Notes to "Convergence analysis of ODE models for accelerated first-order methods via positive semidefinite kernels" by Jungbin Kim and Insoon Yang

Jungbin Kim

March 10, 2026

The least intuitive idea in [1] is the "reparametrization to time-varying functions" introduced in Section 4.2:

$$\hat{f}_t(x) := \lambda^G(t)\hat{f}(x) - \left\langle \int_0^t \dot{\lambda}^G(\tau)\nabla\hat{f}(X(\tau))\,d\tau, x \right\rangle. \tag{1}$$

This idea emerged from a pattern observed while comparing the PEP analyses and the Lyapunov analyses for the five ODE models in [1, Section 4, Appendix E]. Extracting the essence of this pattern gives (1).

1 A guide to arriving at this idea

To understand this note, the reader needs to be familiar with the key ideas in [1] and the related works. I have not cited other papers which can be found in the bibliography of [1].

1.1 PEP analysis

I begin by describing what the PEP in [1, §4] is intended to do. The presentation here is slightly different from [1] (for example, by ignoring constant factors), but the underlying logic is the same. Suppose that the objective function f is μ-strongly convex. Define $\hat{f}(x) := f(x) - \frac{\mu}{2}\|x - x_0\|^2$.

Given a multiplier $\lambda^G(t)$, the convergence of the given ODE model will be proven by taking the weighted integral of the following two families of equalities and inequalities over t:

$$\begin{aligned} 0 &= \frac{d}{dt}\left\{\hat{f}(X(t)) - \hat{f}(X(T))\right\} - \left\langle \nabla\hat{f}(X(t)), \dot{X}(t)\right\rangle && \text{with weight } \lambda^G(t), \\ 0 &\geq \hat{f}(X(t)) - \hat{f}(X(T)) - \left\langle \nabla\hat{f}(X(t)), X(t) - X(T)\right\rangle && \text{with weight } \dot{\lambda}^G(t). \end{aligned}$$

Indeed, when $\lambda^G(0) = 1$ and $\lambda^G(T) < \infty$, we have

$$0 \geq -\hat{f}(X(0)) + \hat{f}(X(T)) - \int_0^T \left\langle \nabla\hat{f}(X(t)), \lambda^G(t)\dot{X}(t) + \dot{\lambda}^G(t)\left(X(t) - X(T)\right)\right\rangle ds.$$

If the goal is to establish the convergence rate $\|\dot{X}(T)\|^2 \le \nu(\hat{f}(x_0) - \hat{f}(X(T)))$, then it suffices to show

$$-\int_0^T \left\langle \nabla \hat{f}(X(t)), \lambda^G(t)\dot{X}(t) + \dot{\lambda}^G(t)\left(X(t) - X(T)\right)\right\rangle ds - \frac{1}{\nu}\|\dot{X}(T)\|^2 \ge 0. \tag{2}$$

PEP is a methodology to verify the non-negativity of the LHS of (2) by expressing it as a positive semidefinite quadratic form.

1.2 Lyapunov analysis

How is Lyapunov analysis typically done? Suppose we have an energy function

$$\mathcal{E}(t) = \lambda^G(t)\left(\hat{f}(X(t)) - \hat{f}(X(T))\right) + A(t),$$

where we want to show that $\mathcal{E}(t)$ is non-increasing. How is it usually proved? For a fixed $t \in (0, T)$, computing the time derivative of $\mathcal{E}(t)$ gives

$$\dot{\mathcal{E}}(t) = \dot{\lambda}^G(t)\left(\hat{f}(X(t)) - \hat{f}(X(T))\right) + \lambda^G(t)\frac{d}{dt}\left\{\hat{f}(X(t)) - \hat{f}(X(T))\right\} + \dot{A}(t).$$

By using

$$\begin{aligned} 0 &= \frac{d}{dt}\left\{\hat{f}(X(t)) - \hat{f}(X(T))\right\} - \left\langle \nabla \hat{f}(X(t)), \dot{X}(t)\right\rangle, \\ 0 &\ge \hat{f}(X(t)) - \hat{f}(X(T)) - \left\langle \nabla \hat{f}(X(t)), X(t) - X(T)\right\rangle, \end{aligned}$$

we obtain

$$\dot{\mathcal{E}}(t) \le \left\langle \nabla \hat{f}(X(t)), \lambda^G(t)\dot{X}(t) + \dot{\lambda}^G(t)\left(X(t) - X(T)\right)\right\rangle + \dot{A}(t). \tag{3}$$

Typically, the right-hand side of (3) is written as the negative of a sum of squares to conclude that $\dot{\mathcal{E}}(t) \le 0$.

1.3 Computing PEP kernel from Lyapunov analysis

The "PEP kernel" is the LHS of (2) (which is considered as a quadratic form) written as a kernel. When a Lyapunov analysis is known, we can rewrite the LHS of (2) as

$$-\int_0^T (\text{RHS of (3)})\, dt + A(T) - A(0) - \frac{1}{\nu}\|\dot{X}(T)\|^2. \tag{4}$$

If each term has a simple kernel expression, this provides an easy way to compute the PEP kernel.

1.4 Example: AGM-SC ODE

The ODE model for AGM-SC is

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

or equivalently,

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

The goal is to show the following convergence guarantee:

$$\left\|\dot{X}(T)\right\|^2 \leq 2e^{-\sqrt{\mu}T}\left(\hat{f}(x_0) - \hat{f}(X(T))\right).$$

A Lyapunov analysis can be found in [1, Appendix I.3]. The energy function is defined as

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

Applying the convexity of f and calculating, one obtains

$$\dot{\mathcal{E}}(t) \leq -\frac{\sqrt{\mu}e^{\sqrt{\mu}t}}{2}\left\|\dot{X}(t)\right\|^2.$$

The inequality $\mathcal{E}(T) \leq \mathcal{E}(0)$ simplifies to

$$\frac{e^{\sqrt{\mu}T}}{2}\left\|\dot{X}(T)\right\|^2 \leq \hat{f}(X(0)) - \hat{f}(X(T)),$$

which is the desired result.

This analysis fits the template in Section 1.2. For this example, we have

$$\begin{aligned}\lambda^G(t) &= e^{\sqrt{\mu}t},\\ A(t) &= \frac{\mu e^{\sqrt{\mu}t}}{2}\left(\left\|X(t) - x_0\right\|^2 - \left\|X(T) - x_0\right\|^2 + \left\|X(t) + \frac{1}{\sqrt{\mu}}\dot{X}(t) - X(T)\right\|^2\right).\end{aligned}$$

Can we compute the PEP kernel using the formula (4)? For this example, we have

$$\begin{aligned}\nu &= 2e^{-\sqrt{\mu}T},\\ \text{(RHS of (3))} &= -\frac{\sqrt{\mu}e^{\sqrt{\mu}t}}{2}\left\|\dot{X}(t)\right\|^2,\\ A(0) &= 0,\\ A(T) &= \frac{e^{\sqrt{\mu}T}}{2}\left\|\dot{X}(T)\right\|^2.\end{aligned}$$

After substitution, the last three terms in (4) cancel out. The PEP kernel is thus the kernel expression for the quantity

$$\int_0^T \frac{\sqrt{\mu}e^{\sqrt{\mu}s}}{2}\left\|\dot{X}(s)\right\|^2 ds,$$

where s is used instead of t to avoid overlapping notation.

Since $\dot{X}(s) = -\int_0^s H^G(s,r)\nabla\hat{f}(X(r))\,dr$, the kernel expression for $\|\dot{X}(s)\|^2$ in the basis $\{\nabla\hat{f}(X(t))\}$ is given by

$$K(t,\tau;s) = \begin{cases} H^G(s,t)H^G(s,\tau), & \text{if } 0 \le t,\tau \le s, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the PEP kernel, which is the kernel expression for $\int_0^T \frac{\sqrt{\mu}e^{\sqrt{\mu}s}}{2}\|\dot{X}(s)\|^2\,ds$, can be obtained by substituting $H^G(t,\tau) = (1+\sqrt{\mu}\tau - \sqrt{\mu}t)e^{\sqrt{\mu}(\tau-t)}$ (see [1, Appendix F.1]) into

$$\text{“}S^G(t,\tau)\text{”} = \int_0^T \frac{\sqrt{\mu}e^{\sqrt{\mu}s}}{2}K(t,\tau;s)\,ds = \begin{cases} \int_t^T \frac{\sqrt{\mu}e^{\sqrt{\mu}s}}{2}H^G(s,t)H^G(s,\tau)\,ds & \text{if } 0 \le \tau \le t \le T, \\ \int_\tau^T \frac{\sqrt{\mu}e^{\sqrt{\mu}s}}{2}H^G(s,t)H^G(s,\tau)\,ds & \text{if } 0 \le t < \tau \le T. \end{cases} \tag{5}$$

The kernel “$S^G(t,\tau)$” calculated here is not as clean as its counterpart $S^F(t,\tau)$, which is given by (see [1, §3.3])

$$S^F(t,\tau) = \begin{cases} \frac{\mu}{2}e^{-2\sqrt{\mu}T}(e^{\sqrt{\mu}\tau}-1), & \text{if } 0 \le \tau \le t \le T, \\ \frac{\mu}{2}e^{-2\sqrt{\mu}T}(e^{\sqrt{\mu}t}-1), & \text{if } 0 \le t < \tau \le T. \end{cases}$$

Thus, there was an impression that “$S^G(t,\tau)$” is not the “right” PEP kernel. Using the basis $\{\nabla f(X(t))\}$ instead of $\{\nabla\hat{f}(X(t))\}$ will provide a cleaner expression, but it still feels like we have not found the right approach yet.

The most natural candidate for the “right” PEP kernel $S^G(t,\tau)$ is the anti-transpose of $S^F(t,\tau)$, namely

$$S^G(t,\tau) = \begin{cases} \frac{\mu}{2}e^{-2\sqrt{\mu}T}(e^{\sqrt{\mu}(T-t)}-1), & \text{if } 0 \le \tau \le t \le T, \\ \frac{\mu}{2}e^{-2\sqrt{\mu}T}(e^{\sqrt{\mu}(T-\tau)}-1), & \text{if } 0 \le t < \tau \le T. \end{cases}$$

From this and (5), can we find the “right” H-kernel $\bar{H}^G(t,\tau)$? While there are many candidates for $\bar{H}^G(t,\tau)$ which gives the “right” PEP kernel $S^G(t,\tau)$, the simplest one is $\bar{H}^G(t,\tau) = e^{-\sqrt{\mu}t}$ (ignoring a constant factor). One might temporarily believe that this is the “right” H-kernel.

What is this? We know that H^G is a kernel representing the operator which maps $\nabla\hat{f}(X(t))$ to $-\dot{X}(t)$. Let's interpret $\bar{H}^G$ as a kernel representing an operator that maps some $g(t)$ to $-\dot{X}(t)$. By applying the technique from [1, Appendix F] to translate this H-kernel into a second-order ODE, one arrives at

$$\ddot{X}(t) + \sqrt{\mu}\dot{X}(t) + e^{-\sqrt{\mu}t}g(t) = 0.$$

What is the relationship between $g(t)$ and $\nabla f(X(t))$? We have $g(t) = e^{\sqrt{\mu}t}(\sqrt{\mu}\dot{X}(t) + \nabla f(X(t)))$. This expression does not look unfamiliar. The energy function $\mathcal{E}(t)$ can be rewritten as

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

where

$$B(y;t) = e^{\sqrt{\mu}t}\left(f(y) - f(X(T)) + \frac{\mu}{2}\|X(t)-X(T)\|^2 + \left\langle\sqrt{\mu}\dot{X}(t), y-X(T)\right\rangle\right). \tag{6}$$

We have

$$[\nabla_y B(y;t)]_{y=X(t)} = e^{\sqrt{\mu}t}\left(\nabla f(X(t)) + \sqrt{\mu}\dot{X}(t)\right) = g(t).$$

Thus, the kernel $\bar{H}^G$ represents the operator that maps the gradient of $B(y;t)$ at $X(t)$ (as a function of t) to $-\dot{X}(t)$.

At this point, one can also see that $S^G(t,\tau)$ is structured as $\bar{H}^G(t,\tau)$ minus something squared: ignoring constant factors, we have $S^G(t,\tau) = e^{-\sqrt{\mu}t} - e^{-\sqrt{\mu}T}$, where the first term is $\bar{H}^G(t,\tau)$ and the second term is $-e^{-\sqrt{\mu}T}$ times $\mathbf{1}_{[0,T]}(t)\mathbf{1}_{[0,T]}(\tau)$. As one can understand from the proof in [1, Appendix D], this is because

$$\begin{aligned}
&\left[\frac{\partial}{\partial t}B(y;t)\right]_{y=X(t)} \\
&\quad= \sqrt{\mu}e^{\sqrt{\mu}t}\left(f(X(t)) - f(X(T)) + \frac{\mu}{2}\left\|X(t)-X(T)\right\|^2 + \left\langle \ddot{X}(t) + 2\sqrt{\mu}\dot{X}(t), X(t)-X(T)\right\rangle\right) \\
&\quad= \sqrt{\mu}e^{\sqrt{\mu}t}\left(f(X(t)) - f(X(T)) + \frac{\mu}{2}\left\|X(t)-X(T)\right\|^2 - \langle \nabla f(X(t)), X(t)-X(T)\rangle\right) \\
&\quad= \sqrt{\mu}e^{\sqrt{\mu}t}\left(\hat{f}(X(t)) - \hat{f}(X(T)) - \left\langle \nabla \hat{f}(X(t)), X(t)-X(T)\right\rangle\right).
\end{aligned}$$

The observations we have made so far apply just as well to other examples in [1, Section 4, Appendix E]. Can we extract the essence of the argument? What is the general expression for $B(y;t)$? The following will work:

$$\begin{aligned}
B(y;t) = \lambda^G(t)\left(f(y) - f(X(T)) + \frac{\mu}{2}\left\|X(t)-X(T)\right\|^2\right) \\
- \left\langle \int_0^t \dot{\lambda}^G(\tau)\nabla f(X(\tau))\,d\tau + \mu\int_0^t \lambda^G(\tau)\dot{X}(\tau)\,d\tau, y - X(T)\right\rangle
\end{aligned}$$

In particular, this recovers (6) by substituting $\lambda^G(t) = e^{\sqrt{\mu}t}$ and $\dot{X}(t) = -\int_0^t e^{2\sqrt{\mu}(\tau-t)}\nabla f(X(\tau))\,d\tau$.

By applying integration by parts as $\int_0^t \lambda^G(\tau)\dot{X}(\tau)\,d\tau = \lambda^G(t)(X(t)-x_0) - \int_0^t \dot{\lambda}^G(\tau)(X(\tau)-x_0)\,d\tau$ and rearranging the terms, one will obtain

$$B(y,t) = \lambda^G(t)\left(\hat{f}(y) - \hat{f}(X(T)) + \frac{\mu}{2}\left\|y-X(t)\right\|^2\right) - \left\langle \int_0^t \dot{\lambda}^G(\tau)\nabla\hat{f}(X(\tau))\,d\tau, y - X(T)\right\rangle.$$

The term $\lambda^G(t)\frac{\mu}{2}\|y-X(t)\|^2$ can be omitted without affecting the validity of the argument above, as both $\frac{\partial}{\partial t}$ and ∇_y of this term vanish at $y = X(t)$. Thus, one can set

$$\begin{aligned}
B(y,t) &= \lambda^G(t)\left(\hat{f}(y) - \hat{f}(X(T))\right) - \left\langle \int_0^t \dot{\lambda}^G(\tau)\nabla\hat{f}(X(\tau))\,d\tau, y - X(T)\right\rangle \\
&= \left(\hat{f}_t(y) - \hat{f}_t(X(T))\right).
\end{aligned}$$

This will give the proof of [1, Thm. 2].

References

[1] Jungbin Kim and Insoon Yang. Convergence analysis of ODE models for accelerated first-order methods via positive semidefinite kernels. *Advances in Neural Information Processing Systems*, 36, 2024.