Convergence Analysis of ODE Models for Accelerated First-Order Methods via Positive Semidefinite Kernels: Supplementary Material

Jungbin Kim
Seoul National University
kjb2952@snu.ac.kr

Insoon Yang
Seoul National University
insoonyang@snu.ac.kr

A Basic Volterra theory

In this section, we review the basic theory of Volterra integral equations. For a more comprehensive treatment of this subject, see [2, Chapter 2]. A *Volterra integral kernel* is an integral kernel k defined on $D := \{(t, \tau) : 0 \leq \tau \leq t \leq T\}$. We assume that k is continuous on D. A *Volterra integral operator* $\mathcal{V} : C([0, T]; \mathbb{R}^d) \to C([0, T]; \mathbb{R}^d)$ associated with the kernel k is defined as follows:

$$(\mathcal{V}f)(t) = \int_0^t k(t, \tau) f(\tau)\, d\tau. \tag{32}$$

Definition 1 ([2, Definition 2.1.1]). *The* resolvent kernel $R : D \to \mathbb{R}$ *corresponding to the given Volterra integral kernel k is defined by either of the following resolvent equations:*

$$R(t, \tau) = k(t, \tau) + \int_\tau^t k(t, v) R(v, \tau)\, dv \tag{33a}$$

$$R(t, \tau) = k(t, \tau) + \int_\tau^t R(t, v) k(v, \tau)\, dv. \tag{33b}$$

Proposition 5 ([2, Theorem 2.1.2]). *Let R be the resolvent kernel corresponding to the Volterra integral kernel k. Then, for any $g \in C([0, T]; \mathbb{R}^d)$, the integral equation*

$$g(t) = y(t) - (\mathcal{V}y)(t)$$

has a unique solution $y \in C([0, T]; \mathbb{R}^d)$, and this solution is given by

$$y(t) = g(t) + \int_0^t R(t, s) g(s)\, ds.$$

B Technical lemmas

In this section, we present lemmas that will be used in the proof of Theorem 1. The following lemma can be regarded as a continuous-time analogue of [4, Lemma 1].

Lemma 1. *Let $k \in L^2([0, T]^2; \mathbb{R})$ be a positive semidefinite Hilbert-Schmidt kernel. Let K^1 and K^d be the Hilbert-Schmidt integral operators on $L^2([0, T]; \mathbb{R})$ and $L^2([0, T]; \mathbb{R}^d)$, respectively, which are associated with the kernel k. Let $a \in L^2([0, T]; \mathbb{R})$ and let $v = (v_1, \dots, v_d)$ be a unit vector in $\mathbb{R}^d$. Then, we have*

$$\inf_{x \in L^2([0,T];\mathbb{R}^d)} \left\{ \frac{1}{2} \langle K^d x, x \rangle + \langle a(t) v, x(t) \rangle \right\} = \inf_{\xi \in L^2([0,T];\mathbb{R})} \left\{ \frac{1}{2} \langle K^1 \xi, \xi \rangle + \langle a, \xi \rangle \right\},$$

where the inner product on the left-hand side is defined in $L^2([0, T]; \mathbb{R}^d)$, and the inner product on the right-hand side is defined in $L^2([0, T]; \mathbb{R})$.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

Proof. We define a quadratic function P on $L^2([0,T];\mathbb{R}^d)$ as follows:

$$\begin{aligned} P(x) &= \frac{1}{2}\left\langle K^d x, x\right\rangle_{L^2([0,T];\mathbb{R}^d)} + \langle a(t)v, x(t)\rangle_{L^2([0,T];\mathbb{R}^d)} \\ &= \sum_{i=1}^d \left(\frac{1}{2}\left\langle K^1 x_i, x_i\right\rangle_{L^2([0,T];\mathbb{R})} + v_i \left\langle a(t), x_i(t)\right\rangle_{L^2([0,T];\mathbb{R})}\right). \end{aligned}$$

Then, for any $\xi \in L^2([0,T];\mathbb{R})$, we have

$$\begin{aligned} &\frac{1}{2}\langle K^1\xi, \xi\rangle_{L^2([0,T];\mathbb{R})} + \langle a, \xi\rangle_{L^2([0,T];\mathbb{R})} \\ &= \sum_{i=1}^d v_i^2 \left(\frac{1}{2}\langle K^1\xi, \xi\rangle_{L^2([0,T];\mathbb{R})} + \langle a, \xi\rangle_{L^2([0,T];\mathbb{R})}\right) \\ &= \sum_{i=1}^d \left(\frac{1}{2}\left\langle K^1\left(v_i\xi\right), v_i\xi\right\rangle_{L^2([0,T];\mathbb{R})} + v_i\left\langle a(t), v_i\xi(t)\right\rangle_{L^2([0,T];\mathbb{R})}\right) \\ &= P\left(v_1\xi, \ldots, v_d\xi\right). \end{aligned}$$

Thus, we have

$$\begin{aligned} \inf_{\xi\in L^2([0,T];\mathbb{R})}\left\{\frac{1}{2}\langle K^1\xi,\xi\rangle + \langle a,\xi\rangle\right\} &= \inf_{\xi\in L^2([0,T];\mathbb{R})} P\left(v_1\xi,\ldots,v_d\xi\right) \\ &\geq \inf_{x\in L^2([0,T];\mathbb{R}^d)} P(x) \\ &= \inf_{x\in L^2([0,T];\mathbb{R}^d)}\left\{\frac{1}{2}\langle K^d x, x\rangle + \langle a(t)v, x(t)\rangle\right\}. \end{aligned}$$

We now prove that the reverse direction of this inequality also holds. Let $\mathcal{I} = \{i \in \{1,\ldots,d\} : v_i \neq 0\}$. Then, for any $x \in L^2([0,T];\mathbb{R}^d)$, we have

$$\begin{aligned} P(x) &= \sum_{i\in\mathcal{I}}\left(\frac{1}{2}\left\langle K^1 x_i, x_i\right\rangle + v_i\left\langle a, x_i\right\rangle\right) + \sum_{i\notin\mathcal{I}}\left(\frac{1}{2}\left\langle K^1 x_i, x_i\right\rangle\right) \\ &\geq \sum_{i\in\mathcal{I}} v_i^2\left(\frac{1}{2}\left\langle K^1\frac{x_i}{v_i}, \frac{x_i}{v_i}\right\rangle + \left\langle a, \frac{x_i}{v_i}\right\rangle\right) \\ &\geq \sum_{i\in\mathcal{I}}\left(v_i^2 \inf_{\xi\in L^2([0,T];\mathbb{R})}\left(\frac{1}{2}\left\langle K^1\xi,\xi\right\rangle + \langle a,\xi\rangle\right)\right) \\ &= \inf_{\xi\in L^2([0,T];\mathbb{R})}\left\{\frac{1}{2}\langle K^1\xi,\xi\rangle + \langle a,\xi\rangle\right\}. \end{aligned}$$

Taking the infimum of both sides yields

$$\inf_{x\in L^2([0,T];\mathbb{R}^d)}\left\{\frac{1}{2}\langle K^d x, x\rangle + \langle a(t)v, x(t)\rangle\right\} \geq \inf_{\xi\in L^2([0,T];\mathbb{R})}\left\{\frac{1}{2}\langle K^1\xi,\xi\rangle + \langle a,\xi\rangle\right\},$$

which completes the proof. □

The following lemma shows that the non-negativity of a quadratic function on $L^2([0,T];\mathbb{R}^d)$ can be translated into the positive semidefininteness of a specific Hilbert-Schmidt kernel.

Lemma 2. *Let $k \in L^2([0,T]^2;\mathbb{R})$ be a symmetric Hilbert-Schmidt kernel and K be the corresponding Hilbert-Schmidt integral operator on $L^2([0,T];\mathbb{R}^d)$. Let $b \in L^2([0,T];\mathbb{R}^d)$ and $c > 0$. Then, the inequality*

$$Q(x) = \frac{1}{2}\langle Kx, x\rangle_{L^2([0,T];\mathbb{R}^d)} + \langle b, x\rangle_{L^2([0,T];\mathbb{R}^d)} + c \geq 0 \tag{34}$$

holds for all $x \in L^2([0,T];\mathbb{R}^d)$ if and only if the following kernel is positive semidefinite:

$$(t,\tau) \mapsto ck(t,\tau) - \frac{1}{2}b(t)b(\tau). \tag{35}$$

Proof. ($\Leftarrow$) Assume that the kernel (35) is positive semidefinite. Then, for any $x \in L^2([0,T];\mathbb{R}^d)$, we have

$$\begin{aligned}0 &\le \int_0^T \int_0^T \left(ck(t,\tau) - \frac{1}{2} b(t)b(\tau) \right) x(t)x(\tau)\,dtd\tau \\ &= c\int_0^T \int_0^T k(t,\tau)x(t)x(\tau)\,dtd\tau - \frac{1}{2}\left(\int_0^T b(t)x(t)\,dt\right)^2 \\ &= c\langle Kx, x\rangle - \frac{1}{2}\langle b, x\rangle^2.\end{aligned}$$

Thus, we have

$$\frac{1}{2}\langle Kx, x\rangle + \langle b, x\rangle + c = \frac{1}{2c}\left(c\langle Kx, x\rangle - \frac{1}{2}\langle b, x\rangle^2\right) + c\left(\frac{1}{2c}\langle b, x\rangle + 1\right)^2 \ge 0. \tag{36}$$

($\Rightarrow$) We prove the contrapositive of the statement. Assume that the kernel (35) is not positive semidefinite, i.e., there exists $x \in L^2([0,T];\mathbb{R}^d)$ such that $c\langle Kx, x\rangle - \frac{1}{2}\langle b, x\rangle^2 < 0$. We consider two cases: when (i) $\langle b, x\rangle = 0$ and (ii) $\langle b, x\rangle \neq 0$. If $\langle b, x\rangle = 0$, then we have $\langle Kx, x\rangle < 0$. For an arbitrary $\alpha \in \mathbb{R}$, we have $Q(\alpha x) = \frac{\alpha^2}{2}\langle Kx, x\rangle + c$, which is negative for a sufficiently large α. If $\langle b, x\rangle \neq 0$, then we have $\frac{1}{2c}\langle b, \alpha x\rangle + 1 = 0$ for some $\alpha \in \mathbb{R}$. For such α, by the equality in (36), we have $Q(\alpha x) = \frac{\alpha^2}{2c}\left(c\langle Kx, x\rangle - \frac{1}{2}\langle b, x\rangle^2\right) < 0$. □

C Proof of Theorem 1

To prove the theorem, we introduce a generalization of the continuous PEP presented in Section 3, which aims to obtain a convergence rate on $\tilde{f}(X(T)) - \tilde{f}(x^*)$, where $\tilde{f}(x) = f(x) - \frac{\mu}{2}\|x - x^*\|^2$. Note that the continuous PEP presented here covers the continuous PEP in Section 3 as a special case when $\mu = 0$. In order to prevent any notational overlap, we denote the constant ν given in the theorem statement by ν_{given}.

Consider the following dynamical system:

$$\dot{X}(t) = -\int_0^t H(t,\tau)\nabla f(X(\tau))\,d\tau, \tag{37}$$

In Appendix C.1, we show its equivalence to the following form:

$$\dot{X}(t) = -\int_0^t H^F(t,\tau)\nabla \hat{f}(X(\tau))\,d\tau, \tag{38}$$

where $\hat{f}(x) := f(x) - \frac{\mu}{2}\|x - x_0\|^2$. Suppose we want to obtain a convergence guarantee in the form of

$$\tilde{f}(X(T)) - \tilde{f}(x^*) \le \lambda^F(0)\left(\tilde{f}(x_0) - \tilde{f}(x^*)\right) + \rho\,\|x_0 - x^*\|^2, \tag{39}$$

where λ^F is the given function in the theorem statement. If we define the exact PEP as follows:

$$\begin{aligned}\max_{\substack{f\in\mathcal{F}_0(\mathbb{R}^d;\mathbb{R})\\ X\in C^1([0,T];\mathbb{R}^d)}} \quad & \frac{\tilde{f}(X(T)) - \tilde{f}(x^*)}{\|x_0 - x^*\|^2} - \lambda^F(0)\frac{\tilde{f}(x_0) - \tilde{f}(x^*)}{\|x_0 - x^*\|^2} \\ \text{subject to} \quad & X \text{ is a solution to (38) with } X(0) = x_0 \\ & x^* \text{ is a minimizer of } f,\end{aligned} \tag{Exact PEP-F}$$

then the convergence guarantee (39) holds with $\rho = \mathrm{val}(\text{Exact PEP-F})$.

We now relax this problem by using the technique outlined in Sections 3.1 and 3.2. We first observe that

$$\begin{aligned}\nabla\tilde{f}(x) &= \nabla f(x) - \mu(x - x^*) \\ &= \nabla f(x) - \mu(x - x_0) + \mu(x^* - x_0)\end{aligned}$$

$$= \nabla \hat{f}(x) + \mu(x^* - x_0).$$

Define two functions $\varphi : [0,T] \to \mathbb{R}$ and $\gamma : [0,T] \to \mathbb{R}^d$ as follows:

$$\begin{aligned}
\varphi(t) &= \frac{1}{\|x_0 - x^*\|^2}\left(\tilde{f}(X(t)) - \tilde{f}(x^*)\right), \\
\gamma(t) &= \frac{1}{\|x_0 - x^*\|}\nabla \tilde{f}(X(t)) = \frac{1}{\|x_0 - x^*\|}\nabla \hat{f}(X(t)) + \mu v,
\end{aligned}$$

where $v = (x^* - x_0)/\|x^* - x_0\|$. Then, we can derive the following equalities and inequalities:

$$\begin{aligned}
\dot{\varphi}(t) &= \frac{1}{\|x_0 - x^*\|^2}\langle \nabla \tilde{f}(X(t)), \dot{X}(t)\rangle \\
&= -\left\langle \gamma(t), \int_0^t H^F(t,\tau)\left(\gamma(\tau) - \mu v\right) d\tau \right\rangle, \\
\varphi(t) &\le \frac{1}{\|x_0 - x^*\|^2}\langle \nabla \tilde{f}(X(t)), X(t) - x^*\rangle \\
&= \frac{1}{\|x_0 - x^*\|^2}\langle \nabla \tilde{f}(X(t)), X(t) - x_0 + x_0 - x^*\rangle \\
&= \frac{1}{\|x_0 - x^*\|^2}\left\langle \nabla \tilde{f}(X(t)), \int_0^t \dot{X}(s)\, ds + x_0 - x^* \right\rangle \\
&= -\left\langle \gamma(t), v + \int_0^t \int_0^s H^F(s,\tau)\left(\gamma(\tau) - \mu v\right) d\tau ds \right\rangle \\
&= -\left\langle \gamma(t), v + \int_0^t \int_\tau^t H^F(s,\tau)\left(\gamma(\tau) - \mu v\right) ds d\tau \right\rangle.
\end{aligned} \tag{40}$$

Thus, Exact PEP-F can be relaxed as follows:

$$\begin{aligned}
\max_{\substack{\varphi \in C^1([0,T];\mathbb{R}) \\ \gamma \in C\left([0,T];\mathbb{R}^d\right) \\ v \in \mathbb{R}^d,\ \|v\|=1}} \quad & \varphi(T) - \lambda^F(0)\varphi(0) \\
\text{subject to} \quad & \dot{\varphi}(t) + \left\langle \gamma(t), \int_0^t H^F(t,\tau)\left(\gamma(\tau) - \mu v\right) d\tau \right\rangle = 0 \qquad \forall t \in (0,T) \\
& \varphi(t) + \left\langle \gamma(t), v + \int_0^t \int_\tau^t H^F(s,\tau)\left(\gamma(\tau) - \mu v\right) ds d\tau \right\rangle \le 0 \quad \forall t \in (0,T).
\end{aligned} \tag{Relaxed PEP-F}$$

Since for any feasible solution to Exact PEP-F, there is a corresponding feasible solution to (Relaxed PEP-F) with the same objective value, we have $\mathrm{val}(\text{Relaxed PEP-F}) \ge \mathrm{val}(\text{Exact PEP-F})$. Therefore, the convergence guarantee (39) holds with $\rho = \mathrm{val}(\text{Relaxed PEP-F})$.

To obtain an upper bound of $\mathrm{val}(\text{Relaxed PEP-F})$, we use Lagrangian duality. With the two *Lagrange multipliers* $\lambda_1 \in C^1([0,T];\mathbb{R})$ and $\lambda_2 \in C([0,T];[0,\infty))$, the Lagrangian function $\mathcal{L}$ is defined as

$$\begin{aligned}
&\mathcal{L}(\varphi, \gamma, v; \lambda_1, \lambda_2) \\
&= \varphi(T) - \lambda^F(0)\varphi(0) - \int_0^T \lambda_1(t)\left(\dot{\varphi}(t) + \left\langle \gamma(t), \int_0^t H^F(t,\tau)\left(\gamma(\tau) - \mu v\right) d\tau \right\rangle_{\mathbb{R}^d}\right) dt \\
&\quad - \int_0^T \lambda_2(t)\left(\varphi(t) + \left\langle \gamma(t), v + \int_0^t \int_\tau^t H^F(s,\tau)\left(\gamma(\tau) - \mu v\right) ds d\tau \right\rangle_{\mathbb{R}^d}\right) dt \\
&= \varphi(T) - \lambda^F(0)\varphi(0) - \int_0^T \lambda_1(t)\left(\dot{\varphi}(t) + \left\langle \gamma(t), \int_0^t H^F(t,\tau)\gamma(\tau)\, d\tau \right\rangle_{\mathbb{R}^d}\right) dt \\
&\quad - \int_0^T \lambda_2(t)\left(\varphi(t) + \left\langle \gamma(t), \int_0^t \int_\tau^t H^F(s,\tau)\gamma(\tau)\, ds d\tau \right\rangle_{\mathbb{R}^d}\right) dt \\
&\quad - \int_0^T \left(-\mu\lambda_1(t)\int_0^t H^F(t,\tau)\, d\tau\right)\langle \gamma(t), v\rangle_{\mathbb{R}^d}\, dt - \int_0^T \lambda_2(t)\,\langle \gamma(t), v\rangle_{\mathbb{R}^d}\, dt
\end{aligned}$$

$$
\begin{aligned}
&\quad - \int_0^T \left(-\mu\lambda_2(t) \int_0^t \int_\tau^t H^F(s,\tau)\, dsdt \right) \langle \gamma(t), v\rangle_{\mathbb{R}^d}\ dt \\
&= \mathcal{L}_1(\varphi;\lambda_1,\lambda_2) + \mathcal{L}_2(\gamma, v;\lambda_1,\lambda_2),
\end{aligned}
$$

where

$$
\mathcal{L}_1(\varphi;\lambda_1,\lambda_2) = \varphi(T) - \lambda^F(0)\varphi(0) - \int_0^T \lambda_1(t)\dot{\varphi}(t)\, dt - \int_0^T \lambda_2(t)\varphi(t)\, dt
$$

and

$$
\begin{aligned}
\mathcal{L}_2(\gamma, v;\lambda_1,\lambda_2) = &- \int_0^T \lambda_1(t) \left\langle \gamma(t), \int_0^t H^F(t,\tau)\gamma(\tau)\, d\tau \right\rangle_{\mathbb{R}^d} dt \\
&- \int_0^T \lambda_2(t) \left\langle \gamma(t), \int_0^t \int_\tau^t H^F(s,\tau)\gamma(\tau)\, dsd\tau \rangle \right\rangle_{\mathbb{R}^d} dt \\
&- \int_0^T \left(-\mu\lambda_1(t) \int_0^t H^F(t,\tau)\, d\tau \right) \langle \gamma(t), v\rangle_{\mathbb{R}^d}\ dt \\
&- \int_0^T \lambda_2(t) \langle \gamma(t), v\rangle_{\mathbb{R}^d}\ dt \\
&- \int_0^T \left(-\mu\lambda_2(t) \int_0^t \int_\tau^t H^F(s,\tau)\, dsdt \right) \langle \gamma(t), v\rangle_{\mathbb{R}^d}\ dt.
\end{aligned}
$$

When expressed in the inner products in function spaces, these functions can be written as

$$
\begin{aligned}
\mathcal{L}_1(\varphi;\lambda_1,\lambda_2) &= \varphi(T) - \lambda^F(0)\varphi(0) - \langle \lambda_1, \dot{\varphi}\rangle_{L^2([0,T];\mathbb{R})} - \langle \lambda_2, \varphi\rangle_{L^2([0,T];\mathbb{R})} \\
\mathcal{L}_2(\gamma, v;\lambda_1,\lambda_2) &= -\frac{1}{2} \left\langle K^d\gamma, \gamma\right\rangle_{L^2([0,T];\mathbb{R}^d)} - 2 \left\langle \alpha(t)v, \gamma(t)\right\rangle_{L^2([0,T];\mathbb{R}^d)},
\end{aligned}
$$

where K^d is the Hilbert-Schmidt integral operator on $L^2([0,T];\mathbb{R}^d)$, associated with the symmetric kernel k defined by

$$
k(t,\tau) = \lambda_1(t)H(t,\tau) + \lambda_2(t) \int_\tau^t H(s,\tau)\, ds, \quad t \geq \tau, \tag{41}
$$

and α is a function of time, defined as

$$
\alpha(t) = \frac{1}{2} \left(-\mu\lambda_1(t) \int_0^t H^F(t,\tau)\, d\tau + \lambda_2(t) \left(1 - \mu \int_0^t \int_\tau^t H^F(s,\tau)\, dsd\tau \right) \right). \tag{42}
$$

Let the dual function be defined as $\mathrm{Dual}(\lambda_1,\lambda_2) = \sup_{\varphi,\gamma,v} \mathcal{L}(\varphi,\gamma,v;\lambda_1,\lambda_2)$. By weak duality, for any feasible solution (λ_1,λ_2) to the dual problem, we have $\mathrm{Dual}(\lambda_1,\lambda_2) \geq \mathrm{val}(\text{Relaxed PEP-F})$. Consequently, we have the convergence guarantee (39) with $\rho = \mathrm{Dual}(\lambda_1,\lambda_2)$.

We proceed by computing the dual objective function. Because the function $\mathcal{L}_1$ can be written as

$$
\begin{aligned}
\mathcal{L}_1(\varphi,\lambda_1,\lambda_2) &= \varphi(T) - \lambda^F(0)\varphi(0) - \int_0^T \lambda_1(t)\dot{\varphi}(t)\, dt - \int_0^T \lambda_2(t)\varphi(t)\, dt \\
&= \varphi(T) - \lambda^F(0)\varphi(0) - [\lambda_1(t)\varphi(t)]_{t=0}^T + \int_0^T \left(\dot{\lambda}_1(t) - \lambda_2(t) \right) \varphi(t)\, dt \\
&= (1 - \lambda_1(T))\, \varphi(T) + \left(\lambda_1(0) - \lambda^F(0) \right) \varphi(0) + \int_0^T \left(\dot{\lambda}_1(t) - \lambda_2(t) \right) \varphi(t)\, dt,
\end{aligned}
$$

we can see that if any of the following conditions holds: (i) $\lambda_1(0) \neq \lambda^F(0)$, (ii) $\lambda_1(T) \neq 1$, or (iii) $\dot{\lambda}_1(t) \neq \lambda_2(t)$ for some t, then $\sup_\varphi \mathcal{L}_1(\varphi;\lambda_1,\lambda_2) = \infty$, which implies $\mathrm{Dual}(\lambda_1,\lambda_2) = \infty$. On the other hand, if $\lambda_1(0) = \lambda^F(0)$, $\lambda_1(T) = 1$, and $\dot{\lambda}_1(t) = \lambda_2(t)$ for all t, then we have $\sup_\varphi \mathcal{L}_1(\varphi;\lambda_1,\lambda_2) = 0$. In this case, using Lemma 1, we can compute the dual objective function as follows:

$$
\sup_{\varphi,\gamma,v} \mathcal{L}(\varphi,\gamma,v;\lambda_1,\lambda_2)
$$

$$
\begin{aligned}
&= \sup_{\gamma \in L^2([0,T];\mathbb{R}^d), v\in\mathbb{R}^d, \|v\|=1} \mathcal{L}_2(\gamma, v; \lambda_1, \lambda_2) \\
&= - \inf_{\gamma \in L^2([0,T];\mathbb{R}^d), v\in\mathbb{R}^d, \|v\|=1} \left\{ \frac{1}{2} \left\langle K^d \gamma, \gamma \right\rangle_{L^2([0,T];\mathbb{R}^d)} + 2 \left\langle \alpha(t) v, , \gamma(t) \right\rangle_{L^2([0,T];\mathbb{R}^d)} \right\} \\
&= - \inf_{\xi \in L^2([0,T];\mathbb{R})} \left\{ \frac{1}{2} \left\langle K^1 \xi, \xi \right\rangle_{L^2([0,T];\mathbb{R})} + 2 \left\langle \alpha, \xi \right\rangle_{L^2([0,T];\mathbb{R})} \right\}.
\end{aligned}
$$

where K^1 is the Hilbert-Schmidt integral operator on $L^2([0,T];\mathbb{R})$, associated with the symmetric kernel k defined in (41). Using Lemma 2, we further simplify as

$$
\begin{aligned}
\sup_{\varphi,\gamma,v} \mathcal{L}(\varphi, \gamma, v; \lambda_1, \lambda_2) &= - \sup_{\nu\in\mathbb{R}} \left\{ -\nu : \frac{1}{2} \left\langle K^1 \xi, \xi \right\rangle_{L^2([0,T];\mathbb{R})} + 2 \left\langle \alpha, \xi \right\rangle_{L^2([0,T];\mathbb{R})} + \nu \geq 0 \right\} \\
&= \inf_{\nu\in\mathbb{R}} \left\{ \nu : \frac{1}{2} \left\langle K^1 \xi, \xi \right\rangle_{L^2([0,T];\mathbb{R})} + 2 \left\langle \alpha, \xi \right\rangle_{L^2([0,T];\mathbb{R})} + \nu \geq 0 \right\} \\
&= \inf_{\nu\in\mathbb{R}} \left\{ \nu : S_{\lambda_1,\lambda_2,\nu} \succeq 0 \right\}
\end{aligned}
$$

with $S_{\lambda_1,\lambda_2,\nu}(t,\tau) = \nu k(t,\tau) - 2\alpha(t)\alpha(\tau)$, where the kernel k is defined in (41) and the function α is defined in (42). Therefore, the Lagrangian dual objective function of Relaxed PEP-F can be written as

$$
\operatorname{Dual}(\lambda_1, \lambda_2) = \begin{cases} \inf_{\nu\in\mathbb{R}} \left\{ \nu : S_{\lambda_1,\lambda_2,\nu} \succeq 0 \right\} & \text{if } \lambda_1(0) = 0,\ \lambda_1(T) = 1, \dot{\lambda}_1(t) = \lambda_2(t) \\ \infty & \text{otherwise.} \end{cases} \tag{43}
$$

To complete the proof of Theorem 1, we note that weak duality implies $\operatorname{Dual}(\lambda_1, \lambda_2) \geq \operatorname{val}(\text{Relaxed PEP-F})$ for all feasible dual variables (λ_1, λ_2). We choose $\lambda_1(t) = \lambda^F(t)$ and $\lambda_2(t) = \dot{\lambda}^F(t)$. Then, by the assumption in the theorem statement, we have $\alpha(t) = \alpha^F(t)$ and $S_{\lambda_1,\lambda_2,\nu_{\text{given}}} \succeq 0$ and. Thus, $\operatorname{Dual}(\lambda_1, \lambda_2) \leq \nu_{\text{given}}$, which implies $\nu_{\text{given}} \geq \operatorname{val}(\text{Relaxed PEP-F})$. Because the guarantee (39) holds with $\rho = \operatorname{val}(\text{Relaxed PEP-F})$ as mentioned before, it also holds with $\rho = \nu_{\text{given}}$. This completes the proof. □

C.1 Equivalence between the expressions (37) and (38)

Note that

$$
\begin{aligned}
\nabla \hat{f}(X(t)) &= \nabla f(X(t)) - \mu \left(X(t) - x_0 \right) \\
&= \nabla f(X(t)) - \mu \int_0^t \dot{X}(s) ds \\
&= \nabla f(X(t)) + \mu \int_0^t \int_0^s H(s,\tau) \nabla f(X(\tau)) \, d\tau ds \\
&= \nabla f(X(t)) + \mu \int_0^t \int_\tau^t H(s,\tau) \, ds \nabla f(X(\tau)) \, d\tau.
\end{aligned}
$$

Denote $g(t) = \nabla f(X(t))$ and $\hat{g}(t) = \nabla \hat{f}(X(t))$. Then, we have $\hat{g}(t) = g(t) - (\mathcal{V}g)(t)$, where $\mathcal{V}$ is the Volterra integral operator associated with the kernel $k(t,\tau) = -\mu \int_\tau^t H(s,\tau) \, ds$ (see Appendix A). By Proposition 5, we obtain

$$
g(t) = \hat{g}(t) + \int_0^t R(t,s) \hat{g}(s) \, ds,
$$

where R is the resolvent kernel corresponding to the kernel k (see Appendix A). Now, we can rewrite (37) as follows:

$$
\begin{aligned}
\dot{X}(t) &= - \int_0^t H(t,\tau) g(\tau) \, d\tau \\
&= - \int_0^t H(t,\tau) \left(\hat{g}(\tau) + \int_0^\tau R(\tau,s) \hat{g}(s) \, ds \right) d\tau
\end{aligned}
$$

$$
\begin{aligned}
&= -\int_0^t H(t,\tau)\hat{g}(\tau)\, d\tau - \int_0^t \int_0^s H(t,s)R(s,\tau)\hat{g}(\tau)\, d\tau ds \\
&= -\int_0^t H(t,\tau)\hat{g}(\tau)\, d\tau - \int_0^t \left(\int_\tau^t H(t,s)R(s,\tau)\, ds \right) \hat{g}(\tau)\, d\tau \\
&= -\int_0^t \left(H(t,\tau) + \int_\tau^t H(t,s)R(s,\tau)\, ds \right) \hat{g}(\tau)\, d\tau,
\end{aligned}
$$

which is (38) with $H^F(t,\tau) = H(t,\tau) + \int_\tau^t H(t,s)R(s,\tau)\, ds$.

Therefore, the form (37) can be transformed into the form (38). Conversely, a similar argument shows that (38) can be written as (37) with $H(t,\tau) = H^F(t,\tau) - \int_\tau^t H^F(t,s)k(s,\tau)\, ds$. Therefore, these two equivalent forms are in a one-to-one correspondence.

D Proof of Theorem 2

To prove the theorem, we introduce a variant of the continuous PEP. For simplicity, we assume $t_{\text{end}} = T$. Consider the following dynamical system defined in the theorem statement:

$$
\dot{X}(t) = -\int_0^t H^G(t,\tau)\nabla \hat{f}(X(\tau))\, d\tau. \tag{44}
$$

In Appendix D.1, we show its equivalence to the following form:

$$
\dot{X}(t) = -\int_0^t \bar{H}^G(t,\tau)\nabla \hat{f}_\tau(X(\tau))\, d\tau. \tag{45}
$$

Suppose we want to obtain a convergence guarantee in the form of

$$
\left\| \int_0^T \alpha^G(\tau)\nabla \hat{f}_\tau(X(\tau))\, d\tau \right\|^2 \le \rho \left(\hat{f}\,(x_0) - \hat{f}\,(X(T)) \right). \tag{46}
$$

We define the exact performance estimation problem as follows:

$$
\begin{aligned}
\max_{\substack{\hat{f}\in\mathcal{F}_0(\mathbb{R}^d;\mathbb{R})\\ X\in C^1([0,T];\mathbb{R}^d)}} \quad & \frac{\left\| \int_0^T \alpha^G(\tau)\nabla \hat{f}_\tau(X(\tau))\, d\tau \right\|^2}{M} \\
\text{subject to} \quad & X \text{ is a solution to (45) with } X(0) = x_0.
\end{aligned} \tag{Exact PEP-G}
$$

where $M = \hat{f}\,(x_0) - \hat{f}\,(X(T))$. Then, the convergence guarantee (46) holds with $\rho = \text{val(Exact PEP-G)}$.

Note that the time-varying function $\hat{f}_t(x)$ parametrized by $t \in (0,T)$ was defined as follows:

$$
\hat{f}_t(x) := \lambda^G(t)\hat{f}(x) - \left\langle \int_0^t \dot{\lambda}^G(\tau)\nabla \hat{f}(X(\tau))\, d\tau, x \right\rangle.
$$

The following property plays a crucial role in the proof:

$$
\begin{aligned}
&\left[\frac{\partial}{\partial t} \left\{ \hat{f}_t(y) - \hat{f}_t(X(T)) \right\} \right]_{y=X(t)} \\
&\qquad\qquad = \dot{\lambda}^G(t) \left(\hat{f}(X(t)) - \hat{f}(X(T)) - \left\langle \nabla \hat{f}(X(t)), X(t) - X(T) \right\rangle \right).
\end{aligned}
$$

We now relax Exact PEP-G by introducing three functions $\varphi : [0,T] \to \mathbb{R}$, $\gamma : [0,T] \to \mathbb{R}^d$, and $N : [0,T] \to \mathbb{R}$ defined as follows:

$$
\varphi(t) = \frac{1}{M} \left(\hat{f}_t(X(t)) - \hat{f}_t(X(T)) \right),
$$

$$\gamma(t) = \frac{1}{\sqrt{M}} \nabla \hat{f}_t(X(t)),$$
$$N(t) = \frac{1}{M} \left[\hat{f}(X(t)) - \hat{f}(X(T)) - \left\langle \nabla \hat{f}(X(t)), X(t) - X(T) \right\rangle \right].$$

It follows from the chain rule and the convexity of $\hat{f}$ that

$$\begin{aligned} \dot{\varphi}(t) &= \frac{1}{M} \frac{d}{dt} \left\{ \hat{f}_t(X(t)) - \hat{f}_t(X(T)) \right\} \\ &= \frac{1}{M} \left[\frac{\partial}{\partial t} \left\{ \hat{f}_t(y) - \hat{f}_t(X(T)) \right\} \right]_{y=X(t)} + \frac{1}{M} \left\langle \nabla \hat{f}_t(X(t)), \dot{X}(t) \right\rangle \\ &= \dot{\lambda}^G(t) N(t) - \left\langle \gamma(t), \int_0^t \bar{H}^G(t,\tau)\gamma(\tau)\, d\tau \right\rangle, \\ N(t) &\leq 0. \end{aligned} \tag{47}$$

Therefore, Exact PEP-G can be relaxed as follows:

$$\begin{array}{lll} \displaystyle\max_{\substack{\varphi \in C^1([0,T];\mathbb{R}) \\ \gamma \in C([0,T];\mathbb{R}^d) \\ N \in C([0,T];\mathbb{R})}} & \displaystyle\left\| \int_0^T \alpha^G(\tau)\gamma(\tau)\, d\tau \right\|^2 & \\ \text{subject to} & \displaystyle\dot{\varphi}(t) - \dot{\lambda}^G(t) N(t) + \left\langle \gamma(t), \int_0^t \bar{H}^G(t,\tau)\gamma(\tau)\, d\tau \right\rangle = 0 & \forall t \in (0,T) \\ & N(t) \leq 0 & \forall t \in (0,T). \end{array} \tag{Relaxed PEP-G}$$

Since any feasible solution to Exact PEP-G corresponds to a feasible solution to Relaxed PEP-G with the same objective value, it follows that $\mathrm{val}(\text{Relaxed PEP-G}) \geq \mathrm{val}(\text{Exact PEP-G})$. Consequently, the convergence guarantee (46) holds with $\rho = \mathrm{val}(\text{Relaxed PEP-G})$.

To obtain an upper bound of $\mathrm{val}(\text{Relaxed PEP-G})$, we use Lagrangian duality. With the Lagrange multipliers $\lambda_1 \in C^1([0,T];\mathbb{R})$ and $\lambda_2 \in C([0,T];[0,\infty))$, we define the Lagrangian function $\mathcal{L}$ as follows:

$$\begin{aligned} &\mathcal{L}(\varphi, \gamma, N; \lambda_1, \lambda_2) \\ &= \left\| \int_0^T \alpha^G(\tau)\gamma(\tau)\, d\tau \right\|^2_{\mathbb{R}^d} \\ &\quad - \int_0^T \lambda_1(t) \left(\dot{\varphi}(t) - \dot{\lambda}^G(t) N(t) + \left\langle \gamma(t), \int_0^t \bar{H}^G(t,\tau)\gamma(\tau)\, d\tau \right\rangle_{\mathbb{R}^d} \right) dt \\ &\quad - \int_0^T \lambda_2(t) N(t)\, dt. \end{aligned}$$

Let the dual function be $\mathrm{Dual}(\lambda_1, \lambda_2) = \sup_{\varphi,\gamma,N} \mathcal{L}(\varphi, \gamma, N; \lambda_1, \lambda_2)$. By weak duality, for any feasible solution (λ_1, λ_2) to the dual problem, we have $\mathrm{Dual}(\lambda_1, \lambda_2) \geq \mathrm{val}(\text{Relaxed PEP-G})$. Consequently, the convergence guarantee (46) holds with $\rho = \mathrm{Dual}(\lambda_1, \lambda_2)$. Because it is sufficient to obtain an upper bound of $\mathrm{val}(\text{Relaxed PEP-G})$ for our purpose, we compute the dual objective function value for a specific choice of dual variables: $\lambda_1(t) = \nu$ and $\lambda_2(t) = \nu \dot{\lambda}^G(t)$.

Note that, by the definition of φ, we have $\varphi(0) = 1$ and $\varphi(T) = 0$, leading to $\int_0^T \nu \dot{\varphi}(t)\, dt = \nu(\varphi(T) - \varphi(0)) = -\nu$. Hence, the Lagrangian function can be simplified as

$$\begin{aligned} \mathcal{L}(\varphi, \gamma, N; \lambda_1, \lambda_2) &= \left\| \int_0^T \alpha^G(\tau)\gamma(\tau)\, d\tau \right\|^2_{\mathbb{R}^d} + \nu \\ &\quad - \nu \int_0^T \left\langle \gamma(t), \int_0^t \bar{H}^G(t,\tau)\gamma(\tau)\, d\tau \right\rangle_{\mathbb{R}^d} dt \end{aligned}$$

$$
\begin{aligned}
&= \nu + \int_0^T \int_0^T \alpha^G(t)\alpha^G(\tau)\gamma(t)\gamma(\tau)\, dt d\tau \\
&\quad - \nu \int_0^T \left\langle \gamma(t), \int_0^t \bar{H}^G(t,\tau)\gamma(\tau)\, d\tau \right\rangle_{\mathbb{R}^d} dt \\
&= \nu - \frac{1}{2} \langle K\gamma, \gamma \rangle_{L^2([0,T];\mathbb{R}^d)},
\end{aligned}
$$

where K is the Hilbert-Schmidt integral operator associated with the symmetric kernel k defined as

$$
k(t,\tau) = \nu \bar{H}^G(t,\tau) - 2\alpha^G(t)\alpha^G(\tau), \quad t \ge \tau,
$$

which is positive semidefinite by the assumption made in the theorem statement. Thus, we have $\mathrm{Dual}(\lambda_1, \lambda_2) = \nu$ for $\lambda_1(t) = \nu$ and $\lambda_2(t) = \nu \dot{\lambda}^G(t)$. This implies $\nu \ge \mathrm{val}(\text{Relaxed PEP-G})$. Since the convergence guarantee (46) holds with $\rho = \mathrm{val}(\text{Relaxed PEP-G})$ as mentioned before, the result follows. $\square$

D.1 Equivalence between the expressions (44) and (45)

Note that

$$
\nabla \hat{f}_t(X(t)) = \lambda^G(t)\nabla \hat{f}(X(t)) - \int_0^t \dot{\lambda}^G(\tau)\nabla \hat{f}(X(\tau))\, d\tau. \tag{48}
$$

Denote $\tilde{g}(t) = \lambda^G(t)\nabla \hat{f}(X(t))$ and $\bar{g}(t) = \nabla \hat{f}_t(X(t))$. Then, we have

$$
\bar{g}(t) = \tilde{g}(t) - \int_0^t \frac{\dot{\lambda}^G(\tau)}{\lambda^G(\tau)} \tilde{g}(\tau)\, d\tau = \tilde{g}(t) - (\mathcal{V}\tilde{g})(t),
$$

where $\mathcal{V}$ is the Volterra integral operator associated with the kernel $k(t,\tau) = \frac{\dot{\lambda}^G(\tau)}{\lambda^G(\tau)}$ (see Appendix A). By Proposition 5, we obtain

$$
\tilde{g}(t) = \bar{g}(t) + \int_0^t R(t,s)\bar{g}(s)\, ds,
$$

where R is the resolvent kernel corresponding to the kernel k (see Appendix A). We can now express (44) as follows:

$$
\begin{aligned}
\dot{X}(t) &= -\int_0^t H^G(t,\tau)\nabla \hat{f}(X(\tau))\, d\tau \\
&= -\int_0^t \frac{H^G(t,\tau)}{\lambda^G(\tau)} \tilde{g}(\tau)\, d\tau \\
&= -\int_0^t \frac{H^G(t,\tau)}{\lambda^G(\tau)} \left(\bar{g}(\tau) + \int_0^\tau R(\tau,s)\bar{g}(s)\, ds \right) d\tau \\
&= -\int_0^t \frac{H^G(t,\tau)}{\lambda^G(\tau)} \bar{g}(\tau)\, d\tau - \int_0^t \int_0^\tau \frac{H^G(t,\tau)}{\lambda^G(\tau)} R(\tau,s)\bar{g}(s)\, ds d\tau \\
&= -\int_0^t \frac{H^G(t,\tau)}{\lambda^G(\tau)} \bar{g}(\tau)\, d\tau - \int_0^t \int_0^s \frac{H^G(t,s)}{\lambda^G(s)} R(s,\tau)\bar{g}(\tau)\, d\tau ds \\
&= -\int_0^t \frac{H^G(t,\tau)}{\lambda^G(\tau)} \bar{g}(\tau)\, d\tau - \int_0^t \left(\int_\tau^t \frac{H^G(t,s)R(s,\tau)}{\lambda^G(s)}\, ds \right) \bar{g}(\tau)\, d\tau \\
&= -\int_0^t \left(\frac{H^G(t,\tau)}{\lambda^G(\tau)} + \int_\tau^t \frac{H^G(t,s)R(s,\tau)}{\lambda^G(s)}\, ds \right) \bar{g}(\tau)\, d\tau,
\end{aligned}
$$

which is (45) with $\bar{H}^G(t,\tau) = \frac{H^G(t,\tau)}{\lambda^G(\tau)} + \int_\tau^t \frac{H^G(t,s)R(s,\tau)}{\lambda^G(s)}\, ds$.

We can explicitly express the resolvent kernel R as follows:

$$
R(t,\tau) = \frac{\lambda^G(t)\dot{\lambda}^G(\tau)}{\lambda^G(\tau)^2}.
$$

To verity this, we check the resolvent equation (33a) as follows:

$$
\begin{aligned}
&R(t,\tau)-k(t,\tau)-\int_{\tau}^{t}k(t,v)R(v,\tau)\,dv\\
&=\frac{\lambda^G(t)\dot{\lambda}^G(\tau)}{\lambda^G(\tau)^2}-\frac{\dot{\lambda}^G(\tau)}{\lambda^G(\tau)}-\int_{\tau}^{t}\frac{\dot{\lambda}^G(v)}{\lambda^G(v)}\frac{\lambda^G(v)\dot{\lambda}^G(\tau)}{\lambda^G(\tau)^2}\,dv\\
&=\frac{\lambda^G(t)\dot{\lambda}^G(\tau)}{\lambda^G(\tau)^2}-\frac{\dot{\lambda}^G(\tau)}{\lambda^G(\tau)}-\frac{\dot{\lambda}^G(\tau)}{\lambda^G(\tau)^2}\int_{\tau}^{t}\dot{\lambda}^G(v)\,dv\\
&=\frac{\lambda^G(t)\dot{\lambda}^G(\tau)}{\lambda^G(\tau)^2}-\frac{\dot{\lambda}^G(\tau)}{\lambda^G(\tau)}-\frac{\dot{\lambda}^G(\tau)}{\lambda^G(\tau)^2}\left(\lambda^G(t)-\lambda^G(\tau)\right)\\
&=0.
\end{aligned}
$$

Thus, we obtain the explicit form for $\bar{H}^G$ as follows:

$$
\begin{aligned}
\bar{H}^G(t,\tau)&=\frac{H^G(t,\tau)}{\lambda^G(\tau)}+\int_{\tau}^{t}\frac{H^G(t,s)R(s,\tau)}{\lambda^G(s)}\,ds\\
&=\frac{1}{\lambda^G(\tau)}H^G(t,\tau)+\frac{\dot{\lambda}^G(\tau)}{\lambda^G(\tau)^2}\int_{\tau}^{t}H^G(t,s)\,ds\\
&=\frac{1}{\lambda^G(\tau)}H^G(t,\tau)-\frac{d}{d\tau}\left\{\frac{1}{\lambda^G(\tau)}\right\}\int_{\tau}^{t}H^G(t,s)\,ds.
\end{aligned}
\tag{49}
$$

Therefore, the form (44) can be transformed into the form (45). Conversely, a similar argument shows that (45) can be written as (44) with $H^G(t,\tau)=\lambda^G(\tau)\bar{H}^G(t,\tau)-\dot{\lambda}^G(\tau)\int_{\tau}^{t}\bar{H}^G(t,s)\,ds$. Therefore, these two equivalent forms are in a one-to-one correspondence.

E Derivation of novel ODE models

In this section, we derive the limiting ODEs for the *triple momentum method* (TMM) [13] and the *information-theoretic exact method* (ITEM) [12].

E.1 TMM ODE

The *triple momentum method* (TMM), proposed in [13], is defined by the following update rule (we follow the form in [12, Section 2.2]):

$$
\begin{aligned}
y_k&=\frac{1-\sqrt{\mu s}}{1+\sqrt{\mu s}}\left(y_{k-1}-s\nabla f\left(y_{k-1}\right)\right)+\left(1-\frac{1-\sqrt{\mu s}}{1+\sqrt{\mu s}}\right)z_k\\
z_{k+1}&=\sqrt{\mu s}\left(y_k-\frac{1}{\mu}\nabla f\left(y_k\right)\right)+\left(1-\sqrt{\mu s}\right)z_k.
\end{aligned}
\tag{TMM}
$$

Here, $s=1/L$, where L is the smoothness parameter. To derive the limiting ODE for TMM, we assume that the iterates y_k and z_k are approximated by smooth curves as $Y\left(t_k\right)\approx y_k$ and $Z\left(t_k\right)\approx z_k$, where $t_k=k\sqrt{s}$. Substituting $y_k=Y(t_k)$, $y_{k-1}=Y(t_k)-\sqrt{s}\dot{Y}(t_k)$, $z_k=Z(t_k)$, and $z_{k+1}=Z(t_k)+\sqrt{s}\dot{Z}(t_k)$ into TMM yields

$$
Y=\frac{1-\sqrt{\mu s}}{1+\sqrt{\mu s}}\left(Y-\sqrt{s}\dot{Y}-s\nabla f\left(Y-\sqrt{s}\dot{Y}\right)\right)+\left(1-\frac{1-\sqrt{\mu s}}{1+\sqrt{\mu s}}\right)Z \tag{50a}
$$

$$
Z+\sqrt{s}\dot{Z}=\sqrt{\mu s}\left(Y-\frac{1}{\mu}\nabla f\left(Y\right)\right)+\left(1-\sqrt{\mu s}\right)Z, \tag{50b}
$$

where we omitted the input t_k for the curves Y and Z. After some calculations, the equation (50a) can be rewritten as

$$
\dot{Y}=-\frac{2\sqrt{\mu}}{1-\sqrt{\mu s}}Y-\sqrt{s}\nabla f\left(Y-\sqrt{s}\dot{Y}\right)+\frac{2\sqrt{\mu}}{1-\sqrt{\mu s}}Z,
$$

and the equation (50b) can be rewritten as

$$\dot{Z} = \sqrt{\mu}\left(Y - Z - \frac{1}{\mu}\nabla f\left(Y\right)\right).$$

Thus, taking the limit $s \to 0$ gives the following system of ODEs:

$$\begin{aligned}\dot{Y} &= 2\sqrt{\mu}(Z - Y)\\ \dot{Z} &= \sqrt{\mu}\left(Y - Z - \frac{1}{\mu}\nabla f\left(Y\right)\right).\end{aligned}$$

This can be equivalently formulated as the following ODE:

$$\ddot{Y} + 3\sqrt{\mu}\dot{Y} + 2\nabla f(Y) = 0,$$

which corresponds to TMM ODE in Section 3.3.

Comparison with the low-resolution TMM ODE in [11]. In [11], a *low-resolution* limiting ODE for TMM was derived as follows:

$$\ddot{X} + 2\sqrt{\tilde{\mu}}\dot{X} + \nabla f(X) = 0,$$

where $\tilde{\mu}$ is a constant greater than the strong convexity parameter μ. This ODE model differs from TMM ODE presented in our work. The reason for this difference lies in the choice of time stepsize. We use a time stepsize of $1/\sqrt{L}$, where L is the smoothness parameter used in TMM, whereas [11] employed a different time stepsize, specifically $\sqrt{\frac{2-\sqrt{\mu/L}}{L}}$.

E.2 ITEM ODE

The *information-theoretic exact method* (ITEM), proposed in [12], is defined by the following update rule:

$$\begin{aligned}A_{k+1} &= \frac{(1+\mu s)A_k + 2\left(1 + \sqrt{\left(1+A_k\right)\left(1+\mu s A_k\right)}\right)}{(1-\mu s)^2}\\ y_k &= \frac{A_k}{(1-\mu s)A_{k+1}}\left(y_{k-1} - s\nabla f\left(y_{k-1}\right)\right) + \left(1 - \frac{A_k}{(1-\mu s)A_{k+1}}\right)z_k\\ z_{k+1} &= z_k + \frac{\mu s\left((1-\mu s)^2 A_{k+1} - (1+\mu s)A_k\right)}{2(1+\mu s+\mu s A_k)}\left(y_k - z_k - \frac{1}{\mu}\nabla f\left(y_k\right)\right),\end{aligned} \tag{ITEM}$$

where $A_0 = 0$. Here, $s = 1/L$, where L is the smoothness parameter. To derive the limiting ODE for ITEM, we assume that the iterates A_k, y_k, and z_k are approximated by smooth functions and smooth curves as $A(t_k) \approx sA_k$, $Y\left(t_k\right) \approx y_k$, and $Z\left(t_k\right) \approx z_k$, where $t_k = k\sqrt{s}$.

We first compute the smooth function $A : [0,\infty) \to [0,\infty)$ that approximates the sequence $\{A_k\}$. Substituting $A_k = A(t_k)/s$ and $A_{k+1} = \frac{1}{s}A(t_k) + \frac{1}{\sqrt{s}}\dot{A}(t_k)$ into the updating rule of $\{A_k\}$, we obtain

$$\frac{1}{s}A + \frac{1}{\sqrt{s}}\dot{A} = \frac{(1+\mu s)\frac{1}{s}A + 2\left(1 + \sqrt{\left(1+\frac{1}{s}A\right)\left(1+\mu A\right)}\right)}{(1-\mu s)^2}, \tag{51}$$

where we omitted the input t_k. After some calculations, (51) can be rewritten as

$$\dot{A} = 2\sqrt{A\left(1+\mu A\right)} + o\left(\sqrt{s}\right).$$

Letting $s \to 0$ and solving the differential equation with the initial condition $A(0) = 0$, we have $A(t) = \frac{1}{\mu}\sinh^2\left(\sqrt{\mu}t\right)$. Note that this shows that the sequence $\{A_k\}$ can be approximated as $A_k \approx \frac{1}{\mu s}\sinh^2\left(\sqrt{\mu s}k\right)$.

Substituting $y_k = Y(t_k)$, $y_{k-1} = Y(t_k) - \sqrt{s}\dot{Y}(t_k)$, $z_k = Z(t_k)$, and $z_{k+1} = Z(t_k) + \sqrt{s}\dot{Z}(t_k)$ into ITEM yields

$$Y = \frac{A_k}{(1-\mu s)A_{k+1}}\left(Y - \sqrt{s}\dot{Y} - s\nabla f(Y - \sqrt{s}\dot{Y})\right) + \left(1 - \frac{A_k}{(1-\mu s)A_{k+1}}\right)Z \tag{52a}$$

$$Z+\sqrt{s}\dot{Z}=Z+\frac{\mu s\left((1-\mu s)^2A_{k+1}-(1+\mu s)A_k\right)}{2(1+\mu s+\mu sA_k)}\left(Y-Z-\frac{1}{\mu}\nabla f\left(Y\right)\right). \tag{52b}$$

To obtain the limiting ODE for this discrete-time method, we first compute the limits of the coefficients in (52) as follows:

$$\begin{aligned}\frac{1}{\sqrt{s}}\left(1-\frac{A_k}{(1-\mu s)A_{k+1}}\right)&=\frac{1}{\sqrt{s}}\left(\frac{(1-\mu s)A_{k+1}-A_k}{(1-\mu s)A_{k+1}}\right)\\&=\frac{A_{k+1}-A_k}{\sqrt{s}A_{k+1}}+o\left(\sqrt{s}\right)\\&=\frac{\dot{A}(t_k)}{A(t_k)}+o\left(\sqrt{s}\right)\\&=2\sqrt{\mu}\coth\left(\sqrt{\mu}t_k\right)+o\left(\sqrt{s}\right),\end{aligned}$$

and

$$\begin{aligned}\frac{1}{\sqrt{s}}\frac{\mu s\left((1-\mu s)^2A_{k+1}-(1+\mu s)A_k\right)}{2(1+\mu s+\mu sA_k)}&=\frac{\mu\left(sA_{k+1}-sA_k\right)}{2\sqrt{s}\left(1+\mu A(t_k)\right)}+o\left(\sqrt{s}\right)\\&=\frac{\mu\left(A(t_{k+1})-A(t_k)\right)}{2\sqrt{s}\left(1+\mu A(t_k)\right)}+o\left(\sqrt{s}\right)\\&=\frac{\mu\dot{A}(t_k)}{2\left(1+\mu A(t_k)\right)}+o\left(\sqrt{s}\right)\\&=\sqrt{\mu}\tanh\left(\sqrt{\mu}t_k\right)+o\left(\sqrt{s}\right).\end{aligned}$$

Then, following the argument in Appendix E.1, we can show that taking the limit $s\to 0$ in (52) yields the following system of ODEs:

$$\begin{aligned}\dot{Y}&=2\sqrt{\mu}\coth\left(\sqrt{\mu}t\right)\left(Z-Y\right)\\\dot{Z}&=\sqrt{\mu}\tanh\left(\sqrt{\mu}t\right)\left(Y-Z-\frac{1}{\mu}\nabla f(Y)\right).\end{aligned}$$

This is equivalent to the following ODE:

$$\ddot{Y}+3\sqrt{\mu}\coth\left(\sqrt{\mu}t\right)\dot{Y}+2\nabla f(Y)=0,$$

which coincides with ITEM ODE in Section 3.3.

F Equivalence verification of different forms of dynamics

In this paper, we frequently encounter the following form of integro-differential equation:

$$\dot{X}(t)=-\int_0^t H(t,\tau)g(\tau)\,d\tau. \tag{53}$$

Using this expression, we can derive an expression for $X(t)-X(0)$ as follows:

$$\begin{aligned}X(t)-X(0)&=\int_0^t\dot{X}(s)\,ds\\&=-\int_0^t\int_0^s H(s,\tau)g(\tau)\,d\tau ds\\&=-\int_0^t\int_\tau^t H(s,\tau)\,dsg(\tau)\,d\tau.\end{aligned} \tag{54}$$

Applying the Leibniz integral rule, we can transform the integro-differential equation (53) into the following second-order ODE (see [7, Appendix B.2.3]):

$$\ddot{X}(t)+\int_0^t\frac{\partial H(t,\tau)}{\partial t}g(\tau)\,d\tau+H(t,t)g(t)=0. \tag{55}$$

The following proposition is useful for establishing the equivalence of different forms of given ODE model.

Proposition 6. *If there exist functions $b(t)$, $B(t)$, and $c(t)$ such that one of the following conditions holds:*

$$\frac{\partial H(t,\tau)}{\partial t} = -b(t)H(t,\tau) - B(t)\int_\tau^t H(s,\tau)\,ds,\ H(t,t) = c(t), \tag{56}$$

or

$$\begin{cases} \frac{\partial^2 H(t,\tau)}{\partial t^2} = -b(t)\frac{\partial H(t,\tau)}{\partial t} - \left(b'(t) + B(t)\right)H(t,\tau), \\ H(t,t) = c(t),\ \frac{\partial H(s,\tau)}{\partial s}\Big|_{(s,\tau)=(t,t)} = -b(t)c(t), \end{cases} \tag{57}$$

then the integro-differential equation (53) *is equivalent to the following ODE:*

$$\ddot{X}(t) + b(t)\dot{X}(t) + B(t)(X(t) - X(0)) + c(t)g(t) = 0. \tag{58}$$

Proof. Substituting (56) into (55), we have

$$\begin{aligned} 0 &= \ddot{X}(t) + \int_0^t \left[-b(t)H(t,\tau) - B(t)\int_\tau^t H(s,\tau)\,ds\right] g(\tau)\,d\tau + H(t,t)g(t) \\ &= \ddot{X}(t) - b(t)\int_0^t H(t,\tau)g(\tau)\,d\tau - B(t)\int_0^t\int_\tau^t H(s,\tau)\,ds g(\tau)\,d\tau + H(t,t)g(t) \\ &= \ddot{X}(t) + b(t)\dot{X}(t) + B(t)(X(t) - X(0)) + c(t)g(t). \end{aligned}$$

Thus, if the condition (56) holds, then the dynamics (53) is equivalent to (58). It is easy to check that the condition (57) implies the condition (56), which completes the proof. □

The following proposition is an immediate consequence of Proposition 6 and can also be found in [7, Appendix B.2.3].

Proposition 7. *The second-order ODE* (58) *with $B(t) = 0$ is equivalent to the dynamics* (53) *with the H-kernel defined as follows:*

$$H(t,\tau) = c(\tau)\exp\left(-\int_\tau^t b(s)\,ds\right).$$

In the following subsections, we derive equivalent formulations of the dynamical systems considered in this paper. A detailed explanation is provided in Appendix F.1, and only the essential steps of the proofs are presented in the subsequent subsections.

F.1 Equivalent forms of AGM-SC ODE

In [7], it was shown that AGM-SC ODE can be expressed as (53) with $g(t) = \nabla f(X(t))$ and

$$H(t,\tau) = e^{2\sqrt{\mu}(\tau - t)}.$$

Equivalent form for applying Theorem 1. Since we have $\nabla\hat{f}(X(t)) = \nabla f(X(t)) - \mu\left(X(t) - x_0\right)$, we can rewrite AGM-SC ODE as follows:

$$\ddot{X}(t) + 2\sqrt{\mu}\dot{X}(t) + \mu\left(X(t) - x_0\right) + \nabla\hat{f}(X(t)) = 0. \tag{59}$$

By substituting $b(t) = 2\sqrt{\mu}$, $B(t) = \mu$, and $c(t) = 1$ into Proposition 6, we obtain the following initial value problem:

$$\begin{cases} \frac{\partial^2 H^F(t,\tau)}{\partial t^2} + 2\sqrt{\mu}\frac{\partial H^F(t,\tau)}{\partial t} + \mu H^F(t,\tau) = 0, \\ H^F(t,t) = 1,\ \frac{\partial H^F(s,\tau)}{\partial s}\Big|_{(s,\tau)=(t,t)} = -2\sqrt{\mu}. \end{cases} \tag{60}$$

We claim that the following kernel is a solution to this problem:

$$H^F(t,\tau) = \left(1 + \sqrt{\mu}(\tau - t)\right)e^{\sqrt{\mu}(\tau - t)}. \tag{61}$$

Verification of the initial conditions is straightforward:

$$H^F(t,t) = \left(1 + \sqrt{\mu}(t - t)\right)e^{\sqrt{\mu}(t-t)} = 1,$$

$$\left.\frac{\partial H^F(s,\tau)}{\partial s}\right|_{(s,\tau)=(t,t)} = \left[\left(-2\sqrt{\mu} - \mu(\tau - s)\right) e^{\sqrt{\mu}(\tau-s)}\right]_{(s,\tau)=(t,t)} = -2\sqrt{\mu}.$$

To show that (61) is a solution to the given PDE, it suffices to show that the kernels $H_1(t,\tau) = e^{-\sqrt{\mu}t}$ and $H_2(t,\tau) = te^{-\sqrt{\mu}t}$ satisfy the given PDE (without considering initial conditions). This can be verified as follows:

$$\begin{aligned}
\frac{\partial^2 H_1(t,\tau)}{\partial t^2} + 2\sqrt{\mu}\frac{\partial H_1(t,\tau)}{\partial t} + \mu H_1(t,\tau) &= \left(\mu e^{-\sqrt{\mu}t}\right) + 2\sqrt{\mu}\left(-\sqrt{\mu}e^{-\sqrt{\mu}t}\right) + \mu\left(e^{-\sqrt{\mu}t}\right) \\
&= 0, \\
\frac{\partial^2 H_2(t,\tau)}{\partial t^2} + 2\sqrt{\mu}\frac{\partial H_2(t,\tau)}{\partial t} + \mu H_2(t,\tau) &= \left(-2\sqrt{\mu} + \mu t\right) e^{-\sqrt{\mu}t} \\
&\quad + 2\sqrt{\mu}\left(\left(1 - \sqrt{\mu}t\right) e^{-\sqrt{\mu}t}\right) + \mu\left(te^{-\sqrt{\mu}t}\right) \\
&= 0.
\end{aligned}$$

Therefore, AGM-SC ODE can be equivalently expressed as the integro-differential equation (9) with the kernel H^F defined in (61).

Equivalent form for applying Theorem 2. As shown above, AGM-SC ODE can be equivalently expressed as the integro-differential equation (20) with $H^G(t,\tau) = \left(1 + \sqrt{\mu}(\tau - t)\right) e^{\sqrt{\mu}(\tau-t)}$. We begin by expressing $\dot{X}(t)$ and $X(t) - x_0$ in terms of $\nabla \hat{f}(X(t))$ as follows:

$$\begin{aligned}
\dot{X}(t) &= -\int_0^t H^G(t,\tau)\nabla \hat{f}(X(\tau))\, d\tau \\
&= -\int_0^t \left(1 + \sqrt{\mu}(\tau - t)\right) e^{\sqrt{\mu}(\tau-t)}\nabla \hat{f}(X(\tau))\, d\tau, \\
X(t) - x_0 &= -\int_0^t \int_\tau^t H^G(s,\tau)\, ds \nabla \hat{f}(X(\tau))\, d\tau \\
&= -\int_0^t \left[-(\tau - s)e^{\sqrt{\mu}(\tau-s)}\right]_{s=\tau}^t \nabla \hat{f}(X(\tau))\, d\tau \\
&= \int_0^t (\tau - t)^{\sqrt{\mu}(\tau-t)}\nabla \hat{f}(X(\tau))\, d\tau.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
-\sqrt{\mu}\int_0^t e^{\sqrt{\mu}\tau}\nabla \hat{f}(X(\tau))\, d\tau &= -\sqrt{\mu}e^{\sqrt{\mu}t}\int_0^t e^{\sqrt{\mu}(\tau-t)}\nabla \hat{f}(X(\tau))\, d\tau \\
&= -\sqrt{\mu}e^{\sqrt{\mu}t}\int_0^t \left(1 + \sqrt{\mu}(\tau - t)\right) e^{\sqrt{\mu}(\tau-t)}\nabla \hat{f}(X(\tau))\, d\tau \\
&\quad + \mu e^{\sqrt{\mu}t}\int_0^t (\tau - t)^{\sqrt{\mu}(\tau-t)}\nabla \hat{f}(X(\tau))\, d\tau \\
&= \sqrt{\mu}e^{\sqrt{\mu}t}\dot{X}(t) + \mu e^{\sqrt{\mu}t}(X(t) - x_0).
\end{aligned} \tag{62}$$

Now, with $\lambda^G(t) = e^{\sqrt{\mu}t}$, the transformation (48) can be rewritten as follows:

$$\begin{aligned}
\nabla \hat{f}_t(X(t)) &= e^{\sqrt{\mu}t}\nabla \hat{f}(X(t)) - \sqrt{\mu}\int_0^t e^{\sqrt{\mu}\tau}\nabla \hat{f}(X(\tau))\, d\tau \\
&= e^{\sqrt{\mu}t}\left(\nabla \hat{f}(X(t)) + \sqrt{\mu}\dot{X}(t) + \mu(X(t) - x_0)\right).
\end{aligned}$$

Thus, we can rewrite (59) as

$$\ddot{X}(t) + \sqrt{\mu}\dot{X}(t) + e^{-\sqrt{\mu}t}\nabla \hat{f}_t(X(t)) = 0. \tag{63}$$

By applying Proposition 7, we can show that this ODE is equivalent to the integro-differential equation (21) with

$$\begin{aligned} \bar{H}^G(t,\tau) &= e^{-\sqrt{\mu}\tau} \exp\left(-\int_\tau^t \sqrt{\mu}\, ds\right) \\ &= e^{-\sqrt{\mu}\tau} \exp\left(\sqrt{\mu}(\tau - t)\right) \\ &= e^{-\sqrt{\mu}t}. \end{aligned} \tag{64}$$

F.2 Equivalent forms of the unified AGM ODE

In [7], it was shown that Unified AGM ODE can be expressed as (53) with $g(t) = \nabla f(X(t))$ and

$$H(t,\tau) = \frac{\sinh_\tau^3 \cosh_\tau}{\sinh_t^3 \cosh_t},$$

where $\sinh_t$ and $\cosh_t$ denote the corresponding hyperbolic functions with the argument $\frac{\sqrt{\mu}}{2}t$.

Equivalent form for applying Theorem 1. We can rewrite Unified AGM ODE as follows:

$$\ddot{X}(t) + \frac{\sqrt{\mu}}{2}\left(\tanh_t + 3\coth_t\right)\dot{X}(t) + \mu\left(X(t) - x_0\right) + \nabla \hat{f}(X(t)) = 0.$$

By substituting $b(t) = \frac{\sqrt{\mu}}{2}\left(\tanh_t + 3\coth_t\right)$, $B(t) = \mu$, and $c(t) = 1$ into Proposition 6, we obtain the following initial value problem:

$$\begin{cases} \frac{\partial^2 H^F(t,\tau)}{\partial t^2} + \frac{\sqrt{\mu}}{2}\left(\tanh_t + 3\coth_t\right)\frac{\partial H^F(t,\tau)}{\partial t} + \left(\mu + \frac{\mu}{4}\operatorname{sech}_t^2 - \frac{3\mu}{4}\operatorname{csch}_t^2\right) H^F(t,\tau) = 0, \\ H^F(t,t) = 1, \ \left.\frac{\partial H^F(s,\tau)}{\partial s}\right|_{(s,\tau)=(t,t)} = -\frac{\sqrt{\mu}}{2}\left(\tanh_t + 3\coth_t\right). \end{cases} \tag{65}$$

The following kernel is a solution to this problem:

$$H^F(t,\tau) = \left(1 + \coth_t^2 \log\left(\operatorname{sech}_t^2\right) - \coth_t^2 \log\left(\operatorname{sech}_\tau^2\right)\right) \frac{\sinh_\tau \cosh_\tau}{\sinh_t \cosh_t}. \tag{66}$$

In order to show this, it is enough to verify the initial conditions and show that the kernels $H_1(t,\tau) = \left(1 + \coth_t^2 \log\left(\operatorname{sech}_t^2\right)\right)\operatorname{sech}_t \operatorname{csch}_t$ and $H_2(t,\tau) = \coth_t \operatorname{csch}_t^2$ satisfy the given PDE. We omit the detailed proofs, as they only involve calculations.

F.3 Equivalent forms of TMM ODE

Using Proposition 7, we can show that TMM ODE can be expressed as (53) with $g(t) = \nabla f(X(t))$ and

$$H(t,\tau) = 2e^{3\sqrt{\mu}(\tau - t)}.$$

Equivalent form for applying Theorem 1. TMM ODE can be rewritten as follows:

$$\ddot{X}(t) + 3\sqrt{\mu}\dot{X}(t) + 2\mu\left(X(t) - x_0\right) + 2\nabla \hat{f}(X(t)) = 0.$$

Substituting $b(t) = 3\sqrt{\mu}$, $B(t) = 2\mu$, and $c(t) = 2$ into Proposition 6 yields the following initial value problem:

$$\begin{cases} \frac{\partial^2 H^F(t,\tau)}{\partial t^2} + 3\sqrt{\mu}\frac{\partial H^F(t,\tau)}{\partial t} + 2\mu H^F(t,\tau) = 0, \\ H^F(t,t) = 2, \ \left.\frac{\partial H^F(s,\tau)}{\partial s}\right|_{(s,\tau)=(t,t)} = -6\sqrt{\mu}. \end{cases} \tag{67}$$

The following kernel is a solution to this problem:

$$H^F(t,\tau) = -2e^{\sqrt{\mu}(\tau - t)} + 4e^{2\sqrt{\mu}(\tau - t)}. \tag{68}$$

To show this, it suffices to verify the initial conditions and show that the kernels $H_1(t,\tau) = e^{-\sqrt{\mu}t}$ and $H_2(t,\tau) = e^{-2\sqrt{\mu}t}$ satisfy the given PDE. We omit the detailed proofs, as they only involve calculations.

Equivalent form for applying Theorem 2. As shown above, TMM ODE can be equivalently expressed as (20) with $H^G(t,\tau) = -2e^{\sqrt{\mu}(\tau-t)} + 4e^{2\sqrt{\mu}(\tau-t)}$. We start by expressing $\dot{X}(t)$ and $X(t) - x_0$ in terms of $\nabla\hat{f}(X(t))$:

$$\dot{X}(t) = -\int_0^t \left(-2e^{\sqrt{\mu}(\tau-t)} + 4e^{2\sqrt{\mu}(\tau-t)}\right) \nabla\hat{f}(X(\tau))\, d\tau,$$

$$X(t) - x_0 = -\int_0^t \left(\frac{2}{\sqrt{\mu}}e^{\sqrt{\mu}(\tau-t)} - \frac{2}{\sqrt{\mu}}e^{2\sqrt{\mu}(\tau-t)}\right) \nabla\hat{f}(X(\tau))\, d\tau.$$

After performing the calculations, we obtain

$$-2\sqrt{\mu}\int_0^t e^{2\sqrt{\mu}\tau}\nabla\hat{f}(X(\tau))\, d\tau = \sqrt{\mu}e^{2\sqrt{\mu}t}\dot{X}(t) + \mu e^{2\sqrt{\mu}t}(X(t) - x_0). \tag{69}$$

Thus, we can express the transformation (48) with $\lambda^G(t) = e^{2\sqrt{\mu}t}$ as follows:

$$\nabla\hat{f}_t(X(t)) = e^{2\sqrt{\mu}t}\left(\nabla\hat{f}(X(t)) + \sqrt{\mu}\dot{X}(t) + \mu(X(t) - x_0)\right).$$

Thus, we can rewrite (59) as

$$\ddot{X}(t) + \sqrt{\mu}\dot{X}(t) + 2e^{-2\sqrt{\mu}t}\nabla\hat{f}_t(X(t)) = 0. \tag{70}$$

By Proposition 7, this ODE is equivalent to the integro-differential equation (21) with

$$\begin{aligned}\bar{H}^G(t,\tau) &= 2e^{-2\sqrt{\mu}\tau}\exp\left(-\int_\tau^t \sqrt{\mu}\, ds\right)\\ &= 2e^{-2\sqrt{\mu}\tau}\exp\left(\sqrt{\mu}(\tau-t)\right)\\ &= 2e^{-\sqrt{\mu}(t+\tau)}.\end{aligned} \tag{71}$$

F.4 Equivalent forms of ITEM ODE

Using Proposition 7, we can show that ITEM ODE can be expressed as (53) with $g(t) = \nabla f(X(t))$ and

$$H(t,\tau) = \frac{2\sinh_\tau^3}{\sinh_t^3},$$

where $\sinh_t$ denotes the corresponding hyperbolic function with the argument $\sqrt{\mu}t$.

Equivalent form for applying Theorem 1. We can rewrite ITEM ODE as follows:

$$\ddot{X}(t) + 3\sqrt{\mu}\coth_t \dot{X}(t) + 2\mu\left(X(t) - x_0\right) + 2\nabla\hat{f}(X(t)) = 0.$$

By substituting $b(t) = 3\sqrt{\mu}\coth_t$, $B(t) = 2\mu$, and $c(t) = 2$ into Proposition 6, we obtain the following initial value problem:

$$\begin{cases}\frac{\partial^2 H^F(t,\tau)}{\partial t^2} + 3\sqrt{\mu}\coth_t \frac{\partial H^F(t,\tau)}{\partial t} + \left(2\mu - 3\mu\,\mathrm{csch}_t^2\right)H^F(t,\tau) = 0,\\ H^F(t,t) = 2,\ \left.\frac{\partial H^F(s,\tau)}{\partial s}\right|_{(s,\tau)=(t,t)} = -6\sqrt{\mu}\coth_t.\end{cases} \tag{72}$$

The following kernel is a solution to this problem:

$$H^F(t,\tau) = 4\sinh_\tau\cosh_\tau\coth_t\mathrm{csch}_t^2 + 2\sinh_\tau\mathrm{csch}_t\left(1 - 2\coth_t^2\right). \tag{73}$$

In order to show this, it is enough to verify the initial conditions and show that the kernels $H_1(t,\tau) = \coth_t\mathrm{csch}_t^2$ and $H_2(t,\tau) = \mathrm{csch}_t\left(1 - 2\coth_t^2\right)$ satisfy the given PDE. We omit the detailed proofs, as they only involve calculations.

F.5 Equivalent forms of OGM-G ODE

[7] showed that OGM-G ODE can be expressed as (53) with $g(t) = \nabla f(X(t))$ and

$$H(t,\tau) = \frac{(T-t)^3}{(T-\tau)^3}.$$

Equivalent form for applying Theorem 2. Since $\mu = 0$, we have $\hat{f}(x) = f(x)$, and consequently, $H^G(t,\tau) = H(t,\tau)$. Note that

$$-\int_0^t \frac{2T^2}{(T-\tau)^3}\nabla f(X(\tau))\,d\tau = -\frac{2T^2}{(T-t)^3}\int_0^t \frac{(T-t)^3}{(T-\tau)^3}\nabla f(X(\tau))\,d\tau = \frac{2T^2}{(T-t)^3}\dot{X}(t). \quad (74)$$

Thus, with $\lambda^G(t) = T^2/(T-t)^2$, the transformation (48) can be rewritten as follows:

$$\begin{aligned}\nabla \hat{f}_t(X(t)) &= \frac{T^2}{(T-t)^2}\nabla f(X(t)) - \int_0^t \frac{2T^2}{(T-\tau)^3}\nabla f(X(\tau))\,d\tau \\ &= \frac{T^2}{(T-t)^2}\nabla f(X(t)) + \frac{2T^2}{(T-t)^3}\dot{X}(t).\end{aligned}$$

Thus, we can rewrite OGM-G ODE as

$$\ddot{X}(t) + \frac{1}{T-t}\dot{X}(t) + \frac{(T-t)^2}{T^2}\nabla \hat{f}_t(X(t)) = 0. \quad (75)$$

By applying Proposition 7, we can show that this ODE is equivalent to the integro-differential equation (21) with

$$\begin{aligned}\bar{H}^G(t,\tau) &= \frac{(T-\tau)^2}{T^2}\exp\left(-\int_\tau^t \frac{1}{T-s}\,ds\right) \\ &= \frac{(T-\tau)^2}{T^2}\exp\left(\log(T-t) - \log(T-\tau)\right) \\ &= \frac{(T-t)(T-\tau)}{T^2}.\end{aligned} \quad (76)$$

F.6 Equivalent forms of the unified AGM-G ODE

In [7], it was shown that Unified AGM-G ODE can be expressed as (53) with $g(t) = \nabla f(X(t))$ and

$$H(t,\tau) = \frac{\sinh^3_{T-t}\cosh_{T-t}}{\sinh^3_{T-\tau}\cosh_{T-\tau}},$$

where $\sinh_{T-t}$ and $\cosh_{T-t}$ denote the corresponding hyperbolic functions with the argument $\frac{\sqrt{\mu}}{2}(T-t)$.

Equivalent form for applying Theorem 2. We begin by showing that Unified AGM-G ODE can be equivalently expressed as the integro-differential equation (20) with

$$H^G(t,\tau) = \left(1 + \coth^2_{T-\tau}\log\left(\operatorname{sech}^2_{T-\tau}\right) - \coth^2_{T-\tau}\log\left(\operatorname{sech}^2_{T-t}\right)\right)\frac{\sinh_{T-t}\cosh_{T-t}}{\sinh_{T-\tau}\cosh_{T-\tau}}.$$

This can be proven by showing that it satisfies the following initial value problem, for which we omit the detailed calculations:

$$\begin{cases}\frac{\partial^2 H^F(t,\tau)}{\partial t^2} + \frac{\sqrt{\mu}}{2}\left(\tanh_{T-t} + 3\coth_{T-t}\right)\frac{\partial H^F(t,\tau)}{\partial t} \\ \qquad + \left(\mu - \frac{\mu}{4}\operatorname{sech}^2_{T-t} + \frac{3\mu}{4}\operatorname{csch}^2_{T-t}\right)H^F(t,\tau) = 0, \\ H^F(t,t) = 1, \ \left.\frac{\partial H^F(s,\tau)}{\partial s}\right|_{(s,\tau)=(t,t)} = -\frac{\sqrt{\mu}}{2}\left(\tanh_{T-t} + 3\coth_{T-t}\right).\end{cases} \quad (77)$$

We proceed to translate the expression from (20) into (21). After performing the calculations, we can express $\dot{X}(t)$ and $X(t) - x_0$ in terms of $\nabla \hat{f}(X(t))$ as follows:

$$\begin{aligned}\dot{X}(t) = -\int_0^t &\left(1 + \coth^2_{T-\tau}\log\left(\operatorname{sech}^2_{T-\tau}\right) - \coth^2_{T-\tau}\log\left(\operatorname{sech}^2_{T-t}\right)\right) \\ &\times \frac{\sinh_{T-t}\cosh_{T-t}}{\sinh_{T-\tau}\cosh_{T-\tau}}\nabla \hat{f}(X(\tau))\,d\tau,\end{aligned}$$

$$X(t)-x_0=\int_0^t\left[\left(1+\coth^2_{T-\tau}\log\left(\operatorname{sech}^2_{T-\tau}\right)-\coth^2_{T-\tau}\log\left(\operatorname{sech}^2_{T-t}\right)\right)\right.$$
$$\left.\times\frac{\cosh^2_{T-t}}{\sqrt{\mu}\sinh_{T-\tau}\cosh_{T-\tau}}-\frac{\sinh^2_{T-t}\coth_{T-\tau}\operatorname{csch}^2_{T-\tau}}{\sqrt{\mu}}\right]\nabla\hat{f}(X(\tau))\,d\tau.$$

This implies

$$-\sqrt{\mu}\int_0^t\frac{\coth_{T-\tau}\operatorname{csch}^2_{T-\tau}}{\operatorname{csch}^2_T}\nabla\hat{f}(X(\tau))\,d\tau$$
$$=\frac{\sqrt{\mu}\coth_{T-t}\operatorname{csch}^2_{T-t}}{\operatorname{csch}^2_T}\dot{X}(t)+\mu\frac{\operatorname{csch}^2_{T-t}}{\operatorname{csch}^2_T}(X(t)-x_0). \tag{78}$$

As a result, we can express the transformation (48) with $\lambda^G(t)=\frac{\operatorname{csch}^2_{T-t}}{\operatorname{csch}^2_T}$, as follows:

$$\nabla\hat{f}_t(X(t))=\frac{\operatorname{csch}^2_{T-t}}{\operatorname{csch}^2_T}\left(\nabla\hat{f}(X(t))+\sqrt{\mu}\coth_{T-t}\dot{X}(t)+\mu(X(t)-x_0)\right).$$

We can then rewrite Unified AGM-G ODE as follows:

$$\begin{aligned}0&=\ddot{X}(t)+\frac{\sqrt{\mu}}{2}\left(\tanh_{T-t}+3\coth_{T-t}\right)\dot{X}(t)+\nabla f(X(t))\\&=\ddot{X}(t)+\frac{\sqrt{\mu}}{2}\left(\tanh_{T-t}+3\coth_{T-t}\right)\dot{X}(t)+\mu(X(t)-x_0)+\nabla\hat{f}(X(t))\\&=\ddot{X}(t)+\frac{\sqrt{\mu}}{2}\left(\tanh_{T-t}+\coth_{T-t}\right)\dot{X}(t)+\frac{\sinh^2_{T-t}}{\sinh^2_T}\nabla\hat{f}_t(X(t)).\end{aligned} \tag{79}$$

By Proposition 7, this ODE is equivalent to the integro-differential equation (21) with the kernel $\bar{H}^G$ given by

$$\begin{aligned}\bar{H}^G(t,\tau)&=\frac{\sinh^2_{T-\tau}}{\sinh^2_T}\exp\left(-\frac{\sqrt{\mu}}{2}\int_\tau^t\left(\tanh_{T-s}+\coth_{T-s}\right)ds\right)\\&=\frac{\sinh^2_{T-\tau}}{\sinh^2_T}\exp\left(\left[\log\left(\cosh_{T-s}\right)+\log\left(\sinh_{T-s}\right)\right]_{s=\tau}^t\right)\\&=\frac{\sinh_{T-t}\cosh_{T-t}\tanh_{T-\tau}}{\sinh^2_T}.\end{aligned} \tag{80}$$

F.7 Equivalent forms of ITEM-G ODE

Using Proposition 7, we can show that ITEM-G ODE can be expressed as (53) with $g(t)=\nabla f(X(t))$ and

$$H(t,\tau)=\frac{2\sinh^3_{T-t}}{\sinh^3_{T-\tau}},$$

where $\sinh_{T-t}$ denotes the corresponding hyperbolic functions with the argument $\sqrt{\mu}(T-t)$.

Equivalent form for applying Theorem 2. We first show that ITEM-G ODE can be equivalently expressed as the integro-differential equation (20) with

$$H^G(t,\tau)=4\sinh_{T-t}\cosh_{T-t}\coth_{T-\tau}\operatorname{csch}^2_{T-\tau}+2\sinh_{T-t}\operatorname{csch}_{T-\tau}\left(1-2\coth^2_{T-\tau}\right).$$

This can be verified by showing that it is a solution to the following initial value problem, for which we omit the calculations:

$$\begin{cases}\frac{\partial^2H^F(t,\tau)}{\partial t^2}+3\sqrt{\mu}\coth_{T-t}\frac{\partial H^F(t,\tau)}{\partial t}+\left(2\mu+3\mu\operatorname{csch}^2_t\right)H^F(t,\tau)=0,\\ H^F(t,t)=2,\ \left.\frac{\partial H^F(s,\tau)}{\partial s}\right|_{(s,\tau)=(t,t)}=-6\sqrt{\mu}\coth_{T-t}.\end{cases} \tag{81}$$

We now proceed to translate the expression from (20) into (21). After performing the calculations, we can express $\dot{X}(t)$ and $X(t)-x_0$ in terms of $\nabla \hat{f}(X(t))$ as shown below:

$$\begin{aligned}
\dot{X}(t) = -\int_0^t \Big[&4\sinh_{T-t}\cosh_{T-t}\coth_{T-\tau}\operatorname{csch}^2_{T-\tau} \\
&+ 2\sinh_{T-t}\operatorname{csch}_{T-\tau}\left(1-2\coth^2_{T-\tau}\right)\Big]\nabla\hat{f}(X(\tau))\,d\tau, \\
X(t)-x_0 = \int_0^t \Bigg[&\frac{4}{\sqrt{\mu}}\cosh^2_{T-t}\coth_{T-\tau}\operatorname{csch}^2_{T-\tau} + \frac{2}{\sqrt{\mu}}\cosh_{T-t}\operatorname{csch}_{T-\tau}\left(1-2\coth^2_{T-\tau}\right) \\
&- \frac{2\sinh^2_{T-t}\coth_{T-\tau}\operatorname{csch}^2_{T-\tau}}{\sqrt{\mu}}\Bigg]\nabla\hat{f}(X(\tau))\,d\tau,
\end{aligned}$$

which implies

$$\begin{aligned}
&-2\sqrt{\mu}\int_0^t \frac{\coth_{T-\tau}\operatorname{csch}^2_{T-\tau}}{\operatorname{csch}^2_T}\nabla\hat{f}(X(\tau))\,d\tau \\
&\qquad = \frac{\sqrt{\mu}\coth_{T-t}\operatorname{csch}^2_{T-t}}{\operatorname{csch}^2_T}\dot{X}(t) + \mu\frac{\operatorname{csch}^2_{T-t}}{\operatorname{csch}^2_T}(X(t)-x_0). \qquad (82)
\end{aligned}$$

Consequently, we can express the transformation (48) with $\lambda^G(t) = \frac{\operatorname{csch}^2_{T-t}}{\operatorname{csch}^2_T}$, as follows:

$$\nabla\hat{f}_t(X(t)) = \frac{\operatorname{csch}^2_{T-t}}{\operatorname{csch}^2_T}\left(\nabla\hat{f}(X(t)) + \sqrt{\mu}\coth_{T-t}\dot{X}(t) + \mu(X(t)-x_0)\right).$$

We can then rewrite ITEM-G ODE as follows:

$$\begin{aligned}
0 &= \ddot{X}(t) + 3\sqrt{\mu}\coth_{T-t}\dot{X}(t) + 2\nabla f(X(t)) \\
&= \ddot{X}(t) + 3\sqrt{\mu}\coth_{T-t}\dot{X}(t) + 2\mu(X(t)-x_0) + 2\nabla\hat{f}(X(t)) \\
&= \ddot{X}(t) + \sqrt{\mu}\coth_{T-t}\dot{X}(t) + \frac{2\sinh^2_{T-t}}{\sinh^2_T}\nabla\hat{f}_t(X(t)).
\end{aligned} \tag{83}$$

By Proposition 7, this ODE is equivalent to the integro-differential equation (21) with

$$\begin{aligned}
\bar{H}^G(t,\tau) &= \frac{2\sinh^2_{T-\tau}}{\sinh^2_T}\exp\left(-\sqrt{\mu}\int_\tau^t \coth_{T-s}\,ds\right) \\
&= \frac{2\sinh^2_{T-\tau}}{\sinh^2_T}\exp\left(\left[\log\left(\sinh_{T-s}\right)\right]_{s=\tau}^t\right) \\
&= \frac{2\sinh_{T-t}\sinh_{T-\tau}}{\sinh^2_T}.
\end{aligned} \tag{84}$$

G Computation of PEP kernels

In this section, we compute the PEP kernels for various ODE models considered in this paper.

G.1 AGM-SC ODE for minimizing function values

In Appendix F.1, we showed that AGM-SC ODE can be equivalently expressed as the integro-differential equation (9) with the kernel H^F defined in (61). For this kernel, we compute the following integrals:

$$\begin{aligned}
\int_\tau^t H^F(s,\tau)\,ds &= \int_\tau^t \left(1+\sqrt{\mu}(\tau-s)\right)e^{\sqrt{\mu}(\tau-s)}\,ds \\
&= \left[-(\tau-s)e^{\sqrt{\mu}(\tau-s)}\right]_{s=\tau}^t
\end{aligned}$$

$$
\begin{aligned}
&= -(\tau - t)e^{\sqrt{\mu}(\tau - t)}, \\
\int_0^t \int_0^s H^F(s,\tau)\, d\tau ds &= \int_0^t \int_\tau^t H^F(s,\tau)\, ds d\tau \\
&= -\int_0^t (\tau - t)e^{\sqrt{\mu}(\tau-t)}\, d\tau \\
&= \left[-\frac{1}{\sqrt{\mu}}(\tau - t)e^{\sqrt{\mu}(\tau - t)} + \frac{1}{\mu}e^{\sqrt{\mu}(\tau - t)} \right]_{\tau=0}^t \\
&= \frac{1}{\mu} - \frac{1}{\sqrt{\mu}}te^{-\sqrt{\mu}t} - \frac{1}{\mu}e^{-\sqrt{\mu}t}.
\end{aligned}
$$

Using these results, we compute the kernel $\frac{\partial}{\partial t}\left\{\lambda^F(t)\int_\tau^t H^F(s,\tau)\, ds\right\}$ and the function $\alpha^F(t)$ as follows:

$$
\begin{aligned}
\frac{\partial}{\partial t}\left\{\lambda^F(t)\int_\tau^t H^F(s,\tau)\, ds\right\} &= \frac{\partial}{\partial t}\left\{e^{\sqrt{\mu}(t-T)}\left(-(\tau - t)e^{\sqrt{\mu}(\tau - t)}\right)\right\} \\
&= \frac{\partial}{\partial t}\left\{(t - \tau)e^{\sqrt{\mu}(\tau - T)}\right\} \\
&= e^{\sqrt{\mu}(\tau - T)}, \\
\alpha^F(t) &= \frac{1}{2}\frac{d}{dt}\left\{\lambda^F(t)(1 - \mu\int_0^t\int_0^s H^F(s,\tau)\, d\tau ds)\right\} \\
&= \frac{1}{2}\frac{d}{dt}\left\{e^{\sqrt{\mu}(t-T)}\left(\sqrt{\mu}te^{-\sqrt{\mu}t} + e^{-\sqrt{\mu}t}\right)\right\} \\
&= \frac{1}{2}\frac{d}{dt}\left\{\sqrt{\mu}te^{-\sqrt{\mu}T} + e^{-\sqrt{\mu}T}\right\} \\
&= \frac{1}{2}\sqrt{\mu}e^{-\sqrt{\mu}T}.
\end{aligned}
$$

Therefore, the PEP kernel (10) can be computed as

$$
\begin{aligned}
S^F(t,\tau) &= \nu\frac{\partial}{\partial t}\left\{\lambda^F(t)\int_\tau^t H^F(s,\tau)\, ds\right\} - 2\alpha^F(t)\alpha^F(\tau) \\
&= \nu e^{\sqrt{\mu}(\tau - T)} - \frac{\mu}{2}e^{-2\sqrt{\mu}T}
\end{aligned}
$$

for $t \geq \tau$.

G.2 Unified AGM ODE

In Appendix F.2, we established the equivalence between AGM-SC ODE and the integro-differential equation (9) with the kernel H^F defined in (66). For this kernel, we can verify that

$$
\begin{aligned}
\int_\tau^t H^F(s,\tau)\, ds &= \int_\tau^t \left(1 + \coth_s^2 \log\left(\operatorname{sech}_s^2\right) - \coth_s^2 \log\left(\operatorname{sech}_\tau^2\right)\right) \frac{\sinh_\tau \cosh_\tau}{\sinh_s \cosh_s}\, ds \\
&= \sinh_\tau \cosh_\tau \int_\tau^t \frac{1 + \coth_s^2 \log\left(\operatorname{sech}_s^2\right)}{\sinh_s \cosh_s}\, ds \\
&\quad - \sinh_\tau \cosh_\tau \log\left(\operatorname{sech}_\tau^2\right) \int_\tau^t \coth_s \operatorname{csch}_s^2\, ds \\
&= \sinh_\tau \cosh_\tau \left[-\frac{1}{\sqrt{\mu}} \operatorname{csch}_s^2 \log\left(\operatorname{sech}_s^2\right)\right]_{s=\tau}^t \\
&\quad - \sinh_\tau \cosh_\tau \log\left(\operatorname{sech}_\tau^2\right) \left[-\frac{1}{\sqrt{\mu}} \operatorname{csch}_s^2\right]_{s=\tau}^t \\
&= \frac{1}{\sqrt{\mu}} \operatorname{csch}_t^2 \sinh_\tau \cosh_\tau \left(\log\left(\operatorname{sech}_\tau^2\right) - \log\left(\operatorname{sech}_t^2\right)\right),
\end{aligned}
$$

$$
\begin{aligned}
\int_0^t \int_0^s H^F(s,\tau)\, d\tau ds &= \int_0^t \int_\tau^t H^F(s,\tau)\, ds d\tau \\
&= \frac{\operatorname{csch}_t^2}{\sqrt{\mu}} \int_0^t \sinh_\tau \cosh_\tau \log\left(\operatorname{sech}_\tau^2\right)\, d\tau \\
&\quad - \frac{\operatorname{csch}_t^2 \log\left(\operatorname{sech}_t^2\right)}{\sqrt{\mu}} \int_0^t \sinh_\tau \cosh_\tau \, d\tau \\
&= \frac{\operatorname{csch}_t^2}{\sqrt{\mu}} \left[\frac{1}{\sqrt{\mu}} \left(\sinh_\tau^2 + \cosh_\tau^2 \log\left(\operatorname{sech}_\tau^2\right)\right) \right]_{\tau=0}^t \\
&\quad - \frac{\operatorname{csch}_t^2 \log\left(\operatorname{sech}_t^2\right)}{\sqrt{\mu}} \left[\frac{\sinh_\tau^2}{\sqrt{\mu}} \right]_{\tau=0}^t \\
&= \frac{1}{\mu} \left(1 + \operatorname{csch}_t^2 \log\left(\operatorname{sech}_t^2\right)\right).
\end{aligned}
$$

Thus, the kernel $\frac{\partial}{\partial t}\left\{\lambda^F(t)\int_\tau^t H^F(s,\tau)\, ds\right\}$ and the function $\alpha^F(t)$ can be computed as follows:

$$
\begin{aligned}
\frac{\partial}{\partial t}\left\{\lambda^F(t)\int_\tau^t H^F(s,\tau)\, ds\right\} &= \frac{\partial}{\partial t}\left\{ \frac{\sinh_t^2}{\sinh_T^2} \frac{1}{\sqrt{\mu}} \operatorname{csch}_t^2 \sinh_\tau \cosh_\tau \left(\log\left(\operatorname{sech}_\tau^2\right) - \log\left(\operatorname{sech}_t^2\right)\right) \right\} \\
&= \frac{\partial}{\partial t}\left\{ \frac{1}{\sqrt{\mu}\sinh_T^2} \sinh_\tau \cosh_\tau \left(\log\left(\operatorname{sech}_\tau^2\right) - \log\left(\operatorname{sech}_t^2\right)\right) \right\} \\
&= -\frac{\sinh_\tau \cosh_\tau}{\sqrt{\mu}\sinh_T^2} \frac{\partial}{\partial t}\left\{\log\left(\operatorname{sech}_t^2\right)\right\} \\
&= \frac{\tanh_t \sinh_\tau \cosh_\tau}{\sinh_T^2}, \\
\alpha^F(t) &= \frac{1}{2}\frac{d}{dt}\left\{\lambda^F(t)(1-\mu\int_0^t\int_0^s H^F(s,\tau)\, d\tau ds)\right\} \\
&= \frac{1}{2}\frac{d}{dt}\left\{ \frac{\sinh_t^2}{\sinh_T^2} \left(1 - \mu\frac{1}{\mu}\left(1 + \operatorname{csch}_t^2 \log\left(\operatorname{sech}_t^2\right)\right)\right) \right\} \\
&= -\frac{1}{2\sinh_T^2}\frac{d}{dt}\left\{\log\left(\operatorname{sech}_t^2\right)\right\} \\
&= \frac{\sqrt{\mu}\tanh_t}{2\sinh_T^2}.
\end{aligned}
$$

Therefore, the PEP kernel (10) can be computed as

$$
\begin{aligned}
S^F(t,\tau) &= \nu \frac{\partial}{\partial t}\left\{\lambda^F(t)\int_\tau^t H^F(s,\tau)\, ds\right\} - 2\alpha^F(t)\alpha^F(\tau) \\
&= \nu \frac{\tanh_t \sinh_\tau \cosh_\tau}{\sinh_T^2} - \frac{\mu \tanh_t \tanh_\tau}{2\sinh_T^4} \\
&= \left(\nu - \frac{\mu}{2}\operatorname{csch}_T^2\right) \frac{\tanh_t \tanh_\tau}{\sinh_T^2} + \nu \frac{\tanh_t \sinh_\tau \sinh_\tau^2}{\sinh_T^2}
\end{aligned}
$$

for $t \geq \tau$.

G.3 TMM ODE for minimizing function values

In Appendix F.3, we showed that TMM ODE can be equivalently expressed as the integro-differential equation (9) with the kernel H^F defined in (68). For this kernel, we compute the following integrals:

$$
\int_\tau^t H^F(s,\tau)\, ds = \int_\tau^t \left(-2e^{\sqrt{\mu}(\tau-s)} + 4e^{2\sqrt{\mu}(\tau-s)}\right)\, ds
$$

$$
\begin{aligned}
&= \frac{2}{\sqrt{\mu}} \left[e^{\sqrt{\mu}(\tau - s)} - e^{2\sqrt{\mu}(\tau - s)} \right]_{s=\tau}^{t} \\
&= \frac{2}{\sqrt{\mu}} \left(e^{\sqrt{\mu}(\tau - t)} - e^{2\sqrt{\mu}(\tau - t)} \right),
\end{aligned}
$$

$$
\begin{aligned}
\int_0^t \int_0^s H^F(s,\tau)\, d\tau ds &= \int_0^t \int_\tau^t H^F(s,\tau)\, ds d\tau \\
&= \frac{2}{\sqrt{\mu}} \int_0^t \left(e^{\sqrt{\mu}(\tau - t)} - e^{2\sqrt{\mu}(\tau - t)} \right) d\tau \\
&= \frac{1}{\mu} \left[2e^{\sqrt{\mu}(\tau - t)} - e^{2\sqrt{\mu}(\tau - t)} \right]_{\tau=0}^{t} \\
&= \frac{1}{\mu} \left(1 - 2e^{-\sqrt{\mu}t} + e^{-2\sqrt{\mu}t} \right).
\end{aligned}
$$

Thus, the kernel $\frac{\partial}{\partial t} \left\{ \lambda^F(t) \int_\tau^t H^F(s,\tau)\, ds \right\}$ and the function $\alpha^F(t)$ can be computed as follows:

$$
\begin{aligned}
\frac{\partial}{\partial t} \left\{ \lambda^F(t) \int_\tau^t H^F(s,\tau)\, ds \right\} &= \frac{\partial}{\partial t} \left\{ e^{2\sqrt{\mu}(t-T)} \frac{2}{\sqrt{\mu}} \left(e^{\sqrt{\mu}(\tau - t)} - e^{2\sqrt{\mu}(\tau - t)} \right) \right\} \\
&= \frac{2}{\sqrt{\mu}} \frac{\partial}{\partial t} \left\{ e^{\sqrt{\mu}(t+\tau-2T)} - e^{2\sqrt{\mu}(\tau - T)} \right\} \\
&= 2e^{\sqrt{\mu}(t+\tau-2T)}, \\
\alpha^F(t) &= \frac{1}{2} \frac{d}{dt} \left\{ \lambda^F(t)(1 - \mu \int_0^t \int_0^s H^F(s,\tau)\, d\tau ds) \right\} \\
&= \frac{1}{2} \frac{d}{dt} \left\{ e^{2\sqrt{\mu}(t-T)} \left(1 - \mu \frac{1}{\mu} \left(1 - 2e^{-\sqrt{\mu}t} + e^{-2\sqrt{\mu}t} \right) \right) \right\} \\
&= \frac{1}{2} \frac{d}{dt} \left\{ 2e^{\sqrt{\mu}(t-2T)} - e^{-2\sqrt{\mu}T} \right\} \\
&= \sqrt{\mu} e^{\sqrt{\mu}(t-2T)}.
\end{aligned}
$$

Therefore, the PEP kernel (10) can be computed as

$$
\begin{aligned}
S^F(t,\tau) &= \nu \frac{\partial}{\partial t} \left\{ \lambda^F(t) \int_\tau^t H^F(s,\tau)\, ds \right\} - 2\alpha^F(t)\alpha^F(\tau) \\
&= 2\nu e^{\sqrt{\mu}(t+\tau-2T)} - 2\mu e^{\sqrt{\mu}(t+\tau-4T)} \\
&= 2 \left(\nu - \mu e^{-2\sqrt{\mu}T} \right) e^{\sqrt{\mu}(t+\tau-2T)}
\end{aligned}
$$

for $t \geq \tau$.

G.4 ITEM ODE

In Appendix F.4, we showed that ITEM ODE can be equivalently expressed as the integro-differential equation (9) with the kernel H^F defined in (73). For this kernel, we can verify that

$$
\begin{aligned}
\int_\tau^t H^F(s,\tau)\, ds &= \int_\tau^t \left(4 \sinh_\tau \cosh_\tau \coth_s \operatorname{csch}_s^2 + 2 \sinh_\tau \operatorname{csch}_s \left(1 - 2\coth_s^2 \right) \right) ds \\
&= 4 \sinh_\tau \cosh_\tau \left[-\frac{1}{2\sqrt{\mu}} \operatorname{csch}_s^2 \right]_{s=\tau}^{t} \\
&\quad + 2 \sinh_\tau \left[\frac{1}{\sqrt{\mu}} \coth_s \operatorname{csch}_s \right]_{s=\tau}^{t} \\
&= \frac{2}{\sqrt{\mu}} \left(-\operatorname{csch}_t^2 \sinh_\tau \cosh_\tau + \coth_\tau + \coth_t \operatorname{csch}_t \sinh_\tau - \coth_\tau \right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{2}{\sqrt{\mu}}\left(-\operatorname{csch}_t^2 \sinh_\tau \cosh_\tau + \coth_t \operatorname{csch}_t \sinh_\tau\right), \\
\int_0^t \int_0^s H^F(s,\tau)\, d\tau ds &= \int_0^t \int_\tau^t H^F(s,\tau)\, ds d\tau \\
&= \frac{2}{\sqrt{\mu}} \int_0^t \left(-\operatorname{csch}_t^2 \sinh_\tau \cosh_\tau + \coth_t \operatorname{csch}_t \sinh_\tau\right)\, d\tau \\
&= \frac{1}{\mu}\left[-\operatorname{csch}_t^2 \sinh_\tau^2 + 2\coth_t \operatorname{csch}_t \cosh_\tau\right]_{\tau=0}^t \\
&= \frac{1}{\mu}\left(-1 + 2\coth_t^2 - 2\coth_t \operatorname{csch}_t\right) \\
&= \frac{1}{\mu}\left(1 + 2\operatorname{csch}_t^2 - 2\coth_t \operatorname{csch}_t\right).
\end{aligned}
$$

Thus, the kernel $\frac{\partial}{\partial t}\left\{\lambda^F(t)\int_\tau^t H^F(s,\tau)\, ds\right\}$ and the function $\alpha^F(t)$ can be computed as follows:

$$
\begin{aligned}
\frac{\partial}{\partial t}\left\{\lambda^F(t)\int_\tau^t H^F(s,\tau)\, ds\right\} &= \frac{\partial}{\partial t}\left\{\frac{\sinh_t^2}{\sinh_T^2}\frac{2}{\sqrt{\mu}}\left(-\operatorname{csch}_t^2 \sinh_\tau \cosh_\tau + \coth_t \operatorname{csch}_t \sinh_\tau\right)\right\} \\
&= \frac{\partial}{\partial t}\left\{-\frac{2\sinh_\tau \cosh_\tau}{\sqrt{\mu}\sinh_T^2} + \frac{2\cosh_t \sinh_\tau}{\sqrt{\mu}\sinh_T^2}\right\} \\
&= \frac{2\sinh_t \sinh_\tau}{\sinh_T^2}, \\
\alpha^F(t) &= \frac{1}{2}\frac{d}{dt}\left\{\lambda^F(t)(1 - \mu\int_0^t \int_0^s H^F(s,\tau)\, d\tau ds)\right\} \\
&= \frac{1}{2}\frac{d}{dt}\left\{\frac{\sinh_t^2}{\sinh_T^2}\left(1 - \mu\frac{1}{\mu}\left(1 + 2\operatorname{csch}_t^2 - 2\coth_t \operatorname{csch}_t\right)\right)\right\} \\
&= \frac{1}{\sinh_T^2}\frac{d}{dt}\left\{-1 + \cosh_t\right\} \\
&= \frac{\sqrt{\mu}\sinh_t}{\sinh_T^2}.
\end{aligned}
$$

Therefore, the PEP kernel (10) can be computed as

$$
\begin{aligned}
S^F(t,\tau) &= \nu\frac{\partial}{\partial t}\left\{\lambda^F(t)\int_\tau^t H^F(s,\tau)\, ds\right\} - 2\alpha^F(t)\alpha^F(\tau) \\
&= \frac{2\nu \sinh_t \sinh_\tau}{\sinh_T^2} - \frac{2\mu \sinh_t \sinh_\tau}{\sinh_T^4} \\
&= 2\operatorname{csch}_T^2\left(\nu - \mu \operatorname{csch}_T^2\right)\sinh_t \sinh_\tau
\end{aligned}
$$

for $t \geq \tau$.

G.5 OGM-G ODE

In Appendix F.5, we showed that OGM-G ODE can be equivalently expressed as the integro-differential equation (21), using the kernel $\bar{H}^G$ defined in (76). Therefore, with $C(t_{\text{end}}) = 1/(T - t_{\text{end}})$ and $\alpha^G(t) = C(t_{\text{end}})\bar{H}^G(t_{\text{end}}, t)$, we can compute the PEP kernel (23) as follows:

$$
\begin{aligned}
S^G(t,\tau) &= \nu\bar{H}^G(t,\tau) - 2\alpha^G(t)\alpha^G(\tau) \\
&= \nu\bar{H}^G(t,\tau) - 2C(t_{\text{end}})^2\bar{H}^G(t_{\text{end}},t)\bar{H}^G(t_{\text{end}},\tau) \\
&= \nu\frac{(T-t)(T-\tau)}{T^2} - \frac{2}{(T-t_{\text{end}})^2}\frac{(T-t_{\text{end}})(T-t)}{T^2}\frac{(T-t_{\text{end}})(T-\tau)}{T^2} \\
&= \left(\nu - \frac{2}{T^2}\right)\frac{(T-t)(T-\tau)}{T^2}
\end{aligned}
$$

for $t \geq \tau$.

G.6 AGM-SC ODE for minimizing velocity norms

In Appendix F.1, we showed that OGM-G ODE can be equivalently expressed as the integro-differential equation (21), using the kernel $\bar{H}^G$ defined in (64). Therefore, with $t_{\text{end}} = T$, $C(T) = \frac{\sqrt{\mu}}{2}$, and $\alpha^G(t) = C(T)\bar{H}^G(T,t)$, we can compute the PEP kernel (23) as follows:

$$\begin{aligned}
S^G(t,\tau) &= \nu \bar{H}^G(t,\tau) - 2\alpha^G(t)\alpha^G(\tau) \\
&= \nu \bar{H}^G(t,\tau) - 2C(T)^2 \bar{H}^G(T,t)\bar{H}^G(T,\tau) \\
&= \nu e^{-\sqrt{\mu}t} - \frac{\mu}{2} e^{-2\sqrt{\mu}T}
\end{aligned}$$

for $t \geq \tau$.

G.7 Unified AGM-G ODE

In Appendix F.6, we showed that Unified AGM-G ODE can be equivalently formulated as the integro-differential equation (21), using the kernel $\bar{H}^G$ defined in (80). Therefore, with $C(t_{\text{end}}) = \frac{\sqrt{\mu}}{2} \operatorname{sech}_{T-t_{\text{end}}} \operatorname{csch}_{T-t_{\text{end}}}$ and $\alpha^G(t) = C(t_{\text{end}})\bar{H}^G(t_{\text{end}}, t)$, we can compute the PEP kernel (23) as follows:

$$\begin{aligned}
S^G(t,\tau) &= \nu \bar{H}^G(t,\tau) - 2\alpha^G(t)\alpha^G(\tau) \\
&= \nu \bar{H}^G(t,\tau) - 2C(t_{\text{end}})^2 \bar{H}^G(t_{\text{end}},t)\bar{H}^G(t_{\text{end}},\tau) \\
&= \nu \frac{\sinh_{T-t}\cosh_{T-t}\tanh_{T-\tau}}{\sinh_T^2} \\
&\quad - \frac{\mu}{2}\operatorname{sech}^2_{T-t_{\text{end}}}\operatorname{csch}^2_{T-t_{\text{end}}} \frac{(\sinh_{T-t_{\text{end}}}\cosh_{T-t_{\text{end}}})^2 \tanh_{T-t}\tanh_{T-\tau}}{\sinh_T^4} \\
&= \left(\nu - \frac{\mu}{2}\operatorname{csch}^2_T\right)\frac{\tanh_{T-t}\tanh_{T-\tau}}{\sinh^2_T} + \nu \frac{\sinh^2_{T-t}\tanh_{T-t}\tanh_{T-\tau}}{\sinh^2_T}
\end{aligned}$$

for $t \geq \tau$.

Verification of $C(t_{\text{end}})\dot{X}(t_{\text{end}}) \to -\frac{1}{2}\nabla f(X(T))$. In order to apply the Dirac delta function argument, we show that the solution $X : [0,T) \to \mathbb{R}^d$ to Unified AGM-G ODE can be continuously extended to $t = T$. We employ a similar argument to the one presented in [10, Appendix D.3]. By using the energy function in [7, Appendix F.5], we can show that $\|\dot{X}(t)\|^2$ is bounded. Thus, X is uniformly continuous, implying that X can be continuously extended to $t = T$.

Note that Unified AGM-G ODE can be expressed in the following form (see [7]):

$$\dot{X}(t) = -\int_0^t H(t,\tau)\nabla f(X(\tau))\,d\tau, \quad H(t,\tau) = \frac{\sinh^3_{T-t}\cosh_{T-t}}{\sinh^3_{T-\tau}\cosh_{T-\tau}}.$$

Thus, we have

$$\begin{aligned}
C(t_{\text{end}})\dot{X}(t_{\text{end}}) &= -C(t_{\text{end}})\int_0^{t_{\text{end}}} H(t_{\text{end}},\tau)\nabla f(X(\tau))\,d\tau \\
&= -\int_0^{t_{\text{end}}} \frac{\sqrt{\mu}\sinh^2_{T-t_{\text{end}}}}{2\sinh^3_{T-\tau}\cosh_{T-\tau}}\nabla f(X(\tau))\,d\tau \\
&= -\frac{1}{2}\int_0^T \alpha_t(\tau)\nabla f(X(\tau))\,d\tau,
\end{aligned}$$

where

$$\alpha_t(\tau) = \frac{\sqrt{\mu}\sinh^2_{T-t_{\text{end}}}}{\sinh^3_{T-\tau}\cosh_{T-\tau}}\mathbf{1}_{[0,t_{\text{end}}]}.$$

Now, it suffices to show $\alpha_t \to \delta_T$. To show this, we need to verify the following three conditions: (i) $\alpha_t(\tau) \geq 0$, (ii) $\int_0^T \alpha_t(\tau)\,d\tau \to 1$ as $t \to T$, and (iii) for every $\eta \in (0,T)$, we have $\int_0^\eta \alpha_t(\tau)\,d\tau \to$

0 as $t \to T$. Checking the conditions (i) and (iii) is straightforward, so we only show (ii). The integral can be computed as follows:

$$\int_0^T \alpha_t(\tau)\, d\tau = \sqrt{\mu} \sinh^2_{T-t_{\text{end}}} \int_0^{t_{\text{end}}} \frac{1}{\sinh^3_{T-\tau} \cosh_{T-\tau}}\, d\tau$$
$$= \sinh^2_{T-t_{\text{end}}} \left[\operatorname{csch}^2_{T-\tau} + \log(\tanh^2_{T-\tau}) \right]_{\tau=0}^{t_{\text{end}}},$$

which converges to 1 as $t_{\text{end}} \to T$.

G.8 TMM ODE for minimizing velocity norms

In Appendix F.3, we showed that TMM ODE can be equivalently expressed as the integro-differential equation (21), using the kernel $\bar{H}^G$ defined in (71). Therefore, with $t_{\text{end}} = T$, $C(T) = \frac{\sqrt{\mu}}{2}$, and $\alpha^G(t) = C(T)\bar{H}^G(T,t)$, we can compute the PEP kernel (23) as follows:

$$\begin{aligned} S_{\lambda,\nu}(t,\tau) &= \nu \bar{H}^G(t,\tau) - 2C(T)^2 \bar{H}^G(T,t)\bar{H}^G(T,\tau) \\ &= 2\nu e^{-\sqrt{\mu}(t+\tau)} - 2\mu e^{-\sqrt{\mu}(t+\tau+2T)} \\ &= 2\left(\nu - \mu e^{-2\sqrt{\mu}T}\right) e^{-\sqrt{\mu}(t+\tau)} \end{aligned}$$

for $t \geq \tau$.

G.9 ITEM-G ODE

In Appendix F.7, we showed that ITEM-G ODE can be equivalently formulated as the integro-differential equation (21), using the kernel $\bar{H}^G$ defined in (84). Therefore, with $C(t_{\text{end}}) = \frac{\sqrt{\mu}}{2} \operatorname{csch}_{T-t_{\text{end}}}$ and $\alpha^G(t) = C(t_{\text{end}})\bar{H}^G(t_{\text{end}}, t)$, we can compute the PEP kernel (23) as follows:

$$\begin{aligned} S^G(t,\tau) &= \nu \bar{H}^G(t,\tau) - 2\alpha^G(t)\alpha^G(\tau) \\ &= \nu \bar{H}^G(t,\tau) - 2C(t_{\text{end}})^2 \bar{H}^G(t_{\text{end}},t)\bar{H}^G(t_{\text{end}},\tau) \\ &= \frac{2\nu \sinh_{T-t}\sinh_{T-\tau}}{\sinh^2_T} - 2\mu \operatorname{csch}^2_{T-t_{\text{end}}} \frac{\sinh^2_{T-t_{\text{end}}} \sinh_{T-t}\sinh_{T-\tau}}{\sinh^4_T} \\ &= 2\operatorname{csch}^2_T \left(\nu - \mu \operatorname{csch}^2_T\right) \sinh_{T-t}\sinh_{T-\tau} \end{aligned}$$

for $t \geq \tau$.

Verification of $C(t_{\text{end}})\dot{X}(t_{\text{end}}) \to -\frac{1}{2}\nabla f(X(T))$. In order to apply the Dirac delta function argument, we first show that the solution $X : [0,T) \to \mathbb{R}^d$ to ITEM-G ODE can be continuously extended to $t = T$. We employ a similar argument to the one presented in [10, Appendix D.3]. By using the conservation of the following quantity over time:

$$\frac{1}{2}\left\|\dot{X}(t)\right\|^2 + 3\sqrt{\mu}\int_0^t \coth_{T-s}\left\|\dot{X}(s)\right\|^2 ds + 2(f(X(t) - f(x^*)),$$

we can see that $\|\dot{X}(t)\|^2$ is bounded. Thus, X is uniformly continuous, implying that X can be continuously extended to $t = T$.

Note that ITEM-G ODE can be expressed in the following form (see Appendix F.7):

$$\dot{X}(t) = -\int_0^t H(t,\tau)\nabla f(X(\tau))\, d\tau, \quad H(t,\tau) = \frac{2\sinh^3_{T-t}}{\sinh^3_{T-\tau}}.$$

Thus, we obtain

$$\begin{aligned} C(t_{\text{end}})\dot{X}(t_{\text{end}}) &= -C(t_{\text{end}})\int_0^{t_{\text{end}}} H(t_{\text{end}},\tau)\nabla f(X(\tau))\, d\tau \\ &= -\int_0^{t_{\text{end}}} \frac{\sqrt{\mu}\sinh^2_{T-t_{\text{end}}}}{\sinh^3_{T-\tau}} \nabla f(X(\tau))\, d\tau \end{aligned}$$

$$= -\frac{1}{2}\int_0^T \alpha_t(\tau)\nabla f(X(\tau))\,d\tau,$$

where

$$\alpha_t(\tau) = \frac{2\sqrt{\mu}\sinh^2_{T-t_{\mathrm{end}}}}{\sinh^3_{T-\tau}}\mathbf{1}_{[0,t_{\mathrm{end}}]}$$

Now, it suffices to show $\alpha_t \to \delta_T$. To show this, we need to verify the following three conditions: (i) $\alpha_t(\tau) \geq 0$, (ii) $\int_0^T \alpha_t(\tau)\,d\tau \to 1$ as $t \to T$, and (iii) for every $\eta \in (0,T)$, we have $\int_0^\eta \alpha_t(\tau)\,d\tau \to 0$ as $t \to T$. Checking the conditions (i) and (iii) is straightforward, so we only prove (ii). The integral can be computed as follows:

$$\begin{aligned}\int_0^T \alpha_t(\tau)\,d\tau &= 2\sqrt{\mu}\sinh^2_{T-t_{\mathrm{end}}}\int_0^{t_{\mathrm{end}}} \frac{1}{\sinh^3_{T-\tau}}\,d\tau \\ &= \sinh^2_{T-t_{\mathrm{end}}}\left[\coth\operatorname{csch}_{T-\tau} + \log(\tanh^2_{\frac{T-\tau}{2}})\right]_{\tau=0}^{t_{\mathrm{end}}},\end{aligned}$$

which converges to 1 as $t_{\mathrm{end}} \to T$.

H Missing details from Section 3

H.1 Optimality of the dual variables selected in the proof of Proposition 2

In this subsection, we show that the multiplier functions $\lambda_1(t) = t^2/T^2$ and $\lambda_2(t) = 2t/T^2$, chosen in the proof of Proposition 2, is indeed an optimal solution for the dual problem $\min_{\lambda_1,\lambda_2}\mathrm{Dual}(\lambda_1,\lambda_2)$. Since a dual feasible solution corresponds to a convergence proof using the corresponding weighted integral of inequalities, and vice versa, this indicates that the obtained guarantee cannot be improved.

We can exclude the case $\dot{\lambda}_1 \neq \lambda_2$, because $\mathrm{Dual}(\lambda_1,\lambda_2) = \infty$ for that case. Let $\Lambda \in C^1([0,T];[0,\infty)$ with $\Lambda(0) = 0$ and $\Lambda(T) = 1$. The PEP kernel (8) with $\lambda_1(t) = \Lambda(t)$, $\lambda_2(t) = \dot{\Lambda}(t)$, and $H(t,\tau) = \tau^3/t^3$ can be computed as follows:[1]

$$\begin{aligned}S_{\Lambda,\dot{\Lambda},\nu}(t,\tau) &= \nu\left(\Lambda(t)H(t,\tau) + \dot{\Lambda}(t)\int_\tau^t H(s,\tau)\,ds\right) - \frac{1}{2}\dot{\Lambda}(t)\dot{\Lambda}(\tau) \\ &= \nu\left(\Lambda(t)\frac{\tau^3}{t^3} + \frac{\dot{\Lambda}(t)}{2}\left(\tau - \frac{\tau^3}{t^2}\right)\right) - \frac{1}{2}\dot{\Lambda}(t)\dot{\Lambda}(\tau)\end{aligned}$$

for $t \geq \tau$.

Claim 1. *The following statements hold:*

(a) *For $\nu = 2/T^2$, we have $S_{\Lambda,\dot{\Lambda},\nu}(t,\tau) \succeq 0$ if and only if $\Lambda(t) = t^2/T^2$,*

(b) *For $\nu \in (0, 2/T^2)$, we have $S_{\Lambda,\dot{\Lambda},\nu}(t,\tau) \not\succeq 0$ for all $\Lambda \in C^1([0,T];[0,\infty))$ such that $\Lambda(0) = 0$ and $\Lambda(T) = 1$.*

Assuming that Claim 1 holds, we proceed to show that the multiplier functions $(\lambda_1(t),\lambda_2(t)) = (t^2/T^2, 2t/T^2)$ form an optimal solution to the dual problem $\min_{\lambda_1,\lambda_2}\mathrm{Dual}(\lambda_1,\lambda_2)$. Because $\mathrm{Dual}(\lambda_1,\lambda_2) = \inf_{\nu\in(0,\infty)}\{\nu : S_{\lambda_1,\lambda_2,\nu} \succeq 0\}$, we have $\mathrm{Dual}(t^2/T^2, 2t/T^2) = 2/T^2$ by Claim 1 (a). In addition, for $\Lambda(t)$ which is not equal to $t \mapsto t^2/T^2$, we have $\mathrm{Dual}(\Lambda,\dot{\Lambda}) \geq 2/T^2$ by Claim 1 (b). Therefore, we conclude that the multiplier functions $(\lambda_1(t),\lambda_2(t)) = (t^2/T^2, 2t/T^2)$ minimize the dual function $\mathrm{Dual}(\lambda_1,\lambda_2)$.

We now prove Claim 1. In the proof of Proposition 2, we showed that $\Lambda(t) = t^2/T^2$ implies $S_{\Lambda,\dot{\Lambda},2/T^2}(t,\tau) \succeq 0$. Assume $S_{\Lambda,\dot{\Lambda},2/T^2}(t,\tau) \succeq 0$. To prove that $\Lambda(t) = t^2/T^2$, we first note that by Proposition 1 (e), we have

$$0 \leq S_{\Lambda,\dot{\Lambda},2/T^2}(t,t) = \frac{2}{T^2}\Lambda(t) - \frac{1}{2}\left(\dot{\Lambda}(t)\right)^2$$

[1] Note that the assumption $\lambda_2(t) = \dot{\lambda}_1(t)$ does not restrict the generality of the analysis, as failing to satisfy this condition leads to $\mathrm{Dual}(\lambda_1,\lambda_2) = \infty$.

for all $t \in (0, T)$. Thus, we have

$$\frac{d}{dt}\left\{\sqrt{\Lambda(t)}\right\} = \frac{\dot{\Lambda}(t)}{2\sqrt{\Lambda(t)}} \leq \frac{1}{T} \tag{85}$$

for all $t \in (0, T)$ with $\Lambda(t) \neq 0$. By the assumptions, we have $\sqrt{\Lambda(0)} = 0$ and $\sqrt{\Lambda(1)} = 1$. It is easy to check that the only function $L \in C^1([0, T]; \mathbb{R})$ satisfying $L(0) = 0$, $L(T) = 1$, and $\dot{L}(t) \leq 1/T$ is $L(t) = t/T$. Therefore, we have $\Lambda(t) = t^2/T^2$, which proves Claim 1 (a). When $\nu \in (0, 2/T^2)$, we can use a similar argument as above to show that $S_{\Lambda,\dot{\Lambda},\nu}(t, \tau) \not\succeq 0$ for $\Lambda(t) = t^2/T^2$ (in this case, the right-hand side of (85) becomes a constant smaller than $1/T$). Claim 1 (b) immediately follows from the fact that $S_{\Lambda,\dot{\Lambda},\nu} \succeq 0$ implies $S_{\Lambda,\dot{\Lambda},\tilde{\nu}} \succeq 0$ for all $\tilde{\nu} \geq \nu$.[2]

H.2 Correspondence between continuous and discrete PEP

In this subsection, we establish the connection between the continuous PEP presented in Section 3 and the discrete PEP proposed in the work of Drori and Teboulle [4], by showing that the former can be seen as the continuous-time limit of the latter.

H.2.1 Review of discrete PEP

We first review the *discrete PEP* presented in [4]. Note that the formulation and notation used here differ slightly from those in [4]. The general form of the discrete-time method we consider is represented by the following system of equations:

$$\begin{aligned} x_1 &= x_0 - \frac{1}{L} h_{1,0} \nabla f(x_0) \\ x_2 &= x_1 - \frac{1}{L}\left(h_{2,0} \nabla f(x_0) + h_{2,1} \nabla f(x_1)\right) \\ &\vdots \\ x_N &= x_{N-1} - \frac{1}{L}\left(h_{N,0} \nabla f(x_0) + \cdots + h_{N,N-1} \nabla f(x_{N-1})\right). \end{aligned} \tag{86}$$

The exact PEP is defined as follows:

$$\begin{array}{ll} \underset{\substack{f \in \mathcal{F}_0(\mathbb{R}^d;\mathbb{R}) \\ x_0,\ldots,x_N \in \mathbb{R}^d}}{\max} & \dfrac{f(x_N) - f(x^*)}{\|x_0 - x^*\|^2} \\ \text{subject to} & \{x_i\}_{i=0}^N \text{ is updated by the rule (86)} \\ & x^* \text{ is a minimizer of } f. \end{array} \tag{Exact Discrete PEP}$$

With two sequences $\{\varphi_i\}_{i=0}^N$ in $\mathbb{R}$ and $\{\gamma_i\}_{i=0}^N$ in $\mathbb{R}^d$, defined as

$$\begin{aligned} \varphi_i &:= \frac{1}{L\|x_0 - x^*\|^2}\left(f(x_i) - f(x^*)\right) \\ \gamma_i &:= \frac{1}{L\|x_0 - x^*\|} \nabla f(x_i), \end{aligned} \tag{87}$$

we can relax Exact Discrete PEP using a similar argument as in Section 3. Define the matrix $G \in \mathbb{R}^{(N+1)\times d}$ as $G = [\gamma_0, \ldots, \gamma_N]^T$. Then, we have the following relaxation of Exact Discrete PEP:

$$\begin{array}{lll} \underset{\substack{G \in \mathbb{R}^{(N+1)\times d} \\ \varphi \in \mathbb{R}^{N+1} \\ v \in \mathbb{R}^d,\ \|v\|=1}}{\max} & L\delta_N & \\ \text{subject to} & \operatorname{Tr}\left\{G^T A_{i-1,i} G\right\} \leq \varphi_{i-1} - \varphi_i & \forall i \in \{1, \ldots, N\} \\ & \operatorname{Tr}\left\{G^T D_i G + v u_i^T G\right\} \leq -\varphi_i & \forall i \in \{0, \ldots, N\}, \end{array} \tag{Relaxed Discrete PEP}$$

[2] This becomes clear by observing that the positive semidefiniteness of $S_{\Lambda,\dot{\Lambda},\nu}$ is equivalent to the positive semidefiniteness of $\frac{1}{\nu} S_{\Lambda,\dot{\Lambda},\nu}$, and the fact that the kernel $(t, \tau) \mapsto \dot{\Lambda}(t)\dot{\Lambda}(\tau)$ is positive semidefinite.

where the $(N+1) \times (N+1)$ matrices $A_{i-1,i}$ and D_i are defined as follows:

$$A_{i-1,i} := \frac{1}{2}(u_{i-1} - u_i)(u_{i-1} - u_i)^T + \frac{1}{2}\sum_{k=0}^{i-1} h_{i,k}\left(u_i u_k^T + u_k u_i^T\right),$$

$$D_i := \frac{1}{2}u_i u_i^T + \frac{1}{2}\sum_{j=1}^{i}\sum_{k=0}^{j-1} h_{j,k}\left(u_i u_k^T + u_k u_i^T\right).$$

With the two *Lagrange multipliers*, $\lambda_1 = (\lambda_1^1, \ldots, \lambda_1^N) \in \mathbb{R}_{\geq 0}^N$ and $\lambda_2 = (\lambda_2^0, \ldots, \lambda_2^N) \in \mathbb{R}_{\geq 0}^{N+1}$, the following dual objective function can be obtained (see [4, Section 4.2]):

$$\mathrm{Dual}(\lambda_1, \lambda_2) = \begin{cases} \inf_{\nu \in \mathbb{R}}\left\{\nu : \bar{S}_{\lambda_1,\lambda_2,\nu} \succeq 0\right\} & \text{if } (\lambda_1, \lambda_2) \in \Xi \\ \infty & \text{otherwise,} \end{cases}$$

with

$$\begin{aligned} \Xi = \Big\{&(\lambda_1, \lambda_2) \in \mathbb{R}_{\geq 0}^N \times \mathbb{R}_{\geq 0}^{N+1} : \lambda_2^0 = \lambda_1^1, \\ &\lambda_1^N + \lambda_2^N = 1,\ \lambda_1^i - \lambda_1^{i+1} + \lambda_2^i = 0\ \forall i \in \{1, \ldots, N-1\}\Big\} \end{aligned} \tag{88}$$

and

$$\bar{S}_{\lambda_1,\lambda_2,\nu} = \begin{bmatrix} \bar{\bar{S}}_{\lambda_1,\lambda_2,\nu} & \tau \\ \tau^T & \frac{2\nu}{L} \end{bmatrix},$$

where $\bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}$ is an $(N+1) \times (N+1)$ matrix defined as follows (the explicit form can be found in [5]):[3]

$$\begin{aligned} \bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}(i,j) &= 2\left(\sum_{i=1}^{N} \lambda_1^i A_{i-1,i} + \sum_{i=0}^{N} \lambda_2^i D_i\right) \\ &= \begin{cases} \left(\lambda_1^i + \lambda_2^i\right) h_{i,j} + \lambda_2^i \sum_{k=j+1}^{i-1} h_{k,j}, & 2 \leq i \leq N,\ 0 \leq j \leq i-2 \\ \left(\lambda_1^i + \lambda_2^i\right) h_{i,j} - \lambda_1^i, & 1 \leq i \leq N,\ j = i-1 \\ 2\lambda_1^{i+1}, & 0 \leq i \leq N-1,\ j = i \\ 1, & i = j = N. \end{cases} \end{aligned} \tag{89}$$

Using a well-known property of the Schur complement (see [1, Appendix A.5.5]), we can show that the condition $\bar{S}_{\lambda_1,\lambda_2,\nu} \succeq 0$ is equivalent to the following condition:

$$S_{\lambda_1,\lambda_2,\nu} := \nu \bar{\bar{S}}_{\lambda_1,\lambda_2,\nu} - \frac{L}{2}\lambda_2 \lambda_2^T \succeq 0. \tag{90}$$

Thus, the dual objective function can be rewritten as follows:

$$\mathrm{Dual}(\lambda_1, \lambda_2) = \begin{cases} \inf_{\nu \in \mathbb{R}}\left\{\nu : S_{\lambda_1,\lambda_2,\nu} \succeq 0\right\} & \text{if } (\lambda_1, \lambda_2) \in \Xi \\ \infty & \text{otherwise.} \end{cases}$$

We now demonstrate how the discrete PEP can be used to establish a convergence guarantee for a given discrete-time method. Let $\nu_{\text{feas}} \in (0, \infty)$ be given, and suppose that the matrix $S_{\lambda_1,\lambda_2,\nu_{\text{feas}}}$ defined in (90) is positive semidefinite, with appropriate multiplier vectors λ_1 and λ_2. Using a similar argument as in Section 3, we can show that $\mathrm{val}(\text{Exact Discrete PEP}) \leq \mathrm{val}(\text{Relaxed Discrete PEP}) \leq \mathrm{Dual}(\lambda_1, \lambda_2) \leq \nu_{\text{feas}}$. Consequently, this implies the following convergence guarantee:

$$f(x_N) - f(x^*) \leq \nu_{\text{feas}} \left\|x_0 - x^*\right\|^2.$$

Therefore, the discrete PEP transforms the task of establishing the convergence guarantee for a discrete-time method into the verification of the positive semidefiniteness of a specific matrix.

[3] Here, the (i, j)th entry of the matrix A is denoted as $A(i, j)$.

Example: AGM. As an example of the application of the discrete PEP presented above, Kim and Fessler [5] used it to analyze the convergence rate of Nesterov's accelerated gradient method (AGM) [8], which can be represented as follows:

$$\begin{aligned}\theta_{i+1} &= \frac{1+\sqrt{1+4\theta_i^2}}{2}\\ y_{i+1} &= x_i - \frac{1}{L}\nabla f(x_i)\\ x_{i+1} &= y_{i+1} + \frac{\theta_i - 1}{\theta_{i+1}}(y_{i+1}-y_i),\end{aligned} \tag{AGM}$$

where $x_0 = y_0$ and $\theta_0 = 1$. Since this method can be expressed in the form of (86) (see [5, Section 3]), we can apply the discrete PEP above. With the Lagrange multipliers $\lambda_1^i = \theta_{i-1}^2/\theta_N^2$ and $\lambda_2^i = \theta_i/\theta_N^2$, the matrix $\bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}(i,j)$ defined in (89) can be computed as follows (see [5, Section 5]):

$$\begin{aligned}\bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}(i,j) &= \begin{cases}\frac{\theta_i\theta_j}{\theta_N^2}, & 1 \le i \le N,\ 0 \le j \le i-1\\ \frac{2\theta_i^2}{\theta_N^2}, & 0 \le i \le N-1,\ j = i\\ 1, & i = j = N\end{cases}\\ &= \frac{1}{\theta_N^2}\left([\theta_0,\dots,\theta_N]\,[\theta_0,\dots,\theta_N]^T + 2\mathrm{diag}\left\{\left[\theta_0^2,\dots,\theta_{N-1}^2,0\right]^T\right\}\right).\end{aligned} \tag{91}$$

Hence, the matrix $S_{\lambda_1,\lambda_2,\nu}$ defined in (90) with $\nu = \frac{L}{2\theta_N^2}$ can be computed as follows:

$$\begin{aligned}S_{\lambda_1,\lambda_2,\nu}(i,j) &= \nu\bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}(i,j) - \frac{L}{2}\lambda_2^i\lambda_2^j\\ &= \frac{L}{2\theta_N^4}\left([\theta_0,\dots,\theta_N]\,[\theta_0,\dots,\theta_N]^T + 2\mathrm{diag}\left\{\left[\theta_0^2,\dots,\theta_{N-1}^2,0\right]^T\right\}\right)\\ &\quad - \frac{L}{2\theta_N^4}\,[\theta_0,\dots,\theta_N]\,[\theta_0,\dots,\theta_N]^T\\ &= \frac{L}{\theta_N^4}\mathrm{diag}\left\{\left[\theta_0^2,\dots,\theta_{N-1}^2,0\right]^T\right\}\\ &= \frac{L}{\theta_N^4}\begin{bmatrix}\theta_0^2 & \cdots & 0 & 0\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \cdots & \theta_{N-1}^2 & 0\\ 0 & \cdots & 0 & 0\end{bmatrix},\end{aligned} \tag{92}$$

which is clearly positive semidefinite. Consequently, we have $\mathrm{Dual}(\lambda_1,\lambda_2) \le \frac{L}{2\theta_N^2}$, implying the following well-known convergence guarantee for AGM:

$$f(x_N) - f(x^*) \le \frac{L}{2\theta_N^2}\left\|x_0 - x^*\right\|^2.$$

H.2.2 Continuous PEP as the limit of discrete PEP

In this subsection, we informally show that the continuous PEP presented in Section 3 can be considered as the continuous-time limit of the discrete PEP presented in Appendix H.2.1. By making the approximation $t \approx k\sqrt{s}$ and $T \approx N\sqrt{s}$, where $s = 1/L$, we establish a correspondence between the iterations $k \in \{0,\dots,N\}$ and the time $t \in [0,T]$. Under this approximation, we can see that the iterates $\{x_i\}$ of the discrete-time method (86) converge to the solution X to the following integro-differential equation:

$$\dot{X}(t) = \int_0^t H(t,\tau)\nabla f(X(\tau))\,d\tau,$$

where $H(t,\tau) = \lim_{s\to 0} h_{\frac{t}{\sqrt{s}},\frac{\tau}{\sqrt{s}}}$, under the approximation $X(k\sqrt{s}) = x_k$ (see [7, Appendix B.2.3]). It follows that the sequences $\{\varphi_i\}$ and $\{\gamma_i\}$ defined in (87) converge to the functions φ and γ

defined in Section 3.1 under the approximations $\varphi(k\sqrt{s}) \approx \varphi_k/s$ and $\gamma(k\sqrt{s}) \approx \gamma_k/s$. Using the approximations $\lambda_1(k\sqrt{s}) \approx \lambda_1^k$ and $\lambda_2(k\sqrt{s}) \approx \lambda_2^k/\sqrt{s}$, we can identify the Lagrangian multipliers in the discrete PEP and the continuous PEP. Consequently, the condition $(\lambda_1, \lambda_2) \in \Xi$, where Ξ is the set defined in (88), is transformed into the following conditions: $\lambda_1(0) = 0$, $\lambda_1(T) = 1$, and $\dot{\lambda}_1(t) = \lambda_2(t)$ for all $t \in (0, T)$. The limiting kernel $\bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}(t,\tau) := \lim_{s\to 0} \bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}(t/\sqrt{s}, \tau/\sqrt{s})$ of the matrix defined in (89) can be computed as follows:

$$
\begin{aligned}
\bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}(t,\tau) &:= \lim_{s\to 0} \bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}\left(\frac{t}{\sqrt{s}}, \frac{\tau}{\sqrt{s}}\right) \\
&= \lim_{s\to 0}\left(\left(\lambda_1^{t/\sqrt{s}} + \lambda_2^{t/\sqrt{s}}\right) h_{\frac{t}{\sqrt{s}},\frac{\tau}{\sqrt{s}}} + \lambda_2^{t/\sqrt{s}} \sum_{k=\frac{\tau}{\sqrt{s}}+1}^{\frac{t}{\sqrt{s}}-1} h_{k,\frac{\tau}{\sqrt{s}}}\right) \\
&= \lim_{s\to 0}\left(\left(\lambda_1(t) + \sqrt{s}\lambda_2(t)\right) H(t,\tau) + \sqrt{s}\lambda_2(t) \sum_{k=1}^{\frac{t-\tau}{\sqrt{s}}-1} H\left(\tau + \sqrt{s}k, \tau\right)\right) \\
&= \lambda_1(t) H(t,\tau) + \lambda_2(t) \int_\tau^t H(s,\tau)\, ds
\end{aligned}
\tag{93}
$$

for $t \geq \tau$. Note that this kernel coincides with the kernel k introduced in Section 3.2. As a result, we can compute the limiting kernel $S_{\lambda_1,\lambda_2,\nu} := \lim_{s\to 0} S_{\lambda_1,\lambda_2,\nu}(t/\sqrt{s}, \tau/\sqrt{s})$ of the matrix defined in (90) as follows:

$$
\begin{aligned}
S_{\lambda_1,\lambda_2,\nu}(t,\tau) &= \nu \lim_{s\to 0} \bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}\left(\frac{t}{\sqrt{s}}, \frac{\tau}{\sqrt{s}}\right) - \lim_{s\to 0} \frac{1}{2s} \lambda_2^{t/\sqrt{s}} \lambda_2^{\tau/\sqrt{s}} \\
&= \nu\left(\lambda_1(t) H(t,\tau) + \lambda_2(t) \int_\tau^t H(s,\tau)\, ds\right) - \frac{1}{2}\lambda_2(t)\lambda_2(\tau)
\end{aligned}
\tag{94}
$$

for $t \geq \tau$. Note that this kernel coincides with the PEP kernel defined in (8). Consequently, it follows from a limiting argument that the positive semidefininteness of the matrix (90) translates into the positive semidefiniteness of the PEP kernel (8). Therefore, we observe that the continuous PEP presented in Section 3 serves as a continuous-time counterpart to the discrete PEP in Appendix H.2.1.

Example: AGM ODE. To provide an example for the relationship between the continuous PEP and the discrete PEP discussed above, we consider AGM ODE, which is the continuous-time limit of AGM. Before delving into the calculations, we note that the sequence $\{\theta_i\}$ involved in the update rule of AGM can be approximated as $\theta_i = (i+2)/2$.

It is known that the limiting kernel of the difference matrix $(h_i j)$ for AGM can be computed as $\lim_{s\to 0} h_{\frac{t}{\sqrt{s}},\frac{\tau}{\sqrt{s}}} = \tau^3/t^3$ (see [7]), which coincides with the H-kernel for AGM ODE presented in Section 3. We now compute the continuous-time counterparts of the multipliers vectors $\lambda_1^i = \theta_{i-1}^2/\theta_N^2$ and $\lambda_2^i = \theta_i/\theta_N^2$:

$$
\begin{aligned}
\lambda_1(t) &= \lim_{s\to 0} \lambda_1^{t\sqrt{s}} = \lim_{s\to 0} \frac{\theta_{t/\sqrt{s}-1}^2}{\theta_{T/\sqrt{s}}^2} = \frac{t^2}{T^2}, \\
\lambda_2(t) &= \lim_{s\to 0} \frac{\lambda_2^{t\sqrt{s}}}{\sqrt{s}} = \lim_{s\to 0} \frac{\theta_{t/\sqrt{s}}}{\sqrt{s}\theta_{T/\sqrt{s}}^2} = \frac{2t}{T^2},
\end{aligned}
$$

which coincides with the multiplier functions considered in the proof of Proposition 2. Next, we compute the kernel $\bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}$ defined in (93):

$$
\begin{aligned}
\bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}(t,\tau) &= \lambda_1(t) H(t,\tau) + \lambda_2(t) \int_\tau^t H(s,\tau)\, ds \\
&= \frac{t^2}{T^2}\frac{\tau^3}{t^3} + \frac{2t}{T^2} \int_\tau^t \frac{\tau^3}{s^3}\, ds \\
&= \frac{t\tau}{T^2}
\end{aligned}
$$

for $t \geq \tau$. We can observe that this kernel is the limiting kernel of the matrix $\bar{\bar{S}}_{\lambda_1,\lambda_2,\nu}$ defined in (91). With $\nu = \lim_{s\to 0} \frac{1}{2s\theta_N^2} = \frac{2}{T^2}$, we can see that the kernel $S_{\lambda_1,\lambda_2,\nu}(t,\tau)$ defined in (94) is the zero kernel, which aligns with the fact that the limiting kernel of the matrix $S_{\lambda_1,\lambda_2,\nu}$ defined in (92) is the zero kernel. Hence, we can conclude that the continuous PEP applied to AGM ODE is consistent with the discrete PEP applied to AGM.

Before concluding this section, we highlight the significance and practicality of our continuous PEP framework. We can observe that the analysis of the continuous PEP applied to AGM ODE, involves shorter and simpler computations compared to the analysis of the discrete PEP applied to AGM. This observation suggests that the continuous PEP can serve as an accessible model for analyzing the discrete PEP. In other words, by examining the continuous PEP for the corresponding continuous-time dynamics, one can gain insights and guidance in analyzing the discrete PEP for a given discrete-time method.

In particular, one non-trivial step in analyzing the discrete PEP is choosing appropriate Lagrangian multipliers that makes the matrix (90) positive semidefinite. In existing literature, this step is typically performed by numerically solving the dual problem, assisted by computers (see, for example, [6, 12]). However, in the continuous PEP, this step can be relatively straightforward. For instance, in Appendix H.1, we analytically derived the optimal dual variables, namely $\lambda_1(t) = t^2/T^2$ and $\lambda_2(t) = 2t/T^2$, for AGM ODE. After analyzing the continuous PEP for AGM ODE, one can attempt to set the Lagrange multiplier vectors in the discrete PEP for AGM by discretizing the Lagrange multiplier functions $\lambda_1(t) = t^2/T^2$ and $\lambda_2(t) = 2t/T^2$. Indeed, the discretization $\lambda_1^i = \theta_{i-1}^2/\theta_N^2$ and $\lambda_2^i = \theta_i/\theta_N^2$ works well, as shown in Appendix H.2.1.

H.3 Convergence rate of TMM ODE matches the known rate of TMM

The well-known convergence rate of TMM is as follows (see [3, Theorem 4.19]):

$$\begin{aligned} &f(y_N) - f(x^*) - \frac{s}{2}\|\nabla f(y_N)\|^2 - \frac{\mu}{2(1-\mu s)}\|y_N - s\nabla f(y_N) - x^*\|^2 \\ &\leq (1-\sqrt{\mu s})^{2N}\bigg(f(y_0) - f(x^*) - \frac{s}{2}\|\nabla f(y_0)\|^2 \\ &\qquad - \frac{\mu}{2(1-\mu s)}\|y_0 - s\nabla f(y_0) - x^*\|^2 + + \frac{\mu}{1-\mu s}\|y_0 - x^*\|^2\bigg). \end{aligned} \tag{95}$$

In Appendix E.1, we showed that TMM ODE is the limiting ODE of TMM under the approximations $t \approx \sqrt{s}k$, $T \approx \sqrt{s}N$, and $Y(\sqrt{s}k) \approx y_k$. It is straightforward to check that taking the limit $s \to 0$ in the inequality (95) gives

$$f(Y(T)) - f(x^*) - \frac{\mu}{2}\|Y(T) - x^*\|^2 \leq e^{-2\sqrt{\mu}T}\left(f(y_0) - f(x^*) + \frac{\mu}{2}\|y_0 - x^*\|^2\right),$$

which coincides with the convergence guarantee of TMM ODE obtained in Section 3.3.

H.4 Convergence rate of ITEM ODE matches the known rate of ITEM

In [12], the convergence rate of ITEM is shown as follows:

$$\begin{aligned} &f(y_N) - f(x^*) - \frac{s}{2}\|\nabla f(y_N)\|^2 \\ &\qquad - \frac{\mu}{2(1-\mu s)}\|y_N - s\nabla f(y_N) - x^*\|^2 \leq \frac{1}{s(1-\mu s)A_N}\|y_0 - x^*\|^2. \end{aligned} \tag{96}$$

In Appendix E.2, we showed that ITEM ODE is the limiting ODE of ITEM under the approximations $t \approx \sqrt{s}k$, $T \approx \sqrt{s}N$, and $Y(\sqrt{s}k) \approx y_k$. Because we can approximate the sequence $\{A_k\}$ as $A_k \approx \frac{1}{\mu s}\sinh^2\left(\sqrt{\mu s}k\right)$, we can check that taking the limit $s \to 0$ in the inequality 96 gives

$$f(Y(T)) - f(x^*) - \frac{\mu}{2}\|Y(T) - x^*\|^2 \leq \mu\,\mathrm{csch}^2\left(\sqrt{\mu}T\right)\|y_0 - x^*\|^2,$$

which coincides with the convergence guarantee of ITEM ODE obtained in Section 3.3.

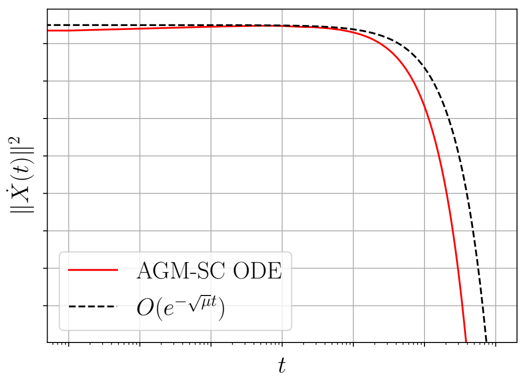

Figure 1: Comparison between the actual performance of AGM-SC ODE and the theoretical convergence guarantee (28).

I Missing details from Section 4

I.1 Numerical experiment for the convergence guarantee (28)

To empirically validate the convergence guarantee (28), we consider the minimization of the following objective function:

$$f(x_1, x_2) = 2 \times 10^{-2} x_1^2 + 5 \times 10^{-3} x_2^2,$$

starting from the initial point $x_0 = (1, 1)$. This simple problem was used in [9]. Note that f is μ-strongly convex with $\mu = 0.01$. The result is shown in Figure 1.

I.2 Novel ODE models for minimizing velocity and gradient norm

TMM ODE. We analyze the convergence rate of TMM ODE on the squared velocity norm. This ODE model is the anti-transposed dynamics of itself because it can be expressed as (20) with $H^G(t,\tau) = -2e^{\sqrt{\mu}(\tau-t)} + 4e^{2\sqrt{\mu}(\tau-t)}$. In Theorem 2, we choose $\lambda^G(t) = e^{-2\sqrt{\mu}t}$. By setting $t_{\text{end}} = T$ and $\alpha^G(t) = C(T)\bar{H}^G(T,t)$ with $C(T) = \sqrt{\mu}/2$, we compute the PEP kernel (23) as (see Appendix G.8)

$$S^G(t,\tau) = 2\left(\nu - \mu e^{-2\sqrt{\mu}T}\right) e^{-\sqrt{\mu}(t+\tau)}, \tag{97}$$

which is the anti-transpose of the PEP kernel (15). Thus, the kernel (97) is positive semidefinite for $\nu = \mu e^{-2\sqrt{\mu}T}$. Therefore, Theorem 2 guarantees that TMM ODE achieves the following convergence guarantee:

$$\left\| \frac{\sqrt{\mu}}{2}\dot{X}(T) \right\|^2 \leq \mu e^{-2\sqrt{\mu}T} \sup_{x\in\mathbb{R}^d} \left\{ \hat{f}(x_0) - \hat{f}(x) \right\}, \tag{98}$$

which is a novel result.

ITEM-G ODE. We consider the following novel ODE model:

$$\ddot{X} + 3\sqrt{\mu}\coth_{T-t}\dot{X} + 2\nabla f(X) = 0, \tag{ITEM-G ODE}$$

where $\coth_{T-t}$ denotes the corresponding hyperbolic function with the argument $\sqrt{\mu}(T-t)$. This ODE model is the anti-transposed dynamics of ITEM ODE, as it can be expressed as (20) with $H^G(t,\tau) = 4\sinh_{T-t}\cosh_{T-t}\coth_{T-\tau}\operatorname{csch}^2_{T-\tau} + 2\sinh_{T-t}\operatorname{csch}_{T-\tau}(1 - 2\coth^2_{T-\tau})$ (see Appendix F.7). We choose $\lambda^G(t) = \frac{\operatorname{csch}^2_{T-t}}{\operatorname{csch}^2_T}$. By setting $t_{\text{end}} < T$ and $\alpha^G(t) = C(t_{\text{end}})\bar{H}^G(t_{\text{end}}, t)$ with $C(t_{\text{end}}) = \frac{\sqrt{\mu}}{2}\operatorname{csch}_{T-t_{\text{end}}}$, the PEP kernel (23) is given by (see Appendix G.9):

$$S^G(t,\tau) = 2\operatorname{csch}^2_T\left(\nu - \mu\operatorname{csch}^2_T\right)\sinh_{T-t}\sinh_{T-\tau}, \tag{99}$$

which is the anti-transpose of (17). Thus, the kernel (99) is positive semidefinite for $\nu = \mu \operatorname{csch}_T^2$. Therefore, Theorem 2 implies that ITEM-G ODE achieves the following convergence guarantee:

$$\left\| \frac{\sqrt{\mu}}{2 \sinh_{T-t_{\text{end}}}} \dot{X}(t_{\text{end}}) \right\|^2 \leq \mu \operatorname{csch}_T^2 \sup_{x \in \mathbb{R}^d} \left\{ \hat{f}(x_0) - \hat{f}(x) \right\}.$$

By using a similar argument as in the case of OGM-G ODE, we can show that the left-hand side of this inequality converges to $\|\nabla f(X(T))\|^2/4$ as $t_{\text{end}} \to T$ (see Appendix G.9). Consequently, we have the following convergence rate on $\|\nabla f(X(T))\|^2$:

$$\|\nabla f(X(T))\|^2 \leq 4\mu \operatorname{csch}_T^2 \sup_{x \in \mathbb{R}^d} \left\{ \hat{f}(x_0) - \hat{f}(x) \right\}, \tag{100}$$

which is a novel result.

I.3 Lyapunov analysis for minimizing velocity and gradient norm

In Section 4.2 and Appendix I.2, we analyzed the convergence rates of various ODE models on velocity or gradient norms, within the continuous PEP framework. In this subsection, we present an alternative approach to obtain the same convergence guarantees. The proof relies on Lyapunov functions and L'Hôpital's rule, similar to the convergence analysis of OGM-G ODE and Unified AGM-G ODE presented in [10, 7]. We provide detailed computational steps for the analysis of OGM-G ODE and AGM-SC ODE, and present only the essential steps for the remaining examples.

Because $\left[\frac{\partial}{\partial t} \left\{ \hat{f}_t(y) - \hat{f}_t(X(T)) \right\} \right]_{y=X(t)} \leq 0$ (see Appendix D), we have

$$\begin{aligned} \frac{d}{dt} \left\{ \hat{f}_t(X(t)) - \hat{f}_t(X(T)) \right\} &= \left[\frac{\partial}{\partial t} \left\{ \hat{f}_t(y) - \hat{f}_t(X(T)) \right\} \right]_{y=X(t)} \\ &\quad + \left[\left\langle \nabla_y \left\{ \hat{f}_t(y) - \hat{f}_t(X(T)) \right\}, \dot{X}(t) \right\rangle \right]_{y=X(t)} \\ &\leq \left\langle \nabla \hat{f}_t(X(t)), \dot{X}(t) \right\rangle. \end{aligned}$$

OGM-G ODE. We bound the time derivative of $\hat{f}_t(X(t)) - \hat{f}_t(X(T))$ as follows:

$$\begin{aligned} \frac{d}{dt} \left\{ \hat{f}_t(X(t)) - \hat{f}_t(X(T)) \right\} &\leq \left\langle \nabla \hat{f}_t(X(t)), \dot{X}(t) \right\rangle \\ &= -\frac{T^2}{(T-t)^2} \left\langle \ddot{X}(t) + \frac{1}{T-t} \dot{X}(t), \dot{X}(t) \right\rangle \\ &= -\frac{d}{dt} \left\{ \frac{T^2}{2(T-t)^2} \left\| \dot{X}(t) \right\|^2 \right\}, \end{aligned}$$

where the second equality follows from (75). Therefore, the following energy function is decreasing:

$$\mathcal{E}(t) = \hat{f}_t(X(t)) - \hat{f}_t(X(T)) + \frac{T^2}{2(T-t)^2} \left\| \dot{X}(t) \right\|^2.$$

We now show that this energy function is equivalent to the one provided in [10]. By (74), we have

$$\begin{aligned} \tilde{f}_t(y) &= \frac{T^2}{(T-t)^2} \hat{f}(y) - \left\langle \int_0^t \frac{2T^2}{(T-t)^3} \nabla \hat{f}(X(\tau))\, d\tau, y \right\rangle \\ &= \frac{T^2}{(T-t)^2} \hat{f}(y) + \left\langle \frac{2T^2}{(T-t)^3} \dot{X}(t), y \right\rangle. \end{aligned}$$

Consequently, we can rewrite the energy function as follows:

$$\begin{aligned} \mathcal{E}(t) &= \hat{f}_t(X(t)) - \hat{f}_t(X(T)) + \frac{T^2}{2(T-t)^2} \left\| \dot{X}(t) \right\|^2 \\ &= \frac{T^2}{(T-t)^2} \left(f(X(t)) - f(X(T)) \right) \end{aligned}$$

$$
\begin{aligned}
&+ \frac{2T^2}{(T-t)^3}\left\langle \dot X(t), X(t)-X(T)\right\rangle + \frac{T^2}{2(T-t)^2}\left\|\dot X(t)\right\|^2 \\
&= \frac{T^2}{(T-t)^2}\left(f(X(t))-f(X(T))\right) \\
&- \frac{2T^2}{(T-t)^4}\|X(t)-X(T)\|^2 + \frac{2T^2}{(T-t)^4}\left\|X(t)+\frac{T-t}{2}\dot X(t)-X(T)\right\|^2,
\end{aligned}
$$

which recovers the known energy function for OGM-G ODE in [10].

AGM-SC ODE. We bound the time derivative of $\hat f_t(X(t)) - \hat f_t(X(T))$ as follows:

$$
\begin{aligned}
\frac{d}{dt}\left\{\hat f_t(X(t)) - \hat f_t(X(T))\right\} &\le \left\langle \nabla \hat f_t(X(t)), \dot X(t)\right\rangle \\
&= -e^{\sqrt{\mu}t}\left\langle \ddot X(t) + \sqrt{\mu}\dot X(t), \dot X(t)\right\rangle \\
&= -\frac{d}{dt}\left\{\frac{e^{\sqrt{\mu}t}}{2}\left\|\dot X(t)\right\|^2\right\} - \frac{\sqrt{\mu}e^{\sqrt{\mu}t}}{2}\left\|\dot X(t)\right\|^2.
\end{aligned}
$$

where the second equality follows from (63). Therefore, the following function is decreasing:

$$
\mathcal{E}(t) = \hat f_t(X(t)) - \hat f_t(X(T)) + \frac{e^{\sqrt{\mu}t}}{2}\left\|\dot X(t)\right\|^2.
$$

We now provide an alternative expression for this energy function. By (62), we have

$$
\begin{aligned}
\tilde f_t(y) &= e^{\sqrt{\mu}t}\hat f(y) - \sqrt{\mu}\left\langle \int_0^t e^{\sqrt{\mu}\tau}\nabla \hat f(X(\tau))\,d\tau, y\right\rangle \\
&= e^{\sqrt{\mu}t}\hat f(y) + \left\langle \sqrt{\mu}e^{\sqrt{\mu}t}\dot X(t) + \mu e^{\sqrt{\mu}t}(X(t)-x_0), y\right\rangle \\
&= e^{\sqrt{\mu}t} f(y) - \frac{\mu e^{\sqrt{\mu}t}}{2}\|y-x_0\|^2 + \left\langle \sqrt{\mu}e^{\sqrt{\mu}t}\dot X(t) + \mu e^{\sqrt{\mu}t}(X(t)-x_0), y\right\rangle.
\end{aligned}
$$

Consequently, we can rewrite the energy function as follows:

$$
\begin{aligned}
\mathcal{E}(t) &= e^{\sqrt{\mu}t}\left(f(X(t)) - \frac{\mu}{2}\|X(t)-x_0\|^2 - f(X(T)) + \frac{\mu}{2}\|X(T)-x_0\|^2\right) \\
&\quad + \left\langle \sqrt{\mu}e^{\sqrt{\mu}t}\dot X(t) + \mu e^{\sqrt{\mu}t}\left(X(t)-x_0\right), X(t)-X(T)\right\rangle + \frac{e^{\sqrt{\mu}t}}{2}\left\|\dot X(t)\right\|^2 \\
&= e^{\sqrt{\mu}t}\left(f(X(t)) - f(X(T)) + \frac{\mu}{2}\|X(t)-X(T)\|^2\right) \\
&\quad + \left\langle \sqrt{\mu}e^{\sqrt{\mu}t}\dot X(t), X(t)-X(T)\right\rangle + \frac{e^{\sqrt{\mu}t}}{2}\left\|\dot X(t)\right\|^2 \\
&= e^{\sqrt{\mu}t}\left(f(X(t)) - f(X(T)) + \frac{\mu}{2}\left\|X(t)+\frac{1}{\sqrt{\mu}}\dot X(t) - X(T)\right\|^2\right).
\end{aligned}
$$

We now have

$$
\frac{e^{\sqrt{\mu}T}}{2}\left\|\dot X(T)\right\|^2 = \mathcal{E}(T) \le \mathcal{E}(0) = \hat f(x_0) - \hat f(X(T)),
$$

which recovers the convergence guarantee (28) obtained in Section 4.2.

Unified AGM-G ODE. We have

$$
\begin{aligned}
\frac{d}{dt}\left\{\hat f_t(X(t)) - \hat f_t(X(T))\right\} &\le \left\langle \nabla \hat f_t(X(t)), \dot X(t)\right\rangle \\
&= -\frac{\sinh_T^2}{\sinh_{T-t}^2}\left\langle \ddot X(t) + \frac{\sqrt{\mu}}{2}\left(\coth_{T-t} + \tanh_{T-t}\right)\dot X(t), \dot X(t)\right\rangle \\
&= -\frac{d}{dt}\left\{\frac{\sinh_T^2}{2\sinh_{T-t}^2}\left\|\dot X(t)\right\|^2\right\} - \frac{\sqrt{\mu}\sinh_T^2}{2\sinh_{T-t}\cosh_{T-t}}\left\|\dot X(t)\right\|^2
\end{aligned}
$$

where the second equality follows from (79). Therefore, the following function is decreasing:

$$\mathcal{E}(t)=\hat{f}_t(X(t))-\hat{f}_t(X(T))+\frac{\sinh_T^2}{2\sinh_{T-t}^2}\left\|\dot{X}(t)\right\|^2.$$

We now show that this energy function is equivalent to the one provided in [7]. By (78), we have

$$\begin{aligned}\tilde{f}_t(y)&=\frac{\sinh_T^2}{\sinh_{T-t}^2}f(y)-\frac{\mu}{2}\frac{\sinh_T^2}{\sinh_{T-t}^2}\|y-x_0\|^2\\&+\left\langle\sqrt{\mu}\frac{\sinh_T^2}{\sinh_{T-t}^2\tanh_{T-t}}\dot{X}(t)+\mu\frac{\sinh_T^2}{\sinh_{T-t}^2}\left(X(t)-x_0\right),y\right\rangle\end{aligned}$$

After performing the calculations, we can rewrite the energy function as follows:

$$\begin{aligned}\mathcal{E}(t)&=\frac{\sinh_T^2}{\sinh_{T-t}^2}\left(f(X(t))-f(X(T))\right)-\frac{\mu\sinh_T^2}{2\sinh_{T-t}^4}\|X(t)-X(T)\|^2\\&+\frac{\mu\sinh_T^2}{2\sinh_{T-t}^2\tanh_{T-t}^2}\left\|X(t)+\frac{\tanh_{T-t}}{\sqrt{\mu}}\dot{X}(t)-X(T)\right\|^2,\end{aligned}$$

which recovers the known energy function for Unified AGM-G ODE in [7].

TMM ODE. We first bound the time derivative of $\hat{f}_t(X(t))-\hat{f}_t(X(T))$ as

$$\begin{aligned}\frac{d}{dt}\left\{\hat{f}_t(X(t))-\hat{f}_t(x^*)\right\}&\le\left\langle\nabla\hat{f}_t(X(t)),\dot{X}(t)\right\rangle\\&=-\frac{e^{2\sqrt{\mu}t}}{2}\left\langle\ddot{X}(t)+\sqrt{\mu}\dot{X}(t),\dot{X}(t)\right\rangle\\&=-\frac{d}{dt}\left\{\frac{e^{2\sqrt{\mu}t}}{4}\left\|\dot{X}(t)\right\|^2\right\},\end{aligned}$$

where the second equality follows from (70). Therefore, the energy function:

$$\mathcal{E}(t)=\hat{f}_t(X(t))-\hat{f}_t(X(T))+\frac{e^{2\sqrt{\mu}t}}{4}\left\|\dot{X}(t)\right\|^2$$

is decreasing. We now provide another expression for this energy function. By (69), we have

$$\tilde{f}_t(y)=e^{2\sqrt{\mu}t}f(y)-\frac{\mu}{2}e^{2\sqrt{\mu}t}\|y-x_0\|^2+\left\langle\sqrt{\mu}e^{2\sqrt{\mu}t}\dot{X}(t)+\mu e^{2\sqrt{\mu}t}\left(X(t)-x_0\right),y\right\rangle.$$

After performing some calculations, we can rewrite the energy function as follows:

$$\begin{aligned}\mathcal{E}(t)&=e^{2\sqrt{\mu}t}\left(f(X(t))-f(X(T))\right)-\frac{\mu}{2}e^{2\sqrt{\mu}t}\|X(t)-X(T)\|^2\\&+\mu e^{2\sqrt{\mu}t}\left\|X(t)+\frac{1}{2\sqrt{\mu}}\dot{X}(t)-X(T)\right\|^2.\end{aligned}$$

We now have

$$\frac{e^{2\sqrt{\mu}T}}{4}\left\|\dot{X}(T)\right\|^2=\mathcal{E}(T)\le\mathcal{E}(0)=\hat{f}(x_0)-\hat{f}(X(T)),$$

which recovers the convergence guarantee (98) obtained in Appendix I.2.

ITEM-G ODE. We bound the time derivative of $\hat{f}_t(X(t))-\hat{f}_t(X(T))$ as follows:

$$\begin{aligned}\frac{d}{dt}\left\{\hat{f}_t(X(t))-\hat{f}_t(X(T))\right\}&\le\left\langle\nabla\hat{f}_t(X(t)),\dot{X}(t)\right\rangle\\&=-\frac{\sinh_T^2}{2\sinh_{T-t}^2}\left\langle\ddot{X}(t)+\sqrt{\mu}\coth_{T-t}\dot{X}(t),\dot{X}(t)\right\rangle\\&=-\frac{d}{dt}\left\{\frac{\sinh_T^2}{4\sinh_{T-t}^2}\left\|\dot{X}(t)\right\|^2\right\},\end{aligned}$$

where the second equality follows from (83). Therefore, the following function is decreasing:

$$\mathcal{E}(t) = \hat{f}_t(X(t)) - \hat{f}_t(X(T)) + \frac{\sinh_T^2}{4\sinh_{T-t}^2}\left\|\dot{X}(t)\right\|^2.$$

We now provide an alternative expression for this energy function. By (82), we have

$$\begin{aligned}\tilde{f}_t(y) &= \frac{\sinh_T^2}{\sinh_{T-t}^2} f(y) - \frac{\mu}{2}\frac{\sinh_T^2}{\sinh_{T-t}^2}\|y-x_0\|^2 \\ &\quad + \left\langle \sqrt{\mu}\frac{\sinh_T^2}{\sinh_{T-t}^2\tanh_{T-t}}\dot{X}(t) + \mu\frac{\sinh_T^2}{\sinh_{T-t}^2}(X-x_0), y\right\rangle.\end{aligned}$$

After performing some calculations, we obtain the following equivalent form for the energy function:

$$\begin{aligned}\mathcal{E}(t) &= \frac{\sinh_T^2}{\sinh_{T-t}^2}(f(X(t)) - f(X(T))) - \frac{\mu\sinh_T^2\left(1+\cosh_{T-t}^2\right)}{2\sinh_{T-t}^4}\|X(t)-X(T)\|^2 \\ &\quad + \frac{\mu\sinh_T^2}{\sinh_{T-t}^2\tanh_{T-t}^2}\left\|X(t) + \frac{\tanh_{T-t}}{2\sqrt{\mu}}\dot{X}(t) - X(T)\right\|^2.\end{aligned}$$

To derive a convergence rate on $\|\nabla f(X(T))\|^2$, we compute $\lim_{t\nearrow T}\mathcal{E}(t)$ by employing a similar argument to the one presented in [10, Appendix D.5]. Note that the solution $X : [0,T) \to \mathbb{R}^d$ can be continuously extended to $t = T$ (see Appendix G.9). We first show $\ddot{X}(T) = \nabla f(X(T))$ as follows:

$$\begin{aligned}0 &= \lim_{t\nearrow T}\left(\ddot{X}(t) + 3\sqrt{\mu}\coth_{T-t}\dot{X}(t) + 2\nabla f(X(t))\right) \\ &= \ddot{X}(T) + 3\lim_{t\nearrow T}\frac{\dot{X}(t)}{T-t} + 2\nabla f(X(T)) \\ &= \ddot{X}(T) - 3\lim_{t\nearrow T}\ddot{X}(t) + 2\nabla f(X(T)) \\ &= -2\ddot{X}(T) + 2\nabla f(X(T)),\end{aligned}$$

where we used L'Hôpital's rule for the third equality. By using L'Hôpital's rule again, we have

$$\begin{aligned}\lim_{t\nearrow T}\frac{f(X(t)) - f(X(T))}{\sinh_{T-t}^2} &= \lim_{t\nearrow T}\frac{f(X(t)) - f(X(T))}{\mu(T-t)^2} = \lim_{t\nearrow T}\frac{\left\langle\nabla f(X(t)), \dot{X}\right\rangle}{2\mu(t-T)} \\ &= \frac{1}{2\mu}\left\langle\nabla f(X(t)), \ddot{X}(T)\right\rangle = \frac{1}{2\mu}\|\nabla f(X(T))\|^2, \\ \lim_{t\nearrow T}\frac{X(t)-X(T)}{\sinh_{T-t}^2} &= \lim_{t\nearrow T}\frac{X(t)-X(T)}{\mu(T-t)^2} = \lim_{t\nearrow T}\frac{\dot{X}(t)}{2\mu(t-T)} \\ &= \frac{1}{2\mu}\ddot{X}(T) = \frac{1}{2\mu}\nabla f(X(T)).\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}\lim_{t\nearrow T}\mathcal{E}(t) &= \sinh_T^2\lim_{t\nearrow T}\frac{f(X(t)) - f(X(T))}{\sinh_{T-t}^2} - \mu\sinh_T^2\left\|\lim_{t\nearrow T}\frac{X(t)-X(T)}{\sinh_{T-t}^2}\right\|^2 \\ &\quad + \mu\sinh_T^2\left\|\lim_{t\nearrow T}\frac{X(t)-X(T)}{\sinh_{T-t}^2} + \lim_{t\nearrow T}\frac{\dot{X}(t)}{2\sqrt{\mu}\sinh_{T-t}}\right\|^2 \\ &= \frac{\sinh_T^2}{2\mu}\|\nabla f(X(T))\|^2 - \frac{\sinh_T^2}{4\mu}\|\nabla f(X(T))\|^2 \\ &\quad + \mu\cosh_{T-t}^2\sinh_T^2\left\|\frac{1}{2\mu}\nabla f(X(T)) - \frac{1}{2\mu}\nabla f(X(T))\right\|^2\end{aligned}$$

$$= \frac{\sinh_T^2}{4\mu} \left\| \nabla f(X(T)) \right\|^2 .$$

We can now derive the convergence guarantee on $\|\nabla f(X(T))\|^2$ as follows:

$$\frac{\sinh_T^2}{4\mu} \left\| \nabla f(X(T)) \right\|^2 = \lim_{t \nearrow T} \mathcal{E}(t) \leq \mathcal{E}(0) = \hat{f}(x_0) - \hat{f}(X(T)),$$

which recovers the convergence guarantee (100) obtained in Appendix I.2.

J Missing details from Section 5

J.1 Proof of Proposition 3

For all $t, \tau \in (0, T)$, we have

$$\begin{aligned}
S^F(t,\tau) &= \nu \left(\lambda^F(t) H^F(t,\tau) + \dot{\lambda}^F(t) \int_\tau^t H^F(s,\tau)\, ds \right) - 2\alpha^F(t)\alpha^F(\tau) \\
&= \nu \left(\frac{1}{\lambda^G(T-t)} H^G(T-\tau, T-t) + \frac{d}{dt} \left\{ \frac{1}{\lambda^G(T-t)} \right\} \int_\tau^t H^G(T-\tau, T-s)\, ds \right) \\
&\quad - 2\alpha^G(T-t)\alpha^G(T-\tau) \\
&= \nu \bar{H}^G(T-\tau, T-t) - 2\alpha^G(T-t)\alpha^G(T-\tau) \\
&= S^G(T-\tau, T-t),
\end{aligned}$$

where the third equality follows from (49). □

J.2 Proof of Proposition 4

We first rewrite $\int_0^{t_{\mathrm{end}}} \alpha^G(t) \nabla \hat{f}_t(X(t))\, dt$ in terms of $\nabla \hat{f}(X(t))$:

$$\begin{aligned}
&\int_0^{t_{\mathrm{end}}} \alpha^G(t) \nabla \hat{f}_t(X(t))\, dt \\
&= \int_0^{t_{\mathrm{end}}} \alpha^G(t) \left(\lambda^G(t) \nabla \hat{f}(X(t)) - \int_0^t \dot{\lambda}^G(\tau) \nabla \hat{f}(X(\tau))\, d\tau \right) dt \\
&= \int_0^{t_{\mathrm{end}}} \alpha^G(t)\lambda^G(t) \nabla \hat{f}(X(t))\, dt - \int_0^{t_{\mathrm{end}}} \int_t^{t_{\mathrm{end}}} \alpha^G(\tau) \dot{\lambda}^G(t) \nabla \hat{f}(X(t))\, d\tau dt \\
&= \int_0^{t_{\mathrm{end}}} \left(\alpha^G(t)\lambda^G(t) - \left(\int_t^{t_{\mathrm{end}}} \alpha^G(\tau)\, d\tau \right) \dot{\lambda}^G(t) \right) \nabla \hat{f}(X(t)) dt \\
&= \int_0^{t_{\mathrm{end}}} \frac{d}{dt} \left\{ \lambda^G(t) \int_t^{t_{\mathrm{end}}} \left(-\alpha^G(\tau) \right) d\tau \right\} \nabla \hat{f}(X(t))\, dt.
\end{aligned}$$

Claim 2. *The function* $t \mapsto -\frac{d}{dt} \left\{ \lambda^G(t) \int_t^{t_{\mathrm{end}}} \alpha^G(\tau)\, d\tau \right\} \mathbf{1}_{[0, t_{\mathrm{end}}]}(t)$ *converges to the function* $t \mapsto \frac{1}{2}\delta_T(t) - \frac{\mu}{2} \int_t^T H^G(s,t)\, ds$ *as* $t_{\mathrm{end}} \to T$*, where* δ_T *is the Dirac delta function centered at* $t = T$*.*

Assuming that Claim 2 holds, we complete the proof of Proposition 4. As $t_{\mathrm{end}} \to T$, we have

$$\begin{aligned}
\int_0^{t_{\mathrm{end}}} \alpha^G(t) \nabla \hat{f}_t(X(t))\, dt &= - \int_0^{t_{\mathrm{end}}} \frac{d}{dt} \left\{ \lambda^G(t) \int_t^{t_{\mathrm{end}}} \alpha^G(\tau)\, d\tau \right\} \nabla \hat{f}(X(t))\, dt \\
&\to \int_0^T \left(\frac{1}{2}\delta_T(t) - \frac{\mu}{2} \int_t^T H^G(s,t)\, ds \right) \nabla \hat{f}(X(t))\, dt \\
&= \frac{1}{2} \nabla \hat{f}(X(T)) - \frac{\mu}{2} \int_0^T \int_t^T H^G(s,t) \nabla \hat{f}(X(t))\, ds dt
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\nabla \hat{f}(X(T)) + \frac{\mu}{2}(X(T) - x_0) \\
&= \frac{1}{2}\nabla f(X(T)),
\end{aligned}
$$

where the third equality follows from (54).

We now prove Claim 2. Note that $\alpha^G(t) = \alpha^F(T-t) = -\frac{d}{dt}\{\frac{A(t)}{\lambda^G(t)}\} = \frac{\dot{\lambda}^G(t)A(t)}{\lambda^G(t)^2} - \frac{\dot{A}(t)}{\lambda^G(t)}$, where $A(t) = \frac{1}{2} - \frac{\mu}{2}\int_t^T \int_\tau^T H^G(s,\tau)\, ds d\tau$. Now, we have

$$
\begin{aligned}
&\frac{d}{dt}\left\{\lambda^G(t)\int_t^{t_{\text{end}}} \left(-\alpha^G(\tau)\right)\, d\tau\right\} \\
&= \dot{\lambda}^G(t)\left(\frac{A(t_{\text{end}})}{\lambda^G(t_{\text{end}})} - \frac{A(t)}{\lambda^G(t)}\right) + \lambda^G(t)\left(\frac{\dot{\lambda}^G(t)}{\lambda^G(t)^2}A(t) - \frac{\dot{A}(t)}{\lambda^G(t)}\right) \\
&= \frac{\dot{\lambda}^G(t)A(t_{\text{end}})}{\lambda^G(t_{\text{end}})} - \dot{A}(t) \\
&= \frac{\dot{\lambda}^G(t)}{2\lambda^G(t_{\text{end}})}\left(1 - \mu\int_{t_{\text{end}}}^T \int_\tau^T H^G(s,\tau)\, ds d\tau\right) - \frac{\mu}{2}\int_t^T H^G(s,t)\, ds.
\end{aligned}
$$

As $t_{\text{end}} \to T$, the function $t \mapsto \frac{\dot{\lambda}^G(t)}{\lambda^G(t_{\text{end}})}\mathbf{1}_{[0,t_{\text{end}}]}$ converges to δ_T, and $\int_{t_{\text{end}}}^T \int_\tau^T H^G(s,\tau)\, ds d\tau$ converges to 0. This completes the proof. $\square$

References

[1] Stephen P Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge university press, 2004.

[2] Hermann Brunner. *Collocation Methods for Volterra Integral and Related Functional Differential Equations*, volume 15. Cambridge university press, 2004.

[3] Alexandre d'Aspremont, Damien Scieur, and Adrien Taylor. Acceleration methods. *arXiv preprint arXiv:2101.09545*, 2021.

[4] Yoel Drori and Marc Teboulle. Performance of first-order methods for smooth convex minimization: A novel approach. *Mathematical Programming*, 145(1):451–482, 2014.

[5] Donghwan Kim and Jeffrey A Fessler. Optimized first-order methods for smooth convex minimization. *Mathematical programming*, 159(1):81–107, 2016.

[6] Donghwan Kim and Jeffrey A Fessler. Optimizing the efficiency of first-order methods for decreasing the gradient of smooth convex functions. *Journal of optimization theory and applications*, 188(1):192–219, 2021.

[7] Jungbin Kim and Insoon Yang. Unifying Nesterov's accelerated gradient methods for convex and strongly convex objective functions. In *International Conference on Machine Learning*, pages 16897–16954. PMLR, 2023.

[8] Yurii E Nesterov. A method for solving the convex programming problem with convergence rate $o(1/k^2)$. In *Dokl. akad. nauk Sssr*, volume 269, pages 543–547, 1983.

[9] Weijie Su, Stephen Boyd, and Emmanuel J Candes. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17:1–43, 2016.

[10] Jaewook J Suh, Gyumin Roh, and Ernest K Ryu. Continuous-time analysis of accelerated gradient methods via conservation laws in dilated coordinate systems. In *International Conference on Machine Learning*, pages 20640–20667. PMLR, 2022.

[11] Boya Sun, Jemin George, and Solmaz Kia. High-resolution modeling of the fastest first-order optimization method for strongly convex functions. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 4237–4242. IEEE, 2020.

[12] Adrien Taylor and Yoel Drori. An optimal gradient method for smooth strongly convex minimization. *Mathematical Programming*, pages 1–38, 2022.

[13] Bryan Van Scoy, Randy A Freeman, and Kevin M Lynch. The fastest known globally convergent first-order method for minimizing strongly convex functions. *IEEE Control Systems Letters*, 2(1):49–54, 2017.