Automatica 142 (2022) 110359

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Proportional–integral projected gradient method for conic optimization ☆

Yue Yu [a,*,1], Purnanand Elango [b,2], Ufuk Topcu [a,1], Behçet Açıkmeşe [b,1]

[a] Oden Institute for Computational Engineering and Sciences, The University of Texas, Austin, TX, 78712, United States of America
[b] Department of Aeronautics and Astronautics, University of Washington, Seattle, WA, 98195, United States of America

ARTICLE INFO

Article history:
Received 23 August 2021
Received in revised form 13 December 2021
Accepted 9 March 2022
Available online 11 May 2022

Keywords:
Convex optimization
First-order methods
Optimal control

ABSTRACT

Conic optimization is the minimization of a differentiable convex objective function subject to conic constraints. We propose a novel primal–dual first-order method for conic optimization, named proportional–integral projected gradient method (PIPG). PIPG ensures that both the primal–dual gap and the constraint violation converge to zero at the rate of $O(1/k)$, where k is the number of iterations. If the objective function is strongly convex, PIPG improves the convergence rate of the primal–dual gap to $O(1/k^2)$. Further, unlike any existing first-order methods, PIPG also improves the convergence rate of the constraint violation to $O(1/k^3)$. We demonstrate the application of PIPG in constrained optimal control problems.

1. Introduction

Conic optimization is the minimization of a differentiable convex objective function subject to conic constraints:

$$\begin{aligned} &\underset{z}{\text{minimize}} && f(z) \\ &\text{subject to} && Hz - g \in \mathbb{K},\ \ z \in \mathbb{D}, \end{aligned} \tag{1}$$

where $z \in \mathbb{R}^n$ is the solution variable, $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable and convex objective function, $\mathbb{K} \subset \mathbb{R}^m$ is a closed convex cone and $\mathbb{D} \subset \mathbb{R}^n$ is a closed convex set, $H \in \mathbb{R}^{m\times n}$ and $g \in \mathbb{R}^m$ are constraint parameters. With proper choice of cone $\mathbb{K}$, conic optimization (1) generalizes linear programming, quadratic programming, second-order cone programming, and semi-definite programming (Ben-Tal & Nemirovski, 2001; Boyd & Vandenberghe, 2004). Conic optimization has found applications in various areas, including signal processing (Luo & Yu, 2006), machine learning (Andersen et al., 2011), robotics (Majumdar et al., 2020), and aerospace engineering (Eren et al., 2017; Liu et al., 2017; Malyuta et al., 2021).

The goal of numerically solving optimization (1) is to compute a solution $z^\star \in \mathbb{R}^n$ that achieves, up to a given numerical tolerance, zero violation of the constraints in (1) and zero *primal–dual gap*; the latter implies that $z^\star$ is an optimal solution of optimization (1) (Boyd et al., 2011; Chambolle & Pock, 2011, 2016b; He & Yuan, 2012). To this end, numerical methods iteratively compute a solution whose constraint violation and primal–dual gap are nonzero at first but converge to zero as the number of iteration k increases.

Due to their low computational cost, first-order methods have attracted increasing attention in conic optimization (Boyd et al., 2011; Chambolle & Pock, 2016a; Lan et al., 2011; O'Donoghue et al., 2016; Yu, Elango et al., 2020). Unlike second-order methods, such as interior point methods (Andersen et al., 2003; Nesterov & Nemirovskii, 1994), first-order methods do not rely on computing matrix inverses. They consequently are suitable for implementation with limited computational resources.

The existing first-order methods solve optimization (1) by solving two different equivalent problems. The first equivalent problem is the following optimization with equality constraints (Boyd et al., 2011; O'Donoghue et al., 2016; Stellato et al., 2020; Yu, Elango et al., 2020):

$$\begin{aligned} &\underset{z,y}{\text{minimize}} && f(z) \\ &\text{subject to} && Hz - y = g,\ \ y \in \mathbb{K},\ \ z \in \mathbb{D}. \end{aligned} \tag{2}$$

In particular, the alternating direction method of multipliers (ADMM) solves optimization (1) by computing one projection onto cone $\mathbb{K}$ and multiple projections onto set $\mathbb{D}$ in each iteration.

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Delin Chu under the direction of Editor Ian R. Petersen.

* Corresponding author.
E-mail addresses: yueyu@utexas.edu (Y. Yu), pelango@uw.edu (P. Elango), utopcu@utexas.edu (U. Topcu), behcet@uw.edu (B. Açıkmeşe).

[1] The work of Y. Yu and U. Topcu is supported by the National Science Foundation under Grant CNS-1836900, and the Air Force Research Laboratory under Grant FA9550-19-1-0169.

[2] The work of P. Elango and B. Açıkmeşe is supported by the National Science Foundation under Grant ECCS-1931744, the Office of Naval Research under Grant N00014-20-1-2288, and the Air Force Office of Scientific Research under Grant FA9550-20-1-0053.

https://doi.org/10.1016/j.automatica.2022.110359
0005-1098/

Table 1
Comparison of different first-order methods for conic optimization (1).

Algorithms	f is smooth & convex				f is smooth & strongly convex			
	# of proj. per iter.		Convergence rates		# of proj. per iter.		Convergence rates	
	$\mathbb{D}$	$\mathbb{K}$ or $\mathbb{K}^\circ$	Primal–dual gap	Constraint violation	$\mathbb{D}$	$\mathbb{K}$ or $\mathbb{K}^\circ$	Primal–dual gap	Constraint violation
ADMM	$O(1/\sqrt{\epsilon})$	1	$O(1/k)$	$O(1/k)$	$O(\ln(1/\epsilon))$	1	$O(1/k)$	$O(1/k)$
PIPGeq	1	1	$O(1/k)$	$O(1/k)$	1	1	$O(1/k)$	$O(1/k)$
PDHG	1	1	$O(1/k)$	N/A	1	1	$O(1/k^2)$	N/A
This work	1	1	$O(1/k)$	$O(1/k)$	1	1	$O(1/k^2)$	$O(1/k^3)$

$\epsilon > 0$ is a tunable accuracy tolerance in ADMM, $\mathbb{K}^\circ$ denotes the polar cone of $\mathbb{K}$.

ADMM ensures that both the constraint violation and the primal–dual gap converge to zero at rate of $O(1/k)$, where k is the number of iterations (Boyd et al., 2011; Eckstein, 1989; Fortin & Glowinski, 2000; Gabay & Mercier, 1976; He & Yuan, 2012; Wang & Banerjee, 2014). The proportional–integral projected gradient method for equality constrained optimization (PIPGeq) achieves the same convergence rates as ADMM, while computing one projection onto cone $\mathbb{K}$ and *only one* projection onto set $\mathbb{D}$ in each iteration (Yu, Elango et al., 2020). Although variants of ADMM (Goldstein et al., 2014; Kadkhodaie et al., 2015; Ouyang et al., 2015; Xu, 2017) and PIPGeq (Xu, 2017; Yu, Elango et al., 2020) can achieve accelerated convergence rates for strongly convex objective functions, such acceleration is not possible for optimization (2) because the objective function in (2) is independent of variable y and, as a result, not strongly convex.

Another problem equivalent to optimization (1) is the following saddle-point problem, where $\mathbb{K}^\circ$ is the polar cone of $\mathbb{K}$ (Chambolle & Pock, 2011, 2016b):

$$\underset{z\in\mathbb{D}}{\text{minimize}} \quad \underset{w\in\mathbb{K}^\circ}{\text{maximize}} \quad f(z) + \langle Hz - g, w\rangle. \tag{3}$$

In particular, the primal–dual hybrid-gradient method (PDHG) solves saddle-point problem (3) by computing one projection onto cone $\mathbb{K}^\circ$ and one projection onto set $\mathbb{D}$ in each iteration. PDHG ensures that the primal–dual gap converges to zero at the rate of $O(1/k)$ for convex f, and at an accelerated rate of $O(1/k^2)$ for strongly convex f (Chambolle & Pock, 2016a, 2016b). However, since the constraint $Hz-g \in \mathbb{K}$ is not explicitly considered in (3), the existing convergence results on PDHG do not provide any convergence rates of the violation of this constraint (Chambolle & Pock, 2016a, 2016b).

We compare the per-iteration computation and the convergence rates of ADMM, PIPGeq and PDHG in Table 1. None of these methods *simultaneously* have accelerated convergence rates (*i.e.*, better than $O(1/k)$) for strongly convex f and guaranteed convergence rates on the constraint violation. To our best knowledge, whether there exists a first-order method that achieves both convergence results remains an open question.

We answer this question affirmatively by proposing a novel primal–dual first-order method for conic optimization, named proportional–integral projected gradient method (PIPG). By combining the idea of proportional–integral feedback control and projected gradient method, PIPG ensures the following convergence results.

(1) For convex f, both the primal–dual gap and the constraint violation converge to zero at the rate of $O(1/k)$.
(2) For strongly convex f, the convergence rate can be improved to $O(1/k^2)$ for the primal–dual gap and $O(1/k^3)$ for the constraint violation.

PIPG generalizes both PDHG with constant step sizes (Chambolle & Pock, 2016b, Alg. 1) and PIPGeq (Yu, Elango et al., 2020). Compared with the existing methods, PIPG has the following advantages; see Table 1 for an overview. In terms of per-iteration cost, it computes one projection onto cone $\mathbb{K}^\circ$ and one projection onto set $\mathbb{D}$, which is the same as PIPGeq and PDHG, and fewer times of projections than ADMM. In terms of its convergence rates, to our best knowledge, the $O(1/k^3)$ convergence rate of constraint violation has never been achieved before for general conic optimization. We numerically demonstrate these advantages of PIPG on several constrained optimal control problems.

The rest of the paper is organized as follows. After some preliminary results on convex analysis, Section 2 reviews existing first-order conic optimization methods. Section 3 introduces PIPG along with its convergence results. Section 4 demonstrates PIPG via numerical examples on constrained optimal control. Finally, Section 5 concludes and comments on future work.

2. Preliminaries and related work

This section reviews some results in convex analysis and first-order conic optimization methods.

2.1. Notation and preliminaries

We let $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{R}_+$ denote the set of positive integer, real, and non-negative real numbers, respectively. For two vectors $z, z' \in \mathbb{R}^n$, $\langle z, z'\rangle$ denotes their inner product, $\|z\| := \sqrt{\langle z, z\rangle}$ denotes the ℓ_2 norm of z, and $\|\cdot\|_\infty$ denotes the ℓ_∞ norm of z, *i.e.*, the maximum absolute value of the entries of z. We let $\mathbf{1}_n$ and $\mathbf{0}_n$ denote the n-dimensional vectors of all 1's and all 0's, respectively. We also let I_n and $0_{m\times n}$ denote the $n \times n$ identity matrix and the $m \times n$ zero matrix, respectively. When their dimensions are clear from the context, we omit the subscripts and simply write vector $\mathbf{1}, \mathbf{0}$ and matrix I, 0. For a matrix $H \in \mathbb{R}^{m\times n}$, $H^\top$ denotes its transpose, $|||H|||$ denotes its largest singular value. For a square matrix $M \in \mathbb{R}^{n\times n}$, $\exp(M)$ denotes the matrix exponential of M, and $\|z\|_M := \sqrt{\langle z, Mz\rangle}$ for all $z \in \mathbb{R}^n$. Given two sets $\mathbb{S}_1$ and $\mathbb{S}_2$, $\mathbb{S}_1 \times \mathbb{S}_2$ denotes their Cartesian product.

Let $z, z' \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. The *Bregman divergence* from z to z' associated with function f is given by

$$B_f(z, z') := f(z) - f(z') - \langle \nabla f(z'), z - z'\rangle. \tag{4}$$

We say function f is μ-strongly convex for some $\mu \in \mathbb{R}_+$ if $B_f(z, z') \geqslant \frac{\mu}{2}\left\|z - z'\right\|^2$ for all $z, z' \in \mathbb{R}^n$. If $\mu = 0$, we say function f is convex. We say function f is λ-smooth for some $\lambda \in \mathbb{R}_+$ if $B_f(z, z') \leqslant \frac{\lambda}{2}\left\|z - z'\right\|^2$ for all $z, z' \in \mathbb{R}^n$.

Let $\mathbb{D} \subset \mathbb{R}^n$ be a closed convex set, *i.e.*, $\mathbb{D}$ contains all of its boundary points and $\gamma z + (1 - \gamma)z' \in \mathbb{D}$ for any $\gamma \in [0, 1]$ and $z, z' \in \mathbb{D}$. The *projection* of $z \in \mathbb{R}^n$ onto set $\mathbb{D}$ is given by

$$\pi_{\mathbb{D}}[z] := \underset{y\in\mathbb{D}}{\text{argmin}} \, \|z - y\| . \tag{5}$$

Let $\mathbb{K} \subset \mathbb{R}^m$ be a closed convex cone, *i.e.*, $\mathbb{K}$ is a closed convex set and $\gamma w \in \mathbb{K}$ for any $w \in \mathbb{K}$ and $\gamma \in \mathbb{R}_+$. The *polar cone* of $\mathbb{K}$ is also a closed convex cone given by

$$\mathbb{K}^\circ := \{w \in \mathbb{R}^m | \langle w, y\rangle \leqslant 0, \forall y \in \mathbb{K}\}. \tag{6}$$

2.2. Related work

We briefly review three existing first-order primal–dual conic optimization methods: ADMM, PIPGeq, and PDHG. In the following, let α, β, γ denote positive scalar step sizes, and $\{\alpha^j\}_{j\in\mathbb{N}}$, $\{\beta^j\}_{j\in\mathbb{N}}$, $\{\gamma^j\}_{j\in\mathbb{N}}$ denote sequences of positive scalar step sizes. For simplicity, we assume all methods are terminated after a fixed number of iterations, denoted by $k \in \mathbb{N}$.

2.2.1. Alternating direction method of multipliers

As a special case of Douglas–Rachford splitting method (Eckstein, 1989; Fortin & Glowinski, 2000), alternating direction method of multipliers (ADMM) solves optimization (1) by solving the equivalent optimization (2) using Algorithm 1 (Boyd et al., 2011; Gabay & Mercier, 1976; He & Yuan, 2012).

Algorithm 1 ADMM

Input: $k, \alpha, z^1 \in \mathbb{D}, y^1 \in \mathbb{K}, w^1 \in \mathbb{R}^m$

1: **for** $j = 1, 2, \ldots, k-1$ **do**

2: $\quad z^{j+1} = \underset{z\in\mathbb{D}}{\operatorname{argmin}} f(z) + \frac{\alpha}{2}\left\|Hz - y^j - g + w^j\right\|^2$

3: $\quad y^{j+1} = \pi_{\mathbb{K}}[Hz^{j+1} - g + w^j]$

4: $\quad w^{j+1} = w^j + Hz^{j+1} - y^{j+1} - g$

5: **end for**

Generally, the minimization in the line 2 of Algorithm 1 can only be solved approximately up to a numerical tolerance $\epsilon > 0$ using iterative methods. Such methods need to compute at least $O(1/\sqrt{\epsilon})$ projections onto set $\mathbb{D}$ if f is merely convex, and $O(\ln \frac{1}{\epsilon})$ projections if function f is strongly convex; see Nesterov (2018, Chp. 2) for a detailed discussion.

There has been many variants of ADMM developed in the literature. However, none of them lead to any significant benefits for optimization in (2). For example, Ouyang et al. (2015) and Xu (2017, Alg. 1) simplified the minimization in the line 2 of Algorithm 1 by approximating function f using its linearization. However, solving the resulting approximate minimization still requires multiple projections onto set $\mathbb{D}$. On the other hand, although the convergence of ADMM can be accelerated when the objective function is strongly convex (Goldstein et al., 2014; Kadkhodaie et al., 2015; Ouyang et al., 2015; Xu, 2017), such acceleration does not apply to the optimization (2). The reason is because the objective function in (2) is not strongly convex with respect to (in fact, does not depend on) variable y.

2.2.2. Proportional–integral projected gradient method for equality constrained optimization

Motivated by applications in model predictive control, the proportional–integral projected gradient method for equality constrained optimization (PIPGeq) solves optimization (1) by solving the equivalent optimization (2) using Algorithm 2.

Algorithm 2 PIPGeq

Input: $k, \alpha, \beta, z^1 \in \mathbb{D}, y^1 \in \mathbb{K}, w^1 \in \mathbb{R}^m$.

1: **for** $j = 1, 2, \ldots, k-1$ **do**

2: $\quad v^{j+1} = w^j + \beta(Hz^j - y^j - g)$

3: $\quad z^{j+1} = \pi_{\mathbb{D}}[z^j - \alpha(\nabla f(z^j) + H^\top v^{j+1})]$

4: $\quad y^{j+1} = \pi_{\mathbb{K}}[y^j + \alpha v^{j+1}]$

5: $\quad w^{j+1} = w^j + \beta(Hz^{j+1} - y^{j+1} - g)$

6: **end for**

Unlike line 2 in Algorithm 1, line 3 in Algorithm 2 computes only one projection onto set $\mathbb{D}$ instead of multiple times. As a result, PIPGeq can achieve the same convergence rates as those of ADMM while lowering the per-iteration computation cost (Xu, 2017; Yu, Elango et al., 2020).

2.2.3. Primal–dual hybrid gradient method

Motivated by applications in computational imaging, the primal–dual hybrid gradient method (PDHG) was first introduced in Chambolle and Pock (2011) and later extended to a three-operator splitting method (Chambolle & Pock, 2016b). To solve optimization (1), PDHG solves the equivalent convex–concave saddle point problem (3) instead. If function f is merely convex, PDHG uses Algorithm 3. If function f is μ-strongly convex for some $\mu > 0$, then PDHG uses Algorithm 4 instead.

Algorithm 3 PDHG with constant step sizes

Input: $k, \alpha, \beta, z^1 \in \mathbb{D}, w^1 \in \mathbb{K}^\circ$.

for $j = 1, 2, \ldots, k-1$ **do**

$\quad z^{j+1} = \pi_{\mathbb{D}}[z^j - \alpha(\nabla f(z^j) + H^\top w^j)]$

$\quad w^{j+1} = \pi_{\mathbb{K}^\circ}[w^j + \beta(H(2z^{j+1} - z^j) - g)]$

end for

Algorithm 4 PDHG with varying step sizes

Input: $k, \{\alpha^j, \beta^j, \gamma^j\}_{j=1}^k, \mu, z^1 \in \mathbb{D}, w^1 \in \mathbb{K}^\circ$.

for $j = 1, 2, \ldots, k-1$ **do**

$\quad w^{j+1} = \pi_{\mathbb{K}^\circ}[w^j + \beta^j(H(z^j + \gamma^j(z^j - z^{j-1})) - g)]$

$\quad z^{j+1} = \pi_{\mathbb{D}}\left[z^j - \frac{\alpha^j}{\mu\alpha^j + 1}(\nabla f(z^j) + H^\top w^{j+1})\right]$

end for

The primal–dual gap converges to zero at the rate of $O(1/k)$ and $O(1/k^2)$ for Algorithm 3 and Algorithm 4, respectively (Chambolle & Pock, 2016b). However, to our best knowledge, there is no convergence result on the constraint violation for either Algorithm 3 or Algorithm 4.

3. Proportional–integral projected gradient method

We introduce a novel first-order primal–dual method, named *proportional–integral projected gradient method (PIPG)*, for conic optimization (1), and discuss its convergence rates in terms of the constraint violation and the primal–dual gap.

Algorithm 5 summarizes the proposed method, where $k \in \mathbb{N}$ is the maximum number of iterations, and $\{\alpha^j\}_{j=1}^k$ and $\{\beta^j\}_{j=1}^k$ are sequences of positive scalar step sizes that will be specified later. We note that, instead of maximum number of iterations, one can use alternative stopping criterions, such as the distance between $Hz^j - g$ and cone $\mathbb{K}$ reaching a given tolerance.

Algorithm 5 PIPG

Input: $k, \{\alpha^j, \beta^j\}_{j=1}^k, z^1 \in \mathbb{D}, v^1 \in \mathbb{K}^\circ$.

for $j = 1, 2, \ldots, k-1$ **do**

$\quad w^{j+1} = \pi_{\mathbb{K}^\circ}[v^j + \beta^j(Hz^j - g)]$

$\quad z^{j+1} = \pi_{\mathbb{D}}[z^j - \alpha^j(\nabla f(z^j) + H^\top w^{j+1})]$

$\quad v^{j+1} = w^{j+1} + \beta^j H(z^{j+1} - z^j)$

end for

Later we will show that different convex combinations of $\{z^1, z^2, \ldots, z^k\}$ computed by ALgorithm 5 guarantee different convergence rates under different assumptions.

The name PIPG is due to the following observations. First, if $\mathbb{K} = \{\mathbf{0}_n\}$, then $\mathbb{K}^\circ = \mathbb{R}^m$ and line 2 and line 4 in Algorithm 5 become the following:

$$w^{j+1} = v^j + \beta^j(Hz^j - g), \tag{7a}$$

$$v^{j+1} = v^j + \beta^j(Hz^{j+1} - g). \tag{7b}$$

By using (7b) one can show that

$$v^j = v^1 + \textstyle\sum_{i=2}^{j} \beta^{i-1}(Hz^i - g).$$

Hence v^j is a weighted summation, or numerical *integration*, of $Hz^i - g$ from $i = 2$ to $i = j$. Further, (7a) states that w^j adds a *proportional* term of $Hz^j - g$ to v^j, hence w^j in (7a) is a *proportional–integral* term of $Hz^j - g$. Second, if H is a zero matrix, then line 3 in Algorithm 5 becomes a projected gradient method that minimizes f over set $\mathbb{D}$ (Nesterov, 2018, Sec. 2.2.5). Therefore Algorithm 5 can be interpreted as a combination of proportional–integral feedback control and the projected gradient method. Similar idea has also been popular in equality constrained optimization (Wang & Elia, 2010; Yu & Açıkmeşe, 2020; Yu, Açıkmeşe et al., 2020; Yu, Elango et al., 2020).

Remark 1. Notice that the w^{j+1} in (7a) is otherwise identical to the v^{j+1} in (7b) except that (7a) uses z^j whereas (7b) uses z^{j+1}. Such a scheme is also known as a *prediction–correction* step, which has been popular in many first-order primal–dual methods, including the extra-gradient and mirror-prox method (Korpelevich, 1977; Nemirovski, 2004; Nesterov, 2007), the accelerated linearized ADMM (Ouyang et al., 2015; Xu, 2017), the primal–dual fixed point methods (Chen et al., 2013, 2016; Krol et al., 2012; Yan, 2018) and the accelerated mirror descent method (Cohen et al., 2018).

Next, we will show the convergence results of Algorithm 5. To this end, we will frequently use the following *quadratic distance function* to closed convex cone $\mathbb{K}$:

$$d_{\mathbb{K}}(w) := \underset{v\in\mathbb{K}}{\text{minimize}}\ \tfrac{1}{2}\|w - v\|^2, \tag{8}$$

which is continuously differentiable and convex (Nesterov, 2018, Lem. 2.2.9). We will also use the following *Lagrangian function*:

$$L(z, w) := f(z) + \langle Hz - g, w\rangle. \tag{9}$$

We make the following assumptions on optimization (1).

Assumption 1.

(1) Function $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. There exists $\mu, \lambda \in \mathbb{R}_+$ with $\mu \leqslant \lambda$ such that f is μ-strongly convex and λ-smooth, *i.e.*, $\frac{\mu}{2}\|z - z'\|^2 \leqslant B_f(z, z') \leqslant \frac{\lambda}{2}\|z - z'\|^2$ for all $z, z' \in \mathbb{R}^n$.
(2) Set $\mathbb{D} \subset \mathbb{R}^n$ and cone $\mathbb{K} \subset \mathbb{R}^m$ are closed and convex.
(3) There exist $z^\star \in \mathbb{D}$ and $w^\star \in \mathbb{K}^\circ$ such that $L(z^\star, \overline{w}) \leqslant L(z^\star, w^\star) \leqslant L(\overline{z}, w^\star)$ for all $\overline{z} \in \mathbb{D}$ and $\overline{w} \in \mathbb{K}^\circ$.

Under the above assumptions, the quantity $L(\overline{z}, w^\star) - L(z^\star, \overline{w})$, also known as the *primal–dual gap* evaluated at $(\overline{z}, \overline{w})$, is nonnegative (Boyd et al., 2011; Chambolle & Pock, 2011, 2016b; He & Yuan, 2012). The following proposition provides a sufficient condition on $\overline{z}$ and $\overline{w}$ under which the primal–dual gap $L(\overline{z}, w^\star) - L(z^\star, \overline{w})$ equals zero and $\overline{z}$ is an optimal solution of optimization (1).

Proposition 1. *If there exist $\overline{z} \in \mathbb{D}$ and $\overline{w} \in \mathbb{K}^\circ$ such that*

$$L(\overline{z}, w) - L(z, \overline{w}) \leqslant 0, \tag{10}$$

for all $z \in \mathbb{D}$ and $w \in \mathbb{K}^\circ$, then $\overline{z}$ is an optimal solution of optimization (1), i.e. $H\overline{z} - g \in \mathbb{K}$ and $f(\overline{z}) \leqslant f(z)$ for any $z \in \mathbb{D}$ such that $Hz - g \in \mathbb{K}$.

Proof. See Appendix A.

The following lemma proves a key inequality for our later discussions.

Lemma 1. *Suppose that Assumption 1 holds and $\{w^j, z^j, v^j\}_{j=1}^k$ is computed using Algorithm 5 where $\alpha^j, \beta^j > 0$ and $\alpha^j(\lambda + \sigma\beta^j) = 1$ for some $\sigma \geqslant \|\!|H|\!\|^2$ and all $j = 1, 2, \ldots, k$. Then*

$$\begin{aligned}&\beta^j d_{\mathbb{K}}(Hz^j - g) + L(z^{j+1}, w) - L(z, w^{j+1})\\ &\leqslant \left(\tfrac{1}{2\alpha^j} - \tfrac{\mu}{2}\right)\|z^j - z\|^2 + \tfrac{1}{2\beta^j}\|v^j - w\|^2\\ &\quad - \tfrac{1}{2\alpha^j}\|z^{j+1} - z\|^2 - \tfrac{1}{2\beta^j}\|v^{j+1} - w\|^2,\end{aligned}$$

for all $z \in \mathbb{D}$, $w \in \mathbb{K}^\circ$, and $j = 1, 2, \ldots, k$.

Proof. See Appendix B.

Equipped with Lemma 1, we are ready to prove the convergence results of Algorithm 5. The idea is to first summing up the inequality in Lemma 1 corresponding to different value of j, then using the Jensen's inequality. We start with the case where $\mu = 0$, *i.e.*, function f is merely convex.

Theorem 1. *Suppose that Assumption 1 holds with $\mu = 0$, and $\{w^j, z^j, v^j\}_{j=1}^k$ is computed using Algorithm 5 with $\alpha^j = \frac{1}{\beta\sigma+\lambda}$ and $\beta^j = \beta$ and all $j = 1, 2, \ldots, k$, where $\beta > 0$ and $\sigma \geqslant \|\!|H|\!\|^2$. Let $\tilde{z}^k := \frac{1}{k}\sum_{j=1}^k z^j$, $\overline{z}^k := \frac{1}{k}\sum_{j=1}^k z^{j+1}$, $\overline{w}^k := \frac{1}{k}\sum_{j=1}^k w^{j+1}$, and $V^1(z, w) := \frac{1}{2\alpha}\|z^1 - z\|^2 + \frac{1}{2\beta}\|v^1 - w\|^2$ for all $z \in \mathbb{D}$ and $w \in \mathbb{K}^\circ$. Then $\tilde{z}^k, \overline{z}^k \in \mathbb{D}$, $\overline{w}^k \in \mathbb{K}^\circ$, and*

$$\begin{aligned}d_{\mathbb{K}}(H\tilde{z}^k - g) &\leqslant \tfrac{V^1(z^\star, w^\star)}{\beta k},\\ L(\overline{z}^k, w) - L(z, \overline{w}^k) &\leqslant \tfrac{V^1(z, w)}{k},\end{aligned}$$

for all $z \in \mathbb{D}$, $w \in \mathbb{K}^\circ$.

Proof. See Appendix C.

If $\mu > 0$, *i.e.*, function f is strongly convex, then we can further improve the convergence results in Theorem 1 as follows.

Theorem 2. *Suppose that Assumption 1 holds with $\mu > 0$ and $\{w^j, z^j, v^j\}_{j=1}^k$ is computed using Algorithm 5 with $\alpha^j = \frac{2}{(j+1)\mu+2\lambda}$ and $\beta^j = \frac{(j+1)\mu}{2\sigma}$ for some $\sigma > \|\!|H|\!\|^2$ and all $j = 1, 2, \ldots, k$. Let $\tilde{z}^k := \frac{3}{k(k^2+6k+11)}\sum_{j=1}^k (j+1)(j+2)z^j$, $\overline{z}^k := \frac{2}{k(k+5)}\sum_{j=1}^k (j+2)z^{j+1}$, $\overline{w}^k := \frac{2}{k(k+5)}\sum_{j=1}^k (j+2)w^{j+1}$, and $V^1(z, w) := \frac{\mu+2\lambda}{4}\|z^1 - z\|^2 + \frac{\sigma}{\mu}\|v^1 - w\|^2$ for all $z \in \mathbb{D}$ and $w \in \mathbb{K}^\circ$. Then $\tilde{z}^k, \overline{z}^k \in \mathbb{D}$, $\overline{w}^k \in \mathbb{K}^\circ$, and*

$$\begin{aligned}d_{\mathbb{K}}(H\tilde{z}^k - g) &\leqslant \tfrac{12\lambda\sigma V^1(z^\star, w^\star)}{\mu^2 k(k^2+6k+11)},\\ L(\overline{z}^k, w) - L(z, \overline{w}^k) &\leqslant \tfrac{4\lambda V^1(z, w)}{\mu k(k+5)},\end{aligned}$$

for all $z \in \mathbb{D}$, $w \in \mathbb{K}^\circ$.

Proof. See Appendix D.

Remark 2. Unlike the results in Chambolle and Pock (2016b), Theorems 1 and 2 prove not only the convergence of the primal–dual gap, but also the convergence of the constraints violation. In addition, if $\alpha^j \equiv \alpha$ and $\beta^j \equiv \beta$ for $j = 1, 2, \ldots, k$, then one can show that Algorithm 5 is equivalent to Algorithm 3; in other words, the results in Theorem 1 also apply to Algorithm 3.

Remark 3. When using varying step sizes, Algorithm 5 differs from Algorithm 4 in the relation between step sizes and the iteration number: the one in Algorithm 5 is explicit, whereas the one in Algorithm 4 is implicitly defined by a recursive formula (Chambolle & Pock, 2016b, Sec. 5.2). Furthermore, we can prove the convergence rate of the constraint violation for Algorithm 5, whereas similar rate for Algorithm 4, to our best knowledge, does not exist in the literature.

4. Applications to constrained optimal control

We demonstrate the application of PIPG to constrained optimal control problems. In Section 4.1, we show how to formulate a typical constrained optimal control problem as an instance of conic optimization (1), and provide examples from mechanical engineering and robotics. In Section 4.2, we demonstrate the performance of PIPG via said examples, and compare it against the existing methods reviewed in Section 2.2.

4.1. Constrained optimal control

We consider the following constrained optimal control problem:

$$\begin{aligned} \underset{\{u_t, x_{t+1}\}_{t=0}^{\tau-1}}{\text{minimize}} \quad & \tfrac{1}{2}\sum_{t=0}^{\tau-1}(\|x_{t+1}-\hat{x}_{t+1}\|_Q^2 + \|u_t - \hat{u}_t\|_R^2) \\ \text{subject to} \quad & x_{t+1} = Ax_t + Bu_t + h, \;\; 0 \leqslant t \leqslant \tau-1, \\ & \|u_{t+1} - u_t\|_\infty \leqslant \gamma, \;\; 0 \leqslant t \leqslant \tau-2, \\ & C_t x_t - a_t \geqslant 0, \; x_t \in \mathbb{X}, \;\; 1 \leqslant t \leqslant \tau, \\ & D_t u_t - b_t \geqslant 0, \; u_t \in \mathbb{U}, \;\; 0 \leqslant t \leqslant \tau-1. \end{aligned} \tag{11}$$

The above optimization minimizes the quadratic distance between $\{x_{t+1}, u_t\}_{t=0}^{\tau-1}$ and a reference trajectory $\{\hat{x}_{t+1}, \hat{u}_t\}_{t=0}^{\tau-1}$, subject to state, input, and input rate constraints. One can transform optimization (11) into a special case of optimization (1) using particular choices of the parameters; see Yu et al. (2021, App. E) for details.

We will provide two illustrating examples of optimization (11) from mechanical engineering and robotics applications. Due to the limit of space, we will only provide the choices of set $\mathbb{X}$ and set $\mathbb{U}$; for the detailed values of the other problem parameters (*e.g.*, A and B), see Yu et al. (2021, Sec. 4.1) and Yu et al. (2021, App. F).

4.1.1. Oscillating masses control

We consider the problem of controlling a one-dimensional oscillating masses dynamical system using external forcing (Jerez et al., 2014; Kögel & Findeisen, 2011; Wang & Boyd, 2009); see Fig. 1 for an illustration. The system consists of a sequence of N masses connected by springs to each other, and to walls on either side. Each mass has value 1, and each spring has a spring constant of 1. The state of the system includes the position and velocity of all masses; The input of the system includes an external force applied to each mass.

Based on the oscillating masses dynamical system, we formulate an instance of optimization (11) as follows. The quadratic objective function penalizes any nonzero position, velocity, and external forcing of the N masses. The constraints include interval constraints on the position, the velocity, the external force and its change rate (Yu et al., 2021, Sec. 4.1.1). In this case, we have

$$\begin{aligned} \mathbb{X} &= \{r \in \mathbb{R}^N \mid \|r\|_\infty \leqslant 2\} \times \{s \in \mathbb{R}^N \mid \|s\|_\infty \leqslant 2\}, \\ \mathbb{U} &= \{u \in \mathbb{R}^N \mid \|u\|_\infty \leqslant 2\}. \end{aligned} \tag{12}$$

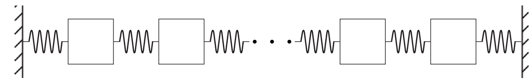


Fig. 1. The oscillating masses system.

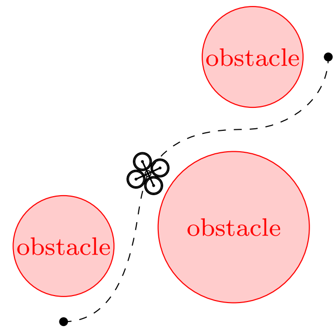


Fig. 2. The quadrotor path planning problem.

4.1.2. Quadrotor path planning

We consider the problem of flying a quadrotor from its initial position to a target position while avoiding collision with cylindrical obstacles; see Fig. 2 for an illustration. For the quadrotor system dynamics, we consider the point mass model introduced in Szmuk et al. (2018, 2017). The state of the system includes the position and velocity of the quadrotor; The input of the system is the thrust vector provided by the on-board propellers.

Based on the quadrotor dynamical system, we formulate an instance of optimization (11) as follows. The quadratic objective function penalizes the deviation from a reference state trajectory and nonzero inputs. The constraints include direction and magnitude constraints on the thrust input, magnitude constraints on the velocity, and linear constraints on the position that ensures collision avoidance with three cylindrical obstacles (Yu et al., 2021, Sec. 4.1.2). In this case, we have

$$\begin{aligned} \mathbb{U} &= \left\{u \in \mathbb{R}^3 \mid \|u\|\cos(\pi/4) \leqslant [u]_3, \|u\| \leqslant 5\right\}, \\ \mathbb{X} &= \{r \in \mathbb{R}^3 \mid \|r\|_\infty \leqslant 3\} \times \{s \in \mathbb{R}^3 \mid \|s\| \leqslant 5\}, \end{aligned} \tag{13}$$

where $[u]_3$ denotes the third element in vector u. Note that set $\mathbb{U}$ is convex; it is the intersection of a second order cone (Bauschke & Combettes, 2017, Ex. 29.12) and a norm ball, both are convex sets themselves. Furthermore, the projection onto set $\mathbb{U}$ can be computed in closed form Bauschke et al. (2018, Thm.7.1).

4.2. Numerical experiments

We demonstrate the numerical performance of PIPG using the two examples of optimization (11), namely the oscillating masses problem and the quadrotor path planning problem discussed in Section 4.1.

Note that interior point methods (IPM) can also efficiently solve optimization (11) (Wang & Boyd, 2009). However, comprehensive studies have already shown that ADMM outperforms IPM in solving optimal control problems (O'Donoghue et al., 2013) and general conic optimization problems (O'Donoghue et al., 2016). Therefore, we will use ADMM, rather than IPM, as the benchmark method for PIPG.

The key step of implementing PIPG method is to compute the projection onto cone $\mathbb{K}^\circ$ and set $\mathbb{D}$. These projections can be computed efficiently for the following reasons. First, projections onto many common closed convex cones and sets can be computed using simple formulas, see Bauschke and Combettes (2017, Chp. 29) for some popular examples. Second, let $\mathbb{D}_1 \subset \mathbb{R}^{n_1}$ and

$\mathbb{D}_2 \subset \mathbb{R}^{n_2}$ be closed convex sets, $x_1 \in \mathbb{R}^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$. Then one can verify the following:

$$\pi_{\mathbb{D}_1 \times \mathbb{D}_2}\left[\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right] = \begin{bmatrix} \pi_{\mathbb{D}_1}[x_1] \\ \pi_{\mathbb{D}_2}[x_2] \end{bmatrix}.$$

Therefore, projections onto sets that are Cartesian products of sets with simple projection formulas, such as the set $\mathbb{X}$ in (12) and (13), also admit simple formulas.

We compare the performance of PIPG, ADMM, PIPGeq and PDHG using optimization (11) as follows. We initialize all methods using vectors whose entries are sampled from the standard normal distribution. We compare the performance of different methods using the convergence of the following two quantities: $\text{error}^j_{\text{opt}} := \frac{\|z^j - z^\star\|^2}{\|z^\star\|^2}$ and $\text{error}^j_{\text{fea}} := \frac{d_{\mathbb{K}}(Hz^j - g)}{\|z^\star\|^2}$, where $z^j \in \mathbb{D}$ is the candidate solution computed of optimization (11) at the jth iteration for $j = 2, 3, \ldots, k$, and $z^\star$ be the ground truth optimal solution of optimization (11) computed using commercial software Mosek (MOSEK ApS, 2019). In addition, we also consider a restarting variant of PIPG and PDHG where the iteration number j is periodically reset to 1. Such restarting scheme is a popular heuristics for improving practical convergence performance of primal–dual methods (Su et al., 2016; Xu, 2017).

The convergence results of different methods in terms of e^j_{opt} and e^j_{opt} using 100 independent random initializations are illustrated in Fig. 3. For "PIPG + restart" and "PDHG + restart", the iteration number j denotes the total number of iterations (*i.e.*, the number of times that the algorithm computes a projection onto cone $\mathbb{K}^\circ$) and not the number of times we restarted. From these results we can see that PIPG has similar convergence as PDHG – although the latter does not have guaranteed convergence rates for the constraint violation – and clearly outperforms the other methods, especially when combined with the restarting heuristics. Note that, although the performance of ADMM is close to PIPG in the oscillating masses example, the per-iteration cost of ADMM is much higher than PIPG, as shown in Table 1. Therefore, PIPG still has clear advantage against ADMM.

5. Conclusions

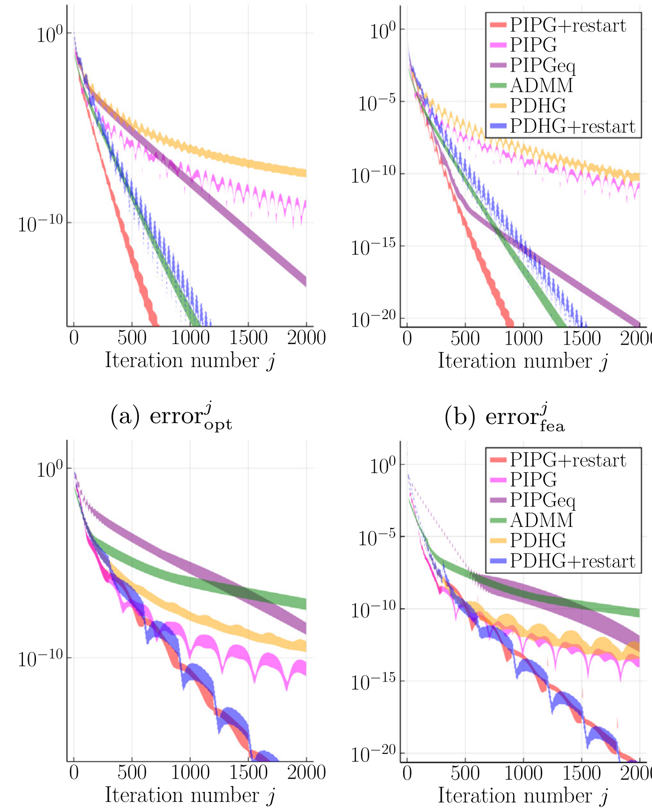
We propose a novel primal–dual first-order method for conic optimization, named PIPG. We prove the convergence rates of PIPG in terms of the constraint violation and the primal–dual gap. We demonstrate the application of PIPG using two examples in constrained optimal control problems.

However, the current work still leaves several questions open. First, the convergence of the constraint violation for PDHG is empirically similar to that of PIPG. Is it possible that PDHG enjoys similar guarantees as PIPG in terms of the convergence of constraint violation? Second, the constant step sizes in Theorem 1 are different from those in Chambolle and Pock (2016b). What are the most general class of step sizes for PIPG and PDHG? Third, PIPG ensures that the constraints violation converges to zero at the rate of either $O(1/k)$ or $O(1/k^3)$. Is it possible to further improve these rates without compromising the convergence of the primal–dual gap? Fourth, although PIPG shows some advantages against the existing first-order methods, it is still unclear whether it can outperform the state-of-the-art second order solvers, *e.g.*, MOSEK, in terms of computation time. We aim to answer these open questions in our future work.

Appendix A. Proof of Proposition 1

First, if (10) holds, then we immediately have

$$L(\bar{z}, w) \leqslant L(\bar{z}, \overline{w}) \leqslant L(z, \overline{w}) \tag{A.1}$$



(a) $\text{error}^j_{\text{opt}}$ (b) $\text{error}^j_{\text{fea}}$ (c) $\text{error}^j_{\text{opt}}$ (d) $\text{error}^j_{\text{fea}}$

Fig. 3. Comparison of different methods for oscillating masses problem (top row) and quadrotor path planning problem (bottom row). The shaded region shows the range of 100 different simulation results using independent random initializations.

for all $z \in \mathbb{D}$ and $w \in \mathbb{K}^\circ$. The first inequality above states that $-L(\bar{z}, \overline{w}) \leqslant -L(\bar{z}, w)$ for all $w \in \mathbb{K}^\circ$, which, due to Rockafellar (2015, Thm. 27.4), implies that

$$\langle H\bar{z} - g, w - \overline{w}\rangle \leqslant 0 \tag{A.2}$$

for all $w \in \mathbb{K}^\circ$. By letting $w = 0$ and $w = 2\overline{w}$ in (A.2), we conclude that

$$\langle H\bar{z} - g, \overline{w}\rangle = 0. \tag{A.3}$$

Combining (A.2) and (A.3) gives $\langle H\bar{z} - g, w\rangle \leqslant 0$ for all $w \in \mathbb{K}^\circ$. Hence $H\bar{z} - g \in (\mathbb{K}^\circ)^\circ = \mathbb{K}$, where the last step is due to Rockafellar and Wets (2009, Cor. 6.21).

Second, let z be such that $z \in \mathbb{D}$ and $Hz - g \in \mathbb{K}$. Since $\overline{w} \in \mathbb{K}^\circ$, using (6) we can show

$$L(z, \overline{w}) = f(z) + \langle Hz - g, \overline{w}\rangle \leqslant f(z). \tag{A.4}$$

Further, using (A.1) and (A.3) we can show

$$f(\bar{z}) = L(\bar{z}, \overline{w}) \leqslant L(z, \overline{w}). \tag{A.5}$$

By combining (A.4) and (A.5) we have $f(\bar{z}) \leqslant f(z)$. Since z is otherwise arbitrary except that $z \in \mathbb{D}$ and $Hz - g \in \mathbb{K}$, the proof is completed.

Appendix B. Proof of Lemma 1

We start with some basic results that will be useful later. By using (4), one can verify the following identity:

$$\begin{aligned} &\langle \nabla f(z) - \nabla f(z'), z'' - z\rangle \\ &= B_f(z'', z') - B_f(z'', z) - B_f(z, z'), \quad \forall z, z', z'' \in \mathbb{R}^n. \end{aligned} \tag{B.1}$$

If $f = \|\cdot\|^2$, the above identify becomes the following:

$$2\langle z - z', z'' - z\rangle = \|z'' - z'\|^2 - \|z'' - z\|^2 - \|z - z'\|^2. \tag{B.2}$$

We now start the main proof. Let z, w, j be an arbitrary element in set $\mathbb{D}$, cone $\mathbb{K}^\circ$, and set $\{1, 2, \ldots, k\}$, respectively. We start with constructing an upper bound for $L(z^{j+1}, w) - L(z, w^{j+1})$. To this end, first we use (9) and (4) to show the following identities

$$\begin{aligned} &L(z^{j+1}, w) - L(z, w) \\ &= B_f(z^{j+1}, z) + \langle \nabla f(z) + H^\top w, z^{j+1} - z\rangle, \end{aligned} \tag{B.3}$$

$$L(z, w) - L(z, w^{j+1}) = \langle Hz - g, w - w^{j+1}\rangle. \tag{B.4}$$

Second, by applying (Nesterov, 2018, Lem. 2.2.7) to the two projections in lines 2 and 3 in Algorithm 5 we can show the following two inequalities

$$0 \leqslant \langle w^{j+1} - v^j - \beta^j(Hz^j - g), w - w^{j+1}\rangle, \tag{B.5a}$$

$$0 \leqslant \langle z^{j+1} - z^j + \alpha^j(\nabla f(z^j) + H^\top w^{j+1}), z - z^{j+1}\rangle. \tag{B.5b}$$

Third, line 4 in Algorithm 5 implies the following

$$0 = \langle v^{j+1} - w^{j+1} - \beta^j H(z^{j+1} - z^j), w - v^{j+1}\rangle. \tag{B.6}$$

Summing up (B.3), (B.4), $\frac{1}{\beta^j}\times$(B.5a), $\frac{1}{\alpha^j}\times$(B.5b) and $\frac{1}{\beta^j}\times$(B.6) gives the following inequality

$$\begin{aligned} &L(z^{j+1}, w) - L(z, w^{j+1}) \\ &\leqslant B_f(z^{j+1}, z) + \langle \nabla f(z) - \nabla f(z^j), z^{j+1} - z\rangle \\ &+ \tfrac{1}{\alpha^j}\langle z^{j+1} - z^j, z - z^{j+1}\rangle + \tfrac{1}{\beta^j}\langle w^{j+1} - v^j, w - w^{j+1}\rangle \\ &+ \tfrac{1}{\beta^j}\langle v^{j+1} - w^{j+1}, w - v^{j+1}\rangle \\ &+ \langle v^{j+1} - w^{j+1}, H(z^{j+1} - z^j)\rangle. \end{aligned} \tag{B.7}$$

Our next step is to bound the inner product terms in (B.7). First, we use (B.1) and (B.2) to show the following identities:

$$\begin{aligned} &\langle \nabla f(z) - \nabla f(z^j), z^{j+1} - z\rangle \\ &= B_f(z^{j+1}, z^j) - B_f(z^{j+1}, z) - B_f(z, z^j), \end{aligned} \tag{B.8}$$

$$\begin{aligned} &2\langle z^{j+1} - z^j, z - z^{j+1}\rangle \\ &= \|z^j - z\|^2 - \|z^{j+1} - z\|^2 - \|z^{j+1} - z^j\|^2, \end{aligned} \tag{B.9}$$

$$\begin{aligned} &2\langle w^{j+1} - v^j, w - w^{j+1}\rangle \\ &= \|v^j - w\|^2 - \|w^{j+1} - w\|^2 - \|w^{j+1} - v^j\|^2, \end{aligned} \tag{B.10}$$

$$\begin{aligned} &2\langle v^{j+1} - w^{j+1}, w - v^{j+1}\rangle \\ &= \|w^{j+1} - w\|^2 - \|v^{j+1} - w\|^2 - \|v^{j+1} - w^{j+1}\|^2. \end{aligned} \tag{B.11}$$

Second, by completing the square we can show

$$\begin{aligned} &2\beta^j\langle v^{j+1} - w^{j+1}, H(z^{j+1} - z^j)\rangle \\ &\leqslant \|v^{j+1} - w^{j+1}\|^2 + (\beta^j)^2\|H(z^{j+1} - z^j)\|^2. \end{aligned} \tag{B.12}$$

Notice that now all inner product terms in (B.7) can be upper bounded. Finally, we further simplify these upper bounds. To this end, first we use the item 1 in Assumption 1 and the fact that $|||H|||^2 \leqslant \sigma$ to show the following

$$B_f(z^{j+1}, z^j) \leqslant \tfrac{\lambda}{2}\|z^{j+1} - z^j\|^2, \tag{B.13a}$$

$$-B_f(z, z^j) \leqslant -\tfrac{\mu}{2}\|z^j - z\|^2, \tag{B.13b}$$

$$\|H(z^{j+1} - z^j)\|^2 \leqslant \sigma\|z^{j+1} - z^j\|^2. \tag{B.13c}$$

Second, we let $y^j := \frac{1}{\beta^j}(v^j + \beta^j(Hz^j - g) - w^{j+1})$. Applying (Rockafellar & Wets, 2009, Cor. 6.21) and Bauschke and Combettes (2017, Thm. 6.30) to the projection in line 2 of Algorithm 5 we can show that $\beta^j y^j \in (\mathbb{K}^\circ)^\circ = \mathbb{K}$. Since $\mathbb{K}$ is a cone and $\beta^j > 0$, we know $y^j \in \mathbb{K}$. Therefore, using (8) and definition of y^j we can show

$$d_{\mathbb{K}}(Hz^j - g) \leqslant \tfrac{1}{2}\|Hz^j - g - y^j\|^2 = \tfrac{1}{2(\beta^j)^2}\|w^{j+1} - v^j\|^2. \tag{B.14}$$

Finally, summing up (B.7), (B.8), $\frac{1}{2\alpha^j}\times$ (B.9), $\frac{1}{2\beta^j}\times$ (B.10), $\frac{1}{2\beta^j}\times$ (B.11), $\frac{1}{2\beta^j}\times$ (B.12), (B.13a), (B.13b), $\frac{\beta^j}{2}\times$(B.13c), and $\beta^j\times$(B.14), and using the assumption that $\alpha^j(\lambda + \sigma\beta^j) = 1$ we obtain the desired results.

Appendix C. Proof of Theorem 1

Let z, w, j be an arbitrary element in set $\mathbb{D}$, $\mathbb{K}^\circ$ and $\{1, 2, \ldots, k\}$, respectively. Let $V^j(z, w) = \frac{1}{2\alpha}\|z^j - z\|^2 + \frac{1}{2\beta}\|w^j - w\|^2$. Since $\alpha^j = \frac{1}{\beta\sigma+\lambda}$ and $\beta^j = \beta$, the inequality in Lemma 1 implies the following:

$$\begin{aligned} &L(z^{j+1}, w) - L(z, w^{j+1}) + \beta d_{\mathbb{K}}(Hz^j - g) \\ &\leqslant V^j(z, w) - V^{j+1}(z, w), \end{aligned}$$

for all $z \in \mathbb{D}$, $w \in \mathbb{K}^\circ$, and $j = 1, 2, \ldots, k$. Summing up this inequality for $j = 1, \ldots, k$ gives

$$\begin{aligned} &\textstyle\sum_{j=1}^k \big(L(z^{j+1}, w) - L(z, w^{j+1}) + \beta d_{\mathbb{K}}(Hz^j - g)\big) \\ &\leqslant V^1(z, w) - V^{k+1}(z, w) \leqslant V^1(z, w), \end{aligned} \tag{C.1}$$

for all $z \in \mathbb{D}$ and $w \in \mathbb{K}^\circ$, where the last step is because $V^{k+1}(z, w) \geqslant 0$. From (8) and item 3 Assumption 1 we know that $d_{\mathbb{K}}(Hz^j - g)$ and $L(z^{j+1}, w^\star) - L(z^\star, w^{j+1})$ are non-negative for all j. Hence (C.1) implies the following

$$\begin{aligned} \textstyle\sum_{j=1}^k \big(L(z^{j+1}, w) - L(z, w^{j+1})\big) &\leqslant V^1(z, w), \\ \beta \textstyle\sum_{j=1}^k d_{\mathbb{K}}(Hz^j - g) &\leqslant V^1(z^\star, w^\star), \end{aligned}$$

for all $z \in \mathbb{D}$, $w \in \mathbb{K}^\circ$, where the second inequality is obtained by letting $z = z^\star$ and $w = w^\star$ in (C.1).

Finally, applying the Jensen's inequality (Nesterov, 2018, Lem. 3.1.1) to convex function $L(\cdot, w)$, $-L(z, \cdot)$, and $d_{\mathbb{K}}(\cdot)$ in the above two inequalities, respectively, we obtain the desired results.

Appendix D. Proof of Theorem 2

Let z, w, j be an arbitrary element in set $\mathbb{D}$, $\mathbb{K}^\circ$ and $\{1, 2, \ldots, k\}$, respectively. Let $V^j(z, w) = \frac{1}{2\alpha^{j-1}}\|z^j - z\|^2 + \frac{1}{2\beta^{j-1}}\|v^j - w\|^2$. Since $\alpha^j = \frac{2}{(j+1)\mu+2\lambda}$ and $\beta^j = \frac{(j+1)\mu}{2\sigma}$, the inequality in Lemma 1 implies the following:

$$\begin{aligned} &L(z^{j+1}, w) - L(z, w^{j+1}) + \tfrac{(j+1)\mu}{2\sigma}d_{\mathbb{K}}(Hz^j - g) \\ &\leqslant \tfrac{1}{2}(\tfrac{1}{\alpha^j} - \mu)\|z^j - z\|^2 + \tfrac{1}{2\beta^j}\|v^j - w\|^2 - V^{j+1}(z, w), \end{aligned} \tag{D.1}$$

for all $z \in \mathbb{D}$, $w \in \mathbb{K}^\circ$, and $j = 1, 2, \ldots, k$. Let $\kappa = \lambda/\mu \geqslant 1$, then one can verify the following

$$\begin{aligned} (\tfrac{1}{\alpha^j} - \mu)(j + 2\kappa) &= \tfrac{1}{\alpha^{j-1}}(j + 2\kappa - 1), \\ \tfrac{1}{\beta^j}(j + 2\kappa) &\leqslant \tfrac{1}{\beta^{j-1}}(j + 2\kappa - 1). \end{aligned} \tag{D.2}$$

Hence multiplying (D.1) with $(j + 2\kappa)$ then substituting in (D.2) we can show

$$\begin{aligned} &(j + 2\kappa)(L(z^{j+1}, w) - L(z, w^{j+1})) + \tfrac{(j+1)(j+2\kappa)\mu}{2\sigma}d_{\mathbb{K}}(Hz^j - g) \\ &\leqslant (j + 2\kappa - 1)V^j(z, w) - (j + 2\kappa)V^{j+1}(z, w), \end{aligned}$$

for all $z \in \mathbb{D}$, $w \in \mathbb{K}^\circ$, and $j = 1, 2, \ldots, k$. Summing up this inequality for $j = 1, 2, \ldots, k$ gives

$$\begin{aligned} &\textstyle\sum_{j=1}^k (j + 2\kappa)(L(z^{j+1}, w) - L(z, w^{j+1})) \\ &+ \textstyle\sum_{j=1}^k \tfrac{(j+1)(j+2\kappa)\mu}{2\sigma}d_{\mathbb{K}}(Hz^j - g) \\ &\leqslant 2\kappa V^1(z, w) - (k + 2\kappa)V^{k+1}(z, w) \leqslant 2\kappa V^1(z, w), \end{aligned} \tag{D.3}$$

for all $z \in \mathbb{D}$ and $w \in \mathbb{K}^\circ$, where the last step is because $V^{k+1}(z, w) \geqslant 0$. From (8) and item 3 in Assumption 1 we know that $d_{\mathbb{K}}(Hz^j - g)$ and $\ell(z^{j+1}, w^{j+1}; z^\star, w^\star)$ are non-negative for all j. Hence the above inequality implies the following

$$\sum_{j=1}^k (j+2)\big(L(z^{j+1}, w) - L(z, v^{j+1})\big) \leqslant 2\kappa V^1(z, w),$$
$$\sum_{j=1}^k \tfrac{(j+1)(j+2)\mu}{2\sigma} d_{\mathbb{K}}(Hz^j - g) \leqslant 2\kappa V^1(z^\star, w^\star),$$

for all $z \in \mathbb{D}$ and $w \in \mathbb{K}^\circ$, where we used the fact that $\kappa \geqslant 1$, and the second inequality is obtained by letting $z = z^\star$ and $w = w^\star$ in (D.3).

Finally, applying the Jensen's inequality (Nesterov, 2018, Lem. 3.1.1) to convex function $L(\cdot, w)$, $-L(z, \cdot)$, and $d_{\mathbb{K}}(\cdot)$ in the above two inequalities, respectively, we obtain the desired results.

References

Andersen, M., Dahl, J., Liu, Z., Vandenberghe, L., Sra, S., Nowozin, S., & Wright, S. (2011). Interior-point methods for large-scale cone programming. In *Optim. mach. learn., vol. 5583*. MIT Press Cambridge, MA.

Andersen, E. D., Roos, C., & Terlaky, T. (2003). On implementing a primal-dual interior-point method for conic quadratic optimization. *Mathematical Programming*, *95*(2), 249–277.

Bauschke, H. H., Bui, M. N., & Wang, X. (2018). Projecting onto the intersection of a cone and a sphere. *SIAM Journal on Optimization*, *28*(3), 2158–2188.

Bauschke, H. H., & Combettes, P. L. (2017). *Convex analysis and monotone operator theory in Hilbert spaces, vol. 408*. Springer.

Ben-Tal, A., & Nemirovski, A. (2001). *Lectures on modern convex optimization: Analysis, algorithms, and engineering applications*. SIAM.

Boyd, S., Parikh, N., & Chu, E. (2011). *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc.

Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Chambolle, A., & Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, *40*(1), 120–145.

Chambolle, A., & Pock, T. (2016a). An introduction to continuous optimization for imaging. *Acta Numerica*, *25*, 161–319.

Chambolle, A., & Pock, T. (2016b). On the ergodic convergence rates of a first-order primal–dual algorithm. *Mathematical Programming*, *159*(1–2), 253–287.

Chen, P., Huang, J., & Zhang, X. (2013). A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration. *Inverse Problems*, *29*(2), Article 025011.

Chen, P., Huang, J., & Zhang, X. (2016). A primal-dual fixed point algorithm for minimization of the sum of three convex separable functions. *Fixed Point Theory and Applications*, *2016*(1), 1–18.

Cohen, M., Diakonikolas, J., & Orecchia, L. (2018). On acceleration with noise-corrupted gradients. In *Int. conf. mach. learn.* (pp. 1019–1028). PMLR.

Eckstein, J. (1989). *Splitting methods for monotone operators with applications to parallel optimization* (Ph.D. thesis), Massachusetts Inst. Technol..

Eren, U., Prach, A., Koçer, B. B., Raković, S. V., Kayacan, E., & Açıkmeşe, B. (2017). Model predictive control in aerospace systems: Current state and opportunities. *Journal of Guidance, Control, and Dynamics*, *40*(7), 1541–1566.

Fortin, M., & Glowinski, R. (2000). *Augmented lagrangian methods: Applications to the numerical solution of boundary-value problems*. Elsevier.

Gabay, D., & Mercier, B. (1976). A dual algorithm for the solution of non-linear variational problems via finite element approximation. *Computers & Mathematics with Applications*, *2*(1), 17–40.

Goldstein, T., O'Donoghue, B., Setzer, S., & Baraniuk, R. (2014). Fast alternating direction optimization methods. *SIAM Journal on Imaging Sciences*, *7*(3), 1588–1623.

He, B., & Yuan, X. (2012). On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method. *SIAM Journal on Numerical Analysis*, *50*(2), 700–709.

Jerez, J. L., Goulart, P. J., Richter, S., Constantinides, G. A., Kerrigan, E. C., & Morari, M. (2014). Embedded online optimization for model predictive control at megahertz rates. *IEEE Transactions on Automatic Control*, *59*(12), 3238–3251.

Kadkhodaie, M., Christakopoulou, K., Sanjabi, M., & Banerjee, A. (2015). Accelerated alternating direction method of multipliers. In *Proc int. conf. knowl. discovery data mining* (pp. 497–506).

Kögel, M., & Findeisen, R. (2011). Fast predictive control of linear systems combining Nesterov's gradient method and the method of multipliers. In *Proc. IEEE conf. decision control and eur. control conf.* (pp. 501–506). IEEE.

Korpelevich, G. (1977). Extragradient method for finding saddle points and other problems. *Matekon*, *13*(4), 35–49.

Krol, A., Li, S., Shen, L., & Xu, Y. (2012). Preconditioned alternating projection algorithms for maximum a posteriori ECT reconstruction. *Inverse Problems*, *28*(11), Article 115005.

Lan, G., Lu, Z., & Monteiro, R. D. (2011). Primal-dual first-order methods with $\mathcal{O}(1/\epsilon)$ iteration-complexity for cone programming. *Mathematical Programming*, *126*(1), 1–29.

Liu, X., Lu, P., & Pan, B. (2017). Survey of convex optimization for aerospace applications. *Astrodynamics*, *1*(1), 23–40.

Luo, Z.-Q., & Yu, W. (2006). An introduction to convex optimization for communications and signal processing. *IEEE Journal on Selected Areas in Communications*, *24*(8), 1426–1438.

Majumdar, A., Hall, G., & Ahmadi, A. A. (2020). Recent scalability improvements for semidefinite programming with applications in machine learning, control, and robotics. *Annual Review of Control, Robotics, and Autonomous Systems*, *3*, 331–360.

Malyuta, D., Yu, Y., Elango, P., & Açıkmeşe, B. (2021). Advances in trajectory optimization for space vehicle control. arXiv preprint arXiv:2108.02335 [math.OC].

MOSEK ApS (2019). The MOSEK optimization toolbox for MATLAB manual. Version 9.0..

Nemirovski, A. (2004). Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, *15*(1), 229–251.

Nesterov, Y. (2007). Dual extrapolation and its applications to solving variational inequalities and related problems. *Mathematical Programming*, *109*(2), 319–344.

Nesterov, Y. (2018). *Lectures on convex optimization, vol. 137*. Springer.

Nesterov, Y., & Nemirovskii, A. (1994). *Interior-point polynomial algorithms in convex programming*. SIAM.

O'Donoghue, B., Chu, E., Parikh, N., & Boyd, S. (2016). Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, *169*(3), 1042–1068.

O'Donoghue, B., Stathopoulos, G., & Boyd, S. (2013). A splitting method for optimal control. *IEEE Transactions on Control Systems Technology*, *21*(6), 2432–2442.

Ouyang, Y., Chen, Y., Lan, G., & Pasiliao, E., Jr. (2015). An accelerated linearized alternating direction method of multipliers. *SIAM Journal on Imaging Sciences*, *8*(1), 644–681.

Rockafellar, R. T. (2015). *Convex analysis*. Princeton University Press.

Rockafellar, R. T., & Wets, R. J.-B. (2009). *Variational analysis, vol. 317*. Springer Science & Business Media.

Stellato, B., Banjac, G., Goulart, P., Bemporad, A., & Boyd, S. (2020). OSQP: An operator splitting solver for quadratic programs. *Mathematical Programming Computation*, *12*(4), 637–672.

Su, W., Boyd, S., & Candes, E. J. (2016). A differential equation for modeling nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, *17*(1), 5312–5354.

Szmuk, M., Pascucci, C. A., & Açıkmeşe, B. (2018). Real-time quad-rotor path planning for mobile obstacle avoidance using convex optimization. In *Proc. int. conf. intell. robots and syst.* (pp. 1–9). IEEE.

Szmuk, M., Pascucci, C. A., Dueri, D., & Açıkmeşe, B. (2017). Convexification and real-time on-board optimization for agile quad-rotor maneuvering and obstacle avoidance. In *Proc. int. conf. intell. robots syst.* (pp. 4862–4868). IEEE.

Wang, H., & Banerjee, A. (2014). Bregman alternating direction method of multipliers. In *Proc. adv. neural inf. process. syst., vol. 4, no. January* (pp. 2816–2824).

Wang, Y., & Boyd, S. (2009). Fast model predictive control using online optimization. *IEEE Transactions on Control Systems Technology*, *18*(2), 267–278.

Wang, J., & Elia, N. (2010). Control approach to distributed optimization. In *Proc. Allerton conf. commun. control comput.* (pp. 557–561). IEEE.

Xu, Y. (2017). Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming. *SIAM Journal on Optimization*, *27*(3), 1459–1484.

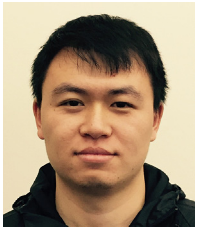
Yan, M. (2018). A new primal–dual algorithm for minimizing the sum of three functions with a linear operator. *Journal of Scientific Computing*, *76*(3), 1698–1717.

Yu, Y., & Açıkmeşe, B. (2020). RLC circuits-based distributed mirror descent method. *IEEE Control Systems Letters*, *4*(3), 548–553.

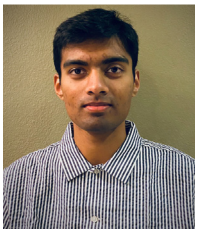
Yu, Y., Açıkmeşe, B., & Mesbahi, M. (2020). Mass–spring–damper networks for distributed optimization in non-Euclidean spaces. *Automatica*, *112*, Article 108703.

Yu, Y., Elango, P., & Açıkmeşe, B. (2020). Proportional-integral projected gradient method for model predictive control. *IEEE Control Systems Letters*.

Yu, Y., Elango, P., Topcu, U., & Açıkmeşe, B. (2021). Proportional-integral projected gradient method for conic optimization. arXiv preprint arXiv:2108.10260 [math.OC].



Yue Yu is a postdoctoral scholar at the Oden Institute for Computational Engineering and Sciences at The University of Texas at Austin. In 2020, Yue obtained his Doctor of Philosophy degree from the William E. Boeing Department of Aeronautics and Astronautics at University of Washington. Yue's research interests include optimization, optimization-based control, network games, and data-driven control.



Purnanand Elango is a Ph.D. student at the Department of Aeronautics and Astronautics in the University of Washington, Seattle. He received B.Tech and M.Tech in Aerospace Engineering from the Indian Institute of Technology Madras. His research interests include real-time trajectory optimization and robust control for space applications.

Ufuk Topcu is an Associate Professor in the Department of Aerospace Engineering and Engineering Mechanics at The University of Texas at Austin. He is a core faculty member at the Oden Institute for Computational Engineering and Sciences. Ufuk obtained his Doctor of Philosophy degree from the University of California, Berkeley in 2008. Prior to joining The University of Texas at Austin, he was with the Department of Electrical and Systems Engineering at the University of Pennsylvania. He was a postdoctoral scholar at California Institute of Technology until 2012. Ufuk's research focuses on the theoretical and algorithmic aspects of the design and verification of autonomous systems, typically in the intersection of formal methods, reinforcement learning, and control theory.

Behçet Açıkmeşe is a professor at University of Washington (UW), Seattle. He received his Ph.D. in Aerospace Engineering from Purdue University. Before joining UW, he was a senior technologist at JPL and a lecturer at Caltech, and later a faculty member at the University of Texas at Austin. At JPL, he developed control algorithms for planetary landing, spacecraft formation flying, and asteroid and comet sample return missions. He developed the ''flyaway" control algorithms in Mars Science Laboratory (MSL) mission, which landed on Mars in 2012. His research interests include optimization and its applications in control theory, Markov decision processes, nonlinear and robust control, and game theory.