A Passivity-Based Method for Accelerated Convex Optimisation

Namhoon Cho, and Hyo-Sang Shin

Abstract— This study presents a constructive methodology for designing accelerated convex optimisation algorithms in continuous-time domain. The two key enablers are the classical concept of passivity in control theory and the time-dependent change of variables that maps the output of the internal dynamic system to the optimisation variables. The Lyapunov function associated with the optimisation dynamics is obtained as a natural consequence of specifying the internal dynamics that drives the state evolution as a passive linear time-invariant system. The passivity-based methodology provides a general framework that has the flexibility to generate convex optimisation algorithms with the guarantee of different convergence rate bounds on the objective function value. The same principle applies to the design of online parameter update algorithms for adaptive control by re-defining the output of internal dynamics to allow for the feedback interconnection with tracking error dynamics.

I. INTRODUCTION

The analysis of accelerated optimisation algorithms in the continuous-time domain has been the area of active research in recent years. Well-known results include the analysis of Nesterov's accelerated gradient descent method provided in [1] by taking the limit of infinitesimal step-size. In relation to continuous-time optimisation as well as direct adaptive control, [2] presented four different algorithms and studied their stability properties. The recent work of [3] provided a much more generalised understanding towards various accelerated convex optimisation methods by highlighting the role of using a time-dilated coordinate in the analysis. However, the scope of [1]–[3] is focused on the second-order optimisation dynamics of specific forms.

Meanwhile, there have been attempts to approach optimisation from robust control perspectives [4]–[6]. The control-theoretic concept of dissipativity was noticed to be central in the construction of associated Lyapunov function, which in turn enables rigorous convergence analysis. Although the role of dissipativity theory in the analysis and design of optimisation algorithms was illustrated very clearly, the scope of [4]–[6] alone falls short of providing a constructive procedure for achieving accelerated convergence in the design of a new optimisation algorithm.

Motivated by the viewpoints provided by the previous works, we propose a constructive methodology for streamlined design and analysis of accelerated convex optimisation

Namhoon Cho and Hyo-Sang Shin are with the Centre for Autonomous and Cyber-Physical Systems, Faculty of Engineering and Applied Sciences, Cranfield University, Cranfield, Bedfordshire, MK43 0AL, United Kingdom. e-mail: {n.cho, h.shin}@cranfield.ac.uk

Hyo-Sang Shin is with Cho Chun Shik Graduate School of Mobility, Korea Advanced Institute of Science and Technology, Daejeon, 34051, South Korea. e-mail: hyosangshin@kaist.ac.kr

algorithms. The proposed approach combines the utility of the two design philosophies, namely, the coordinate transform of [3] and the dissipativity viewpoint of [6]. Central to the new development is the idea that the passive input-output pair of a open-loop stable linear time-invariant dynamic system can be defined to be the gradient of an objective function and the vector of optimisation variables, respectively. Our new method characterises the structure of passive optimisation dynamics to a generalised extent while guaranteeing a prescribed order for the convergence rate bound on the objective function value.

II. PRELIMINARIES

This section introduces several basic concepts as the mathematical preliminaries and the notation.

A. Convex Optimisation

Consider a differentiable and convex function $f : \mathbb{R}^n \to \mathbb{R}$. The optimisation problem of our interest is described as

$$\underset{\theta \in \mathbb{R}^n}{\text{minimise}} \quad f(\theta) \tag{1}$$

If the solution of the problem in Eq. (1) exists, let $\theta_* \in \mathbb{R}^n$ represents the optimal solution. Also, let $f_* \triangleq \inf_{\theta \in \mathbb{R}^n} f(\theta)$ be the optimal value of the objective function.

The Bregman divergence associated with f for points $p, q \in \mathbb{R}^n$ is defined as

$$D_f(p, q) \triangleq f(p) - f(q) - \langle \nabla f(q), p - q \rangle \tag{2}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. By definition of convexity, for a differentiable and convex function f, we have

$$D_f(p, q) \geq 0 \tag{3}$$

for all $p, q \in \mathbb{R}^n$.

B. Stability and Passivity

Consider a linear time-invariant dynamic system whose state space realisation is given by

$$\mathcal{G} : \quad \begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{4}$$

with constant matrices A, B, C, and D, where x represents the state, u represents the input, and y represents the output. One may consider $\mathcal{G}[\cdot]$ as an operator that maps u to y as $y = \mathcal{G}[u]$.

The following lemma is a classical result of linear system theory which relates the stability of a system that has no input-output with a Lyapunov equation.

Lemma 1 (Lyapunov Equation, Theorem 3.7 in [7]). *A matrix A is Hurwitz if and only if for every positive definite symmetric matrix Q there exists a positive definite symmetric matrix P that satisfies the Lyapunov equation*

$$A^T P + PA = -Q \tag{5}$$

Moreover, if A is Hurwitz, then P is the unique solution.

Passivity is a fundamental concept in classical control theory which formally describes the property of a system that does not generate energy internally but only stores and dissipates energy supplied from outside [8].

Definition 1 (Passivity, Definition 5.3 in [7]). Consider a dynamic system described by

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)) \\ y(t) &= h(x(t), u(t)) \end{aligned} \tag{6}$$

The system in Eq. (6) is passive if there exists a continuously differentiable positive semidefinite storage function $V(x)$ such that

$$u^T y \geq \dot{V} = \langle \nabla_x V, f(x, u) \rangle \qquad \forall (x, u) \tag{7}$$

Moreover, it is strictly passive if $u^T y \geq \dot{V} + \phi(x)$ for some positive definite function ϕ.

The following lemma is another classical result of linear system theory which relates the passivity of a system that has input-output with a set of algebraic equations extending Eq. (5).

Lemma 2 (Kalman-Yakubovich-Popov, Lemma 5.3 and 5.4 in [7]). *Let $G(s) = C(sI - A)^{-1} B + D$ be a square transfer function matrix, where (A, B) is controllable and (A, C) is observable. Then, $G(s)$ is (strictly) positive real if and only if there exist matrices $P = P^T > 0$, L, and W, (and a positive constant ϵ) such that*

$$PA + A^T P = -L^T L (-\epsilon P) \tag{8}$$

$$PB = C^T - L^T W \tag{9}$$

$$W^T W = D + D^T \tag{10}$$

Moreover, the system with minimal realisation (A, B, C, D) is (strictly) passive if $G(s)$ is (strictly) positive real. The storage function certifying passivity of the system is given by

$$V = \frac{1}{2} x^T P x \tag{11}$$

Proof. For V given by Eq. (11), straightforward substitution of Eqs. (8)-(10) verifies

$$\begin{aligned} u^T y - \dot{V} &= u^T (Cx + Du) - x^T P (Ax + Bu) \\ &= \frac{1}{2} (Lx + Wu)^T (Lx + Wu) + \frac{1}{2} \epsilon x^T P x \quad (12) \\ &\geq \epsilon V \end{aligned}$$

$\square$

Readers may refer to [9] for more detailed discussion on the equivalence between the frequency-domain notion of strict positive realness and the time-domain notion of strict passivity for linear dynamic systems.

C. Partial Derivative of Time-Dependent Function

Motivated by [3], consider a scalar-valued function $U(v(t), t)$ where vector-valued $v(t)$ is differentiable with respect to t. The chain rule states that

$$\frac{d}{dt} U(v(t), t) = \langle \nabla_v U(v, t), \dot{v}(t) \rangle + \frac{\partial}{\partial t} U(v, t) \tag{13}$$

Integrating Eq. (13) from t_0 to t gives

$$\begin{aligned} &\int_{t_0}^{t} \langle \nabla_v U(v, \tau), \dot{v}(\tau) \rangle \, d\tau \\ &= U(v(t), t) - U(v(t_0), t_0) - \int_{t_0}^{t} \frac{\partial}{\partial \tau} U(v, \tau) \, d\tau \end{aligned} \tag{14}$$

III. PASSIVITY-BASED DESIGN OF OPTIMISERS

A. Generic Method

Suppose that $U(v(t), t) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is a lower-bounded function (which will be specified later) whose value is nonnegative for every t and defined with respect to some variable $v(t) \in \mathbb{R}^n$ (which will also be specified later). Let us introduce a variable $w(t) \in \mathbb{R}^n$ where $n = \dim(\theta)$ which evolves over time according to

$$\begin{aligned} w^{(m)}(t) &+ a_{m-1} w^{(m-1)}(t) + \cdots + a_1 \dot{w}(t) + a_0 w(t) \\ &= -\alpha(t) \nabla_v U(v(t), t) \end{aligned} \tag{15}$$

where $\{a_i\}_{i=0}^{n-1}$ is a set of constant coefficients, $\alpha(t) \neq 0$ for $\forall t \geq t_0$, and $w^{(i)}$ denotes the i-th time-derivative of w.

Let us refer to the system in Eq. (15) as the generator dynamics $\mathcal{G}$ for which state and input are defined as

$$x = \begin{bmatrix} w \\ \dot{w} \\ \ddot{w} \\ \vdots \\ w^{(m-1)} \end{bmatrix} \in \mathbb{R}^{mn} \tag{16}$$

$$u = -\alpha \nabla_v U(v, t) \in \mathbb{R}^n$$

Then, the ordinary differential equation in Eq. (15) can be rewritten as a linear system of the form in Eq. (4) with the controllable canonical form realisation given by

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{m-1} \end{bmatrix} \otimes I_n$$

$$B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \otimes I_n \tag{17}$$

where $\otimes$ represents the Kronecker product.

Suppose that A is Hurwitz by appropriate choice of $\{a_i\}_{i=0}^{n-1}$ and Q is an arbitrarily chosen symmetric positive definite matrix. Given strictly stable A and B as defined in

Eq. (17), Lemma 1 states that there exists symmetric positive definite P satisfying Eq. (5). As noted in [10], given the definition of input u, Lemma 2 implies that the arbitrariness of Q carries over to the possibility of constructing an infinite number of passive systems by compatible choices of outputs. Let the output has no direct feedthrough term, i.e., $D = 0$. Then, Lemma 2 verifies that the compatible output can be obtained by setting

$$C = B^T P \qquad (18)$$

In this setup, the time-derivative of the storage function $V = \frac{1}{2}x^T P x$ can be written as

$$\dot{V} = x^T P (Ax + Bu) = x^T P B u - \frac{1}{2} x^T Q x$$
$$= \langle y, u \rangle - \frac{1}{2} x^T Q x \qquad (19)$$

Equivalently, if one defines

$$W(t) \triangleq V(t) - \int_{t_0}^{t} \langle y(\tau), u(\tau) \rangle d\tau \qquad (20)$$

then Eq. (19) verifies that

$$\dot{W} = \dot{V} - \langle y, u \rangle = -\frac{1}{2} x^T Q x$$
$$\leq 0 \qquad (21)$$

as per the definition of passivity stated in Definition 1. (We will come back to Eq. (21) after defining the input-output map.)

Integrating Eq. (19) gives

$$V(t) - V(t_0)$$
$$= \int_{t_0}^{t} \langle y(\tau), u(\tau) \rangle d\tau - \frac{1}{2} \int_{t_0}^{t} x^T(\tau) Q x(\tau) d\tau \qquad (22)$$

Considering the way that the input u is defined in Eq. (16) and the integral relation shown in Eq. (14), a natural choice is to define the variable $v(t)$ as the integral of scaled output, that is,

$$\dot{v}(t) = \alpha(t) y(t) = \alpha(t) C x(t) \qquad (23)$$

With this choice, Eq. (22) can be rewritten by using Eq. (14) as

$$V(t) - V(t_0) = -U(v(t), t) + U(v(t_0), t_0)$$
$$+ \int_{t_0}^{t} \frac{\partial}{\partial \tau} U(v, \tau) d\tau - \frac{1}{2} \int_{t_0}^{t} x^T(\tau) Q x(\tau) d\tau \qquad (24)$$

Note that if an invertible preconditioner matrix M is introduced in the definition of input as $u = -\alpha M \nabla_v U$, then the inner product between input and output does not change by defining $\dot{v} = \alpha M^{-1} y$.

We are now at the stage to relate i) the system variable $v(t)$ to the optimisation variable θ and ii) the auxiliary function U to the objective function f. The purpose is to associate the construction of the passive dynamic system with the design of an optimisation algorithm to solve the problem in Eq. (1). In a similar manner to [3], let us introduce a coordinate transform given by

$$v(t) = \gamma(t) (\theta(t) - \theta_c) \qquad (25)$$

with a strictly monotonically increasing positive-valued function $\gamma(t) \in \mathbb{R}_{>0}$, i.e., $\dot{\gamma}(t) > 0$ for $\forall t \geq t_0$, and some constant $\theta_c \in \mathbb{R}^n$. Since $\theta(v(t), t) = \gamma^{-1}(t) v(t) + \theta_c$, we have

$$\frac{\partial \theta(v, t)}{\partial t} = -\gamma^{-2}(t) \dot{\gamma}(t) v(t)$$
$$= -\gamma^{-1}(t) \dot{\gamma}(t) (\theta(v, t) - \theta_c) \qquad (26)$$

In anticipation of the Bregman divergence $D_f(\cdot, \cdot)$ appearing in the subsequent derivation, let us also define the conjugate function U as

$$U(v(t), t)$$
$$= \gamma(t) [f(\theta(v(t), t)) - f(\theta_c)] + \psi(v(t), t) \qquad (27)$$

with a function $\psi(v(t), t)$ which satisfies $\psi(v(t), t) \geq \underline{\psi}$ for some constant $\underline{\psi}$ and $\frac{\partial}{\partial t}\psi(v(t), t) \leq 0$ for $\forall t \geq t_0$, and the choice of θ_c compatible with the nonnegativity of $f(\theta(v(t), t)) - f(\theta_c)$ such as $\theta_c = \theta_*$. Then, using Eq. (26) in the partial differentiation of Eq. (27) with respect to t yields

$$\frac{\partial}{\partial t} U(v(t), t) = \dot{\gamma}(t) [f(\theta(v(t), t)) - f(\theta_c)]$$
$$+ \gamma(t) \left\langle \nabla_\theta f(\theta), \frac{\partial \theta(v, t)}{\partial t} \right\rangle + \frac{\partial}{\partial t} \psi(v(t), t) \qquad (28)$$
$$= -\dot{\gamma}(t) D_f(\theta_c, \theta(v, t)) + \frac{\partial}{\partial t} \psi(v(t), t)$$
$$\leq 0$$

Note that the gradient of U in Eq. (27) with respect to v can be written as

$$\nabla_v U(v(t), t)$$
$$= \gamma(t) \nabla_v f(\gamma^{-1}(t) v(t) + \theta_c) + \nabla_v \psi(v, t) \qquad (29)$$
$$= \nabla_\theta f(\theta(v(t), t)) + \nabla_v \psi(v, t)$$

Substituting Eq. (28) back into Eq. (24) and rearranging the equation gives a conservation relation as

$$E \equiv U(v(t_0), t_0) + V(t_0)$$
$$= U(v(t), t) + V(t)$$
$$+ \int_{t_0}^{t} \dot{\gamma}(\tau) D_f(\theta_c, \theta(v, \tau)) d\tau$$
$$- \int_{t_0}^{t} \frac{\partial}{\partial \tau} \psi(v, \tau) d\tau + \frac{1}{2} \int_{t_0}^{t} x^T(\tau) Q x(\tau) d\tau \qquad (30)$$
$$= \gamma(t) [f(\theta(v(t), t)) - f(\theta_c)] + \psi(v(t), t)$$
$$+ V(t) + \int_{t_0}^{t} \dot{\gamma}(\tau) D_f(\theta_c, \theta(v, \tau)) d\tau$$
$$- \int_{t_0}^{t} \frac{\partial}{\partial \tau} \psi(v, \tau) d\tau + \frac{1}{2} \int_{t_0}^{t} x^T(\tau) Q x(\tau) d\tau$$

Since $P > 0$ so $V(t) \geq 0$, $Q > 0$, $\psi(v(t), t) \geq \underline{\psi}$ and $\frac{\partial}{\partial t}\psi(v(t), t) \leq 0$ by definition, $\dot{\gamma}(t) > 0$ by definition, and $D_f(\theta_c, \theta) \geq 0$ as discussed in Eq. (3), the conservation law of Eq. (30) indicates that

$$E \geq \gamma(t) [f(\theta(v(t), t)) - f(\theta_c)] + \underline{\psi} \qquad (31)$$

As $\gamma(t) > 0$ for $\forall t \geq t_0$ by definition, Eq. (31) ensures the convergence rate bound for the objective function value as

$$f(\theta(t)) - f(\theta_c) \leq \frac{E - \underline{\psi}}{\gamma(t)}$$

$$= \frac{\gamma_0 [f(\theta_0) - f(\theta_c)] + \psi_0 - \underline{\psi} + \frac{1}{2}x_0^T P x_0}{\gamma(t)} \quad (32)$$

where $\theta(t_0) = \theta_0$, $x(t_0) = x_0$, $\psi(v(t_0), t_0) = \psi_0$, and $\gamma(t_0) = \gamma_0$. One may also establish that

$$E \geq \psi(v(t), t) = \psi\left(\gamma(t)\tilde{\theta}(t), t\right) \quad (33)$$

where $\tilde{\theta}(t) \triangleq \theta(t) - \theta_c$. Note that Eq. (33) can be utilised to certify the boundedness of $\tilde{\theta}(t)$ with appropriate choice of the function ψ. For example, let $\psi(v, t) = \psi\left(\gamma\tilde{\theta}, t\right) = R(t)\left\|\tilde{\theta}\right\|_p^q$ where $R(t) > 0$, $\dot{R}(t) \leq 0$, $p \geq 1$, and $q > 0$. Then, considering the absolute homogeneity of a norm and the fact that $\gamma(t) > 0$ by definition, Eq. (33) ensures the convergence rate bound for the optimisation variable as

$$\left\|\tilde{\theta}(t)\right\|_p \leq \frac{1}{\gamma(t)}\left(\frac{E}{R(t)}\right)^{\frac{1}{q}} \quad (34)$$

It is impossible to obtain the actual optimisation variable θ directly from the relation in Eq. (25). For implementation of the process, we instead opt to the propagation of the following dynamics from a given initial guess θ_0.

$$\dot{\theta}(t) = \frac{d}{dt}\theta(v(t), t) = \frac{d}{dt}\left(\gamma^{-1}(t)v(t)\right)$$
$$= -\gamma^{-2}(t)\dot{\gamma}(t)v(t) + \gamma^{-1}(t)\alpha(t)y(t) \quad (35)$$

As a summary, Fig. 1 shows the block diagram representation of the passivity-based convex optimisation algorithm.
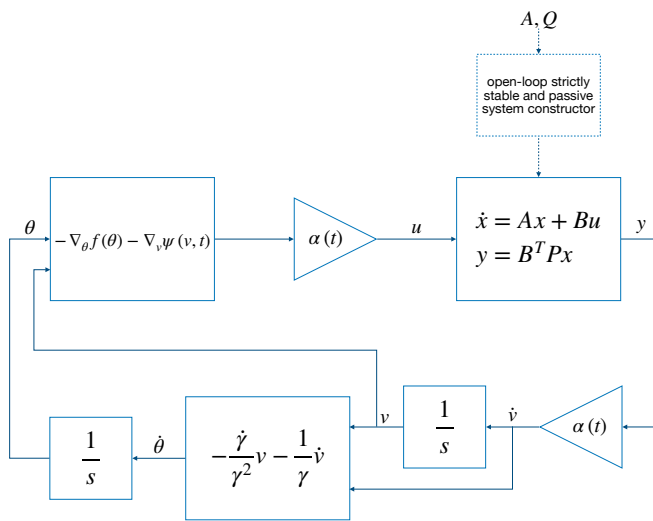

Fig. 1: Block Diagram of Passivity-Based Convex Optimiser

B. Particular Case: $m = 1$

Consider the particular case of $m = 1$. In this case, the definitions of the variables reduce to $x = w$, $A = -a_0 I$ for some $a_0 > 0$, $B = I$, $C = B^T P = P = \frac{Q}{2a_0} > 0$, $y = Px$, and $u = -\alpha(\nabla_\theta f(\theta) + \nabla_v\psi(v, t))$. As a result, the optimisation dynamics can be written as

$$\dot{x}(t) = -a_0 x(t) - \alpha(t)(\nabla_\theta f(\theta) + \nabla_v\psi(v, t)) \quad (36)$$
$$\dot{v}(t) = \alpha(t)Px(t) \quad (37)$$
$$\dot{\theta}(t) = -\gamma^{-2}(t)\dot{\gamma}(t)v(t) + \gamma^{-1}(t)\alpha(t)Px(t) \quad (38)$$

From Eq. (37), we have

$$x = \alpha^{-1}P^{-1}\dot{v}$$
$$\dot{x} = -\alpha^{-2}\dot{\alpha}P^{-1}\dot{v} + \alpha^{-1}P^{-1}\ddot{v} \quad (39)$$

Also, from Eq. (25), we have

$$v = \gamma\tilde{\theta}$$
$$\dot{v} = \dot{\gamma}\tilde{\theta} + \gamma\dot{\theta} \quad (40)$$
$$\ddot{v} = \ddot{\gamma}\tilde{\theta} + 2\dot{\gamma}\dot{\theta} + \gamma\ddot{\theta}$$

Substituting Eq. (39) into Eq. (36) and rearranging the result leads to

$$\ddot{v} + \left(a_0 - \frac{\dot{\alpha}}{\alpha}\right)\dot{v} + P\alpha^2(\nabla_\theta f(\theta) + \nabla_v\psi(v, t)) = 0 \quad (41)$$

Again, substituting Eq. (40) into Eq. (41) and rearranging the result gives

$$\ddot{\theta} + \left(\frac{2\dot{\gamma}}{\gamma} + a_0 - \frac{\dot{\alpha}}{\alpha}\right)\dot{\theta} + \left(\frac{\ddot{\gamma}}{\gamma} + \left(a_0 - \frac{\dot{\alpha}}{\alpha}\right)\frac{\dot{\gamma}}{\gamma}\right)\tilde{\theta}$$
$$+ \frac{P\alpha^2}{\gamma}(\nabla_\theta f(\theta) + \nabla_v\psi(v, t)) = 0 \quad (42)$$

The equation above clearly shows that the optimisation dynamics for this particular case is a second-order dynamics in θ which can still capture many existing algorithms through the choice of a_0, $\gamma(t)$, $\alpha(t)$, and $\psi(v(t), t)$.

1) *Example 1:* If we take $\psi(v(t), t) = \frac{1}{2}\frac{R(t)}{\gamma^2(t)}\|v(t)\|_2^2 + \delta$ for a function $R(t)$ such that $\dot{R}(t) \leq 0$ and some constant δ, then Eq. (42) turns into

$$\ddot{\theta} + \left(\frac{2\dot{\gamma}}{\gamma} + a_0 - \frac{\dot{\alpha}}{\alpha}\right)\dot{\theta}$$
$$+ \left(\frac{\ddot{\gamma}}{\gamma} + \left(a_0 - \frac{\dot{\alpha}}{\alpha}\right)\frac{\dot{\gamma}}{\gamma} + \frac{P\alpha^2 R}{\gamma^2}\right)\tilde{\theta}$$
$$+ \frac{P\alpha^2}{\gamma}\nabla_\theta f(\theta) = 0 \quad (43)$$

2) *Example 2:* If we choose α to be a constant and take $\psi(v, t) = 0$, the time-derivatives of α disappear from Eq. (42), and we get

$$\ddot{\theta} + \frac{2\dot{\gamma} + a_0\gamma}{\gamma}\dot{\theta} + \frac{\ddot{\gamma} + a_0\dot{\gamma}}{\gamma}\tilde{\theta} + \frac{P\alpha^2}{\gamma}\nabla_\theta f(\theta) = 0 \quad (44)$$

3) Example 3: Suppose that $\gamma(t) = \frac{c}{2}\alpha^2(t)$ for some constant $c > 0$. In this setup, we have

$$\frac{\dot{\gamma}}{\gamma} = \frac{2\dot{\alpha}}{\alpha}$$
$$\frac{\ddot{\gamma}}{\gamma} = 2\frac{\dot{\alpha}^2 + \alpha\ddot{\alpha}}{\alpha^2} \tag{45}$$

Plugging Eq. (45) into Eq. (42) brings

$$\ddot{\theta} + \left(a_0 + \frac{3\dot{\alpha}}{\alpha}\right)\dot{\theta} + \frac{2(\ddot{\alpha} + a_0\dot{\alpha})}{\alpha}\tilde{\theta}$$
$$+ \frac{2P}{c}(\nabla_\theta f(\theta) + \nabla_v \psi(v,t)) = 0 \tag{46}$$

Let $\psi(v,t) = 0$ for simplicity and consider the limiting case of $a_0 \to 0$. By choosing $\alpha(t) = kt$ with some nonzero constant k and setting $c = 2P$, Eq. (46) reduces to

$$\ddot{\theta}(t) + \frac{3}{t}\dot{\theta}(t) + \nabla_\theta f(\theta(t)) = 0 \tag{47}$$

which recovers the well-known result presented in [1], [3] as the continuous-time limit of the Nesterov's accelerated gradient descent method. Since $\gamma(t) = Pk^2t^2$ in this case, Eq. (32) with $\theta_c = \theta_*$ certifies the convergence rate as

$$f(\theta(t)) - f_* \leq \frac{E}{Pk^2t^2} \sim \mathcal{O}\left(\frac{1}{t^2}\right) \tag{48}$$

Remark 1. Design parameters of the accelerated gradient flow dynamics in Eq. (42) should be chosen carefully to ensure realisability and robustness against noisy gradients. First, the coefficient of the third term in Eq. (42) should be zero to make the algorithm implementable since $\tilde{\theta}$ cannot be measured. Second, establishing uniform asymptotic stability is advantageous to ensure robustness of the time-varying optimisation dynamics against disturbances [11], [12].

C. Particular Case: $m = 0$

The particular case of $m = 0$ corresponds to the case where the generator $\mathcal{G}$ is a static function, i.e., memoryless map, between the input-output pairs. In this special case, the system $y = \mathcal{G}[u]$ is defined to be passive if $u^T y \geq 0$, and the notion of storage function is no longer necessary. By understanding that Eq. (15) becomes an algebraic relation of the form $w = u$, a passive map can be constructed naturally by defining the output as $y = w = u$ since $u^T y = u^T u \geq 0$. Without changing the definition of the signals and functions introduced to be compatible with the passivity of the input-output pair, the main conservation result given by Eqs. (30) reduces to

$$E \equiv U(v(t_0), t_0)$$
$$= \gamma(t)[f(\theta(v(t),t)) - f(\theta_c)] + \psi(v(t),t)$$
$$+ \int_{t_0}^t \dot{\gamma}(\tau) D_f(\theta_c, \theta(v,\tau)) d\tau - \int_{t_0}^t \frac{\partial}{\partial\tau}\psi(v,\tau) d\tau \tag{49}$$

which verifies

$$E \geq \gamma(t)[f(\theta(v(t),t)) - f(\theta_c)] + \underline{\psi} \tag{50}$$
$$E \geq \psi(v(t),t) \tag{51}$$

Also, the optimisation dynamics in θ can be expressed by combining Eqs. (23), (29), (40), and the relation $y = u$ as

$$\dot{\theta} + \frac{\dot{\gamma}}{\gamma}\tilde{\theta} + \frac{\alpha^2}{\gamma}(\nabla_\theta f(\theta) + \nabla_v \psi(v,t)) = 0 \tag{52}$$

which recovers the simple gradient flow dynamics given by $\dot{\theta} + \nabla_\theta f(\theta) = 0$ if $\gamma = \alpha^2$ is constant and $\psi(v,t) = 0$.

IV. EXTENSION TO ADAPTIVE CONTROL

Adaptive control deals with stable tracking of an uncertain dynamic system by means of online parameter update. Hence, the corresponding closed-loop dynamics generally consists of tracking error and parameter estimation error subsystems. If the uncertain part of the plant dynamics can be represented with a linear parametric model, parameter update for stable adaptive control can be interpreted as a convex optimisation process which involves interaction with an external system.

Consider the common tracking error model of an adaptive control system given by

$$\dot{e}(t) = A_m e(t) + B_p \Phi(x_p)\tilde{\theta}(t) \tag{53}$$

where A_m is a constant Hurwitz system matrix of the closed-loop plant dynamics, B_p is the constant control effectiveness matrix of the plant, Φ represents the known basis function for the linear model of the plant uncertainty, and $x_p(t)$ denotes the plant state. It is assumed that the true uncertainty is represented with a constant true parameter $\theta_* = \theta_c$. Lemma 1 states that there exists $P_e = P_e^T > 0$ verifying $A_m^T P_e + P_e A_m = -Q_e$ for any given $Q_e = Q_e^T > 0$ and the corresponding Lyapunov function is $V_e \triangleq \frac{1}{2}e^T P_e e$. The time-derivative of the Lyapunov function can be written as

$$\dot{V}_e = -\frac{1}{2}e^T Q_e e + e^T P_e B_p \Phi\tilde{\theta}$$
$$= -\frac{1}{2}e^T Q_e e + \langle\tilde{\theta}, \nabla_\theta \dot{V}_e\rangle \tag{54}$$

One may notice that $\dot{V}_e$ alone is sign-indefinite.

To design a parameter update algorithm based on accelerated gradient flow, it is necessary to allow the optimisation dynamics to have a port for interaction with exogenous signal. With this background, let us redefine the output of the generator $\mathcal{G}$ inside the optimiser by introducing a new variable z to replace Eq. (23) with

$$\dot{v}(t) + z(t) = \alpha(t)y(t) \tag{55}$$

while keeping the input definition unchanged as shown in Eq. (16). With this setup, Eq. (21) can be rewritten as

$$\dot{W} = \dot{V} - \langle y, u\rangle = \dot{V} + \langle\dot{v} + z, \nabla_v U(v,t)\rangle$$
$$= \dot{V} + \frac{d}{dt}U(v,t) - \frac{\partial}{\partial t}U(v,t) + \langle z, \nabla_v U(v,t)\rangle$$
$$= -\frac{1}{2}x^T Q x \tag{56}$$
$$\leq 0$$

Since we already have Eq. (28) that states $\frac{\partial}{\partial t} U(v,t) \leq 0$, the following inequality also holds

$$
\begin{aligned}
\dot{W} + \frac{\partial}{\partial t} U(v,t) &= \dot{V} + \frac{d}{dt} U(v,t) + \langle z, \nabla_v U(v,t) \rangle \\
&= -\frac{1}{2} x^T Q x - \dot{\gamma} D_f(\theta_c, \theta) + \frac{\partial}{\partial t} \psi(v,t) \\
&\leq 0
\end{aligned}
\tag{57}
$$

In the followings, let $\psi(v,t) = 0$ for simplicity, the choice which verifies $\nabla_v U(v,t) = \nabla_\theta f(\theta)$ according to Eq. (29). Then, by rearranging Eq. (57), we have

$$
\begin{aligned}
\dot{V} &+ \frac{d}{dt} U(v,t) \\
&= -\frac{1}{2} x^T Q x - \dot{\gamma} D_f(\theta_c, \theta) - \langle z, \nabla_\theta f(\theta) \rangle \\
&\leq - \langle z, \nabla_\theta f(\theta) \rangle
\end{aligned}
\tag{58}
$$

Now, let us define the Lyapunov function for the total closed-loop system by

$$
V_T \triangleq V + U + V_e
\tag{59}
$$

Also, let us define the additional variable as $z = \tilde{\theta} = \frac{v}{\gamma}$ and choose the objective function as $f(\theta) = \dot{V}_e + L(\theta)$ where L is a loss function representing the goal of uncertainty learning which should be measurable. One example for the learning loss based on regressor extension is $L(\theta) = \frac{1}{2} \tilde{Y}^T \tilde{Y} = \frac{1}{2} \tilde{\theta}^T \Omega^2 \tilde{\theta}$ where $\Omega = \mathcal{F}[\Phi^T \Phi]$ with some stable linear filter $\mathcal{F}$. Then, by using Eqs. (54) and (58), the time-derivative of Eq. (59) satisfies

$$
\begin{aligned}
\dot{V}_T &= -\frac{1}{2} x^T Q x - \frac{1}{2} e^T Q_e e - \dot{\gamma} D_f(\theta_c, \theta) - \tilde{\theta}^T \Omega^2 \tilde{\theta} \\
&\leq 0
\end{aligned}
\tag{60}
$$

for $\forall t \geq t_0$. Integrating Eq. (60) yields

$$
\begin{aligned}
E_{AC} &\equiv V(t_0) + U(v(t_0), t_0) + V_e(t_0) \\
&\geq V(t) + U(v(t), t) + V_e(t) \\
&\geq U(v(t), t) \\
&= \gamma(t)[f(\theta(t)) - f(\theta_c)]
\end{aligned}
\tag{61}
$$

As $\gamma(t) > 0$ for $\forall t \geq t_0$ by definition, Eq. (61) ensures the convergence rate bound for the objective function value as

$$
f(\theta(t)) - f(\theta_c) \leq \frac{E_{AC}}{\gamma(t)}
\tag{62}
$$

If the plant controller is driven by an exogenous reference command under which the system experiences a sufficient but not necessarily persistent excitation, then we can introduce a mild assumption that $\Omega > 0$ will be satisfied after some finite time t_e. Therefore, assuming the presence of interval excitation, Eq. (60) indicates the negative definiteness of $\dot{V}_T$ for $\forall t \geq t_e$. Consequentially, the convergence of $\left(x, e, \tilde{\theta} \right)$ towards $(0, 0, 0)$ can be established.

V. Conclusion

This study developed a systematic method for constructing an accelerated algorithm in continuous-time domain to solve convex optimisation problems. The proposed method takes a control-theoretic approach which is to synthesise a passive dynamics as an internal generator with compatible choice of the input-output pair. The classical notion of passivity along with the Kalman-Yakubovich-Popov lemma plays a key role not only in the definition of output of the generator dynamics but also in the characterisation of the storage function. The convergence rate bound for the function value was established in relation to the time-varying factor that relates the integral of scaled output with the optimisation variable. As an example, the ordinary differential equation model for the Nesterov's accelerated gradient descent method is shown as a particular case. The passivity-based optimisation methodology has its significance as i) a theoretical bridge between optimisation and control and also as ii) a generic framework that can be utilised to design convex optimisation algorithms.

References

[1] W. Su, S. Boyd, and E. J. Candès, "A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights," *Journal of Machine Learning Research*, vol. 17, no. 153, pp. 1–43, 2016. [Online]. Available: http://jmlr.org/papers/v17/15-084.html

[2] J. E. Gaudio, A. M. Annaswamy, M. A. Bolender, E. Lavretsky, and T. E. Gibson, "A Class of High Order Tuners for Adaptive Systems," *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 391–396, 2021.

[3] J. J. Suh, G. Roh, and E. K. Ryu, "Continuous-Time Analysis of Accelerated Gradient Methods via Conservation Laws in Dilated Coordinate Systems," in *39th International Conference on Machine Learning*, Baltimore, MD, USA, July 2022, pp. 20 640–20 667. [Online]. Available: https://proceedings.mlr.press/v162/suh22a.html

[4] B. Hu and L. Lessard, "Control Interpretations for First-Order Optimization Methods," in *American Control Conference*, Seattle, WA, USA, May 2017, pp. 3114–3119.

[5] ——, "Dissipativity Theory for Nesterov's Accelerated Method," in *34th International Conference on Machine Learning*, Sydney, Australia, August 2017, pp. 1549–1557. [Online]. Available: https://proceedings.mlr.press/v70/hu17a.html

[6] L. Lessard, "The Analysis of Optimization Algorithms: A Dissipativity Approach," *IEEE Control Systems Magazine*, vol. 42, no. 3, pp. 58–72, 2022.

[7] H. K. Khalil, *Nonlinear Control*. Harlow, Essex, UK: Pearson Education, 2015.

[8] N. Kottenstette, M. J. McCourt, M. Xia, V. Gupta, and P. J. Antsaklis, "On Relationships Among Passivity, Positive Realness, and Dissipativity in Linear Systems," *Automatica*, vol. 50, no. 4, pp. 1003–1016, 2014.

[9] D. d. S. Madeira and J. Adamy, "On the Equivalence Between Strict Positive Realness and Strict Passivity of Linear Systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 3091–3095, 2016.

[10] J.-J. E. Slotine and W. Li, *Applied Nonlinear Control*. Englewood Cliffs, NJ, USA: Prentice Hall, 1991, ch. 4, pp. 126–142.

[11] J. I. Poveda and N. Li, "Inducing Uniform Asymptotic Stability in Non-Autonomous Accelerated Optimization Dynamics via Hybrid Regularization," in *58th IEEE Conference on Decision and Control*, Nice, France, December 2019, pp. 3000–3005.

[12] J. I. Poveda and A. R. Teel, "The Heavy-Ball ODE with Time-Varying Damping: Persistence of Excitation and Uniform Asymptotic Stability," in *American Control Conference*, Denver, CO, USA, July 2020, pp. 773–778.