

$$\nabla \varphi(\mathbb{R}^n) := \{\nabla \varphi(x) : x \in \mathbb{R}^n\} \subseteq \mathbb{R}^n.$$

We aim to show that $\det(\nabla^2 \varphi) \equiv 0$ if and only if $\nabla \varphi(\mathbb{R}^n)$ is contained in an affine hyperplane. First, we recall that the range of the gradient of a convex function is *almost convex*.

Lemma 4.10 (cf. [Min64, Corollary 2]). *Let $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ be convex and continuously differentiable. Then the range of the gradient of φ is almost convex, i.e., we have $\text{int}(\text{conv}(\nabla \varphi(\mathbb{R}^n))) \subseteq \nabla \varphi(\mathbb{R}^n)$.*

Proof. The *convex conjugate* of $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ is the map

$$\varphi^* : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}, \quad v \mapsto \sup_{x \in \mathbb{R}^n} \{\langle v, x \rangle - \varphi(x)\}.$$

The function φ^* is itself convex (since it is the supremum of affine functions). Furthermore, $\varphi^* \not\equiv \infty$ (i.e., φ^* is a *proper* convex function) [Roc70, Theorem 12.2]. Lastly, for any $x, v \in \mathbb{R}^n$, v is a subgradient of φ at x if and only if x is a subgradient of φ^* at v [Roc70, Theorem 23.5].

Now, since φ is differentiable, we get

$$\nabla \varphi(\mathbb{R}^n) = \{v \in \mathbb{R}^n : \varphi^* \text{ has a subgradient at } v\}. \quad (17)$$

We give a convex set C such that $\text{int}(C) \subseteq \nabla \varphi(\mathbb{R}^n) \subseteq C$. This immediately implies that $\nabla \varphi(\mathbb{R}^n)$ is almost convex, since we then have $\text{conv}(\nabla \varphi(\mathbb{R}^n)) \subseteq C$, and thus also $\text{int}(\text{conv}(\nabla \varphi(\mathbb{R}^n))) \subseteq \text{int}(C)$.

Let $C := \{v \in \mathbb{R}^n \mid \varphi^*(v) \neq \infty\}$ be the *effective domain* of φ^* . Since $\varphi^* \not\equiv \infty$, we find that φ^* has no subgradient at any $v \in \mathbb{R}^n$ with $\varphi^*(v) = \infty$. Hence, using (17), we get $\nabla \varphi(\mathbb{R}^n) \subseteq C$. On the other hand, because φ^* is a proper convex function, it has a subgradient at any point in the relative interior of its effective domain [Roc70, Theorem 23.4]. Thus, $\text{int}(C) \subseteq \text{relint}(C) \subseteq \nabla \varphi(\mathbb{R}^n)$, which completes the proof. \square

Then, we observe that $\nabla^2 \varphi$ is the Jacobian of $\nabla \varphi$. Thus, if $\det(\nabla^2 \varphi) \equiv 0$, we find that $\nabla \varphi(\mathbb{R}^n)$ has Lebesgue measure zero by Sard's theorem.

Theorem 4.11 (Sard's theorem). *Let $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be continuously differentiable. Let*

$$X := \{x \in \mathbb{R}^n : \det(\nabla\psi(x)) = 0\}.$$

Then $\psi(X) \subseteq \mathbb{R}^n$ has Lebesgue measure zero.

Proof of Proposition 4.9. First, assume that $\nabla_v \varphi \equiv c$ for some $v \in \mathbb{R}^n \setminus \{0\}$, $c \in \mathbb{R}$. Then, $\nabla_w \nabla_v \varphi(x) = 0$ for all $w \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$, meaning $\nabla^2 \varphi(x) \cdot v = 0$. In particular, $\det(\nabla^2 \varphi(x)) = 0$.

Now, assume $\det(\nabla^2 \varphi) \equiv 0$. By the above, this implies that $\nabla\varphi(\mathbb{R}^n)$ is an almost convex set of Lebesgue measure 0. But that means $\nabla\varphi(\mathbb{R}^n)$ is contained in an affine hyperplane, for, if not, the interior of its convex hull would contain (the interior of) a simplex of positive measure. Thus, there is a $v \in \mathbb{R}^n \setminus \{0\}$, and $c \in \mathbb{R}$ such that $\langle y, v \rangle = c$ for all $y \in \nabla\varphi(\mathbb{R}^n)$. In other words, $\nabla_v \varphi \equiv c$. \square

Proof of Proposition 4.2. With Proposition 4.9 in hand, we are able to identify the subspace $\mathcal{U} \subseteq \mathbb{R}^n$ and vector $w \in \mathbb{Q}^n$ that feature in Proposition 4.2. Let $f \in \mathbb{Q}[x]$ be a convex polynomial. Consider the joint support

$$S = \{0\} \cup \bigcup_{i=1}^n \text{supp}(\partial f / \partial x_i) \subseteq \mathbb{N}^n$$

of the partial derivatives of f , which is of size $|S| \leq n \cdot |\text{supp}(f)| + 1$. (Here, we have added 0 $\in \mathbb{N}^n$ for technical reasons that become clear shortly.) Set $\mathcal{R} = \text{span}\{x^\alpha : \alpha \in S\} \subseteq \mathbb{R}[x]$. We view the gradient of f as a linear operator $\mathbb{R}^n \rightarrow \mathcal{R}$, namely

$$\nabla f : v \mapsto \nabla_v f = \sum_{i=1}^n v_i \cdot \frac{\partial f}{\partial x_i} \in \mathcal{R} \quad (v \in \mathbb{R}^n).$$

Importantly, this operator can be expressed as a rational matrix $[\nabla f]$ of size $|S| \times n$, where the i th column represents the expansion of $\partial f / \partial x_i$ in the (partial) monomial basis. Moreover, this matrix has bit size at most $\text{poly}(\text{enc}(f))$. Write $\mathbb{1} \in \mathbb{Q}^{|S|}$ for the expansion of the constant polynomial $x \mapsto 1$ (which exists as $0 \in S$). In light of Proposition 4.9, we consider the subspace

$$\mathcal{L} := \{v \in \mathbb{R}^n : \nabla_v f \equiv c \text{ for some } c \in \mathbb{R}\} = \{v \in \mathbb{R}^n : [\nabla f] \cdot v \in \text{span}\{\mathbb{1}\}\}.$$

Since \mathcal{L} is the inverse image of a subspace of dimension 1, we can write

$$\mathcal{L} = \ker([\nabla f]) \oplus \text{span}\{w\},$$

where $w \in \ker([\nabla f])^\perp$ is either zero, or satisfies $[\nabla f] \cdot w = -\mathbb{1}$. Then, we set $\mathcal{U} = \mathcal{L}^\perp$, and write $\mathcal{V} = \ker([\nabla f])$, $\mathcal{W} = \text{span}\{w\}$. By construction, we have

$$f(x) = f(x_{\mathcal{U}} + x_{\mathcal{V}} + x_{\mathcal{W}}) = f(x_{\mathcal{U}}) - \langle w, x \rangle.$$

We can compute w , and an orthogonal basis U_1, \dots, U_k for \mathcal{U} in polynomial time in $\text{enc}(f)$. In particular, w and the matrix $U \in \mathbb{Q}^{k \times n}$ whose rows are U_1, \dots, U_k have polynomial bit size in $\text{enc}(f)$. This follows from standard (bit-)complexity results in linear algebra, see Section A.

For $1 \leq i \leq k$, write $\bar{U}_i = U_i / \|U_i\|^2$, and note that $\text{bl}(\bar{U}_i) \leq \text{poly}(\text{enc}(f))$. Consider the k -variate polynomial $\hat{f} \in \mathbb{Q}[y_1, \dots, y_k]$ defined by $\hat{f}(y) = f(\sum_{i=1}^k y_i \cdot \bar{U}_i)$. Note that \hat{f} is convex, and that $\text{enc}(\hat{f}) \leq \text{poly}(\text{enc}(f))$. Furthermore, for any $x \in \mathbb{R}^n$,

$$\hat{f}(Ux) = f\left(\sum_{i=1}^k \frac{\langle x, U_i \rangle}{\|U_i\|^2} \cdot U_i\right) = f(x_{\mathcal{U}}).$$

It remains to show that $h_{\hat{f}} \not\equiv 0$. By [Proposition 4.9](#), this is equivalent to showing it has no directions of linearity. Suppose that it did, i.e., that there is a nonzero $\hat{v} \in \mathbb{R}^k$ such that $\nabla_{\hat{v}} \hat{f} \equiv c$ for some $c \in \mathbb{R}$. Then, writing $e_j \in \mathbb{R}^n$ for the j -th standard basis vector,

$$\nabla_{e_j} \hat{f}(y) = \nabla_{\bar{U}_j} f \left(\sum_{i=1}^k y_i \cdot \bar{U}_i \right).$$

Now, write $u = \sum_{i=1}^k \hat{v}_i \bar{U}_i \in \mathcal{U} = \mathcal{L}^\perp$. Note that $u \neq 0$. Then, as $\langle w, u \rangle = 0$, for any $x \in \mathbb{R}^n$,

$$\nabla_u f(x) = \nabla_u f \left(\sum_{i=1}^k \langle x, U_i \rangle \cdot \bar{U}_i \right) = \nabla_{\hat{v}} \hat{f}(Ux) = c,$$

a contradiction. \square

5 Applications in convex polynomial programming

In this section, we give a full proof of [Theorem 1.1](#) and [Corollary 1.2](#), which we restate below.

Theorem 1.1. *Let $f \in \mathbb{Q}[x]$ be a convex polynomial. Let P be a (nonempty) convex polyhedron. Then, either f is unbounded from below on P , or it attains its minimum on P at a point x^* of norm*

$$\log \|x^*\| \leq \text{poly}(\mathbf{enc}(f), \mathbf{enc}(P)).$$

Corollary 1.2. *Let $f \in \mathbb{Q}[x]$ be a convex polynomial. Let P be a (nonempty) convex polyhedron. Let $\varepsilon > 0$. We can decide in time $\text{poly}(\mathbf{enc}(f), \mathbf{enc}(P), \log(1/\varepsilon))$ whether f is unbounded from below on P , and, if not, output a point $\tilde{x} \in P$ with $f(\tilde{x}) \leq \min_{x \in P} f(x) + \varepsilon$.*

For the proof of [Theorem 1.1](#) and [Corollary 1.2](#) we need [Theorem 1.3](#), which we also restate below for convenience.

Theorem 1.3. *Let $f \in \mathbb{Q}[x]$ be a convex polynomial. Then, there is a subspace $\mathcal{U} \subseteq \mathbb{R}^n$, and a rational vector $w \in \mathcal{U}^\perp$ (possibly $w = 0$), such that, writing $x_{\mathcal{U}}$ for the projection of x onto \mathcal{U} ,*

$$f(x) = f(x_{\mathcal{U}}) - \langle w, x \rangle,$$

where $f(x_{\mathcal{U}})$ has a μ -strongly convex lower bound on \mathcal{U} . That is, there is a matrix $U \in \mathbb{Q}^{k \times n}$ whose rows are orthogonal and span \mathcal{U} , and a k -variate, μ -strongly convex, quadratic polynomial $q \in \mathbb{Q}[y_1, \dots, y_k]$ such that $f(x_{\mathcal{U}}) \geq q(Ux)$ for all $x \in \mathbb{R}^n$. Moreover, the matrix U and the vector w can be computed in polynomial time in $\mathbf{enc}(f)$, the lower bound q satisfies $\mathbf{enc}(q) \leq \text{poly}(\mathbf{enc}(f))$, and $\mu \geq 2^{-\text{poly}(\mathbf{enc}(f))}$.

In particular, this gives us the following lower bound on f :

$$f(x) \geq q(0) - \|\nabla q(0)\| \cdot \|Ux\| - \langle w, x \rangle + \frac{\mu}{2} \|Ux\|^2. \quad (18)$$

We define $\mathcal{W} := \text{span}\{w\}$ for w as in [Theorem 1.3](#). In order to prove [Theorem 1.1](#) and [Corollary 1.2](#), we also need the following three lemmas. [Lemma 5.1](#) shows how we can efficiently detect whether f is unbounded from below. [Lemma 5.2](#) then shows that any minimizer needs to have small norm in the subspace $\mathcal{U} \oplus \mathcal{W}$. Finally, [Lemma 5.3](#) shows how to lift this norm bound in a subspace to the full space.

Lemma 5.1. Let $f \in \mathbb{Q}[x]$ be a convex polynomial and let $P = \{Ax \leq b\}$ be a (nonempty) polyhedron. Let \mathcal{U} and w as in [Theorem 1.3](#). Then, f is unbounded from below on P , if and only if there exists $x^0 \in \mathbb{R}^n$ with $Ax^0 \leq 0$, $x^0 \in \mathcal{U}^\perp$ (or equivalently $Ux^0 = 0$ for U as in [Theorem 1.3](#)) and $\langle w, x^0 \rangle = 1$.

Lemma 5.2. Let $f \in \mathbb{Q}[x]$ be a convex polynomial and let $P = \{Ax \leq b\}$ be a (nonempty) polyhedron. Let \mathcal{U} and w as in [Theorem 1.3](#) and let $x \in P$. Assume that x satisfies $\|x_{\mathcal{U} \oplus \mathcal{W}}\| > 2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$. Then there is a point $x' \in P$ with $\|x'\| \leq 2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$ and $f(x') < f(x)$. In particular, if f is bounded from below on P , then every minimizer $x^* \in P$ of f on P satisfies $\|x^*_{\mathcal{U} \oplus \mathcal{W}}\| \leq 2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$.

Lemma 5.3. Let $P = \{Ax \leq b\}$ be a (nonempty) polyhedron. Let \mathcal{Z} be a linear subspace of \mathbb{R}^n spanned by orthogonal vectors x^1, \dots, x^k . Let $x^* \in P$. There is an $x' \in P$ such that $x^*_{\mathcal{Z}} = x'_{\mathcal{Z}}$ and

$$\|x'\| \leq \text{poly}\left(\|x^*_{\mathcal{Z}}\|, 2^{\text{poly}(\text{enc}(P), \text{bl}(\mathcal{Z}))}\right),$$

where $\text{bl}(\mathcal{Z}) = \text{bl}(x^1) + \dots + \text{bl}(x^k)$.

We now first give a formal proof of [Theorem 1.1](#) and [Corollary 1.2](#) given the lemmas and then prove [Lemmas 5.1](#) to [5.3](#) in [Sections 5.1](#) to [5.3](#).

Proof of Theorem 1.1. Assume that f is bounded from below (otherwise we are immediately done). We claim that

$$\inf_{x \in P} f(x) = \inf_{x \in P: \|x\| \leq R} f(x) \tag{19}$$

for some $R = 2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$. Since the set on the right-hand side is compact and f is continuous, this will allow us to conclude that f attains its minimum. Moreover, it attains it at a point x^* with $\|x^*\| \leq R$, which will complete the proof. Thus, it remains to prove (19). Assume by contradiction that this is not true. Then, there is $x^0 \in P$ with

$$f(x^0) < \inf_{x \in P: \|x\| \leq R} f(x).$$

By [Lemma 5.2](#), we can assume without loss of generality that $\|(x^0)_{\mathcal{U} \oplus \mathcal{W}}\| \leq 2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$. Otherwise, we could replace x^0 by a point $x^{0'}$ that satisfies this and has $f(x^{0'}) < f(x^0)$. By [Lemma 5.3](#) (setting $\mathcal{Z} = \mathcal{U} \oplus \mathcal{W}$ and x^1, \dots, x^k being the rows of U and w), we can now find an $x' \in P$ such that $x^0_{\mathcal{U} \oplus \mathcal{W}} = x'_{\mathcal{U} \oplus \mathcal{W}}$ and

$$\|x'\| \leq \text{poly}\left(\|x^0_{\mathcal{U} \oplus \mathcal{W}}\|, 2^{\text{poly}(\text{enc}(P), \text{bl}(\mathcal{U} \oplus \mathcal{W}))}\right) \leq 2^{\text{poly}(\text{enc}(f), \text{enc}(P))}. \tag{20}$$

Here we used that by [Theorem 1.3](#) we have $\text{bl}(\mathcal{U} \oplus \mathcal{W}) \leq \text{poly}(\text{enc}(f))$. Choosing R as this upper bound, we have

$$f(x') = f(x'_{\mathcal{U} \oplus \mathcal{W}}) = f(x^0_{\mathcal{U} \oplus \mathcal{W}}) = f(x^0) < \inf_{x \in P: \|x\| \leq R} f(x),$$

which contradicts (20). Thus, (19) holds, which completes the proof. \square

Proof of Corollary 1.2. Consider [Algorithm 1](#), which we want to use to show [Corollary 1.2](#). Correctness follows from [Lemma 5.1](#) and [Theorem 1.1](#). The runtime is $\text{poly}(\text{enc}(f), \text{enc}(P), \log(1/\varepsilon))$. Indeed, by [Theorem 1.3](#), line 3 can be done in time $\text{poly}(\text{enc}(f))$. The check in line 4 can be done in time $\text{poly}(\text{enc}(f), \text{enc}(P))$ by [Proposition 3.1](#) since it checks for feasibility of a linear program and the bit lengths of A , U and w are bounded by $\text{poly}(\text{enc}(f), \text{enc}(P))$. Finally, the ellipsoid method from [Proposition 3.2](#) in line 6 runs in time $\text{poly}(\text{enc}(f), \text{enc}(P), \log(R), \log(1/\varepsilon))$, which is also $\text{poly}(\text{enc}(f), \text{enc}(P), \log(1/\varepsilon))$ since we have $R = 2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$. \square

Algorithm 1 Algorithm for convex polynomial programming

- 1: **Input:** A polynomial f , a matrix $A \in \mathbb{Q}^{m \times n}$, a vector $b \in \mathbb{Q}^m$ (defining a nonempty polyhedron $P = \{Ax \leq b\}$), an error parameter ε
 - 2: **Output:** Either $-\infty$ if f is unbounded on P or a point $\tilde{x} \in P$ with $f(\tilde{x}) \leq \min_{x \in \mathbb{R}^n} f(x) + \varepsilon$
 - 3: Compute U and w as in [Theorem 1.3](#).
 - 4: **if** there exists $x^0 \in \mathbb{R}^n$ with $Ax^0 \leq 0$, $Ux^0 = 0$ and $\langle w, x^0 \rangle = 1$ **then return** $-\infty$
 - 5: **else**
 - 6: Run the ellipsoid method from [Proposition 3.2](#) with input $f, P, R = 2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$ (as in [Theorem 1.1](#)) and ε and **return** the output of the ellipsoid method.
 - 7: **end if**
-

5.1 Deciding unboundedness: Proof of [Lemma 5.1](#)

In this section, we prove [Lemma 5.1](#), i.e. we show that f is unbounded on P if and only if there exists an $x^0 \in \mathbb{R}^n$ with $Ax^0 \leq 0$, $x^0 \in U^\perp$ and $\langle w, x^0 \rangle = 1$. We first show that this condition is sufficient.

Proof of [Lemma 5.1](#) (part 1: condition is sufficient). Let $x \in P$ and consider the points $x + \lambda x^0$ for $\lambda \geq 0$. We have

$$A(x + \lambda x^0) = Ax + \lambda Ax^0 \leq b + \lambda \cdot 0 = b$$

and thus $x + \lambda x^0 \in P$ for any $\lambda \geq 0$. Furthermore

$$f(x + \lambda x^0) = f((x + \lambda x^0)_U) - \langle w, x + \lambda x^0 \rangle.$$

Since $x^0 \in U^\perp$, we have $(x + \lambda x^0)_U = x_U$. Thus, using $\langle w, x^0 \rangle = 1$,

$$f(x + \lambda x^0) = f(x_U) - \langle w, x \rangle - \lambda,$$

showing that $f(x + \lambda x^0) \rightarrow -\infty$ as $\lambda \rightarrow \infty$. Hence, if the condition is satisfied, then f is unbounded from below on P . \square

To show that the condition is also necessary, assume that f is unbounded from below on P . Consider a sequence $(x^k)_{k \geq 1} \subseteq P$ with $\lim_{k \rightarrow \infty} f(x^k) = -\infty$. Since the unit sphere S^{n-1} is compact, a subsequence of the normalized points $x^k / \|x^k\|$ converges to a point x^0 . Unfortunately, x^0 could be in $(U \oplus W)^\perp$ (in particular meaning that $\langle w, x^0 \rangle = 0$). Instead, we get the following claim about the projected normalized points. (Note that, since $f(x) = f(x_{U \oplus W})$ for all $x \in \mathbb{R}^n$, the statements $\lim_{k \rightarrow \infty} f(x^k) = -\infty$ and $\lim_{k \rightarrow \infty} f(x_{U \oplus W}^k) = -\infty$ are equivalent.)

Claim 5.4. *Consider a sequence $(x^k)_{k \geq 1}$ such that $\lim_{k \rightarrow \infty} f(x^k) = -\infty$. Then a subsequence of the normalized projected points $x_{U \oplus W}^k / \|x_{U \oplus W}^k\|$ converges to a point $y^0 \in W$ with $\langle w, y^0 \rangle > 0$.*

Before we prove this claim, we show how to apply it to prove that the condition of [Lemma 5.1](#) is also necessary. For this, we want to argue that we can lift this limit to a point in the full space satisfying the conditions of [Lemma 5.1](#). In order to do so, we need the following two propositions. Recall that the recession cone $\text{recc}(C)$ of a convex set C is the set of directions in which C extends to infinity, i.e.

$$\text{recc}(C) = \{y \in \mathbb{R}^n : x + \lambda y \in C \ \forall x \in C \ \forall \lambda \geq 0\}.$$

Proposition 5.5 ([Roc70], (Special case of Corollary 8.3.4]). *Let C be a (nonempty) closed convex set. Let \mathcal{Z} be a subspace. Consider the projection*

$$C_{\mathcal{Z}} = \{x \in \mathcal{Z} : \exists x' \in \mathcal{Z}^\perp \text{ with } x + x' \in C\}.$$

Then the projection of the recession cone of C is the recession cone of the projection $C_{\mathcal{Z}}$.

Proposition 5.6 ([Roc70, Theorem 19.3]). *The projection of the polyhedron $P = \{Ax \leq b\}$ to a subspace \mathcal{Z} is again a polyhedron, i.e., it can be described as*

$$P_{\mathcal{Z}} = \{x \in \mathcal{Z} : \exists x' \in (\mathcal{Z})^\perp \text{ with } x + x' \in P\} = \{\hat{A}x \leq \hat{b}\}.$$

Note that the dimensions of \hat{A} and \hat{b} can be exponential in n , but since we are only interested in proving existence of an x^0 as in Lemma 5.1 this is not a problem. We now show how to use Claim 5.4 and Propositions 5.5 and 5.6 to prove Lemma 5.1.

Proof of Lemma 5.1 (part 2: condition is necessary). Since f is unbounded, there is a sequence $(x^k)_{k \geq 1}$ of points in P such that $\lim_{k \rightarrow \infty} f(x^k) = -\infty$. Thus, using Claim 5.4 we get a $y^0 \in \mathcal{W}$ with

$$y^0 = \lim_{k \rightarrow \infty} \frac{x_{\mathcal{U} \oplus \mathcal{W}}^k}{\|x_{\mathcal{U} \oplus \mathcal{W}}^k\|}.$$

Note that we can assume without loss of generality that y^0 is the limit of all the points (and not just of a subsequence). Since $y^0 \in \mathcal{W}$, we in particular have $y^0 \in \mathcal{U}^\perp$.

We want to show that we can lift y^0 to a point x^0 (meaning to an x^0 such that $x^0 - y^0 \in (\mathcal{U} \oplus \mathcal{W})^\perp$) with $Ax^0 \leq 0$. We have $x_{\mathcal{U} \oplus \mathcal{W}}^k \in P_{\mathcal{U} \oplus \mathcal{W}}$. By Proposition 5.6, $P_{\mathcal{U} \oplus \mathcal{W}} = \{\hat{A}x \leq \hat{b}\}$ for some \hat{A} and \hat{b} . Since we also have $\lim_{k \rightarrow \infty} f(x_{\mathcal{U} \oplus \mathcal{W}}^k) = -\infty$, it follows that $\lim_{k \rightarrow \infty} \|x_{\mathcal{U} \oplus \mathcal{W}}^k\| = \infty$ (f is bounded over any compact set) and thus also

$$\hat{A}y^0 = \lim_{k \rightarrow \infty} \frac{1}{\|x_{\mathcal{U} \oplus \mathcal{W}}^k\|} \hat{A}x_{\mathcal{U} \oplus \mathcal{W}}^k \leq \lim_{k \rightarrow \infty} \frac{1}{\|x_{\mathcal{U} \oplus \mathcal{W}}^k\|} \hat{b} = 0.$$

Thus, y^0 is in the recession cone of $P_{\mathcal{U} \oplus \mathcal{W}}$, which by Proposition 5.5 is the projection of the recession cone $\text{recc}(P) = \{x \in \mathbb{R}^n : Ax \leq 0\}$ of P . This implies that we can lift y^0 to a point $x^0 \in \text{recc}(P)$.

We then have $Ax^0 \leq 0$ by definition of $\text{recc}(P)$. Since x^0 and y^0 only differ in $(\mathcal{U} \oplus \mathcal{W})^\perp$, we still have $x^0 \in \mathcal{U}^\perp$. Since $\langle w, x^0 \rangle = \langle w, y^0 \rangle > 0$, by rescaling, we can get $\langle w, x^0 \rangle = 1$, which completes the proof. \square

It remains to prove Claim 5.4.

Proof of Claim 5.4. Consider the projections $x_{\mathcal{U} \oplus \mathcal{W}}^k$. We have $\lim_{k \rightarrow \infty} f(x_{\mathcal{U} \oplus \mathcal{W}}^k) = -\infty$. Thus, by compactness of the unit sphere, a subsequence of the normalized points $x_{\mathcal{U} \oplus \mathcal{W}}^k / \|x_{\mathcal{U} \oplus \mathcal{W}}^k\|$ converges to a point $y^0 \in \mathcal{U} \oplus \mathcal{W}$. Without loss of generality, we can replace the sequence $(x^k)_{k \geq 1}$ by the elements corresponding to this subsequence and thus assume $\lim_{k \rightarrow \infty} x_{\mathcal{U} \oplus \mathcal{W}}^k / \|x_{\mathcal{U} \oplus \mathcal{W}}^k\| = y^0$.

By (18), we can conclude two things: First, we need to have $\lim_{k \rightarrow \infty} \langle w, x_{\mathcal{U} \oplus \mathcal{W}}^k \rangle = \infty$. Without loss of generality we can thus assume $\langle w, x_{\mathcal{U} \oplus \mathcal{W}}^k \rangle > 0$ for all k , which implies $\langle w, y^0 \rangle \geq 0$. Second, we need to have

$$\lim_{k \rightarrow \infty} \frac{\left\| U \frac{x_{\mathcal{U} \oplus \mathcal{W}}^k}{\|x_{\mathcal{U} \oplus \mathcal{W}}^k\|} \right\|}{\left\langle w, \frac{x_{\mathcal{U} \oplus \mathcal{W}}^k}{\|x_{\mathcal{U} \oplus \mathcal{W}}^k\|} \right\rangle} = \lim_{k \rightarrow \infty} \frac{\|Ux_{\mathcal{U} \oplus \mathcal{W}}^k\|}{\langle w, x_{\mathcal{U} \oplus \mathcal{W}}^k \rangle} = 0.$$

Indeed, if $\langle w, x_{U \oplus W}^k \rangle \leq c \|Ux_{U \oplus W}^k\|$ for all k and a constant c independent of k , then, by (18),

$$f(x_{U \oplus W}^k) \geq q(0) - (\|\nabla q(0)\| + c) \cdot \|Ux_{U \oplus W}^k\| + \frac{\mu}{2} \|Ux_{U \oplus W}^k\|^2.$$

The right hand side is globally lower bounded since it is a quadratic polynomial with positive leading coefficient, which contradicts $\lim_{k \rightarrow \infty} f(x_{U \oplus W}^k) = -\infty$.

In particular, since $\left\langle w, \frac{x_{U \oplus W}^k}{\|x_{U \oplus W}^k\|} \right\rangle \leq \|w\|$ this furthermore implies

$$\|Uy^0\| = \lim_{k \rightarrow \infty} \left\| U \frac{x_{U \oplus W}^k}{\|x_{U \oplus W}^k\|} \right\| = 0$$

and thus $Uy^0 = 0$ or in other word $y^0 \in U^\perp$. Hence, $y^0 \in W$ and since

$$\|y^0\| = \lim_{k \rightarrow \infty} \left\| \frac{x_{U \oplus W}^k}{\|x_{U \oplus W}^k\|} \right\| = 1,$$

we also need to have $\langle w, y^0 \rangle > 0$ (as opposed to just $\langle w, y^0 \rangle \geq 0$), which completes the proof. \square

5.2 Bounding the norm of minimizers in a subspace: Proof of Lemma 5.2

In order to prove Lemma 5.2, we want to use (18) and show that if $\|Ux\|$ or $|\langle w, x \rangle|$ is large, then x cannot be a minimizer. We need Farkas' lemma to get a certificate for the fact that there is no x^0 as in Lemma 5.1.

Proposition 5.7 (Farkas' lemma, see e.g. [Sch86, Corollary 7.1e]). *Let $C \in \mathbb{R}^{M \times N}$ be a matrix and $d \in \mathbb{R}^M$ be a vector. Exactly one of the following two statements hold:*

- The system $Cx \leq d$ has a solution.
- There is a vector $y \in \mathbb{R}_{\geq 0}^M$ with $C^\top y = 0$ and $d^\top y < 0$.

Applying this to the system from Lemma 5.1, we get the following.

Lemma 5.8. *Let $f \in \mathbb{Q}[x]$ be a convex polynomial and let $P = \{Ax \leq b\}$ be a polyhedron. Let U and w as in Theorem 1.3. If f is bounded from below on P , then there exist vectors $\lambda \in \mathbb{R}_{\geq 0}^n$ and $z \in \mathbb{R}^k$ such that*

$$w = A^\top \lambda + U^\top z. \tag{21}$$

Furthermore, there exist such λ and z with $\text{bl}(\lambda), \text{bl}(z) \leq \text{poly}(\text{enc}(f), \text{bl}(A))$.

Proof. Since f is bounded from below on P , there does not exist a vector as in Lemma 5.1, i.e. there is no x^0 with $Ax^0 \leq 0$, $Ux^0 = 0$ and $w^\top x^0 = 1$. By applying Proposition 5.7 to the system

$$C = \begin{bmatrix} A \\ U \\ -U \\ w^\top \\ -w^\top \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

we get that there is a vector $y \geq 0$ and $C^\top y = 0$ and $d^\top y < 0$. Decomposing

$$y = \begin{bmatrix} \lambda' \\ y^1 \\ y^2 \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

for $\lambda' \in \mathbb{R}_{\geq 0}^n$, $y^1, y^2 \in \mathbb{R}_{\geq 0}^k$, $\alpha_1, \alpha_2 \in \mathbb{R}_{\geq 0}$, we get

$$A^\top \lambda' + U^\top (y^1 - y^2) + (\alpha_1 - \alpha_2)w = 0 \text{ and } \alpha_1 - \alpha_2 < 0.$$

Rescaling this by $\frac{1}{\alpha_2 - \alpha_1} > 0$ and defining $\lambda = \frac{1}{\alpha_2 - \alpha_1} \lambda'$ and $z = \frac{1}{\alpha_2 - \alpha_1} (y^1 - y^2)$, we get

$$A^\top \lambda + U^\top z = w.$$

By [Proposition 3.1](#), we can even get λ and z with

$$\text{bl}(\lambda), \text{bl}(z) \leq \text{poly}(\text{bl}(C), \text{bl}(d)) = \text{poly}(\text{bl}(A), \text{bl}(U), \text{bl}(w)) = \text{poly}(\text{enc}(f), \text{bl}(A)),$$

where the last step used that $\text{bl}(U), \text{bl}(w) \leq \text{poly}(\text{enc}(f))$ by [Theorem 1.3](#). Note that, since rescaling y by positive scalars does not change feasibility, we can replace $d^\top y < 0$ by $d^\top y = -1$ (thus making it a linear program as in [Proposition 3.1](#)). \square

We can now prove [Lemma 5.2](#).

Proof of Lemma 5.2. Let $\lambda \in \mathbb{R}_{\geq 0}^n$ and $z \in \mathbb{R}^k$ as in [Lemma 5.8](#). For any $x \in \mathbb{R}^n$ we get

$$\langle w, x \rangle = \langle A^\top \lambda, x \rangle + \langle U^\top z, x \rangle = \langle \lambda, Ax \rangle + \langle z, Ux \rangle.$$

For $x \in P$ we have $Ax \leq b$, which together with $\lambda \geq 0$ implies

$$\langle \lambda, Ax \rangle \leq \lambda^\top b.$$

Thus, we get for any $x \in P$

$$|\langle w, x \rangle| \leq \|\lambda\| \cdot \|b\| + \|z\| \cdot \|Ux\|. \quad (22)$$

Using (22) in (18), we get for all $x \in P$

$$\begin{aligned} f(x) &\geq q(0) - \|\nabla q(0)\| \cdot \|Ux\| - \|\lambda\| \cdot \|b\| - \|z\| \cdot \|Ux\| + \frac{\mu}{2} \|Ux\|^2 \\ &= (q(0) - \|\lambda\| \cdot \|b\|) + (-\|\nabla q(0)\| - \|z\|) \cdot \|Ux\| + \frac{\mu}{2} \|Ux\|^2 \end{aligned}$$

Fix $a \in P$ with $\|a\| \leq 2^{\text{poly}(\text{enc}(P))}$ (such a point exists by [Proposition 3.1](#)). We are interested when the lower bound is at least $f(a)$, i.e. when

$$(q(0) - \|\lambda\| \cdot \|b\| - f(a)) + (-\|\nabla q(0)\| - \|z\|) \cdot \|Ux\| + \frac{\mu}{2} \|Ux\|^2 \geq 0$$

Since this is a quadratic polynomial in $\|Ux\|$ with positive leading coefficient, as $\|Ux\| \rightarrow \infty$ this is positive. Note that for $x = a$ the lower bound needs to be non-positive, i.e. this polynomial in

$\|Ux\|$ has at least one root. Thus, if $\|Ux\|$ is larger than the largest root, it needs to be positive. That is, whenever

$$\|Ux\| > \frac{\|\nabla q(0)\| + \|z\| + \sqrt{(\|\nabla q(0)\| + \|z\|)^2 + 2\mu(f(a) + \|\lambda\| \cdot \|b\| - q(0))}}{\mu}$$

for some $x \in P$, then we have $f(x) > f(a)$. Furthermore, combining this with (22), whenever

$$|\langle w, x \rangle| > \|\lambda\| \cdot \|b\| + \|z\| \cdot \frac{\|\nabla q(0)\| + \|z\| + \sqrt{(\|\nabla q(0)\| + \|z\|)^2 + 2\mu(f(a) + \|\lambda\| \cdot \|b\| - q(0))}}{\mu}$$

for some $x \in P$, then we also have $f(x) > f(a)$.

Thus, if $\|Ux\|$ or $|\langle w, x \rangle|$ are large, then $f(x) > f(a)$. We now want to quantify this in terms of the bit length of the input. We do this term by term:

- By [Theorem 1.3](#), we have $\text{bl}(q) \leq \text{poly}(\text{enc}(f))$ and thus $|q(0)|, \|\nabla q(0)\| \leq 2^{\text{poly}(\text{enc}(f))}$.
- By [Theorem 1.3](#), we have $\mu \geq 2^{-\text{poly}(\text{enc}(f))}$.
- By [Lemma 5.8](#), we have $\text{bl}(z) \leq \text{poly}(\text{bl}(A), \text{bl}(U), \text{bl}(w)) \leq \text{poly}(\text{enc}(f), \text{bl}(A))$ and thus $\|z\| \leq 2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$.
- By [Lemma 5.8](#), we have $\text{bl}(\lambda) \leq \text{poly}(\text{bl}(A), \text{bl}(U), \text{bl}(w)) \leq \text{poly}(\text{enc}(f), \text{bl}(A))$ and thus $\|\lambda\| \leq 2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$.
- Since we have $\|a\| \leq 2^{\text{poly}(\text{enc}(P))}$, we can bound $|f(a)| \leq 2^{\text{poly}(\text{enc}(f))} \cdot \|a\|^d$ and thus we get $|f(a)| \leq 2^{\text{poly}(\text{enc}(f), \text{enc}(P))}$.
- We have $\|b\| \leq 2^{\text{poly}(\text{enc}(P))}$.

Putting this all together, we get that for all $x \in P$ we have

$$\|Ux\| > 2^{\text{poly}(\text{enc}(f), \text{enc}(P))} \quad \text{or} \quad |\langle w, x \rangle| > 2^{\text{poly}(\text{enc}(f), \text{enc}(P))} \implies f(x) > f(a). \quad (23)$$

To complete the proof, it remains to connect $\|Ux\|$ and $|\langle w, x \rangle|$ to $\|x_{U \oplus W}\|$. Note that we have

$$x_{U \oplus W} = \sum_{i=1}^k \frac{(Ux)_i}{\|U_i\|^2} U_i + \frac{\langle w, x \rangle}{\|w\|^2} w,$$

where U_i are the rows of U (that are orthogonal and span \mathcal{U}). Thus, we have

$$\|x_{U \oplus W}\| \leq \sum_{i=1}^k \frac{|(Ux)_i|}{\|U_i\|} + \frac{|\langle w, x \rangle|}{\|w\|}.$$

First, since $\text{bl}(U) \leq \text{poly}(\text{enc}(f))$ by [Theorem 1.3](#) we have $\|U_i\| \geq 2^{-\text{poly}(\text{enc}(f))}$. Furthermore, we have that $\sum_{i=1}^k |(Ux)_i| = \|Ux\|_1 \leq \sqrt{k} \|Ux\|$. Thus, we get that

$$\sum_{i=1}^k \frac{|(Ux)_i|}{\|U_i\|} \leq 2^{\text{poly}(\text{enc}(f))} \|Ux\|.$$

Second, by [Theorem 1.3](#), we have $\text{bl}(w) \leq \text{poly}(\text{enc}(f))$ and thus $\|w\| \geq 2^{-\text{poly}(\text{enc}(f))}$. Together, this gives

$$\|x_{U \oplus W}\| \leq 2^{\text{poly}(\text{enc}(f))} (\|Ux\|_2 + |\langle w, x \rangle|)$$

and hence, by (23), we can conclude that, for all $x \in P$,

$$\|x_{U \oplus W}\| > 2^{\text{poly}(\text{enc}(f), \text{enc}(P))} \implies f(x) > f(a).$$

□

5.3 Finding a minimizer of small norm: Proof of Lemma 5.3

It remains to prove Lemma 5.3, which shows that we can lift a point $x^* \in P$ with small norm in some subspace \mathcal{Z} to a point $x' \in P$, whose norm in the entire space is small.

Proof of Lemma 5.3. Consider the polyhedron

$$\begin{aligned} P' &= \{x \in \mathbb{R}^n : Ax \leq b, \langle x, x^i \rangle = \langle x^*, x^i \rangle \forall 1 \leq i \leq k\} \\ &= \{x \in \mathbb{R}^n : A'x \leq b'\} \end{aligned}$$

for

$$A' = \begin{bmatrix} A \\ x^{1^\top} \\ -x^{1^\top} \\ \vdots \\ x^{k^\top} \\ -x^{k^\top} \end{bmatrix} \quad \text{and} \quad b' = \begin{bmatrix} b \\ \langle x^*, x^1 \rangle \\ -\langle x^*, x^1 \rangle \\ \vdots \\ \langle x^*, x^k \rangle \\ -\langle x^*, x^k \rangle \end{bmatrix}.$$

Note that $x' \in P'$ if and only if $x' \in P$ and $x_{\mathcal{Z}}^* = x'_{\mathcal{Z}}$. Thus, it remains to show that there is a $x' \in P'$ with $\|x'\| \leq \text{poly}(\|x_{\mathcal{Z}}^*\|, 2^{\text{poly}(\text{enc}(P), \text{bl}(\mathcal{Z}))})$.

Notice that we cannot apply Proposition 3.1 immediately since the vector b' might not be rational. Instead, we want to pick a vertex of P' and argue that it has small norm (even though it might also not be rational). However, it could be that P' has no vertices. Namely, P' has no vertices if and only if $\text{rank}(A') < n$ or in other words $\ker(A') \neq \{0\}$ [Sch86, section 8.5]. Let y^1, \dots, y^ℓ be a basis for $\ker(A')$. We have $\ell \leq n$ and since the y^j are solutions to the linear system $A'y = 0$, they satisfy $\text{bl}(y^j) \leq \text{poly}(\text{bl}(A), \text{bl}(x^1), \dots, \text{bl}(x^k)) \leq \text{poly}(\text{bl}(A), \text{bl}(\mathcal{Z}))$ [Sch86, Corollary 3.2d]. Consider the polyhedron

$$\begin{aligned} P'' &= \{x \in \mathbb{R}^n : Ax \leq b, \langle x, x^i \rangle = \langle x^*, x^i \rangle \forall 1 \leq i \leq k, \langle x, y^j \rangle = 0 \forall 1 \leq j \leq \ell\} \\ &= \{x \in \mathbb{R}^n : A''x \leq b''\}, \end{aligned}$$

where

$$A'' = \begin{bmatrix} A \\ x^{1^\top} \\ -x^{1^\top} \\ \vdots \\ x^{k^\top} \\ -x^{k^\top} \\ y^1 \\ -y^1 \\ \vdots \\ y^\ell \\ -y^\ell \end{bmatrix} \quad \text{and} \quad b'' = \begin{bmatrix} b \\ \langle x^*, x^1 \rangle \\ -\langle x^*, x^1 \rangle \\ \vdots \\ \langle x^*, x^k \rangle \\ -\langle x^*, x^k \rangle \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

We claim that the projection $x^{*'} := x^*_{\text{span}\{y^1, \dots, y^\ell\}^\perp}$ of x^* to $\text{span}\{y^1, \dots, y^\ell\}^\perp$ is in P'' , i.e. that P'' is feasible. Since $x^* \in P'$ and $y^1, \dots, y^\ell \in \ker(A')$, we have $A'x^{*'} \leq b'$. Clearly, also $\langle x^{*'}, y^j \rangle = 0$, so we indeed have $x^{*'} \in P''$. Furthermore, since $P'' \subseteq P'$, it is sufficient to find a point $x' \in P''$ that satisfies $\|x'\| \leq \text{poly}(\|x_{\mathcal{Z}}^*\|, 2^{\text{poly}(\text{enc}(P), \text{bl}(\mathcal{Z}))})$.

Note that since the y^j generate $\ker(A')$, we have $\ker(A'') = \{0\}$. Thus, $\text{rank}(A'') = n$ and P'' has a vertex x' . This vertex is the solution to a subsystem

$$\hat{A}x' = \hat{b},$$

where \hat{A} contains n linearly independent rows of A'' and \hat{b} contains the corresponding elements of b'' (see e.g. [Sch86, equation (23)]). This means that we have

$$x' = \hat{A}^{-1}\hat{b}$$

and thus also

$$\|x'\| \leq \|\hat{A}^{-1}\| \|\hat{b}\|.$$

Note that $\text{bl}(\hat{A}) \leq \text{bl}(A'') \leq \text{poly}(\text{bl}(A), \text{bl}(\mathcal{Z}))$. We also have $\text{bl}(\hat{A}^{-1}) \leq \text{poly}(\text{bl}(\hat{A}))$ [Sch86, Corollary 3.2a]. Thus, we get

$$\|\hat{A}^{-1}\| \leq \sqrt{n} \|\hat{A}^{-1}\|_F \leq 2^{\text{poly}(\text{bl}(\hat{A}))} \leq 2^{\text{poly}(\text{bl}(A), \text{bl}(\mathcal{Z}))}.$$

For this, note that $n \leq \text{bl}(\hat{A})$. Furthermore, we have

$$\|\hat{b}\|^2 \leq \|b\|^2 + \sum_{i=1}^k \langle x^*, x^i \rangle^2 \leq \|b\|^2 + \sum_{i=1}^k \|x^*\| \|x^i\| \leq \text{poly}(\|x^*\|, 2^{\text{poly}(\text{bl}(b), \text{bl}(\mathcal{Z}))}).$$

Combining these, we get

$$\|x'\| \leq \text{poly}(\|x^*\|, 2^{\text{poly}(\text{enc}(P), \text{bl}(\mathcal{Z}))}),$$

which completes the proof. \square

Acknowledgments

We thank Amir Ali Ahmadi for bringing the problem of efficient minimization of convex polynomials to our attention. We thank Rares-Darius Buhai, Jarosław Błasiok, Hongjie Chen, Jingqiu Ding and Yiding Hua for valuable discussions at an early stage of the project. We thank Abraar Chaudhry for a discussion on quadratic lower bounds for convex polynomials. We thank Kevin Shu for a discussion on polynomials with vanishing Hessian determinant, and in particular for pointing us to the Schwartz-Zippel lemma.

This research was supported by the Swiss National Science Foundation (SNSF), grant no. 10004947.

A Linear algebra in the bit model

The following proposition shows that one can compute an orthogonal basis of small bit size for the inverse image of a (one-dimensional) subspace under a rational matrix, and for its complement. This captures the linear algebra required in our proof of [Theorem 1.3](#).

Proposition A.1. *Let $A \in \mathbb{Q}^{M \times N}$, $M \geq N$, and $b \in \mathbb{Q}^M$. Define the subspaces*

$$\mathcal{L} = \{v \in \mathbb{R}^N : A \cdot v \in \text{span}\{b\}\}, \quad \mathcal{U} = \mathcal{L}^\perp.$$

We then have $\mathcal{L} = \ker(A) \oplus \text{span}\{w\}$, where $w \in \ker(A)^\perp$ is either zero or $A \cdot w = b$. We can compute the vector w ; an orthogonal basis v_1, \dots, v_ℓ for $\ker(A)$; an orthogonal basis u_1, \dots, u_k for \mathcal{U} in polynomial time in $\text{bl}(A) + \text{bl}(b)$. In particular, these have polynomial bit size in $\text{bl}(A) + \text{bl}(b)$.

[Proposition A.1](#) follows from standard results on linear system solving and Gram-Schmidt orthogonalization in the bit model. We include a brief proof for completeness. We rely on the *Hermite normal form* of an integer matrix, which is an integer analog of the echelon form. The same conclusion could be reached using a polynomial-time algorithm to compute the echelon form (over \mathbb{Q}), as described, e.g., in [\[Sch86\]](#).

Lemma A.2 (Hermite normal form). *Let $A \in \mathbb{Z}^{M \times N}$, $M \geq N$ be of column rank R . We can compute in time polynomial in $\text{bl}(A)$ a decomposition*

$$A \cdot [U | K] = [H | 0],$$

where $U \in \mathbb{Z}^{N \times R}$, $K \in \mathbb{Z}^{N \times (N-R)}$ satisfy $[U | K] \in \text{GL}_N(\mathbb{Z})$, and $H \in \mathbb{Z}^{M \times R}$ is lower triangular. In particular these matrices have bit length polynomial in $\text{bl}(A)$.

Proof. Apply [\[Sto00, Proposition 6.3\]](#) and [\[Sto00, Proposition 6.6\]](#) to A^\top . \square

Lemma A.3 (Gram-Schmidt orthogonalization). *Let $v_1, \dots, v_k \in \mathbb{Q}^N$. Then, there exist pairwise orthogonal vectors $u_1, \dots, u_k \in \mathbb{Q}^N$, such that*

$$\text{span}\{u_1, \dots, u_i\} = \text{span}\{v_1, \dots, v_i\} \quad (\forall i \leq k).$$

Moreover, these vectors can be computed in polynomial time in N, k and $B := \max_i \text{bl}(v_i)$.

Proof. The only thing to check is that the recursively defined coefficients that appear in the Gram-Schmidt procedure do not grow too large. Such an analysis is carried out, e.g., in [\[LLL82\]](#) or [\[EKM96\]](#). \square

Proof of [Proposition A.1](#). After multiplication by an integer of bit length $\text{poly}(\text{bl}(A))$, we may assume without loss of generality that A is an integer matrix. Let R be the column rank of A . Compute the Hermite normal form $A \cdot [U | K] = [H | 0]$ as in [Lemma A.2](#). Note that the columns of $[U | K]$ form a basis for \mathbb{R}^N , and that the columns of K form a basis for $\ker(A)$. The lower triangular system $H \cdot y = b$ can be solved via forward substitution; this either yields a solution $y \in \mathbb{Q}^R$ (of bit length at most $\text{poly}(\text{bl}(H))$), or shows that no solution exists at all. In the former case, $\hat{w} := U \cdot y$ satisfies $A \cdot \hat{w} = b$ (and $\text{bl}(\hat{w}) \leq \text{poly}(\text{bl}(U), \text{bl}(H))$). It remains to orthogonalize: Apply [Lemma A.3](#) to the columns of K , \hat{w} , and the columns of U , in that order. This yields orthogonal vectors $v_1, \dots, v_{N-R}, w, u_1, \dots, u_R$ (of appropriately bounded bit length). Note that v_1, \dots, v_{N-R} span $\ker(A)$, that v_1, \dots, v_{N-R}, w span \mathcal{L} , and that u_1, \dots, u_R span \mathcal{L}^\perp . Note that either w , or exactly one of the u_i will be zero. Discarding that vector completes the proof. \square

B A polynomial with identically vanishing hessian

We provide an example adapted from [\[GN76, GR09\]](#) of a polynomial whose Hessian is everywhere singular, but which does not have a direction of linearity. Let $p(x) = x_1x_4^2 + 2x_2x_4x_5 + x_3x_5^2$. Then,

$$\nabla^2 p(x) = \begin{pmatrix} 0 & 0 & 0 & 2x_4 & 0 \\ 0 & 0 & 0 & 2x_5 & 2x_4 \\ 0 & 0 & 0 & 0 & 2x_5 \\ 2x_4 & 2x_5 & 0 & 2x_1 & 2x_2 \\ 0 & 2x_4 & 2x_5 & 2x_2 & 2x_3 \end{pmatrix}.$$

One may verify that the Hessian of p is singular for every $x \in \mathbb{R}^n$. On the other hand, we have

$$\nabla p(x) = (x_4^2, 2x_4x_5, x_5^2, 2x_1x_4 + 2x_2x_5, 2x_2x_4 + 2x_3x_5),$$

and so p has no direction of linearity (its directional derivatives are not affinely dependent). But, we do have an algebraic dependency of degree 2, namely

$$\frac{\partial p}{\partial x_1} \cdot \frac{\partial p}{\partial x_3} = x_4^2 \cdot x_5^2 = \frac{1}{4}(2x_4x_5)^2 = \frac{1}{4} \left(\frac{\partial p}{\partial x_2} \right)^2.$$

C Witnesses for polynomial programs

In this appendix, we give a rational univariate convex polynomial of degree 4 that has an irrational minimizer (see [Example C.1](#)). This shows that for the decision problem [\(D\)](#) for convex polynomials of degree at least 4, x^* is not always a compact witness. We also give a rational univariate convex polynomial of degree 6 that has minimum value 0, but attains this at exactly one *irrational* point (see [Example C.2](#)). This shows that for convex polynomials of degree at least 6, for some polynomials there is no compact witness at all. Finally, we give a proof that such a (univariate) polynomial cannot exist for degree 4. We show that for a rational univariate convex polynomial of degree 4, if the minimum value is rational, then also the minimizer is rational (see [Lemma C.3](#)). Thus, for convex polynomials of degree 4, it could be that there is always a compact witness even though the minimizer might not be: If the minimizer is irrational, then the set $\{x : f(x) \leq 0\}$ does not contain just the minimizer since the minimum value is not 0 (it is irrational). Thus, there will at least be a rational point in this set and it is unclear whether there might always be a compact witness.

Example C.1. Consider the polynomial

$$f(x) = x^4 + x.$$

The second derivative is positive for all x , thus f is convex. The (unique) minimizer of f is the point x^* that satisfies $4x^{*3} + 1 = 0$, i.e., $x^* = -1/\sqrt[3]{4}$, which is irrational.

Example C.2. Consider the polynomial

$$f(x) = (x^3 + x + 1)^2 = x^6 + 2x^4 + 2x^3 + x^2 + 2x + 1.$$

The second derivative of f is given by

$$f''(x) = 30x^4 + 24x^2 + 12x + 2 = 30x^4 + 6x^2 + 2 \cdot (3x + 1)^2 \geq 0 \quad (\forall x \in \mathbb{R})$$

and hence f is a convex polynomial.

We now want to determine the minimal value and the minimizer of f . Note that f is non-negative. To determine the minimum and the minimizer, we want to examine the term $g(x) = x^3 + x + 1$. As a cubic polynomial it has at least one real root. Thus, the minimum value of f is 0 and the minimizer(s) of f are exactly the root(s) of g .

Cardano's formula states that a cubic polynomial $x^3 + px + q$ with $\frac{q^2}{4} + \frac{p^3}{27} > 0$ has in fact exactly one real root. Furthermore, by the rational root theorem, if there is a rational root it is an integer divisor of q . This implies that g (for which $p = q = 1$) has exactly one real root x^* . Since ± 1 (the

only divisors of $q = 1$) are not roots, this root is not rational, i.e. $x^* \in \mathbb{R} \setminus \mathbb{Q}$. This x^* is the (unique) minimizer of f . In fact, we can even compute x^* explicitly:

$$x^* = \sqrt[3]{-\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{27}}} + \sqrt[3]{-\frac{1}{2} - \sqrt{\frac{1}{4} + \frac{1}{27}}}.$$

Hence, f is a convex polynomial with minimum value 0, but this is attained at exactly one (irrational) point x^* .

Lemma C.3. *Let $f \in \mathbb{Q}[x]$ be a convex (univariate) polynomial of degree 4. Assume the minimum value of f is rational. Then also the minimizer of f is rational.*

Proof. Let $x^* \in \mathbb{R}$ be the minimizer of f . By convexity, x^* is the unique minimizer and the unique critical point of f . Consider the two polynomials (recall that by assumption $f(x^*) \in \mathbb{Q}$)

$$g_1(x) = f(x) - f(x^*) \in \mathbb{Q}[x] \quad \text{and} \quad g_2(x) = f'(x) \in \mathbb{Q}[x].$$

We have that x^* is the unique (real) root of these two polynomials. Consider $m \in \mathbb{Q}[x]$ the monic polynomial of minimal degree that has x^* as a root. Since x^* is a root of g_1 and g_2 , we have $m \mid g_1$ and $m \mid g_2$.⁷ Thus, there are polynomials $h_1, h_2 \in \mathbb{Q}[x]$ such that $g_1 = m \cdot h_1$ and $g_2 = m \cdot h_2$.

Note that $\deg(h_1) = \deg(h_2) + 1$. Thus one of $\deg(h_1)$ and $\deg(h_2)$ is odd. Let $i \in \{1, 2\}$ be the index such that $\deg(h_i)$ is odd. Then h_i has a real root. Since x^* is the unique real root of g_i , this root needs to be x^* . But then we need to have $m \mid h_i$ and there is a polynomial \tilde{h}_i such that $g_i = m^2 \cdot \tilde{h}_i$. This implies $4 \geq \deg(g_i) \geq 2 \deg(m)$ and thus $\deg(m) \leq 2$.

If $\deg(m) = 1$, then $m(x) = x + a_0$ and thus $x^* = -a_0 \in \mathbb{Q}$. If $\deg(m) = 2$, then $m(x) = x^2 + a_1x + a_0$. The roots of this polynomial are $\frac{-a_1 \pm \sqrt{a_1^2 - 4a_0}}{2}$. Since x^* is the unique (real) root of g_1 (and thus of m), we need to have that $a_1^2 - 4a_0 = 0$ and thus $x^* = \frac{-a_1}{2} \in \mathbb{Q}$, which completes the proof. \square

D Ellipsoid method

In this appendix, we give a proof of [Proposition 3.2](#) about minimizing a convex polynomial over a polyhedron.

Proposition 3.2. *Let $R > 0$, $\varepsilon > 0$. Let $f \in \mathbb{Q}[x]$ be a polynomial. Let $P = \{Ax \leq b\}$ be a (nonempty) polyhedron. Then, there exists an algorithm that outputs $\tilde{x} \in P$ such that $f(\tilde{x}) \leq \min_{x \in P \cap B_R(0)} f(x) + \varepsilon$ in time $\text{poly}(\text{enc}(f), \text{enc}(P), \log(R), \log(1/\varepsilon))$.*

All the ideas presented in this appendix are standard results on the ellipsoid method and can for example be found in [[GLS93](#), [Vis21](#)]. We combine these results to get the exact statement of [Proposition 3.2](#). We include a full proof for completeness and in order to carry out the bit complexity analysis.

⁷Formally, consider the field extension $\mathbb{Q}(x^*)/\mathbb{Q}$. Since $g_1(x^*) = 0$ and $g_1 \in \mathbb{Q}[x]$, x^* is algebraic over \mathbb{Q} . Then, m is the minimal polynomial of x^* over \mathbb{Q} , i.e. the unique monic irreducible polynomial $m \in \mathbb{Q}[x]$ that has x^* as a root [[Rot15](#), Theorem A-3.87]. Furthermore, m generates the ideal of all polynomials in $\mathbb{Q}[x]$ that have x^* as a root. Thus, m divides any polynomial g that has x^* as a root [[Rot15](#), Proof of Theorem A-3.87].

D.1 Additional preliminaries on the ellipsoid method

In this section, we discuss the necessary preliminaries to apply the ellipsoid method to our case. The section is adapted from different definitions and theorems from [GLS93]. We first define what a strong separation oracle is.

Definition D.1 (Strong separation oracle). A strong separation oracle for a convex set $K \subseteq \mathbb{R}^n$ is an oracle that, given as input $y \in \mathbb{Q}^n$ either asserts $y \in K$ or finds a separating hyperplane $c \in \mathbb{Q}^n$ such that $\max_{x \in K} c^\top x < c^\top y$.

Remark D.2. We always assume that given an input $y \in \mathbb{Q}^n$, the bit length of the output of the oracle is at most polynomial in $\text{bl}(y)$ for a fixed polynomial. See also [GLS93, Assumption 1.2.1].

For a polyhedron, there is always a strong separation oracle that runs in polynomial time in the bit length of the input and in $\text{enc}(P)$.

Lemma D.3 ([GLS93, Example 2.16]). *For a polyhedron $P = \{Ax \leq b\}$ there is a strong separation oracle that runs on an input of bit length k in time polynomial in k and $\text{enc}(P)$.*

The idea for this separation oracle is to check for all constraints of P whether y satisfies them. If all constraints are satisfied, then $y \in P$. Otherwise, any violated constraint is a separating hyperplane. We now state a result on how we can apply the ellipsoid method to solve feasibility problems given strong separation oracles.

Proposition D.4 ([GLS93, Theorem 3.2.1]). *There is an algorithm with the following guarantees: Given as input $r > 0$, $R > 0$ and a strong separation oracle to a closed convex set $K \subseteq B_R(0)$, the algorithm outputs one of the following:*

1. a vector $a \in K$ ⁸
2. an ellipsoid E such that $K \subseteq E$ and $\text{vol}(E) \leq r$.

In particular, if $\text{vol}(K) > r$, then the algorithm outputs a vector in K . The number of oracle calls and the runtime⁹ of the algorithm are polynomial in n , $\log(1/r)$ and $\log(R)$.

Remark D.5. When we apply Proposition D.4, the oracle runs in polynomial time in $\text{enc}(f)$, $\text{enc}(P)$, k on an input of bit length k . Therefore, the total runtime of the algorithm will be polynomial in $\text{enc}(f)$, $\text{enc}(P)$, $\log(1/r)$ and $\log(R)$.

To apply the statement above, we also need the following two statements that will allow us to reduce to the case where the polynomial has volume (i.e. the full-dimensional case).

Proposition D.6 ([GLS93, Theorems 6.4.9 and 6.5.5]). *There is an algorithm with the following guarantees: Let $P = \{Ax \leq b\} \subseteq \mathbb{R}^n$ be a polyhedron and let φ be a bound on the maximum bit length of any constraint of P . Given as input n , φ and a strong separation oracle to P , the algorithms output affinely independent vectors v_1, \dots, v_k such that $\text{aff}(P) = \text{aff}(\{v_1, \dots, v_k\})$. The number of oracle calls and the runtime of the algorithm is polynomial in n and φ .*

⁸We note that the algorithm in [GLS93, Theorem 3.2.1], if given a strong separation oracle (instead of a weak one), in fact outputs $y \in K$ (as opposed to just $d(y, K) \leq r$).

⁹Here, the runtime does not include the oracle call, but it does include the time needed to write down the input for the oracle and read the output of the oracle.

Proposition D.7 ([GLS93, Lemmas 3.1.33 and 3.1.35]). *Let $P = \{Ax \leq b\}$ be a polyhedron. If P is full-dimensional, then*

$$\text{vol}(P) \geq 2^{-\text{poly}(\text{enc}(P))}.$$

In fact, there exist $v_1, \dots, v_{n+1} \in P$ with $\|v_i\| \leq 2^{\text{poly}(\text{enc}(P))}$ such that

$$\text{vol}(\text{conv}(\{v_1, \dots, v_{n+1}\})) \geq 2^{-\text{poly}(\text{enc}(P))}.$$

D.2 Proof of Proposition 3.2

In order to prove [Proposition 3.2](#) we want to reduce to the full-dimensional case, for which we then can apply [Proposition D.4](#). We first prove the full-dimensional case of [Proposition 3.2](#).

Proof of Proposition 3.2 (full-dimensional case). Consider the set $F_\tau := \{x \in \mathbb{R}^n : f(x) \leq \tau\}$. Our goal is to apply [Proposition D.4](#) to the sets $P \cap B_R(0) \cap F_\tau$ for different values of τ . Thus, we need to give a strong separation oracle for this set and find a value for r such that $\text{vol}(P \cap B_R(0) \cap F_\tau) \geq r$.

Strong separation oracle. By [Lemma D.3](#) we have a strong separation oracle for P . Clearly, there is also a strong separation oracle for $B_R(0)$ (check if $\|y\| \leq R$; if not, y is a separating hyperplane) and thus it remains to argue that we can get a strong separation oracle for F_τ . Given $y \in \mathbb{Q}^n$, we can determine whether $f(y) \leq \tau$ and thus whether $y \in F_\tau$. If $y \notin F_\tau$, then the gradient $\nabla f(y)$ is a separating hyperplane. This is true since $f(x) - f(y) \geq \nabla f(y)^\top (x - y)$ for all $x \in \mathbb{R}^n$ and thus for $x \in F_\tau$ we have $f(x) \leq \tau < f(y)$ and thus $\max_{x \in F_\tau} \nabla f(y)^\top x < \nabla f(y)^\top y$. Thus, we have strong separation oracles for P , $B_R(0)$ and F_τ and thus also a strong separation oracle for $P \cap B_R(0) \cap F_\tau$. This separation oracle runs in time $\text{poly}(\text{enc}(f), \text{enc}(P), \text{bl}(\tau), \log(R), \text{bl}(y))$ if given as input a point $y \in \mathbb{Q}^n$.

Bound on volume. Next, we want to bound $\text{vol}(P \cap B_R(0) \cap F_\tau)$. Let x^* be a minimizer of f on $P \cap B_R(0)$ and let $f^* = f(x^*)$. Note that f is Lipschitz on $B_R(0)$ with Lipschitz constant $L = 2^{\text{poly}(\text{enc}(f), \log(R))}$. If $\tau = f^* + c$, then we want to argue that we can lower bound the volume in terms of c . Define $c' = c/L$. Since P is full-dimensional, by [Proposition D.7](#), there are $v_1, \dots, v_{n+1} \in P$ with $\|v_i\| \leq 2^{\text{poly}(\text{enc}(P))}$ such that

$$\text{vol}(\text{conv}(\{v_1, \dots, v_{n+1}\})) \geq 2^{-\text{poly}(\text{enc}(P))}.$$

Without loss of generality, we assume $R \geq \|v_i\|$ for all i such that $v_i \in P \cap B_R(0)$.¹⁰ Let $d_i = \|x^* - v_i\|$ and let $d_{\max} = \max_{i \in \{1, \dots, n+1\}} d_i$. We have, for some $i \in \{1, \dots, n\}$,

$$d_{\max} = \|x^* - v_i\| \leq \|x^*\| + \|v_i\| \leq R + 2^{\text{poly}(\text{enc}(P))}.$$

If $d_{\max} \leq c'$, then, by Lipschitzness of f , we have

$$f(v_i) \leq f(x^*) + L\|v_i - x^*\| \leq f^* + L \cdot c' = f + c = \tau.$$

Hence, $v_i \in P \cap B_R(0) \cap F_\tau$ and thus

$$\text{vol}(P \cap B_R(0) \cap F_\tau) \geq \text{vol}(\text{conv}(\{v_1, \dots, v_{n+1}\})) \geq 2^{-\text{poly}(\text{enc}(P))}.$$

¹⁰We can do this because $\|v_i\| \leq 2^{\text{poly}(\text{enc}(P))}$. Thus, if $R < 2^{\text{poly}(\text{enc}(P))}$, we can replace R by $R' = 2^{\text{poly}(\text{enc}(P))}$, which satisfies $\log(R') = \text{poly}(\text{enc}(f))$.

If $d_{\max} > c'$, consider the points $\hat{v}_i := x^* + \frac{c'}{d_{\max}}(v_i - x^*) \in P \cap B_R(0)$ (since $x^*, v_i \in P \cap B_R(0)$). Then, we have, again by Lipschitzness of f ,

$$f(\hat{v}_i) \leq f(x^*) + L \left\| \frac{c'}{d_{\max}}(v_i - x^*) \right\| \leq f^* + L \cdot c' \cdot \frac{d_i}{d_{\max}} \leq f^* + L \cdot c' = \tau$$

and hence $\hat{v}_i \in F_\tau$. Thus, $\text{conv}(\{\hat{v}_1, \dots, \hat{v}_{n+1}\}) \subseteq P \cap B_R(0) \cap F_\tau$. We have

$$\begin{aligned} \text{vol}(\text{conv}(\{\hat{v}_1, \dots, \hat{v}_{n+1}\})) &= \text{vol}\left(\text{conv}\left(\left\{x^* + \frac{c'}{d_{\max}}(v_i - x^*): i \in \{1, \dots, n+1\}\right\}\right)\right) \\ &= \left(\frac{c'}{d_{\max}}\right)^n \text{vol}(\text{conv}(\{v_i : i \in \{1, \dots, n+1\}\})) \\ &\geq \left(\frac{c'}{d_{\max}}\right)^n \cdot 2^{-\text{poly}(\text{enc}(P))}. \end{aligned}$$

Thus, we get that

$$\text{vol}(P \cap B_R(0) \cap F_\tau) \geq \text{vol}(\text{conv}(\{\hat{v}_1, \dots, \hat{v}_{n+1}\})) \geq \left(R + 2^{\text{poly}(\text{enc}(P))}\right)^{-n} \cdot (c')^n \cdot 2^{-\text{poly}(\text{enc}(P))}.$$

Thus, by choosing

$$r = \min \left\{ 2^{-\text{poly}(\text{enc}(P))}, \left(R + 2^{\text{poly}(\text{enc}(P))}\right)^{-n} \cdot \left(\frac{\varepsilon}{2L}\right)^n \cdot 2^{-\text{poly}(\text{enc}(P))} \right\}$$

we get that

$$\tau > f_{\min} + \frac{\varepsilon}{2} \implies \text{vol}(P \cap B_R(0) \cap F_\tau) > r, \quad (24)$$

since if $\tau > f_{\min} + \frac{\varepsilon}{2}$, then $c > \frac{\varepsilon}{2}$ and thus $\text{vol}(P \cap B_R(0) \cap F_\tau) > r$. Note that, using that the Lipschitz constant L satisfies $L = 2^{\text{poly}(\text{enc}(f), \log(R))}$, we get

$$\log(1/r) = \text{poly}(n, \text{enc}(P), \log(R), \log(1/\varepsilon)).$$

Algorithm. Using binary search, we want to use this to get a point $\tilde{x} \in P$ with the guarantee that $f(\tilde{x}) \leq \min_{x \in P \cap B_R(0)} f(x) + \varepsilon$. Consider [Algorithm 2](#). This algorithm and the analysis are a standard way to move from using the ellipsoid method to solve a feasibility problem to solving an optimization problem. It can for example be found in [\[Vis21, Chapter 13\]](#). We include it here for completeness and in order to carry out the bit complexity arguments in detail, which is needed for our result.

We claim that the algorithm satisfies the following two invariants:

- For τ_r , we always have that the output of the algorithm from [Proposition D.4](#) belongs to case (i) (given r, R and a separation oracle to $P \cap B_R(0) \cap F_{\tau_r}$ as input).
- For τ_ℓ , we always have $\tau_\ell \leq f^* + \frac{\varepsilon}{2}$.

This is true in the beginning because by Lipschitzness of f and since $f(0) \leq 2^{\text{poly}(\text{enc}(f), \log(R))}$, we know that on $B_R(0)$ the value of f is in $[-2^{\text{poly}(\text{enc}(f), \log(R))}, 2^{\text{poly}(\text{enc}(f), \log(R))}]$. Thus, for the first choice of τ_r , we have $F_{\tau_r} \supseteq B_R(0)$ and hence $\text{vol}(P \cap B_R(0) \cap F_{\tau_r}) = \text{vol}(P \cap B_R(0)) > r$ (using (24) for $\tau = \infty$). For the first choice of τ_ℓ , we have $\tau_\ell \leq f^*$. Furthermore, this stays true during the

Algorithm 2

- 1: **Input:** A polynomial f , a matrix $A \in \mathbb{Q}^{m \times n}$, a vector $b \in \mathbb{Q}^m$, a bound R , an error parameter ε
- 2: **Output:** A point $\tilde{x} \in P$ with $f(\tilde{x}) \leq \min_{x \in P \cap B_R(0)} f(x) + \varepsilon$.
- 3: Let $\tau_\ell = -2^{\text{poly}(\text{enc}(f), \log(R))}$ and $\tau_r = 2^{\text{poly}(\text{enc}(f), \log(R))}$.
- 4: **while** $\tau_r - \tau_\ell \geq \frac{\varepsilon}{2}$ **do**
- 5: $\tau = \frac{\tau_r + \tau_\ell}{2}$
- 6: Run the algorithm from [Proposition D.4](#) with input r, R and a strong separation oracle for the set $P \cap B_R(0) \cap F_\tau$.
- 7: **if** output is $x \in P \cap B_R(0) \cap F_\tau$ (i.e. we are in case (i)) **then**
- 8: $\tau_r = \frac{\tau_r + \tau_\ell}{2}$
- 9: **else** (i.e. we are in case (ii))
- 10: $\tau_\ell = \frac{\tau_r + \tau_\ell}{2}$
- 11: **end if**
- 12: **end while**
- 13: Run the algorithm from [Proposition D.4](#) with input r, R and a strong separation oracle for the set $P \cap B_R(0) \cap F_{\tau_r}$.
- 14: **return** the point \tilde{x} that is output by this algorithm

whole algorithm because we only update τ_r in case (i) of the algorithm from [Proposition D.4](#), i.e. exactly when the invariant stays true. If we update τ_ℓ , we are in case (ii) of the algorithm from [Proposition D.4](#). Then we need to have $\text{vol}(P \cap B_R(0) \cap \tau_\ell) < r$ (for the updated τ_ℓ) and hence $\tau_\ell \leq f_{\min} + \frac{\varepsilon}{2}$ by (24).

Thus, the algorithm is well-defined (i.e., after the while-loop the algorithm from [Proposition D.4](#) does in fact output a point \tilde{x}). This point is in $P \cap B_R(0) \cap F_{\tau_r}$ for the final value of τ_r . Hence, $f(\tilde{x}) \leq f^* + \tau_r$. Furthermore, we have $\tau_r \leq \tau_\ell + \frac{\varepsilon}{2} \leq f^* + \varepsilon$ and thus we output a point $\tilde{x} \in P \cap B_R(0)$ with

$$f(\tilde{x}) \leq f^* + \varepsilon = \min_{x \in P \cap B_R(0)} f(x) + \varepsilon.$$

It remains to argue the runtime of the algorithm. We have $\text{poly}(\text{enc}(f), \log(R), \log(1/\varepsilon))$ iterations of the while-loop (we start with $\tau_r - \tau_\ell = 2^{\text{poly}(\text{enc}(f), \log(R))}$, end with $\tau_r - \tau_\ell \leq \frac{\varepsilon}{2}$ and we half this distance in every step). This then also implies that $\text{bl}(\tau_r), \text{bl}(\tau_\ell) \leq \text{poly}(\text{enc}(f), \log(R), \log(1/\varepsilon))$. So, the separation oracle for $P \cap B_R(0) \cap F_{\frac{\tau_r+\tau_\ell}{2}}$ needs time $\text{poly}(\text{enc}(f), \text{enc}(P), \log(R), \log(1/\varepsilon), k)$ on an input of bit length k . Thus, one execution of the algorithm from [Proposition D.4](#) takes time

$$\text{poly}(n, \log(R), \log(1/r), \text{enc}(f), \text{enc}(P), \log(1/\varepsilon)).$$

Using $\log(1/r) = \text{poly}(n, \text{enc}(P), \log(R), \log(1/\varepsilon))$ and $n \leq \text{enc}(f)$, this runtime is

$$\text{poly}(\text{enc}(f), \text{enc}(P), \log(R), \log(1/\varepsilon)).$$

Thus, since we have $\text{poly}(\text{enc}(f), \log(R), \log(1/\varepsilon))$ iterations the overall runtime of the algorithm is also

$$\text{poly}(\text{enc}(f), \text{enc}(P), \log(R), \log(1/\varepsilon)),$$

which completes the proof of [Proposition 3.2](#) for the full-dimensional case. \square

Finally, we want to prove [Proposition 3.2](#) for a general (not necessarily full-dimensional) polyhedron by reducing this to the full-dimensional case.

Proof of Proposition 3.2 (general case). By [Proposition D.6](#), we can compute vectors v_1, \dots, v_k such that $\text{aff}(P) = \text{aff}(\{v_1, \dots, v_k\})$. Then, we also have that

$$\text{aff}(P) = v_1 \oplus \text{span}\{v_2 - v_1, \dots, v_k - v_1\}.$$

By [Lemma A.3](#), we can replace the $v_2 - v_1, \dots, v_k - v_1$ by orthogonal vector $\hat{v}_2, \dots, \hat{v}_k$ (this is not strictly necessary but makes the following argument simpler). Define the following affine map

$$L : \mathbb{R}^{k-1} \rightarrow \text{aff}(P) \subseteq \mathbb{R}^n, \quad x' \mapsto v_1 + \sum_{i=2}^k x'_{i-1} \hat{v}_i.$$

This map is a bijection from \mathbb{R}^{k-1} to $\text{aff}(P)$. In fact, the inverse is

$$L^{-1} : \text{aff}(P) \rightarrow \mathbb{R}^{k-1}, \quad x \mapsto (\langle x - v_1, \hat{v}_2 \rangle / \|\hat{v}_2\|^2, \dots, \langle x - v_1, \hat{v}_k \rangle / \|\hat{v}_k\|^2). \quad (25)$$

We can write L as $L(x') = v_1 + Bx'$ for an appropriate matrix $B \in \mathbb{Q}^{n \times (k-1)}$. We define

$$P' = L^{-1}(P) = \{x' \in \mathbb{R}^{k-1} : v_1 + Bx' \in P\} = \{x' \in \mathbb{R}^{k-1} : (AB)x' \leq b - Av_1\} \subseteq \mathbb{R}^{k-1}$$

and

$$f' : \mathbb{R}^{k-1} \rightarrow \mathbb{R}, \quad x' \mapsto f'(x') = f(L(x')).$$

Note that P' is a polyhedron in \mathbb{R}^{k-1} and f' is a $(k-1)$ -variate polynomial.

On the polyhedron P , the algorithm from [Proposition D.6](#) runs in time $\text{poly}(\mathbf{enc}(P))$ (recall that by [Lemma D.3](#) the separation oracle for P runs in time $\text{poly}(k, \mathbf{enc}(P))$ on an input of size k). Also the orthogonalization then takes time $\text{poly}(\mathbf{enc}(P))$ by [Lemma A.3](#). Thus, we can bound the bit length of B as $\text{bl}(B) \leq \text{poly}(\mathbf{enc}(P))$ and thus also $\mathbf{enc}(P') \leq \text{poly}(\mathbf{enc}(P))$. Furthermore, this also implies that $\mathbf{enc}(f') \leq \text{poly}(\mathbf{enc}(f), \mathbf{enc}(P))$.

Since the v_1, \dots, v_k are affinely independent, P' is full-dimensional ($L^{-1}(v_1), \dots, L^{-1}(v_k)$ are affinely independent points in $\text{aff}(P')$ and thus by a dimension argument, $\text{aff}(P') = \mathbb{R}^{k-1}$). Let $R > 0$. Note that for $x \in P \cap B_R(0)$, using (25), we have

$$\|L^{-1}(x)\| \leq \sum_{i=2}^k \frac{|\langle x - v_1, \hat{v}_i \rangle|^2}{\|\hat{v}_i\|^2} \leq (R + \|v_1\|) \cdot \sum_{i=2}^k \frac{1}{\|\hat{v}_i\|}.$$

Since $\text{bl}(v_1), \text{bl}(\hat{v}_i) \leq \text{poly}(\mathbf{enc}(P))$, we have $\|v_1\|, \|\hat{v}_2\|^{-1}, \dots, \|\hat{v}_k\|^{-1} \leq 2^{\text{poly}(\mathbf{enc}(P))}$. Thus, by defining $R' = (R + 1) \cdot 2^{\text{poly}(\mathbf{enc}(P))}$ we get that $L^{-1}(P \cap B_R(0)) \subseteq P' \cap B_{R'}(0)$.

Thus, by the proof of the full-dimension case (applied with R' , ε , f' and P'), we can compute a point $\tilde{x}' \in P'$ with $f(\tilde{x}') \leq \min_{x' \in P' \cap B_{R'}(0)} f'(x') + \varepsilon$ in time

$$\text{poly}(\mathbf{enc}(f'), \mathbf{enc}(P'), \log(R'), \log(1/\varepsilon)) \leq \text{poly}(\mathbf{enc}(f), \mathbf{enc}(P), \log(R), \log(1/\varepsilon)).$$

Since $P \cap B_R(0) \subseteq L(P' \cap B_{R'}(0))$, we have $\min_{x' \in P' \cap B_{R'}(0)} f'(x') \leq \min_{x \in P \cap B_R(0)} f(x)$. Given a point \tilde{x}' as above, we thus get a point $\tilde{x} = L(\tilde{x}') \in P$ with

$$f(\tilde{x}) = f'(\tilde{x}') \leq \min_{x' \in P' \cap B_{R'}(0)} f'(x') + \varepsilon \leq \min_{x \in P \cap B_R(0)} f(x) + \varepsilon. \quad \square$$

References

- [ACZ24] Amir Ali Ahmadi, Abraar Chaudhry, and Jeffrey Zhang, *Higher-order Newton methods with polynomial work per iteration*, Advances in Mathematics **452** (2024), 109808. [5](#), [6](#), [8](#), [15](#)
- [AOPT11] Amir Ali Ahmadi, Alex Olshevsky, Pablo A. Parrilo, and John N. Tsitsiklis, *NP-hardness of deciding convexity of quartic polynomials and related problems*, Mathematical Programming **137** (2011), no. 1–2, 453–476. [2](#), [6](#)
- [AP13] Amir Ali Ahmadi and Pablo A. Parrilo, *A complete characterization of the gap between convexity and sos-convexity*, SIAM Journal on Optimization **23** (2013), no. 2, 811–833. [6](#)
- [AZ22a] Amir Ali Ahmadi and Jeffrey Zhang, *Complexity aspects of local minima and related notions*, Advances in Mathematics **397** (2022), 108119. [5](#)
- [AZ22b] ———, *On the complexity of finding a local minimizer of a quadratic function over a polytope*, Mathematical Programming **195** (2022), no. 1–2, 783–792. [5](#)
- [BMPV25] Alex Bortolotti, Monaldo Mastrolilli, Marilena Palomba, and Luis Felipe Vargas, *On the bit size of sum-of-squares proofs for symmetric formulations*, arXiv:2509.06928, 2025. [6](#)
- [BMV25] Alex Bortolotti, Monaldo Mastrolilli, and Luis Felipe Vargas, *On the degree automatability of sum-of-squares proofs*, arXiv:2504.17756, 2025. [6](#)
- [BS14] Boaz Barak and David Steurer, *Sum-of-squares proofs and the quest toward optimal algorithms*, Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV, Kyung Moon Sa, Seoul, 2014, pp. 509–533. [6](#)
- [BSS89] Lenore Blum, Mike Shub, and Steve Smale, *On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines*, Bulletin of the American Mathematical Society (N.S.) **21** (1989), no. 1, 1–46. [5](#)
- [CC60] Charles W. Clenshaw and Andrew R. Curtis, *A method for numerical integration on an automatic computer*, Numer. Math. **2** (1960), 197–205. [8](#)
- [Dan63] George B. Dantzig, *Linear programming and extensions*, Princeton University Press, Princeton, NJ, 1963. [1](#), [4](#), [5](#)
- [dKV16] Etienne de Klerk and Frank Vallentin, *On the Turing model complexity of interior point methods for semidefinite programming*, SIAM Journal on Optimization **26** (2016), no. 3, 1944–1961. [6](#)
- [Edm67] Jack Edmonds, *Systems of distinct representatives and linear algebra*, Journal of Research of the National Bureau of Standards – Section B Mathematics and Mathematical Physics **71B** (1967), no. 4, 241. [1](#), [4](#), [5](#)
- [EKM96] Úlfar Erlingsson, Erich Kaltofen, and David Musser, *Generic gram-schmidt orthogonalization by exact division*, Proceedings of the 1996 International Symposium on Symbolic and Algebraic Computation (New York, NY, USA), ISSAC ’96, Association for Computing Machinery, 1996, p. 275–282. [27](#)

- [GLS81] Martin Grötschel, László Lovász, and Alexander Schrijver, *The ellipsoid method and its consequences in combinatorial optimization*, Combinatorica **1** (1981), 169–197. [1](#), [6](#)
- [GLS93] ———, *Geometric algorithms and combinatorial optimization*, second ed., Algorithms and Combinatorics, vol. 2, Springer-Verlag, Berlin, 1993. [1](#), [13](#), [14](#), [29](#), [30](#), [31](#)
- [GN76] Paul Gordan and Max Noether, *Ueber die algebraischen formen, deren Hesse'sche determinante identisch verschwindet*, Mathematische Annalen **10** (1876), no. 4, 547–568. [10](#), [27](#)
- [GPS23] Sander Gribling, Sven Polak, and Lucas Slot, *A note on the computational complexity of the moment-SOS hierarchy for polynomial optimization*, Proceedings of the 2023 International Symposium on Symbolic and Algebraic Computation (New York, NY, USA), ISSAC ’23, Association for Computing Machinery, 2023, pp. 280–288. [6](#)
- [GR09] Alice Garbagnati and Flavia Repetto, *A geometrical approach to Gordan-Noether's and Franchetta's contributions to a question posed by Hesse*, Collectanea mathematica **60** (2009), no. 1, 27–41. [10](#), [27](#)
- [GW95] Michel X. Goemans and David P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, Journal of the Association for Computing Machinery **42** (1995), no. 6, 1115–1145. [6](#)
- [Hes51] Otto Hesse, *Über die Bedingung, unter welche eine homogene ganze Function von nunabhängigen Variabeln durch lineäre Substitutionem von n andern unabhängigen Variabeln auf eine homogene Function sich zurück-führen lässt, die eine Variable weniger enthält*, Journal für die reine und angewandte Mathematik **42** (1851), 117–124. [2](#), [10](#)
- [Imh63] J. P. Imhof, *On the method for numerical integration of Clenshaw and Curtis*, Numer. Math. **5** (1963), 138–141. [8](#)
- [Kha79] L. G. Khachiyan, *A polynomial algorithm in linear programming*, Soviet Mathematics – Doklady **1** (1979), 191–194. [1](#), [4](#), [5](#)
- [KP80] Richard M. Karp and Christos H. Papadimitriou, *On linear characterizations of combinatorial optimization problems*, 21st Annual Symposium on Foundations of Computer Science (sfcs 1980), IEEE, October 1980, p. 1–9. [1](#)
- [KTK80] M.K. Kozlov, S.P. Tarasov, and L.G. Khachiyan, *The polynomial solvability of convex quadratic programming*, USSR Computational Mathematics and Mathematical Physics **20** (1980), no. 5, 223–228. [1](#), [4](#), [5](#)
- [Las01] Jean B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM Journal on Optimization **11** (2001), no. 3, 796–817. [6](#)
- [Las08] ———, *Representation of nonnegative convex polynomials*, Archiv der Mathematik **91** (2008), no. 2, 126–130. [6](#)
- [Lau08] Monique Laurent, *Sums of squares, moment matrices and optimization over polynomials*, pp. 157–270, Springer New York, 2008. [5](#)
- [LLL82] Arjen. K. Lenstra, Hendrik W. Lenstra, and László Lovász, *Factoring polynomials with rational coefficients*, Mathematische Annalen **261** (1982), no. 4, 515–534. [27](#)

- [Min64] George Minty, *On the monotonicity of the gradient of a convex function*, Pacific Journal of Mathematics **14** (1964), no. 1, 243–247. [9](#), [16](#)
- [Nes19] Yurii Nesterov, *Implementable tensor methods in unconstrained convex optimization*, Mathematical Programming **186** (2019), no. 1–2, 157–183. [5](#)
- [O'D17] Ryan O'Donnell, *SOS Is Not Obviously Automatizable, Even Approximately*, 8th Innovations in Theoretical Computer Science Conference (ITCS 2017) (Dagstuhl, Germany) (Christos H. Papadimitriou, ed.), Leibniz International Proceedings in Informatics (LIPIcs), vol. 67, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017, pp. 59:1–59:10. [1](#), [6](#)
- [Par00] Pablo A. Parrilo, *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*, Ph.D. thesis, California Institute of Technology, 2000. [6](#)
- [PK97] Lorant Porkolab and Leonid Khachiyan, *On the complexity of semidefinite programs*, Journal of Global Optimization **10** (1997), no. 4, 351–365. [1](#), [6](#)
- [PR81] M.W. Padberg and M.R. Rao, *The Russian method for linear inequalities III: bounded integer programming*, Tech. Report RR-0078, INRIA, May 1981. [1](#)
- [PT24] Gábor Pataki and Aleksandr Touzov, *How do exponential size solutions arise in semidefinite programming?*, SIAM Journal on Optimization **34** (2024), no. 1, 977–1005. [1](#), [6](#)
- [Ram97] Motakuri V. Ramana, *An exact duality theory for semidefinite programming and its complexity implications*, Mathematical Programming **77** (1997), no. 1, 129–162. [1](#), [6](#)
- [Roc70] R. Tyrrell Rockafellar, *Convex analysis*, Princeton Mathematical Series, vol. No. 28, Princeton University Press, Princeton, NJ, 1970. [16](#), [21](#)
- [Rot15] Joseph J. Rotman, *Advanced modern algebra. Part 1*, third ed., Graduate Studies in Mathematics, vol. 165, American Mathematical Society, Providence, RI, 2015. [29](#)
- [RW17] Prasad Raghavendra and Benjamin Weitz, *On the Bit Complexity of Sum-of-Squares Proofs*, 44th International Colloquium on Automata, Languages, and Programming (ICALP 2017) (Dagstuhl, Germany) (Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl, eds.), Leibniz International Proceedings in Informatics (LIPIcs), vol. 80, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017, pp. 80:1–80:13. [6](#), [16](#)
- [Sah74] Sartaj Sahni, *Computationally related problems*, SIAM Journal on Computing **3** (1974), no. 4, 262–279. [4](#), [5](#)
- [Sch80] Jacob T. Schwartz, *Fast probabilistic algorithms for verification of polynomial identities*, Journal of the ACM **27** (1980), no. 4, 701–717. [15](#)
- [Sch86] Alexander Schrijver, *Theory of linear and integer programming*, Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons, Ltd., Chichester, 1986, A Wiley-Interscience Publication. [13](#), [22](#), [25](#), [26](#), [27](#)
- [Sma00] Steve Smale, *Mathematical problems for the next century*, La Gaceta de la Real Sociedad Matemática Española **3** (2000), no. 3, 413–434. [5](#)
- [Sto00] Arne Storjohann, *Algorithms for matrix canonical forms*, Ph.D. thesis, ETH Zurich, 2000. [27](#)

- [Vav90] Stephen A. Vavasis, *Quadratic programming is in NP*, Information Processing Letters **36** (1990), no. 2, 73–77. [4](#), [5](#)
- [Vis21] Nisheeth K. Vishnoi, *Algorithms for convex optimization*, 1st ed. ed., Cambridge University Press, 2021. [13](#), [14](#), [29](#), [32](#)
- [YN76] D. Yudin and A. S. Nemirovski, *Informational complexity and effective methods of solution of convex extremal problems*, Econom. Math. Methods **12** (1976), 357–369. [1](#)