

Ellipsoid Separation in Data-Dependent Time via Infimal Displacement Vectors

Matthew Hough · Stephen A. Vavasis

Abstract The infimal displacement vector of the nonexpansive operator underlying the primal-dual hybrid gradient algorithm (PDHG) is known to encode useful information for infeasible optimization problems. In particular, for the ellipsoid separation problem (ESP), where the goal is to find a separating hyperplane between two collections of ellipsoids, the infimal displacement vector corresponding to an infeasible second-order cone program (SOCP) encodes a separating hyperplane. We show FISTA computes a sequence of approximations to the dual component of the infimal displacement vector and establish that a solution to the ESP can be obtained in $\mathcal{O}(\delta_{\mathcal{A}}^{-2})$ iterations. Here $\delta_{\mathcal{A}}$ denotes the distance between the convex hulls of the two ellipsoid collections.

Keywords primal-dual hybrid gradient · infimal displacement vector · FISTA · ellipsoid separation problem

1 Introduction

Recent work has shown that when an optimization problem is infeasible or unbounded, the iterates of the primal-dual hybrid gradient algorithm (PDHG) can encode useful information through the *infimal displacement vector* (IDV) [1,6]. In particular, in [1] the authors show that the IDV is able to certify infeasibility and unboundedness of linear programs. Jiang et al. [6] further show that the IDV associated with an infeasible second-order cone program (SOCP) can be used to solve the ellipsoid separation problem (ESP), which seeks a hyperplane that strictly separates two finite collections of ellipsoids.

This work was supported in part by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Matthew Hough · Stephen A. Vavasis
Department of Combinatorics & Optimization, University of Waterloo, Waterloo, ON N2L 3G1, Canada
E-mail: mhough@uwaterloo.ca
E-mail: vavasis@uwaterloo.ca

This problem arises in robust classification, where each data point is represented by an ellipsoid describing uncertainty in its location [8].

For the infeasible SOCP considered in [6, Section 6.2], differences between successive PDHG iterates converge to the IDV, and thus solve the ESP asymptotically. In this paper, we reformulate the SOCP solved by PDHG as a least-squares problem with second-order cone constraints so that it can be solved by FISTA, and show that the FISTA residuals converge to the dual component of the IDV which encodes a separating hyperplane for the ESP. We establish that a separating hyperplane can be obtained in finitely many iterations and derive a data-dependent iteration complexity for obtaining such a hyperplane. In particular, we show that a separating hyperplane can be obtained in $\mathcal{O}(\delta_{\mathcal{A}}^{-2})$ iterations, where $\delta_{\mathcal{A}}$ is the distance between the convex hulls of the two ellipsoid collections. Thus, the number of iterations required depends on the geometry of the problem, with better-separated ellipsoids admitting faster convergence to a separator, highlighting a novel role for the IDV.

2 Primal-dual hybrid gradient

The primal-dual hybrid gradient algorithm (PDHG) is an algorithm for solving saddle point problems of the form

$$\inf_{\mathbf{x} \in \mathbb{R}^n} \sup_{\mathbf{y} \in \mathbb{R}^m} \mathbf{y}^T A \mathbf{x} + f(\mathbf{x}) - g(\mathbf{y}) \quad (1)$$

where $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is linear and f and g are proper, lower semicontinuous, convex functions. Let $(\mathbf{x}_0, \mathbf{y}_0) \in \mathbb{R}^n \times \mathbb{R}^m$ and suppose the parameters $\sigma, \tau > 0$ are chosen to satisfy $\sigma\tau\|A\|^2 < 1$. PDHG generates its iterates by repeatedly applying the operator T :

$$(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) = T(\mathbf{x}_k, \mathbf{y}_k) = \begin{pmatrix} \text{prox}_{\sigma f}(\mathbf{x}_k - \sigma A^T \mathbf{y}_k) \\ \text{prox}_{\tau g}(\mathbf{y}_k + \tau A(2 \text{prox}_{\sigma f}(\mathbf{x}_k - \sigma A^T \mathbf{y}_k) - \mathbf{x}_k)) \end{pmatrix}.$$

Under appropriate constraint qualifications, $(\mathbf{x}^*, \mathbf{y}^*)$ is a fixed point of T if and only if $(\mathbf{x}^*, \mathbf{y}^*)$ is optimal for (1) [6, Proposition 3.1]. In the inconsistent case, where $\mathbf{0} \notin \text{range}(\text{Id} - T)$, the iterates $(\mathbf{x}_k, \mathbf{y}_k)$ diverge [5, 2]. However, the successive differences $(\mathbf{x}_k, \mathbf{y}_k) - (\mathbf{x}_{k+1}, \mathbf{y}_{k+1})$ converge to the IDV $\mathbf{v} = (\mathbf{v}_R, \mathbf{v}_D)$, defined as

$$\mathbf{v} := \underset{\mathbf{w} \in \text{cl}(\text{range}(\text{Id} - T))}{\text{argmin}} \|\mathbf{w}\|_M,$$

where $\|\mathbf{w}\|_M = \sqrt{\mathbf{w}^T M \mathbf{w}}$ is the norm induced by the matrix

$$M = \begin{pmatrix} \sigma^{-1} I & -A^T \\ -A & \tau^{-1} I \end{pmatrix}.$$

The assumption $\sigma\tau\|A\|^2 < 1$ ensures that M is positive definite.

3 Ellipsoid separation as an SOCP

Given two finite collections of ellipsoids in $\mathbb{R}^d$,

$$\{C_1, \dots, C_j\}, \quad \{D_1, \dots, D_l\},$$

the ellipsoid separation problem asks whether there exists a hyperplane strictly separating the two collections. We write the ellipsoids as

$$\begin{aligned} C_i &= \{\mathbf{z} : \|A_i^{-1}(\mathbf{z} - \mathbf{c}_i)\| \leq 1\}, \quad i = 1, \dots, j, \\ D_i &= \{\mathbf{z} : \|B_i^{-1}(\mathbf{z} - \mathbf{d}_i)\| \leq 1\}, \quad i = 1, \dots, l, \end{aligned}$$

where $A_i, B_i \in \mathbb{R}^{d \times d}$ are nonsingular and $\mathbf{c}_i, \mathbf{d}_i \in \mathbb{R}^d$ are the ellipsoid centers. We make the assumption throughout that the data is scaled so that

$$\max \left\{ \max_i \{\|\mathbf{c}_i\|, \|A_i\|\}, \max_i \{\|\mathbf{d}_i\|, \|B_i\|\} \right\} = 1. \quad (\text{A})$$

We seek $\mathbf{w} \in \mathbb{R}^d$ and $s, t \in \mathbb{R}$ such that

$$\begin{aligned} \mathbf{w}^T \mathbf{z} &\leq -s, \quad \forall \mathbf{z} \in C_i, \quad i = 1, \dots, j, \\ \mathbf{w}^T \mathbf{z} &\geq t, \quad \forall \mathbf{z} \in D_i, \quad i = 1, \dots, l, \end{aligned}$$

with $s + t > 0$. In this case, any hyperplane $\{\mathbf{z} : \mathbf{w}^T \mathbf{z} = h\}$ with $h \in (-s, t)$ strictly separates the two collections. By [6, Lemma 6.18], these halfspace containment conditions are equivalent to

$$\begin{pmatrix} s + \mathbf{c}_i^T \mathbf{w} \\ A_i^T \mathbf{w} \end{pmatrix} \in -C_2^{d+1}, \quad \begin{pmatrix} t - \mathbf{d}_i^T \mathbf{w} \\ B_i^T \mathbf{w} \end{pmatrix} \in -C_2^{d+1},$$

for all $i = 1, \dots, j$ and $i = 1, \dots, l$, respectively, where C_2^{d+1} denotes the second-order cone. Consequently, the ellipsoid separation problem can be formulated as the following homogeneous second-order cone program:

$$\begin{aligned} &\max_{s, t, \mathbf{w}} s + t \\ &\text{s.t. } \mathcal{A}^T \begin{pmatrix} s \\ t \\ \mathbf{w} \end{pmatrix} \preceq_C \mathbf{0}, \end{aligned} \quad (2)$$

where the matrix $\mathcal{A}^T \in \mathbb{R}^{(j+l)(d+1) \times (d+2)}$ is

$$\mathcal{A}^T := \begin{pmatrix} \bar{A}_1^T \\ \vdots \\ \bar{A}_j^T \\ \bar{B}_1^T \\ \vdots \\ \bar{B}_l^T \end{pmatrix},$$

with

$$\bar{A}_i^T := \begin{pmatrix} 1 & 0 & \mathbf{c}_i^T \\ \mathbf{0} & \mathbf{0} & A_i^T \end{pmatrix}, \quad \bar{B}_i^T := \begin{pmatrix} 0 & 1 & -\mathbf{d}_i^T \\ \mathbf{0} & \mathbf{0} & B_i^T \end{pmatrix},$$

and the notation $\mathbf{a} \preceq_{\mathcal{C}} \mathbf{b}$ means that $\mathbf{a} - \mathbf{b} \in -\mathcal{C}$, where

$$\mathcal{C} := \underbrace{C_2^{d+1} \times \cdots \times C_2^{d+1}}_{(j+l) \text{ times}}.$$

Due to homogeneity, the objective does not yield a finite maximum; rather, feasibility with $s+t > 0$ certifies that the halfspaces are disjoint and a separator exists. Otherwise, the optimum is 0, certifying that no separating hyperplane exists.

4 FISTA for ellipsoid separation

It is not apparent how to apply FISTA [4,7] directly to (2), since there is no closed-form expression to project onto the constraint set $\{(s, t, \mathbf{w}) : \mathcal{A}^T(s, t, \mathbf{w}) \preceq_{\mathcal{C}} \mathbf{0}\}$. It turns out that the dual SOCP admits a constraint set that can be efficiently projected onto. Introduce the dual variables $\lambda_1, \dots, \lambda_j, \mu_1, \dots, \mu_l \in \mathbb{R}$ and $\mathbf{p}_1, \dots, \mathbf{p}_j, \mathbf{q}_1, \dots, \mathbf{q}_l \in \mathbb{R}^d$, and define

$$\mathbf{x} := \begin{pmatrix} \lambda_1 \\ \mathbf{p}_1 \\ \vdots \\ \lambda_j \\ \mathbf{p}_j \\ \mu_1 \\ \mathbf{q}_1 \\ \vdots \\ \mu_l \\ \mathbf{q}_l \end{pmatrix}.$$

The dual SOCP is then

$$\begin{aligned} \min_{\mathbf{x}} \quad & 0 \\ \text{s.t.} \quad & \mathcal{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \succeq_{\mathcal{C}} \mathbf{0}, \end{aligned} \tag{3}$$

where $\mathbf{b} = (1, 1, \mathbf{0})$. Written explicitly, the constraint $\mathcal{A}\mathbf{x} = \mathbf{b}$ becomes

$$\begin{aligned} \lambda_1 + \cdots + \lambda_j &= 1, \\ \mu_1 + \cdots + \mu_l &= 1, \\ \sum_{i=1}^j (\lambda_i \mathbf{c}_i + A_i \mathbf{p}_i) - \sum_{i=1}^l (\mu_i \mathbf{d}_i - B_i \mathbf{q}_i) &= \mathbf{0}, \end{aligned} \tag{4}$$

while the constraint $\mathbf{x} \succeq_{\mathcal{C}} \mathbf{0}$ becomes

$$\begin{aligned} \|\mathbf{p}_i\| &\leq \lambda_i & \forall i = 1, \dots, j \\ \|\mathbf{q}_i\| &\leq \mu_i & \forall i = 1, \dots, l. \end{aligned}$$

The dual feasibility problem (3) has a simple geometric interpretation. The constraint $\mathcal{A}\mathbf{x} = \mathbf{b}$, together with $\mathbf{x} \in \mathcal{C}$, is feasible if and only if

$$\text{conv}(C_1 \cup C_2 \cup \dots \cup C_j) \cap \text{conv}(D_1 \cup D_2 \cup \dots \cup D_l) \neq \emptyset.$$

Consequently, the ellipsoid separation problem has a solution if and only if the dual conic feasibility problem is infeasible. To obtain a separating hyperplane, we apply FISTA to the problem

$$\begin{aligned} \min_{\mathbf{x}} f(\mathbf{x}) &:= \frac{1}{2} \|\mathcal{A}\mathbf{x} - \mathbf{b}\|^2 \\ \text{s.t. } \mathbf{x} &\succeq_{\mathcal{C}} \mathbf{0}. \end{aligned} \quad (5)$$

This formulation is amenable to FISTA because ∇f is Lipschitz continuous with constant $L = \|\mathcal{A}\|^2$ and projection onto the product of second-order cones $\mathcal{C}$ can be computed in linear time (cf. [3, Example 6.37]).

Let $\mathbf{x}_k$ denote the iterates generated by FISTA and define the residuals

$$\mathbf{v}_k := \mathbf{b} - \mathcal{A}\mathbf{x}_k.$$

We next establish how these residuals can be used to construct a separating hyperplane, and subsequently bound the number of iterations required to obtain one.

In the following analysis, we consider the infimal displacement vector of PDHG applied to (3) with $\tau = 1$ and $\sigma > 0$ satisfying $\sigma\|\mathcal{A}\|^2 < 1$.

Theorem 1 *For any minimizer $\bar{\mathbf{x}}$ of (5), the dual component of the PDHG infimal displacement vector, $\mathbf{v}_D$, satisfies*

$$\mathbf{v}_D = \mathbf{b} - \mathcal{A}\bar{\mathbf{x}}.$$

Furthermore, the FISTA residuals converge to the dual part of the IDV: $\mathbf{v}_k \rightarrow \mathbf{v}_D$.

Proof Jiang et al. note that $\ker(\mathcal{A}) \cap \mathcal{C} = \{\mathbf{0}\}$ [6, Section 6.2], so we may take $\tau = 1$ in [6, Lemma 6.8 (vi)] to obtain

$$\mathbf{v}_D = \text{proj}_{\mathbf{b} - \mathcal{A}(\mathcal{C})}(\mathbf{0}) = \text{proj}_{(\mathcal{A}(\mathcal{C}))^\circ}(\mathbf{b}). \quad (6)$$

In particular, a minimizer $\bar{\mathbf{x}}$ of (5) exists, and every such minimizer satisfies $\mathbf{v}_D = \mathbf{b} - \mathcal{A}\bar{\mathbf{x}}$. For any $\mathbf{z} \in \mathcal{A}(\mathcal{C})$, the first-order optimality conditions of the projection give

$$\mathbf{v}_D^T (\mathbf{v}_D - (\mathbf{b} - \mathbf{z})) \leq 0. \quad (7)$$

We may write

$$\|\mathcal{A}\mathbf{x} - \mathbf{b}\|^2 - \|\mathcal{A}\bar{\mathbf{x}} - \mathbf{b}\|^2 = \|\mathcal{A}\mathbf{x} - \mathcal{A}\bar{\mathbf{x}}\|^2 + 2(\mathcal{A}\bar{\mathbf{x}} - \mathbf{b})^T (\mathcal{A}\mathbf{x} - \mathcal{A}\bar{\mathbf{x}}),$$

and substitute $\mathbf{b} - \mathcal{A}\bar{\mathbf{x}} = \mathbf{v}_D$ to obtain

$$\|\mathcal{A}\mathbf{x} - \mathbf{b}\|^2 - \|\mathcal{A}\bar{\mathbf{x}} - \mathbf{b}\|^2 = \|\mathcal{A}\mathbf{x} - \mathcal{A}\bar{\mathbf{x}}\|^2 - 2\mathbf{v}_D^T(\mathbf{v}_D - (\mathbf{b} - \mathcal{A}\mathbf{x})).$$

Let $\mathbf{v} := \mathbf{b} - \mathcal{A}\mathbf{x}$. Applying (7) gives the inequality:

$$\begin{aligned} \|\mathcal{A}\mathbf{x} - \mathbf{b}\|^2 - \|\mathcal{A}\bar{\mathbf{x}} - \mathbf{b}\|^2 &= \|\mathcal{A}\mathbf{x} - \mathcal{A}\bar{\mathbf{x}}\|^2 - 2\mathbf{v}_D^T(\mathbf{v}_D - (\mathbf{b} - \mathcal{A}\mathbf{x})) \\ &\geq \|\mathcal{A}\mathbf{x} - \mathcal{A}\bar{\mathbf{x}}\|^2 \\ &= \|\mathbf{v} - \mathbf{v}_D\|^2. \end{aligned}$$

Taking $\mathbf{x} = \mathbf{x}_k$ where $\mathbf{x}_k$ represents the k th FISTA iterate, we may use the fact that the iterates of FISTA converge in objective value, so $\mathbf{v}_k \rightarrow \mathbf{v}_D$. $\square$

The preceding theorem establishes that the FISTA residuals converge to the dual component of the PDHG infimal displacement vector. The significance of this result is that $\mathbf{v}_D$ encodes a separating hyperplane whenever the two collections of ellipsoids are strictly separable. We show this in the following lemma.

Lemma 1 *Suppose the two collections of ellipsoids are strictly separable, and denote*

$$\mathbf{v}_D = \begin{pmatrix} s^* \\ t^* \\ \mathbf{w}^* \end{pmatrix}.$$

Then $s^* + t^* = \|\mathbf{v}_D\|^2 > 0$, and

$$\begin{aligned} (\mathbf{w}^*)^T \mathbf{z} &\leq -s^*, & \forall \mathbf{z} \in C_1 \cup \dots \cup C_j \\ (\mathbf{w}^*)^T \mathbf{z} &\geq t^*, & \forall \mathbf{z} \in D_1 \cup \dots \cup D_l. \end{aligned}$$

That is, any hyperplane $\{\mathbf{z} : (\mathbf{w}^*)^T \mathbf{z} = h\}$ with $h \in (-s^*, t^*)$ strictly separates the two collections.

Proof Taking $\tau = 1$ in [6, Lemma 6.8 (viii)] gives $\|\mathbf{v}_D\|^2 = \mathbf{b}^T \mathbf{v}_D$. But $\mathbf{b} = (1, 1, \mathbf{0})$, so

$$\|\mathbf{v}_D\|^2 = \mathbf{b}^T \mathbf{v}_D = s^* + t^*.$$

By (6), $\mathbf{v}_D \in (\mathcal{A}(\mathcal{C}))^\circ$, thus $\mathcal{A}^T \mathbf{v}_D \in \mathcal{C}^\circ$. Therefore $\mathbf{v}_D$ is feasible for (2) with $s^* + t^* > 0$. The constraints of (2) encode

$$\begin{pmatrix} s + \mathbf{c}_i^T \mathbf{w} \\ A_i^T \mathbf{w} \end{pmatrix} \in -C_2^{d+1}, \quad \begin{pmatrix} t - \mathbf{d}_i^T \mathbf{w} \\ B_i^T \mathbf{w} \end{pmatrix} \in -C_2^{d+1},$$

for all $i = 1, \dots, j$ and $i = 1, \dots, l$, respectively. The conclusion follows by the equivalence between the halfspace containment conditions and the second-order cone membership shown above [6, Lemma 6.18]. $\square$

We now have a procedure for obtaining a separating hyperplane from the optimal FISTA residual $\mathbf{v}_D = (s^*, t^*, \mathbf{w}^*)$. Simply choose any $h \in (-s^*, t^*)$ and the corresponding separating hyperplane is $\{\mathbf{z} : (\mathbf{w}^*)^T \mathbf{z} = h\}$. However, $\mathbf{v}_D$ is only guaranteed to be obtained from FISTA in the limit. Our goal is to find a separator after finitely many iterations, without computing $\mathbf{v}_D$ exactly. To this end, write

$$\mathbf{v}_k = \begin{pmatrix} s_k \\ t_k \\ \mathbf{w}_k \end{pmatrix},$$

and consider the candidate separating hyperplane

$$H_k := \{\mathbf{z} : \mathbf{w}_k^T \mathbf{z} = m_k\}, \quad m_k := \frac{-s_k + t_k}{2}.$$

The next theorem establishes how close $\mathbf{v}_k$ must be to $\mathbf{v}_D$ in order to guarantee that H_k separates the two collections of ellipsoids.

Theorem 2 *Suppose the two collections of ellipsoids are strictly separable. The residual $\mathbf{v}_k$ defines a strict separator $\{\mathbf{z} : \mathbf{w}_k^T \mathbf{z} = m_k\} = H_k$ provided*

$$\|\mathbf{v}_k - \mathbf{v}_D\| < \frac{\|\mathbf{v}_D\|^2}{\sqrt{18}}$$

Proof For any $\mathbf{z} \in C_1 \cup \dots \cup C_j$, we have that $(\mathbf{w}^*)^T \mathbf{z} \leq -s^*$. This inequality holds true if and only if

$$\mathbf{w}_k^T \mathbf{z} \leq m_k + (\mathbf{w}_k - \mathbf{w}^*)^T \mathbf{z} - s^* - m_k.$$

Let

$$r := \|\mathbf{w}_k - \mathbf{w}^*\| \cdot \|\mathbf{z}\| - \frac{s^* - s_k}{2} + \frac{t^* - t_k}{2} - \frac{s^* + t^*}{2}.$$

By Cauchy-Schwarz, any $\mathbf{z} \in C_1 \cup \dots \cup C_j$ satisfies

$$\begin{aligned} \mathbf{w}_k^T \mathbf{z} &= m_k + (\mathbf{w}_k - \mathbf{w}^*)^T \mathbf{z} + \mathbf{w}^{*T} \mathbf{z} - m_k \\ &\leq m_k + \|\mathbf{w}_k - \mathbf{w}^*\| \cdot \|\mathbf{z}\| - s^* - m_k \\ &= m_k + r. \end{aligned}$$

It follows that if we have $r < 0$, then $\mathbf{w}_k^T \mathbf{z} < m_k$. Under the normalization assumption (A), we have that $\|\mathbf{z}\| \leq 2$, and it is easy to verify that $r < 0$ is implied by the condition

$$s^* + t^* > |s^* - s_k| + |t^* - t_k| + 4\|\mathbf{w}_k - \mathbf{w}^*\|. \quad (8)$$

The implication $\mathbf{z} \in D_1 \cup \dots \cup D_l \implies \mathbf{w}_k^T \mathbf{z} > m_k$ holds under the same assumption. Therefore, the satisfaction of the condition (8) implies that the residual $\mathbf{v}_k$ defines a strict separator.

Now observe the bound

$$|s^* - s_k| + |t^* - t_k| + 4\|\mathbf{w}_k - \mathbf{w}^*\| \leq \sqrt{18}\|\mathbf{v}_k - \mathbf{v}_D\|$$

holds, so a new condition for $\mathbf{v}_k$ defining a strict separator is

$$\|\mathbf{v}_k - \mathbf{v}_D\| < \frac{s^* + t^*}{\sqrt{18}}.$$

From Lemma 1, $s^* + t^* = \|\mathbf{v}_D\|^2$, which completes the proof. $\square$

We have shown that sufficiently accurate FISTA residuals yield separating hyperplanes. In the next section we derive a bound on the number of FISTA iterations required before $\mathbf{v}_k$ is sufficiently accurate per Theorem 2.

5 A data-dependent iteration complexity bound

In order to derive a data-dependent bound on the number of FISTA iterations needed to obtain a separating hyperplane, we relate $\|\mathbf{v}_D\|$ to the distance between the convex hulls of the two ellipsoid collections.

Our strategy is to use a minimizer of (5) to construct a point in the convex hull of the C_i ellipsoids and another in the convex hull of the D_i ellipsoids, whose distance can be bounded in terms of $\|\mathbf{v}_D\|$. Combining this geometric bound with the convergence rate of FISTA provides an iteration complexity bound that depends on the separation between the two collections.

Let $\mathbf{x}^* \in \mathcal{C}$ such that $\mathbf{b} - \mathcal{A}\mathbf{x}^* = \mathbf{v}_D$, which exists by Theorem 1. Let

$$\mathbf{x}^* := \begin{pmatrix} \lambda_1^* \\ \mathbf{p}_1^* \\ \vdots \\ \lambda_j^* \\ \mathbf{p}_j^* \\ \mu_1^* \\ \mathbf{q}_1^* \\ \vdots \\ \mu_l^* \\ \mathbf{q}_l^* \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{x}} := \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\mathbf{p}}_1 \\ \vdots \\ \hat{\lambda}_j \\ \hat{\mathbf{p}}_j \\ \hat{\mu}_1 \\ \hat{\mathbf{q}}_1 \\ \vdots \\ \hat{\mu}_l \\ \hat{\mathbf{q}}_l \end{pmatrix},$$

where $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_j^*)$, $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \dots, \hat{\lambda}_j)$, $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_l^*)$, $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \dots, \hat{\mu}_l)$, and

$$\begin{aligned} \hat{\boldsymbol{\lambda}} &:= \text{proj}_{\Delta^j}(\boldsymbol{\lambda}^*) \\ \hat{\boldsymbol{\mu}} &:= \text{proj}_{\Delta^l}(\boldsymbol{\mu}^*) \\ \hat{\mathbf{p}}_i &:= \begin{cases} \frac{\hat{\lambda}_i}{\lambda_i^*} \mathbf{p}_i^*, & \hat{\lambda}_i < \lambda_i^*; \\ \mathbf{p}_i^*, & \text{otherwise} \end{cases} \\ \hat{\mathbf{q}}_i &:= \begin{cases} \frac{\hat{\mu}_i}{\mu_i^*} \mathbf{q}_i^*, & \hat{\mu}_i < \mu_i^*; \\ \mathbf{q}_i^*, & \text{otherwise,} \end{cases} \end{aligned}$$

where $\Delta^t = \{\boldsymbol{\lambda} \in \mathbb{R}^t : \lambda_i \geq 0, \forall i \in [t], \sum_{i=1}^t \lambda_i = 1\}$. By construction, $\hat{\mathbf{x}} \in \mathcal{C}$.

Lemma 2 $\|\boldsymbol{\lambda}^* - \hat{\boldsymbol{\lambda}}\|_1 = |1 - \mathbf{e}^T \boldsymbol{\lambda}^*| = |s^*|$ and $\|\boldsymbol{\mu}^* - \hat{\boldsymbol{\mu}}\|_1 = |1 - \mathbf{e}^T \boldsymbol{\mu}^*| = |t^*|$.

Proof We will prove $\|\boldsymbol{\lambda}^* - \hat{\boldsymbol{\lambda}}\|_1 = |1 - \mathbf{e}^T \boldsymbol{\lambda}^*| = |s^*|$ only, since the proof of the second expression is analogous.

By definition of $\mathcal{A}$ and the identity $(s^*, t^*, \mathbf{w}^*) = \mathbf{v}_D = \mathbf{b} - \mathcal{A}\mathbf{x}^*$, we have $s^* = 1 - \mathbf{e}^T \boldsymbol{\lambda}^*$.

Recall $\hat{\boldsymbol{\lambda}} := \text{proj}_{\Delta^j}(\boldsymbol{\lambda}^*)$ is given by

$$\hat{\lambda}_i = \max\{\lambda_i^* - \theta, 0\},$$

where $\theta \in \mathbb{R}$ is chosen to make $\sum_{i=1}^j \hat{\lambda}_i = 1$. Since $\lambda_i^* \geq \|\mathbf{p}_i^*\|$, we have $\lambda_i^* \geq 0$.

If $\sum_{i=1}^j \lambda_i^* \geq 1$, then $\theta \geq 0$, so for all i

$$\hat{\lambda}_i \leq \lambda_i^*$$

On the other hand, if $\sum_{i=1}^j \lambda_i^* < 1$, then $\theta < 0$, so for all i

$$\hat{\lambda}_i \geq \lambda_i^*.$$

It follows that all $\lambda_i^* - \hat{\lambda}_i$ must have the same sign for all i . So we may write

$$\|\boldsymbol{\lambda}^* - \hat{\boldsymbol{\lambda}}\|_1 = \sum_{i=1}^j |\lambda_i^* - \hat{\lambda}_i| = \left| \sum_{i=1}^j (\lambda_i^* - \hat{\lambda}_i) \right| = |1 - \mathbf{e}^T \boldsymbol{\lambda}^*|.$$

□

Lemma 3 $\|\hat{\mathbf{p}}_i - \mathbf{p}_i^*\| \leq |\hat{\lambda}_i - \lambda_i^*|$ and $\|\hat{\mathbf{q}}_i - \mathbf{q}_i^*\| \leq |\hat{\mu}_i - \mu_i^*|$.

Proof If $\hat{\lambda}_i \geq \lambda_i^*$, then the inequality is clear. If $\hat{\lambda}_i < \lambda_i^*$, then $\hat{\mathbf{p}}_i = \hat{\lambda}_i \mathbf{p}_i^* / \lambda_i^*$, so $\|\hat{\mathbf{p}}_i - \mathbf{p}_i^*\| = \|(\hat{\lambda}_i - \lambda_i^*) \mathbf{p}_i^* / \lambda_i^*\| \leq |\hat{\lambda}_i - \lambda_i^*|$, since $\|\mathbf{p}_i^*\| \leq \lambda_i^*$. The proof of $\|\hat{\mathbf{q}}_i - \mathbf{q}_i^*\| \leq |\hat{\mu}_i - \mu_i^*|$ is analogous. □

Theorem 3 Define $\delta_{\mathcal{A}}$ to be the distance between the two convex hulls

$$\delta_{\mathcal{A}} = \text{dist} \left(\text{conv} \left(\bigcup_{i=1}^j C_i \right), \text{conv} \left(\bigcup_{i=1}^l D_i \right) \right).$$

Then $\delta_{\mathcal{A}} \leq 3\|\mathbf{v}_D\|$.

Proof Since $\mathbf{b} - \mathcal{A}\mathbf{x}^* = \mathbf{v}_D = (s^*, t^*, \mathbf{w}^*)$, we have that $\mathbf{b} - \mathcal{A}\hat{\mathbf{x}} = (0, 0, \mathbf{r})$, where

$$\mathbf{r} := \mathbf{w}^* + \sum_{i=1}^j ((\lambda_i^* - \hat{\lambda}_i) \mathbf{c}_i + A_i(\mathbf{p}_i^* - \hat{\mathbf{p}}_i)) + \sum_{i=1}^l (-(\mu_i^* - \hat{\mu}_i) \mathbf{d}_i + B_i(\mathbf{q}_i^* - \hat{\mathbf{q}}_i)).$$

From Lemma 2,

$$\sum_{i=1}^j |\lambda_i^* - \hat{\lambda}_i| \leq |s^*|, \quad \sum_{i=1}^l |\mu_i^* - \hat{\mu}_i| \leq |t^*|,$$

so from Lemma 3, we have that

$$\sum_{i=1}^j \|\mathbf{p}_i^* - \hat{\mathbf{p}}_i\| \leq |s^*|, \quad \sum_{i=1}^l \|\mathbf{q}_i^* - \hat{\mathbf{q}}_i\| \leq |t^*|.$$

Also by the normalization assumption (A), $\|\mathbf{c}_i\| \leq 1$, $\|A_i\| \leq 1$, $\|\mathbf{d}_i\| \leq 1$, $\|B_i\| \leq 1$. These inequalities allow us to conclude that

$$\|\mathbf{r}\| \leq \|\mathbf{w}^*\| + 2|s^*| + 2|t^*| \leq 3\|\mathbf{v}_D\|.$$

Since $\hat{\mathbf{x}} \in \mathcal{C}$, $\|\hat{\mathbf{p}}_i\| \leq \hat{\lambda}_i$. Define for all $i \in [j]$ with $\hat{\lambda}_i > 0$

$$\mathbf{z}_i := \mathbf{c}_i + \frac{1}{\hat{\lambda}_i} A_i \hat{\mathbf{p}}_i.$$

Then

$$\|A_i^{-1}(\mathbf{z}_i - \mathbf{c}_i)\| \leq 1,$$

so $\mathbf{z}_i \in C_i$. Hence, we can define

$$\mathbf{u} := \sum_{i=1}^j (\hat{\lambda}_i \mathbf{c}_i + A_i \hat{\mathbf{p}}_i),$$

such that $\mathbf{u} = \sum_{i:\hat{\lambda}_i > 0} \hat{\lambda}_i \mathbf{z}_i$. Since $\sum_{i=1}^j \hat{\lambda}_i = 1$ and $\hat{\lambda}_i \geq 0$, it follows that

$$\mathbf{u} \in \text{conv} \left(\bigcup_{i=1}^j C_i \right).$$

A similar construction shows that

$$\mathbf{v} := \sum_{i=1}^l (\hat{\mu}_i \mathbf{d}_i - B_i \hat{\mathbf{q}}_i) \in \text{conv} \left(\bigcup_{i=1}^l D_i \right).$$

By the definition of $\mathbf{r}$, we have $\mathbf{r} = \mathbf{v} - \mathbf{u}$. Consequently,

$$\delta_{\mathcal{A}} \leq \|\mathbf{u} - \mathbf{v}\| = \|\mathbf{r}\| \leq 3\|\mathbf{v}_D\|.$$

□

Corollary 1 *Suppose the two collections of ellipsoids are strictly separable. The residual $\mathbf{v}_k = \mathbf{b} - \mathcal{A}\mathbf{x}_k = (s_k, t_k, \mathbf{w}_k)$ gives rise to the strict separator H_k , provided*

$$\|\mathbf{v}_k - \mathbf{v}_D\| < \frac{\delta_{\mathcal{A}}^2}{27\sqrt{2}}.$$

Proof We have shown in Theorem 2 that $\mathbf{v}_k$ is a strict separator provided $\|\mathbf{v}_k - \mathbf{v}_D\| < \|\mathbf{v}_D\|^2/\sqrt{18}$. We have also shown in Theorem 3 that we can lower bound $\|\mathbf{v}_D\|$ by the distance between the two convex hulls $\delta_{\mathcal{A}}$: $\|\mathbf{v}_D\| \geq \delta_{\mathcal{A}}/3$. Hence, to ensure $\|\mathbf{v}_k - \mathbf{v}_D\| < \|\mathbf{v}_D\|^2/\sqrt{18}$, we require that

$$\|\mathbf{v}_k - \mathbf{v}_D\| < \frac{\delta_{\mathcal{A}}^2}{27\sqrt{2}}.$$

□

Theorem 4 *Suppose the two collections of ellipsoids are strictly separable. FISTA applied to the reformulation of the ESP (5) with $\mathbf{x}_0 = \mathbf{0}$ finds a separating hyperplane in at most*

$$\left\lceil \frac{108\sqrt{6(j+l)}}{\delta_{\mathcal{A}}^2} \right\rceil$$

iterations.

Proof From Corollary 1, we know that $\mathbf{v}_k = (s_k, t_k, \mathbf{w}_k)$ encodes a separating hyperplane when $\|\mathbf{v}_k - \mathbf{v}_D\| < \delta_{\mathcal{A}}^2/(27\sqrt{2})$. We want to know when $\mathbf{v}_k$ lies in an open ball of radius $\delta_{\mathcal{A}}^2/(27\sqrt{2})$ around $\mathbf{v}_D$. Recall (cf. [3, Theorem 10.34]) that for iterates $\mathbf{x}_k$ generated by FISTA and any $\bar{\mathbf{x}} \in \{\mathbf{x} \in \mathcal{C} : \mathbf{b} - \mathcal{A}\mathbf{x} = \mathbf{v}_D\}$,

$$f(\mathbf{x}_k) - f(\bar{\mathbf{x}}) \leq \frac{2L\|\mathbf{x}_0 - \bar{\mathbf{x}}\|^2}{(k+1)^2},$$

where $\mathbf{x}_0$ is the starting point of the algorithm and $L = \|\mathcal{A}\|^2$ is the Lipschitz constant for ∇f . We thus obtain

$$f(\mathbf{x}_k) - f(\bar{\mathbf{x}}) \leq \frac{2\|\mathcal{A}\|^2\|\mathbf{x}_0 - \bar{\mathbf{x}}\|^2}{(k+1)^2}.$$

From the proof of Theorem 1, we have $\|\mathbf{v}_k - \mathbf{v}_D\| \leq \sqrt{2}\sqrt{f(\mathbf{x}_k) - f(\bar{\mathbf{x}})}$, so applying this to the above we obtain the bound

$$\|\mathbf{v}_k - \mathbf{v}_D\| \leq \frac{2\|\mathcal{A}\|\|\mathbf{x}_0 - \bar{\mathbf{x}}\|}{k+1}.$$

For any $\mathbf{x} \in \mathcal{C}$,

$$\begin{aligned} \|\mathbf{x}\|^2 &= \sum_{i=1}^j (\lambda_i^2 + \|\mathbf{p}_i\|^2) + \sum_{i=1}^l (\mu_i^2 + \|\mathbf{q}_i\|^2) \\ &\leq 2 \sum_{i=1}^j \lambda_i^2 + 2 \sum_{i=1}^l \mu_i^2 \\ &\leq 2 \left(\left(\sum_{i=1}^j \lambda_i \right)^2 + \left(\sum_{i=1}^l \mu_i \right)^2 \right) \\ &\leq 2\|\mathcal{A}\mathbf{x}\|^2. \end{aligned}$$

From (6), $\mathcal{A}\bar{\mathbf{x}} = \mathbf{b} - \mathbf{v}_D = \mathbf{b} - \text{proj}_{\mathbf{b}-\mathcal{A}(C)}(\mathbf{0}) = \text{proj}_{\mathcal{A}(C)}(\mathbf{b})$, and $\mathbf{0} \in \mathcal{A}(C)$, so by the nonexpansiveness of the projection

$$\|\mathbf{x}_0 - \bar{\mathbf{x}}\| \leq \sqrt{2}\|\mathcal{A}\bar{\mathbf{x}}\| \leq \sqrt{2}\|\text{proj}_{\mathcal{A}(C)}(\mathbf{b}) - \text{proj}_{\mathcal{A}(C)}(\mathbf{0})\| \leq \sqrt{2}\|\mathbf{b}\| = 2.$$

It follows that we can guarantee that $\mathbf{v}_k$ lies in an open ball of radius $\delta_{\mathcal{A}}^2/27\sqrt{2}$ around $\mathbf{v}_D$ when $4\|\mathcal{A}\|/(k+1) < \delta_{\mathcal{A}}^2/(27\sqrt{2})$. Under Assumption A, a sufficient condition is

$$k \geq \frac{108\sqrt{6(j+l)}}{\delta_{\mathcal{A}}^2} \geq \frac{108\sqrt{2}\|\mathcal{A}\|}{\delta_{\mathcal{A}}^2}.$$

□

Verifying separation. Suppose the collections of ellipsoids are separable. For the FISTA residual $\mathbf{v}_k = (s_k, t_k, \mathbf{w}_k)$, the hyperplane H_k strictly separates the collections of ellipsoids if and only if

$$\begin{aligned} \mathbf{c}_i^T \mathbf{w}_k + \|A_i^T \mathbf{w}_k\| &< m_k, \quad i = 1, \dots, j, \\ \mathbf{d}_i^T \mathbf{w}_k - \|B_i^T \mathbf{w}_k\| &> m_k, \quad i = 1, \dots, l, \end{aligned}$$

where the expressions on the left hand side are the maximum and minimum, respectively, of $\mathbf{w}_k^T \mathbf{z}$ over the corresponding ellipsoids (see, for example, [6, Lemma 6.18]). It follows that separation can be verified efficiently at each FISTA iteration without the knowledge of $\mathbf{v}_D$.

References

1. Applegate, D., Díaz, M., Lu, H., Lubin, M.: Infeasibility detection with primal-dual hybrid gradient for large-scale linear programming. *SIAM Journal on Optimization* **34**(1), 459–484 (2024). DOI 10.1137/22M1510467. URL <https://doi.org/10.1137/22M1510467>
2. Baillon, J.B., Bruck, R.E., Reich, S.: On the Asymptotic Behavior of Nonexpansive Mappings and Semigroups in Banach Spaces. *Houston Journal of Mathematics* **4**(1), 1–9 (1978)
3. Beck, A.: First-Order Methods in Optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA (2017). DOI 10.1137/1.9781611974997
4. Beck, A., Teboulle, M.: A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences* **2**(1), 183–202 (2009). DOI 10.1137/080716542. URL <https://doi.org/10.1137/080716542>
5. Bruck, R.E., Reich, S.: Nonexpansive Projections and Resolvents of Accretive Operators in Banach Spaces. *Houston Journal of Mathematics* **3**(4), 459–470 (1977)
6. Jiang, T., Moursi, W.M., Vavasis, S.A.: Range of the displacement operator of PDHG with applications to quadratic and conic programming (2023)
7. Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course, *Applied Optimization*, vol. 87. Kluwer Academic Publishers, Boston (2004). DOI 10.1007/978-1-4419-8853-9
8. Shivaswamy, P.K., Bhattacharyya, C., Smola, A.J.: Second order cone programming approaches for handling missing and uncertain data. *Journal of Machine Learning Research* **7**(47), 1283–1314 (2006). URL <http://jmlr.org/papers/v7/shivaswamy06a.html>