A Contraction Theory Approach to Optimization Algorithms from Acceleration Flows

Pedro Cisneros-Velarde
University of Illinois at Urbana-Champaign
Department of Computer Science
pacisne@gmail.com

Francesco Bullo
Univerisy of California, Santa Barbara
Department of Mechanical Engineering
bullo@ucsb.edu

Abstract

Much recent interest has focused on the design of optimization algorithms from the discretization of an associated optimization flow, i.e., a system of differential equations (ODEs) whose trajectories solve an associated optimization problem. Such a design approach poses an important problem: how to find a principled methodology to design and discretize appropriate ODEs. This paper aims to provide a solution to this problem through the use of contraction theory. We first introduce general mathematical results that explain how contraction theory guarantees the stability of the implicit and explicit Euler integration methods. Then, we propose a novel system of ODEs, namely the Accelerated-Contracting-Nesterov flow, and use contraction theory to establish it is an optimization flow with exponential convergence rate, from which the linear convergence rate of its associated optimization algorithm is immediately established. Remarkably, a simple explicit Euler discretization of this flow corresponds to the Nesterov acceleration method. Finally, we present how our approach leads to performance guarantees in the design of optimization algorithms for time-varying optimization problems.

1 INTRODUCTION

Problem Statement and Motivation There has been a recent interest in studying systems of ODEs that solve an optimization problem – also known as *optimization flows* – with the understanding that their study can lead to the analysis and design of discrete-time solvers of optimization problems also known as *optimization algorithms*. This interest is motivated by the fact that analyzing a system of ODEs can be much simpler than analyzing a discrete system. Indeed, the ambitious goal of this research area is to find a general theory mapping properties of ODEs into corresponding properties for discrete updates" — as quoted from the seminal work (Su et al., 2016). Our paper aims to provide a solution to this problem. Ideally, the desired pipeline is to first design an optimization flow – using all the machinery of dynamical systems analysis – with good stability and convergence properties, and then formulate a principled way of guaranteeing such good properties translate to its associated optimization algorithm through discretization.

A first problem in the literature is that the analysis of the optimization algorithm is commonly done separately or independently from the analysis of its associated optimization flow (e.g., see (Wibisono et al., 2016; Wilson et al., 2018; Shi et al., 2021; Zhang et al., 2018; Shi et al., 2019; Muehlebach and Jordan, 2019)), instead of the former analysis following directly as a consequence of the latter one. For example, separate Lyapunov analyses have been made for optimization flows and their associated algorithms. This problem diminishes one of the very first motivations of analyzing a system of ODEs, namely, that its analysis should directly establish properties of its associated discretization.

Moreover, it is highly important to formulate an optimization flow with good convergence properties so that its discretization may also lead to an optimization algorithm maintaining such good convergence properties, without becoming numerically unstable (Wibisono et al., 2016; Shi et al., 2019). A long-standing objective is to find *rate-matching* discretizations, i.e., discretizations that preserve the good

Proceedings of the 25[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151.

convergence rate performance from the optimization flow to the optimization algorithm.

Our paper aims to propose a solution to the aforementioned two problems by using *contraction theory* and by proposing the Accelerated-Contracting-Nesterov optimization flow, respectively. Contraction theory provides a principled approach for directly establishing the convergence properties of the optimization algorithm directly from the analysis of its associated optimization flow. For example, our approach directly translates the exponential convergence rate of the proposed optimization flow to a linear convergence rate for the optimization algorithm.

We also mention that, to the best of our knowledge, no previous work in the literature relevant to the design of (discrete-time) optimization algorithms from optimization flows has considered the case where the objective function associated to the optimization problem is *time-varying*, i.e., where the optimal solutions vary through time and form a trajectory. In this setting, the objective is to design an optimization flow and optimization algorithms that are able to track the time-varying optimal solution up to some bounded error. In this paper, we use contraction theory to directly establish the tracking performance of our proposed optimization flow and its associated optimization algorithms. Time-varying optimization has found diverse applications in machine learning, signal processing, robotics, etc.; e.g., see (Simonetto et al., 2020) and the references therein. We must mention the existence of *prediction-correction methods* (Bastianello et al., 2020) which consist on sampling a continuous-time time-varying optimization problem on fixed intervals of time and then finding approximate solutions to the full optimization problem within the sampling intervals. These methods are not within our problem's scope because we are interested in obtaining optimization algorithms whose dynamics directly track the solution trajectory of the problem.

Literature Review There is a large growing literature on the analysis and design of optimization algorithms based on ODEs, so we limit ourselves to mentioning some representative works. The seminal works (Su et al., 2016; Shi et al., 2021) propose systems of ODEs based on continuous approximations or versions of the Heavy-ball and the Nesterov acceleration methods, providing a thorough analysis of these optimization flows with the purpose of gaining more insights about the momentum or acceleration phenomenon in their discrete-time counterparts. Examples of works that design optimization algorithms by direct discretization of optimization flows include: the use of optimization flows derived from variational approaches (Wibisono et al., 2016), the use of both explicit and implicit Euler discretizations (Wilson et al., 2018; Shi et al., 2019), the use of Runge-Kutta integrators (Zhang et al., 2018), the use of semi-implicit Euler integration Muehlebach and Jordan (2019), the use of opportunistic state triggering (Vaquero and Cortes, 2019), the use of symplectic methods (França et al., 2020; Shi et al., 2019). Most of the previous works are based on Lyapunov analysis, as well as on the discretization of the ODEs proposed in the works (Su et al., 2016; Shi et al., 2021) or a generalization of them.

To the best of our knowledge, only the approaches based on symplectic integration and opportunistic state triggering allows for a principled way of discretizing optimization flows while preserving good convergence properties and without the need to do an independent analysis of the discretizations. In this way, these works are closer in spirit to the approach in this paper; however, in contrast to them, our paper uses the classic implicit and explicit Euler integration methods.

Contraction theory is a mathematical tool that analyzes the *incremental stability* of dynamical systems, i.e, whether two different trajectories of a dynamical system converge or at least do not diverge towards each other, and has a long history of development in the control theory community (Coppel, 1965; Vidyasagar, 2002; Lohmiller and Slotine, 1998; Manchester and Slotine, 2017). An introduction and survey can be found in (Aminzare and Sontag, 2014). Contraction theory has been used in the analysis of optimization flows associated to certain classes of constrained optimization problems (Nguyen et al., 2018; Cisneros-Velarde et al., 2021), distributed systems (Boffi and Slotine, 2020), and connections with gradient flows has been made (Wensing and Slotine, 2020). Parallel to this development, an important concept that implies contraction, known as the one-sided Lipschitz condition, has been used in numerical analysis (Hairer et al., 1993). Contraction theory has recently found applications in neural networks (Revay and Manchester, 2020).

Contributions Our contributions are as follows:

- We establish how the contraction analysis of a dynamical system immediately characterizes the stability of its associated implicit and explicit Euler integrations. We consider the case where the dynamical system is contracting and time-varying without necessarily possessing the same equilibrium point at all times; the time-invariant case follows immediately. In the case of the implicit Euler method, compared to the work (Desoer and Haneda, 1972), our results: 1) do not assume the vector field to be continuously differentiable; 2) consider the case where the contraction rate can be time-varying; and 3) use the one-sided Lipschitz condition. In the case of the explicit Euler method, our results generalize the classic forward step method from the theory of monotone operators (Ryu and Boyd, 2016) because we consider the case of time-varying contraction rates.

For the rest of contributions, we remark that we consider a differentiable, strongly convex objective function with Lipschitz smoothness — a commonly used class of functions for the theoretical analysis of optimization flows and their discretization schemes, e.g., (Su et al., 2016; Wilson et al., 2018; Shi et al., 2021; Vaquero and Cortes, 2019; Shi et al., 2019; Muehlebach and Jordan, 2019), and even for other state-of-the-art solvers for time-varying optimization, e.g., (Bastianello et al., 2020).

- We formulate a system of ODEs called the *ACcelerated-COntracting-NESTerov* (ACCONEST) flow, and we prove it is an optimization flow with exponential convergence using contraction theory, from which we directly prove that both its implicit and explicit Euler integrations have linear convergence. For the implicit integration, rate-matching and acceleration is guaranteed. For the appropriate integration step, the explicit Euler integration of this system is the Nesterov acceleration method and thus has rate-matching and acceleration too.
- Our contraction analysis for the ACCONEST flow implies the existence of a simple quadratic Lyapunov function for the optimization flow, as opposed to other more complex Lyapunov functions or energy functionals used in previous works, e.g., see (Su et al., 2016; Shi et al., 2019). Proving that the optimization flow is contracting provides additional robustness results that a Lyapunov analysis does not generally imply, such as strong input-to-state stability guarantees and finite input-state gains (Davydov et al., 2021). Moreover, a contracting system is guaranteed to have fast correction after transient perturbations to the trajectory of the solution, since initial conditions are forgotten.
- Finally, we extend our analysis to the case of time-varying optimization, i.e., where the objective function varies through time. Contraction theory provides guarantees for the tracking of the time-varying optimization solution based upon the analysis done for the time-invariant case. We show that the ACCONEST flow and its discretizations have a tracking error that is uniformly ultimately bounded.

Paper Organization Section 2 has notation and preliminary concepts. Section 3 has results on contraction theory and stability of discretizations. Section 4 uses contraction theory to analyze a proposed optimization flow and its associated optimization algorithms. Section 5 analyzes the time-varying case. Section 6 is the conclusion.

The proofs for the results in Section 3, Section 4, and Section 5 are found in the supplementary material.

2 PRELIMINARIES AND NOTATION

2.1 Notation and Definitions

Consider $A \in \mathbb{R}^{n\times n}$. If A only has real eigenvalues, let $\lambda_{\min}(A)$ be its minimum eigenvalue and $\lambda_{\max}(A)$ its maximum one. Let I_n be the $n \times n$ identity matrix. Let $\|\cdot\|_p$ denote the ℓ_p-norm, and when the argument of a norm is a matrix, we refer to its respective induced norm. The matrix measure associated to norm $\|\cdot\|$ is $\mu(A) = \lim_{h\to 0^+} \frac{\|I_n+hA\|-1}{h}$. Given invertible $Q \in \mathbb{R}^{n\times n}$, let $\|\cdot\|_{2,Q}$ be the weighted ℓ_2-norm $\|x\|_{2,Q} = \|Qx\|_2$, $x \in \mathbb{R}^n$, and whose associated matrix measure is $\mu_{2,Q}(A) = \mu_2(QAQ^{-1})$.

Let 1_n and 0_n be the all-ones and all-zeros column vector with n entries respectively. Given $x_i \in \mathbb{R}^{k_i}$, let $(x_1,\dots,x_N) = \begin{bmatrix} x_1^\top & \dots & x_N^\top \end{bmatrix}$.

Consider a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$. We say f is *L-smooth* if $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x-y\|_2$, $L > 0$, for any $x, y \in \mathbb{R}^n$; and *μ-strongly convex* if $\mu\|x-y\|_2^2 \leq (\nabla f(x) - \nabla f(y))^\top (x-y)$, $\mu > 0$, for any $x, y \in \mathbb{R}^n$. Let $\mathcal{S}^1_{\mu,L}$ be the set of continuously differentiable functions on $\mathbb{R}^n$ that are L-smooth and μ-strongly convex, and denote by $\kappa := \frac{L}{\mu}$ the condition number of any function belonging to this set.

Consider a dynamical system $\dot{x} = g(x,t)$ with *vector field* $g : \mathbb{R}^n \times [t_0,\infty) \to \mathbb{R}^n$, $x \in \mathbb{R}^n$, $t \geq t_0$. We assume that a solution exists under any initial condition. Let $t \mapsto \phi(t, t_0, x_0)$ be the trajectory of the system starting from $x_0 \in \mathbb{R}^n$ at time $t_0 \geq 0$. We say g is *uniformly ℓ-Lipschitz continuous* with respect to norm $\|\cdot\|$ if $\|g(x,t) - g(y,t)\| \leq \ell\|x-y\|$, $\ell > 0$, for any $x, y \in \mathbb{R}^n$, and for every $t > 0$. The *implicit Euler* discretization or method with constant discretization or integration step-size $h > 0$ of this dynamical system with $t_0 = 0$ is: $y_{k+1} = y_k + hg(y_{k+1}, (k+1)h)$, $k = 0, 1, \dots$, with given $y_0 \in \mathbb{R}^n$. The *explicit Euler* discretization or method with constant discretization or integration step-size $h > 0$ of this dynamical system with $t_0 = 0$ is: $y_{k+1} = y_k + hg(y_k, kh)$, $k = 0, 1, \dots$, with given $y_0 \in \mathbb{R}^n$.

2.2 Review of Basic Results on Contraction Theory

The following results will be useful throughout the paper; e.g., see references (Ladas and Lakshmikantham, 1972; Davydov et al., 2021).

Theorem 2.1 (Characterizations of contracting systems)**.** *Consider the dynamical system $\dot{x} = g(x,t)$ with $x \in \mathbb{R}^n$. Pick a symmetric positive-definite $P \in \mathbb{R}^{n\times n}$ and an integrable function $\bar{\gamma} : [t_0, \infty) \to \mathbb{R}$. Then, the following statements are equivalent*

1. $\|\phi(t,t_0,x_0)-\phi(t,t_0,y_0)\|_{2,P^{1/2}} \leq e^{-\int_{t_0}^t \bar{\gamma}(s)ds}\|x_0-y_0\|_{2,P^{1/2}}$;
2. *the* one-sided Lipschitz condition $(y-x)^\top P(g(x,t)-g(y,t)) \leq -\bar{\gamma}(t)\|x-y\|^2_{2,P^{1/2}}$ *holds for every* $x,y \in \mathbb{R}^n$ *and* $t \geq t_0$.

If f is continuously differentiable, then the previous statements are also equivalent to

3. $\mu_{2,P^{1/2}}(Dg(x,t)) \leq -\bar{\gamma}(t)$ *for every* $x \in \mathbb{R}^n$ *and* $t \geq t_0$; *where* $D(\cdot,t)$ *is the Jacobian of* $g(\cdot,t)$.

Any dynamical system that satisfies any of the conditions in Theorem 2.1 with $\int_{t_0}^t \bar{\gamma}(s)ds > 0$ for any $t \geq t_0$ is said to be contracting (with respect to $\|\cdot\|_{2,P^{1/2}}$) or have the *contraction* property, and $\bar{\gamma}(t)$ is the contraction rate. The contraction rate is time-invariant when $\bar{\gamma}(t) \equiv \gamma$.

The central and most challenging part of determining if a system is contracting is to find an adequate symmetric positive-definite matrix $P \in \mathbb{R}^{n\times n}$ that establishes its contraction respect to $\|\cdot\|_{2,P^{1/2}}$.

Theorem 2.2 (Characterization for homogeneous systems)**.** *Consider the homogeneous or time-invariant dynamical system $\dot{x} = g(x)$ with $x \in \mathbb{R}^n$. Pick a symmetric positive-definite $P \in \mathbb{R}^{n\times n}$. If the system is contracting with contraction rate $\gamma > 0$ and respect to $\|\cdot\|_{2,P^{1/2}}$, then*

1. *there exists a unique equilibrium point* x^*;
2. x^* *is exponentially globally stable with Lyapunov functions* $V(x) = \|x-x^*\|^2_{2,P^{1/2}}$ *and* $V(x) = \|g(x)\|^2_{2,P^{1/2}}$; *and*
3. *the exponential convergence rate is* γ.

In light of Theorem 2.1, statements 2 and 3 of Theorem 2.2 can be equivalently expressed as $\|\phi(t,t_0,x_0)-x^*\| \leq e^{-\gamma(t-t_0)}\|x_0-x^*\|$ for any $x_0 \in \mathbb{R}^n$.

3 CONTRACTION THEORY AND STABILITY OF EULER DISCRETIZATIONS

The next result establishes how the contraction analysis of a dynamical system can immediately characterize the stability of its associated implicit and implicit Euler integration.

Theorem 3.1 (Stability of the implicit Euler integration)**.** *Consider the system $\dot{x} = g(x,t)$, $x \in \mathbb{R}^n$, $t \geq 0$, and that it is contracting with contraction rate $\bar{\gamma}(t) > 0$ for all $t \geq 0$ and respect to some appropriate norm $\|\cdot\|_{2,P^{1/2}}$, $P \in \mathbb{R}^{n\times n}$. Let (y_k) be a sequence generated by the implicit Euler discretization with constant integration step-size $h > 0$.*

1. *If there exists x^* such that $g(x^*,t) = 0_n$ for all $t \geq 0$, then*

$$\|y_k - x^*\|_{2,P^{1/2}} \leq \left(\prod_{m=1}^k (1+h\bar{\gamma}(mh))^{-1}\right)\|y_0 - x^*\|_{2,P^{1/2}}, \quad (1)$$

for $k = 1,2,\ldots$; and if additionally the contraction rate is time-invariant $\bar{\gamma}(t) \equiv \gamma > 0$, then there is linear convergence

$$\|y_k - x^*\|_{2,P^{1/2}} \leq (1+h\gamma)^{-k}\|y_0 - x^*\|_{2,P^{1/2}}. \quad (2)$$

2. *If there exists a curve $t \mapsto x^*(t)$ such that $g(x^*(t),t) = 0_n$ for any $t \geq 0$, then*

$$\begin{aligned}&\|y_k - x^*(kh)\|_{2,P^{1/2}} \\ &\leq \left(\prod_{m=1}^k (1+h\bar{\gamma}(mh))^{-1}\right)\|y_0 - x^*(0)\|_{2,P^{1/2}} \\ &\quad + \sum_{m=1}^k \prod_{r=m}^k (1+h\bar{\gamma}(rh))^{-1} \\ &\qquad \times \|x^*(mh) - x^*((m-1)h)\|_{2,P^{1/2}},\end{aligned} \quad (3)$$

for $k = 1,2,\ldots$; and if additionally the contraction rate is time-invariant $\bar{\gamma}(t) \equiv \gamma > 0$ and $\sup_{k=1,2,\ldots}\|x^*(kh) - x^*((k-1)h)\|_{2,P^{1/2}} \leq \rho$ for some constant $\rho \geq 0$, then

$$\begin{aligned}&\|y_k - x^*(kh)\|_{2,P^{1/2}} \\ &\leq (1+h\gamma)^{-k}\|y_0 - x^*(0)\|_{2,P^{1/2}} + \rho\sum_{m=1}^k (1+h\gamma)^{-m},\end{aligned} \quad (4)$$

and so

$$\limsup_{k\to\infty}\|y_k - x^*(kh)\|_{2,P^{1/2}} \leq \frac{\rho}{h\gamma}. \quad (5)$$

Remark 3.2 (Comparison with the work (Desoer and Haneda, 1972)). *Statement 1 in Theorem 3.1 was proved in (Desoer and Haneda, 1972) for the case where the vector field of the dynamical system is continuously differentiable and using the condition 3 of Theorem 2.1 along with properties of matrix measures. We also remark that Theorem 3.1 considers the case where the contraction rate can be time-varying, unlike (Desoer and Haneda, 1972).*

Remark 3.3 (Rate-matching for the Implicit Euler integration). *Assume the conditions of statement 1 of Theorem 3.1, with the system having a time-invariant contraction rate $\gamma > 0$. Now, observe that $e^{\gamma h} \geq (1+\gamma h) \implies (1+\gamma h)^{-k} \leq e^{-\gamma h k}$ for $k = 1, 2 \dots$. In order words, the linear convergence of the implicit Euler integration is upper bounded by the exponential convergence of its ODE counterpart for the discretized time $t = kh$ (see statement 1 of Theorem 2.1), i.e., there is rate-matching.*

Theorem 3.4 (Stability of the explicit Euler integration). *Consider the system $\dot{x} = g(x,t)$, $x \in \mathbb{R}^n$, $t \geq 0$, with g being uniformly ℓ-Lipschitz continuous, and that it is contracting with contraction rate $\bar{\gamma}(t)$ such that $\inf_{t\geq 0} \bar{\gamma}(t) > 0$ and $\ell > \sup_{t \geq 0} \bar{\gamma}(t)$, both respect to some appropriate norm $\|\cdot\|_{2,P^{1/2}}$, $P \in \mathbb{R}^{n\times n}$. Let (y_k) be a sequence generated by the explicit Euler discretization with constant integration step-size $0 < h < \frac{2}{\ell^2} \inf_{t\geq 0} \bar{\gamma}(t)$.*

1. *If there exists x^* such that $g(x^*, t) = 0_n$ for all $t \geq 0$, then*

$$\|y_k - x^*\|_{2,P^{1/2}} \leq \left(\prod_{m=0}^{k-1} (1 - 2h\bar{\gamma}(mh) + h^2\ell^2)^{1/2} \right) \times \|y_0 - x^*\|_{2,P^{1/2}}, \quad (6)$$

   for $k = 1, 2, \dots$; and if additionally the contraction rate is time-invariant $\bar{\gamma}(t) \equiv \gamma > 0$, then there is linear convergence

$$\|y_k - x^*\|_{2,P^{1/2}} \leq (1 - 2h\gamma + h^2\ell^2)^{k/2} \|y_0 - x^*\|_{2,P^{1/2}}. \quad (7)$$

2. *If there exists a curve $t \mapsto x^*(t)$ such that $g(x^*(t), t) = 0_n$ for any $t \geq 0$, then*

$$\begin{aligned} \|y_k - x^*(kh)\|_{2,P^{1/2}} &\leq \left(\prod_{m=0}^{k-1} (1 - 2h\bar{\gamma}(mh) + h^2\ell^2)^{1/2} \right) \times \|y_0 - x^*(0)\|_{2,P^{1/2}} \\ &+ 1_{\{k>1\}} \sum_{m=1}^{k-1} \prod_{r=m}^{k-1} (1 - 2h\bar{\gamma}(rh) + h^2\ell^2)^{1/2} \times \|x^*(mh) - x^*((m-1)h)\|_{2,P^{1/2}} \\ &+ \|x^*(kh) - x^*((k-1)h)\|_{2,P^{1/2}}, \quad (8) \end{aligned}$$

   for $k = 1, 2, \dots$; where $1_{\{k>1\}} = 1$ if $k > 1$ and $1_{\{k>1\}} = 0$ otherwise. Additionally, if the contraction rate is time-invariant $\bar{\gamma}(t) \equiv \gamma > 0$ and $\sup_{k=1,2,\dots} \|x^*(kh) - x^*((k-1)h)\|_{2,P^{1/2}} \leq \rho$ for some constant $\rho \geq 0$, then

$$\begin{aligned} \|y_k - x^*(kh)\|_{2,P^{1/2}} &\leq (1 - 2h\gamma + h^2\ell^2)^{k/2} \|y_0 - x^*(0)\|_{2,P^{1/2}} \\ &+ \rho \sum_{m=0}^{k-1} (1 - 2h\gamma + h^2\ell^2)^{m/2}, \quad (9) \end{aligned}$$

   for $k = 1, 2, \dots$, and so

$$\limsup_{k\to\infty} \|y_k - x^*(kh)\|_{2,P^{1/2}} \leq \frac{\rho}{1 - (1 - 2h\gamma + h^2\ell^2)^{1/2}}. \quad (10)$$

3. *In all the previous cases where the system has contraction rate γ, the optimal choice for the time-step that gives the fastest convergence rate is $h^* = \frac{\gamma}{\ell^2}$; i.e., in equations (7), (9) and (10), we have $1 - 2h^*\gamma + h^{*2}\ell^2 = 1 - \frac{\gamma^2}{\ell^2}$.*

Remark 3.5 (Connection with monotone operator theory). *Consider the homogeneous dynamical system $\dot{x} = g(x)$, $x \in \mathbb{R}^n$, with time-invariant contraction rate $\gamma > 0$. If the vector field g is seen as an operator (over its argument), then the one-sided Lipschitz condition in item 2 of Theorem 2.1 is equivalent to $-g$ being* strongly monotone *with parameter γ. Therefore, the classic proof for finding zeros of a strongly monotone operator (Ryu and Boyd, 2016) (also called the forward step method) follows as a particular case of the proof of item 1 of Theorem 3.4 for the case of time-invariant contraction rates.*

Remark 3.6 (Improving the contraction rate for Theorem 3.4). *Consider Theorem 2.1 and its assumptions, with the system having a time-invariant contraction rate $\gamma > 0$. A closer look to the proof shows that the* *linear convergence rate depends on the inequality*

$$\begin{aligned}&-2h\gamma\|y_k - x^*\|^2_{2,P^{1/2}} + h^2\|g(y_k, kh)\|^2_{2,P^{1/2}}\\&\qquad\leq (-2h\gamma + h^2\ell^2)\|y_k - x^*\|^2_{2,P^{1/2}}.\end{aligned} \tag{11}$$

The use of the Lipschitz constant ℓ is done in order to provide a general bound for our proof; however, it may be possible to upper bound the left-hand side of (11) *and obtain a less conservative result, and thus a better contraction rate. However, this bound will depend on the particular vector field g.*

4 THE ACCELERATED-CONTRACTING-NESTEROV FLOW AND ITS EULER DISCRETIZATIONS

Consider an objective function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying $f \in \mathcal{S}^1_{\mu,L}$. The works (Su et al., 2016; Shi et al., 2021) propose the following system of ODEs:

$$\begin{aligned}\dot{x}_1 &= c\,x_2\\ \dot{x}_2 &= -a\,x_2 - b\nabla f(x_1)\end{aligned} \tag{12}$$

with $x_1, x_2 \in \mathbb{R}^n$, as an optimization flow related to the heavy-ball method for specific values of the scalars $a, b, c > 0$. Then, the work (Muehlebach and Jordan, 2019) proposed to modify (12) by including a *displaced gradient*:

$$\begin{aligned}\dot{x}_1 &= c\,x_2\\ \dot{x}_2 &= -a\,x_2 - b\nabla f(x_1 + dx_2)\end{aligned} \tag{13}$$

with additional scalar $d > 0$. The state x_1 is the *position* of the system whose trajectory should converge to the minimum of f, and x_2 is its instantaneous velocity. In this paper, for any fixed $t \geq 0$, we interpret $x_1(t) + dx_2(t)$ as the *predicted* position that the system has if it constantly follows the velocity $x_2(t)$ for a period of time of length d. Thus, we can think of system (13) as having its acceleration $\dot{x}_2$ being influenced by the gradient at a predicted position, instead of its current position as in (12).

Now, recall that in a gradient flow the negative gradient affects the direction of the instantaneous velocity, i.e., $\dot{x}_1 = -\nabla f(x_1)$. Then, we propose to include such effect of the gradient in (13) with the expectation that this may improve the convergence rate of the system:

$$\begin{aligned}\dot{x}_1 &= c\,x_2 - e\nabla f(x_1 + dx_2)\\ \dot{x}_2 &= -a\,x_2 - b\nabla f(x_1 + dx_2)\end{aligned} \tag{14}$$

with additional scalar $e > 0$. Systems (12), (13), and (14) have the unique equilibrium point $(x^*, 0_n)$ with $x^* = \arg\min_{z\in\mathbb{R}^n} f(z)$.

Now, we make the change of variables $\bar{x}_1 = x_1$ and $\bar{x}_2 = x_1 + dx_2$, and set $a = \frac{2}{\sqrt{\kappa}+1}$, $b = \frac{\sqrt{\kappa}-1}{2L}$, $c = d = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$ and $e = \frac{1}{L}$. Then, we obtain the *ACcelerated-COntracting-NESTerov* (ACCONEST) flow

$$\begin{aligned}\begin{bmatrix}\dot{\bar{x}}_1\\ \dot{\bar{x}}_2\end{bmatrix} &:= F_{\mathrm{Acc}}(\bar{x}_1, \bar{x}_2)\\ &= \begin{bmatrix}\bar{x}_2 - \bar{x}_1 - \frac{1}{L}\nabla f(\bar{x}_2)\\ \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}(\bar{x}_2 - \bar{x}_1) - \frac{2\sqrt{\kappa}}{(\sqrt{\kappa}+1)L}\nabla f(\bar{x}_2)\end{bmatrix},\end{aligned} \tag{15}$$

which has unique equilibrium point (x^*, x^*); i.e., at equilibrium, the system's current position $\bar{x}_1$ and predicted position $\bar{x}_2$ must be equal to the optimal value.

We now analyze the ACCONEST flow using contraction theory and the results introduced in the previous section.

Theorem 4.1 (Analysis of the ACCONEST flow and its optimization algorithms)**.** *Consider the ACCONEST flow* (15) *with $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{S}^1_{\mu,L}$; and let the optimal value $x^* = \arg\min_{z\in\mathbb{R}^n} f(z)$.*

1. *The ACCONEST flow is contracting with rate $\sqrt{\frac{\mu}{L}}$ and respect to $\|\cdot\|_{2,P^{1/2}}$ with $P = \begin{bmatrix}\gamma\frac{\sqrt{\kappa}}{\sqrt{\kappa}+1} & -1\\ -1 & \frac{\sqrt{\kappa}+1}{\sqrt{\kappa}}\end{bmatrix} \otimes I_n$, $1 < \gamma \leq 1 + \frac{1}{\kappa}$.*
2. *The ACCONEST flow has global exponential convergence to $\begin{bmatrix}1\\1\end{bmatrix} \otimes x^*$.*
3. *Let $((y^{(1)}_k, y^{(2)}_k))$ be the sequence generated by the implicit Euler discretization of the ACCONEST flow* (15) *with integration step-size h. Then, there is linear convergence to the optimal value characterized by*
$$f(y^{(1)}_k) - f(x^*) \leq C\left(1 + h\sqrt{\frac{\mu}{L}}\right)^{-2k} \tag{16}$$
with C being some constant that depends on the initial conditions $y^{(1)}_0, y^{(2)}_0 \in \mathbb{R}^n$.
4. *Let $((y^{(1)}_k, y^{(2)}_k))$ be the sequence generated by the explicit Euler discretization of the ACCONEST flow* (15) *with integration step-size $h^* := \frac{1}{\sqrt{\kappa}(7+5\beta+6\beta^2)}\frac{\lambda_{\min}(P)}{\lambda_{\max}(P)}$, $\beta := \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$. Then, there is linear convergence to the optimal value characterized by*
$$f(y^{(1)}_k) - f(x^*) \leq C\left(1 - h^*\sqrt{\frac{\mu}{L}}\right)^{k} \tag{17}$$
with C being some constant that depends on the initial conditions $y^{(1)}_0, y^{(2)}_0 \in \mathbb{R}^n$.

Remark 4.2 (About Theorem 4.1)**.** *Some remarks are in order:*

- ***Rate-matching and acceleration:*** *Note that, as pointed out in Remark 3.3, we have rate-matching for the implicit Euler integration (starting with step-size equal to one) and, consequently, there is acceleration because the contraction rate of the optimization flow* (15) *is* $\sqrt{\frac{\mu}{L}}$*. Unfortunately, for the explicit Euler case, we could not prove rate-matching in Theorem 4.1; however, the simple proof of Proposition 4.4 shows that rate-matching and acceleration can be established (since* $(1-c)^{-k} \leq e^{-ck}$ *for any* $0 < c < 1$ *and non-negative integer* k*).*
- ***Connection with Lyapunov analysis:*** *As a consequence of Theorem 2.2, the optimization flow* (15) *can have its exponential convergence certified by two simple quadratic Lyapunov functions as stated in Corollary 4.3. This is in contrast to other works that have analyzed optimization flows using more complicated Lyapunov functions, e.g., (Su et al., 2016; Shi et al., 2021; Vaquero and Cortes, 2019; Shi et al., 2019).*

Corollary 4.3 (Quadratic Lyapunov functions for the optimization flow (15))**.** *Consider the ACCONEST flow* (15) *with* $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{S}^1_{\mu,L}$*. Then, the system is an optimization flow whose exponential convergence to the optimization solution with rate* $\sqrt{\frac{\mu}{L}}$ *can be established via the Lyapunov functions* $V(\bar{x}_1, \bar{x}_2) = \|(\bar{x}_1 - x^*, \bar{x}_2 - x^*)^\top\|^2_{2,P^{1/2}}$ *and* $V(\bar{x}_1, \bar{x}_2) = \|F_{\mathrm{Acc}}(\bar{x}_1, \bar{x}_2)\|^2_{2,P^{1/2}}$ *with* $P = \begin{bmatrix} \gamma\frac{\sqrt{\kappa}}{\sqrt{\kappa}+1} & -1 \\ -1 & \frac{\sqrt{\kappa}+1}{\sqrt{\kappa}} \end{bmatrix} \otimes I_n$, $1 < \gamma \leq 1 + \frac{1}{\kappa}$*.*

Proposition 4.4 (Acceleration from the explicit Euler integration of the ACCONEST flow (15))**.** *Consider the ACCONEST flow* (15) *with* $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{S}^1_{\mu,L}$*. Let* $((y^{(1)}_k, y^{(2)}_k))$ *be the sequence generated by its explicit Euler discretization with integration step-size 1. Then, there is linear convergence to the optimal value characterized by*

$$f(y^{(1)}_k) - f(x^*) \leq C \left(1 - \sqrt{\frac{\mu}{L}}\right)^k \tag{18}$$

with C *being some constant that depends on the initial conditions* $y^{(1)}_0, y^{(2)}_0 \in \mathbb{R}^n$*.*

Proof. We compute the update for the explicit Euler integration of the optimization flow (15) with integration step-size 1 which, after some algebraic operations, becomes

$$y^{(1)}_{k+1} = y^{(2)}_k - \frac{1}{L}\nabla f(y^{(2)}_k)$$
$$y^{(2)}_{k+1} = y^{(1)}_{k+1} + \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}(y^{(1)}_{k+1} - y^{(1)}_k)$$

which is the Nesterov acceleration method itself (Nesterov, 2018). Thus, the linear convergence (18) is guaranteed. □

The previous proposition tells us that if we perform an explicit Euler integration to the optimization flow (15) with integration step-size 1, then we obtain the Nesterov acceleration method itself. Thus, we have used contraction theory in Theorem 4.1 to analyze an optimization flow associated to the Nesterov accelerated method (hence the name of Accelerated-Contracting-Nesterov flow). Moreover, this optimization flow has its exponential rate being identical to Nesterov's acceleration rate, which makes it quite different in structure and convergence rate from other flows in the literature (e.g., the ones of the form of (12)), which have not been proved to be contracting.

Remark 4.5 (Comparing the ACCONEST flow to high-resolution ODEs)**.** *For the class of objective functions in* $\mathcal{S}^1_{\mu,L}$*, the works (Su et al., 2016; Shi et al., 2021), and other consecutive ones that are based on the use of their equations such as (Vaquero and Cortes, 2019; Shi et al., 2019), analyze the so-called high-resolution ODEs, which have some specific fixed values for the constants of the system* (12)*. We also remark that the high-resolution ODE associated to the Nesterov acceleration method considers the Hessian* $\nabla^2 f(x_1)$ *in its structure whereas the ACCONEST flow* (15)*, which also corresponds to the Nesterov acceleration method, makes no use of second order information, let alone any twice-differentiability assumption. Lyapunov analysis of this high-resolution ODE and its discretizations has shown that acceleration does not occur under the explicit Euler integration (Shi et al., 2019).*

5 SOLVING TIME-VARYING OPTIMIZATION WITH THE ACCELERATED-CONTRACTING-NESTEROV FLOW AND ITS EULER DISCRETIZATIONS

Assume that the optimization problem is time-varying

$$\min_{z\in\mathbb{R}^n} f(z,t) \tag{19}$$

and that, for every $t \geq 0$, $f(\cdot,t) \in \mathcal{S}^1_{\mu,L}$. Similar to (15), we consider the following system which we call the time-varying ACCONEST flow:

$$\begin{bmatrix} \dot{\bar{x}}_1 \\ \dot{\bar{x}}_2 \end{bmatrix} = \begin{bmatrix} \bar{x}_2 - \bar{x}_1 - \frac{1}{L}\nabla f(\bar{x}_2,t) \\ \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}(\bar{x}_2 - \bar{x}_1) - \frac{2\sqrt{\kappa}}{(\sqrt{\kappa}+1)L}\nabla f(\bar{x}_2,t) \end{bmatrix} \quad (20)$$

with $\bar{x}_1, \bar{x}_2 \in \mathbb{R}^n$.

Given a fixed time $t \geq 0$, let $x^*(t) = \arg\min_{x\in\mathbb{R}^n} f(x,t)$. Then, $(x^*(t))_{t\geq 0}$ defines the *optimizer trajectory*, i.e., the trajectory of the solution to the optimization problem through time. The following result establishes the performance of the time-varying ACCONEST flow (20) and its discretizations in tracking the optimizer trajectory.

Theorem 5.1 (Performance of the time-varying ACCONEST flow and its optimization algorithms for time-varying optimization)**.** *Consider the time-varying ACCONEST flow* (20) *with* $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $f(\cdot,t) \in \mathcal{S}^1_{\mu,L}$ *for every* $t \geq 0$. *Set* $z(t) := (x_1(t), x_2(t))^\top$ *and* $z^*(t) := (x^*(t), x^*(t))^\top$.

1. *The time-varying ACCONEST flow is contracting with rate* $\sqrt{\frac{\mu}{L}}$ *and respect to* $\|\cdot\|_{2,P^{1/2}}$ *with* $P = \begin{bmatrix} \gamma\frac{\sqrt{\kappa}}{\sqrt{\kappa}+1} & 1 \\ 1 & \frac{\sqrt{\kappa}+1}{\sqrt{\kappa}} \end{bmatrix} \otimes I_n$, $1 < \gamma \leq 1 + \frac{1}{\kappa}$.

2. *Consider that* $x \mapsto f(t,x)$ *is twice continuously differentiable for any* $t \geq 0$ *and that* $t \mapsto \nabla f(t,x)$ *is continuously differentiable with* $\|\dot{\nabla} f(x,t)\|_2 \leq \rho$, $\rho > 0$, *for any* $(t,x) \in [0,\infty) \times \mathbb{R}^n$. *Then,*

$$\begin{aligned} &\|z(t) - z^*(t)\|_{2,P^{1/2}} \\ &\leq \left(\|z(0) - z^*(0)\|_{2,P^{1/2}} - \frac{\sqrt{2\lambda_{\max}(P)}\rho}{\mu}\sqrt{\frac{L}{\mu}}\right) \\ &\quad \times e^{-\sqrt{\frac{\mu}{L}}t} + \frac{\sqrt{2\lambda_{\max}(P)}\rho}{\mu}\sqrt{\frac{L}{\mu}}, \quad t \geq 0, \end{aligned} \quad (21)$$

and so the tracking error is uniformly ultimately bounded as

$$\limsup_{t\to\infty} \|z(t) - z^*(t)\|_{2,P^{1/2}} \leq \frac{\sqrt{2\lambda_{\max}(P)}\rho}{\mu}\sqrt{\frac{L}{\mu}}. \quad (22)$$

Now, consider that $\sup_{k=1,2,\dots} \|\nabla f(x,kh) - \nabla f(x,(k-1)h)\|_2 \leq \rho$, $\rho > 0$, *for any* $x \in \mathbb{R}^n$.

3. *Let* $(y_k := (y_k^{(1)}, y_k^{(2)}))$ *be the sequence generated by the implicit Euler discretization of the time-varying ACCONEST flow* (20) *with integration step-size h. Then,*

$$\begin{aligned} &\|y_k - z^*(kh)\|_{2,P^{1/2}} \\ &\quad \leq (1 + h\sqrt{\frac{\mu}{L}})^{-k}\|y_0 - z^*\|_{2,P^{1/2}} \\ &\quad + \sqrt{2\lambda_{\max}(P)}\frac{\rho}{\mu}\sum_{m=1}^{k}(1 + h\sqrt{\frac{\mu}{L}})^{-m} \end{aligned} \quad (23)$$

for $k = 1, 2, \dots$; *and so the tracking error is uniformly ultimately bounded as*

$$\limsup_{k\to\infty} \|y_k - z^*(kh)\|_{2,P^{1/2}} \leq \frac{\sqrt{2\lambda_{\max}(P)}\rho}{h\mu}\sqrt{\frac{L}{\mu}}. \quad (24)$$

4. *Let* $(y_k := (y_k^{(1)}, y_k^{(2)}))$ *be the sequence generated by the explicit Euler discretization of the time-varying ACCONEST flow* (20) *with integration step-size* h^* *as in statement 4 of Theorem 4.1. Then,*

$$\begin{aligned} &\|y_k - x^*(kh)\|_{2,P^{1/2}} \\ &\quad \leq \left(1 - h^*\sqrt{\frac{\mu}{L}}\right)^{k/2}\|y_0 - x^*(0)\|_{2,P^{1/2}} \\ &\quad + \sqrt{2\lambda_{\max}(P)}\frac{\rho}{\mu}\sum_{m=0}^{k-1}\left(1 - h^*\sqrt{\frac{\mu}{L}}\right)^{m/2}, \end{aligned} \quad (25)$$

for $k = 1, 2, \dots$; *and so the tracking error is uniformly ultimately bounded as*

$$\limsup_{k\to\infty} \|y_k - x^*(kh)\|_{2,P^{1/2}} \leq \frac{\sqrt{2\lambda_{\max}(P)}\frac{\rho}{\mu}}{1 - (1 - h^*\sqrt{\frac{\mu}{L}})^{1/2}}. \quad (26)$$

Remark 5.2 (About the tracking error)**.** *The bounds in the assumptions for statements 2, 3 and 4 of Theorem 5.1 ensure that the rate at which the time-varying optimization changes is bounded. Indeed, the right-hand sides of equations* (22), (24) *and* (26) *are consistent: the larger (lower) these bounds, the larger (lower) the asymptotic tracking error. Finally, and more importantly, the tracking is better when the contraction rate is larger.*

6 CONCLUSION

In this paper we first presented results on the discretization of systems of ODEs using contraction theory. Then, we proposed the Accelerated-Contracting-Nesterov flow and applied our methodology to the design of optimization algorithms based on implicit and explicit Euler discretizations. Finally, we extended our results to the case of time-varying optimization, where contracting properties of the system implies the tracking of the optimizer up to a bounded asymptotic error for both the optimization flow and algorithms.

An important future direction is to apply our discretization framework to the case where the objective function to minimize is convex and not necessarily strongly convex, and aim to establish the contracting nature of an associated optimization flow to then directly establish the convergence of its discretizations. We conjecture that such optimization flow will have a time-varying contraction rate. For example, it would be interesting to formulate a counterpart of the AC-CONEST flow for the case where f is not necessarily strongly convex — perhaps resulting in a non-autonomous system — and study it from a contraction perspective.

Another valuable future direction is to formulate a systematic mechanism where an optimization flows ODEs are built from an existing optimization algorithm, and then modify such ODEs to improve their convergence rate when discretized according to the contraction theory approach described in this paper.

The use of contraction theory also opens the possibility of other interesting future extensions such as the use of state-dependent metrics and/or semi-norms, and the formulation of families of contracting optimization algorithms according to these extensions.

Acknowledgments

This work was funded in part by AFOSR award FA9550-22-1-0059. This work was mostly done while Pedro Cisneros-Velarde was a doctoral student at the University of California, Santa Barbara. We are very grateful to all the anonymous reviewers.

References

Abraham, R., Marsden, J. E., and Ratiu, T. S. (1988). *Manifolds, Tensor Analysis, and Applications*, volume 75 of *Applied Mathematical Sciences*. Springer, 2 edition.

Aminzare, Z. and Sontag, E. D. (2014). Contraction methods for nonlinear systems: A brief introduction and some open problems. In *IEEE Conf. on Decision and Control*, pages 3835–3847.

Bastianello, N., Simonetto, A., and Carli, R. (2020). Primal and dual prediction-correction methods for time-varying convex optimization. *arXiv preprint arXiv:2004.11709*.

Boffi, N. M. and Slotine, J.-J. E. (2020). A continuous-time analysis of distributed stochastic gradient. *Neural Computation*, 32(1):36–96.

Cisneros-Velarde, P., Jafarpour, S., and Bullo, F. (2021). Distributed and time-varying primal-dual dynamics via contraction analysis. *IEEE Transactions on Automatic Control*.

Coppel, W. A. (1965). *Stability and Asymptotic Behavior of Differential Equations*. Heath.

Davydov, A., Jafarpour, S., and Bullo, F. (2021). Non-Euclidean contraction theory for robust nonlinear stability. *arXiv preprint arXiv:2103.12263*.

Desoer, C. A. and Haneda, H. (1972). The measure of a matrix as a tool to analyze computer algorithms for circuit analysis. *IEEE Transactions on Circuit Theory*, 19(5):480–486.

França, G., Sulam, J., Robinson, D. P., and Vidal, R. (2020). Conformal symplectic and relativistic optimization. *Journal of Statistical Mechanics: Theory and Experiment*, 12.

Hairer, E., Nørsett, S. P., and Wanner, G. (1993). *Solving Ordinary Differential Equations I. Nonstiff Problems*. Springer.

Khalil, H. K. (2002). *Nonlinear Systems*. Prentice Hall, 3 edition.

Ladas, G. E. and Lakshmikantham, V. (1972). *Differential Equations in Abstract Spaces*. Academic Press.

Lohmiller, W. and Slotine, J.-J. E. (1998). On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696.

Manchester, I. R. and Slotine, J.-J. E. (2017). On existence of separable contraction metrics for monotone nonlinear systems. *IFAC-PapersOnLine*, 50(1):8226–8231. 20th IFAC World Congress.

Muehlebach, M. and Jordan, M. I. (2019). A dynamical systems perspective on Nesterov acceleration. In *International Conference on Machine Learning*, pages 4656–4662.

Nesterov, Y. (2018). *Lectures on Convex Optimization*. Springer, 2 edition.

Nguyen, H. D., Vu, T. L., Turitsyn, K., and Slotine, J.-J. E. (2018). Contraction and robustness of continuous time primal-dual dynamics. *IEEE Control Systems Letters*, 2(4):755–760.

Revay, M. and Manchester, I. (2020). Contracting implicit recurrent neural networks: Stable models with improved trainability. In *Conference on Learning for Dynamics and Control*, volume 120, pages 393–403.

Ryu, E. K. and Boyd, S. (2016). Primer on monotone operator methods. *Applied Computational Mathematics*, 15(1):3–43.

Shi, B., Du, S. S., Jordan, M. I., and Su, W. J. (2021). Understanding the acceleration phenomenon via high-resolution differential equations. *Mathematical Programming*.

Shi, B., Du, S. S., Su, W., and Jordan, M. I. (2019). Acceleration via symplectic discretization of high-resolution differential equations. In *Advances in Neural Information Processing Systems*.

Simonetto, A., Dall'Anese, E., Paternain, S., Leus, G., and Giannakis, G. B. (2020). Time-varying convex optimization: Time-structured algorithms and applications. *Proceedings of the IEEE*, 108(11):2032–2048.

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

Vaquero, M. and Cortes, J. (2019). Convergence-rate-matching discretization of accelerated optimization flows through opportunistic state-triggered control. In *Advances in Neural Information Processing Systems*.

Vidyasagar, M. (2002). *Nonlinear Systems Analysis*. SIAM.

Wensing, P. M. and Slotine, J.-J. E. (2020). Beyond convexity — Contraction and global convergence of gradient descent. *PLoS One*, 15(8):1–29.

Wibisono, A., Wilson, A. C., and Jordan, M. I. (2016). A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47):E7351–E7358.

Wilson, A. C., Recht, B., and Jordan, M. I. (2018). A Lyapunov analysis of momentum methods in optimization. *arXiv preprint arXiv:1611.02635*.

Zhang, J., Mokhtari, A., Sra, S., and Jadbabaie, A. (2018). Direct Runge-Kutta discretization achieves acceleration. In *Advances in Neural Information Processing Systems*.

Supplementary Material: A Contraction Theory Approach to Optimization Algorithms from Acceleration Flows

A PROOF OF THEOREM 3.1

We first prove statement (i). From the implicit Euler discretization we have $y_{k+1}-x^* = y_k-x^*+hg(y_{k+1},(k+1)h)$. Then,

$$\begin{aligned}\|y_{k+1}-x^*\|^2_{2,P^{1/2}} &= \langle P^{1/2}(y_k-x^*), P^{1/2}(y_{k+1}-x^*)\rangle + h\langle P^{1/2}g(y_{k+1},(k+1)h), P^{1/2}(y_{k+1}-x^*)\rangle \\ &\leq \langle P^{1/2}(y_k-x^*), P^{1/2}(y_{k+1}-x^*)\rangle - h\bar{\gamma}((k+1)h)\|y_{k+1}-x^*\|^2_{2,P^{1/2}}\end{aligned}$$

where the inequality follows from the one-sided Lipschitz condition and $g(x^*,(k+1)h) = 0_n$. Then, since $\langle P^{1/2}(y_k-x^*), P^{1/2}(y_{k+1}-x^*)\rangle \leq \|y_k-x^*\|_{2,P^{1/2}}\|y_{k+1}-x^*\|_{2,P^{1/2}}$, we obtain

$$\|y_{k+1}-x^*\|_{2,P^{1/2}} \leq (1+h\bar{\gamma}((k+1)h))^{-1}\|y_k-x^*\|_{2,P^{1/2}}. \tag{27}$$

Expanding the previous expression backwards in time leads to $\|y_{k+1}-x^*\|_{2,P^{1/2}} \leq (1+h\bar{\gamma}((k+1)h))^{-1}\dots(1+h\bar{\gamma}((k-m)h))^{-1}\|y_{k-m-1}-x^*\|_{2,P^{1/2}}$ for $m \geq 1$ and $k-m-1 \geq 0$, from which equation (1) from the main paper follows immediately. This finishes the proof of statement (i).

We now prove statement (ii). Since $g(x^*((k+1)h),(k+1)h) = 0_n$, we use (27) to obtain

$$\begin{aligned}&\|y_{k+1}-x^*((k+1)h)\|_{2,P^{1/2}} \\ &\leq (1+h\bar{\gamma}((k+1)h))^{-1}\|y_k-x^*((k+1)h)\|_{2,P^{1/2}} \\ &\leq (1+h\bar{\gamma}((k+1)h))^{-1}(\|y_k-x^*(kh)\|_{2,P^{1/2}} + \|x^*((k+1)h)-x^*(kh)\|_{2,P^{1/2}}).\end{aligned}$$

Expanding the previous expression backwards in time leads to $\|y_{k+1}-x^*((k+1)h)\|_{2,P^{1/2}} \leq (1+h\bar{\gamma}((k+1)h))^{-1}\dots(1+h\bar{\gamma}((k-m)h))^{-1}\|y_{k-m-1}-x^*\|_{2,P^{1/2}} + (1+h\bar{\gamma}((k+1)h))^{-1}\dots(1+h\bar{\gamma}((k-m)h))^{-1}\|x^*((k-m)h)-x^*((k-(m+1))h)\|_{2,P^{1/2}}+\dots+(1+h\bar{\gamma}((k+1)h))^{-1}\|x^*((k+1)h)-x^*(kh)\|_{2,P^{1/2}}$ for $m \geq 1$, $k-m-1 \geq 0$; from which equation (3) from the main paper follows immediately. Now we prove the second part of statement (ii). Observe that under the conditions of the statement,

$$\begin{aligned}\sum_{m=1}^{k}\prod_{\ell=m}^{k}(1+h\bar{\gamma}(mh))^{-1}\|x^*(mh)-x^*((m-1)h)\|_{2,P^{1/2}} &\leq \rho\sum_{m=1}^{k}\prod_{\ell=m}^{k}(1+h\gamma)^{-1} \\ &= \rho\sum_{m=1}^{k}(1+h\gamma)^{-(k-m+1)} = \rho\sum_{m=1}^{k}(1+h\gamma)^{-m},\end{aligned}$$

and this result along with the linear convergence result from statement (i) imply equation (4) from the main paper. Now, $\lim_{k\to\infty}\sum_{m=1}^{k}\frac{1}{(1+h\gamma)^{-m}} = \frac{1}{h\gamma}$ follows from the convergence of geometric series, and this along with $\lim_{k\to\infty}\frac{\|y_0-x^*(0)\|_{2,P^{1/2}}}{(1+h\gamma)^k} = 0$ imply equation (5) from the main paper. This finishes the proof of statement (ii).

B PROOF OF THEOREM 3.4

We first prove statement (i). From the Euler discretization we have $y_{k+1}-x^* = y_k-x^*+hg(y_k,kh)$. Then,

$$\begin{aligned}\|y_{k+1}-x^*\|^2_{2,P^{1/2}} &= \|y_k-x^*+hg(y_k,kh)\|^2_{2,P^{1/2}} \\ &= \|y_k-x^*\|^2_{2,P^{1/2}} + 2h\langle P^{1/2}(y_k-x^*), P^{1/2}g(y_k,kh)\rangle + h^2\|g(y_k,kh)\|^2_{2,P^{1/2}} \\ &\leq \|y_k-x^*\|^2_{2,P^{1/2}} - 2h\bar{\gamma}(kh)\|y_k-x^*\|^2_{2,P^{1/2}} + h^2\|g(y_k,kh)\|^2_{2,P^{1/2}}\end{aligned}$$

where the inequality follows from the one-sided Lipschitz condition. Now, we have $\|g(y_k, kh)\|^2_{2,P^{1/2}} \leq \ell^2\|y_k - x^*\|_2^2$ from g being uniformly ℓ-Lipschitz continuous and $g(x^*, kh) = 0_n$, and so

$$\|y_{k+1} - x^*\|_{2,P^{1/2}} \leq (1 - 2h\bar{\gamma}(kh) + h^2\ell^2)^{1/2}\|y_k - x^*\|_{2,P^{1/2}},$$

and note that the theorem's conditions on the step-size h ensures that $(1 - 2h\bar{\gamma}(kh) + h^2\ell^2)^{1/2} < 1$. Expanding the previous expression backwards in time leads to $\|y_{k+1} - x^*\|_{2,P^{1/2}} \leq (1-2h\bar{\gamma}(kh)+h^2\ell^2)^{1/2}\dots(1-2h\bar{\gamma}((k-m)h) + h^2\ell^2)^{1/2}\|y_{k-m} - x^*\|_{2,P^{1/2}}$ for $m \geq 0$ and $k - m \geq 0$, from which equation (6) from the main paper follows immediately. This finishes the proof of statement (i).

We now prove statement (ii). Note that, using the proof of statement (i), $\|y_{k+1}-x^*(kh)\|^2_{2,P^{1/2}} = \|y_k-x^*(kh)+hg(y_k, kh)\|^2_{2,P^{1/2}} \leq (1 - 2h\bar{\gamma}(kh) + h^2\ell^2)\|y_k - x^*(kh)\|_{2,P^{1/2}}$. Then, we obtain

$$\begin{aligned}
&\|y_{k+1} - x^*((k+1)h)\|_{2,P^{1/2}} \\
&\leq \|y_{k+1} - x^*(kh)\|_{2,P^{1/2}} + \|x^*((k+1)h) - x^*(kh)\|_{2,P^{1/2}} \\
&\leq (1 - 2h\bar{\gamma}(kh) + h^2\ell^2)^{1/2}\|y_k - x^*(kh)\|_{2,P^{1/2}} + \|x^*((k+1)h) - x^*(kh)\|_{2,P^{1/2}}.
\end{aligned}$$

Expanding the previous expression backwards in time leads to $\|y_{k+1} - x^*((k+1)h)\|_{2,P^{1/2}} \leq (1 - 2h\bar{\gamma}(kh) + h^2\ell^2)^{1/2}\dots(1-2h\bar{\gamma}((k-m)h)+h^2\ell^2)^{1/2}\|y_{k-m} - x^*((k-m)h)\|_{2,P^{1/2}} + (1-2h\bar{\gamma}(kh)+h^2\ell^2)^{1/2}\dots(1-2h\bar{\gamma}((k-(m-1))h)+h^2\ell^2)^{1/2}\|x^*((k-(m-1))h) - x^*((k-m)h)\|_{2,P^{1/2}} + \dots + (1-2h\bar{\gamma}(kh)+h^2\ell^2)^{1/2}\|x^*(kh) - x^*((k-1)h)\|_{2,P^{1/2}} + \|x^*((k+1)h) - x^*(kh)\|_{2,P^{1/2}}$ for $m \geq 1$, $k - m - 1 \geq 0$; from which equation (8) from the main paper follows immediately. The proof for the second part of statement (ii) is very similar to the one done for the second part of statement (ii) in Theorem 3.1, and thus is omitted. The proof of statement (iii) is immediately obtained from $h^* = \arg\min_{h>0} -2h\gamma + h^2\ell^2$.

C PROOF OF THEOREM 4.1

Consider any $x_1, x_2, z_1, z_2 \in \mathbb{R}^n$ and let $P = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \otimes I_n$ with $a, c > 0$, $b \in \mathbb{R}$; and let $\beta := \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$. Then, the ACCONEST flow (equation (15) from the main paper) is described by: $\dot{\bar{x}}_1 = \bar{x}_2 - \bar{x}_1 - \frac{1}{L}\nabla f(\bar{x}_2)$ and $\dot{\bar{x}}_2 = \beta(\bar{x}_2 - \bar{x}_1) - \frac{\beta+1}{L}\nabla f(\bar{x}_2)$. After some algebraic work we obtain

$$\begin{aligned}
\eta &:= (F_{\text{Acc}}(x_1, x_2) - F_{\text{Acc}}(z_1, z_2))^\top P \begin{bmatrix} x_1 - z_1 \\ x_2 - z_2 \end{bmatrix} \\
&= (a + b\beta)(x_1 - z_1)^\top(x_2 - z_2) - (a + b\beta)\|x_1 - z_1\|_2^2 + \frac{a + b(\beta+1)}{L}(\nabla f(x_2) - \nabla f(z_2))^\top(x_1 - z_1) \\
&\quad + (b + c\beta)\|x_2 - z_2\|_2^2 - (b + c\beta)(x_1 - z_1)^\top(x_2 - z_2) - \frac{b + c(\beta+1)}{L}(\nabla f(x_2) - \nabla f(z_2))^\top(x_2 - z_2).
\end{aligned}$$

Now, we use the strong convexity property $-(\nabla f(x_2)-\nabla f(z_2))^\top(x_2-z_2) \leq -\mu\|x_2-z_2\|_2^2$ and the L-smoothness of f as $\|\nabla f(x_2) - \nabla f(z_2)\|_2^2 \leq L\|x_2 - z_2\|_2^2$ in the equality $\|\sqrt{\frac{\xi}{4L}}(\nabla f(x_2) - \nabla f(z_2)) + \sqrt{\frac{L}{\xi}}(x_2 - z_2)\|_2^2 = \frac{\xi}{4L}\|\nabla f(x_1) - \nabla f(z_1)\|_2^2 + (\nabla f(x_1) - \nabla f(z_1))^\top(x_2 - z_2) + \frac{L}{\xi}\|x_2 - z_2\|_2^2$, with $\xi > 0$, to obtain

$$\begin{aligned}
\eta &\leq (a + b\beta - (b + c\beta))(x_2 - z_2)^\top(x_1 - z_1) + \left(\frac{a + b(\beta+1)}{\xi} - a - b\beta\right)\|x_1 - z_1\|_2^2 \\
&\quad + \left(\frac{a + b(\beta+1)}{4}\xi + (b + c\beta) - (b + c(\beta+1))\frac{1}{\kappa}\right)\|x_2 - z_2\|_2^2.
\end{aligned}$$

Now, the coefficients multiplying the terms $\|x_1 - z_1\|_2^2$, $\|x_2 - z_2\|_2^2$ and $(x_1 - z_1)^\top(x_2 - z_2)$ must be less or equal than $-\frac{1}{\sqrt{\kappa}}a$, $-\frac{1}{\sqrt{\kappa}}c$ and $-\frac{2}{\sqrt{\kappa}}b$ respectively, in order that the vector field F_{Acc} satisfies the one-sided Lipschitz condition with constant $-\frac{1}{\sqrt{\kappa}}$, i.e., with $\bar{\gamma}(t) \equiv \frac{1}{\sqrt{\kappa}}$ in the statement (ii) of Theorem 2.1 and with $\|\cdot\|_{2,P^{1/2}}$ on the right-hand side of its inequality. Thus, we need to analyze conditions on a, b, c such that the following inequalities hold

$$a\left(\frac{1}{\xi}-1+\frac{1}{\sqrt{\kappa}}\right)+b\left(\frac{\beta+1}{\xi}-\beta\right)\leq 0 \tag{28}$$

$$\frac{a\xi}{4}+b\left(\frac{\xi}{4}(\beta+1)+1-\frac{1}{\kappa}\right)+c\left(\beta+\frac{1}{\sqrt{\kappa}}-\frac{\beta+1}{\kappa}\right)\leq 0 \tag{29}$$

$$a+b\left(\beta-1+\frac{2}{\sqrt{\kappa}}\right)-c\beta\leq 0. \tag{30}$$

We first analyze conditions that ensure equation (28) holds. We choose $\xi=\frac{\sqrt{\kappa}}{\sqrt{\kappa}-1}$ to eliminate the dependency on a (recall that $\kappa>1$). With this value of ξ, the coefficient multiplying b on the left-hand side of (28) becomes

$$\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}}(\beta+1)-\beta=\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}>0;$$

thus to ensure (28) holds, we choose $b<0$. For simplicity, in what follows in the proof, we do the change of variables $b\to -b$ so that we can take $b>0$.

We now analyze conditions that ensure equation (29) holds, which after the change of variables $b\to -b$ and rearranging some terms becomes

$$\frac{\xi}{4}\left(a-b(1+\beta)\right)-b\left(1-\frac{1}{\kappa}\right)+c\left(\beta+\frac{1}{\sqrt{\kappa}}-\frac{\beta+1}{\kappa}\right)\leq 0. \tag{31}$$

Now, by establishing the condition

$$a\leq b(1+\beta) \tag{32}$$

and observing that $\beta+\frac{1}{\sqrt{\kappa}}-\frac{1+\beta}{\kappa}=\frac{\kappa-1}{\sqrt{\kappa}(\sqrt{\kappa}+1)}=\left(1-\frac{1}{\kappa}\right)\left(\frac{\sqrt{\kappa}}{\sqrt{\kappa}+1}\right)<1-\frac{1}{\kappa}$, we conclude that in order to satisfy inequality (31), we need to ensure:

$$-b\left(1-\frac{1}{\kappa}\right)+c\left(1-\frac{1}{\kappa}\right)\left(\frac{\sqrt{\kappa}}{\sqrt{\kappa}+1}\right)\leq 0 \quad\Longrightarrow\quad c\leq\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}}b. \tag{33}$$

Now, recall that we need P to be positive definite, for which the inequality $ac>b^2$ must hold. Therefore, to ensure P is positive definite, we take $c=\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}}b$, which satisfies (33), and choose $a=\gamma\frac{\sqrt{\kappa}}{\sqrt{\kappa}+1}b$ where $\gamma>1$ must be chosen such that (32) is satisfied.

We now analyze conditions that ensure equation (30) holds, which after the change of variables $b\to -b$ becomes

$$a-b\left(\beta-1+\frac{2}{\sqrt{\kappa}}\right)-c\beta\leq 0. \tag{34}$$

Replacing our newly assigned values for a and c in (31), we only need to verify that the following inequality holds

$$\gamma\frac{\sqrt{\kappa}}{\sqrt{\kappa}+1}-\left(\beta-1+\frac{2}{\sqrt{2}}\right)-\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}}\beta\leq 0$$

for an appropriate value of $\gamma>0$. After some algebraic work, we conclude that $\gamma\leq 1+\frac{1}{\kappa}$ satisfies this inequality, besides of defining a value for a which satisfies inequality (32).

Finally, after the change of variables $b\to -b$, we set $b=1$ and conclude that the ACCONEST flow satisfies the one-sided Lipschitz condition of Theorem 2.1 with constant $-\frac{1}{\sqrt{\kappa}}$. Thus, contraction of the system in equation (15) from the main paper follows as stipulated in statement (i). The proof of statement (ii) follows immediately from Theorem 2.1.

Now we prove statement (iii). First, we use the result in statement (i) of Theorem 3.1 to obtain

$$\begin{aligned}\|y_k^{(1)}-x^*\|_2^2&\leq\|y_k^{(1)}-x^*\|_2^2+\|y_k^{(2)}-x^*\|_2^2\\&\leq\frac{1}{\lambda_{\max}(P)}\|(y_k^{(1)}-x^*,y_k^{(2)}-x^*)^\top\|_{2,P^{1/2}}^2\leq(1+h\frac{1}{\sqrt{\kappa}})^{-k}\|(y_0^{(1)}-x^*,y_0^{(2)}-x^*)^\top\|_{2,P^{1/2}}^2.\end{aligned} \tag{35}$$

Using this result, along with $f(y_k^{(1)}) - f(x^*) \le \nabla f(x^*)^\top (y_k^{(1)} - x^*) + \frac{L}{2}\|y_k^{(1)} - x^*\|_2^2 = \frac{L}{2}\|y_k^{(1)} - x^*\|_2^2$ from the L-smoothness of f, we obtain

$$f(y_k^{(1)}) - f(x^*) \le \frac{L}{2\lambda_{\min}(P)} \|(y_0^{(1)} - x^*, y_0^{(2)} - x^*)^\top\|_{2,P^{1/2}}^2 \left(1 + h\frac{1}{\sqrt{\kappa}}\right)^{-2k},$$

which concludes the proof of statement (iii).

Now we prove statement (iv). In order to apply Theorem 3.4, we need to find a Lipschitz constant for the vector field F_{Acc}. Thus,

$$\begin{aligned}
&\|F_{\text{Acc}}(y_k^{(1)}, y_k^{(2)})\|_{2,P^{1/2}}^2 \\
&\le \lambda_{\max}(P)\|F_{\text{Acc}}(y_k^{(1)}, y_k^{(2)})\|_2^2 \\
&\le \lambda_{\max}(P)(7 + 5\beta + 6\beta^2)\|(y_k^{(1)} - x^*, y_k^{(2)} - x^*)^\top\|_2^2 \\
&\le \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}(7 + 5\beta + 6\beta^2)\|(y_k^{(1)} - x^*, y_k^{(2)} - x^*)^\top\|_{2,P^{1/2}}^2.
\end{aligned} \tag{36}$$

We now explain how we obtain the second inequality in (36). Firstly, note that $\|F_{\text{Acc}}(y_k^{(1)}, y_k^{(2)})\|_2^2 = \|y_k^{(2)} - y_k^{(1)} - \frac{1}{L}\nabla f(y_k^{(2)})\|_2^2 + \|\beta(y_k^{(2)} - y_k^{(1)}) - \frac{1+\beta}{L}\nabla f(y_k^{(2)})\|_2^2$. Now, $\|y_k^{(2)} - y_k^{(1)} - \frac{1}{L}\nabla f(y_k^{(2)})\|_2 \le \|y_k^{(2)} - x^*\|_2 + \|y_k^{(1)} - x^*\|_2 + \frac{1}{L}\|\nabla f(y_k^{(2)}) - \nabla f(x^*)\|_2 \le 2\|y_k^{(2)} - x^*\|_2 + \|y_k^{(1)} - x^*\|_2$ by using the triangle inequality and the L-smoothness of f. Then, $\|y_k^{(2)} - y_k^{(1)} - \frac{1}{L}\nabla f(y_k^{(2)})\|_2^2 \le 4\|y_k^{(2)} - x^*\|_2^2 + \|y_k^{(1)} - x^*\|_2^2 + 4\|y_k^{(2)} - x^*\|_2\|y_k^{(1)} - x^*\|_2 \le 6\|y_k^{(2)} - x^*\|_2^2 + 3\|y_k^{(1)} - x^*\|_2^2$, where the last inequality follows from $a^2 + b^2 \ge 2ab$ for any $a, b \in \mathbb{R}$. Likewise, we can obtain $\|\beta(y_k^{(2)} - y_k^{(1)}) - \frac{1+\beta}{L}\nabla f(y_k^{(2)})\|_2^2 \le (1 + 4\beta + 6\beta^2)\|y_k^{(2)} - x^*\|_2^2 + (\beta + 3\beta^2)\|y_k^{(1)} - x^*\|_2^2$. Putting it all together, we obtain the sought inequality: $\|F_{\text{Acc}}(y_k^{(1)}, y_k^{(2)})\|_2^2 = \|y_k^{(2)} - y_k^{(1)} - \frac{1}{L}\nabla f(y_k^{(2)})\|_2^2 + \|\beta(y_k^{(2)} - y_k^{(1)}) - \frac{1+\beta}{L}\nabla f(y_k^{(2)})\|_2^2 \le (7 + 5\beta + 6\beta^2)\|(y_k^{(1)} - x^*, y_k^{(2)} - x^*)^\top\|_2^2$.

Then, since the square of the Lipschitz constant is $\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}(7 + 5\beta + 6\beta^2)$, the optimal integration step-size according to Theorem 3.4 is $h^* := \frac{1}{\sqrt{\kappa}(7+5\beta+6\beta^2)}\frac{\lambda_{\min}(P)}{\lambda_{\max}(P)}$, and then we have that

$$\|(y_k^{(1)} - x^*, y_k^{(2)} - x^*)^\top\|_{2,P^{1/2}} \le (1 - h^*\sqrt{\frac{\mu}{L}})^{k/2}\|(y_0^{(1)} - x^*, y_0^{(2)} - x^*)^\top\|_{2,P^{1/2}}$$

for $k = 1, 2, \dots$. Finally, a similar analysis to the proof of statement (iii) leads to

$$f(y_k^{(1)}) - f(x^*) \le \frac{L}{2\lambda_{\min}(P)}\|(y_0^{(1)} - x^*, y_0^{(2)} - x^*)^\top\|_{2,P^{1/2}}^2 \left(1 - h^*\sqrt{\frac{\mu}{L}}\right)^k,$$

which concludes the proof of statement (iv).

D PROOF OF THEOREM 5.1

The proof of statement (i) is virtually the same as the one of statement (i) of Theorem 4.1 due to the assumption $f(\cdot, t) \in \mathcal{S}_{\mu,L}^1$ for every $t \ge 0$. Now we prove statement (ii). Let us fix any $t \ge 0$ and observe that

$$0_n = -\nabla f(x^*(t), t). \tag{37}$$

We first show that the curve $t \mapsto x^*(t)$ is continuously differentiable. Define the function $g : \mathbb{R}^{n+1} \to \mathbb{R}^n$ as $g(t, x) = -\nabla f(x, t)$. Since $t \mapsto \nabla f(x, t)$ is continuously differentiable, the function g is continuously differentiable on $\mathbb{R}^{n+1}$. Moreover, note that $\nabla_x g(t, x) = -\nabla^2 f(x, t) \preceq -\mu I_n$ which implies that $\nabla_x g(t, x)$ is Hurwitz and therefore nonsingular. Then, the Implicit Function (Abraham et al., 1988, Theorem 2.5.7) implies the solutions $t \mapsto x^*(t)$ of the algebraic equation (37) is continuously differentiable for any $t \ge 0$.

Now, differentiating equation (37) with respect to time,

$$\begin{aligned}
&\implies 0_n = -\nabla^2 f(x^*(t), t)\dot{x^*}(t) - \dot{\nabla} f(x^*(t), t) \\
&\implies 0_m = -\dot{x^*}(t) - (\nabla^2 f(x^*(t), t))^{-1}\dot{\nabla} f(x^*(t), t) \\
&\implies \|\dot{x^*}(t)\|_2 = \|(\nabla^2 f(x^*(t), t))^{-1}\dot{\nabla} f(x^*(t), t)\|_2 \le \|(\nabla^2 f(x^*(t), t))^{-1}\|_2\|\dot{\nabla} f(x^*(t), t)\|_2 \le \frac{\rho}{\mu}
\end{aligned}$$

where the last inequality follows from $\mu I_n \preceq \nabla^2 f(x,t) \preceq L I_n \implies \frac{1}{L} I_n \preceq (\nabla^2 f(x,t))^{-1} \preceq \frac{1}{\mu} I_n$ and thus $\|(\nabla^2 f(x,t))^{-1}\|_2 \leq \frac{1}{\mu}$ for every $(t,x) \in [0,\infty) \times \mathbb{R}^n$.

Now, considering the contraction of the system of equation (20) from the main paper from statement (i), we set $\delta(t) := \left\| \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} - \begin{bmatrix} x^*(t) \\ x^*(t) \end{bmatrix} \right\|_{2,P^{1/2}}$ and use (Nguyen et al., 2018, Lemma 2) to obtain the following differential inequality $\dot{\delta}(t) \leq -\sqrt{\frac{\mu}{L}}\delta(t) + \left\| \begin{bmatrix} \dot{x^*}(t) \\ \dot{x^*}(t) \end{bmatrix} \right\|_{2,P^{1/2}}$. Then, using our previous results,

$$\dot{\delta}(t) \leq -\sqrt{\frac{\mu}{L}}\delta(t) + \sqrt{2\lambda_{\max}(P)}\|\dot{x^*}\|_2 \leq -\sqrt{\frac{\mu}{L}}\delta(t) + \sqrt{2\lambda_{\max}(P)}\frac{\rho}{\mu}.$$

Now, since the function $h(u) = -\sqrt{\frac{\mu}{L}}u + \sqrt{2\lambda_{\max}(P)}\frac{\rho}{\mu}$ is Lipschitz (it is an affine function), we use the Comparison Lemma (Khalil, 2002) to upper bound $\delta(t)$ by the solution to the differential equation $\dot{u}(t) = -\sqrt{\frac{\mu}{L}}u + \sqrt{2\lambda_{\max}(P)}\frac{\rho}{\mu}$ for all $t \geq 0$, from which statement (ii) follows.

Now we prove both statements (iii) and (iv). First, notice that $\|x^*(kh) - x^*((k-1)h)\|_2 \leq \frac{1}{\mu}\|\nabla f(x^*(kh), kh) - \nabla f(x^*((k-1)h), kh)\|_2 = \frac{1}{\mu}\|\nabla f(x^*((k-1)h), kh)\|_2 = \frac{1}{\mu}\|\nabla f(x^*((k-1)h), kh) - \nabla f(x^*((k-1)h), (k-1)h)\|_2 \leq \frac{\rho}{\mu}$, where the first inequality follows from the strong convexity assumption. Using this result, we obtain

$$\left\| \begin{bmatrix} x^*(kh) \\ x^*(kh) \end{bmatrix} - \begin{bmatrix} x^*((k-1)h) \\ x^*((k-1)h) \end{bmatrix} \right\|_{2,P^{1/2}} = \sqrt{2\lambda_{\max}(P)}\|x^*(kh) - x^*((k-1)h)\|_2 \leq \sqrt{2\lambda_{\max}(P)}\frac{\rho}{\mu}. \tag{38}$$

Since the system is contracting from statement (i), we use (38) and statement (ii) of Theorem 3.1 to conclude the proof of statement (iii). Finally, the proof of statement (iv) is very similar: again, since the system is contracting, we use (38) and statement (ii) of Theorem 3.4 to conclude the proof.