# Alternating Linear Minimization: Revisiting von Neumann's alternating projections

**Gábor Braun** BRAUN@ZIB.DE
*Zuse Institute Berlin*
*Berlin, Germany*

**Sebastian Pokutta** POKUTTA@ZIB.DE
*Zuse Institute Berlin and*
*Institute of Mathematics, Technische Universität Berlin*
*Berlin, Germany*

**Robert Weismantel** ROBERT.WEISMANTEL@IFOR.MATH.ETHZ.CH
*Department of Mathematics, ETH Zürich*
*Zürich, Switzerland*

**Abstract**

In 1933 von Neumann proved a beautiful result that one can approximate a point in the intersection of two convex sets by alternating projections, i.e., successively projecting on one set and then the other. This algorithm assumes that one has access to projection operators for both sets. In this note, we consider the much weaker setup where we have only access to linear minimization oracles over the convex sets and present an algorithm to find a point in the intersection of two convex sets.

## 1. Introduction

We are interested in the following problem and its variants: Given two (compact) convex sets $P$ and $Q$, compute

$$x \in P \cap Q, \tag{1.1}$$

if such an $x$ exists otherwise certify disjointness. Von Neumman's *alternating projection algorithm* (see von Neumann [1949], a reprint of lecture notes first distributed in 1933; see also, e.g., Ginat [2018]) is a very simple algorithm to compute such a point in the intersection of $P$ and $Q$: we alternatingly project onto each of the sets and we are guaranteed to converge to a point in the intersection if such a point exists. This idea has sparked a lot of follow-up work and eventually led to splitting algorithms (see e.g., Combettes & Pesquet [2011] for a great overview), where feasible regions are treated separately and then the results are combined into a convergent scheme. All these approaches rely on projections onto the convex sets, or generalizations of projection, so-called proximal operators.

A different more recent line of work considers the so-called composite optimization framework, where the goal is to minimize the sum of a smooth function and a non-smooth function over some feasible region. To solve Problem (1.1), one minimizes the sum of the convex function $\|x - y\|^2$ and the indicator function of $x \in P$ (which has value 0 if $x$ is in $P$ and has value $+\infty$ otherwise) and the indicator function of $y \in Q$. Composite optimization is usually solved by smoothening the non-smooth summands, e.g., Argyriou et al. [2014] [later rediscovered in Yurtsever et al., 2018] smoothens via the *Moreau envelope* [Bauschke & Combettes, 2017, Moreau, 1965]. This smoothed problem can then be solved with various algorithms, in particular Frank–Wolfe algorithms as done in Argyriou et al. [2014], however computing gradients of the smoothened function involves a proximal operator.

Most closely related to our work are Willner [1968] that compute the distance between two polytopes via the vanilla Frank–Wolfe algorithm with an explicit line search and Wolfe [1976] that computes the nearest point in a polytope using Wolfe's algorithm which can be understood as a quadratic extension of the original Frank–Wolfe approach and which is closely related to the minimum norm point algorithm of Fujishige & Zhan [1990]. Frank–Wolfe algorithms (also called conditional gradients) have the appeal that they access the feasible region only through linear minimization, we refer the interested reader to the survey of Braun et al. [2022] for a broad exposition and background. For our purposes, the Block-coordinate Frank–Wolfe algorithm from Lacoste-Julien et al. [2013] is the starting point, which minimizes a convex function over a product of convex sets, randomly alternating among the convex sets (the blocks). Below we will use its version, the *Cyclic Block-Coordinate Conditional Gradient algorithm (CBCG)* of Beck et al. [2015], which alternates between the convex sets in a cyclic fashion, and thus is a direct analogue of alternating projections with good performance in actual computations [Beck et al., 2015].

While all these lines of work are highly interesting and relevant, here we are concerned with a much simpler and more elementary question. What happens if we change the access model to the feasible regions, i.e., rather than having access to projection operators, we assume that we only have access to *linear minimization oracles (LMO)* for each of the two sets. These oracles, given a linear objective $c \in \mathbb{R}^n$ as input, return

$$\operatorname*{argmin}_{x \in P} \langle c, x \rangle \qquad \text{and} \qquad \operatorname*{argmin}_{y \in Q} \langle c, y \rangle \,,$$

respectively. Compared to projection operators, linear minimization tends to be considerably cheaper, which is one of our motivations; see e.g., Combettes & Pokutta [2021a] for a discussion. The other is that of linear minimization oracles as being somewhat dual to separation oracles. The question is now, whether we can still compute a point in the intersection, having only access to $P$ and $Q$ through their linear minimization oracles.

We answer this question in the affirmative and obtain an *alternating linear minimization* algorithm, that computes $x \in P \cap Q$ or certifies disjointness for two such compact convex sets $P$ and $Q$ whose iteration complexity (up to constants depending on $P$ and $Q$) is basically identical to von Neumann's algorithm. We do only one LMO call per alternation rather than approximating the projection via the Frank–Wolfe algorithm e.g., via a Douglas–Ratchford scheme as done in Millán et al. [2021], typically requiring many LMO calls per each approximation of the projection operation.

## Contribution

We specialize the Cyclic Block-Coordinate Conditional Gradient algorithm to von Neumann's setting, similar to the specialization of the vanilla Frank–Wolfe algorithm in Willner [1968], arriving at our *Alternating Linear Minimizations* Algorithm 3. This algorithm is the direct analog to von Neumann's alternating projection algorithm, however not using projections in each iteration but using linear minimizations only. Surprisingly, we obtain (up to constants) the same rate of convergence of $O(1/t)$ for the distance to the intersection after $t$ linear minimizations as von Neumann's original algorithm albeit with a different dependence on the geometric parameters of the two sets (as expected): we obtain roughly $O\left(\frac{(D_P+D_Q)^2}{t}\right)$ vs $O\left(\frac{\operatorname{dist}(y_0,P\cap Q)^2}{t}\right)$, where $D_p$ and $D_Q$ are diameters of the sets $P$ and $Q$ and $y_0$ is the starting point of the algorithm chosen arbitrarily; see Proposition 4.1 for details. We also outline in Remark 4.7 how our algorithm can be used to not only approximately but exactly compute $x \in P \cap Q$. Moreover, our algorithm certifies non-existence of $x \in P \cap Q$ in case $P \cap Q = \emptyset$ and generates a certifying hyperplane. An interesting corollary of our algorithm is that we can check whether two polytopes (approximately) intersect without an explicit description of the polytopes by inequalities, requiring only access via a linear minimization oracle. This might be of independent interest in discrete geometry. We then provide an adaptive variant in Algorithm 4 that does not require knowledge of problem parameters and decides whether $P \cap Q = \emptyset$ (and provides $x \in P \cap Q$ if exists) add an additive logarithmic cost. Alternating projections converge in finitely many steps under favorable circumstances, see Behling et al. [2021], e.g., if $P$ is a polytope and $Q$ a hyperplane, but there are polytopes $P$ and $Q$ for which they don't converge in finitely many steps, see e.g., [Behling et al., 2021, Example 5.2].

In the following let $\operatorname{dist}(X, Y)$ denote the Euclidean distance between sets $X$ and $Y$, and $\operatorname{dist}(x, X)$ the distance between a point $x$ and a set $X$. The *Frank–Wolfe gap* of a differentiable function $f: P \to \mathbb{R}$ at a point $x \in P$ is

$$\max_{v \in P} \langle \nabla f(x), x - v \rangle . \tag{1.2}$$

Let $\operatorname{vert}(P)$ denote the set of vertices of a polytope $P$. Let $\operatorname{conv}(X)$ denote the convex hull of a set $X$ of points.

All other notation and notions are standard as to be found in Braun et al. [2022].

Outline

In Section 2 we provide a quick recap of von Neumann's alternating projection method, also called POCS, with a short proof for motivation. In Section 3 we then recall the Cyclic Block-Coordinate Conditional Gradient algorithm of Beck et al. [2015] which is the basis for our alternating linear minimization algorithm, which we present in Section 4.

# 2. von Neumann's Alternating Projections

We first briefly recall von Neumann's original algorithm (see von Neumann [1949]; a reprint of his earlier lecture notes from 1933) and present a very elementary proof of its convergence. While von Neumann's original proof applies only to linear subspaces, we consider projections to convex sets; in this setup the algorithm has also been called POCS (Projections Onto Convex Sets) and its convergence is well-known, see e.g., Bauschke & Combettes [2017, Corollary 5.24]. Here we provide an explicit convergence rate.

We consider two closed convex sets $P, Q \subseteq \mathbb{R}^n$ with projectors $\Pi_P$ and $\Pi_Q$, respectively, where for a given $z \in \mathbb{R}^n$ we define $\Pi_P(z) = z_P \doteq \operatorname{argmin}_{x \in P} \|x - z\|^2$ and similarly for $\Pi_Q(z)$, i.e., $z_P$ and $z_Q$ are the projections of $z$ onto $P$ and $Q$, respectively, under the 2-norm. We shall use the following inequality for projection [Dattorro, 2005, proof of Theorem E.9.3.0.1]:

$$\|x - y\|^2 \geq \|\Pi_P(x) - \Pi_P(y)\|^2 + \|x - \Pi_P(x) - y + \Pi_P(y)\|^2 \tag{2.1}$$

generalizing for $x \in P$ the well-known inequality $\|x - y\|^2 \geq \|x - \Pi_P(y)\|^2 + \|\Pi_P(y) - y\|^2$, which is equivalent to $\langle x - \Pi_P(y), \Pi_P(y) - y \rangle \geq 0$. If $P$ and $Q$ are compact then by continuity of distance there is $x^* \in P$ and $y^* \in Q$ with minimal distance, obviously $y^* = \Pi_Q(x^*)$ and $x^* = \Pi_P(y^*)$. We call $x^* - y^*$ the *distance vector* between $P$ and $Q$, which is easily seen by the above inequality to be independent of the choice of $x^*$ and $y^*$. Indeed, for any $x \in P$ and $y \in Q$, by the above we have $\langle x - x^*, x^* - y^* \rangle \geq 0$

$$\begin{aligned} \langle x - y, x^* - y^* \rangle &= \langle x - x^*, x^* - y^* \rangle + \langle x^* - y^*, x^* - y^* \rangle + \langle y^* - y, x^* - y^* \rangle \\ &\geq 0 + \operatorname{dist}(P, Q)^2 + 0 = \operatorname{dist}(P, Q)^2. \end{aligned} \tag{2.2}$$

Thus if $x \in P$ and $y \in Q$ also have minimal distance, i.e., $\|x - y\| = \operatorname{dist}(P, Q) = \|x^* - y^*\|$, then necessarily $x - y = x^* - y^*$.

**Algorithm 1** POCS, von Neumann's Alternating Projections

**Input:** Point $y_0 \in \mathbb{R}^n$, $\Pi_P$ projector onto $P \subseteq \mathbb{R}^n$ and $\Pi_Q$ projector onto $Q \subseteq \mathbb{R}^n$.
**Output:** Iterates $x_1, y_1 \ldots \in \mathbb{R}^n$

**for** $t = 0$ **to** $\ldots$ **do**
  $x_{t+1} \leftarrow \Pi_P(y_t)$
  $y_{t+1} \leftarrow \Pi_Q(x_{t+1})$

**Proposition 2.1.** *Let $P$ and $Q$ be compact convex sets and let $x_1, y_1 \ldots, x_T, y_T \in \mathbb{R}^n$ be the sequence of iterates of Algorithm 1. Let $Q_{\min}$ be the sets of points of $Q$ with minimal distance to $P$, and $d$ the distance vector between $Q$ and $P$, i.e., $d = z - \Pi_P(z)$ for all $z \in Q_{\min}$. Then the iterates converge: $x_t \to x$ and $y_t \to y$ to some $x \in P$ and $y \in Q$ with $x - y = d$. Moreover,*

$$\min_{t=0,\ldots,T-1} \|y_t - x_{t+1} - d\|^2 + \|x_{t+1} - y_{t+1} + d\|^2 \leq \frac{1}{T} \sum_{t=0}^{T-1} \left(\|y_t - x_{t+1} - d\|^2 + \|x_{t+1} - y_{t+1} + d\|^2\right) \leq \frac{\operatorname{dist}(y_0, Q_{\min})^2}{T}. \tag{2.3}$$

*In particular, if $P$ and $Q$ intersect (i.e., $Q_{\min} = P \cap Q$ and $d = 0$) then*

$$\|x_T - y_T\|^2 \leq \frac{1}{T} \sum_{t=0}^{T-1} \left(\|y_t - x_{t+1}\|^2 + \|x_{t+1} - y_{t+1}\|^2\right) \leq \frac{\operatorname{dist}(y_0, P \cap Q)^2}{T}. \tag{2.4}$$

*Proof.* Let $z_2 \in Q_{\min}$ and $z_1 = \Pi_P(z_2) \in P$. Obviously, $\|z_1 - z_2\| = \operatorname{dist}(P, Q)$, and hence $z_2 = \Pi_Q(z_1)$ and $d = z_2 - z_1$.

We estimate the summands in the claim via the projection inequality (2.1):

$$\|x_{t+1} - y_{t+1} + d\|^2 = \|x_{t+1} - y_{t+1} - z_1 + z_2\|^2 \leq \|x_{t+1} - z_1\|^2 - \|y_{t+1} - z_2\|^2 \tag{2.5a}$$
$$\|y_t - x_{t+1} - d\|^2 = \|y_t - x_{t+1} - z_2 + z_1\|^2 \leq \|y_t - z_2\|^2 - \|x_{t+1} - z_1\|^2. \tag{2.5b}$$

Summing up provides a telescope sum:

$$\begin{aligned} &\sum_{t=0}^{T-1} \left(\|x_{t+1} - y_{t+1} + d\|^2 + \|y_t - x_{t+1} - d\|^2\right) \\ &\qquad \leq \sum_{t=0}^{T-1} \left[(\|x_{t+1} - z_1\|^2 - \|y_{t+1} - z_2\|^2) + (\|y_t - z_2\|^2 - \|x_{t+1} - z_1\|^2)\right] \\ &\qquad\qquad = \|y_0 - z_2\|^2 - \|y_T - z_2\|^2 \leq \|y_0 - z_2\|^2. \end{aligned}$$

Minimizing over $z_2$ proves the second inequality of Equation (2.3). The first inequality of Equation (2.3) is obvious.

From Equation (2.3) it follows that $x_t - y_t$ converges to $d$. Next we prove that the individual sequences $x_t$ and $y_t$ converge. As $Q$ is compact, the sequence $y_t$ has some accumulation point $y$. Then $x \doteq y + d$ is an accumulation point of $x_t$ and thus $x \in P$ and $y \in Q$ with $x - y = d$, hence $y \in Q_{\min}$.

By Equations (2.5), we have $\|y_t - z_2\|^2 \geq \|x_{t+1} - \Pi_P(z_2)\|^2 \geq \|y_{t+1} - z_2\|^2$ for every $z_2 \in Q_{\min}$, in particular, $\|y_t - z_2\|$ is a decreasing sequence. This means for the accumulation point $z_2 \doteq y$ that $\|y_t - y\|$ must converge to 0, i.e., $y_t$ converges to $y$. Therefore $x_t$ converges to $y + d = x$.

Finally, when $P$ and $Q$ intersect, i.e., $d = 0$ then as projections select minimal distance points, the distances are decreasing: $\|x_t - y_t\|^2 \geq \|x_{t+1} - y_t\|^2 \geq \|x_{t+1} - y_{t+1}\|^2$, and therefore the minimum in (2.3) simplifies. □

## 3. The Cyclic Block-Coordinate Conditional Gradient algorithm

Here we recall the Cyclic Block-Coordinate Conditional Gradient algorithm (CBCG) of Beck et al. [2015] as Algorithm 2, using the notation of its predecessor from Lacoste-Julien et al. [2013]. The algorithm solves the optimization problem:

$$\min_{(x_0,\ldots,x_{k-1})\in P_0\times\cdots\times P_{k-1}} f(x_0,\ldots,x_{k-1}), \tag{3.1}$$

where $f$ is a convex function and the $P_i$ are convex sets. This problem can be also solved with the original Frank–Wolfe algorithm [Frank & Wolfe, 1956], however, for this paper the cyclic variant in Algorithm 2 is more suitable. CBCG makes a single linear minimization in only one block $P_i$ in every iteration, and blocks are selected in a simple cyclic order.

We follow a similar notation as in [Lacoste-Julien et al., 2013]. The vector $s_{[i]}$ denotes the vector with $s$ in $P_i$ and 0 elsewhere, i.e., it is the natural generalization of the coordinate vectors. We now state the convergence rate of the Cyclic Block-Coordinate Conditional Gradient algorithm from Beck et al. [2015] with a slight improvement in the constants for the agnostic step rule: the denominator is improved from $t+2$ to $t+1$, and the constant factor 6.75 for the dual gap is a bit smaller.

**Algorithm 2** Cyclic Block-Coordinate Conditional Gradient algorithm [Beck et al., 2015]

**Input:** Points $x_i^0 \in P_i$, LMO for $P_i \subseteq \mathbb{R}^{n_i}$, $i = 0,\ldots,k-1$ and $0 < \gamma_0,\ldots,\gamma_t,\ldots \leq 1$.
**Output:** Iterates $x^1,\ldots \in P_0\times\cdots\times P_{k-1}$

1: **for** $t = 0$ **to** $\ldots$ **do**
2: $\quad i \leftarrow t \bmod k$
3: $\quad v^t \leftarrow \operatorname{argmin}_{x\in P_i} \langle \nabla_{P_i} f(x^t), x\rangle$
4: $\quad x^{t+1} \leftarrow x^t + \gamma_t (v^t - x_i^t)_{[i]}$

**Theorem 3.1** (Convergence of Cyclic Block-Coordinate Conditional Gradient algorithm [Beck et al., 2015, cf Theorem 4.5])**.** *Let $P_i$ be a compact convex set with diameter $D_i$ for $i = 0,1,\ldots,k-1$. Let $f\colon P_0\times\cdots\times P_{k-1} \to \mathbb{R}$ be an $L$-smooth convex function, which is moreover partially $L_i$-smooth in $P_i$ for $i = 0,1,\ldots,k-1$. Let $D \doteq \sqrt{\sum_{i=0}^{k-1} D_i^2}$ be the diameter of the feasible region $P_0\times\cdots\times P_{k-1}$. Then the iterates of Algorithm 2 with the choice $\gamma_t = \frac{2}{\lfloor t/k\rfloor+2}$ satisfy*

$$f(x^{kt}) - f(x^*) \leq \frac{2}{t+2}\left(\sum_{i=0}^{k-1}\frac{L_i D_i^2}{2} + 2LD\sum_{i=0}^{k-1} D_i\right),$$
$$\min_{1\leq t\leq T}\max_{y\in P_0\times\cdots\times P_{k-1}} \langle \nabla f(x^{kt}), x^{kt} - y\rangle \leq \frac{6.75}{T+2}\left(\sum_{i=0}^{k-1}\frac{L_i D_i^2}{2} + 2LD\sum_{i=0}^{k-1} D_i\right).$$

*where $x^*$ is a minimizer of $f$.*

*Proof.* By Beck et al. [2015, Lemma 4.3]

$$f(x^{kt+k}) - f(x^{kt}) - \gamma_{kt}\langle \nabla f(x^{kt}), u^{kt} - x^{kt}\rangle \leq \gamma_{kt}^2\left(\sum_{i=0}^{k-1}\frac{L_i D_i^2}{2} + 2LD\sum_{i=0}^{k-1} D_i\right). \tag{3.2}$$

This differs from the standard recursion for the vanilla Frank–Wolfe algorithm only by replacing the smoothness constant with the parenthesized expression on the right multiplied by 2. Therefore the rest of the proof is standard, see e.g., Jaggi [2013] or Braun et al. [2022]. ◻

**Theorem 3.2** (Convergence of Cyclic Block-Coordinate Conditional Gradient algorithm [Beck et al., 2015, Theorem 4.13]). *Let $P_i$ be a compact convex set with diameter $D_i$ for $i = 0, 1, \ldots, k-1$. Let $f\colon P_0 \times \cdots \times P_{k-1} \to \mathbb{R}$ be an $L$-smooth convex function, which is moreover partially $L_i$-smooth and partially $G_i$-Lipschitz in $P_i$ for $i = 0, 1, \ldots, k-1$. Let $D \doteq \sqrt{\sum_{i=0}^{k-1} D_i^2}$ be the diameter of the feasible region $P_0 \times \cdots \times P_{k-1}$. Then the iterates of Algorithm 2 with the short step size rule $\gamma_t = \min\{\langle \nabla_{P_i} f(x^t), x_i^t - v^t \rangle / (L_i \|x_i^t - v^t\|^2), 1\}$ satisfy*

$$f(x^{kt}) - f(x^*) \leq \frac{4k}{t+4} \left( \max_{i=0,\ldots,k-1} \{L_i D_i^2, G_i D_i\} + \frac{k L^2 D^2}{\min_{i=0,\ldots,k-1} L_i} \right),$$

$$\min_{1 \leq t \leq T} \max_{y \in P_0 \times \cdots \times P_{k-1}} \langle \nabla f(x^{kt}), x^{kt} - y \rangle \leq \frac{8k}{T+4} \left( \max_{i=0,\ldots,k-1} \{L_i D_i^2, G_i D_i\} + \frac{k L^2 D^2}{\min_{i=0,\ldots,k-1} L_i} \right).$$

*where $x^*$ is a minimizer of $f$.*

Note that typically the short-step step-size rule used in Theorem 3.2 performs much better in actual computations than the agnostic step-size rule used in Theorem 3.1. Therefore we will state our results for both choices in the remainder.

# 4. Alternating Linear Minimizations

We are returning to our original problem of interest: Let $P$ and $Q$ be compact convex sets contained in the same ambient space $\mathbb{R}^n$, and the goal is to find points in them with minimal distance. In contrast to von Neumann's approach, we will assume that we have only access to two Linear Minimization Oracles (LMOs) for the compact convex sets $P \subseteq \mathbb{R}^n$ and $Q \subseteq \mathbb{R}^n$, respectively, i.e., given a vector $c \in \mathbb{R}^n$ the oracles return:

$$v \leftarrow \operatorname*{argmin}_{x \in P} \langle c, x \rangle \qquad \text{and} \qquad w \leftarrow \operatorname*{argmin}_{y \in Q} \langle c, y \rangle .$$

We adapt Algorithm 2 to find points in $P$ and $Q$ with approximate minimal distance, using the objective function $f(x, y) = \|x - y\|^2$ over $P \times Q$.

**Algorithm 3** Alternating Linear Minimizations (ALM)

**Input:** Points $x_0 \in P$, $y_0 \in Q$, LMO over $P, Q \subseteq \mathbb{R}^n$
**Output:** Iterates $x_1, y_1 \ldots \in \mathbb{R}^n$

```
for t = 0 to ... do
    u_t ← argmin_{x∈P} ⟨x_t − y_t, x⟩
    x_{t+1} ← x_t + γ_{t,1} · (u_t − x_t)
    v_t ← argmin_{y∈Q} ⟨y_t − x_{t+1}, y⟩
    y_{t+1} ← y_t + γ_{t,2} · (v_t − y_t)
```

Theorem 3.1 provides the following guarantees:

**Proposition 4.1** (Intersection of two sets). *Let $P$ and $Q$ be compact convex sets. Then Algorithm 3 with $\gamma_{t,1} = \gamma_{t,2} \doteq \frac{2}{t+2}$ (agnostic step-size rule) generates iterates $z_t \doteq \frac{1}{2}(x_t + y_t)$, such that*

$$\max\{\operatorname{dist}(z_t, P)^2, \operatorname{dist}(z_t, Q)^2\} \leq \frac{\|x_t - y_t\|^2}{4} \leq \frac{(1 + 2\sqrt{2})(D_P^2 + D_Q^2)}{t+2} + \frac{\operatorname{dist}(P, Q)^2}{4} \tag{4.1}$$

$$\min_{1 \leq t \leq T} \max_{x \in P, y \in Q} \|x_t - y_t\|^2 - \langle x_t - y_t, x - y \rangle \leq \frac{6.75(1 + 2\sqrt{2})}{T+2} (D_P^2 + D_Q^2). \tag{4.2}$$

*With $\gamma_{t,1} \doteq \min\{\langle x_t - y_t, x_t - u_t\rangle / \|x_t - u_t\|^2, 1\}$ and $\gamma_{t,2} \doteq \min\{\langle y_t - x_{t+1}, y_t - v_t\rangle / \|y_t - v_t\|^2, 1\}$ (short-step step-size rule) the iterates of Algorithm 3 satisfy*

$$\max\{\operatorname{dist}(z_t, P)^2, \operatorname{dist}(z_t, Q)^2\} \leq \frac{\|x_t - y_t\|^2}{4} \leq \frac{4c}{t+4} + \frac{\operatorname{dist}(P,Q)^2}{4} \tag{4.3}$$

$$\min_{1\leq t\leq T} \max_{x\in P, y\in Q} \|x_t - y_t\|^2 - \langle x_t - y_t, x - y\rangle \leq \frac{16c}{T+4} \tag{4.4}$$

$$c \doteq (D_P + D_Q + \operatorname{dist}(P,Q)) \cdot \max\{D_P, D_Q\} + 2D^2. \tag{4.5}$$

*Proof.* First note that $\operatorname{dist}(z_t, P)^2 \leq \|x_t - z_t\|^2 = \left\|x_t - \frac{1}{2}(x_t + y_t)\right\|^2 = \|x_t - y_t\|^2 / 4$ and similarly for $\operatorname{dist}(z_t, P) \leq \|x_t - y_t\| / 2$. Combining this with Theorem 3.1, we obtain the claim. □

The rates obtained in Proposition 4.1 are essentially tight up to constant factors. This can be seen, when choosing $P$ to be a polytope and $Q$ to be a single point. Then the problem under consideration reduces to the approximate Carathéodory problem (and our algorithm to the vanilla Frank–Wolfe algorithm) for which matching lower bounds are known [Combettes & Pokutta, 2021b, Mirrokni et al., 2017].

It is informative to compare the rate obtained in Proposition 4.1 with the rate in Proposition 2.1.

**Remark 4.2** (Comparison to von Neumann's alternating projection algorithm)**.** For simplicity let us consider the case where $P \cap Q \neq \emptyset$. Via Proposition 2.1 we then obtain the following guarantee for von Neumann's alternating projection algorithm:

$$\|x_T - y_T\|^2 \leq \frac{\operatorname{dist}(y_0, P\cap Q)^2}{T}. \tag{4.6}$$

With $z_t \doteq \frac{1}{2}(x_t + y_t)$, we have

$$\max\{\operatorname{dist}(z_{t+1}, P)^2, \operatorname{dist}(z_{t+1}, Q)^2\} \leq \|x_{t+1} - z_{t+1}\|^2 + \|y_{t+1} - z_{t+1}\|^2 \leq \frac{1}{2}\|x_{t+1} - y_{t+1}\|^2.$$

Combining with the above and estimating generously we obtain for von Neumann's alternating projection method:

$$\min_{t=0,\ldots,T-1} \max\{\operatorname{dist}(z_{t+1}, P)^2, \operatorname{dist}(z_{t+1}, Q)^2\} \leq \frac{\operatorname{dist}(y_0, P\cap Q)^2}{T}. \tag{4.7}$$

In contrast to this, the LMO-based approach guarantees via Proposition 4.1 after $T$ many iterations:

$$\max\{\operatorname{dist}(z_T, P)^2, \operatorname{dist}(z_T, Q)^2\} \leq \frac{(1+2\sqrt{2})(D_P^2 + D_Q^2)}{T+2}.$$

So both approaches essentially guarantee an $O(1/T)$ rate of convergence, however the constants are quite different and not necessarily comparable, depending on the regime. The vanilla Frank–Wolfe algorithm applied directly to the function $f(x,y) = \|x-y\|^2$ would essentially lead to the same convergence guarantee as Algorithm 3 (up to a small constant factor of $\sqrt{2}$), however the processing would not be alternating between the blocks anymore.

Finally, running von Neumann's original algorithm with a näive simulation of the projection operator via the vanilla Frank–Wolfe method (which only requires LMOs) would roughly require an additional number of $\Omega(t)$ LMO calls per iteration. This assumes the target accuracy is at least $O(1/t)$ which might be required to solve the projection problem with reasonably high accuracy. Thus the overall convergence rate in the number of LMO calls would be roughly $O(1/\sqrt{T})$, i.e., much slower than our Algorithm 3.

We can also use the same approach to certify that the intersection of two sets $P$ and $Q$ is empty. To this end, we simply run the algorithm until the function value $f(x_t, y_t)$ is strictly larger than the primal or dual gap.

**Corollary 4.3** (Certifying that $P \cap Q = \emptyset$)**.** *Let $P, Q$ be compact convex sets with diameters $D_P$ and $D_Q$, respectively. If some iterates of Algorithm 3 using $\gamma_{t,1} = \gamma_{t,2} = 2/(t+2)$ satisfy*

$$\|x_t - y_t\|^2 > \frac{4(1+2\sqrt{2})(D_P^2 + D_Q^2)}{t+2}$$

*then $P$ and $Q$ are disjoint. If $P$ and $Q$ are disjoint then the above condition is satisfied after at most*

$$\frac{8(1+2\sqrt{2})(D_P^2 + D_Q^2)}{\operatorname{dist}(P,Q)^2}$$

*block-LMO calls.*

*Under the short step rule, $P$ and $Q$ are disjoint if for some iterate $t$*

$$\|x_t - y_t\|^2 > \frac{16}{t+4}\Big((D_P + D_Q) \cdot \max\{D_P, D_Q\} + 2D^2\Big). \tag{4.8}$$

*If $P$ and $Q$ are disjoint then this condition is satisfied after at most the following number of block-LMO calls*

$$\frac{32}{(t+4)\operatorname{dist}(P,Q)^2}\Big((D_P + D_Q) \cdot \max\{D_P, D_Q\} + 2D^2\Big). \tag{4.9}$$

*Proof.* We prove only the claims for the agnostic step step size rule $\gamma_{t,1} = \gamma_{t,2} = 2/(t+2)$ as the proof for the short step rule is similar.

First we prove the first claim using the following chain of inequalities, the first inequality by Proposition 4.1, the second inequality because $x_t \in P$ and $y_t \in Q$:

$$\operatorname{dist}(P,Q)^2 \geq \|x_t - y_t\|^2 - \frac{4(1+2\sqrt{2})(D_P^2 + D_Q^2)}{t+2} \geq \operatorname{dist}(P,Q)^2 - \frac{4(1+2\sqrt{2})(D_P^2 + D_Q^2)}{t+2}.$$

By the first inequality, $\|x_t - y_t\|^2 > \frac{4(1+2\sqrt{2})(D_P^2+D_Q^2)}{t+2}$ obviously implies $\operatorname{dist}(P,Q) > 0$, i.e., that $P$ and $Q$ are disjoint. By the second inequality, $\|x_t - y_t\|^2 > \frac{4(1+2\sqrt{2})(D_P^2+D_Q^2)}{t+2}$ for $t > 4(D_P+D_Q)^2/\operatorname{dist}(P,Q)^2$. As every iteration of Algorithm 3 costs two block-LMO calls, this provides the upper bound $8(1 + 2\sqrt{2})(D_P^2 + D_Q^2)/\operatorname{dist}(P,Q)^2$ on the number of block-LMO calls.

For the second claim, note that Equation (4.10) obviously implies that $P$ and $Q$ are disjoint. Let us assume that $\min_{x\in P, y\in Q} \langle x_t - y_t, x - y\rangle \leq 0$ for all $1 \leq t \leq T$. By (4.2),

$$\operatorname{dist}(P,Q)^2 \leq \min_{1\leq t\leq T} \|x_t - y_t\|^2 \leq \min_{1\leq t\leq T} \max_{x\in P, y\in Q} \|x_t - y_t\|^2 - \langle x_t - y_t, x - y\rangle \leq \frac{6.75(1+2\sqrt{2})}{T+2}(D_P^2 + D_Q^2).$$

Therefore $T + 2 \leq \frac{6.75(1+2\sqrt{2})}{\operatorname{dist}(P,Q)^2}(D_P^2 + D_Q^2)$. Thus, Inequality (4.10) is satisfied by some among the first $\frac{6.75(1+2\sqrt{2})}{\operatorname{dist}(P,Q)^2}(D_P^2 + D_Q^2)$ iterations. The upper bound on the block LMO calls to satisfy the inequality is twice this number since every iteration costs two block LMO calls as above. ∎

With a slightly worse constant factor we can also certify and test for disjointness without knowing $D_P$ and $D_Q$ explicitly.

**Corollary 4.4** (Certifying $P \cap Q = \emptyset$ without parameters)**.** *Let $P, Q$ be disjoint compact convex sets with diameters $D_p$ and $D_Q$, respectively. Then executing Algorithm 3, after at most*

$$\frac{13.5(1+2\sqrt{2})(D_P^2 + D_Q^2)}{\operatorname{dist}(P,Q)^2} \qquad \text{with step sizes } \gamma_{t,1} = \gamma_{t,2} = 2/(t+2)$$

$$\frac{32}{\operatorname{dist}(P,Q)^2}\Big((D_P + D_Q + \operatorname{dist}(P,Q)) \cdot \max\{D_P, D_Q\} + 2D^2\Big) \qquad \text{with short step rule}$$

*block-LMO calls, some (of the already seen!) iteration $t$ provides the following certificate for disjointness, which does not require explicit bounds on $D_P$ and $D_Q$ (in contrast to the above):*

$$\min_{x \in P, y \in Q} \langle x_t - y_t, x - y \rangle > 0. \tag{4.10}$$

*Moreover, this inequality is guaranteed to hold for every iteration $t > 4(1+2\sqrt{2})(D_P^2 + D_Q^2)(D_P + D_Q)^2/\operatorname{dist}(P,Q)^4$, i.e., after $8(1+2\sqrt{2})(D_P^2 + D_Q^2)(D_P + D_Q)^2/\operatorname{dist}(P,Q)^4$ block-LMO calls under the agnostic step size rule $\gamma_{t,1} = \gamma_{t,2} = 2/(t+2)$, and after $32[(D_P + D_Q + \operatorname{dist}(P,Q)) \cdot \max\{D_P, D_Q\} + 2D^2](D_P + D_Q)^2/\operatorname{dist}(P,Q)^4$ block-LMO calls under the short step rule.*

*Proof.* Clearly, Inequality (4.10) implies disjointness of $P$ and $Q$, and by Proposition 4.1 it is satisfied for some of the first claimed number of iterations. So we only need to prove that the inequality holds for every late iteration.

We prove the claim only for the agnostic step size rule $\gamma_{t,1} = \gamma_{t,2} = 2/(t+2)$, as the proof for the short step rule is similar.

Let $d = x^* - y^*$ be the distance vector between $P$ and $Q$ with $x^* \in P$ and $y^* \in Q$. Using Inequality (2.2),

$$\|x_t - y_t - d\|^2 = \|x_t - y_t\|^2 - 2\langle x_t - y_t, d\rangle + \|d\|^2 \le \|x_t - y_t\|^2 - \operatorname{dist}(P,Q)^2.$$

Applying Inequality (2.2) again, we obtain for any $x \in P$ and $y \in Q$

$$\begin{aligned}
\langle x_t - y_t, x - y\rangle &= \langle x_t - y_t - d, x - y - d\rangle + \langle d, x - y\rangle + \langle x_t - y_t, d\rangle - \langle d, d\rangle \\
&\ge -\|x_t - y_t - d\|(D_p + D_Q) + 2\operatorname{dist}(P,Q)^2 - \operatorname{dist}(P,Q)^2 \\
&\ge \operatorname{dist}(P,Q)^2 - \sqrt{\|x_t - y_t\|^2 - \|d\|^2}(D_p + D_Q) \\
&\ge \operatorname{dist}(P,Q)^2 - \frac{2\sqrt{(1+2\sqrt{2})(D_P^2 + D_Q^2)}(D_P + D_q)}{\sqrt{t+2}}.
\end{aligned}$$

Thus Inequality (4.10) holds for $t > 4(1+2\sqrt{2})(D_P^2 + D_Q^2)(D_P + D_Q)^2/\operatorname{dist}(P,Q)^4$ as claimed. ◻

**Remark 4.5** ($V$-representation vs. $H$-representation vs. implicit description)**.** Given two polytopes $P$ and $Q$ in $H$-representation, i.e., each given as a set of defining inequalities, then it is well-known to decide whether $P \cap Q = \emptyset$ and compute $x \in P \cap Q$ via Farkas' lemma and linear programming techniques.

If the polytopes are given in $V$-representation (i.e., as finite set whose convex hull is the polytope), then we can resolve the question similarly: Let $P = \operatorname{conv}(U)$ and $Q = \operatorname{conv}(W)$. We consider the extended formulations arising from $U$ and $W$ and check the feasibility of the linear program

$$\left\{(\lambda, \kappa) : \sum_{u \in U} \lambda_u u = \sum_{w \in W} \kappa_w w, \sum_{u \in U} \lambda_u = \sum_{w \in W} \kappa_w = 1, \lambda, \kappa \ge 0\right\}.$$

The mixed case can be decided similarly.

If $P$ and $Q$ are only given implicitly via linear optimization oracles, it is *a priori* not obvious how to decide whether the polytopes intersect. However, we can use Algorithm 3 to, at least, approximately decide this question and use Remark 4.7 to make the test exact.

The careful reader will have realized that Algorithm 3 only provides an approximate answer as to whether $x \in P \cap Q$ exists and in fact, it only decides whether there exists a point $x \in \mathbb{R}^n$ that is close to both $P$ and $Q$. For the case where $P$ and $Q$ are polytopes, we will now show that there exists $\varepsilon_{PQ} > 0$ depending on $P$ and $Q$, such that if we obtain iterates $x_t$ and $y_t$ via Algorithm 3 with $\|x_t - y_t\| \leq \varepsilon_{PQ}$, then we can compute an actual point $x \in P \cap Q$.

To this end we start with a simple observation.

**Observation 4.6.** Let $P, Q \subseteq \mathbb{R}^n$ be polytopes. There exists $\varepsilon_{PQ} > 0$, so that for all $U \subseteq \operatorname{vert}(P), V \subseteq \operatorname{vert}(Q)$ with $\operatorname{dist}(\operatorname{conv}(U), \operatorname{conv}(V)) < \varepsilon_{PQ}$, it holds $\operatorname{conv}(U) \cap \operatorname{conv}(V) \neq \emptyset$.

*Proof.* This easily follows from the fact that polytopes having only a finite number of vertices, so that the simple definition provides a positive number:

$$\varepsilon_{PQ} := \min\{\operatorname{dist}(\operatorname{conv}(U), \operatorname{conv}(V)) : U \subseteq \operatorname{vert}(P), V \subseteq \operatorname{vert}(Q), \operatorname{conv}(U) \cap \operatorname{conv}(V) = \emptyset\}. \qquad \square$$

With this observation we obtain now:

**Remark 4.7** (Recovery of $x \in P \cap Q$ by linear programming)**.** Once we obtain $x_t$ and $y_t$ with $\|x_t - y_t\| < \varepsilon_{PQ}$ via Algorithm 3, we can recover $x \in P \cap Q$ by means of linear programming. To this end let $U \subseteq \operatorname{vert}(P)$ be all extreme points returned by the LMO for $P$ throughout the execution of Algorithm 3 and define $V \subseteq \operatorname{vert}(Q)$ accordingly. From Observation 4.6 it follows that $\operatorname{conv}(U) \cap \operatorname{conv}(V) \neq \emptyset$ and we can solve the linear feasibility program:

$$\sum_{u \in U} \lambda_u u = \sum_{v \in V} \kappa_u v \tag{4.11a}$$
$$\sum_{u \in U} \lambda_u = 1 \tag{4.11b}$$
$$\sum_{v \in V} \kappa_u = 1 \tag{4.11c}$$
$$\lambda_u \geq 0 \qquad \forall u \in U \tag{4.11d}$$
$$\kappa_v \geq 0 \qquad \forall v \in V, \tag{4.11e}$$

which can be solved with the ellipsoid method efficiently. This linear feasibility problem is guaranteed to be feasible and the obtained solution satisfies $x := \sum_{u \in U} \lambda_u u = \sum_{v \in V} \kappa_u v \in P \cap Q$.

If $\varepsilon_{PQ}$ is not known ahead of time, which will usually be the case, we can simply try to solve the above linear program whenever the distance $\|x_t - y_t\|$ is halved. This leads to a logarithmic overhead of about $O(\log 1/\varepsilon_{PQ})$ solved linear programs.

In fact we can turn Algorithm 3 into an exact adaptive algorithm (Algorithm 4) for polytopes not requiring knowledge of any parameters such as $\varepsilon_{PQ}$ and diameters if we permit ourselves an additive logarithmic overhead in terms of LMO calls and linear programs. The algorithm can be also run for more general regions however then the guarantees are not as clean.

**Algorithm 4** Alternating Linear Minimizations (ALM) (adaptive variant)

**Input:** Points $x_0 \in P$, $y_0 \in Q$, LMO over $P, Q \subseteq \mathbb{R}^n$
**Output:** $x \in P \cap Q$ or "disjoint"

**for** $t = 0$ **to** $\ldots$ **do**
  $u_t \leftarrow \operatorname{argmin}_{x \in P} \langle x_t - y_t, x \rangle$
  $x_{t+1} \leftarrow x_t + \gamma_{t,1} \cdot (u_t - x_t)$
  $v_t \leftarrow \operatorname{argmin}_{y \in Q} \langle y_t - x_{t+1}, y \rangle$
  $y_{t+1} \leftarrow y_t + \gamma_{t,2} \cdot (v_t - y_t)$
  **if** $t = 2^k$ for some $k$ **then**
    **if** $\min_{x \in P, y \in Q} \langle x_{t+1} - y_{t+1}, x - y \rangle > 0$ **then**
      **return** "disjoint" and certificate $\langle x_{t+1} - y_{t+1}, x - y \rangle > 0$
    **else**
      Solve linear program (4.11).
      **if** feasible **then**
        **return** a solution $x \in P \cap Q$

We will now state the guarantees for Algorithm 4 for the case of polytopes. We count solving (4.11) as one block-LMO call.

**Proposition 4.8** (Adaptive variant). *Let $P, Q$ be polytopes with diameters $D_P$ and $D_Q$, respectively. Executing Algorithm 4, then:*

1. *If $P \cap Q \neq \emptyset$, then after no more than*

$$\frac{16(1 + 2\sqrt{2})(D_P^2 + D_Q^2)}{\varepsilon_{PQ}^2}$$

*block-LMO calls, the algorithm returns $x \in P \cap Q$ with the agnostic step-size rule and no more than*

$$\frac{64((D_P + D_Q + \operatorname{dist}(P, Q)) \cdot \max\{D_P, D_Q\} + 2D^2)}{\varepsilon_{PQ}^2}$$

*block-LMO calls with the short-step step-size rule.*

2. *If $P \cap Q = \emptyset$, then after no more than*

$$16(1 + 2\sqrt{2})(D_P^2 + D_Q^2)\frac{(D_P + D_Q)^2}{\operatorname{dist}(P, Q)^4}$$

*block-LMO calls the algorithm certifies $P \cap Q = \emptyset$ with the agnostic step-size rule and no more than*

$$64((D_P + D_Q + \operatorname{dist}(P, Q)) \cdot \max\{D_P, D_Q\} + 2D^2)\frac{(D_P + D_Q)^2}{\operatorname{dist}(P, Q)^4}$$

*block-LMO calls with the short-step step-size rule.*

*Proof.* If $P \cap Q \neq \emptyset$ and using the agnostic step size rule, then by Proposition 4.1, after $T \doteq 4(1 + 2\sqrt{2})(D_P^2 + D_Q^2)/\varepsilon_{PQ}^2 - 2$ iterations one has $\|x_t - y_t\| < \varepsilon_{PQ}$ for all $t \geq T$, and hence by Remark 4.7 solving (4.11) provides a point $x$ in the intersection. Due to delayed tests, the algorithm executes at most $2T$ iterations, makes one linear minimization in each of them, and 3 additional linear minimizations in at most $\log_2 T + 2$ iterations, however among these one is identical to the linear minimization in the following round. Thus at most $2T + 1 + 2(\log_2 T + 2) \leq 4(T + 2)$ linear minimizations are made.

The other bounds are obtained similarly. The case $P \cap Q = \emptyset$ follows via Corollary 4.4, where the bound on block-LMO calls is twice the bound on iterations. $\square$

## Acknowledgements

Research reported in this paper was partially supported by the DFG Cluster of Excellence MATH+ (EXC-2046/1, project id 390685689) funded by the Deutsche Forschungsgemeinschaft (DFG). The third author is supported by the Einstein Foundation Berlin.

## References

Argyriou, A., Signoretto, M., and Suykens, J. A. K. Hybrid conditional gradient-smoothing algorithms with applications to sparse and low rank regularization. In *Regularization, Optimization, Kernels, and Support Vector Machines*, pp. 53–82. Chapman & Hall/CRC, 2014. ISBN 1482241390, 9781482241396.

Bauschke, H. H. and Combettes, P. L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, 2 edition, 2017. ISBN 978-3-319-48310-8 (hardcover); 978-3-319-83911-0 (softcover); 978-3-319-48311-5 (ebook). doi: 10.1007/978-3-319-48311-5.

Beck, A., Pauwels, E., and Sabach, S. The cyclic block conditional gradient method for convex optimization problems. *SIAM Journal on Optimization*, 25(4):2024–2049, 2015. ISSN 1052-6234 (print); 1095-7189 (online). doi: 10.1137/15M1008397.

Behling, R., Bello-Cruz, Y., and Santos, L.-R. Infeasibility and error bound imply finite convergence of alternating projections. *SIAM Journal on Optimization*, 31(4):2863–2892, 2021.

Braun, G., Carderera, A., Combettes, C. W., Hassani, H., Karbasi, A., Mokthari, A., and Pokutta, S. Conditional gradient methods, 8 2022.

Combettes, C. W. and Pokutta, S. Complexity of Linear Minimization and Projection on Some Sets. *Operations Research Letters*, 49, 7 2021a.

Combettes, C. W. and Pokutta, S. Revisiting the approximate Carathéodory problem via the Frank–Wolfe algorithm. *Mathematical Programming*, 2021b. doi: 10.1007/s10107-021-01735-x.

Combettes, P. L. and Pesquet, J.-C. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

Dattorro, J. *Convex Optimization † Euclidean Distance Geometry*. M$\varepsilon\beta$oo Publishing USA, PO Box 12, Palo Alto, California 94302, 2005. ISBN 978-0-578-16140-2. URL https://ccrma.stanford.edu/%7Edattorro/0976401304.pdf. v2019.10.28.

Frank, M. and Wolfe, P. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956.

Fujishige, S. and Zhan, P. A dual algorithm for finding the minimum-norm point in a polytope. *Journal of the Operations Research Society of Japan*, 33(2):188–195, 1990.

Ginat, O. The method of alternating projections. *arXiv preprint arXiv:1809.05858*, 2018.

Jaggi, M. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning*, pp. 427–435. PMLR, 2013.

Lacoste-Julien, S., Jaggi, M., Schmidt, M., and Pletscher, P. Block-coordinate frank-wolfe optimization for structural svms. In *International Conference on Machine Learning*, pp. 53–61. PMLR, 2013.

Millán, R. D., Ferreira, O., and Ugon, J. Approximate douglas-rachford algorithm for two-sets convex feasibility problems. *arXiv preprint arXiv:2105.13005*, 2021.

Mirrokni, V., Leme, R. P., Vladu, A., and Wong, S. C.-w. Tight bounds for approximate Carathéodory and beyond. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70, pp. 2440–2448. PMLR, 2017.

Moreau, J. J. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–279, 1965. doi: 10.24033/bsmf.1625. URL https://hal.archives-ouvertes.fr/hal-01740635.

von Neumann, J. On rings of operators. reduction theory. *Annals of Mathematics*, pp. 401–485, 1949.

Willner, L. B. On the distance between polytopes. *Quarterly of Applied Mathematics*, 26(2):207–212, July 1968. URL https://www.jstor.org/stable/43635795.

Wolfe, P. Finding the nearest point in a polytope. *Mathematical Programming*, 11(1):128–149, 1976.

Yurtsever, A., Fercoq, O., Locatello, F., and Cevher, V. A conditional gradient framework for composite convex minimization with applications to semidefinite programming. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, volume 80, pp. 5727–5736. PMLR, 2018.