SIAM J. COMPUT.
Vol. 49, No. 1, pp. 138–169

THE MINIMUM EUCLIDEAN-NORM POINT IN A CONVEX POLYTOPE: WOLFE'S COMBINATORIAL ALGORITHM IS EXPONENTIAL*

JESÚS A. DE LOERA†, JAMIE HADDOCK‡, AND LUIS RADEMACHER†

Abstract. The complexity of Philip Wolfe's method for the minimum Euclidean-norm point problem over a convex polytope has remained unknown since he proposed the method in 1974. The method is important because it is used as a subroutine for one of the most practical algorithms for submodular function minimization. We present the first example that Wolfe's method takes exponential time. Additionally, we improve previous results to show that linear programming reduces in strongly polynomial time to the minimum norm point problem over a simplex.

Key words. convex quadratic optimization, Wolfe's method, linear programming, strongly polynomial time algorithms, lower bounds

AMS subject classifications. 52B12, 68Q17, 68Q25

DOI. 10.1137/18M1221072

The fundamental algorithmic problem we consider here[1] is: Given a convex polytope $P \subset \mathbb{R}^d$, find the point $\mathbf{x} \in P$ of minimum Euclidean norm, i.e., the *closest point to the origin* or what we call its *minimum norm point* for short. We denote the Euclidean norm by $\|\cdot\|$ throughout. We assume P is presented as the convex hull of finitely many points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ (not necessarily in convex position). We wish to find

$$\begin{aligned} &\text{argmin} \quad \|\mathbf{x}\| \\ &\text{subject to} \quad \mathbf{x} = \sum_{k=1}^{n} \lambda_k \mathbf{p}_k, \\ &\qquad \sum_{k=1}^{n} \lambda_k = 1, \\ &\qquad \lambda_k \geq 0, \text{ for } k = 1, 2, \ldots, n. \end{aligned}$$

Finding the minimum norm point in a polytope is a basic auxiliary step in several algorithms arising in many areas of optimization and machine learning; a subroutine

*Received by the editors October 16, 2018; accepted for publication (in revised form) November 14, 2019; published electronically February 18, 2020. An extended abstract of this work appeared in *Proceedings of STOC* 2018.

https://doi.org/10.1137/18M1221072

Funding: This work was partially supported by NSF grant DMS-1440140 while the authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the fall 2017 semester. The first and second authors were also partially supported by NSF grant DMS-1522158. The second author was also partially supported by the University of California, Davis, dissertation fellowship. The third author was also partially supported by NSF grants CCF-1657939 and CCF-1422830.

†Department of Mathematics, University of California, Davis, Davis, CA 95616 (deloera@math.ucdavis.edu, lrademac@ucdavis.edu).

‡Department of Mathematics, University of California, Los Angeles, Los Angeles, CA 90095 (jhaddock@math.ucla.edu).

[1] An extended abstract of this work appeared in *Proceedings of STOC* 2018 [10].

for solving the minimum norm point problem can be used to compute the projection of an arbitrary point to a polytope (indeed, $\mathrm{argmin}_{\mathbf{x}\in P}\|\mathbf{x}-\mathbf{a}\|$ is the same as $\mathbf{a}+\mathrm{argmin}_{\mathbf{y}\in P-\mathbf{a}}\|\mathbf{y}\|$). The minimum norm point problem additionally appears in combinatorial optimization, e.g., the nearest point problem for transportation polytopes [2, 5], and as a vital subroutine in Bárány and Onn's approximation algorithm to solve the colorful linear programming problem [3]. One of the most important reasons to study this problem is because the minimum norm point problem can be used as a subroutine for submodular function minimization through projection onto the base polytope, as proposed by Fujishige [12]. Submodular minimization is useful in machine learning, where applications such as large scale learning and vision require efficient and accurate solutions [1, 26]. The problem also appears in optimal loading of recursive neural networks [7]. The Fujishige–Wolfe algorithm is currently considered an important practical algorithm in applications [6, 13, 14]. Furthermore, Fujishige, Hayashi, and Isotani first observed that linear programs may be solved by solving the minimum norm point problem [13], so this simple geometric problem is also relevant to the theory of algorithmic complexity of linear optimization.

One may ask about the complexity of closely related problems. First, it is worth remembering that L_p norm minimization over a polyhedron is NP-hard for $0 \le p < 1$ (see [17] and the references therein), while for $p \ge 1$ the convexity of the norm allows for computation of an ϵ-approximate solution in time polynomial in $\log(1/\epsilon)$ [4, 22]. We do not aim to be comprehensive in this discussion; see [19] for a detailed discussion of the complexity of convex programming. When $p = 1, 2$, L_p norm minimization over a polyhedron given by rational data has a rational solution (see [30] for details), so the minimum norm point can be computed exactly in polynomial time. Meanwhile, we now prove that the seemingly similar problem of finding the minimum norm *vertex* of a convex polytope given by inequalities is NP-hard. Here the distinction between point and vertex distinguishes the polynomially solvable problem from the NP-hard problem. The reduction for hardness is from the *directed Hamiltonian path problem*: Given a directed graph $G = (V, A)$ and two distinct vertices $s, t \in V$, one aims to decide whether G contains a directed Hamiltonian path from s to t. To be precise, we argue that the directed Hamiltonian path problem reduces to the problem of determining whether there is a vertex of squared norm at most a given value K. It is well known that there is a polytope represented by inequalities which has some vertices corresponding to the characteristic vectors of directed paths joining s to t in G. Let P be the directed st-path polytope of a directed graph $G = (V, A)$ and vertices s, t mentioned in [15]. The polytope is defined by the inequalities and equations

$$\begin{aligned}
\sum_{j:(s,j)\in A} x_{sj} - \sum_{j:(j,s)\in A} x_{js} &= 1,\\
\sum_{j:(t,j)\in A} x_{tj} - \sum_{j:(j,t)\in A} x_{jt} &= -1,\\
\sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} &= 0 \text{ for all } i \in V - \{s,t\},\\
0 \le x_{ij} &\le M \text{ for all } (i,j) \in A,
\end{aligned}$$

where there is a variable x_{ij} for every edge (i, j) in A. For $M \geq |V|$, there are two types of vertices of P. Either the vertices of P are characteristic vectors of the simple directed st-paths in G [15] or at least one of the entries of the vertex must be $x_{ij} = M$. Hamiltonian paths in G correspond to 0–1 vectors in P with exactly $|V|-1$ ones. Now we introduce the *reflected st-path polytope*, $\bar{P}$, by applying the affine transformation $y_{ij} = 1 - x_{ij}$. Note that the vertices of this polytope correspond to the vertices of polytope P. The vertices of $\bar{P}$ corresponding to directed st-paths are 0–1 vectors with at least $|A| - |V| + 1$ ones. Moreover, they have exactly $|A| - |V| + 1$ ones if and only if they correspond to a directed st-Hamiltonian path in the graph. The vertices of $\bar{P}$ corresponding to vertices of P with at least one entry equal to M must have at least one entry equal to $-M + 1$, so these have squared norm at least $(M - 1)^2$. We choose $M := \max\{|A| + 2, |V|\}$ so that $(M - 1)^2 \geq (|A| + 1)^2 > |A| - |V| + 1$. Thus, the minimum norm vertex of $\bar{P}$ has squared norm less than or equal to $|A|-|V|+1$ if and only if it corresponds to a directed st-Hamiltonian path in G. Taking $K := |A|-|V|+1$ completes the correctness of the hardness reduction.

In this paper, we focus on combinatorial algorithms that rely on the structure of the polytope. There are several reasons to study the complexity of combinatorial algorithms for the minimum norm problem. On one hand, the minimum norm problem can indeed be solved in strongly polynomial time for some polytopes, most notably in network-flow and transportation polytopes (see [2, 5, 34], and references therein, for details). On the other hand, while linear programming reduces to the minimum norm problem, it is unknown whether linear programming can be solved in strongly polynomial time [32]; thus the complexity of the minimum norm point problem could also impact the algorithmic efficiency of linear programming and optimization in general. For all these reasons it is natural to ask whether a strongly polynomial time algorithm exists for the minimum norm point problem for general polytopes.

Our contributions.

- In 1974, Philip Wolfe proposed a combinatorial method that can solve the minimum norm point problem exactly [35, 36]. Since then, the complexity of Wolfe's method has not been understood. In section 1 we present our main contribution and give the first example in which Wolfe's method has an exponential number of iterations. This is akin to the well-known Klee–Minty examples on which the simplex method has an exponential number of iterations [21].
- As we mentioned earlier, an enticing reason to explore the complexity of the minimum norm problem is its intimate link to the complexity of linear programming. It is known that linear programming can be polynomially reduced to the minimum norm point problem [13]. In section 2, we strengthen earlier results showing that linear optimization is strongly polynomial time reducible to the minimum norm point problem on a simplex.

1. Wolfe's method exhibits exponential behavior. For convenience of the reader and to set up notation we start with a brief description of Wolfe's method; however, for efficiency of presentation, we refer the reader to [30] for relevant definitions and preliminary results in convex analysis. We will then describe our exponential example in detail, proving the exponential behavior of Wolfe's method. First, we review definitions to describe the method. We denote the line segment between points $\mathbf{x}$ and $\mathbf{y}$ by $[\mathbf{x}, \mathbf{y}]$. Given a set of points $S \subseteq \mathbb{R}^d$, we have two minimum-norm points to consider. One is the *affine minimizer*, which is the point of minimum norm in the affine hull of S, $\operatorname{argmin}_{\mathbf{x}\in\operatorname{aff}(S)}\|\mathbf{x}\|$. The second is the *convex minimizer*, which is the point of minimum norm in the convex hull of S, $\operatorname{argmin}_{\mathbf{x}\in\operatorname{conv}(S)}\|\mathbf{x}\|$. Note that solving for the convex minimizer of a set of points is exactly the problem we are solving, while solving for the affine minimizer of a set of points is easily computable by solving a system of linear equations and may be computed in strongly polynomial time; see [30, section 3.3] and references therein.

1.1. A brief review of Wolfe's combinatorial method. Wolfe's combinatorial method solves the minimum norm point problem over a polytope, $P = \operatorname{conv}(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n) \subset \mathbb{R}^d$, and was introduced by Wolfe in [36]. The method iteratively solves the minimum norm point problem over a sequence of subsets of no more than $d+1$ affinely independent points from $\mathbf{p}_1, \dots, \mathbf{p}_n$ and it checks to see if the solution to the subproblem is a solution to the problem over P using the following lemma due to Wolfe. We call this *Wolfe's criterion*; a visualization is provided in Figure 1.1.

LEMMA 1.1 (Wolfe's criterion [36]). *Let* $P = \operatorname{conv}(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n) \subset \mathbb{R}^d$. *Let* $\mathbf{x} \in P$. *Then* $\mathbf{x}$ *is the minimum norm point in* P *if and only if*

$$\mathbf{x}^\top \mathbf{p}_j \geq \|\mathbf{x}\|^2 \quad \textit{for all} \quad j \in [n].$$

Note that this tells us that if there exists a point $\mathbf{p}_j$ so that $\mathbf{x}^\top\mathbf{p}_j < \|\mathbf{x}\|^2$ (i.e., the hyperplane $\{\mathbf{y} : \mathbf{x}^\top\mathbf{y} = \|\mathbf{x}\|^2\}$ does not weakly separate P from $\mathbf{0}$), then $\mathbf{x}$ is not the minimum norm point in P. We say that $\mathbf{p}_j$ violates Wolfe's criterion and using this point should decrease the norm of the minimum norm point of the current subproblem. This criterion is a special case of the general optimality criterion in differentiable convex optimization which says that $\mathbf{x}^*$ is optimal over a convex set if and only if $\nabla f(\mathbf{x}^*)^\top(\mathbf{x} - \mathbf{x}^*) \geq 0$ for all feasible $\mathbf{x}$ [4, p. 139]; to apply this result consider the objective function $f(\boldsymbol{\lambda}) = \frac{1}{2}\|Q\boldsymbol{\lambda}\|^2$ over the feasible region $\{\boldsymbol{\lambda} : \sum_i \lambda_i = 1, \boldsymbol{\lambda} \geq \mathbf{0}\}$ where Q is the matrix with the points $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$ as columns.

It should be observed that just as Wolfe's criterion is a rule to decide optimality over $\operatorname{conv}(P)$, one has a very similar rule for deciding optimality over the affine hull, $\operatorname{aff}(P)$.

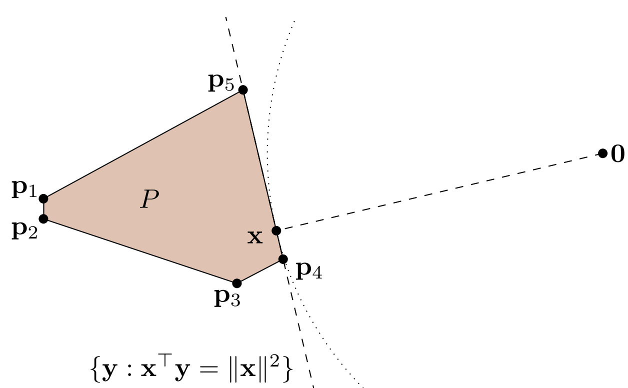


FIG. 1.1. *A visualization of Wolfe's criterion. Note that* $\{\mathbf{y} : \mathbf{x}^\top\mathbf{y} = \|\mathbf{x}\|^2\}$ *weakly separates* P *from* $\mathbf{0}$*, so* $\mathbf{x}$ *is the minimum norm point in* P.

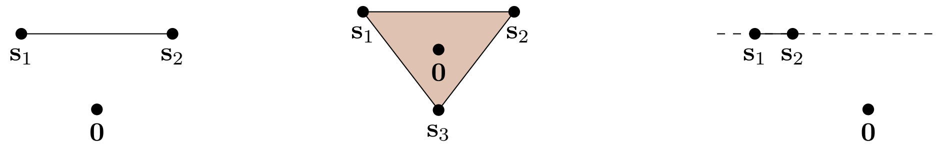


FIG. 1.2. *Left: Example of corral. Middle: Example of corral. Right: Not a corral.*

LEMMA 1.2 (Wolfe's criterion for the affine hull). *Let $P = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\} \subseteq \mathbb{R}^d$ be a nonempty finite set of points. Let $\mathbf{x} \in \operatorname{aff}(P)$. Then $\mathbf{x}$ is the minimum norm point in $\operatorname{aff}(P)$ if and only if for all $\mathbf{p}_i \in P$ we have $\mathbf{p}_i^\top \mathbf{x} = \|\mathbf{x}\|^2$.*

This result follows from [4, p. 139] applied to objective function $f(\boldsymbol{\lambda}) = \frac{1}{2}\|Q\boldsymbol{\lambda}\|^2$ over the feasible region $\{\boldsymbol{\lambda} : \sum_i \lambda_i = 1\}$ where Q is the matrix with the points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ as columns. Finally, the linear inequality must actually be an equality, since if there is feasible $\boldsymbol{\lambda}$ with $\nabla f(\boldsymbol{\lambda}^*)(\boldsymbol{\lambda} - \boldsymbol{\lambda}^*) > 0$, then the feasible affine space also contains $\boldsymbol{\lambda}'$ with $\nabla f(\boldsymbol{\lambda}^*)(\boldsymbol{\lambda}' - \boldsymbol{\lambda}^*) < 0$, which is a contradiction.

We say a set of affinely independent points S is a *corral* if the affine minimizer of S lies in the relative interior of $\operatorname{conv}(S)$; see Figure 1.2 for examples. Requiring the affine minimizer to lie in the relative interior of the convex hull ensures that corrals are of minimal size and without points unnecessary for expressing the affine minimizer as a convex combination of the corral. Note that singletons are always corrals. Carathéodory's theorem implies that the minimum norm point of P will lie in the convex hull of some corral of points among $\mathbf{p}_1, \ldots, \mathbf{p}_n$. The goal of Wolfe's method is to search for a corral containing the (unique) minimizing point.

The pseudocode in Method 1.1 below presents the iterations of Wolfe's method. We have additionally provided a visualization of the iterations of the method on an example in Figure 1.3; this example is the same as in [36]. It is worth noticing that some steps of the method can be implemented in more than one way and Wolfe proved that all of them lead to a correct, terminating algorithm (for example, the choice of the initial point in line 2). We therefore use the word *method* to encompass all these variations and we discuss specific choices when they are relevant to our analysis of the method.

The subset of points being considered as the *potential corral* is maintained in the set C. Iterations of the outer loop, where points are added to C, are called *major cycles* and iterations of the inner loop, where points are removed from C, are called *minor cycles*. The potential corral, C, is named so because at the beginning of a major cycle it is guaranteed to be a corral, while within the minor cycles it may or may not be a corral. Intuitively, a major cycle of Wolfe's method inserts an *improving point* which violates Wolfe's criterion ($\mathbf{p}_j$ so that $\mathbf{x}^\top \mathbf{p}_j < \|\mathbf{x}\|^2$) into C, then the minor cycles remove points until C is a corral, and this process is repeated until no points are improving and C is guaranteed to be a corral containing the minimizer.

It can be shown that this method terminates because the norm of the convex minimizer of the corrals visited strictly decreases and thus no corral is visited twice [36]. Like [6], we sketch the argument in [36]. One may see that the norms of the iterates strictly decrease by noting that there are two possible updates to the current iterate $\mathbf{x}$, either at the end of a major cycle or at the end of a minor cycle. Let C be

Method 1.1 Wolfe's method [36].

1: **procedure** WOLFE($\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$)
2: Initialize $\mathbf{x} = \mathbf{p}_i$ for some $i \in [n]$, initial corral $C = \{\mathbf{p}_i\}$, $I = \{i\}$, $\lambda = \mathbf{e}_i$, $\alpha = \mathbf{0}$.
3: **while** $\mathbf{x} \neq \mathbf{0}$ and there exists $\mathbf{p}_j$ with $\mathbf{x}^\top \mathbf{p}_j < \|\mathbf{x}\|^2$ **do**
4: Add $\mathbf{p}_j$ to the potential corral: $C = C \cup \{\mathbf{p}_j\}$, $I = I \cup \{j\}$.
5: Find the affine minimizer of C, $\mathbf{y} = \operatorname{argmin}_{\mathbf{y} \in \operatorname{aff}(C)} \|\mathbf{y}\|$, and the affine coefficients, α.
6: **while** $\mathbf{y}$ is not a strict convex combination of C; $\alpha_i \leq 0$ for some $i \in I$ **do**
7: Find $\mathbf{z}$, closest point to $\mathbf{y}$ on $[\mathbf{x}, \mathbf{y}] \cap \operatorname{conv}(C)$; $\mathbf{z} = \theta \mathbf{y} + (1 - \theta)\mathbf{x}$, $\theta = \min_{i \in I : \alpha_i \leq 0} \frac{\lambda_i}{\lambda_i - \alpha_i}$.
8: Select $\mathbf{p}_i \in \{\mathbf{p}_k \in C : \theta \alpha_k + (1 - \theta)\lambda_k = 0\}$.
9: Remove this point from C; $C = C - \{\mathbf{p}_i\}$, $I = I - \{i\}$, $\alpha_i = 0$, $\lambda_i = 0$.
10: Update $\mathbf{x} = \mathbf{z}$ and the convex coefficients, λ, of $\mathbf{x}$ for C; solve $\mathbf{x} = \sum_{\mathbf{p}_i \in C} \lambda_i \mathbf{p}_i$ for λ.
11: Find the affine minimizer of C, $\mathbf{y} = \operatorname{argmin}_{\mathbf{y} \in \operatorname{aff}(C)} \|\mathbf{y}\|$ and the affine coefficients, α.
12: **end while**
13: Update $\mathbf{x} = \mathbf{y}$ and $\lambda = \alpha$.
14: **end while**
15: **Return** $\mathbf{x}$.
16: **end procedure**

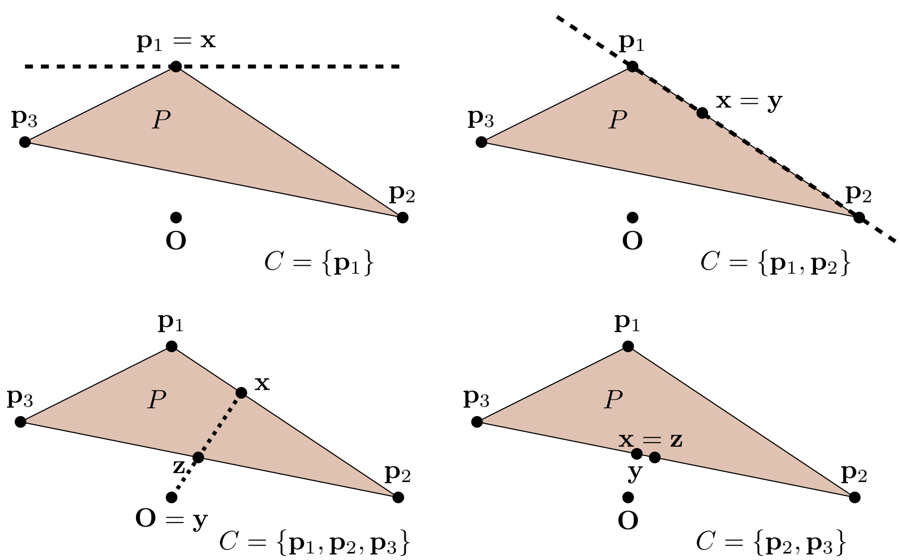


FIG. 1.3. *Example of several iterations of Wolfe's method on a polytope in* $\mathbb{R}^2$. *Upper left: Line* 2 *of Method* 1.1. *Upper right: This example is the same as in* [36].

the corral at the beginning of a major cycle (line 3 of Method 1.1) and let $\mathbf{x}$ be the current convex minimizer of C, then the affine minimizer of $C \cup \{\mathbf{p}_i\}$, $\mathbf{y}$, has norm strictly less than that of $\mathbf{x}$ by Lemma 1.2, uniqueness of the affine minimizer, and the fact that $\mathbf{p}_i^\top \mathbf{x} < \|\mathbf{x}\|^2$, where $\mathbf{p}_i$ is the added point. Now, either $\mathbf{x}$ is updated to $\mathbf{y}$ or a minor cycle begins. Let S be the potential corral at the beginning of the first minor cycle (line 6 of Method 1.1) of a major cycle, let $\mathbf{x}$ be the current iterate (which is a convex combination of points of S but not the convex minimizer of S), and let $\mathbf{y}$ be the affine minimizer of S. Note that $\mathbf{z}$ is a proper convex combination of $\mathbf{x}$ and $\mathbf{y}$ and since $\|\mathbf{y}\| < \|\mathbf{x}\|$, we have $\|\mathbf{z}\| < \|\mathbf{x}\|$. Thus, we see that within a major cycle the first update of $\mathbf{x}$ decreases its norm (at either line 10 or line 13 of Method 1.1). Note that the number of minor cycles within any major cycle is bounded by $d+1$, where d is the dimension of the space. Thus, the total number of iterations is bounded by the number of corrals visited multiplied by $d+1$. It is nevertheless not clear how the number of corrals grows. A trivial bound is given by $\sum_{i=1}^{d+1}\binom{n}{i}$. We think this problem should be investigated further.

Within the method, there are two moments at which one may choose which points to add to the potential corral. Observe that at line 2 of the pseudocode, one may choose which initial point to add to the potential corral. In this paper we will only consider one *initial rule*, which is to initialize with the point of minimum norm. Observe that at line 4 of the pseudocode, there are several potential choices of which point to add to the potential corral. Two important examples of *insertion rules* are, first, the *minnorm rule* which dictates that one chooses, out of the improving points for the potential corral, to add the point $\mathbf{p}_j$ of minimum norm. Second, the *linopt rule* dictates that one chooses, out of the improving points for the potential corral, to add the point $\mathbf{p}_j$ minimizing $\mathbf{x}^\top \mathbf{p}_j$. Notice that insertion rules are to Wolfe's method what *pivot rules* are to the simplex method (see [33] for a summary). Additionally, in line 8 of Method 1.1, there can be choices of which points to remove. However, removal rules of this type are less important to the analysis of Wolfe's method as any points which are able to be removed will be removed before the end of the major cycle.

As with pivot rules, there are advantages and disadvantages of insertion rules. For example, the minnorm rule has the advantage that its implementation only requires an initial ordering of the points; then in each iteration it needs only to search for an improving point in order of increasing norm and to add the first found. However, the linopt insertion rule has the advantage that if the polytope is given in H-representation (intersection of halfspaces) rather than V-representation (convex hull of points), one may still perform Wolfe's method by using linear programming to find $\mathbf{p}_j$ minimizing $\mathbf{x}^\top \mathbf{p}_j$ over the polytope. In other words, Wolfe's method does not need to have the list of vertices explicitly given but suffices to have a linear optimization oracle that provides the new vertex to be inserted. This feature of Wolfe's method means that each iteration can be implemented efficiently even for certain polyhedra having too many vertices and facets, specifically over zonotopes (presented as a Minkowski sum of segments) [13] and over the base polyhedron of a submodular function [12]. It also works for some polytopes that do not admit small linear programming formulations such as the matching polytope, which has an efficient linear optimization oracle; see [28, 29] and references therein. This is not a property that the minnorm rule shares; we must have the polytope in explicit V-representation to perform Wolfe's method with the minnorm rule.

Wolfe's method with the linopt insertion rule was independently discovered by Lawson and Hanson in an essentially equivalent form, known as the Lawson–Hanson algorithm for nonnegative least squares [25, section 23.3]). They are similar to other active-set methods in convex optimization. In particular, von Neumann's algorithm for determining whether the origin lies in a convex polytope [9], Gilbert's procedure for computing the minimum of a quadratic form on a convex set [18], and the Frank–Wolfe method for convex optimization [11] make use of the same active-set selection criterion, the linopt insertion rule. Several of these methods and their variants have been shown to have sublinear or linear convergence (with appropriate assumptions on the location of the minimizer); see e.g., [23, 24, 27]. In [24], the authors additionally studied Wolfe's method with the linopt insertion rule, showing that the method converges linearly with a rate $O(e^{-\rho t})$, where ρ is an eccentricity parameter of the polytope. This improved upon prior results by [6], which had provided a sublinear ($O(1/t)$) rate. Note, however, that the parameter ρ defined in [24] may be exponentially small in the encoding length of the problem, so both results give only a pseudo-polynomial time bound for Wolfe's method with the linopt insertion rule. A simple example exhibiting exponentially small ρ is $A = \{\mathbf{a}_1 = (1/2^k, 0), \mathbf{a}_2 = (-1/2^k, 0), \mathbf{a}_3 = (0, 1)\}$; using the notation of [24], note that $\mathrm{diam}(A) \geq 1$ while

$$\mathrm{PWidth}(A) \leq \mathrm{PdirW}(\{\mathbf{a}_1, \mathbf{a}_2\}, \mathbf{a}_1, (0,0)) = \max_{\mathbf{s} \in A, \mathbf{v} \in \{\mathbf{a}_1, \mathbf{a}_2\}} \langle (1,0), \mathbf{s} - \mathbf{v} \rangle = 1/2^{k-1}$$

so $\rho \leq 1/2^{k-1}$.

In general, the optimal choice of insertion rule depends on the input data. We first present a simple example where the minnorm rule outperforms the linopt rule. That is, the minnorm insertion rule is not in obvious disadvantage to the linopt rule. This is in contrast to the family of examples we present in subsection 1.2, where the minnorm rule takes exponential time, while we expect the linopt rule to take polynomial time in this family.

Consider the simplex P shown in Figure 1.4 (we present the coordinates of vertices in the figure's caption). We list the steps of Wolfe's method on P for the minnorm and linopt insertion rules in Tables 1.1 and 1.2 and demonstrate a single step from each set of iterations in Figure 1.5. Each row lists major cycle and minor cycle iteration number, the vertices in the potential corral, and the value of $\mathbf{x}$ and $\mathbf{y}$ at the end of the iteration (before $\mathbf{x} = \mathbf{y}$ for major cycles). Note that the vertex $\mathbf{p}_4$ is added to the potential corral twice with the linopt insertion rule, as evidenced in Table 1.2.

Currently, there are examples of exponential behavior for the simplex method for all known deterministic pivot rules. It is our aim to provide the same for insertion rules on Wolfe's method. In the next subsection we will present the first exponential-time example using the minnorm insertion rule. We additionally note that while in the previous tables we recorded all sets C constructed by the algorithm (corral and noncorral), we will now track only the corrals the method passes through.

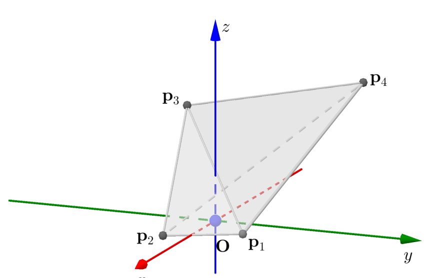


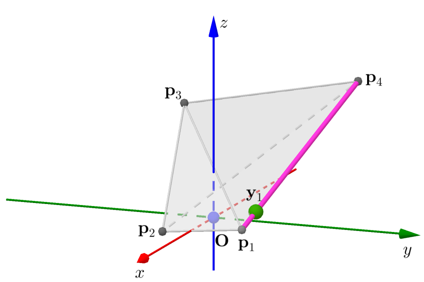
Fig. 1.4. *The simplex* $P = \mathrm{conv}\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4\} \subset \mathbb{R}^3$, *where* $\mathbf{p}_1 = (0.8, 0.9, 0), \mathbf{p}_2 = (1.5, -0.5, 0), \mathbf{p}_3 = (-1, -1, 2)$, *and* $\mathbf{p}_4 = (-4, 1.5, 2)$.

Table 1.1
Iterations for minnorm *insertion rule.*

Major, minor	C	$\mathbf{x}$	$\mathbf{y}$
0, 0	$\{\mathbf{p}_1\}$	$\mathbf{p}_1$	
1, 0	$\{\mathbf{p}_1, \mathbf{p}_2\}$	$\mathbf{p}_1$	$(1, 0.5, 0)$
2, 0	$\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$	$(1, 0.5, 0)$	$(0.3980, 0.199, 0.5473)$
3, 0	$\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4\}$	$(0.3980, 0.199, 0.5473)$	$(0, 0, 0)$
3, 1	$\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_4\}$	$(0.2878, 0.1439, 0.3957)$	$(0.1980, 0.0990, 0.4455)$

Table 1.2
Iterations for linopt *insertion rule.*

Major, minor	C	$\mathbf{x}$	$\mathbf{y}$
0, 0	$\{\mathbf{p}_1\}$	$\mathbf{p}_1$	
1, 0	$\{\mathbf{p}_1, \mathbf{p}_4\}$	$\mathbf{p}_1$	$(0.2219, 0.9723, 0.2409)$
2, 0	$\{\mathbf{p}_1, \mathbf{p}_4, \mathbf{p}_3\}$	$(0.2219, 0.9723, 0.2409)$	$(0.2848, 0.3417, 0.5810)$
2, 1	$\{\mathbf{p}_1, \mathbf{p}_3\}$	$(0.2835, 0.3548, 0.5739)$	$(0.2774, 0.3484, 0.5807)$
3, 0	$\{\mathbf{p}_1, \mathbf{p}_3, \mathbf{p}_2\}$	$(0.2774, 0.3484, 0.5807)$	$(0.3980, 0.199, 0.5473)$
4, 0	$\{\mathbf{p}_1, \mathbf{p}_3, \mathbf{p}_2, \mathbf{p}_4\}$	$(0.3980, 0.199, 0.5473)$	$(0, 0, 0)$
4, 1	$\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_4\}$	$(0.2878, 0.1439, 0.3957)$	$(0.1980, 0.0990, 0.4455)$



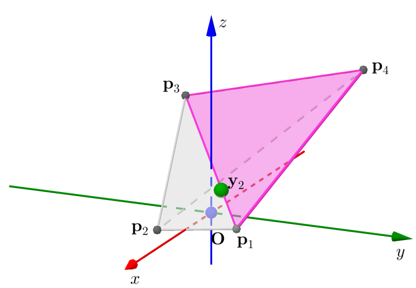


Fig. 1.5. *Left: Major cycle* 1*, minor cycle* 0 *for the linopt rule on* P *illustrates the end of a major cycle; the affine minimizer* $\mathbf{y}_1 \in \operatorname{relint}(\operatorname{conv}\{C\}) = \operatorname{relint}(\operatorname{conv}\{\mathbf{p}_1, \mathbf{p}_4\})$*. Right: Major cycle* 2*, minor cycle* 0 *for the linopt rule on* P *illustrates the beginning of a minor cycle; the affine minimizer* $\mathbf{y}_2 \notin \operatorname{relint}(\operatorname{conv}\{C\}) = \operatorname{relint}(\operatorname{conv}\{\mathbf{p}_1, \mathbf{p}_4, \mathbf{p}_3\})$*, and the vertex* $\mathbf{p}_4$ *will be removed in the next minor cycle.*

1.2. An exponential lower bound for Wolfe's method. To understand our hard instance, it is helpful to consider first a simple instance that shows an inefficiency of Wolfe's method. The example is a set of points where a point leaves, and reenters the current corral: four points in $\mathbb{R}^3$, $(1, 0, 0), (1/2, 1/4, 1), (1/2, 1/4, -1), (-2, 1/4, 0)$. If one labels the points $1, 2, 3, 4$, the sequence of corrals with the minnorm rule is $1, 12, 23, 234, 14$, where point 1 enters, leaves, and reenters (for succinctness, sets of points like $\{a, b, c\}$ may be denoted abc). The idea now is to recursively replace point 1 (that reenters) in this construction by a recursively constructed set of points whose corrals are then considered twice by Wolfe's method. To simplify the proof, our construction uses a variation of this set of four points with an additional point

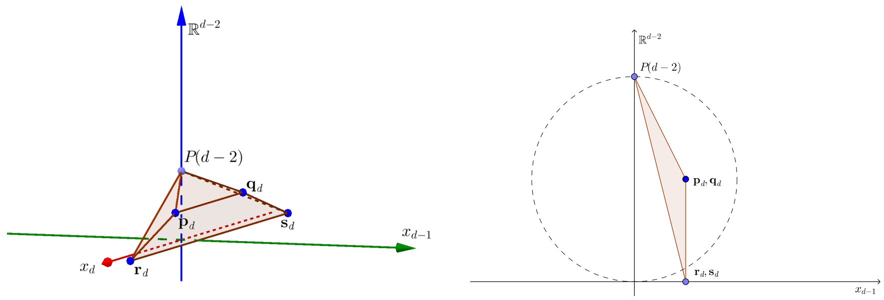


FIG. 1.6. *Left: In this view of $P(d)$, the point labeled $P(d-2)$ represents all points from $P(d-2)$ embedded into $\mathbb{R}^d$. The axis labeled $\mathbb{R}^{d-2}$ represents the $(d-2)$-dimensional subspace,* $\operatorname{span}(P(d-2))$ *projected into* $\operatorname{span}(\mathbf{o}^*_{d-2})$. *Right: A two-dimensional view of $P(d)$ projected along the x_d coordinate axis.*

and modified coordinates. This modified construction is depicted in Figure 1.6, where point 1 corresponds to a set of points $P(d-2)$, points 2 and 3 correspond to points $\mathbf{p}_d, \mathbf{q}_d$, and point 4 corresponds to points $\mathbf{r}_d, \mathbf{s}_d$.

The high-level idea of our exponential lower bound example is the following. We will inductively define a sequence of instances of the minimum norm point problem of increasing dimension. Given an instance in dimension $d-2$, we will add a few dimensions and points so that, when given to Wolfe's method, the new augmented instance in dimension d has about twice the number of corrals of the input instance in dimension $d-2$. More precisely, our augmentation procedure takes an instance $P(d-2)$ in $\mathbb{R}^{d-2}$, adds two new coordinates, and adds four points, $\mathbf{p}_d, \mathbf{q}_d, \mathbf{r}_d, \mathbf{s}_d$, to get an instance $P(d)$ in $\mathbb{R}^d$.

Points $\mathbf{p}_d, \mathbf{q}_d$ are defined so that the method on instance $P(d)$ first goes through every corral given by the points in the prior configuration $P(d-2)$ and then goes to corral $\mathbf{p}_d\mathbf{q}_d$. To achieve this under the minimum norm rule, the four new points have greater norm than any point in $P(d-2)$ and they are in the geometric configuration sketched in Figure 1.6.

At this time, no point in $P(d-2)$ is in the current corral and so, if a point in $P(d-2)$ is part of the optimal corral, it will have to reenter, which is expensive. Points $\mathbf{r}_d, \mathbf{s}_d$ are defined so that $\mathbf{r}_d\mathbf{s}_d$ is a corral after $\mathbf{p}_d\mathbf{q}_d$, but now every point in $P(d-2)$ is improving according to Wolfe's criterion and may enter again. Specifically, every corral in $P(d-2)$, with $\mathbf{r}_d\mathbf{s}_d$ appended, is visited again.

Before we start describing the exponential example in detail, we wish to review preliminary lemmas of independent interest which will be used in the arguments. The first proposition gives us an explicit description of the point of minimum norm in the line through two distinct points and is easily proven via techniques of differentiable convex optimization. Throughout the following results, we will often provide both geometric and analytical proof alternatives.

PROPOSITION 1.3. *Given two distinct points $\mathbf{a}, \mathbf{b}$, the minimum norm point in the line through them is $\lambda\mathbf{a} + (1-\lambda)\mathbf{b}$, where $\lambda = \mathbf{b}^\top(\mathbf{b}-\mathbf{a})/\|\mathbf{b}-\mathbf{a}\|^2$.*

The next lemma demonstrates that orthogonality between finite point sets allows us to easily describe the affine minimizer of their union. Figure 1.7 shows two such situations, one in which the affine hull of the union of the point sets spans all of $\mathbb{R}^3$ and one in which it does not.

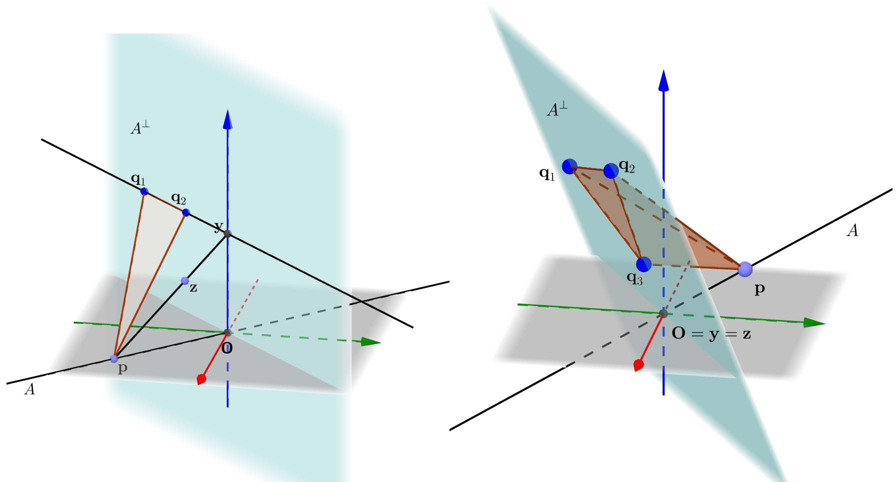


FIG. 1.7. *Examples of Lemma 1.4. Left: The affine hull of $P \cup Q$ is not full-dimensional, and thus the affine minimizer lies at $\mathbf{z}$ along the line segment connecting $\mathbf{x} = \mathbf{p}$ and $\mathbf{y}$. Right: The convex hull of $P \cup Q$ is full-dimensional and thus the affine hull of $P \cup Q$ includes O, which is the affine minimizer.*

LEMMA 1.4. *Let $A \subseteq \mathbb{R}^d$ be a proper linear subspace. Let $P \subseteq A$ be a nonempty finite set. Let $Q \subseteq A^{\perp}$ be another nonempty finite set. Let $\mathbf{x}$ be the minimum norm point in* $\operatorname{aff}(P)$. *Let $\mathbf{y}$ be the minimum norm point in* $\operatorname{aff}(Q)$. *Let $\mathbf{z}$ be the minimum norm point in* $\operatorname{aff}(P \cup Q)$. *We have*

1. $\mathbf{z}$ *is the minimum norm point in* $[\mathbf{x}, \mathbf{y}]$ *and therefore, if $\mathbf{x} \neq \mathbf{0}$ or $\mathbf{y} \neq \mathbf{0}$, then* $\mathbf{z} = \lambda \mathbf{x} + (1-\lambda)\mathbf{y}$ *with* $\lambda = \frac{\|\mathbf{y}\|^2}{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2}$.
2. *If $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{y} \neq \mathbf{0}$, then $\mathbf{z}$ is a strict convex combination of $\mathbf{x}$ and $\mathbf{y}$.*
3. *If $\mathbf{x} \neq \mathbf{0}$, $\mathbf{y} \neq \mathbf{0}$, and P and Q are corrals, then $P \cup Q$ is also a corral.*

Proof. If $\mathbf{x} = \mathbf{y} = \mathbf{0}$, then part 1 follows immediately. If at least one of $\mathbf{x}, \mathbf{y}$ is nonzero, then they are also distinct by the orthogonality assumption. For points $\mathbf{x}, \mathbf{y}$ as in the statement, Proposition 1.3 guarantees the minimum norm point in $\operatorname{aff}(\mathbf{x} \cup \mathbf{y})$ is $\mathbf{z}' = \lambda \mathbf{x} + (1-\lambda)\mathbf{y}$ with $\lambda = \frac{\|\mathbf{y}\|^2}{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2} \in [0,1]$. Thus, $\mathbf{z}'$ is also the minimum norm point in $[\mathbf{x}, \mathbf{y}]$. We will now use the optimality condition in Lemma 1.2 to conclude that $\mathbf{z}' = \mathbf{z}$. Let $\mathbf{p} \in P$. Then $\mathbf{p}^\top \mathbf{z}'$ can be computed in two steps. First project $\mathbf{p}$ onto $\operatorname{span}(\mathbf{x}, \mathbf{y})$ (a subspace that contains $\mathbf{z}'$). This projection is $\mathbf{x}$ by optimality of $\mathbf{x}$. Then project onto $\mathbf{z}'$. This shows that $\mathbf{p}^\top \mathbf{z}' = \mathbf{x}^\top \mathbf{z}' = \|\mathbf{z}'\|^2$.

One may alternatively prove $\mathbf{p}^\top \mathbf{z}' = \|\mathbf{z}'\|^2$ using direct calculations. From the definition of $\mathbf{z}'$ we have $\mathbf{p}^\top \mathbf{z}' = \lambda \mathbf{p}^\top \mathbf{x} = \lambda \|\mathbf{x}\|^2$ by orthogonality of $\mathbf{p}$ and $\mathbf{y}$ and optimality of $\mathbf{x}$, $\lambda \|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{z}'$ by orthogonality of $\mathbf{x}$ and $\mathbf{y}$, and $\mathbf{x}^\top \mathbf{z}' = \|\mathbf{z}'\|^2$ by optimality of $\mathbf{z}'$ over $\operatorname{aff}(\{\mathbf{x}, \mathbf{y}\})$.

A similar calculation shows $\mathbf{q}^\top \mathbf{z}' = \|\mathbf{z}'\|^2$ for any $\mathbf{q} \in Q$. We conclude that $\mathbf{z}'$ is the minimum norm point in $\operatorname{aff}(P \cup Q)$. This proves part 1.

Part 2 follows from our expression for λ above, which is in $(0,1)$ when $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{y} \neq \mathbf{0}$.

Under the assumptions of part 3, we have that $\mathbf{x}$ is a strict convex combination of P and $\mathbf{y}$ is a strict convex combination of Q. This combined with the conclusion of part 2 gives that $\mathbf{z}$ is a strict convex combination of $P \cup Q$. The claim in part 3 follows. $\square$

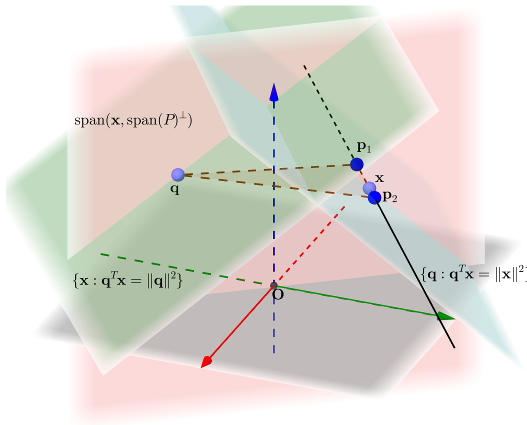


FIG. 1.8. *An example of Lemma* 1.5 *in which point* $\mathbf{q}$ *satisfies all assumptions and* $P \cup \{\mathbf{q}\}$ *is a corral. The hyperplanes are labeled with their defining properties and demonstrate that* $\mathbf{q}^\top \mathbf{x} < \min\{\|\mathbf{x}\|^2, \|\mathbf{q}\|^2\}$. *The minimizer of* $P \cup \{\mathbf{q}\}$ *lies at the intersection of the blue, vertical axis and* $\mathrm{conv}(P \cup \{\mathbf{q}\})$.

The following lemma shows conditions under which if we have a corral and a new point that only has components along the minimum norm point of the corral and along new coordinates, then the corral with the new point added is also a corral. Moreover, the new minimum norm point is a convex combination of the old minimum norm point and the added point. Figure 1.8 gives an example of such a situation in $\mathbb{R}^3$. Denote by $\mathrm{span}\,(M)$ the linear span of the set M.

LEMMA 1.5. *Let* $P \subseteq \mathbb{R}^d$ *be a finite set of points that is a corral. Let* $\mathbf{x}$ *be the minimum norm point in* $\mathrm{aff}(P)$. *Let* $\mathbf{q} \in \mathrm{span}(\mathbf{x}, \mathrm{span}\,(P)^\perp)$, *and assume* $\mathbf{q}^\top \mathbf{x} < \min\{\|\mathbf{q}\|^2, \|\mathbf{x}\|^2\}$. *Then* $P \cup \{\mathbf{q}\}$ *is a corral. Moreover, the minimum norm point* $\mathbf{y}$ *in* $\mathrm{conv}(P \cup \{\mathbf{q}\})$ *is a (strict) convex combination of* $\mathbf{q}$ *and the minimum norm point of* P*:* $\mathbf{y} = \lambda \mathbf{x} + (1-\lambda)\mathbf{q}$ *with* $\lambda = \mathbf{q}^\top(\mathbf{q} - \mathbf{x})/\|\mathbf{q} - \mathbf{x}\|^2$.

Proof. Let $\mathbf{y}$ be the minimum norm point in $\mathrm{aff}(P \cup \{\mathbf{q}\})$. Intuitively, $\mathbf{y}$ should be the minimum norm point in the line through $\mathbf{x}$ and $\mathbf{q}$. We will characterize $\mathbf{y}$ and show that it is a strict convex combination of $P \cup \{\mathbf{q}\}$ (which implies that it is a corral). We define $\mathbf{y} = \lambda \mathbf{x} + (1-\lambda)\mathbf{q}$ with $\lambda = \mathbf{q}^\top(\mathbf{q}-\mathbf{x})/\|\mathbf{q}-\mathbf{x}\|^2$; note Proposition 1.3 guarantees this is the point of minimum norm in $[\mathbf{x}, \mathbf{q}]$. By definition we have $\mathbf{y} \in \mathrm{aff}(P \cup \{\mathbf{q}\})$.

The minimality of the norm of $\mathbf{y}$ follows from the optimality condition in Lemma 1.2. It holds by construction for $\mathbf{q}$. It also holds for $\mathbf{p} \in P$: The projection of $\mathbf{p}$ onto $\mathbf{y}$ can be computed in two steps. First, project onto $\mathrm{span}\,(\mathbf{x}, \mathbf{q})$ (a subspace that contains $\mathbf{y}$), which is $\mathbf{x}$ by optimality of $\mathbf{x}$. Then project onto $\mathbf{y}$. This shows that $\mathbf{p}^\top \mathbf{y} = \mathbf{x}^\top \mathbf{y} = \|\mathbf{y}\|^2$ (the second equality, by optimality of $\mathbf{y}$). We conclude that $\mathbf{y}$ is of minimum norm in $\mathrm{aff}(P \cup \{\mathbf{q}\})$.

One may alternatively prove $\mathbf{p}^\top \mathbf{y} = \|\mathbf{y}\|^2$ using direct calculations. We have $\mathbf{q} = \sum_i \alpha_i \mathbf{q}_i + \beta \mathbf{x}$, where the $\mathbf{q}_i$ are a basis of $\mathrm{span}\,(P)^\perp$. Then we have $\mathbf{p}^\top \mathbf{y} = \lambda \mathbf{p}^\top \mathbf{x} + (1-\lambda)\mathbf{p}^\top \mathbf{q}$ and $\mathbf{x}^\top \mathbf{y} = \lambda \|\mathbf{x}\|^2 + (1-\lambda)\mathbf{x}^\top \mathbf{q}$. Now, the λ-terms are equal by optimality of $\mathbf{x}$ and the $(1-\lambda)$-terms are equal since $\mathbf{p}^\top \mathbf{q} = \beta \mathbf{p}^\top \mathbf{x} = \beta \|\mathbf{x}\|^2$ by orthogonality of $\mathbf{p}$ and $\mathbf{q}_i$, and $\beta \|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{q}$ by orthogonality of $\mathbf{x}$ and $\mathbf{q}_i$. Thus, we have $\mathbf{p}^\top \mathbf{y} = \mathbf{x}^\top \mathbf{y} = \|\mathbf{y}\|^2$, where the second equality follows from optimality of $\mathbf{y}$ over $\mathrm{aff}(\{\mathbf{x}, \mathbf{q}\})$.

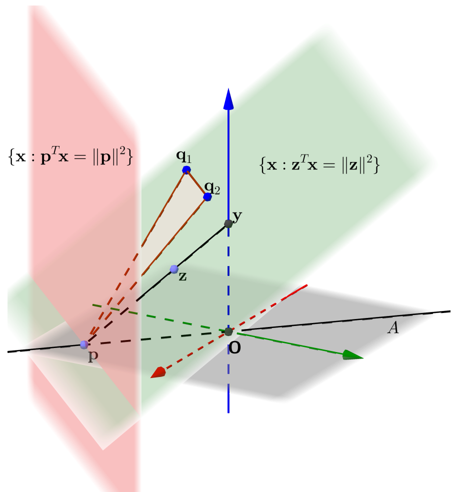


FIG. 1.9. *An example of Lemma* 1.6 *in which adding points Q from $A^\perp$ to points P from A creates a new affine minimizer,* $\mathbf{z}$*, but the points satisfying Wolfe's criterion in A remain the same. Note that both hyperplanes intersect at the affine minimizer of P, so the halfspace intersections with A are the same.*

To conclude that $P \cup \{\mathbf{q}\}$ is a corral, we show that $\mathbf{y}$ is a strict convex combination of points $P \cup \{\mathbf{q}\}$. It is enough to show that $\mathbf{y}$ is a strict convex combination of $\mathbf{x}$ and $\mathbf{q}$. We have $\lambda = \mathbf{q}^\top(\mathbf{q}-\mathbf{x})/\|\mathbf{q}-\mathbf{x}\|^2 = \frac{\|\mathbf{q}\|^2 - \mathbf{q}^\top\mathbf{x}}{\|\mathbf{q}-\mathbf{x}\|^2} > 0$ by assumption. We also have $1-\lambda = -\mathbf{x}^\top(\mathbf{q}-\mathbf{x})/\|\mathbf{q}-\mathbf{x}\|^2 = \frac{\|\mathbf{x}\|^2 - \mathbf{q}^\top\mathbf{x}}{\|\mathbf{q}-\mathbf{x}\|^2} > 0$ by assumption. ∎

Our last lemma shows that if we have points in two orthogonal subspaces, A and $A^\perp$, then adding a point from $A^\perp$ to a set from A does not cause any points from A that previously did not violate Wolfe's criterion (for the affine minimizer) to violate it. Figure 1.9 demonstrates this situation.

LEMMA 1.6. *For a point* $\mathbf{z}$ *define* $H_\mathbf{z} = \{\mathbf{w} \in \mathbb{R}^n : \mathbf{w}^\top\mathbf{z} < \|\mathbf{z}\|^2\}$*. Suppose that we have an instance of the minimum norm point problem in* $\mathbb{R}^d$ *as follows: Some points, P, live in a proper linear subspace A and some points, Q, in $A^\perp$. Let* $\mathbf{x}$ *be the minimum norm point in* $\operatorname{aff}(P)$ *and* $\mathbf{y} \neq \mathbf{0}$ *be the minimum norm point in* $\operatorname{aff}(P \cup Q)$*. Then* $H_\mathbf{y} \cap A = H_\mathbf{x} \cap A$.

Proof. Let B be the span of $\mathbf{x}$ and Q. We first show $\mathbf{y} \in B$. To see this, suppose not. Decompose $\mathbf{y}$ as $\mathbf{y} = \lambda\mathbf{v} + \sum_{\mathbf{q}\in Q}\mu_q\mathbf{q}$, where $\mathbf{v} \in \operatorname{aff}(P)$ and $\lambda + \sum\mu_q = 1$. Decompose $\mathbf{v}$ as $\mathbf{v} = \mathbf{u} + \mathbf{x}$ where $\mathbf{u} \perp \mathbf{x}$ and $\mathbf{u} \in A$ (this is possible because $\mathbf{v} - \mathbf{x}$ is orthogonal to $\mathbf{x}$, by optimality of $\mathbf{x}$, Lemma 1.2). Thus, $\mathbf{y} = \lambda\mathbf{u} + \lambda\mathbf{x} + \sum_{\mathbf{q}\in Q}\mu_q\mathbf{q}$ with $\lambda\mathbf{u}$ orthogonal to $\lambda\mathbf{x} + \sum_{\mathbf{q}\in Q}\mu_q\mathbf{q}$. This implies that $\mathbf{y}' = \lambda\mathbf{x} + \sum_{\mathbf{q}\in Q}\mu_q\mathbf{q}$ has a smaller norm than $\mathbf{y}$ and $\mathbf{y}' \in \operatorname{aff}(P \cup Q)$. This is a contradiction.

To conclude, we have $H_\mathbf{y} \cap A$ is a halfspace in A whose normal is parallel to the projection of $\mathbf{y}$ onto A. (It is helpful to understand how to compute the intersection of a hyperplane with a subspace. If $T_\mathbf{g} = \{\mathbf{w} : \mathbf{w}\cdot\mathbf{g} = 1\}$ and S is a linear subspace, then $T_\mathbf{g} \cap S = \{\mathbf{w} \in S : \mathbf{w} \cdot \operatorname{proj}_S \mathbf{g} = 1\}$. In other words, in order to intersect a hyperplane with a subspace we project the normal.) That is, it is parallel to $\mathbf{x}$. But that halfspace must also contain $\mathbf{x}$ on its boundary. Thus, that halfspace is equal to $H_\mathbf{x} \cap A$.

Once we see that $\mathbf{y} \in B$, so $\mathbf{y} = \lambda \mathbf{x} + \sum_{\mathbf{q} \in Q} \mu_{\mathbf{q}} \mathbf{q}$, we can alternatively prove $H_{\mathbf{y}} \cap A = H_{\mathbf{x}} \cap A$ using direct computations. First, let $\mathbf{w} \in A$ and note that $\mathbf{w}^\top \mathbf{y} = \lambda \mathbf{w}^\top \mathbf{x} + \sum_{\mathbf{q} \in Q} \mu_{\mathbf{q}} \mathbf{w}^\top \mathbf{q} = \lambda \mathbf{w}^\top \mathbf{x}$ since $\mathbf{q} \in A^\perp$. Now, let $\mathbf{p} \in P$ and note that $\|\mathbf{y}\|^2 = \mathbf{p}^\top \mathbf{y}$ by Lemma 1.2 since $\mathbf{p} \in \operatorname{aff}(P \cup Q)$ and $\mathbf{y}$ is the affine minimizer of $P \cup Q$. Similarly, $\|\mathbf{x}\|^2 = \mathbf{p}^\top \mathbf{x}$. Finally, $\mathbf{p}^\top \mathbf{y} = \lambda \mathbf{p}^\top \mathbf{x} + \sum_{\mathbf{q} \in Q} \mu_{\mathbf{q}} \mathbf{p}^\top \mathbf{q} = \lambda \mathbf{p}^\top \mathbf{x}$ since $\mathbf{p} \in A$ and $\mathbf{q} \in A^\perp$, so we have $\|\mathbf{y}\|^2 = \mathbf{p}^\top \mathbf{y} = \lambda \mathbf{p}^\top \mathbf{x} = \lambda \|\mathbf{x}\|^2$. Thus, for $\mathbf{w} \in A$, we see $\mathbf{w}^\top \mathbf{y} < \|\mathbf{y}\|^2$ if and only if $\mathbf{w}^\top \mathbf{x} < \|x\|^2$, so $H_{\mathbf{y}} \cap A = H_{\mathbf{x}} \cap A$. ∎

We will now describe our example in detail. The simplest version of our construction uses square roots and real numbers. We present instead a version with a few additional tweaks so that it only involves rational numbers.

Let $P(1) = \{1\} \subseteq \mathbb{Q}$. For odd $d \geq 1$, let $P(d+2)$ be a list of points in $\mathbb{Q}^{d+2}$ defined inductively as follows: Let $\mathbf{o}_d^*$ denote the minimum norm point in $\operatorname{conv}(P(d))$. Let M_d be a value larger than or equal to the 2-norm of all points in $P(d)$. For a first reading, one can let M_d be the maximum 2-norm among points in $P(d)$, which leads to an instance that shows the exponential complexity but it is not necessarily rational. For a rational instance, one can take $M_d := \max_{\mathbf{p} \in P(d)} \|\mathbf{p}\|_1$, which is a rational upper bound to the maximum 2-norm among the points in $P(d)$. Similarly, let m_d be a positive value that is a lower bound to $\|\mathbf{o}_d^*\|$. Again, for a first reading one can define $m_d = \|\mathbf{o}_d^*\|$, which leads to an instance that shows the exponential complexity but it is not rational. For a rational instance, one can take $m_d = \|\mathbf{o}_d^*\|_\infty$, which is a rational lower bound to $\|\mathbf{o}_d^*\|$. If we want a worst-case lower bound in the Turing machine model (as a function of the bit-length of the input), then we not only need to ensure that the proposed instance has rational entries. We want an instance whose numerators and denominators are integers with binary encodings of length bounded by a polynomial in d. Choices that provably lead to this are $m_d = 1/4^d$ and $M_d = 2d$. This is analyzed in detail at the end of subsection 1.2.

We finally present the example. If we identify $P(d+2)$ with a matrix where the points are rows, then the points in $P(d+2)$ are given by the following block matrix:

$$P(d+2) = \begin{pmatrix} P(d) & 0 & 0 \\ \frac{1}{2}\mathbf{o}_d^* & \frac{m_d}{4} & M_d \\ \frac{1}{2}\mathbf{o}_d^* & \frac{m_d}{4} & -(M_d+1) \\ 0 & \frac{m_d}{4} & M_d+2 \\ 0 & \frac{m_d}{4} & -(M_d+3) \end{pmatrix}. \tag{1.1}$$

The last four rows of the matrix $P(d+2)$ are the points $\mathbf{p}_{d+2}, \mathbf{q}_{d+2}, \mathbf{r}_{d+2}, \mathbf{s}_{d+2}$ of the configuration. For a picture of the case of $P(3)$ see Figure 1.10. We also present the sets C, and points $\mathbf{x}$ and $\mathbf{y}$ defined throughout the iterations of Wolfe's method with the minnorm insertion rule on $P(3)$ in Table 1.3. For comparison only, we include the same for the iterations of Wolfe's method with the linopt insertion rule on $P(3)$ in Table 1.4. In this example, we choose $m_1 = \|\mathbf{o}_1^*\|_\infty$ and $M_1 = \|\mathbf{o}_1^*\|_1$ so that $P(3)$ is made up of the points $\mathbf{o}_1^* = (1, 0, 0)$, $\mathbf{p}_3 = (1/2, 1/4, 1)$, $\mathbf{q}_3 = (1/2, 1/4, -2)$, $\mathbf{r}_3 = (0, 1/4, 3)$, and $\mathbf{s}_3 = (0, 1/4, -4)$. Where appropriate, we have truncated the decimal approximations of the coordinates of the points $\mathbf{x}$ and $\mathbf{y}$.

THEOREM 1.7. *Consider the execution of Wolfe's method with the* minnorm *insertion rule on input $P(d)$ where d is odd and parameters satisfying for all $d' = 1, 3, \ldots, d$: $0 < m_{d'} \leq \|\mathbf{o}_{d'}^*\|$ and $M_{d'} \geq \|\mathbf{p}\|$ for all $\mathbf{p} \in P(d')$. Then the sequence of corrals has length $5 \cdot 2^{\frac{d-1}{2}} - 4$.*

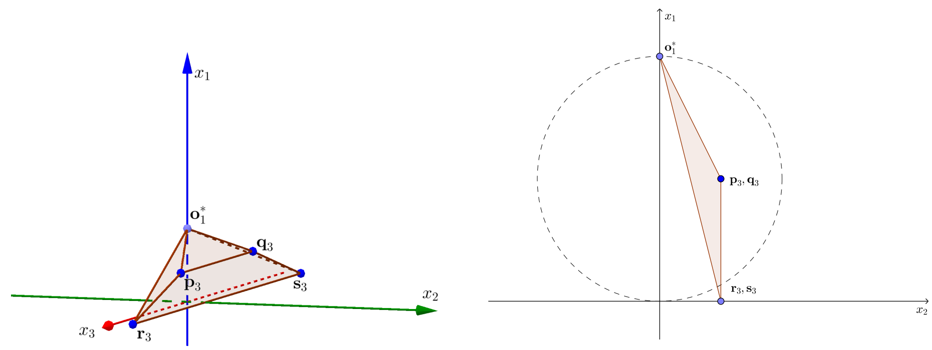


FIG. 1.10. *Left: Three-dimensional view of $P(3)$. Right: A two-dimensional view of $P(3)$ projected along the x_3 coordinate axis.*

TABLE 1.3

Iterations for minnorm *insertion rule on* $P(3) = \{\mathbf{0}^* = (1,0,0),\ \mathbf{p}_3 = (1/2,1/4,1),\ \mathbf{q}_3 = (1/2,1/4,-2),\ \mathbf{r}_3 = (0,1/4,3),\ \mathbf{s}_3 = (0,1/4,-4)\}$. *Cycle i,j denotes the ith major cycle and the jth minor cycle within.*

Cycle	C	$\mathbf{x}$	$\mathbf{y}$
0, 0	$\{\mathbf{o}^*\}$	$\mathbf{o}^*$	
1, 0	$\{\mathbf{o}^*, \mathbf{p}_3\}$	$\mathbf{o}^*$	$(0.810, 0.095, 0.381)$
2, 0	$\{\mathbf{o}^*, \mathbf{p}_3, \mathbf{q}_3\}$	$(0.810, 0.095, 0.381)$	$(0.2, 0.4, 0)$
2, 1	$\{\mathbf{p}_3, \mathbf{q}_3\}$	$(0.5, 0.25, 0.1875)$	$(0.5, 0.25, 0)$
3, 0	$\{\mathbf{p}_3, \mathbf{q}_3, \mathbf{r}_3\}$	$(0.5, 0.25, 0)$	$(0, 0.25, 0)$
3, 1	$\{\mathbf{q}_3, \mathbf{r}_3\}$	$(0.3, 0.25, 0)$	$(0.297, 0.25, 0.0297)$
4, 0	$\{\mathbf{q}_3, \mathbf{r}_3, \mathbf{s}_3\}$	$(0.297, 0.25, 0.0297)$	$(0, 0.25, 0)$
4, 1	$\{\mathbf{r}_3, \mathbf{s}_3\}$	$(0, 0.25, 0)$	$(0, 0.25, 0)$
5, 0	$\{\mathbf{r}_3, \mathbf{s}_3, \mathbf{o}^*\}$	$(0, 0.25, 0)$	$(0.059, 0.235, 0)$

Remark. First note that strictly speaking $P(d) \subset \mathbb{Q}^d$ and that we are defining an embedding of it into $\mathbb{Q}^{d+2}$, for which we have to use a recursive process. To avoid unnecessary indices in what follows, we will abuse the notation. The point $\mathbf{v}_d$ denotes a point both of $P(d)$ and of the subsequent $P(d+2)$, i.e., $\mathbf{v}_d = (\mathbf{v}, 0, 0)$ will be the identical copy of $\mathbf{v}_d$ within $P(d)$, but we add two extra zero coordinates. Depending on the context $\mathbf{v}_d$ will be understood as both a d-dimensional vector and as a $(d+2)$-dimensional vector (e.g., when doing dot products). The points of $P(d)$ become a subset of the point configuration $P(d+2)$ by padding extra zeros. See Figures 1.6 and 1.11, which illustrate this embedding and address our visualizations of these sets in three dimensions.

Proof. Points in $P(d)$ appear in (1.1) sorted by increasing norm, due to the choices of $M_{d'}$ for $d' = 1, 2, \ldots, d-2$. Let $\mathbf{p}_d, \mathbf{q}_d, \mathbf{r}_d, \mathbf{s}_d$ denote the last four points of $P(d)$, respectively. Let $C(d)$ denote the ordered sequence of corrals in the execution of Wolfe's method on $P(d)$. Let $O(d)$ denote the last (optimal) corral in $C(d)$.

Table 1.4
Iterations for linopt *insertion rule on* $P(3) = \{\mathbf{0}^* = (1,0,0),\ \mathbf{p}_3 = (1/2,1/4,1),\ \mathbf{q}_3 = (1/2,1/4,-2),\ \mathbf{r}_3 = (0,1/4,3),\ \mathbf{s}_3 = (0,1/4,-4)\}$. *Cycle* i,j *denotes the* i*th major cycle and the* j*th minor cycle within.*

Cycle	C	$\mathbf{x}$	$\mathbf{y}$
0, 0	$\{\mathbf{o}^*\}$	$\mathbf{o}^*$	
1, 0	$\{\mathbf{o}^*, \mathbf{r}_3\}$	$\mathbf{o}^*$	$(0.901, 0.025, 0.298)$
2, 0	$\{\mathbf{o}^*, \mathbf{r}_3, \mathbf{s}_3\}$	$(0.901, 0.025, 0.298)$	$(0.059, 0.235, 0)$

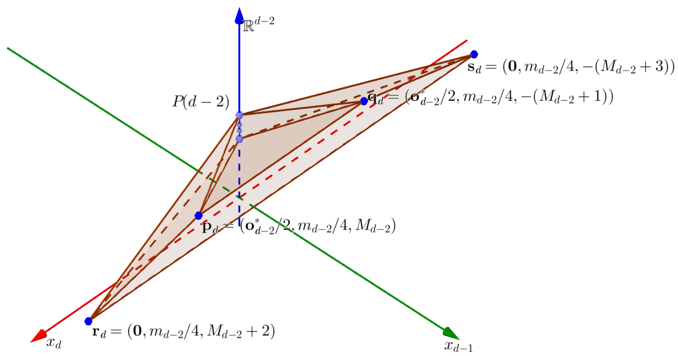


Fig. 1.11. *As described in Figure* 1.6*, the axis labeled* $\mathbb{R}^{d-2}$ *represents the* $(d-2)$*-dimensional subspace* span $(P(d-2))$ *projected onto the one-dimensional subspace* span $(\mathbf{o}^*_{d-2})$*. Here we illustrate that the projection of the set* $P(d-2)$ *forms a "cloud" of points and the convex hull of this projection has many fewer faces than the unprojected convex hull. For simplicity, we will visualize* $P(d-2)$ *and subsets of* $P(d-2)$ *as a single point in* span $(\mathbf{o}^*_{d-2})$ *as in Figure* 1.6*.*

The rest of the proof will establish that the sequence of corrals $C(d)$ is

$$\begin{aligned} C(d-2),\\ O(d-2)\mathbf{p}_d,\\ \mathbf{p}_d\mathbf{q}_d,\\ \mathbf{q}_d\mathbf{r}_d,\\ \mathbf{r}_d\mathbf{s}_d,\\ C(d-2)\mathbf{r}_d\mathbf{s}_d \end{aligned}$$

(where a concatenation such as $C(d-2)\mathbf{r}_d\mathbf{s}_d$ denotes every corral in $C(d-2)$ with $\mathbf{r}_d$ and $\mathbf{s}_d$ added). After this sequence of corrals is established, we solve the resulting recurrence relation: Let $T(d)$ denote the length of $C(d)$. We have $T(1) = 1$, $T(d) = 2T(d-2) + 4$. This implies $T(d) = 5 \cdot 2^{\frac{d-1}{2}} - 4$ (with d odd).

All we must show now to complete the proof of Theorem 1.7 is that $C(d)$ has indeed the stated recursive form. We do this by induction on d. The steps of the proof are written as claims with individual proofs.

By construction, $C(d)$ starts with $C(d-2)$. This happens because points in $C(d)$ are ordered by increasing norm and the proof proceeds inductively as follows: The first corral in $C(d)$ is the minimum norm point in $P(d)$ (equal to the minimum norm point in $P(d-2)$), which is also the first corral in $C(d-2)$. Suppose now that the first t corrals of $C(d)$ coincide with the first t corrals of $C(d-2)$. We will show that corral $t+1$ in $C(d)$ is the same as corral $t+1$ in $C(d-2)$. To see this, it is enough to see that the set of points in $P(d)$ that can enter (improving points) contains the point that enters in $C(d-2)$ (with two zeros appended) and contains no point of smaller norm. This two-part claim is true because the two new zero coordinates play no role in this and points in $P(d) \setminus P(d-2) = \{\mathbf{p}_d, \mathbf{q}_d, \mathbf{r}_d, \mathbf{s}_d\}$ have a larger norm than any other point in $P(d)$.

Once corral $O(d-2)$ is reached (with minimum norm point $\mathbf{o}^*_{d-2}$), the set of improving points, as established by Wolfe's criterion, consists of $\{\mathbf{p}_d, \mathbf{q}_d, \mathbf{r}_d, \mathbf{s}_d\}$, since $\mathbf{p}_d^\top \mathbf{o}^*_{d-2} = \mathbf{q}_d^\top \mathbf{o}^*_{d-2} = \frac{1}{2}\|\mathbf{o}^*_{d-2}\|^2 < \|\mathbf{o}^*_{d-2}\|^2$ and $\mathbf{r}_d^\top \mathbf{o}^*_{d-2} = \mathbf{s}_d^\top \mathbf{o}^*_{d-2} = 0 < \|\mathbf{o}^*_{d-2}\|^2$. Now, because we are using the minimum-norm insertion rule, the next point to enter is $\mathbf{p}_d$.

Claim 1.8. *$O(d-2)\mathbf{p}_d$ is a corral.*

Proof of Claim. This is a special case of Lemma 1.5. The coordinates of point $\mathbf{p}_d$ are $\mathbf{p}_d = (\mathbf{o}^*_{d-2}/2, m_{d-2}/4, M_{d-2})$. We just need to verify the two inequalities in Lemma 1.5: $(\mathbf{o}^*_{d-2})^\top \mathbf{p}_d = \|\mathbf{o}^*_{d-2}\|^2/2 < \|\mathbf{p}_d\|^2$. ∎

Claim 1.9. *The next improving point to enter is $\mathbf{q}_d$.*

Proof of Claim. We first check that no point in $P(d-2)$ can enter. From Lemma 1.5 we know the optimal point $\mathbf{y}$ in corral $O(d-2)\mathbf{p}_d$ explicitly in terms of the optimal point $\mathbf{o}^*_{d-2}$ of $O(d-2)$ and $\mathbf{p}_d$, namely $\mathbf{y}$ is a convex combination $\lambda \mathbf{o}^*_{d-2} + (1-\lambda)\mathbf{p}_d$, with $\lambda = \frac{\|\mathbf{p}_d\|^2 - {\mathbf{p}_d}^\top \mathbf{o}^*_{d-2}}{\|\mathbf{p}_d - \mathbf{o}^*_{d-2}\|^2}$. Let $\mathbf{p} \in P(d-2)$. We check that it cannot enter via Wolfe's criterion. We compute $\mathbf{p}^\top \mathbf{y}$ in two steps: First project $\mathbf{p}$ onto span $(\mathbf{o}^*_{d-2}, \mathbf{p}_d)$ (a subspace that contains $\mathbf{y}$). This projection is longer than $\mathbf{o}^*_{d-2}$ by optimality of $\mathbf{o}^*_{d-2}$. Then project onto $\mathbf{y}$. This shows that $\mathbf{p}^\top \mathbf{y} \geq {\mathbf{o}^*_{d-2}}^\top \mathbf{y} = \|\mathbf{y}\|^2$ and $\mathbf{p}$ cannot enter as it is not an improving point according to Wolfe's criterion.

By construction, $\mathbf{q}_d$ is closer to the origin than $\mathbf{r}_d, \mathbf{s}_d$, so to conclude it is enough to check that $\mathbf{q}_d$ is an improving point per Wolfe's criterion. Compute

$$
\begin{aligned}
\mathbf{y}^\top \mathbf{q}_d &= \lambda(\mathbf{o}^*_{d-2})^\top \mathbf{q}_d + (1-\lambda)\mathbf{p}_d^\top \mathbf{q}_d \\
&\leq \frac{\lambda}{2}\|\mathbf{o}^*_{d-2}\|^2 + (1-\lambda)\left[\frac{1}{4}\|\mathbf{o}^*_{d-2}\|^2 + \frac{1}{16}\|\mathbf{o}^*_{d-2}\|^2 - M_{d-2}^2 - M_{d-2}\right] \\
&\leq \frac{\lambda}{2}\|\mathbf{o}^*_{d-2}\|^2
\end{aligned}
$$

since by construction $M_{d-2} \geq 1$ and $m_{d-2} < \|\mathbf{o}^*_{d-2}\| \leq \|\mathbf{p}_1\| = 1$. On the other hand,

$$
\begin{aligned}
\|\mathbf{y}\|^2 &= \lambda^2\|\mathbf{o}^*_{d-2}\|^2 + (1-\lambda)^2\|\mathbf{p}_d\|^2 + 2\lambda(1-\lambda)\frac{1}{2}\|\mathbf{o}^*_{d-2}\|^2 \\
&= \lambda\|\mathbf{o}^*_{d-2}\|^2 + (1-\lambda)^2\|\mathbf{p}_d\|^2 \\
&\geq \lambda\|\mathbf{o}^*_{d-2}\|^2.
\end{aligned}
$$

Thus, $\mathbf{y}^\top \mathbf{q}_d < \|\mathbf{y}\|^2$, that is, $\mathbf{q}_d$ is an improving point. ∎

Claim 1.10. *The current set of points, $O(d-2) \cup \{\mathbf{p}_d, \mathbf{q}_d\}$, is not a corral. Points in $O(d-2)$ leave one by one. The next corral is $\mathbf{p}_d\mathbf{q}_d$.*

Proof of Claim. Instead of analyzing the iterations of Wolfe's inner loop, we use the key fact, from subsection 1.1, that the inner loop must end with a corral whose distance to the origin is strictly less than the previous corral. We look at the alternatives:

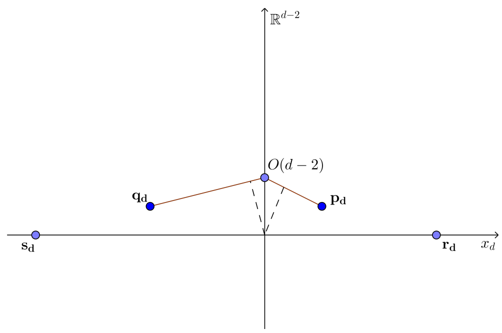


FIG. 1.12. *A projection of the point set in the direction of x_{d-1}. Any corral of the form $S\mathbf{q}_d$ where $S \subset O(d-2)$ would have distance larger than the previous corral, $O(d-2)\mathbf{p}_d$.*

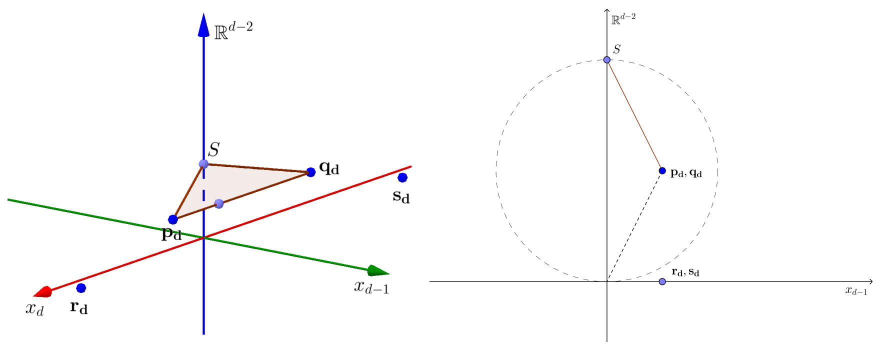


FIG. 1.13. *The minimum norm point in* $\operatorname{conv}(S \cup \{\mathbf{p}_d, \mathbf{q}_d\})$ *is in the line segment between $\mathbf{p}_d$ and $\mathbf{q}_d$.*

This new corral cannot be $O(d-2) \cup \{\mathbf{p}_d\}$ (the previous corral) or any subset of it because it would not decrease the distance. Since $(\mathbf{o}^*_{d-2})^\top \mathbf{q}_d < \|\mathbf{o}^*_{d-2}\|^2/2 < \|\mathbf{o}^*_{d-2}\|^2 < \|\mathbf{q}_d\|^2$, we have by Lemma 1.5 that $O(d-2) \cup \{\mathbf{q}_d\}$ is a corral whose distance to the origin is larger than the distance for $O(d-2) \cup \{\mathbf{p}_d\}$. See Figure 1.12, where we show a projection, and the perpendicular line segments to $\operatorname{conv}(O(d-2), \mathbf{p}_d)$ and $\operatorname{conv}(O(d-2), \mathbf{q}_d)$ are shown in dotted line after projection. Thus, the new corral cannot be $O(d-2) \cup \{\mathbf{q}_d\}$ or any subset of it.

No set of the form $S \cup \{\mathbf{p}_d, \mathbf{q}_d\}$ with $S \subseteq O(d-2)$ and S nonempty can be a corral. To see this, first note that the minimum norm point in $\operatorname{conv}(S \cup \{\mathbf{p}_d, \mathbf{q}_d\})$ is in the segment $[\mathbf{p}_d, \mathbf{q}_d]$, specifically, point $(\mathbf{o}^*_{d-2}/2, m_{d-2}/4, 0)$ (minimality follows from Wolfe's criterion, Lemma 1.1). This implies that the minimum norm point in $\operatorname{aff}(S \cup \{\mathbf{p}_d, \mathbf{q}_d\})$ cannot be in the relative interior of $\operatorname{conv}(S \cup \{\mathbf{p}_d, \mathbf{q}_d\})$ when S is nonempty (see Figure 1.13).

The only remaining nonempty subset is $\{\mathbf{p}_d, \mathbf{q}_d\}$, which is the new corral. ∎

CLAIM 1.11. *The set of improving points is now $\{\mathbf{r}_d, \mathbf{s}_d\}$.*

Proof of Claim. Recall that the optimal point in corral $\{\mathbf{p}_d, \mathbf{q}_d\}$ has coordinates $(\mathbf{o}^*_{d-2}/2, m_{d-2}/4, 0)$. Thus, when computing distances and checking Wolfe's criterion it is enough to do so in the two-dimensional situation depicted in Figure 1.14. Thus, a hyperplane orthogonal to the segment from the origin to $(\mathbf{o}^*_{d-2}/2, m_{d-2}/4, 0)$ is shown in the figure. It leaves the points in $P(d-2)$ above (as $\mathbf{z} \in P(d-2)$

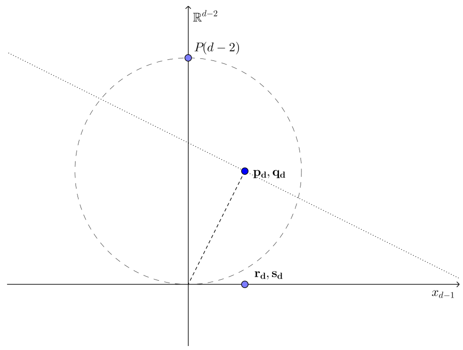


FIG. 1.14. *The set of improving points is now* $\{\mathbf{r}_d, \mathbf{s}_d\}$.

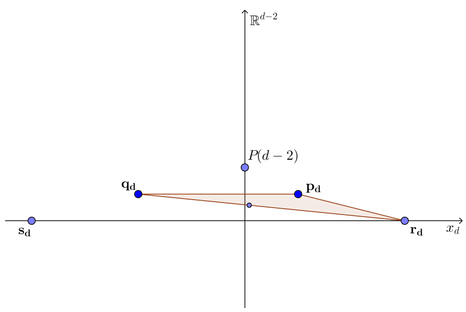


FIG. 1.15. *The set* $\{\mathbf{p}_d, \mathbf{q}_d, \mathbf{r}_d\}$ *is not a corral.*

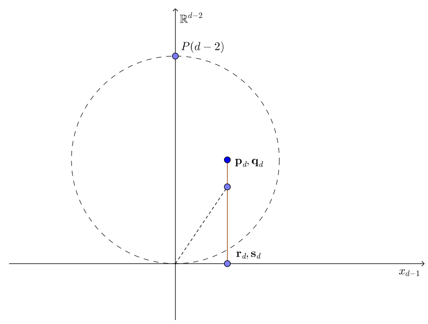


FIG. 1.16. *The only improving point is* $\mathbf{s}_d$.

satisfies $\mathbf{z}^\top(\mathbf{o}^*_{d-2}/2, m_{d-2}/4, 0) = \mathbf{z}^\top\mathbf{o}^*_{d-2}/2 > \|\mathbf{o}^*_{d-2}\|^2/2 > \|(\mathbf{o}^*_{d-2}/2, m_{d-2}/4, 0)\|^2)$ and both $\mathbf{r}_d$ and $\mathbf{s}_d$ below making them the only improving points. Alternatively, note that $\mathbf{r}_d^\top(\mathbf{o}^*_{d-2}/2, m_{d-2}/4, 0) = \mathbf{s}_d^\top(\mathbf{o}^*_{d-2}/2, m_{d-2}/4, 0) = \frac{1}{16}m^2_{d-2} < \|(\mathbf{o}^*_{d-2}/2, m_{d-2}/4, 0)\|^2$, so $\mathbf{r}_d$ and $\mathbf{s}_d$ are available. ∎

Point $\mathbf{r}_d$ enters since it has smallest norm.

CLAIM 1.12. *Point* $\mathbf{p}_d$ *leaves and the next corral is* $\mathbf{q}_d\mathbf{r}_d$.

Proof of Claim. To start, notice that by construction the four points $\mathbf{p}_d, \mathbf{q}_d, \mathbf{r}_d, \mathbf{s}_d$ lie on a common hyperplane, L, parallel to the line spanned by $\mathbf{o}^*_{d-2}$. Thus, one does not need to do distance calculations but rather Figure 1.15 is a faithful representation of the positions of points. The closest point to the origin within L is in the line segment joining $\mathbf{r}_d, \mathbf{s}_d$; thus, as we move parallel to $\mathbf{o}^*_{d-2}$ in L, the closest point to the origin in triangle $\mathbf{p}_d, \mathbf{q}_d, \mathbf{r}_d$ must be in the line segment joining $\mathbf{r}_d$ and $\mathbf{q}_d$. ∎

CLAIM 1.13. *The only improving point now is* $\mathbf{s}_d$.

Proof of Claim. Figures 1.15 and 1.16 provide geometric intuition that the point $\mathbf{s}_d$ is available to add while $\mathbf{p}_d$ and $P(d-2)$ are not. We provide a full proof next.

First, let $\mathbf{x}$ denote the minimum norm point in the line segment joining $\mathbf{r}_d$ and $\mathbf{q}_d$ so we can write $\mathbf{x} = (\alpha\mathbf{o}^*_{d-2}, m_{d-2}/4, \epsilon)$. Furthermore, $\mathbf{x} = \lambda\mathbf{q}_d + (1-\lambda)\mathbf{r}_d$ where

$$\lambda = \frac{\mathbf{r}_d^\top(\mathbf{r}_d - \mathbf{q}_d)}{\|\mathbf{r}_d - \mathbf{q}_d\|^2} = \frac{(M_{d-2}+2)(2M_{d-2}+3)}{\|\mathbf{o}^*_{d-2}\|^2/4 + (2M_{d-2}+3)^2}$$

by Proposition 1.3. Note that we have $\alpha = \lambda/2$. Since $\mathbf{x}^\top(\mathbf{r}_d - \mathbf{q}_d) = 0$, we have $\epsilon = \frac{\alpha\|\mathbf{o}^*_{d-2}\|^2/2}{2M_{d-2}+3}$.

Now, we develop estimates on α and ϵ. First, note that

$$\lambda = \frac{(M_{d-2}+2)(2M_{d-2}+3)}{\|\mathbf{o}^*_{d-2}\|^2/4 + (2M_{d-2}+3)^2} \leq \frac{(M_{d-2}+2)(2M_{d-2}+3)}{(2M_{d-2}+3)^2} = \frac{M_{d-2}+2}{2M_d+3} \leq \frac{3}{5}$$

since $M_{d-2} \geq 1$. Thus, $\alpha = \lambda/2 \leq 3/10$. Additionally, since $\|\mathbf{o}^*_{d-2}\|^2/4 \leq 1/4$ and $(2M_{d-2}+3)^2 \geq 25$, we have $\|\mathbf{o}^*_{d-2}\|^2/4 \leq (2M_{d-2}+3)^2/100$. This yields

$$\lambda = \frac{(M_{d-2}+2)(2M_{d-2}+3)}{\|\mathbf{o}^*_{d-2}\|^2/4 + (2M_{d-2}+3)^2} \geq \frac{100(M_{d-2}+2)}{101(2M_{d-2}+3)} \geq \frac{100(M_{d-2}+3/2)}{101(2M_{d-2}+3)} = \frac{50}{101}.$$

Thus, we have $\alpha = \lambda/2 \geq 25/101$ and $\epsilon \leq 3\|\mathbf{o}^*_{d-2}\|^2/100$.

Now, note that

$$\begin{aligned}
\|\mathbf{x}\|^2 &= \alpha^2\|\mathbf{o}^*_{d-2}\|^2 + m^2_{d-2}/16 + \epsilon^2 \\
&\leq \frac{9}{100}\|\mathbf{o}^*_{d-2}\|^2 + \frac{1}{16}\|\mathbf{o}^*_{d-2}\|^2 + \left(\frac{3}{100}\|\mathbf{o}^*_{d-2}\|^2\right)^2 \\
&< \frac{9}{100}\|\mathbf{o}^*_{d-2}\|^2 + \frac{1}{10}\|\mathbf{o}^*_{d-2}\|^2 + \frac{3}{100}\|\mathbf{o}^*_{d-2}\|^2 = \frac{22}{100}\|\mathbf{o}^*_{d-2}\|^2.
\end{aligned}$$

Thus, for $\mathbf{p} \in P(d-2)$, we have

$$\mathbf{p}^\top\mathbf{x} = \alpha\mathbf{p}^\top\mathbf{o}^*_{d-2} \geq \alpha\|\mathbf{o}^*_{d-2}\|^2 \geq \frac{25}{101}\|\mathbf{o}^*_{d-2}\|^2 > \frac{22}{100}\|\mathbf{o}^*_{d-2}\|^2 > \|\mathbf{x}\|^2,$$

so $\mathbf{p}$ is not available. Similarly,

$$\mathbf{p}_d^\top\mathbf{x} = \frac{\alpha}{2}\|\mathbf{o}^*_{d-2}\|^2 + \frac{m^2_{d-2}}{16} + M_{d-2}\epsilon \geq \alpha^2\|\mathbf{o}^*_{d-2}\|^2 + \frac{m^2_{d-2}}{16} + \epsilon^2 = \|\mathbf{x}\|^2,$$

where the inequality follows from $\alpha < 1/2$, $M_{d-2} \geq 1$, and $\epsilon < 1$. Thus, $\mathbf{p}_d$ is not available. Finally,

$$\mathbf{s}_d^\top\mathbf{x} = \frac{m^2_{d-2}}{16} - (M_{d-2}+3)\epsilon < \frac{m^2_{d-2}}{16} < \|\mathbf{x}\|^2,$$

where the first inequality follows since $M_{d-2}, \epsilon > 0$. Thus, $\mathbf{s}_d$ is available. ∎

Point $\mathbf{s}_d$ enters as the closest improving point to the origin.

Claim 1.14. *Point $\mathbf{q}_d$ leaves. The next corral is $\mathbf{r}_d\mathbf{s}_d$.*

Proof of Claim. We follow the execution of the algorithm. It finds the affine minimizer for $C = \{\mathbf{q}_d, \mathbf{r}_d, \mathbf{s}_d\}$ (the minimum norm point in their affine hull, $\mathbf{y}$ in Method 1.1). It is between $\mathbf{r}_d$ and $\mathbf{s}_d$; this point is $(m_{d-2}/4)\mathbf{e}_{d-1}$ (where $\mathbf{e}_{d-1}$ is the $(d-1)$st canonical vector). While this point is in the convex hull of C, it is not a strict convex combination; see Figure 1.17 for a visualization. Point $\mathbf{q}_d$ is not used in the convex combination and therefore leaves C. The resulting set, $\{\mathbf{r}_d, \mathbf{s}_d\}$, is the next corral. ∎

Claim 1.15. *The set of improving points is now $P(d-2)$ (with two zero coordinates appended).*

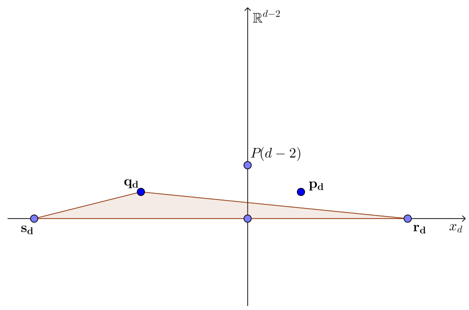


FIG. 1.17. *The point* $\mathbf{q}_d$ *leaves.*

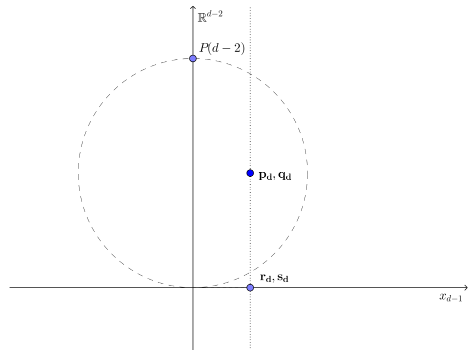


FIG. 1.18. *The improving points are* $P(d-2)$.

Proof of Claim. Now Wolfe's criterion hyperplane contains the four points $\mathbf{p}_d$, $\mathbf{q}_d$, $\mathbf{r}_d$, $\mathbf{s}_d$ by construction leaving $P(d-2)$ on the same side as the origin (see Figure 1.18). Note that for all $\mathbf{p} \in P(d-2)$ we have $\mathbf{p}^\top (m_{d-2}/4)\mathbf{e}_{d-1} = 0 < \|(m_{d-2}/4)\mathbf{e}_{d-1}\|^2$. $\square$

The first (and minimum norm) point in $P(d)$ enters and the next corral is this point together with $\mathbf{r}_d$ and $\mathbf{s}_d$. That is, the next corral is precisely the first corral in $C(d-2)\mathbf{r}_d\mathbf{s}_d$. This is a corral by Lemma 1.4. We will prove inductively that the sequence of corrals from now on is exactly all of $C(d-2)\mathbf{r}_d\mathbf{s}_d$. To see this, we repeatedly invoke Lemma 1.6 after every corral with A equal to the subspace spanned by the first $d-2$ coordinate vectors of $\mathbb{R}^d$. Suppose that the current corral is $C\mathbf{r}_d\mathbf{s}_d$, where C is one of the corrals in $C(d-2)$, and denote the next corral in $C(d-2)$ by C'. Applying Lemma 1.6 with $P = C$ and $Q = \{\mathbf{r}_d, \mathbf{s}_d\}$, we get that the set of improving points for corral $C\mathbf{r}_d\mathbf{s}_d$ contains the set of improving points for corral C. Thus, by the use of the minnorm rule, the point that enters is the same that would enter after corral C. Let $\mathbf{a}$ denote that point.

CLAIM 1.16. *The next corral is* $C'\mathbf{r}_d\mathbf{s}_d$.

Proof of Claim. The current set of points is $C\mathbf{r}_d\mathbf{s}_d\mathbf{a}$. If $C\mathbf{a}$ is a corral, then so is $C\mathbf{r}_d\mathbf{s}_d\mathbf{a} = C'\mathbf{r}_d\mathbf{s}_d$ (by Lemma 1.4, part 3, with $P = C\mathbf{a}$ and $Q = \mathbf{r}_d\mathbf{s}_d$) and the claim holds. If $C\mathbf{a}$ is not a corral, it is enough to prove that the sequence of points removed by the inner loop of Wolfe's method on this set is the same as the sequence on set $C\mathbf{r}_d\mathbf{s}_d\mathbf{a}$. We will show this now by simultaneously analyzing the execution of the inner loop of Wolfe's method (Method 1.1) on $C\mathbf{a}$ and $C\mathbf{r}_d\mathbf{s}_d\mathbf{a}$. We distinguish the two cases with the following notation: variables are written without a bar ($^-$) and with a bar, respectively.

Let $C_0 = C\mathbf{a}$ and $\mathbf{x}_0$ be the point $\mathbf{x}$ defined as at line 5 of Method 1.1 prior to entering the inner loop. Let $\mathbf{x}_1, \dots, \mathbf{x}_k$ be the sequence of current points constructed by the inner loop on $C\mathbf{a}$ at line 10 of Method 1.1. Let $\mathbf{p}_1, \dots, \mathbf{p}_k$ be the sequence of removed points defined at line 8 of Method 1.1. Let $C_1, \dots, C_k$ be the sequence of current sets of points at every iteration defined at line 9 of Method 1.1. Let $\bar{\mathbf{x}}_0, \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_{\bar{k}}$ be the corresponding sequence on $C\mathbf{r}_d\mathbf{s}_d\mathbf{a}$. Let $\bar{\mathbf{p}}_1, \dots, \bar{\mathbf{p}}_{\bar{k}}$ be the corresponding sequence of removed points. Let $\bar{C}_0, \bar{C}_1, \dots, \bar{C}_{\bar{k}}$ be the corresponding sequence of current sets of points. We will show inductively that $k = \bar{k}$, there is a one-to-one correspondence between sequences $(\mathbf{x}_i)$ and $(\bar{\mathbf{x}}_i)$, and $(\mathbf{p}_i) = (\bar{\mathbf{p}}_i)$. More specifically, the correspondence is realized by maintaining the following invariant in the inner loop: $\bar{\mathbf{x}}_i$ is a strict convex combination of $\mathbf{x}_i$ and the minimum norm point in $[\mathbf{r}_d, \mathbf{s}_d]$.

For the base case, from Lemma 1.4, part 2, we have that $\bar{\mathbf{x}}_0$ is a strict convex combination of $\mathbf{x}_0$ (which is the minimum norm point in $\operatorname{conv}(C)$) and the minimum norm point in segment $[\mathbf{r}_d, \mathbf{s}_d]$, specifically $\mathbf{w} := \frac{m_{d-2}}{4}\mathbf{e}_{d-1}$.

For the inductive step, if $\mathbf{x}_i$ is a strict convex combination of the current set of points C_i, then so is $\bar{\mathbf{x}}_i$ of $\bar{C}_i$ and the inner loop ends in both cases with corrals $C_i = C'$ and $\bar{C}_i = C'\mathbf{r}_d\mathbf{s}_d$, respectively. The claim holds. If $\mathbf{x}_i$ is not a strict convex combination of the current set of points C_i, then neither is $\bar{\mathbf{x}}_i$ of $\bar{C}_i$, since it inherits this property from $\mathbf{x}_i$ by Lemma 1.4. The inner loop then continues by computing the minimum norm point $\mathbf{y}$ in $\operatorname{aff}(C_i)$ and $\bar{\mathbf{y}}$ in $\operatorname{aff}(\bar{C}_i)$, respectively. It then finds point $\mathbf{z}$ in $\operatorname{conv}(C_i)$ that is closest to $\mathbf{y}$ in segment $[\mathbf{x}_i, \mathbf{y}]$. It finds $\bar{\mathbf{z}}$, respectively. It then selects a point $\mathbf{p}_i$ to be removed and a point $\bar{\mathbf{p}}_i$, respectively. From Lemma 1.4, part 2, we have that $\bar{\mathbf{y}}$ is a strict convex combination of $\mathbf{y}$ and $\mathbf{w}$.

We will argue that $\bar{\mathbf{z}}$ is a strict convex combination of $\mathbf{z}$ and $\mathbf{w}$. To see this, we note that segment $[\bar{\mathbf{x}}_i, \bar{\mathbf{y}}]$ lies in the hyperplane where the last coordinate is 0. Therefore we only need to intersect it with the part of $\operatorname{conv}(\bar{C}_i)$ that lies in that hyperplane. This part is exactly $\operatorname{conv}(C_i \cup \{\mathbf{w}\})$, which can be written in a more explicit way as the union of all segments of the form $[\mathbf{b}, \mathbf{w}]$ with $\mathbf{b} \in C_i$. Even more, we only need to look at triangle $\mathbf{w}, \mathbf{x}_i, \mathbf{y}$, as all relevant segments lie on it. The intersection of this triangle with $\operatorname{conv}(C_i)$ is segment $[\mathbf{x}_i, \mathbf{z}]$ and therefore the intersection of the triangle with $\operatorname{conv}(\bar{C}_i)$ is simply triangle $\mathbf{x}_i, \mathbf{z}, \mathbf{w}$. This implies that the intersection between segment $[\bar{\mathbf{x}}_i, \bar{\mathbf{y}}]$ and $\operatorname{conv}(\bar{C}_i)$ is the same as the intersection between segment $[\bar{\mathbf{x}}_i, \bar{\mathbf{y}}]$ and triangle $\mathbf{x}_i, \mathbf{z}, \mathbf{w}$. This intersection is an interval $[\bar{\mathbf{x}}_i, \bar{\mathbf{z}}]$ where $\bar{\mathbf{z}}$ is a strict convex combination of $\mathbf{w}$ and $\mathbf{z}$ and $\bar{\mathbf{z}}$ is the closest point to $\bar{\mathbf{y}}$ in that intersection.

It follows that the set of potential points to be removed is the same for the two executions. Specifically, if $\mathbf{z}$ is a strict convex combination of a certain subset C^* of C_i, then $\bar{\mathbf{z}}$ is a strict convex combination of $C^* \cup \{\mathbf{r}_d, \mathbf{s}_d\}$. The sets of points that can potentially be removed are $C_i \setminus C^*$ and $\bar{C}_i \setminus (C^* \cup \{\mathbf{r}_d, \mathbf{s}_d\}) = C_i \setminus C^*$ (the same), respectively. In particular, $\mathbf{p}_i = \bar{\mathbf{p}}_i$ (under a mild consistency assumption on the way a point is chosen when there is more than one choice, for example, "choose the point with smallest index among potential points"). This implies $C_{i+1} = \bar{C}_{i+1}$. Also, $\mathbf{x}_{i+1} = \mathbf{z}$ and $\bar{\mathbf{x}}_{i+1} = \bar{\mathbf{z}}$ is in $[\mathbf{x}_{i+1}, \mathbf{w}]$. This completes the inductive argument about the inner loop and proves the claim. ∎

At this point the current corral is $O(d-2)\mathbf{r}_d\mathbf{s}_d$. To conclude, we need to show that this is the optimal corral according to Wolfe's criterion (Lemma 1.1). From Lemma 1.4 applied to $O(d-2)$ and $\{\mathbf{r}_d, \mathbf{s}_d\}$, the minimum norm point (convex minimizer) for the current corral is a strict convex combination of $(\mathbf{o}^*_{d-2}, 0, 0)$ and $\frac{m_{d-2}}{4}\mathbf{e}_{d-1}$. That is, it can be written in the form $\mathbf{x} = \lambda(\mathbf{o}^*_{d-2}, 0, 0) + (1-\lambda)\frac{m_{d-2}}{4}\mathbf{e}_{d-1}$, $0 < \lambda < 1$. No point in $P(d-2)$ can enter: if $\mathbf{y} \in P(d-2)$, then $\mathbf{y}^\top \mathbf{x} = \lambda \mathbf{y}^\top(\mathbf{o}^*_{d-2}, 0, 0) \geq \lambda \|\mathbf{o}^*_{d-2}\|^2 = (\mathbf{o}^*_{d-2}, 0, 0)^\top \mathbf{x} = \|\mathbf{x}\|^2$ (where the inequality is by optimality of $\mathbf{o}^*_{d-2}$ for $P(d-2)$ and the last equality is by optimality of $\mathbf{x}$ for the current corral). Similarly, $\mathbf{p}_d, \mathbf{q}_d$ cannot enter: by definition $\mathbf{p}_d^\top \mathbf{x} = \mathbf{q}_d^\top \mathbf{x} = \frac{\lambda}{2}\|\mathbf{o}^*_{d-2}\|^2 + (1-\lambda)\frac{m^2_{d-2}}{16} > (1-\lambda)\frac{m^2_{d-2}}{16} = \mathbf{r}_d^\top \mathbf{x} = \|\mathbf{x}\|^2$. Thus, no point can enter and the current corral is optimal.

This completes the proof of Theorem 1.7. ∎

Theorem 1.7 implies that Wolfe's method has exponential worst-case running time on rational inputs in an arithmetic model of computation. For this claim one can specifically use the integer RAM model discussed at the beginning of section 2 and understand the running time as a function of the number of numbers in the input. To conclude that it is exponential in the Turing machine model (as a function of the bit-length of the input, not just the dimension or the number of points), one also needs to argue that the numbers that appear in our hard instances $P(d)$ are not too large. It is enough to show that the bit-lengths of the numbers in our hard instance $P(d)$ (equation (1.1)) are bounded by a polynomial in d. We show this in two steps: We first make explicit, nonrecursive choices of m_d, M_d that grow at most exponentially in d (which makes their bit-lengths grow at most polynomially in d). This bounds the bit-lengths of entries in (1.1) involving m_d and M_d. We then argue that $\mathbf{o}_d^*$ has polynomial bit-length by expressing it nonrecursively as the affine minimizer of $O(d-2)$, whose points are explicitly defined in terms of m_d and M_d. This bounds the bit-lengths of entries in (1.1) involving $\mathbf{o}_d^*$. We make this argument precise in Theorem 1.17 below. Let the size of a rational number, represented by a/b with $a, b \in \mathbb{Z}$, $b \geq 1$, be $\mathrm{size}(a/b) := 1 + \lceil \log(|a|+1) \rceil + \lceil \log b \rceil$.

THEOREM 1.17. *Let $m_d = 1/4^d$ and $M_d = 2d$. Then the* size *of every number in $P(d)$ is bounded by a (universal) polynomial in d. Moreover, m_d, M_d as above satisfy the assumptions of Theorem* 1.7.

Proof. We first show that the choices of parameters m_d, M_d lead to a valid instance (satisfy the conditions of Theorem 1.7). We proceed by induction on odd d. The maximum 2-norm among points in $P(d)$ is $\|\mathbf{s}_d\|$. We have $M_1 = 2$ and $\|\mathbf{s}_1\| = 1$, which gives the base case. We also have $M_d = 2d$ and $\|\mathbf{s}_d\| \leq \|\mathbf{s}_d\|_1 < M_{d-2} + 4$, which gives the inductive step. The desired bound on M_d follows.

Similarly for m_d, we find estimates on $\mathbf{o}_d^*$. We have $\mathbf{o}_1^* = 1$. From the proof of Theorem 1.7, the optimal corral for $P(d)$ is $O(d) = O(d-2)\mathbf{r}_d\mathbf{s}_d$ so that $O(1) = 1$, $O(d) = 1, \mathbf{r}_3, \mathbf{s}_3, \mathbf{r}_5, \mathbf{s}_5, \ldots, \mathbf{r}_d, \mathbf{s}_d$. We just need to determine the minimum norm point in this corral. We will use Lemma 1.4 to determine this point. We have that $O(d-2)$ and $\mathbf{r}_d\mathbf{s}_d$ are corrals contained in orthogonal subspaces with minimum norm points $(\mathbf{o}_{d-2}^*, 0, 0)$ and $\frac{m_{d-2}}{4}\mathbf{e}_{d-1}$, respectively. We conclude that the segment joining these two points contains $\mathbf{o}_d^*$. From Lemma 1.4, part 1, with $\mathbf{x} = (\mathbf{o}_{d-2}^*, 0, 0)$, $\mathbf{y} = \frac{m_{d-2}}{4}\mathbf{e}_{d-1}$ giving $0 < \lambda \leq \frac{m_{d-2}/4}{m_{d-2}+m_{d-2}/4} = 1/17$, we conclude $\mathbf{o}_d^* = \lambda(\mathbf{o}_{d-2}^*, 0, 0) + (1-\lambda)\frac{m_{d-2}}{4}\mathbf{e}_{d-1}$ and $\|\mathbf{o}_d^*\|^2 = \|\lambda(\mathbf{o}_{d-2}^*, 0, 0)\|^2 + \|(1-\lambda)\frac{m_{d-2}}{4}\mathbf{e}_{d-1}\|^2$ so that

$$\|\mathbf{o}_d^*\| \geq \left\|(1-\lambda)\frac{m_{d-2}}{4}\mathbf{e}_{d-1}\right\| \geq \left\|\frac{1}{2}\frac{m_{d-2}}{4}\mathbf{e}_{d-1}\right\| \geq \frac{m_{d-2}}{8}.$$

We also have $m_d = m_{d-2}/16$. The bound $m_d \leq \|\mathbf{o}_d^*\|$ follows.

To conclude that the sizes of numbers are bounded by a polynomial in d, it is enough to notice that, given our explicit choices of constants m_d, M_d of size polynomial in d, there is a less recursive way of defining instance $P(d)$: Write o_{d-2}^* more explicitly as the affine minimizer of $O(d-2) = 1, \mathbf{r}_3, \mathbf{s}_3, \mathbf{r}_5, \mathbf{s}_5, \ldots, \mathbf{r}_{d-2}, \mathbf{s}_{d-2}$, where these points are explicit, not recursively defined, given m_d, M_d. We look at the entries in $P(d)$ (see (1.1)). We have that o_{d-2}^* is the solution to a linear system with entries of size bounded by a polynomial in d (the system to determine the minimum norm point in the affine hull of $1, \mathbf{r}_3, \mathbf{s}_3, \mathbf{r}_5, \mathbf{s}_5, \ldots, \mathbf{r}_{d-2}, \mathbf{s}_{d-2}$). From Cramer's rule, it follows that o_{d-2}^* is a vector with rational entries and the size of every entry is bounded by a polynomial in d. The other entries in $P(d)$ are m_d or M_d with a constant multiplied or added, and their sizes are also bounded by a polynomial in d. The claim follows. ∎

2. Linear optimization reduces to minimum-norm point problems on simplices. We reduce linear programming to the minimum norm point problem over a simplex via a series of strongly polynomial time reductions. The algorithmic problems we consider are defined below. We give definitions for the problems of linear programming (LP), feasibility (FP), bounded feasibility (BFP), V-polytope membership (VPM), zero V-polytope membership (ZVPM), zero V-polytope membership decision (ZVPMD), and distance to a V-simplex (DVS). (The prefix "V-" means that the respective object is specified as the convex hull of a set of points.)

Existing works describe slightly different notions of a "strongly polynomial time algorithm." A source of differences is the treatment of division and inputs that allow rational or real numbers (as opposed to just integers). Some known algorithms are strongly polynomial in some sort of "real RAM" (random access machine) model without a particular concern about divisions and with an analysis of growth of numbers in intermediate calculations (for example, algorithms that do not use division). At the same time, some basic algorithms such as Gaussian elimination are only known to be strongly polynomial when, on rational input, the integers in numerators and denominators of intermediate values are carefully handled and division is implemented in a particular form. For example, the analysis of Gaussian elimination in [19, sections 1.3 and 1.4] defines the result of dividing two integers as "the rational number a/b, and if it is known in advance that a/b is an integer, the integer a/b." These analyses are therefore performed in a sort of integer RAM where rational numbers are encoded as pairs of integers.

Our reductions below use Gaussian elimination and deal with algorithmic problems related to the linear programming problem. As algorithmic issues around linear programming are likely to run into similar issues as Gaussian elimination, we naturally adopt a similar integer RAM for our claims of strongly polynomial time algorithms. The precise definition of an arithmetic RAM that we use for the analysis of strongly polynomial time algorithms is based on the one in [31, section 4.2]. We state our version here for completeness.

An *integer random access machine* (RAM) has a finite set of variables $z_0, \dots, z_k$ and one array, f, of length depending on the input. Each array entry and variable stores an integer. A pair of integers p, q can be interpreted as rational number p/q if the machine needs to handle rational numbers. Initially $z_0, \dots, z_k$ are set to 0 and f contains the input. Each instruction is a finite sequence of resettings of one of the following types, for $i, j, h \in \{1, \dots, k\}$:

$$z_i := f(z_j); f(z_j) := z_i; z_i := z_j + z_h; z_i := z_j - z_h; z_i := z_j z_h; z_i := z_i + 1;$$
$$z_i := z_j/z_h \text{ if it is known in advance that } z_j/z_h \text{ is an integer;}$$
$$z_i := 1 \text{ if } z_j > 0 \text{ and } z_i := 0 \text{ otherwise.}$$

The instructions are numbered $0, 1, \dots, t$, and z_1 is the number of the instruction to be executed. If $z_1 > t$ we stop and return the contents of the array f as output.

An algorithm is an integer RAM. An algorithm runs in time $O(g)$ if it terminates after $O(g(n, s))$ operations (including elementary arithmetic operations), where the input consists of n integers of maximum size s and if the numbers occurring during the execution of the algorithm have size $O(s)$. The algorithm is called a *strongly polynomial time* algorithm if it takes $O(g(n))$ time for some polynomial g in n (the number of integer entries in the input array), where g is independent of s.

In our algorithms below, we assume that numbers in the inputs are rational numbers, each given as a pair of integers (interpreted as numerator and denominator) to an integer RAM. We do not make any claims about inputs that involve arbitrary real numbers. Still, our model guarantees desirable properties of strongly polynomial time algorithms: the number of steps in an integer RAM is independent of the sizes of the input numbers and the algorithms run in polynomial time in the Turing machine model. In particular, the existence of a strongly polynomial time algorithm in the integer RAM model for the distance to a V-simplex problem implies the existence of a strongly polynomial time algorithm in the integer RAM model for linear programming. See [19, 30, 31] for a detailed discussions of strongly polynomial time algorithms.

DEFINITION 2.1. *Consider the following computational problems:*

- LP: *Given a rational* $d \times n$ *matrix* A*, a rational column vector* $\mathbf{b}$*, and a rational row vector* $\mathbf{c}^\top$*, output rational* $\mathbf{x} \in \operatorname{argmax}\{\mathbf{c}^\top\mathbf{x} : A\mathbf{x} \leq \mathbf{b}\}$ *if* $\max\{\mathbf{c}^\top\mathbf{x} : A\mathbf{x} \leq \mathbf{b}\}$ *is finite, otherwise output INFEASIBLE if* $\{\mathbf{x} : A\mathbf{x} \leq \mathbf{b}\}$ *is empty and else output INFINITE.*
- FP: *Given a rational* $d \times n$ *matrix* A *and a rational vector* $\mathbf{b}$*, if* $P := \{\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ *is nonempty, output a rational* $\mathbf{x} \in P$*, otherwise output NO.*
- BFP: *Given a rational* $d \times n$ *matrix* A*, a rational vector* $\mathbf{b}$ *and a rational value* $M > 0$*, if* $P := \{\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}, \sum_{i=1}^n x_i \leq M\}$ *is nonempty, output a rational* $\mathbf{x} \in P$*, otherwise output NO.*
- VPM: *Given a rational* $d \times n$ *matrix* A *and a rational vector* $\mathbf{b}$*, if* $P := \{\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}, \sum_{i=1}^n x_i = 1\}$ *is nonempty, output a rational* $\mathbf{x} \in P$*, otherwise output NO.*
- ZVPM: *Given a rational* $d \times n$ *matrix* A*, if* $P := \{\mathbf{x} : A\mathbf{x} = \mathbf{0}, \mathbf{x} \geq \mathbf{0}, \sum_{i=1}^n x_i = 1\}$ *is nonempty, output a rational* $\mathbf{x} \in P$*, otherwise output NO.*
- ZVPMD: *Given rational points* $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n \in \mathbb{R}^d$*, output YES if* $\mathbf{0} \in \operatorname{conv}\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$ *and NO otherwise.*
- DVS: *Given* $n \leq d+1$ *affinely independent rational points* $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n \in \mathbb{R}^d$ *defining* $(n-1)$*-dimensional simplex* $P = \operatorname{conv}\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$*, output* $d(\mathbf{0}, P)^2$*.*

The main result in this section reduces linear programming to finding the minimum norm point in a (vertex-representation) simplex.

THEOREM 2.2. *LP reduces to DVS in strongly polynomial time.*

To prove each of the lemmas below, we illustrate the problem transformation and its strong polynomiality. The first two reductions are highly classical, while those following are intuitive, but we do not believe they have been written elsewhere.

Below is the sequence of algorithmic reductions that reduce LP to DVS. The first result, which reduces LP to FP, is a classical result; see, e.g., [19, 0.1.49]. This reduction consists of several FP problems. We first check feasibility of the primal and if so, we check feasibility of the primal and dual LP simultaneously with the additional constraint that the objective values are equal. If this FP is feasible, we return the objective function value; otherwise we indicate it is infinite. The reduction requires only the time to construct the FP problems and two calls to the FP oracle, so it is a strongly polynomial reduction. We do not include this proof as it is so similar to standard results. The second result, which reduces FP to BFP, is similar to classical results; see, e.g., [30, Corollary 3.2b]. We include a proof here as the reduction is less intuitive than the former.

LEMMA 2.3. *LP reduces in strongly polynomial time to FP.*

LEMMA 2.4. *FP reduces in strongly polynomial time to BFP.*

Proof. Let $\mathcal{O}$ denote the oracle for BFP. Suppose $A = (a_{ij}/\alpha_{ij})_{i,j=1}^{d,n}$, $\mathbf{b} = (b_j/\beta_j)_{j=1}^{d}$ and define

$$D := \max(\max_{i\in[d],j\in[n]} |\alpha_{ij}|, \max_{k\in[d]} |\beta_k|),$$
$$N := \max(\max_{i\in[d],j\in[n]} |a_{ij}|, \max_{k\in[d]} |b_k|) + 1.$$

If the entry of A is zero, $a_{ij}/\alpha_{ij} = 0$, define $a_{ij} = 0$ and $\alpha_{ij} = 1$. If the entry of b is zero, $b_j/\beta_j = 0$, define $b_j = 0$ and $\beta_j = 1$.

Require: $A \in \mathbb{Q}^{d\times n}, \mathbf{b} \in \mathbb{Q}^d$.

Invoke $\mathcal{O}$ on $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}, \sum_{i=1}^n x_i \leq nD^{d(n+1)\min(d^3,n^3)}N^{d(n+1)}$. If the output is NO, output NO, else output rational $\mathbf{x}$.

CLAIM 2.5. *The FP $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$ is feasible if and only if the BFP $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}, \sum_{i=1}^n x_i \leq nD^{d(n+1)\min(d^3,n^3)}N^{d(n+1)}$ is feasible.*

Proof of Claim. If the BFP is feasible, then clearly the FP is feasible. Suppose the FP is feasible. By the theory of minimal faces of polyhedra, we can reduce this to an FP defined by a square matrix, A, in the following way: By [8, Theorem 1.1], there is a solution, $\mathbf{x}$, with no more than $\min(d, n)$ positive entries so that $A\mathbf{x} = \mathbf{b}$ and the positive entries of $\mathbf{x}$ combine linearly independent columns of A to form $\mathbf{b}$. Let A' denote the matrix containing only these linearly independent columns and $\mathbf{x}'$ denote only the positive entries of $\mathbf{x}$. Then $A'\mathbf{x}' = \mathbf{b}$. Now, note that $A' \in \mathbb{Q}^{d\times m}$, where $m \leq d$. Since the column rank of A' equals the row rank of A', we may remove $d - m$ linearly dependent rows of A' and the corresponding entries of $\mathbf{b}$, forming A'' and $\mathbf{b}'$ so that $A''\mathbf{x}' = \mathbf{b}'$, where $A'' \in \mathbb{Q}^{m\times m}$, $\mathbf{b}' \in \mathbb{Q}^m$, and A'' is a full-rank matrix.

Define $M := \prod_{i,j=1}^m |\alpha''_{i,j}| \prod_{k=1}^m |\beta'_k|$ and note that $M \leq D^{d(n+1)}$. Define $L := \prod_{i,j=1}^m (|a''_{i,j}| + 1) \prod_{k=1}^m (|b'_k| + 1)$ and note that $L \leq N^{d(n+1)}$. Define $\bar{A} = MA''$ and $\bar{\mathbf{b}} = Mb'$ and note that $\bar{A}$ and $\bar{\mathbf{b}}$ are integral. By Cramer's rule, we known that $x'_i = \frac{|\det \bar{A}_i|}{|\det \bar{A}|}$, where $\bar{A}_i$ denotes $\bar{A}$ with the ith column replaced by $\bar{\mathbf{b}}$. By integrality, $|\det \bar{A}| \geq 1$, so $x'_i \leq |\det \bar{A}_i| \leq \prod_{i,j=1}^m M(|a_{ij}| + 1) \prod_{k=1}^m M(|b_k| + 1) = M^{m^3}L \leq D^{d(n+1)\min(d^3,n^3)}N^{d(n+1)}$. Now, note that $\mathbf{x}'$ defines a solution, $\mathbf{x}$, to the original system of equations. Let $x_i = x'_j$ if the jth column of A' was the selected ith column of A and $x_i = 0$ if the ith column of A was not selected. Note then that $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}, \sum_{i=1}^n x_i \leq nD^{d(n+1)\min(d^3,n^3)}N^{d(n+1)}$. ∎

Thus, we have that the FP and BFP are equivalent. To see that this is a strongly polynomial time reduction, note that adding this additional constraint takes time for constructing the number $nD^{d(n+1)\min(d^3,n^3)}N^{d(n+1)}$ plus small constant time. This number takes $d(n+1)$ comparisons and $d(n+1)\min(d^3, n^3)$ multiplications to form. Additionally, this number takes space which is polynomial in the size of the input (polynomial in d,n and size of D, N). ∎

LEMMA 2.6. *BFP reduces in strongly polynomial time to VPM.*

Proof. Let $\mathcal{O}$ denote the oracle for VPM.

Require: $A \in \mathbb{Q}^{d\times n}, b \in \mathbb{Q}^d, 0 < M \in \mathbb{Q}$.

Invoke $\mathcal{O}$ on

$$\begin{bmatrix} MA & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ z \end{bmatrix} = \mathbf{b}, \begin{bmatrix} \mathbf{y} \\ z \end{bmatrix} \geq \mathbf{0}, z + \sum_{i=1}^n y_i = 1. \tag{2.1}$$

If the output is NO, output NO, else output rational $\mathbf{x} = M\mathbf{y}$.

CLAIM 2.7. *A solution*

$$\tilde{\mathbf{w}} := \begin{bmatrix} \mathbf{y} \\ z \end{bmatrix}$$

to (2.1) *gives a solution the BFP instance,* $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}, \sum_{i=1}^n x_i \leq M$ *and vice versa.*

Proof of Claim. Suppose $\tilde{\mathbf{w}}$ satisfies (2.1). Then $\mathbf{x} = M\mathbf{y}$ is a solution to the BFP instance since $A\mathbf{x} = MA\mathbf{y} = \mathbf{b}$ and since $\mathbf{y} \geq \mathbf{0}$, $\mathbf{x} = M\mathbf{y} \geq \mathbf{0}$ and since $\sum_{i=1}^n y_i + z = 1$, we have $\sum_{i=1}^n y_i \leq 1$ so $\sum_{i=1}^n x_i = M \sum_{i=1}^n y_i \leq M$. Suppose $\mathbf{x}$ is a solution to the BFP instance. Then $\mathbf{y} = \frac{1}{M}\mathbf{x}$ and $z = 1 - \sum_{i=1}^n y_i$ satisfies (2.1), since $\begin{bmatrix} MA & 0 \end{bmatrix} \tilde{\mathbf{w}} = MA\mathbf{y} = A\mathbf{x} = \mathbf{b}$, $\mathbf{y} \geq \mathbf{0}$ since $\mathbf{x} \geq \mathbf{0}$ and since $\sum_{i=1}^n x_i \leq M$, we have $\sum_{i=1}^n y_i = \frac{1}{M} \sum_{i=1}^n x_i \leq 1$ so $z \geq 0$. ∎

Clearly, this reduction is simply a rewriting, so the reduction is strongly polynomial time. ∎

LEMMA 2.8. *VPM reduces in strongly polynomial time to ZVPM.*

Proof. Let $\mathcal{O}$ be the oracle for ZVPM.

Require: $A \in \mathbb{Q}^{d\times n}, b \in \mathbb{Q}^d$.
Invoke $\mathcal{O}$ on

$$\begin{bmatrix} \mathbf{a}_1 - \mathbf{b} & \mathbf{a}_2 - \mathbf{b} & \cdots & \mathbf{a}_n - \mathbf{b} \end{bmatrix} \mathbf{x} = \mathbf{0}, \mathbf{x} \geq \mathbf{0}, \sum_{i=1}^n x_i = 1, \tag{2.2}$$

where $\mathbf{a}_i \in \mathbb{Q}^m$ is the ith column of A. If the output is NO, output NO, else output rational $\mathbf{x}$.

CLAIM 2.9. *A solution to* (2.2) *gives a solution to the VPM instance and vice versa.*

Proof of Claim. Note that $\mathbf{x}$ satisfies (2.2) if and only if $0 = \sum_{i=1}^n x_i(\mathbf{a}_i - \mathbf{b}) = \sum_{i=1}^n x_i \mathbf{a}_i - \mathbf{b} \sum_{i=1}^n x_i = A\mathbf{x} - \mathbf{b}$ so $A\mathbf{x} = \mathbf{b}$. Thus, $\mathbf{x}$ is a solution to the VPM instance if and only if $\mathbf{x}$ is a solution to (2.2). ∎

Clearly, this reduction is simply a rewriting, so the reduction is strongly polynomial time. ∎

LEMMA 2.10. *ZVPM reduces in strongly polynomial time to ZVPMD.*

Proof idea. The reduction sequentially asks for every vertex whether it is redundant and if so, it removes it and continues. This process ends with at most $d+1$ vertices so that $\mathbf{x}$ is a strict convex combination of them and the coefficients x_i can be found in this resulting case by solving a linear system. ∎

Proof. Let $\mathcal{O}$ denote the ZVPMD oracle.

Require: $P := \{\mathbf{A}_1, \ldots, \mathbf{A}_n\} \subseteq \mathbb{Q}^d$ where $\mathbf{A}_i$ is the ith column of A.
Invoke $\mathcal{O}$ on P. If the output is NO, output NO.
for $i = 1, \ldots, n$ **do**
  Invoke $\mathcal{O}$ on instance P without $\mathbf{A}_i$. If output is YES, remove $\mathbf{A}_i$ from P.
end for
Let $\mathbf{p}_1, \ldots, \mathbf{p}_m$ be the points in P, where m is the cardinality of P.
Output the solution $\mathbf{x} \in \mathbb{R}^m$ to the linear system $\sum_{i=1}^m x_i = 1$, $\sum_{i=1}^m x_i \mathbf{p}_i = \mathbf{0}$.

Let P^* be the resulting set of points P after the loop in the reduction. The correctness of the reduction will follow from the following claim: P^* contains at most $d+1$ points so that $\mathbf{0}$ is a strict convex combination of (all of) them. We will show the claim in the rest of the proof. By Caratheodory's theorem there is a subset $Q \subseteq P^*$ of at most $d+1$ points so that $\mathbf{0}$ is a strict convex combination of points in Q. We will see that P^* is actually equal to Q. Suppose not, for a contradiction. Let $\mathbf{p} \in P^* \setminus Q$. At the time the loop in the reduction examines $\mathbf{p}$, no point in Q has been removed and therefore $\mathbf{p}$ is redundant and is removed. This is a contradiction. ∎

In the last step of the reduction (Lemma 2.13), we make use of two claims. The first one is so elementary that we omit the proof. The second is a more technical claim and we include the proof.

Claim 2.11. *Given an $m \times n$ matrix A, let B be A with a row of 1s appended. The columns of A are affinely independent if and only if the columns of B are linearly independent. The convex hull of the columns of A is full-dimensional if and only if rank of B is $m+1$.*

Claim 2.12. *Let $P = \operatorname{conv}\{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ be a V-polytope with $\mathbf{p}_i \in \mathbb{Q}^d$. Let T be the maximum of the absolute values of all numerators and denominators of entries in $(\mathbf{p}_i)_{i=1}^n$. If $\mathbf{0} \notin P$, then $d(\mathbf{0}, P) \geq \frac{1}{(dT)^d\sqrt{d}}$.*

Proof. Let $\mathbf{y} = \operatorname{proj}_P(\mathbf{0})$. We have that every facet of P can be written as $\mathbf{a}^\top \mathbf{x} \leq k$, where $\mathbf{a}(\neq \mathbf{0})$ is an integral vector, k is an integer, and the absolute values of the entries of $\mathbf{a}$ as well as k are less than $(dT)^d$ [20, Theorem 3.6]. By assumption at least one these facet inequalities is violated by $\mathbf{0}$. Denote by $\mathbf{a}^\top \mathbf{x} \leq k$ one such inequality. Let $H = \{\mathbf{x} : \mathbf{a}^\top \mathbf{x} = k\}$. We have $\|\mathbf{y}\| = d(0, P) \geq d(0, H)$, and $d(0, H)^2 = k^2/\|\mathbf{a}\|^2 \geq \frac{1}{d(dT)^{2d}}$. The claim follows. ∎

For the proof of Lemma 2.13, we will also need the fact that Gram–Schmidt orthogonalization can be implemented in strongly polynomial time. By "Gram–Schmidt orthogonalization" we mean that given vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$, compute vectors $\mathbf{u}_i = \mathbf{v}_i - \sum_{j=1}^{i-1} \operatorname{proj}_{\mathbf{u}_j}(\mathbf{v}_i)$, where $\operatorname{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{u}^\top \mathbf{v}}{\mathbf{u}^\top \mathbf{u}} \mathbf{u}$. The subset of all nonzero vectors in $\{\mathbf{u}_i : i = 1, \dots, k\}$ is a set of orthogonal (and thus linearly independent) vectors. Moreover, the whole sequence satisfies $\operatorname{span}(\mathbf{u}_1, \dots, \mathbf{u}_l) = \operatorname{span}(\mathbf{v}_1, \dots, \mathbf{v}_l)$ for all $l = 1, \dots, k$.

We could not find in the literature a complete proof that Gram–Schmidt orthogonalization can be computed in strongly polynomial time for rational input. At the same time, it follows easily from results in [19, section 1.4]. For example [19, Corollary 1.4.9] establishes that the determinant of a rational matrix can be computed in strongly polynomial time. The following argument reduces the computation of Gram–Schmidt orthogonalization to the computation of determinants in a way that proves our claim: First, if the input sequence of vectors is linearly independent, the following formulas give the orthogonalized sequence using determinants (see, for example, [16, section IX.6]:

$$\mathbf{u}_j = \frac{1}{D_{j-1}} \det \begin{pmatrix} \mathbf{v}_1^\top \mathbf{v}_1 & \mathbf{v}_2^\top \mathbf{v}_1 & \cdots & \mathbf{v}_j^\top \mathbf{v}_1 \\ \mathbf{v}_1^\top \mathbf{v}_2 & \mathbf{v}_2^\top \mathbf{v}_2 & \cdots & \mathbf{v}_j^\top \mathbf{v}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_1^\top \mathbf{v}_{j-1} & \mathbf{v}_2^\top \mathbf{v}_{j-1} & \cdots & \mathbf{v}_j^\top \mathbf{v}_{j-1} \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_j \end{pmatrix},$$

where $D_0 = 1$ and, for $j \geq 1$,

$$D_j = \det \begin{pmatrix} \mathbf{v}_1^\top \mathbf{v}_1 & \mathbf{v}_2^\top \mathbf{v}_1 & \cdots & \mathbf{v}_j^\top \mathbf{v}_1 \\ \mathbf{v}_1^\top \mathbf{v}_2 & \mathbf{v}_2^\top \mathbf{v}_2 & \cdots & \mathbf{v}_j^\top \mathbf{v}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_1^\top \mathbf{v}_{j-1} & \mathbf{v}_2^\top \mathbf{v}_{j-1} & \cdots & \mathbf{v}_j^\top \mathbf{v}_{j-1} \\ \mathbf{v}_1^\top \mathbf{v}_j & \mathbf{v}_2^\top \mathbf{v}_j & \cdots & \mathbf{v}_j^\top \mathbf{v}_j \end{pmatrix}.$$

In the definition of $\mathbf{u}_j$, the determinant should be interpreted as a formal cofactor expansion along the last row; the last entries of each column are the vectors $\mathbf{v}_i$ and the expansion gives the coefficients of these vectors (as subdeterminants) in the definition of $\mathbf{u}_j$. If the input sequence of vectors is linearly dependent, then this can be detected and handled as the determinants above are computed: Whenever the Gram determinant D_j is zero, the sequence $\{\mathbf{v}_1, \ldots, \mathbf{v}_j\}$ is linearly dependent, the corresponding orthogonalized vector $\mathbf{u}_j$ is 0, and $\mathbf{v}_j$ is skipped in subsequent iterations.

LEMMA 2.13. *ZVPMD reduces in strongly polynomial time to DVS.*

Proof. Given an instance $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ of ZVPMD, we reduce it to an instance of DVS as follows: We lift the points to an affinely independent set in higher dimension, a simplex, by adding small-valued new coordinates. Claim 2.11 allows us to handle affine independence in matrix form. Let A be the $d \times n$ matrix having columns $(\mathbf{p}_i)_{i=1}^n$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_d$ be the rows of A. Let $\mathbf{v}_0 \in \mathbb{R}^n$ be the all-ones vector. We want to add vectors $\mathbf{v}_{d+1}, \ldots, \mathbf{v}_{d+t}$, for some t, so that $\mathbf{v}_0, \ldots, \mathbf{v}_{d+t}$ is of rank n. To this end, we construct an orthogonal basis (but not normalized, to preserve rationality) of the orthogonal complement of span$(\mathbf{v}_0, \ldots, \mathbf{v}_d)$. The basis is obtained by applying the Gram–Schmidt orthogonalization procedure (that is, without the normalization step) to the sequence $\mathbf{v}_0, \ldots, \mathbf{v}_d, \mathbf{e}_1, \ldots, \mathbf{e}_n$. It is known that Gram–Schmidt orthogonalization can be computed in strongly polynomial time (see the paragraph after the proof for a discussion). Denote $\mathbf{v}_{d+1}, \ldots, \mathbf{v}_{d+t}$ the resulting orthogonal basis of the orthogonal complement of span$(\mathbf{v}_0, \ldots, \mathbf{v}_d)$. The matrix with rows $\mathbf{v}_0, \ldots, \mathbf{v}_d, \mathbf{v}_{d+1}, \ldots, \mathbf{v}_{d+t}$ is of rank n and so is the matrix with rows

$$\mathbf{v}_0, \ldots, \mathbf{v}_d, \epsilon \mathbf{v}_{d+1}, \ldots, \epsilon \mathbf{v}_{d+t}$$

for any $\epsilon > 0$ (to be fixed later). Therefore, the n columns of this matrix are linearly independent. Let B be the matrix with rows

$$\mathbf{v}_1, \ldots, \mathbf{v}_d, \epsilon \mathbf{v}_{d+1}, \ldots, \epsilon \mathbf{v}_{d+t}.$$

Let $\mathbf{w}_1, \ldots, \mathbf{w}_n$ be the columns of B. By construction and Claim 2.11 they are affinely independent. Let S be the convex hull of these $(n-1)$-dimensional rational points. Polytope S is a simplex. Let $Q := \operatorname{conv}\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$, and take $\sum_k \lambda_k \mathbf{p}_k$ to be the minimum norm point in Q, and $\sum_k \lambda_k \mathbf{w}_k$ the lifted image of this point in S. Then

$$
\begin{aligned}
d(\mathbf{0},S)^2 \le \|\sum_k \lambda_k \mathbf{w}_k\|^2 &= \|\sum_k \lambda_k \mathbf{p}_k\|^2 + \epsilon^2 \sum_k \lambda_k \sum_{i=d+1}^{d+t} v_{ik}^2 \\
&= d(\mathbf{0},Q)^2 + \epsilon^2 \sum_{i=d+1}^{d+t} \sum_k \lambda_k v_{ik}^2 \\
&\le d(\mathbf{0},Q)^2 + \epsilon^2 \sum_{i=d+1}^{d+t} \|\mathbf{v}_i\|^2 \\
&\le d(\mathbf{0},Q)^2 + \epsilon^2 n
\end{aligned}
$$

(where we use that $\|\mathbf{v}_i\| \le 1$ for $i \ge d+1$, as those vectors are the orthogonalized vectors corresponding to the canonical vectors and each vector in the output of Gram–Schmidt orthogonalization cannot be longer than the corresponding vector in the input).

The reduction proceeds as follows: Let T be the maximum of the absolute values of all numerators and denominators of entries in $(\mathbf{p}_i)_{i=1}^n$. Note that T can be computed in strongly polynomial time: take the integers in the input representing the numerators and denominators, compute their absolute values, and then take the maximum. From Claim 2.12, we have $d(\mathbf{0},Q)^2 \ge \frac{1}{d(dT)^{2d}}$ if $\mathbf{0} \notin Q$. Compute rational $\epsilon > 0$ so that $\epsilon^2 n < \frac{1}{d(dT)^{2d}}$. For example, let $\epsilon := \frac{1}{nd(dT)^d}$. The reduction queries $d(\mathbf{0},S)^2$ for S constructed as above and given by the choice of ϵ we just made. It then outputs YES if $d(\mathbf{0},S)^2 < \frac{1}{d(dT)^{2d}}$ and NO otherwise. Note that $d(\mathbf{0},Q)^2 \le d(\mathbf{0},S)^2$ since if $\sum_k \lambda_k \mathbf{w}_k \in S$ is the point achieving $\|\sum_k \lambda_k \mathbf{w}_k\|^2 = d(\mathbf{0},S)^2$, then by construction of $\mathbf{w}_k$ we have $d(\mathbf{0},Q)^2 \le \|\sum_k \lambda_k \mathbf{p}_k\|^2 \le \|\sum_k \lambda_k \mathbf{w}_k\|^2 = d(\mathbf{0},S)^2$. ∎

3. Conclusions and open questions. We have seen that Wolfe's method using a natural point insertion rule exhibits exponential behavior. We have also shown that the minimum norm point problem for simplices is intimately related to the complexity of linear programming. Our work raises several very natural questions:

- Are there exponential examples for other insertion rules for Wolfe's method? Also, at the moment, the ordering of the points starts with the closest point to the origin, but one could also consider a randomized initial rule or a randomized insertion rule.
- For applications in submodular function minimization, the polytopes one considers are base polytopes and our exponential example is not of this kind. Could there be hope that for base polytopes Wolfe's method performs better?
- It would be interesting to understand the average performance of Wolfe's method. How does it behave for random data? Further randomized analysis of this method would include the smoothed analysis of Wolfe's method or at least the behavior for data following a prescribed distribution.
- We have seen that it is already quite interesting to study the minimum norm point problem for simplices, when we discussed the connection with linear programming. Is there a family of simplices where Wolfe's method takes exponential time?
- Can Wolfe's method be extended to other convex L_p norms for $p \ge 1$? Can we identify the types of objective functions for which computing the affine minimizer is easy?

Acknowledgments. We are truly grateful to the anonymous STOC referees for their many comments and suggestions, which have greatly improved this paper. We also thank Gilberto Calvillo, Deeparnab Chakrabarty, Antoine Deza, Stephanie Jegelka, Matthias Köppe, Tamon Stephen, and John Sullivan for useful suggestions, conversations, and comments about this project.

REFERENCES

[1] F. Bach, *Learning with submodular functions: A convex optimization perspective*, Found. Trends Machine Learning, 6 (2013), pp. 145–373.

[2] A. Bachem and B. Korte, *Minimum norm problems over transportation polytopes*, Linear Algebra Appl., 31 (1980), pp. 103–118.

[3] I. Bárány and S. Onn, *Colourful linear programming and its relatives*, Math. Oper. Res., 22 (1997), pp. 550–567.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[5] G. Calvillo and D. Romero, *On the closest point to the origin in transportation polytopes*, Discrete Appl. Math., 210 (2016), pp. 88–102.

[6] D. Chakrabarty, P. Jain, and P. Kothari, *Provable submodular minimization using Wolfe's algorithm*, in Proceedings of Advances in Neural Information Processing Systems, 2014, pp. 802–809.

[7] V. Chandru, A. Dattasharma, S. S. Keerthi, N. K. Sancheti, and V. Vinay, *Algorithms for the optimal loading of recursive neural nets*, in Proceedings of SODA, 1995, pp. 342–349.

[8] S. N. Chernikov, *The convolution of finite systems of linear inequalities*, USSR Comput. Math. Math. Phys., 5 (1965), pp. 1–24.

[9] G. B. Dantzig, *An ϵ-precise Feasible Solution to a Linear Program with a Convexity Constraint in $1/\epsilon^2$ Iterations Independent of Problem Size*, Tech. report, Stanford University, 1992.

[10] J. A. De Loera, J. Haddock, and L. Rademacher, *The minimum Euclidean-norm point on a convex polytope: Wolfe's combinatorial algorithm is exponential*, in Proceedings of the 50th ACM Symposium on Theory of Computing (STOC), 2018.

[11] M. Frank and P. Wolfe, *An algorithm for quadratic programming*, Naval Res. Logist., 3 (1956), pp. 95–110.

[12] S. Fujishige, *Lexicographically optimal base of a polymatroid with respect to a weight vector*, Math. Oper. Res., 5 (1980), pp. 186–196, https://doi.org/10.1287/moor.5.2.186.

[13] S. Fujishige, T. Hayashi, and S. Isotani, *The Minimum-Norm-Point Algorithm Applied to Submodular Function Minimization and Linear Programming*, Research Institute for Mathematical Sciences, Kyoto University, 2006.

[14] S. Fujishige and S. Isotani, *A submodular function minimization algorithm based on the minimum-norm base*, Pacific J. Optim., 7 (2011), pp. 3–17.

[15] K. Fukuda, T. M. Liebling, and F. Margot, *Analysis of backtrack algorithms for listing all vertices and all faces of a convex polyhedron*, Comput. Geom., 8 (1997), pp. 1–12, https://doi.org/https://doi.org/10.1016/0925-7721(95)00049-6.

[16] F. R. Gantmacher, *The Theory of Matrices*, translated by K. A. Hirsch, Chelsea Publishing, New York, 1959.

[17] D. Ge, X. Jiang, and Y. Ye, *A note on the complexity of L_p minimization*, Math. Program., 129 (2011), pp. 285–299, https://doi.org/10.1007/s10107-011-0470-2.

[18] E. G. Gilbert, *An iterative procedure for computing the minimum of a quadratic form on a convex set*, SIAM J. Control, 4 (1966), pp. 61–80.

[19] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, Springer-Verlag, Berlin, 1988.

[20] M. Grötschel, L. Lovász, and A. Schrijver, *The ellipsoid method and its consequences in combinatorial optimization*, Combinatorica, 1 (1981), pp. 169–197, https://doi.org/10.1007/BF02579273.

[21] V. Klee and G. J. Minty, *How good is the simplex algorithm?*, in Inequalities III Academic Press, New York, 1972, pp. 159–175.

[22] M. K. Kozlov, S. P. Tarasov, and L. G. Khachiyan, *The polynomial solvability of convex quadratic programming*, USSR Comput. Math. Math. Phys., 20 (1980), pp. 223–228.

[23] S. Lacoste-Julien and M. Jaggi, *An Affine Invariant Linear Convergence Analysis for Frank-Wolfe Algorithms*, preprint, arXiv:1312.7864, 2013.

[24] S. Lacoste-Julien and M. Jaggi, *On the Global Linear Convergence of Frank-Wolfe Optimization Variants*, in Proceedings of Advances in Neural Information Processing Systems, 2015, pp. 496–504.

[25] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Classics in Appl. Math. 15, SIAM, Philadelphia, 1995, https://doi.org/10.1137/1.9781611971217.

[26] K. Nagano, Y. Kawahara, and K. Aihara, *Size-constrained submodular minimization through minimum norm base*, in Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011, pp. 977–984.

[27] J. Peña, D. Rodríguez, and N. Soheili, *On the von Neumann and Frank–Wolfe algorithms with away steps*, SIAM J. Optim., 26 (2016), pp. 499–512, https://doi.org/10.1137/15M1009937.

[28] T. Rothvoss, *The matching polytope has exponential extension complexity*, in Proceedings of the 2014 ACM Symposium on Theory of Computing, ACM, New York, 2014, pp. 263–272.

[29] T. Rothvoss, *The matching polytope has exponential extension complexity*, J. ACM, 64 (2017), 41, https://doi.org/10.1145/3127497.

[30] A. Schrijver, *Theory of Linear and Integer Programming*, John Wiley & Sons, Chichester, UK, 1998.

[31] A. Schrijver, *Combinatorial Optimization. Polyhedra and Efficiency. Vol. A: Paths, Flows, Matchings*, Algorithms Combin. 24, Springer-Verlag, Berlin, 2003.

[32] S. Smale, *Mathematical problems for the next century*, Math. Intelligencer, 20 (2000), pp. 7–15.

[33] T. Terlaky and S. Zhang, *Pivot rules for linear programming: A survey on recent theoretical developments*, Ann. Oper. Res., 46 (1993), pp. 203–233, https://doi.org/10.1007/BF02096264.

[34] L. A. Végh, *A strongly polynomial algorithm for a class of minimum-cost flow problems with separable convex objectives*, SIAM J. Comput., 45 (2016), pp. 1729–1761, https://doi.org/10.1137/140978296.

[35] P. Wolfe, *Algorithm for a least-distance programming problem*, Math. Program. Stud., 1 (1974), pp. 190–205.

[36] P. Wolfe, *Finding the nearest point in a polytope*, Math. Program., 11 (1976), pp. 128–149.