The Minimum Euclidean-Norm Point in a Convex Polytope: Wolfe’s Combinatorial Algorithm Is Exponential

Jesús A. De Loera
Department of Mathematics
University of California, Davis
USA
deloera@math.ucdavis.edu

Jamie Haddock
Department of Mathematics
University of California, Davis
USA
jhaddock@math.ucdavis.edu

Luis Rademacher
Department of Mathematics
University of California, Davis
USA
lrademac@ucdavis.edu

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.

CCS CONCEPTS
• Mathematics of computing → Quadratic programming;
• Theory of computation → Linear programming; Quadratic programming;

KEYWORDS
convex quadratic optimization, Wolfe’s method, linear programming, strongly polynomial time algorithms, lower bounds

ACM Reference Format:

1 INTRODUCTION
The fundamental algorithmic problem we consider here is: Given a convex polytope \( P \subset \mathbb{R}^d \), to find the point \( x \in P \) of minimum Euclidean norm, i.e., the closest point to the origin or what we call its minimum norm point for short. We assume \( P \) is presented as the convex hull of finitely many points \( p_1, \ldots, p_n \) (not necessarily in convex position). We wish to find

\[
\text{argmin} \quad \|x\|_2 \\
\text{subject to} \quad x = \sum_{k=1}^n \lambda_k p_k, \\
\sum_{k=1}^n \lambda_k = 1, \lambda_k \geq 0, \quad \text{for } k = 1, \ldots, n.
\]

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 for solving the minimum norm point problem can be used to compute the projection of an arbitrary point to a polytope (indeed, \( \text{argmin}_{x \in P} \|x - a\|_2 \) is the same as \( a + \text{argmin}_{y \in P - a} \|y\|_2 \)). The minimum norm problem additionally appears in combinatorial optimization, e.g., for the nearest point problem for transportation polytopes [Bachem and Korte 1980; Calvillo and Romero 2016] and as a vital subroutine in Bárány and Onn’s approximation algorithm to solve the colorful linear programming problem [Bárány and Onn 1997]. One of the most important reasons to study this problem is because the minimum norm problem can be used as a subroutine for submodular function minimization through projection onto the base polytope, as proposed by Fujishige [Fujishige 1980]. Submodular minimization is useful in machine learning, where applications such as large scale learning and vision require efficient and accurate solutions [Bach et al. 2013; Nagano et al. 2011]. The problem also appears in optimal loading of recursive neural networks [Chandru et al. 1995]. The Fujishige-Wolfe algorithm is currently considered an important practical algorithm in applications [Chakrabarty et al. 2014; Fujishige et al. 2006; Fujishige and Iwata 2011]. Furthermore, Fujishige et al. first observed that linear programs may be solved by solving the minimum norm point problem [Fujishige et al. 2006], so this simple geometric problem is also relevant to the theory of algorithmic complexity of linear optimization.

One may ask about the complexity of other closely related problems. First, it is worth remembering that \( L_p \) norm minimization over a polyhedron is NP-hard for \( 0 \leq p < 1 \) (see [Ge et al. 2011] and the references therein), while for \( p \geq 1 \) the convexity of the norm allows for computation of an \( \epsilon \)-approximate solution in time polynomial in \( \log(1/\epsilon) \). When \( p = 2 \), the solution to \( L_2 \) norm minimization over a polyhedron given by rational data is rational (see [Schrijver 1998] for details), so this solution may be computed exactly in polynomial time. Meanwhile, one can prove that the seemingly similar problem of finding the closest vertex of a convex polytope given by inequalities is NP-hard. The reduction for hardness is to the directed \emph{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 \( t \). It is well-known there is a polytope represented by inequalities which has some vertices corresponding to the characteristic vectors of directed paths joining \( s \) to \( t \) in \( G \). See e.g., [Fukuda et al. 1997] for the details, including the explicit inequality description of this polytope. Finally, by a change of variable \( y_i = 1 - x_i \), changing zeros to ones and vice versa, and an appropriate selection of the
upper bound on the value of $x_i$, the minimum Euclidean norm vertex becomes precisely the "longest path from $s$ to $t'$, solving the directed Hamiltonian path problem.

Since the Euclidean norm is a convex quadratic form, the minimum norm point problem is a special case of convex quadratic optimization problem. Indeed, it is well-known that a convex quadratic programming problem can be approximately solved in polynomial time; that is, some point $x$ within distance $\varepsilon$ of the desired minimizing point $x$ may be found in polynomial time with respect to $\log 1/\varepsilon$. This can be done either through several iterative (convergent) algorithms, such as the Ellipsoid method [Kozlov et al. 1980] and interior-point methods techniques [Boyd and Vandenberghe 2004]. Each of these are methods whose complexity depends upon the desired accuracy. However, an approximate numerical solution is inconvenient when the application requires more information, e.g., if we require to know the face that contains the minimum norm point. Numeric methods that converge to a solution and require further rounding are not as convenient for this need.

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 the 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 [Bachem and Korte 1980; Calvillo and Romero 2016; Végh 2016], 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 [Smale 2000], 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 problem for general polytopes.

2 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 [Schrijver 1998] 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 two important definitions. 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 $S$, $\text{argmin}_{x \in \text{aff}(S)} \|x\|_2$. The second is the convex minimizer which is the point of minimum norm in the convex hull of $S$, $\text{argmin}_{x \in \text{conv}(S)} \|x\|_2$. 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 thus may be computed in strongly-polynomial time; see [Schrijver 1998, Section 3.3] and references therein.

2.1 A Brief Review of Wolfe’s Combinatorial Method

Wolfe’s combinatorial method solves the minimum norm point problem over a polytope, $P = \text{conv}(p_1, p_2, \ldots, p_n) \subseteq \mathbb{R}^d$, and was introduced by P. Wolfe in [Wolfe 1976]. The method iteratively solves the minimum norm point problem over a sequence of subsets of no more than $d + 1$ affinely independent points from $p_1, \ldots, p_n$ 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.

**Lemma 2.1 (Wolfe’s criterion [Wolfe 1976]).** Let $P \subseteq \mathbb{R}^d$ be the convex hull of finitely many points, $P := \text{conv}(p_1, \ldots, p_n)$, then $x \in P$ is the minimum norm point in $P$ if and only if $x^T p_j \geq \|x\|_2^2$ for all $j \in [n]$.

Note that this tells us that if there exists a point $p_j$ so that $x^T p_j < \|x\|_2$ (i.e., the hyperplane $\{y : x^T y = \|x\|_2^2\}$ does not weakly separate $P$ from $0$), then $x$ is not the minimum norm point in $P$. We say that $p_j$ violates Wolfe’s criterion and using this point should decrease the norm of the minimum norm point of the current subproblem.

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

**Lemma 2.2 (Wolfe’s criterion for the affine hull).** Let $P = \{p_1, p_2, \ldots, p_n\} \subseteq \mathbb{R}^d$ be a non-empty finite set of points. Then $x \in \text{aff} P$ is the minimum norm point in $\text{aff} P$ if and only if $p_i \in P$ we have $p_i^T x = \|x\|_2^2$.

We say a set of affinely independent points $S$ is a corral if the affine minimizer of $S$ lies in the relative interior of $\text{conv} S$. 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 $p_1, \ldots, p_n$. The goal of
Wolfe’s method is to search for a corral containing the (unique) minimizing point.

The pseudo-code in Method 1 below presents the iterations of Wolfe’s method. 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.

Method 1 Wolfe’s Method [Wolfe 1976]

```plaintext
1. procedure WOLFE(p1, p2, ..., pn)
2. Initialize x = p1 for some i ∈ [n], initial corral C = {p1}, I = {i}, λ = e_i, α = 0.
3. while x ≠ 0 and there exists p_j with x^T p_j < ∥x∥^2 do
4. Add p_j to the potential corral: C = C ∪ {p_j},
   I = I ∪ {j}.
5. Find the affine minimizer of C,
   y = argmin_{y ∈ aff(C)} ∥y∥_2, and the affine coefficients, α.
6. while y is not a strict convex combination of points in C; α_i ≤ 0 for some i ∈ I do
7. Find z, closest point to y on [x, y] ∩ conv(C);
   z = θy + (1 - θ)x, θ = min_{i∈I, α_i < 0} λ_j/α_j.
8. Select p_j ∈ {p_j ∈ C : θα_j + (1 - θ)l_j = 0}.
9. Remove this point from C; C = C - {p_j},
   I = I - {j}, α_j = 0, λ_j = 0.
10. Update x = z and the convex coefficients, λ, of x for C; solve x = Σ p_j ∈ C λ_j p_j for λ.
11. Find the affine minimizer of C,
   y = argmin_{y ∈ aff(C)} ∥y∥_2 and the affine coefficients, α.
12. end while
13. Update x = y and λ = α.
14. end while
15. Return x.
16. end procedure
```

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 (p_j so that x^T p_j < ∥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 monotonically decreases and thus, no corral is visited twice [Wolfe 1976].

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 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 p_j minimizing x^T p_j. Notice that insertion rules are to Wolfe’s method what pivot rules are to the Simplex Method (see [Terlaky and Zhang 1993] for a summary).

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 need 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 p_j minimizing x^T 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 programming 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) [Fujishige et al. 2006] and over the base polyhedron of a submodular function [Fujishige 1980].

Wolfe’s method with the linopt insertion rule (otherwise called the Lawson-Hanson algorithm for non-negative least squares [Lawson and Hanson 1995, Section 23.3]) is 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 [Dantzig 1992]. Gilbert’s procedure for computing the minimum of a quadratic form on a convex set [Gilbert 1966], and the Frank-Wolfe method for convex optimization [Frank and Wolfe 1956] make use of the same active-set selection criterion, the linopt insertion rule. Several of these methods and their variants have been shown to
Wolfe’s method on $P$ with a rate $O$ using the minnorm insertion rule. In the next subsection we will present the first exponential time example using the linopt insertion rule, as evidenced in Table 2.

Consider the simplex $P = \text{conv}\{p_1, p_2, p_3, p_4\} \subset \mathbb{R}^3$ where $p_1 = (0.8, 0.9, 0), p_2 = (1.5, -0.5, 0), p_3 = (-1, -1, 2)$ and $p_4 = (-4, 1, 5, 2)$.

Table 1: iterations for minnorm insertion rule on simplex $P$. Cycle $i,j$ denotes the $i$th major cycle and the $j$th minor cycle within.

<table>
<thead>
<tr>
<th>Cycle</th>
<th>$C$</th>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>${p_1}$</td>
<td>$p_1$</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>${p_1, p_2}$</td>
<td>$p_1$</td>
<td>$(1, 0, 5, 0)$</td>
</tr>
<tr>
<td>2.0</td>
<td>${p_1, p_2, p_3}$</td>
<td>$(1, 0.5, 0)$</td>
<td>$(0.398, 0.199, 0.547)$</td>
</tr>
<tr>
<td>3.0</td>
<td>${p_1, p_2, p_3, p_4}$</td>
<td>$(0.398, 0.199, 0.547)$</td>
<td>$(0, 0)$</td>
</tr>
<tr>
<td>3.1</td>
<td>${p_1, p_2, p_4}$</td>
<td>$(0.288, 0.144, 0.396)$</td>
<td>$(0.198, 0.099, 0.446)$</td>
</tr>
</tbody>
</table>

Table 2: iterations for linopt insertion rule on simplex $P$. Cycle $i,j$ denotes the $i$th major cycle and the $j$th minor cycle within.

<table>
<thead>
<tr>
<th>Cycle</th>
<th>$C$</th>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>${p_1}$</td>
<td>$p_1$</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>${p_1, p_4}$</td>
<td>$p_1$</td>
<td>$(0.222, 0.972, 0.241)$</td>
</tr>
<tr>
<td>2.0</td>
<td>${p_1, p_2, p_3}$</td>
<td>$(0.222, 0.972, 0.241)$</td>
<td>$(0.285, 0.342, 0.581)$</td>
</tr>
<tr>
<td>2.1</td>
<td>${p_1, p_3}$</td>
<td>$(0.286, 0.355, 0.574)$</td>
<td>$(0.277, 0.348, 0.581)$</td>
</tr>
<tr>
<td>3.0</td>
<td>${p_1, p_2, p_4}$</td>
<td>$(0.277, 0.348, 0.581)$</td>
<td>$(0.398, 0.199, 0.547)$</td>
</tr>
<tr>
<td>4.0</td>
<td>${p_1, p_2, p_3, p_4}$</td>
<td>$(0.398, 0.199, 0.547)$</td>
<td>$(0, 0)$</td>
</tr>
<tr>
<td>4.1</td>
<td>${p_1, p_2, p_4}$</td>
<td>$(0.288, 0.144, 0.396)$</td>
<td>$(0.198, 0.099, 0.446)$</td>
</tr>
</tbody>
</table>

2.2 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: 4 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 corals 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 corals are then considered twice by Wolfe’s method.

To simplify the proof, our construction uses a variation of this set of 4 points with an additional point and modified coordinates. This modified construction is depicted in Figure 4, where point 1 corresponds to a set of points $P(d-2)$, points 2, 3 correspond to points $p_a, q_d$ and point 4 corresponds to points $r_d, s_d$.

The high-level idea of our exponential lower bound example is the following. We will inductively define a sequence of instances of increasing dimension of the minimum norm point problem. Given an instance in dimension $d-2$, we will add a few dimensions and points so that, when given to Wolfe’s method, the number of corals of the new augmented instance in dimension $d$ has about twice the number of corals of the input instance in dimension $d-2$. More precisely, our augmentation procedure takes an instance $P(d-2)$ in...
Figure 3: Top: Major cycle 1, minor cycle 0 for the lineopt rule on $P$ illustrates the end of a major cycle; the affine minimizer $y_1 \in \text{relint(conv}(C)) = \text{relint(conv}(p_1, p_4))$. Bottom: Major cycle 2, minor cycle 0 for the lineopt rule on $P$ illustrates the beginning of a minor cycle; the affine minimizer $y_2 \notin \text{relint(conv}(C)) = \text{relint(conv}(p_1, p_4, p_3))$ and the vertex $p_4$ will be removed in the next minor cycle.

Lemma 2.3. Let $A \subseteq \mathbb{R}^d$ be a proper linear subspace. Let $P \subseteq A$ be a non-empty finite set. Let $Q \subseteq A^\perp$ be another non-empty finite set. Let $x$ be the minimum norm point in aff $P$. Let $y$ be the minimum norm point in aff $Q$. Let $z$ be the minimum norm point in aff$(P \cup Q)$. We have:

1. $z$ is the minimum norm point in $[x, y]$ and therefore, if $x \neq 0$
or $y \neq 0$, then $z = \lambda x + (1 - \lambda) y$ with $\lambda = \frac{\|y\|_2^2}{\|x\|_2^2 + \|y\|_2^2}$.
2. If $x \neq 0$ and $y \neq 0$, then $z$ is a strict convex combination of $x$ and $y$.
3. If $x \neq 0$, $y \neq 0$ and $P$ and $Q$ are corrales, then $P \cup Q$ is also a corral.

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 5 gives an example of such a situation in $\mathbb{R}^3$. Denote by $\text{span}(M)$ the linear span of the set $M$. 

Figure 4: Top: 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, $\text{span}(P(d - 2))$ projected into $\text{span}(\mathbb{R}^{d-2})$. Bottom: A two-dimensional view of $P(d)$ projected along the $x_d$ coordinate axis.
Figure 5: Examples of Lemma 2.3. Top: the affine hull of $P \cup Q$ is not full dimensional, and thus the affine minimizer lies at $z$ along the line segment connecting $x = p$ and $y$. Bottom: 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.

**Figure 6:** An example of Lemma 2.4 in which point $q$ satisfies all assumptions and $P \cup \{q\}$ is a corral. The hyperplanes are labeled with their defining properties and demonstrate that $q^T x < \min\{\|x\|^2, \|q\|^2\}$. The minimizer of $P \cup \{q\}$ lies at the intersection of the blue, vertical axis and $\text{conv}(P \cup \{q\})$.

**Lemma 2.4.** Let $P \subseteq \mathbb{R}^d$ be a finite set of points that is a corral. Let $x$ be the minimum norm point in $\text{aff } P$. Let $q \in \text{span}(x, \text{span}(P)^\perp)$, and assume $q^T x < \min\{\|q\|^2, \|x\|^2\}$. Then $P \cup \{q\}$ is a corral. Moreover, the minimum norm point $y$ in $\text{conv}(P \cup \{q\})$ is a (strict) convex combination of $q$ and the minimum norm point of $P$: $y = \lambda x + (1-\lambda)q$ with $\lambda = q^T(q-x)/\|q-x\|^2$.

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 7 demonstrates this situation.

**Lemma 2.5.** For a point $z$ define $H_z = \{ w \in \mathbb{R}^n : w^T z < \|z\|^2 \}$. 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, $Q$, in $A^\perp$. Let $x$ be the minimum norm point in $\text{aff } P$ and $y$ be the minimum norm point in $\text{aff}(P \cup Q)$. Then $H_y \cap A = H_x \cap A$.

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 $P(1) = \{1\} \subseteq Q$. For odd $d > 1$, let $P(d)$ be a list of points in $Q^d$ defined inductively as follows: Let $o^*_{p,d}$ denote the minimum norm point in $\text{conv}(P(d))$. Let $M_{d} := \max_{p \in P(d)} \|p\|_1$, which is a rational upper bound to the maximum 2-norm among the 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 essentially equivalent instance except that it is not rational.) Similarly, let $m_{d} = \|o^*_{p,d}\|_\infty$, which is a rational lower bound to the minimum norm among points in $\text{conv}(P(d))$. (Again, for a first reading one can define $m_{d} = \|o^*_{p,d}\|_2$, which leads to an essentially equivalent instance, except that it is not rational.)
Figure 7: An example of Lemma 2.5 in which adding points $Q$ from $A^\perp$ to points $P$ from $A$ create a new affine minimizer, $z$, 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.

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

$$P(d) = \begin{pmatrix} P(d-2) & 0 & 0 \\ \frac{1}{2} \mathbf{0}_{d-2} & m_{d-2} & M_{d-2} \\ \frac{1}{2} \mathbf{0}_{d-2} & m_{d-2} & -(M_{d-2} + 1) \\ 0 & m_{d-2} & M_{d-2} + 2 \\ 0 & m_{d-2} & -(M_{d-2} + 3) \end{pmatrix}.$$  

The last four rows of the matrix $P(d)$ are the points $p_d, q_d, r_d, s_d$ of the configuration. For a picture of the case of $P(3)$ see Figure 8. We also present the sets $C$, and points $x$ and $y$ defined throughout the iterations of Wolfe’s method with the minnorm insertion rule on $P(3)$ in Table 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 4.

Remark: First note that strictly speaking $P(d-2) \subset Q^{d-2}$, and that we are defining an embedding of it into $Q^d$, for which we have to use a recursive process. To avoid unnecessary notation, we will abuse the notation. The point $v_{d-2}$ denotes both a point of $P(d-2)$ and of the subsequent $P(d)$, i.e., $v_{d-2} = (v, 0, 0)$ will be the identical copy of $v_{d-2}$ within $P(d)$, but we add two extra zero coordinates. Depending on the context $v_{d-2}$ will be understood as both a $(d-2)$-dimensional vector or as a $d$-dimensional vector (e.g., when doing dot products). The points of $P(d-2)$ become a subset of the point configuration $P(d)$ by padding extra zeros. See Figures 4 and 9 which illustrate this embedding and address our visualizations of these sets in three dimensions.

**Theorem 2.6.** Consider the execution of Wolfe’s method with the minnorm point rule on input $P(d)$ where $d = 2k - 1$. Then the sequence of corraals has length $5 \cdot 2^{k-1} - 4$.

The inductive proof of this result relies upon the three lemmas stated above. Each of these are applied in various steps that characterize the sequence of corraals visited for $P(d)$ in terms of the sequence of corraals visited for $P(d-2)$. Moreover, this characterization shows that the length of the sequence of corraals at least doubles in length for each step from $P(d-2)$ to $P(d)$.
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.

Definition 3.1. Consider the following computational problems:

- LP: Given a rational matrix $A$, a rational column vector $b$, and a rational row vector $c^T$, output rational vector $x \in \text{argmax} \{c^T x : Ax \leq b\}$ if $\max \{c^T x : Ax \leq b\}$ is finite, otherwise output INFEASIBLE if $\{x : Ax \leq b\}$ is empty and else output INFINITE.

- DVS: Given $n \leq d + 1$ affinely independent rational points $p_1, p_2, \ldots, p_n \in \mathbb{R}^d$ defining $(n-1)$-dimensional simplex $P = \text{conv} \{p_1, p_2, \ldots, p_n\}$, output $d(0, P)^2$.

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

**Theorem 3.2.** LP reduces to DVS in strongly-polynomial time.

This result is proved via a series of strongly-polynomial reductions which may be found in the full version [De Loera et al. 2017].

The first reductions, which reduce LP to the problem of deciding if the origin lies in a polytope given in vertex-representation (which we call ZVPMD), are classically known, although we do not believe they have all been written elsewhere. The last reduction, which reduces ZVPMD to DVS, is novel and relies upon a careful application of Gram-Schmidt orthogonalization to construct a lift of the polytope to a simplex in higher dimension. We comment that the application of Gram-Schmidt without normalization causes the bit-size of the numbers involved in the problem to grow, but only polynomially, thus preserving the strong-polynomiality of the reduction; see [Grötschel et al. 1988, Section 1.4], [Schrijver 1998, Section 3.3].

### 4 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 $p \geq 1$? Can we identify the types of objective functions for which computing the affine minimizer is easy?

ACKNOWLEDGMENTS

We 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. 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 author 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 while the third author was in residence at the Simons Institute for the Theory of Computing.

REFERENCES


Satoru Fujishige, Takumi Hayashi, and Shiguro Isotani. 2006. The minimum-norm-point algorithm applied to submodular function minimization and linear programming. Kyoto University. Research Institute for Mathematical Sciences [RIMS].