Automated Lyapunov Analysis of Primal-Dual Optimization Algorithms: An Interpolation Approach

Bryan Van Scoy[1]   John W. Simpson-Porco[2]   Laurent Lessard[3]

Abstract

Primal-dual algorithms are frequently used for iteratively solving large-scale convex optimization problems. The analysis of such algorithms is usually done on a case-by-case basis, and the resulting guaranteed rates of convergence can be conservative. Here we consider a class of first-order algorithms for linearly constrained convex optimization problems, and provide a linear matrix inequality (LMI) analysis framework for certifying worst-case exponential convergence rates. Our approach builds on recent results for interpolation of convex functions and linear operators, and our LMI directly constructs a Lyapunov function certifying the guaranteed convergence rate. By comparing to rates established in the literature, we show that our approach can certify significantly faster convergence for this family of algorithms.

1 Introduction

Primal-dual (or saddle-point) optimization methods have a rich history, dating back to the earliest days of mathematical programming [1, 2]. The core idea — that of sequentially or simultaneously updating both primal and dual variables — is now widely used in algorithms for solving constrained optimization problems, including in interior-point methods [3], the method of multipliers, and in distributed optimization methods such as ADMM [4].

There is significant overlap with the literature on *operator splitting*, as finding a point satisfying the KKT conditions of a constrained optimization problem can be cast as the problem of finding a zero of a sum of monotone operators; see [5–7] for extensive overviews of this perspective. Convergence proofs in this literature generally rely on the construction of bespoke pre-conditioners, followed by applications of convergence results for known fixed-point algorithms (e.g., Krasnoselskii-Mann iterations), or by clever direct construction of Lyapunov-like functions.

While we do not herein study operator splitting methods in generality, part of our goal is to establish some preliminary foundations for more systematic and automated analyses of such algorithms.

Primal-dual methods have attracted attention from controls researchers, particularly in the continuous-time setting where Lyapunov analysis techniques can be applied with relative ease; see [8–16]. This has led to various control applications of the algorithms, such as in energy systems [17–21]. However, only [15] directly addresses the issue of the *rate* of exponential convergence in discrete time, and the analysis presented is focused on a particular algorithm obtained via Euler discretization from the continuous-time version.

By interpreting optimization algorithms as dynamical systems, specifically as robust controllers, integral quadratic constraints (IQCs) have been used to find tight bounds on convergence rates [22, 23]. Alternatively, one can directly generate a set of valid inequalities relating inputs and outputs of the objective function, and solve a *meta*-optimization problem that searches for tight worst-case guarantees. This was applied in a *finite-horizon* setting in the so-called PEP formulation [24], and also in an asymptotic setting [25–27] to directly search for Lyapunov functions that certify a given convergence rate.

To the best of our knowledge, the aforementioned approaches have not previously been applied to analyze primal-dual algorithms for linearly constrained convex optimization. The closest works we found examined over-relaxed ADMM [28], or alternating gradient methods for bilinear games [29] or smooth monotone games [30].

Contributions: We consider a family of first-order primal-dual algorithms for solving linearly constrained convex optimization problems of the form

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ f(x) \quad \text{subject to} \quad Ax = b. \qquad (1)$$

Our main contribution is an automated framework for computing worst-case convergence rates of the algorithm over a class of problem data. We consider smooth and strongly convex f and matrices A with known bounds on the singular values. We show numerically that our analysis improves on known results. We also develop the set of multipliers for the class of smooth strongly convex functions and the class of linear functions with eigenvalues in a closed interval which may be of independent interest.

---

[1]B. Van Scoy is with the Department of Electrical and Computer Engineering at Miami University, Oxford, OH 45056, USA. `bvanscoy@miamioh.edu`

[2]J. W. Simpson-Porco is with the Department of Electrical and Computer Engineering at the University of Toronto, Toronto, ON, M5S 3G4, Canada. `jwsimpson@ece.utoronto.ca`

[3]L. Lessard is with the Department of Mechanical and Industrial Engineering at Northeastern University, Boston, MA 02115, USA. `l.lessard@northeastern.edu`

2 Primal-dual iterations for linearly constrained convex optimization

Consider the optimization problem (1), where the goal is to minimize the objective function $f : \mathbb{R}^n \to \mathbb{R}$ over the affine constraint set $\mathsf{C} := \{x \in \mathbb{R}^n \mid Ax = b\}$ with $A \in \mathbb{R}^{r \times n}$ and $b \in \mathbb{R}^r$. Throughout this work, we make the following assumptions, which (among other things) ensure that the problem is feasible for any $b \in \mathbb{R}^r$ and possess a unique optimal solution $x_\star \in \mathbb{R}^n$.

(i) For known constants $0 < m \leq L < \infty$, the objective function f is m-strongly convex, continuously differentiable, and its gradient $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is globally Lipschitz continuous with Lipschitz constant L; we denote the set of all such functions by $\mathcal{F}(m, L)$, and we denote the condition number as $\kappa(f) = L/m$.

(ii) For known constants $0 < \underline{\sigma} \leq \overline{\sigma} < \infty$, the singular values $\sigma_i(A)$ of the constraint matrix $A \in \mathbb{R}^{r \times n}$ satisfy $\underline{\sigma} \leq \sigma_i(A) \leq \overline{\sigma}$ for all $i \in \{1, \dots, r\}$; we denote the set of all such matrices by $\mathcal{A}(\underline{\sigma}, \overline{\sigma})$, and we denote the condition number as $\kappa(A) = \overline{\sigma}/\underline{\sigma}$.

Note that (ii) implies that A has full row rank, and thus the constraints $Ax = b$ are linearly independent.

Perhaps the most immediate iterative approach for computing the optimal solution of (1) would be projected gradient descent

$$x_{k+1} = \mathrm{Proj}_{\mathsf{C}}(x_k - \alpha \nabla f(x_k)), \qquad k = 0, 1, \dots$$

with step size $\alpha > 0$. In many large-scale applications however, computing this projection is too computationally expensive, and one encounters similar computational bottlenecks if gradient ascent is applied to the dual problem of (1). Instead, iterative methods are sought which rely only on (a small number of) evaluations of ∇f, A, and A^{T} at each iteration [5,6].

Such methods can be developed through Lagrange relaxation of the equality constraint in (1). For $\mu \geq 0$ define the *augmented Lagrangian*

$$L_\mu(x, \lambda) = f(x) + \lambda^{\mathsf{T}}(Ax - b) + \tfrac{\mu}{2}\|Ax - b\|_2^2,$$

with $\lambda \in \mathbb{R}^r$ the dual variable and μ the augmentation parameter. Under the present assumptions, strong duality holds, and $(x_\star, \lambda_\star)$ is primal-dual optimal for (1) if and only if it is a *saddle point* of L_μ. As an iterative method to determine a saddle point, one begins with any initial condition (x_0, λ_0) and performs the gradient decent-ascent iterations

$$
\begin{aligned}
x_{k+1} &= x_k - \alpha_x \nabla_x L_\mu(x_k, \lambda_k) \\
&= x_k - \alpha_x[\nabla f(x_k) + A^{\mathsf{T}}\lambda_k + \mu A^{\mathsf{T}}(Ax_k - b)] \quad \text{(2a)} \\
\lambda_{k+1} &= \lambda_k + \alpha_\lambda \nabla_\lambda L_\mu(x_k, \lambda_k) \\
&= \lambda_k + \alpha_\lambda(Ax_k - b) \quad \text{(2b)}
\end{aligned}
$$

where $\alpha_x, \alpha_\lambda > 0$ are step sizes. For obvious reasons, such algorithms are termed *primal-dual* algorithms. In this work, we consider a slightly more general variation on (2a), given by

$$x_{k+1} = x_k - \alpha_x[\nabla f(x_k) + A^{\mathsf{T}}\lambda_k + \mu A^{\mathsf{T}}(Ax_k - b)] \quad \text{(3a)}$$

$$\tilde{x}_k = x_k + \gamma(x_{k+1} - x_k) \quad \text{(3b)}$$

$$\lambda_{k+1} = \lambda_k + \alpha_\lambda(A\tilde{x}_k - b), \quad \text{(3c)}$$

where $\gamma \in [0, 2]$ is an *extrapolation* parameter. Various algorithms are contained as special cases of (3a) and will serve as points of comparison. Our broad goal is to quantify the *worst-case* asymptotic geometric convergence rates achieved by some selected iterative primal-dual algorithms over all possible instances of problem data $f \in \mathcal{F}(m, L)$ and $A \in \mathcal{A}(\underline{\sigma}, \overline{\sigma})$.

2.1 Literature on known rates

For sufficiently small step sizes, (3a) converges exponentially to the unique saddle point $(x_\star, \lambda_\star)$ of L_μ. A significantly more challenging question is to provide nonconservative estimates of the worst-case asymptotic geometric convergence rate for the method over the class of problem data defined by $(\mathcal{F}, \mathcal{A})$.

We have found two relatively clear comparison points. First, [31] considers (3a) with $\mu = \gamma = 0$. Translating the notation[1], they use a Lyapunov function of the form

$$V(x_k, \lambda_k) = \|x_k - \nabla f^*(-A^{\mathsf{T}}\lambda_k)\|_2 + c\|\lambda_k - \lambda_\star\|_2,$$

where $f^*(z) = \sup_{x \in \mathbb{R}^n} x^{\mathsf{T}}z - f(x)$ is the convex conjugate of f [32] and $c := 2\frac{L}{m^2}\frac{\overline{\sigma}^3}{\underline{\sigma}^2}$. Using stepsizes

$$\alpha_x = \frac{2}{m + L} \quad \text{and} \quad \alpha_\lambda = \frac{m}{(m+L)\left(\frac{\overline{\sigma}^2}{m} + c\overline{\sigma}\right)}, \quad \text{(4)}$$

the decrease condition $V_{k+1} \leq \rho V_k$ holds with

$$\rho = 1 - \frac{1}{12\,\kappa(f)^3\,\kappa(A)^4}. \quad \text{(5)}$$

Second, the authors of [33] consider (3a) with $\gamma = 1$, and for notational simplicity we consider here $\mu = 0$. Using the quadratic Lyapunov function

$$V(x_k, \lambda_k) = (1 - \alpha_x\alpha_\lambda\overline{\sigma}^2)\|x_k - x_\star\|_2^2 + \|\lambda_k - \lambda_\star\|_2^2$$

and the step size conditions $\alpha_x < 1/L$ and $\alpha_\lambda < m/\overline{\sigma}^2$, they establish the decrease condition $V_{k+1} \leq \rho^2 V_k$ with

$$\rho^2 = \max\{1 - \alpha_x m\,(1 - \alpha_x L),\ 1 - \alpha_x\alpha_\lambda\underline{\sigma}^2\}.$$

The bound on the convergence rate is optimized by the step sizes

$$\alpha_x = \begin{cases} \frac{1}{2L} & \text{if } \kappa(A) \leq \sqrt{2} \\ \frac{1 - \kappa(A)^{-2}}{L} & \text{otherwise} \end{cases} \quad \text{(6a)}$$

$$\alpha_\lambda = \begin{cases} \frac{m}{4}\left(\frac{2}{\overline{\sigma}^2} + \frac{1}{\underline{\sigma}^2}\right) & \text{if } \kappa(A) \leq \sqrt{2} \\ \frac{m}{\overline{\sigma}^2} & \text{otherwise}, \end{cases} \quad \text{(6b)}$$

and the convergence factor using these step sizes is

$$\rho^2 = \begin{cases} 1 - \frac{1}{4\,\kappa(f)} & \text{if } \kappa(A) \leq \sqrt{2} \\ 1 - \frac{1}{\kappa(f)}\left(\frac{1}{\kappa(A)^2} - \frac{1}{\kappa(A)^4}\right) & \text{otherwise.} \end{cases} \quad \text{(7)}$$

---

[1]In the notation of [31], our set-up corresponds to the case where $f = 0$; our "A" is their "A^{T}" and our "f" is their "$-g$".

3 Automated convergence analysis

We now return to the primal-dual algorithm (3a) and rewrite it in the form of a linear fractional representation, as traditionally used in robust control [34]. Let $A = U\Sigma V_1^\mathsf{T}$ be a compact SVD of A, where $U \in \mathbb{R}^{r \times r}$ is orthogonal and $V_1 \in \mathbb{R}^{n \times r}$ has orthonormal columns. Let $V_2 \in \mathbb{R}^{n \times (n-r)}$ be the orthogonal completion, so that $V = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \in \mathbb{R}^{n \times n}$ is orthogonal. Here, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$ and we have the inequality $\bar\sigma \geq \sigma_1 \geq \cdots \geq \sigma_r \geq \underline\sigma$. Consider now the invertible change of state

$$p_k = V_1^\mathsf{T} x_k, \qquad q_k = V_2^\mathsf{T} x_k, \qquad \nu_k = -\Sigma U^\mathsf{T} \lambda_k.$$

Routine computations quickly show that

$$\begin{bmatrix} p_{k+1} \\ q_{k+1} \end{bmatrix} = \begin{bmatrix} p_k \\ q_k \end{bmatrix} - \alpha_x V^\mathsf{T} \nabla f(V \begin{bmatrix} p_k \\ q_k \end{bmatrix}) + \alpha_x \begin{bmatrix} \nu_k - \mu\Sigma^2 p_k \\ 0 \end{bmatrix}$$
$$\tilde p_k = p_k + \gamma (p_{k+1} - p_k)$$
$$\nu_{k+1} = \nu_k - \alpha_\lambda \Sigma^2 \tilde p_k.$$

Eliminating $\tilde p_k$ and defining the concatenated state $\xi_k = (p_k, q_k, \nu_k)$, the dynamics can be expressed as

$$\xi_{k+1} = A\xi_k + B_1 \begin{bmatrix} u_k^1 \\ u_k^2 \end{bmatrix} + B_2 \begin{bmatrix} u_k^3 \\ u_k^4 \end{bmatrix}$$
$$\begin{bmatrix} y_k^1 \\ y_k^2 \end{bmatrix} = C_1 \xi_k + D_{11} \begin{bmatrix} u_k^1 \\ u_k^2 \end{bmatrix} + D_{12} \begin{bmatrix} u_k^3 \\ u_k^4 \end{bmatrix}$$
$$\begin{bmatrix} y_k^3 \\ y_k^4 \end{bmatrix} = C_2 \xi_k + D_{21} \begin{bmatrix} u_k^1 \\ u_k^2 \end{bmatrix} + D_{22} \begin{bmatrix} u_k^3 \\ u_k^4 \end{bmatrix}$$

where the matrices are defined by the blocks

$$\begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{bmatrix} = \left[\begin{array}{ccc|cc|cc} I & 0 & \alpha_x I & -\alpha_x I & 0 & -\alpha_x \mu I & 0 \\ 0 & I & 0 & 0 & -\alpha_x I & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & -\alpha_\lambda I \\ \hline I & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & 0 \\ \hline I & 0 & 0 & 0 & 0 & 0 & 0 \\ I & 0 & \gamma\alpha_x I & -\gamma\alpha_x I & 0 & -\gamma\alpha_x \mu I & 0 \end{array}\right]$$

and with inputs (u_k^1, u_k^2) and (u_k^3, u_k^4) defined by

$$\begin{bmatrix} u_k^1 \\ u_k^2 \end{bmatrix} = V^\mathsf{T} \nabla f\left(V \begin{bmatrix} y_k^1 \\ y_k^2 \end{bmatrix}\right), \qquad \begin{bmatrix} u_k^3 \\ u_k^4 \end{bmatrix} = \begin{bmatrix} \Sigma^2 & 0 \\ 0 & \Sigma^2 \end{bmatrix} \begin{bmatrix} y_k^3 \\ y_k^4 \end{bmatrix} \quad (8)$$

and outputs $(y_k^1, y_k^2, y_k^3, y_k^4) = (p_k, q_k, p_k, \tilde p_k)$. To simplify notation in the sequel, we assume without loss of generality that each input and output is one-dimensional; see the lossless dimensionality reduction in [22].

3.1 Lifted dynamics

Let $G(z)$ denote the transfer function corresponding to the state-space matrices above, which maps the set of inputs $(u_k^1, u_k^2, u_k^3, u_k^4)$ to the set of outputs $(y_k^1, y_k^2, y_k^3, y_k^4)$. This system is connected with the feedback in (8) through the gradient of the objective function and the squared matrix of singular values.

To analyze the system, we will replace the feedback in (8) with constraints on the inputs and outputs of ∇f and Σ^2. The more constraints that we use, the tighter

the analysis will be. To obtain more constraints, we will lift the system so that the inputs and outputs are in a higher-dimensional space, and then apply the constraints between all iterates in this lifted space [25].

Given a lifting dimension $\ell \in \{1, 2, \ldots\}$, let

$$\psi(z) = I_4 \otimes \begin{bmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{1-\ell} \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} \psi G \\ \psi \end{bmatrix} \quad (9)$$

where $\otimes$ denotes the Kronecker product. The system $\mathbf{G}$, called the *lifted system*, is the map

$$(u_k^1, u_k^2, u_k^3, u_k^4) \mapsto (Y_k^1, \ldots, Y_k^4, U_k^1, \ldots, U_k^4) \quad (10a)$$

where the lifted iterates are

$$Y_k^i = (y_k^i, y_{k-1}^i, \ldots, y_{k-\ell+1}^i), \quad i \in \{1, 2, 3, 4\} \quad (10b)$$

and similarly for U_k^i. Each lifted iterate Y_k^i and U_k^i is an ℓ-dimensional vector that consists of ℓ lagged iterates, where the lifting dimension is how many past iterates are used. When $\ell = 1$, the lifted system is simply $\mathbf{G} = \begin{bmatrix} G \\ I \end{bmatrix}$.

3.2 Multipliers

To analyze the system, we will replace the feedback (8) with inequalities on the inputs and outputs of the system in the lifted space. We parameterize the set of inequalities using a symmetric block matrix, called a *multiplier*, of the form

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}. \quad (11)$$

We now describe the constraints for both the objective function and constraint matrix.

Objective function. We consider inequalities on the objective function gradient of the form

$$0 \leq \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \left((M_{11})_{ij} \begin{bmatrix} y_{k-i+1}^1 \\ y_{k-i+1}^2 \end{bmatrix}^\mathsf{T} \begin{bmatrix} y_{k-j+1}^1 \\ y_{k-j+1}^2 \end{bmatrix} \right.$$
$$+ 2 (M_{12})_{ij} \begin{bmatrix} y_{k-i+1}^1 \\ y_{k-i+1}^2 \end{bmatrix}^\mathsf{T} \begin{bmatrix} u_{k-j+1}^1 \\ u_{k-j+1}^2 \end{bmatrix}$$
$$\left. + (M_{22})_{ij} \begin{bmatrix} u_{k-i+1}^1 \\ u_{k-i+1}^2 \end{bmatrix}^\mathsf{T} \begin{bmatrix} u_{k-j+1}^1 \\ u_{k-j+1}^2 \end{bmatrix} \right). \quad (12)$$

The multiplier M has dimensions $2\ell \times 2\ell$ and parameterizes all inequalities that are linear in the inner products between inputs and outputs of the gradient. The following result characterizes all multipliers such that this inequality holds for all iterates of the system. We provide the proof in the appendix.

Proposition 1 (Objective multipliers). The quadratic inequality (12) holds for all iterates that satisfy the feedback (8) for some function $f \in \mathcal{F}(m, L)$ if and only if the multiplier has the form

$$M = \begin{bmatrix} -2mL & L+m \\ L+m & -2 \end{bmatrix} \otimes R + \begin{bmatrix} \mathbf{0} & S \\ S^\mathsf{T} & \mathbf{0} \end{bmatrix} \quad (13)$$

for some $\ell \times \ell$ symmetric matrix R such that $R\mathbf{1} = 0$, $R_{ii} \geq 0$ for all i, and $R_{ij} \leq 0$ for all $i \neq j$,[2] and some $\ell \times \ell$ skew-symmetric matrix S satisfying $S\mathbf{1} = \mathbf{0}$. We denote the set of all such matrices as $\mathcal{M}_{\mathcal{F}}(m, L)$.

Constraint matrix. We consider inequalities on the constraint matrix of the form

$$0 \leq \text{tr}\left(M \begin{bmatrix} Y_k^3 \\ Y_k^4 \\ U_k^3 \\ U_k^4 \end{bmatrix} \begin{bmatrix} Y_k^3 \\ Y_k^4 \\ U_k^3 \\ U_k^4 \end{bmatrix}^{\mathsf{T}} \right). \tag{14}$$

The multiplier M has dimensions $4\ell \times 4\ell$ and parameterizes all inequalities that are linear in the inner products between inputs and outputs of the matrix Σ^2 of squared singular values of A. The following result characterizes the set of multipliers, which we prove in the appendix.

Proposition 2 (Constraint multipliers)**.** The quadratic inequality (12) holds for all iterates that satisfy the feedback (8) for some matrix $A \in \mathcal{A}(\underline{\sigma}, \overline{\sigma})$ if the multiplier has the form

$$M = \begin{bmatrix} -2\underline{\sigma}^2\overline{\sigma}^2 R & (\overline{\sigma}^2 + \underline{\sigma}^2)R \\ (\overline{\sigma}^2 + \underline{\sigma}^2)R & -2R \end{bmatrix} + \begin{bmatrix} 0 & S \\ S^{\mathsf{T}} & 0 \end{bmatrix} \tag{15}$$

for some $2\ell \times 2\ell$ symmetric positive semidefinite matrix R, and some $2\ell \times 2\ell$ skew-symmetric matrix S. We denote the set of all such matrices as $\mathcal{M}_{\mathcal{A}}(\underline{\sigma}, \overline{\sigma})$.

3.3 Linear matrix inequality

We now use the lifted system (9) and the multipliers (13) and (15) that characterize the objective function and constraint matrix to construct a linear matrix inequality (LMI) whose feasibility certifies convergence of the primal-dual algorithm (3a) with a specified rate $\rho \in (0, 1)$.

Denote a minimal realization of the lifted system as $\mathbf{G} \sim (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ and let $\mathbf{n}$ denote the dimension of the realization. Recall that the lifted system maps the iterates as in (10). Let $\mathbf{Y_f}$ and $\mathbf{U_f}$ denote the rows of $\mathbf{C}$ and $\mathbf{D}$ corresponding to pairs of inputs and outputs of the gradient, and let $\mathbf{Y_A}$ and $\mathbf{U_A}$ denote rows corresponding to inputs and outputs of Σ^2. We can now state our main result.

Theorem 1 (Analysis)**.** Given $\rho \in (0, 1)$, if there exists an $\mathbf{n} \times \mathbf{n}$ symmetric matrix P and multipliers $M_1, M_2 \in \mathcal{M}_{\mathcal{F}}(m, L)$ and $M_3, M_4 \in \mathcal{M}_{\mathcal{A}}(\underline{\sigma}, \overline{\sigma})$ such that

$$0 \succeq \begin{bmatrix} \mathbf{A}^{\mathsf{T}} P \mathbf{A} - \rho^2 P & \mathbf{A}^{\mathsf{T}} P \mathbf{B} \\ \mathbf{B}^{\mathsf{T}} P \mathbf{A} & \mathbf{B}^{\mathsf{T}} P \mathbf{B} \end{bmatrix} + \begin{bmatrix} \mathbf{Y_f} \\ \mathbf{U_f} \end{bmatrix}^{\mathsf{T}} M_1 \begin{bmatrix} \mathbf{Y_f} \\ \mathbf{U_f} \end{bmatrix}$$

$$+ \begin{bmatrix} \mathbf{Y_A} \\ \mathbf{U_A} \end{bmatrix}^{\mathsf{T}} M_3 \begin{bmatrix} \mathbf{Y_A} \\ \mathbf{U_A} \end{bmatrix} \tag{16a}$$

and

$$0 \preceq \begin{bmatrix} P - I & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \mathbf{Y_f} \\ \mathbf{U_f} \end{bmatrix}^{\mathsf{T}} M_2 \begin{bmatrix} \mathbf{Y_f} \\ \mathbf{U_f} \end{bmatrix}$$

$$+ \begin{bmatrix} \mathbf{Y_A} \\ \mathbf{U_A} \end{bmatrix}^{\mathsf{T}} M_4 \begin{bmatrix} \mathbf{Y_A} \\ \mathbf{U_A} \end{bmatrix}, \tag{16b}$$

then the primal-dual iterations from (3a) converge linearly with rate $O(\rho^k)$ for all objective functions $f \in \mathcal{F}(m, L)$ and all constraint matrices $A \in \mathcal{A}(\underline{\sigma}, \overline{\sigma})$.

Proof. Suppose the LMI (16) is feasible, and consider a trajectory of the primal-dual algorithm (3a). Let $\boldsymbol{\xi}_k$ denote the state of the lifted system $\mathbf{G}$. For each iterate, let a tilde denote the iterate shifted by the fixed point of the system. Now multiply the LMI in (16a) on the right and left by the lifted state and inputs $(\tilde{\boldsymbol{\xi}}_k, \tilde{u}_k^1, \tilde{u}_k^2, \tilde{u}_k^3, \tilde{u}_k^4)$ and its transpose and use the fact that the inequalities (12) and (14) are nonnegative when $M_1, M_2 \in \mathcal{M}_{\mathcal{F}}(m, L)$ and $M_3, M_4 \in \mathcal{M}_{\mathcal{A}}(\underline{\sigma}, \overline{\sigma})$. This produces the inequality $V(\tilde{\boldsymbol{\xi}}_{k+1}) \leq \rho^2 V(\tilde{\boldsymbol{\xi}}_k)$. Likewise, from the LMI (16b), we obtain the inequality $V(\tilde{\boldsymbol{\xi}}_k) \geq \|\tilde{\boldsymbol{\xi}}_k\|^2$. Since the state $\tilde{\xi}_k$ of the original system G is contained in the lifted system, we have that $\|\tilde{\boldsymbol{\xi}}_k\|^2 \geq \|\tilde{\xi}_k\|^2$. Chaining all of these inequalities together gives

$$\|\tilde{\xi}_k\|^2 \leq \|\tilde{\boldsymbol{\xi}}_k\|^2 \leq V(\tilde{\boldsymbol{\xi}}_k) \leq \ldots \leq \rho^{2k} V(\tilde{\boldsymbol{\xi}}_0).$$

Taking the square root gives $\|\tilde{\xi}_k\| \leq c\,\rho^k$ for some $c > 0$, so the iterates converge to the optimizer linearly with rate $O(\rho^k)$. ∎

4 Results

We now compare our analysis with the results from the literature described in Section 2.1. In each case, we choose the objective function condition number $\kappa(f) = 2$ and the Lagrangian augmentation parameter $\mu = 0$. We emphasize that our analysis applies to any values of these parameters, but we choose these values to be able to compare with known bounds.
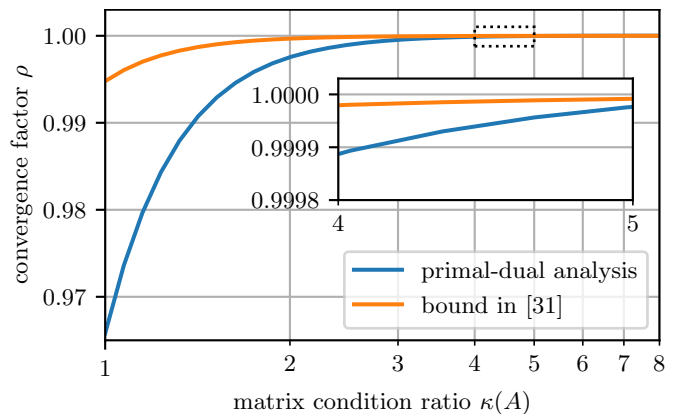

Fig. 1: Comparison with [31]. The algorithm parameters are those in (4) with extrapolation parameter $\gamma = 0$.

4

We use the step size selections from (4) and (6); Figures 1 and 2 plot the convergence factor ρ obtained[3] from our analysis in Theorem 1 along with the corresponding bound from the literature as a function of the matrix condition number $\kappa(A)$. In all cases, our approach certifies a faster rate of convergence. Note that the algorithm in Figure 1 does not use extrapolation, while the algorithm in Figure 2 does use extrapolation, and achieves a much faster convergence rate as a result.
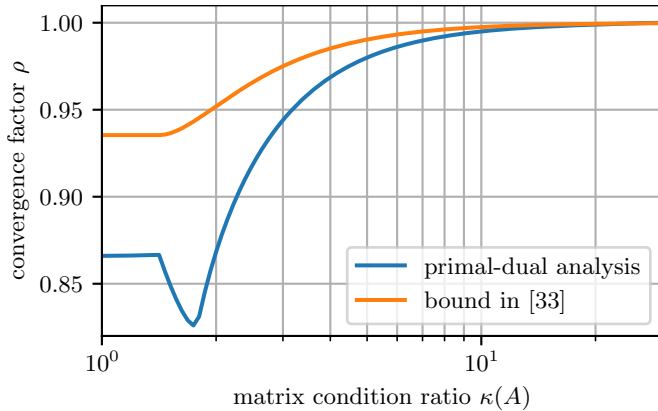

Fig. 2: Comparison with [33]. The algorithm parameters are those in (6) with extrapolation parameter $\gamma = 1$.

The guaranteed convergence factor produced by our approach with step size selection (6) is a non-monotonic function of $\kappa(A)$ in Figure 2. For any *fixed* algorithm, the convergence rate must increase monotonically with the condition numbers $\kappa(f)$ and $\kappa(A)$. However, the step size selection (6) is a function of $\kappa(A)$, and thus the algorithm used is varying at each point on this curve. This highlights that the bounds from the literature are conservative, and suggests that further improvements in the worst-case convergence rate could be obtained by optimizing the step-sizes subject to feasibility of our LMI.

5   Conclusion

Focusing on first-order primal-dual methods for linearly constrained convex optimization, we proposed a systematic method to search for a Lyapunov function that certifies a worst-case convergence rate of the algorithm across all smooth strongly convex functions and all constraint matrices with bounded singular values. Our analysis applies to a range of primal-dual algorithms, including those with extrapolation and Lagrangian augmentation. We compared our numerical results with two bounds from the literature, and our analysis yields better bounds in each case. Future work includes finding algorithm parameters that optimize the rates obtained from our analysis, improving the analysis by leveraging recent interpolation conditions for linear operators [35], and extending the approach to more general operator splitting methods [7].

---

[3]A bisection search was applied to determine the smallest ρ for which the LMIs were feasible.

References

[1] T. Kose, "Solutions of saddle value problems by differential equations," *Econometrica*, vol. 24, no. 1, pp. 59–70, 1956.

[2] K. Arrow, L. Hurwicz, and H. Uzawa, *Studies in linear and non-linear programming.* Stanford University Press, 2006.

[3] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer, 2006.

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers.* Foundations and Trends in Machine Learning, 2010, vol. 3.

[5] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Proc. Mag.*, vol. 32, no. 6, pp. 31–54, 2015.

[6] P. L. Combettes and J.-C. Pesquet, "Fixed point strategies in data science," *IEEE Trans. Signal Proc.*, vol. 69, pp. 3878–3905, 2021.

[7] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, "Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists," *SIAM Review*, 2023, to appear.

[8] D. Feijer and F. Paganini, "Stability of primal–dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.

[9] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal–dual dynamics," *IFAC Syst & Control L*, vol. 87, pp. 10–15, 2016.

[10] J. W. Simpson-Porco, "Input/output analysis of primal-dual gradient algorithms," in *Allerton Conf on Comm, Ctrl & Comp*, Monticello, IL, USA, Sep. 2016, pp. 219–224.

[11] A. Cherukuri, E. Mallada, S. Low, and J. Cortés, "The role of convexity in saddle-point dynamics: Lyapunov function and robustness," *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2449–2464, 2017.

[12] J. W. Simpson-Porco, B. K. Poolla, N. Monshizadeh, and F. Dörfler, "Input-output performance of linear-quadratic saddle-point algorithms with application to distributed resource allocation problems," *IEEE Trans. Autom. Control*, vol. 65, no. 5, pp. 2032–2045, 2019.

[13] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "The proximal augmented lagrangian method for nonsmooth composite optimization," *IEEE Trans. Autom. Control*, vol. 64, no. 7, pp. 2861–2868, 2019.

[14] D. Ding and M. R. Jovanović, "Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented lagrangian," in *Proc. ACC*, 2019, pp. 3414–3419.

[15] ——, "Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented lagrangian: A lyapunov-based approach," in *Proc. IEEE CDC*, 2020, pp. 4836–4841.

[16] X. Chen and N. Li, "Exponential stability of primal-dual gradient dynamics with non-strong convexity," in *Proc. ACC*, 2020, pp. 1612–1618.

[17] N. Li, L. Chen, C. Zhao, and S. H. Low, "Connecting automatic generation control and economic dispatch from an optimization view," in *Proc. ACC*, 2014, pp. 735–740.

[18] A. Cherukuri and J. Cortés, "Initialization-free distributed coordination for economic dispatch under varying loads and generator commitment," *Automatica*, vol. 74, pp. 183–193, 2016.

[19] T. Stegink, C. D. Persis, and A. van der Schaft, "A unifying energy-based approach to stability of power grids with market dynamics," *IEEE Trans. Autom. Control*, vol. 62, no. 6, pp. 2612–2622, 2017.

[20] J. W. Simpson-Porco, B. K. Poolla, N. Monshizadeh, and F. Dörfler, "Quadratic performance of primal-dual methods with application to secondary frequency control of power systems," in *Proc. IEEE CDC*, 2016, pp. 1840–1845.

[21] A. Bernstein and E. Dall'Anese, "Real-time feedback-based optimization of distribution grids: A unified approach," *IEEE Trans. Control Net. Syst.*, vol. 6, no. 3, pp. 1197–1209, 2019.

[22] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM J Optimization*, vol. 26, no. 1, pp. 57–95, 2016.

[23] S. Michalowsky, C. Scherer, and C. Ebenbauer, "Robust and structure exploiting optimisation algorithms: an integral quadratic constraint approach," *Int J Control*, vol. 94, no. 11, pp. 2956–2979, 2021.

[24] A. B. Taylor, J. M. Hendrickx, and F. Glineur, "Exact worst-case performance of first-order methods for composite convex optimization," *SIAM J Optimization*, vol. 27, no. 3, pp. 1283–1313, 2017.

[25] B. Van Scoy and L. Lessard, "Absolute stability via lifting and interpolation," in *Proc. IEEE CDC*, 2022, pp. 6217–6223.

[26] L. Lessard, "The analysis of optimization algorithms: A dissipativity approach," *IEEE Control Syst. Mag.*, vol. 42, no. 3, pp. 58–72, 2022.

[27] B. Hu and L. Lessard, "Dissipativity theory for Nesterov's accelerated method," in *International Conference on Machine Learning*, 2017, pp. 1549–1557.

[28] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. Jordan, "A general analysis of the convergence of ADMM," in *International Conference on Machine Learning*, 2015, pp. 343–352.

[29] G. Zhang, Y. Wang, L. Lessard, and R. B. Grosse, "Near-optimal local convergence of alternating gradient descent-ascent for minimax optimization," in *International Conference on Artificial Intelligence and Statistics*, vol. 151, 2022, pp. 7659–7679.

[30] G. Zhang, X. Bao, L. Lessard, and R. Grosse, "A unified analysis of first-order methods for smooth games via integral quadratic constraints," *The Journal of Machine Learning Research*, vol. 22, no. 1, pp. 4648–4686, 2021.

[31] S. S. Du and W. Hu, "Linear convergence of the primal-dual gradient method for convex-concave saddle point problems without strong convexity," in *International Conference on Artificial Intelligence and Statistics*, vol. 89, 2019, pp. 196–205.

[32] R. T. Rockafellar, *Convex Analysis*. Princteon, NJ: Princeton University Press, 1996.

[33] S. A. Alghunaim and A. H. Sayed, "Linear convergence of primal-dual gradient methods and their performance in distributed optimization," *Automatica*, vol. 117, p. 109003, 2020.

[34] C. Scherer and S. Weiland, *Linear Matrix Inequalites in Control*, 2015. [Online]. Available: https://www.imng.uni-stuttgart.de/mst/files/LectureNotes.pdf

[35] N. Bousselmi, J. M. Hendrickx, and F. Glineur, "Interpolation conditions for linear operators and applications to performance estimation problems," *arXiv:2302.08781*, 2023.

A   Proof of Proposition 1

Let y_i, u_i, f_i for $i \in [\ell] = \{1, \ldots, \ell\}$ denote a sequence of ℓ iterates. From [24], these points are interpolable by an L-smooth and m-strongly convex function if and only if

$$q_{ij} := \begin{bmatrix} y_i \\ y_j \\ u_i \\ u_j \end{bmatrix}^\mathsf{T} \underbrace{\begin{bmatrix} -mL & mL & m & -L \\ mL & -mL & -m & L \\ m & -m & -1 & 1 \\ -L & L & 1 & -1 \end{bmatrix}}_{H} \begin{bmatrix} y_i \\ y_j \\ u_i \\ u_j \end{bmatrix} + h^\mathsf{T} \begin{bmatrix} f_i \\ f_j \end{bmatrix} \geq 0$$

for all $i, j \in [\ell]$ where $h = 2(L-m) \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. In terms of the stacked vectors u, y, f,

$$q_{ij} = \begin{bmatrix} y \\ u \end{bmatrix}^\mathsf{T} H_{ij} \begin{bmatrix} u \\ u \end{bmatrix} + h_{ij}^\mathsf{T} f$$

where $h_{ij} = \begin{bmatrix} e_i & e_j \end{bmatrix} h = 2(L-m)(e_i - e_j)$ and H_{ij} is the matrix

$$\begin{bmatrix} e_i^\mathsf{T} & 0_\ell^\mathsf{T} \\ e_j^\mathsf{T} & 0_\ell^\mathsf{T} \\ 0_\ell^\mathsf{T} & e_i^\mathsf{T} \\ 0_\ell^\mathsf{T} & e_j^\mathsf{T} \end{bmatrix}^\mathsf{T} H \begin{bmatrix} e_i^\mathsf{T} & 0_\ell^\mathsf{T} \\ e_j^\mathsf{T} & 0_\ell^\mathsf{T} \\ 0_\ell^\mathsf{T} & e_i^\mathsf{T} \\ 0_\ell^\mathsf{T} & e_j^\mathsf{T} \end{bmatrix} = \begin{bmatrix} -mL(e_i-e_j)(e_i-e_j)^\mathsf{T} & (e_i-e_j)(me_i-Le_j)^\mathsf{T} \\ (me_i-Le_j)(e_i-e_j)^\mathsf{T} & -(e_i-e_j)(e_i-e_j)^\mathsf{T} \end{bmatrix}$$

where e_i is the ith unit vector in $\mathbb{R}^\ell$. Taking a nonnegative linear combination of the inequalities $q_{ij} \geq 0$, we obtain $Q = \sum_{i \neq j} \lambda_{ij} q_{ij} \geq 0$ for all coefficients $\lambda_{ij} \geq 0$ with $\lambda_{ii} = 0$ for all i. With $\Lambda \in \mathbb{R}_{\geq 0}^{\ell \times \ell}$ the matrix with elements λ_{ij}, straightforward but tedious algebra establishes that

$$Q = \begin{bmatrix} y \\ u \end{bmatrix}^\mathsf{T} \begin{bmatrix} (-mL)R & \frac{m+L}{2}R + \frac{m-L}{2}T \\ \star & -R \end{bmatrix} \begin{bmatrix} y \\ u \end{bmatrix} + (L-m)\mathbf{1}^\mathsf{T}(T - T^\mathsf{T})f,$$

where the $\star$ block can be inferred from symmetry and

$$R := \text{diag}((\Lambda + \Lambda^\mathsf{T})\mathbf{1}) - (\Lambda + \Lambda^\mathsf{T}),$$
$$T := \text{diag}((\Lambda - \Lambda^\mathsf{T})\mathbf{1}) + (\Lambda - \Lambda^\mathsf{T}).$$

Note that R is symmetric doubly hyperdominant with zero row sums, and R and T are independent, as they depend on the symmetric and skew-symmetric parts of Λ, respectively. Now define the skew symmetric matrix $S := \frac{m-L}{2}(\Lambda - \Lambda^\mathsf{T})$ and note that $\frac{m-L}{2}T = \text{diag}(S\mathbf{1}) + S$ and $\frac{m-L}{2}T^\mathsf{T} = \text{diag}(S\mathbf{1}) - S = \text{diag}(S\mathbf{1}) + S^\mathsf{T}$. Then the nonnegative quantity Q is

$$\begin{bmatrix} y \\ u \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} -2mL & m+L \\ m+L & -2 \end{bmatrix} \otimes R + \begin{bmatrix} \mathbf{0} & \text{diag}(S\mathbf{1})+S \\ \star & \mathbf{0} \end{bmatrix} \right) \begin{bmatrix} y \\ u \end{bmatrix} - 4(S\mathbf{1})^\mathsf{T}f,$$

which is of the form (12) if and only if $S\mathbf{1} = \mathbf{0}$. ∎

B   Proof of Proposition 2

Consider a matrix M of this form, and define the matrices

$$M_0 = \begin{bmatrix} -2\,\underline{\sigma}^2\,\overline{\sigma}^2 & \overline{\sigma}^2 + \underline{\sigma}^2 \\ \overline{\sigma}^2 + \underline{\sigma}^2 & -2 \end{bmatrix} \quad \text{and} \quad M_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

so that the multiplier is $M = M_0 \otimes R + M_1 \otimes S$. We first write the inequality (14) as

$$0 \leq \begin{bmatrix} Y \\ U \end{bmatrix}^\mathsf{T} M \begin{bmatrix} Y \\ U \end{bmatrix} \quad \text{where} \quad Y = \begin{bmatrix} Y_k^3 \\ Y_k^4 \end{bmatrix}, \quad U = \begin{bmatrix} U_k^3 \\ U_k^4 \end{bmatrix}.$$

From the feedback (8), we have that $U = (I_{2\ell} \otimes \Sigma^2)Y$, so the matrix $YU^\mathsf{T} = UY^\mathsf{T}$ is symmetric. For the term $M_1 \otimes S$ of the multiplier, the quadratic form is

$$\begin{bmatrix} Y \\ U \end{bmatrix}^\mathsf{T} (M_1 \otimes S) \begin{bmatrix} Y \\ U \end{bmatrix} = Y^\mathsf{T} SU - U^\mathsf{T} SY$$
$$= \text{tr}\big(S(UY^\mathsf{T} - YU^\mathsf{T})\big) = 0.$$

Therefore, this term does not affect the inequality. Without loss of generality, we can take $M_1 \otimes S$ to be symmetric, in which case S is skew-symmetric. For the term $M_0 \otimes R$ of the multiplier, the quadratic form is

$$\begin{bmatrix} Y \\ U \end{bmatrix}^\mathsf{T} (M_0 \otimes R) \begin{bmatrix} Y \\ U \end{bmatrix} = (U - \underline{\sigma}^2 Y)^\mathsf{T} R (\overline{\sigma}^2 Y - U)$$
$$= \big((I \otimes \Sigma^2) - \underline{\sigma}^2 I\big) Y^\mathsf{T} RY \big(\overline{\sigma}^2 - (I \otimes \Sigma^2)\big).$$

This quantity is nonnegative since Σ is a diagonal matrix of singular values in the interval $[\underline{\sigma}, \overline{\sigma}]$ and $Y^\mathsf{T} RY$ is positive semidefinite. Therefore, the inequality (14) holds for iterates that satisfy (8) for all $M \in \mathcal{M}_\mathcal{A}(\underline{\sigma}, \overline{\sigma})$. ∎