Appendix: Proofs and Supplementaries

A Proof for Theorem 1

For convenience of the reader, we report here our generalized model for momentum methods (GM-ODE), motivated in the main paper.

$$\begin{cases} \dot{X} = -m\nabla f(X) - nV \\ \dot{V} = \nabla f(X) - qV. \end{cases} \tag{GM-ODE}$$

Theorem 1 (Continuous-time stability)**.** *Let f be μ-strongly-convex and L-smooth. If $n, m, q \geq 0$ then, for any value of the strong-convexity modulus $\mu \geq 0$, the point $(x^*, 0) \in \mathbb{R}^{2d}$ is globally asymptotically stable for GM-ODE, as*

$$\mathcal{E}(X(t), V(t)) \leq e^{-\gamma_1 t} \cdot \mathcal{E}(X(0), V(0)), \tag{2}$$

where $\gamma_1 := \min\left(\frac{\mu(n+qm)}{2q}, \frac{q}{2}\right)$.

Proof. We propose the Lyapunov function

$$\mathcal{E}(t) = \underbrace{(qm+n)}_{c_1}\big(f(X(t)) - f(x^*)\big) + \underbrace{\frac{n(qm+n)}{4}}_{c_2}\|V(t)\|^2 + \underbrace{\frac{1}{4}}_{c_3}\|q(X(t)-x^*) - nV(t)\|^2, \tag{7}$$

consisting of quadratic and mixing parts

$$\mathcal{E}_1(t) = f(X(t)) - f(x^*), \quad \mathcal{E}_2(t) = \|V(t)\|^2, \quad \mathcal{E}_3(t) = \|-nV(t) + q(X(t)-x^*)\|^2. \tag{8}$$

The derivatives of each quadratic part are

$$\frac{d}{dt}\mathcal{E}_1(t) = -m\|\nabla f(X(t)\|^2 - n\langle \nabla f(X(t)), V(t)\rangle \tag{9}$$

and

$$\frac{d}{dt}\mathcal{E}_2(t) = -2q\|V(t)\|^2 + 2\langle \nabla f(X(t)), V(t)\rangle, \tag{10}$$

along with that of the mixing term:

$$\begin{aligned} \frac{d}{dt}\mathcal{E}_3(t) =& 2\langle -n\dot{V}(t) + q\dot{X}(t), -nV(t) + q(X(t)-x^*)\rangle \\ =& -2(qm+n)\langle \nabla f(X(t)), -nV(t) + q(X(t)-x^*)\rangle \\ =& -2q(qm+n)\langle \nabla f(X(t)), X(t)-x^*\rangle + 2n(qm+n)\langle \nabla f(X(t)), V(t)\rangle \\ \leq& -2q(qm+n)\Big(f(X(t)) - f(x^*)\Big) - \mu q(qm+n)\|X(t)-x^*\|^2 \\ &+ 2n(qm+n)\langle \nabla f(X(t)), V(t)\rangle, \end{aligned} \tag{11}$$

where last inequality is due to the strong convexity. Plugging the value of c_1, c_2 and c_3, we have

$$\frac{d}{dt}\mathcal{E}(t) \leq -\frac{q(n+qm)}{2}\Big(\big(f(X(t)) - f(x^*)\big) + \frac{\mu}{2}\|X(t)-x^*\|^2 + n\|V(t)\|^2\Big). \tag{12}$$

Besides, the mixing term can be upper-bounded by

$$\mathcal{E}_3(t) \leq 2q^2\|X(t)-x^*\|^2 + 2n^2\|V(t)\|^2. \tag{13}$$

Therefore we have $\mathcal{E}(t)$ satisfying

$$\mathcal{E}(t) \leq (qm+n)\big(f(X(t)) - f(x^*)\big) + q^2\|X(t)-x^*\|^2/2 + \Big(n^2/2 + \frac{n(n+qm)}{4}\Big)\|V(t)\|^2, \tag{14}$$

which implies

$$\frac{d}{dt}\mathcal{E}(t) \leq -\min\left\{\frac{\mu(n+qm)}{2q}, \frac{q}{2}\right\} \cdot \mathcal{E}(t). \tag{15}$$

We then conclude using Gronwall's lemma (Khalil and Grizzle, 2002). □

B Proof for Theorem 3

For convenience of the reader, we repeat here the semi-implicit integrator of GM-ODE we seek to study:

$$\text{(SIE)}: \quad \begin{cases} x_{k+1} - x_k = -m\sqrt{s}\nabla f(x_k) - n\sqrt{s}v_k \\ v_{k+1} - v_k = \sqrt{s}\nabla f(x_{k+1}) - q\sqrt{s}v_k. \end{cases}$$

In compact notation, the second iteration can be written as

$$r_1(v_{k+1} - v_k) = \sqrt{s}\nabla f(x_{k+1}) - q\sqrt{s}v_{k+1} \tag{16}$$

or

$$r_1 v_k = v_{k+1} - \sqrt{s}\nabla f(x_{k+1}), \tag{17}$$

where $r_1 = 1 - q\sqrt{s}$.

Theorem 3 (Convergence of SIE)**.** *Assume f L-smooth and μ-strongly-convex. Let $(x_k)_{k=1}^{\infty}$ be the sequence obtained from semi-implicit discretization of GM-ODE with step $\sqrt{s}$. Let*

$$0 < m\sqrt{s} \leq \frac{1}{2L},\ 0 < ns \leq m\sqrt{s},\ 0 < q\sqrt{s} \leq \frac{1}{2}. \tag{3}$$

There exists a constant $C > 0$ such that, for any $k \in \mathbb{N}$, it holds that

$$f(x_k) - f(x^*) \leq \big(1 + \gamma_2\sqrt{s}\big)^{-k} C,$$

where $\gamma_2 := \frac{1}{5}\min\left(\frac{n\mu}{q}, \frac{q}{1+q^2/(nL)}\right)$.

Proof. We propose the discrete Lyapunov function defined as

$$\mathcal{E}(k) = r_1 r_2(f(x_k) - f(x^*)) + \frac{1}{4}\|q(x_{k+1} - x^*) - nr_1 v_k\|^2 + \frac{nr_1^2 r_2}{4}\|v_k\|^2 - \frac{r_1 r_2 m\sqrt{s}}{2}\|\nabla f(x_k)\|^2. \tag{18}$$

We use colors for different parts to keep track of related terms in the derivation. As the first step, thanks to L-Lipshitz smoothness, we have

$$\begin{aligned} f(x_{k+1}) - f(x_k) \leq& \langle \nabla f(x_{k+1}), x_{k+1} - x_k\rangle - \frac{1}{2L}\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 \\ =& -m\sqrt{s}\langle \nabla f(x_k), \nabla f(x_{k+1})\rangle - n\sqrt{s}\langle v_k, \nabla f(x_{k+1})\rangle \\ & - \frac{1}{2L}\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2. \end{aligned} \tag{19}$$

We proceed by computing the difference in $\mathcal{E}$ in two subsequent iterations. Denote $r_2 = n + mq$, we have

$$\mathcal{E}(k+1) - \mathcal{E}(k) \overset{(A)}{\leq} -r_1 r_2 m\sqrt{s}\langle \nabla f(x_k), \nabla f(x_{k+1})\rangle - r_1 r_2 n\sqrt{s}\langle v_k, \nabla f(x_{k+1})\rangle$$

$$
\begin{aligned}
&- \frac{r_1 r_2}{2L}\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 + \frac{1}{4}\|q(x_{k+2} - x_{k+1}) - nr_1(v_{k+1} - v_k)\|^2 \\
&+ \frac{1}{2}\langle q(x_{k+2} - x_{k+1}) - nr_1(v_{k+1} - v_k), q(x_{k+1} - x^*) - nv_{k+1} + n\sqrt{s}\nabla f(x_{k+1})\rangle \\
&+ \frac{nr_1^2 r_2}{4}\|v_{k+1}\|^2 - \frac{nr_2}{4}\|v_{k+1} - \sqrt{s}\nabla f(x_{k+1})\|^2 \\
&- \frac{r_1 r_2 m\sqrt{s}}{2}\Big(\|\nabla f(x_{k+1})\|^2 - \|\nabla f(x_k)\|^2\Big) \\
\overset{(B)}{=}& -r_1 r_2 m\sqrt{s}\langle \nabla f(x_k), \nabla f(x_{k+1})\rangle - r_1 r_2 n\sqrt{s}\langle v_k, \nabla f(x_{k+1})\rangle \\
&- \frac{r_1 r_2}{2L}\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 - \frac{r_2(2n - r_2)}{4}s\|\nabla f(x_{k+1})\|^2 \\
&- \frac{r_2}{2}\sqrt{s}\langle \nabla f(x_{k+1}), q(x_{k+1} - x^*) - nv_{k+1}\rangle \\
&- \frac{nr_2(1 - r_1^2)}{4}\|v_{k+1}\|^2 - \frac{nr_2}{4}s\|\nabla f(x_{k+1})\|^2 + \frac{nr_2}{2}\sqrt{s}\langle \nabla f(x_{k+1}), v_{k+1}\rangle \\
&- \frac{r_1 r_2 m\sqrt{s}}{2}\Big(\|\nabla f(x_{k+1})\|^2 - \|\nabla f(x_k)\|^2\Big) \\
\overset{(C)}{=}& nr_2\sqrt{s}\langle \nabla f(x_{k+1}), v_{k+1}/2 + v_{k+1}/2 - r_1 v_k\rangle \\
&+ \frac{r_1 r_2}{2}m\sqrt{s}\Big(\|\nabla f(x_{k+1})\|^2 - 2\langle \nabla f(x_{k+1}), \nabla f(x_k)\rangle + \|\nabla f(x_k)\|^2\Big) \\
&- \Big(\frac{r_2(2n - r_2)}{4}s + \frac{nr_2}{4}s + r_1 r_2 m\sqrt{s}\Big)\|\nabla f(x_{k+1})\|^2 - \frac{nr_2(1 - r_1^2)}{4}\|v_{k+1}\|^2 \\
&- \frac{r_1 r_2}{2L}\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 - \frac{r_2}{2}q\sqrt{s}\langle \nabla f(x_{k+1}), x_{k+1} - x^*\rangle. \qquad (20)
\end{aligned}
$$

In step (A), we use smoothness of f as stated in Eq. 19 for the blue term. Also, we used the inequality $\|a\|^2 - \|b\|^2 = \|a - b\|^2 + 2\langle a - b, b\rangle$ where $a = q(x_{k+2} - x^*) - nr_1 v_k$ and $b = q(x_{k+1} - x^*) - nr_1 v_k$ to obtain the red term. In particular,

$$
\begin{aligned}
a - b &= q(x_{k+2} - x_{k+1}) - nr_1(v_{k+1} - v_k) \\
&= -mq\sqrt{s}\nabla f(x_{k+1}) - nq\sqrt{s}v_{k+1} - n\sqrt{s}\nabla f(x_{k+1}) + nq\sqrt{s}v_{k+1} \\
&= -r_2\sqrt{s}\nabla f(x_{k+1}). \qquad (21)
\end{aligned}
$$

In step (B), we incorporate the recurrence of SIE. Step (C) is a simple re-arrangement of terms.

We can easily verify the following identities:

$$\sqrt{s}\langle \nabla f(x_{k+1}), v_{k+1} - r_1 v_k\rangle = s\|\nabla f(x_{k+1})\|^2 \qquad (22)$$

and

$$\|\nabla f(x_{k+1})\|^2 - 2\langle \nabla f(x_{k+1}), \nabla f(x_k)\rangle + \|\nabla f(x_k)\|^2 = \|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2. \qquad (23)$$

We have

$$
\begin{aligned}
\mathcal{E}(k+1) - \mathcal{E}(k) \le& nr_2 s\|\nabla f(x_{k+1})\|^2 + \frac{r_1 r_2}{2}m\sqrt{s}\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 \\
&- r_2 s\Big(\frac{2n - r_2}{4} + \frac{n}{4} + \frac{r_1 m}{\sqrt{s}}\Big)\|\nabla f(x_{k+1})\|^2 - \frac{nr_2(1 - r_1^2)}{4}\|v_{k+1}\|^2 \\
&- \frac{r_1 r_2}{2L}\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 - \frac{r_2}{2}q\sqrt{s}\langle \nabla f(x_{k+1}), x_{k+1} - x^*\rangle. \qquad (24)
\end{aligned}
$$

We leverage μ-strong convexity of f to get

$$\langle \nabla f(x_{k+1}), x_{k+1} - x^*\rangle \ge f(x_{k+1}) - f(x^*) + \frac{\mu}{2}\|x_{k+1} - x^*\|^2. \qquad (25)$$

Applying the above inequality to the last term of Eq. 24, we obtain

$$\mathcal{E}(k+1) - \mathcal{E}(k) \le -\frac{r_2}{2}q\sqrt{s}(f(x_{k+1}) - f(x^*)) - \frac{r_2\mu}{4}q\sqrt{s}\|x_{k+1} - x^*\|^2$$

$$
\begin{aligned}
&- \frac{r_1 r_2}{2}(1/L - m\sqrt{s})\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 - \frac{n r_2(1-r_1^2)}{4}\|v_{k+1}\|^2 \\
&- r_2 s\Big(\frac{2n - r_2}{4} + \frac{n}{4} + \frac{r_1 m}{\sqrt{s}} - n\Big)\|\nabla f(x_{k+1})\|^2. \qquad (26)
\end{aligned}
$$

Now we plug in the the value of r_1, r_2 and calculate

$$1 - r_1^2 = 1 - (1 - q\sqrt{s})^2 = q\sqrt{s}(2 - q\sqrt{s}) \geq q\sqrt{s}, \qquad (27)$$

where we used the condition $q\sqrt{s} \leq 1/2$. Next, since $m\sqrt{s} \leq 1/(2L)$, $n \leq m/\sqrt{s}$ and $r_1 = 1 - q\sqrt{s} \geq 1/2$, it holds that

$$\frac{2n - r_2}{4} + \frac{r_1 m}{\sqrt{s}} - \frac{3n}{4} = \frac{n - mq}{4} + \frac{r_1 m}{\sqrt{s}} - \frac{3n}{4} = \frac{r_1 m}{\sqrt{s}} - \frac{n}{2} - \frac{mq}{4} \geq -\frac{mq}{4}. \qquad (28)$$

Hence, the difference between two iterations can be upper-bounded as follows:

$$
\begin{aligned}
\mathcal{E}(k+1) - \mathcal{E}(k) \leq & - \frac{r_2 q\sqrt{s}}{2}\Big(f(x_{k+1}) - f(x^*) + \frac{\mu}{2}\|x_{k+1} - x^*\|^2 + n\|v_{k+1}\|^2/2 - \frac{m\sqrt{s}}{2}\|\nabla f(x_{k+1})\|^2\Big) \\
= & - \frac{r_2 q\sqrt{s}}{2}\Big((1 - r_3)[f(x_{k+1}) - f(x^*)] + \frac{\mu}{2}\|x_{k+1} - x^*\|^2 \\
& + n\|v_{k+1}\|^2/2 + r_3[f(x_{k+1}) - f(x^*) - \frac{1}{2L}\|\nabla f(x_{k+1})\|^2]\Big), \qquad (29)
\end{aligned}
$$

where $r_3 = Lm\sqrt{s} \leq 1/2$ and the bound remains legal since $1 - r_3 \geq 1/2$.

On the other hand, our candidate Lyapunov function at iteration k itself can be upper-bounded as

$$
\begin{aligned}
\mathcal{E}(k) &= r_1 r_2(f(x_k) - f(x^*) + \frac{1}{4}\|q(x_{k+1} - x^*) - n r_1 v_k\|^2 + \frac{n r_1^2 r_2}{4}\|v_k\|^2 - \frac{r_1 r_2 m\sqrt{s}}{2}\|\nabla f(x_k)\|^2 \\
&\overset{(A)}{=} r_1 r_2(f(x_k) - f(x^*)) + \frac{1}{4}\|q(x_k - x^*) - n v_k - mq\sqrt{s}\nabla f(x_k)\|^2 + \frac{n r_1^2 r_2}{4}\|v_k\|^2 \\
&\quad - r_1 r_2 m\sqrt{s}\|\nabla f(x_k)\|^2/2 \\
&\overset{(B)}{\leq} r_1 r_2(f(x_k) - f(x^*)) + q^2\|x_k - x^*\|^2 + n^2\|v_k\|^2 + \frac{q^2 m^2 s}{2}\|\nabla f(x_k)\|^2 + \frac{n r_1^2 r_2}{4}\|v_k\|^2 \\
&\quad - r_1 r_2 m\sqrt{s}\|\nabla f(x_k)\|^2/2 \\
&= r_1 r_2(1 - r_3 + r_4)(f(x_k) - f(x^*)) + q^2\|x_k - x^*\|^2 + (n^2 + n r_1^2 r_2/4)\|v_k\|^2 \\
&\quad + r_1 r_2(r_3 - r_4)[f(x_k) - f(x^*) - \frac{1}{2L}\|\nabla f(x_k)\|^2], \qquad (30)
\end{aligned}
$$

with $r_4 = Lq^2 m^2 s/(r_1 r_2)$. Precisely, step (A) is obtained by replacing SIE update for the term x_{k+1}. (B) is obtained by repeatedly using the inequality $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$. Finally, noting that $f(x_k) - f(x^*) \geq \frac{1}{2L}\|\nabla f(x_k)\|^2$, we have

$$
\begin{aligned}
\mathcal{E}(k) \leq r_2\Big(& r_1(1 - r_3 + r_4)[f(x_k) - f(x^*)] + \frac{q^2}{n}\|x_k - x^*\|^2 + 5n\|v_k\|^2/4 \\
& + r_1(r_3 - r_4)[f(x_k) - f(x^*) - \frac{1}{2L}\|\nabla f(x_k)\|^2]\Big), \qquad (31)
\end{aligned}
$$

since $r_2 = n + mq \geq n$. It is reckoned that $\mathcal{E}(k+1) - \mathcal{E}(k)$ and $\mathcal{E}(k)$ share identical parts except for different coefficients. Now we aim at obtaining following inequality

$$\mathcal{E}(k+1) - \mathcal{E}(k) \leq -\gamma_2\sqrt{s}\mathcal{E}(k+1). \qquad (32)$$

To achieve this, γ_2 should be the minimal ratio for coefficients of each parts of $\mathcal{E}(k+1) - \mathcal{E}(k)$ to those of $\mathcal{E}(k)$. It is easy then to notice that γ_2 should be smaller than $q/5$ and $n\mu/(4q)$. Besides it should also hold that

$$\frac{r_2 q}{2 r_1 r_2}\frac{r_3}{r_3 - r_4} \geq \frac{q}{2}\frac{1 - r_3}{1 - (r_3 - r_4)} = \frac{q}{2}\frac{1 - r_3}{1 - r_3(1 - r_4/r_3)} \geq \frac{q}{2}\frac{1 - 1/2}{1 - 1/2(1 - \frac{q^2}{nL})} \geq \frac{q}{2}\frac{1}{1 + \frac{q^2}{nL}} \geq \gamma_2, \qquad (33)$$

due to the fact $\frac{r_4}{r_3} = \frac{q^2 m\sqrt{s}}{r_1 r_2} \leq \frac{q^2}{nL}$ and $r_3 \leq 1/2$. Therefore $\gamma_2 = \frac{1}{5}\min\{\frac{q}{1+\frac{q^2}{nL}}, \frac{n\mu}{q}\}$ satisfies the above inequality and completes the proof. □

We now use the above result to prove the convergence of QHM iterations (see Sec. 4).

Corollary 5 (Convergence of QHM)**.** *Let f be L-smooth and μ-strongly-convex with $L/\mu \geq 9$. The iterates of enjoy a linear convergence rate for $s \leq \frac{1}{4L}$ and $a \leq 1/2$. In particular, also enjoys convergence rate $O((1-\sqrt{\mu/L})^k)$ for $b = 1-2\sqrt{\mu s}$. Namely, $\exists C > 0$ such that*

$$f(x_k) - f(x^*) \leq \left(1 + a\sqrt{\mu s}/10\right)^{-k} C.$$

Proof. First, we show how one can alternatively write QHM as one-line scheme. The original QHM algorithm is reported here for convenience of the reader

$$\begin{cases} x_{k+1} = x_k - s((1-a)\nabla f(x_k) + a g_{k+1}) \\ g_{k+1} = b g_k + \nabla f(x_k). \end{cases} \tag{QHM}$$

We replace the second line of QHM into the first one :

$$x_{k+1} = x_k - s(1-a)\nabla f(x_k) - s \cdot b \cdot a \cdot g_k - as\nabla f(x_k). \tag{34}$$

Using the first iterate we get:

$$-(x_k - x_{k-1})/s - (1-\alpha)\nabla f(x_{k-1}) = a g_k. \tag{35}$$

Replacing this into the result of first equation, we get:

$$x_{k+1} = x_k - s(1-a)\nabla f(x_k) + b((x_k - x_{k-1}) + s(1-a)\nabla f(x_{k-1})) - as\nabla f(x_k). \tag{36}$$

By rearrangment, we finally obtain

$$x_{k+1} = x_k + b(x_k - x_{k-1}) - s\nabla f(x_k) + sb(1-a)\nabla f(x_{k-1}). \tag{37}$$

The above iterates can be viewed as SIE discretization of GM-ODE with the following specific choice of parameters (see the single sequence of iterates of SIE in the last section):

$$m = (1-a)\sqrt{s}, \quad n = a, \quad q = \frac{1-b}{\sqrt{s}}. \tag{38}$$

Invoking Thm. 3, we get the convergence rate for QHM. More precisely, choosing $b = 1 - 2\sqrt{\mu s}$ we obtain

$$q = 2\sqrt{\mu}. \tag{39}$$

The above choice of parameters obeys the constraints in Thm. 3:

$$m\sqrt{s} = (1-a)s < s \leq \frac{1}{4L}, \quad n = a \leq (1-a) = \frac{m}{\sqrt{s}}, \tag{40}$$

and

$$q\sqrt{s} = 2\sqrt{\mu s} \leq 2\sqrt{sL/9} \leq 1/3, \tag{41}$$

since we assumed $s \leq 1/(4L)$, $a \leq 1/2$ and $L/\mu \geq 9$. The rate — thanks to Thm. 3 — is determined by $\gamma_2 = \frac{1}{5}\min\{\frac{n\mu}{q}, \frac{q}{1+q^2/(nL)})\}$. We conclude the proof by showing that $\gamma_2 = a\sqrt{\mu}/8$ in the case of QHM. First, one can readily check that $(n\mu)/(5q) = a\mu/(10\sqrt{\mu})$ holds due to the choice of parameters. Second, with some patience, one can check that the following chain of inequality holds:

$$\frac{1}{5} \cdot \frac{q}{1+\frac{q^2}{nL}} = \frac{2\sqrt{\mu}}{5 + \frac{20\mu}{aL}} \geq \frac{2a\sqrt{\mu}}{5a + 20/9} \geq \frac{a\sqrt{\mu}}{10}.$$

□

C Proofs for Section 5

As stated in the main paper, we consider the following discretization errors:

$$\Delta_k^{(\text{EE})} := \|X(k\sqrt{s}) - x_k\|, \quad x_k \text{ obtained by EE}$$
$$\Delta_k^{(\text{SIE})} := \|X(k\sqrt{s}) - x_{k+1}\|, \quad x_k \text{ obtained by SIE.}$$

We define $w_k := x_k$ for EE and $w_k := x_{k+1}$ for SIE. We compare the error $\Delta_k = \|X(k\sqrt{s}) - w_k\|$ for $k = 1$ in the next lemma, assuming $\Delta_0 = 0$ and $v_0 = V(0)$. This is also called local (or one-step) integration error.

Lemma 7. *Let f be L-smooth and of class C^2. If $m = O(\sqrt{s})$, then $\Delta_1^{(SIE)} = O(s^{3/2})$ and $\Delta_1^{(EE)} = O(s)$.*

Proof. We introduce the notation $X_k := X(k\sqrt{s})$, $V_k := V(k\sqrt{s})$. Our problem setting requires $w_0 = X_0$ and $v_0 = V_0$. For SIE, $w_k = x_{k+1}$ and we begin from Taylor expansion of X as

$$X_1 - X_0 = \sqrt{s}\dot{X}_0 + s\ddot{X}_0 + O(s^{3/2}), \tag{42}$$

and therefore

$$\begin{aligned}
X_1 - w_1 =& X_1 - X_0 - (w_1 - w_0) + X_0 - w_0 \\
=& X_1 - X_0 - (x_2 - x_1) + X_0 - x_1 \\
=& \sqrt{s}\dot{X}_0 + s\ddot{X}_0 + m\sqrt{s}\nabla f(x_1) + n\sqrt{s}v_1 + O(s^{3/2}) \\
=& \sqrt{s}\Big(- m\nabla f(X_0) - nV_0\Big) + s\frac{d}{dt}\Big(- m\nabla f(X_0) - nV_0\Big) \\
&+ m\sqrt{s}\nabla f(x_1) + n\sqrt{s}\Big(\sqrt{s}\nabla f(x_1) + (1 - q\sqrt{s})v_0\Big) + O(s^{3/2}).
\end{aligned} \tag{43}$$

where in the third equality we used the fact that, by hypothesis, $X_0 - x_1 = 0$. And in particular, since $\frac{d\nabla f(X)}{dt} = \nabla^2 f(X)\dot{X}$,

$$\begin{aligned}
s\frac{d}{dt}\Big(- m\nabla f(X_0) - nV_0\Big) =& - sm\nabla^2 f(X_0)\dot{X}_0 - sn\dot{V}_0 \\
=& - sn\nabla f(X_0) + snqV_0 + sm^2\nabla^2 f(X_0)\nabla f(X_0) + smn\nabla^2 f(X_0)V_0.
\end{aligned} \tag{44}$$

Then it holds that

$$X_1 - w_1 = - (m\sqrt{s} + ns)\Big(\nabla f(X_0) - \nabla f(x_1)\Big) - n\sqrt{s}(V_0 - v_0) + snq(V_0 - v_0) + O(s^{3/2}) \le O(s^{3/2}) \tag{45}$$

and $\Delta_1^{(\text{SIE})} \le O(s^{3/2})$.

We proceed with the EE iterations (remember: $w_k = x_k$). We expand $\Delta_1^{(\text{EE})}$ as

$$\begin{aligned}
X_1 - w_1 =& X_1 - X_0 - (w_1 - w_0) + X_0 - x_0 + O(s^{3/2}) \\
=& X_1 - X_0 - (x_1 - x_0) + X_0 - x_0 + O(s^{3/2}) \\
=& \sqrt{s}\dot{X}_0 + s\ddot{X}_0 + m\sqrt{s}\nabla f(x_0) + n\sqrt{s}v_0 + O(s^{3/2}) \\
=& \sqrt{s}\Big(- m\nabla f(X_0) - nV_k\Big) + m\sqrt{s}\nabla f(x_0) + n\sqrt{s}v_0 \\
&+ sm\Big(m\nabla^2 f(X_0)\nabla f(X_0) + n\nabla^2 f(X_k)V_0\Big) \\
&- sn\Big(\nabla f(X_0) - qV_0\Big) + O(s^{3/2}) \\
=& - m\sqrt{s}\Big(\nabla f(X_0) - \nabla f(x_0)\Big) - n\sqrt{s}\Big(V_0 - v_0\Big) + O(s).
\end{aligned} \tag{46}$$

Therefore, we conclude that $\Delta_1^{(\text{EE})} \le O(s)$. □

Lemma 8. *Let f be μ-strongly-convex and L-smooth. For EE discretization of GM-ODE obeying Eq. 4, the discretization error decays as $\Delta_k^{(EE)} = O((1+\gamma_3\sqrt{s})^{-k})$ where γ_3 is defined in Thm. 6. Furthermore, SIE also enjoys $\Delta_k^{(SIE)} = O((1+\gamma_2\sqrt{s})^{-k})$ where γ_2 is defined in Thm. 3 as long as conditions in Eq. 3 are satisfied.*

Proof. The proof is based on the following consequence of strong convexity

$$\mu\|x - x^*\|^2/2 \leq f(x) - f(x^*). \tag{47}$$

Using the above inequality together with a straightforward application of triangular inequality we complete the proof:

$$\begin{aligned} \|X(k\sqrt{s}) - x_k\| &= \|X(k\sqrt{s}) - x^* + x^* - x_k\| \\ &\leq \|X(k\sqrt{s}) - x^*\| + \|x_k - x^*\| \\ &\leq \sqrt{2}\mu^{-1/2}\left(\left(f(X(k\sqrt{s})) - f(x^*)\right)^{1/2} + \left(f(x_k) - f(x^*)\right)^{1/2}\right). \end{aligned} \tag{48}$$

Replacing the convergence results in Thm. 1, 3, and 6 into the the above bound concludes the proof. □