Supplementary Material for
" A Class of Two-Timescale Stochastic EM Algorithms for Nonconvex Latent Variable Models "

A Proofs for the iSAEM Algorithm

A.1 Proof of Lemma 2

Lemma. *Assume A3, A4. For all* $\mathrm{s} \in \mathsf{S}$,

$$\upsilon_{\min}^{-1}\left\langle \nabla V(\mathrm{s})\,,\,\mathrm{s}-\overline{s}(\overline{\boldsymbol{\theta}}(\mathrm{s}))\right\rangle \geq \|\mathrm{s}-\overline{s}(\overline{\boldsymbol{\theta}}(\mathrm{s}))\|^2 \geq \upsilon_{\max}^{-2}\|\nabla V(\mathrm{s})\|^2. \tag{32}$$

Proof. Using A3 and the fact that we can exchange integration with differentiation and the Fisher's identity, we obtain

$$\begin{aligned}
\nabla_{\mathrm{s}} V(\mathrm{s}) &= \mathrm{J}_{\boldsymbol{\theta}}^{\mathrm{s}}(\mathrm{s})^{\top}\left(\nabla_{\boldsymbol{\theta}}\,\mathrm{r}(\overline{\boldsymbol{\theta}}(\mathrm{s}))+\nabla_{\boldsymbol{\theta}}\mathsf{L}(\overline{\boldsymbol{\theta}}(\mathrm{s}))\right)\\
&= \mathrm{J}_{\boldsymbol{\theta}}^{\mathrm{s}}(\mathrm{s})^{\top}\left(\nabla_{\boldsymbol{\theta}}\psi(\overline{\boldsymbol{\theta}}(\mathrm{s}))+\nabla_{\boldsymbol{\theta}}\,\mathrm{r}(\overline{\boldsymbol{\theta}}(\mathrm{s}))-\mathrm{J}_{\phi}^{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}(\mathrm{s}))^{\top}\overline{s}(\overline{\boldsymbol{\theta}}(\mathrm{s}))\right)\\
&= \mathrm{J}_{\boldsymbol{\theta}}^{\mathrm{s}}(\mathrm{s})^{\top}\,\mathrm{J}_{\phi}^{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}(\mathrm{s}))^{\top}\left(\mathrm{s}-\overline{s}(\overline{\boldsymbol{\theta}}(\mathrm{s}))\right)\,.
\end{aligned} \tag{33}$$

Consider the following vector map:

$$\mathrm{s}\rightarrow\nabla_{\boldsymbol{\theta}}L(\mathrm{s};\boldsymbol{\theta})|_{\boldsymbol{\theta}=\overline{\boldsymbol{\theta}}(\mathrm{s})}=\nabla_{\boldsymbol{\theta}}\psi(\overline{\boldsymbol{\theta}}(\mathrm{s}))+\nabla_{\boldsymbol{\theta}}\,\mathrm{r}(\overline{\boldsymbol{\theta}}(\mathrm{s}))-\mathrm{J}_{\phi}^{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}(\mathrm{s}))^{\top}\mathrm{s}\,.$$

Taking the gradient of the above map w.r.t. s and using assumption A3, we show that:

$$\mathbf{0}=-\,\mathrm{J}_{\phi}^{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}(\mathrm{s}))+\big(\underbrace{\nabla_{\boldsymbol{\theta}}^{2}(\psi(\boldsymbol{\theta})+\mathrm{r}(\boldsymbol{\theta})-\langle\phi(\boldsymbol{\theta})\,,\,\mathrm{s}\rangle)}_{=\mathrm{H}_{L}^{\boldsymbol{\theta}}(\mathrm{s};\boldsymbol{\theta})}|_{\boldsymbol{\theta}=\overline{\boldsymbol{\theta}}(\mathrm{s})}\big)\,\mathrm{J}_{\boldsymbol{\theta}}^{\mathrm{s}}(\mathrm{s})\,.$$

The above yields

$$\nabla_{\mathrm{s}}V(\mathrm{s})=\mathrm{B}(\mathrm{s})(\mathrm{s}-\overline{s}(\overline{\boldsymbol{\theta}}(\mathrm{s})))\,,$$

where we recall $\mathrm{B}(\mathrm{s})=\mathrm{J}_{\phi}^{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}(\mathrm{s}))(\mathrm{H}_{L}^{\boldsymbol{\theta}}(\mathrm{s};\overline{\boldsymbol{\theta}}(\mathrm{s})))^{-1}\,\mathrm{J}_{\phi}^{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}(\mathrm{s}))^{\top}$. The proof of (32) follows directly from the assumption A4. □

A.2 Proof of Theorem 1

Beforehand, We present two intermediary Lemmas important for the analysis of the incremental update of the iSAEM algorithm. The first one gives a characterization of the quantity $\mathbb{E}[S_{\text{tts}}^{(k+1)} - \hat{\mathrm{s}}^{(k)}]$:

Lemma. *Assume A1. The update (1) is equivalent to the following update on the resulting statistics*

$$\hat{\mathrm{s}}^{(k+1)} = \hat{\mathrm{s}}^{(k)} + \gamma_{k+1}(S_{tts}^{(k+1)} - \hat{\mathrm{s}}^{(k)}) \ .$$

Also:

$$\mathbb{E}[S_{tts}^{(k+1)} - \hat{\mathrm{s}}^{(k)}] = \mathbb{E}[\overline{\mathrm{s}}^{(k)} - \hat{\mathrm{s}}^{(k)}] + (1-1/n)\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} \tilde{S}_i^{(\tau_i^k)} - \overline{\mathrm{s}}^{(k)}] + \frac{1}{n}\mathbb{E}[\eta_{i_k}^{(k+1)}] \ ,$$

where $\overline{\mathrm{s}}^{(k)}$ is defined by (4) and $\tau_i^k = \max\{k' : i_{k'} = i,\ k' < k\}$.

Proof. From update (1), we have:

$$\begin{aligned}
S_{\text{tts}}^{(k+1)} - \hat{\mathrm{s}}^{(k)} &= S_{\text{tts}}^{(k)} - \hat{\mathrm{s}}^{(k)} + \frac{1}{n}(\tilde{S}_{i_k}^{(k+1)} - \tilde{S}_{i_k}^{(\tau_i^k)}) \\
&= \overline{\mathrm{s}}^{(k)} - \hat{\mathrm{s}}^{(k)} + S_{\text{tts}}^{(k)} - \overline{\mathrm{s}}^{(k)} - \frac{1}{n}(\tilde{S}_{i_k}^{(\tau_i^k)} - \tilde{S}_{i_k}^{(k+1)}) \ .
\end{aligned}$$

Since $\tilde{S}_{i_k}^{(k+1)} = \overline{\mathrm{s}}_{i_k}(\boldsymbol{\theta}^{(k)}) + \eta_{i_k}^{(k+1)}$ we have

$$S_{\text{tts}}^{(k+1)} - \hat{\mathrm{s}}^{(k)} = \overline{\mathrm{s}}^{(k)} - \hat{\mathrm{s}}^{(k)} + S_{\text{tts}}^{(k)} - \overline{\mathrm{s}}^{(k)} - \frac{1}{n}(\tilde{S}_{i_k}^{(\tau_i^k)} - \overline{\mathrm{s}}_{i_k}(\boldsymbol{\theta}^{(k)})) + \frac{1}{n}\eta_{i_k}^{(k+1)} \ .$$

Taking the full expectation of both side of the equation leads to:

$$\begin{aligned}
\mathbb{E}[S_{\text{tts}}^{(k+1)} - \hat{\mathrm{s}}^{(k)}] = \mathbb{E}[\overline{\mathrm{s}}^{(k)} - \hat{\mathrm{s}}^{(k)}] + \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} \tilde{S}_i^{(\tau_i^k)} - \overline{\mathrm{s}}^{(k)}] \\
- \frac{1}{n}\mathbb{E}[\mathbb{E}[\tilde{S}_i^{(\tau_i^k)} - \overline{\mathrm{s}}_{i_k}(\boldsymbol{\theta}^{(k)})|\mathcal{F}_k]] + \frac{1}{n}\mathbb{E}[\eta_{i_k}^{(k+1)}] \ .
\end{aligned}$$

Since we have $\mathbb{E}[\tilde{S}_i^{(\tau_i^k)}|\mathcal{F}_k] = \frac{1}{n}\sum_{i=1}^{n} \tilde{S}_i^{(\tau_i^k)}$ and $\mathbb{E}[\overline{\mathrm{s}}_{i_k}(\boldsymbol{\theta}^{(k)})|\mathcal{F}_k] = \overline{\mathrm{s}}^{(k)}$, we conclude the proof of the Lemma. $\square$

We also derive the following auxiliary Lemma which sets an upper bound for the quantity $\mathbb{E}[\|S_{\text{tts}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2]$:

Lemma. *For any $k \geq 0$ and consider the iSAEM update in (1), it holds that*

$$\begin{aligned}
\mathbb{E}[\|S_{tts}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2] \leq & 4\mathbb{E}[\|\overline{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] + \frac{2\,\mathrm{L}_{\mathbf{s}}^2}{n^3}\sum_{i=1}^{n}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2] \\
& + 2\frac{c_\eta}{M_k} + 4\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^{n} \tilde{S}_i^{(\tau_i^k)} - \overline{\mathrm{s}}^{(k)}\|^2] \ .
\end{aligned}$$

Proof. Applying the iSAEM update yields:

$$\begin{aligned}\mathbb{E}[\|S_{\text{tts}}^{(k+1)}-\hat{\boldsymbol{s}}^{(k)}\|^2] =&\mathbb{E}[\|S_{\text{tts}}^{(k)}-\hat{\boldsymbol{s}}^{(k)}-\frac{1}{n}(\tilde{S}_{i_k}^{(\tau_i^k)}-\tilde{S}_{i_k}^{(k)})\|^2]\\ \leq& 4\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^{n}\tilde{S}_i^{(\tau_i^k)}-\overline{\mathbf{s}}^{(k)}\|^2]+4\mathbb{E}[\|\overline{\mathbf{s}}^{(k)}-\hat{\boldsymbol{s}}^{(k)}\|^2]+\frac{2}{n^2}\mathbb{E}[\|\overline{\boldsymbol{s}}_{i_k}^{(k)}-\overline{\boldsymbol{s}}_{i_k}^{(t_{i_k}^k)}\|^2]+2\frac{c_\eta}{M_k}\,.\end{aligned}$$

The last expectation can be further bounded by

$$\frac{2}{n^2}\mathbb{E}[\|\overline{\boldsymbol{s}}_{i_k}^{(k)}-\overline{\boldsymbol{s}}_{i_k}^{(t_{i_k}^k)}\|^2]=\frac{2}{n^3}\sum_{i=1}^{n}\mathbb{E}[\|\overline{\boldsymbol{s}}_{i}^{(k)}-\overline{\boldsymbol{s}}_{i}^{(t_i^k)}\|^2]\overset{(a)}{\leq}\frac{2\,\mathrm{L}_{\mathbf{s}}^2}{n^3}\sum_{i=1}^{n}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)}-\hat{\boldsymbol{s}}^{(t_i^k)}\|^2]\,,$$

where (a) is due to Lemma 1 and which concludes the proof of the Lemma.

□

Theorem. *Assume A1-A5. Consider the iSAEM sequence* $\{\hat{\mathrm{s}}^{(k)}\}_{k>0}\in\mathcal{S}$ *obtained with* $\rho_{k+1}=1$ *for any* $k\leq K_m$ *where* K_m *is a positive integer. Let* $\{\gamma_k=1/(k^a\alpha c_1\overline{L})\}_{k>0}$*, where* $a\in(0,1)$*, be a sequence of stepsizes,* $c_1=\upsilon_{\min}^{-1}$*,* $\alpha=\max\{8,1+6\upsilon_{\min}\}$*,* $\overline{L}=\max\{\mathrm{L}_{\mathbf{s}},\mathrm{L}_V\}$*,* $\beta=c_1\overline{L}/n$*. Then:*

$$\upsilon_{\max}^{-2}\sum_{k=0}^{K_m}\tilde{\alpha}_k\mathbb{E}[\|\nabla V(\hat{\boldsymbol{s}}^{(k)})\|^2]\leq\mathbb{E}[V(\hat{\boldsymbol{s}}^{(0)})-V(\hat{\boldsymbol{s}}^{(K_m)})]+\sum_{k=0}^{K_m-1}\tilde{\Gamma}_k\mathbb{E}[\|\eta_{i_k}^{(k)}\|^2]\,.$$

Proof. Under the smoothness of the Lyapunov function V (cf. Lemma 1), we can write:

$$V(\hat{\boldsymbol{s}}^{(k+1)})\leq V(\hat{\boldsymbol{s}}^{(k)})+\gamma_{k+1}\big\langle S_{\text{tts}}^{(k+1)}-\hat{\boldsymbol{s}}^{(k)}\,,\,\nabla V(\hat{\boldsymbol{s}}^{(k)})\big\rangle+\frac{\gamma_{k+1}^2\,\mathrm{L}_V}{2}\|S_{\text{tts}}^{(k+1)}-\hat{\boldsymbol{s}}^{(k)}\|^2\,.$$

Taking the expectation on both sides yields:

$$\mathbb{E}[V(\hat{\boldsymbol{s}}^{(k+1)})]\leq\mathbb{E}[V(\hat{\boldsymbol{s}}^{(k)})]+\gamma_{k+1}\mathbb{E}[\big\langle S_{\text{tts}}^{(k+1)}-\hat{\boldsymbol{s}}^{(k)}\,,\,\nabla V(\hat{\boldsymbol{s}}^{(k)})\big\rangle]+\frac{\gamma_{k+1}^2\,\mathrm{L}_V}{2}\mathbb{E}[\|S_{\text{tts}}^{(k+1)}-\hat{\boldsymbol{s}}^{(k)}\|^2]\,.$$

Using Lemma 4, we obtain:

$$
\begin{aligned}
&\mathbb{E}[\langle S_{\text{tts}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)})\rangle] \\
=&\mathbb{E}[\langle \overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)})\rangle] + \left(1-\frac{1}{n}\right)\mathbb{E}[\langle \frac{1}{n}\sum_{i=1}^n \tilde{S}_i^{(\tau_i^k)} - \overline{\mathbf{s}}^{(k)}\,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)})\rangle] \\
&+ \frac{1}{n}\mathbb{E}[\langle \eta_{i_k}^{(k)}\,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)})\rangle] \\
\overset{(a)}{\leq}& - \upsilon_{\min}\mathbb{E}[\|\overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] + \left(1-\frac{1}{n}\right)\mathbb{E}[\langle \frac{1}{n}\sum_{i=1}^n \tilde{S}_i^{(\tau_i^k)} - \overline{\mathbf{s}}^{(k)}\,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)})\rangle] \\
&+ \frac{1}{n}\mathbb{E}[\langle \eta_{i_k}^{(k)}\,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)})\rangle] \\
\overset{(b)}{\leq}& - \upsilon_{\min}\mathbb{E}[\|\overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] + \frac{1-\frac{1}{n}}{2\beta}\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^n \tilde{S}_i^{(\tau_i^k)} - \overline{\mathbf{s}}^{(k)}\|^2] \\
&+ \frac{\beta(n-1)+1}{2n}\mathbb{E}[\left\|\nabla V(\hat{\boldsymbol{s}}^{(k)})\right\|^2] + \frac{1}{2n}\mathbb{E}[\|\eta_{i_k}^{(k)}\|^2] \\
\overset{(a)}{\leq}& \left(\upsilon_{\max}^2\frac{\beta(n-1)+1}{2n} - \upsilon_{\min}\right)\mathbb{E}[\|\overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] + \frac{1-\frac{1}{n}}{2\beta}\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^n \tilde{S}_i^{(\tau_i^k)} - \overline{\mathbf{s}}^{(k)}\|^2] \\
&+ \frac{1}{2n}\mathbb{E}[\|\eta_{i_k}^{(k)}\|^2]\,,
\end{aligned}
$$

where (a) is due to the growth condition (2) and (b) is due to Young's inequality (with $\beta \to 1$). Note $a_k = \gamma_{k+1}\left(\upsilon_{\min} - \upsilon_{\max}^2\frac{\beta(n-1)+1}{2n}\right)$ and

$$
\begin{aligned}
a_k\mathbb{E}[\|\overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] \leq& \mathbb{E}[V(\hat{\boldsymbol{s}}^{(k)}) - V(\hat{\boldsymbol{s}}^{(k+1)})] + \frac{\gamma_{k+1}^2\,\mathrm{L}_V}{2}\mathbb{E}[\|S_{\text{tts}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2] \\
&+ \frac{\gamma_{k+1}(1-\frac{1}{n})}{2\beta}\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^n \tilde{S}_i^{(\tau_i^k)} - \overline{\mathbf{s}}^{(k)}\|^2] + \frac{\gamma_{k+1}}{2n}\mathbb{E}[\|\eta_{i_k}^{(k)}\|^2]\,.
\end{aligned}
\tag{34}
$$

We now give an upper bound of $\mathbb{E}[\|S_{\text{tts}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2]$ using Lemma 5 and plug it into (34):

$$
\begin{aligned}
(a_k - 2\gamma_{k+1}^2\,\mathrm{L}_V)\mathbb{E}[\|\overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] \leq& \mathbb{E}[V(\hat{\boldsymbol{s}}^{(k)}) - V(\hat{\boldsymbol{s}}^{(k+1)})] \\
&+ \gamma_{k+1}\left(\frac{1}{2\beta}(1-1/n) + 2\gamma_{k+1}\,\mathrm{L}_V\right)\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^n \tilde{S}_i^{(\tau_i^k)} - \overline{\mathbf{s}}^{(k)}\|^2] \\
&+ \gamma_{k+1}\left(\gamma_{k+1}\,\mathrm{L}_V + \frac{1}{2n}\right)\mathbb{E}[\|\eta_{i_k}^{(k)}\|^2] \\
&+ \frac{\gamma_{k+1}^2\,\mathrm{L}_V\,\mathrm{L}_{\mathbf{s}}^2}{n^3}\sum_{i=1}^n \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\tau_i^k)}\|^2]\,.
\end{aligned}
\tag{35}
$$

Next, we observe that

$$
\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(t_i^{k+1})}\|^2] = \frac{1}{n}\sum_{i=1}^n\left(\frac{1}{n}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2] + \frac{n-1}{n}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(\tau_i^k)}\|^2]\right)\,,
$$

where the equality holds as i_k and j_k are drawn independently. For any $\beta > 0$, it holds

$$
\begin{aligned}
&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2] \\
=&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2 + \|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\tau_i^k)}\|^2 + 2\langle \hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\,,\, \hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\tau_i^k)}\rangle] \\
=&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2 + \|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\tau_i^k)}\|^2 - 2\gamma_{k+1}\langle \hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}\,,\, \hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\tau_i^k)}\rangle] \\
\leq&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2 + \|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\tau_i^k)}\|^2 + \frac{\gamma_{k+1}}{\beta}\|\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}\|^2 + \gamma_{k+1}\beta\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\tau_i^k)}\|^2]\ ,
\end{aligned}
$$

where the last inequality is due to Young's inequality. Subsequently, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(\tau_i^{k+1})}\|^2] \\
\leq&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2] + \frac{n-1}{n^2}\sum_{i=1}^{n}\mathbb{E}[(1+\gamma_{k+1}\beta)\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\tau_i^k)}\|^2 + \frac{\gamma_{k+1}}{\beta}\|\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}\|^2]\ .
\end{aligned}
$$

Observe that $\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)} = -\gamma_{k+1}(\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)})$. Applying Lemma 5 yields

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(\tau_i^{k+1})}\|^2] \\
\leq&(\gamma_{k+1}^2 + \frac{n-1}{n}\frac{\gamma_{k+1}}{\beta})\mathbb{E}[\|S_{\text{tts}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2] + \sum_{i=1}^{n}\mathbb{E}[\frac{1-\frac{1}{n}+\gamma_{k+1}\beta}{n}\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\tau_i^k)}\|^2] \\
\leq&4(\gamma_{k+1}^2 + \frac{\gamma_{k+1}}{\beta})\mathbb{E}[\|\overline{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] + 2(\gamma_{k+1}^2 + \frac{\gamma_{k+1}}{\beta})\mathbb{E}[\|\eta_{i_k}^{(k)}\|^2] \\
&+4(\gamma_{k+1}^2 + \frac{\gamma_{k+1}}{\beta})\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^{n}\tilde{S}_i^{(\tau_i^k)} - \overline{\mathbf{s}}^{(k)}\|^2] \\
&+\sum_{i=1}^{n}\mathbb{E}[\frac{1-\frac{1}{n}+\gamma_{k+1}\beta + \frac{2\gamma_{k+1}\mathrm{L}_{\mathbf{s}}^2}{n^2}(\gamma_{k+1}+\frac{1}{\beta})}{n}\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2]\ .
\end{aligned}
$$

Let us define

$$
\Delta^{(k)} := \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\tau_i^k)}\|^2]\ .
$$

From the above, we obtain

$$
\begin{aligned}
\Delta^{(k+1)} \leq\ & (1-\frac{1}{n}+\gamma_{k+1}\beta + \frac{2\gamma_{k+1}\mathrm{L}_{\mathbf{s}}^2}{n^2}(\gamma_{k+1}+\frac{1}{\beta}))\Delta^{(k)} + 4(\gamma_{k+1}^2 + \frac{\gamma_{k+1}}{\beta})\mathbb{E}[\|\overline{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] \\
&+ 2(\gamma_{k+1}^2 + \frac{\gamma_{k+1}}{\beta})\mathbb{E}[\|\eta_{i_k}^{(k)}\|^2] + 4(\gamma_{k+1}^2 + \frac{\gamma_{k+1}}{\beta})\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^{n}\tilde{S}_i^{(\tau_i^k)} - \overline{\mathbf{s}}^{(k)}\|^2]\ .
\end{aligned}
$$

Setting $c_1 = \upsilon_{\min}^{-1}$, $\alpha = \max\{8, 1+6\upsilon_{\min}\}$, $\overline{L} = \max\{\mathrm{L}_{\mathbf{s}}, \mathrm{L}_V\}$, $\gamma_{k+1} = \frac{1}{k\alpha c_1\overline{L}}$, $\beta = \frac{c_1\overline{L}}{n}$, we remark $c_1(k\alpha-1) \geq c_1(\alpha-1) \geq 6$ and we observe that

$$
1-\frac{1}{n}+\gamma_{k+1}\beta + \frac{2\gamma_{k+1}\mathrm{L}_{\mathbf{s}}^2}{n^2}(\gamma_{k+1}+\frac{1}{\beta}) \leq 1 - \frac{c_1(k\alpha-1)-4}{k\alpha n c_1} \leq 1 - \frac{2}{k\alpha n c_1}\ ,
$$

which shows that $1-\frac{1}{n}+\gamma_{k+1}\beta+\frac{2\gamma_{k+1}\mathrm{L}_{\mathbf{s}}^2}{n^2}(\gamma_{k+1}+\frac{1}{\beta})\in(0,1)$ for any $k>0$. Denote $\Lambda_{(k+1)}=\frac{1}{n}-\gamma_{k+1}\beta-\frac{2\gamma_{k+1}\mathrm{L}_{\mathbf{s}}^2}{n^2}(\gamma_{k+1}+\frac{1}{\beta})$ and note that $\Delta^{(0)}=0$, thus the telescoping sum yields:

$$\begin{aligned}\Delta^{(k+1)} \leq\, & 4\sum_{\ell=0}^{k}\prod_{j=\ell+1}^{k}(1-\Lambda_{(j)})(\gamma_{\ell+1}^2+\frac{\gamma_{\ell+1}}{\beta})\mathbb{E}[\|\overline{\boldsymbol{s}}^{(\ell)}-\hat{\boldsymbol{s}}^{(\ell)}\|^2]\\ &+2\sum_{\ell=0}^{k}\prod_{j=\ell+1}^{k}(1-\Lambda_{(j)})(\gamma_{\ell+1}^2+\frac{\gamma_{\ell+1}}{\beta})\mathbb{E}[\|\eta_{i_\ell}^{(\ell)}\|^2]\\ &+4\sum_{\ell=0}^{k}\prod_{j=\ell+1}^{k}(1-\Lambda_{(j)})(\gamma_{\ell+1}^2+\frac{\gamma_{\ell+1}}{\beta})\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^{n}\tilde{S}_i^{(\tau_i^\ell)}-\overline{\mathbf{s}}^{(\ell)}\|^2]\,.\end{aligned}$$

Note $\omega_{k,\ell}=\prod_{j=\ell+1}^{k}(1-\Lambda_{(j)})$ Summing on both sides over $k=0$ to $k=K_m-1$ yields:

$$\begin{aligned}\sum_{k=0}^{K_m-1}\Delta^{(k+1)} =&4\sum_{k=0}^{K_m-1}(\gamma_{k+1}^2+\frac{\gamma_{k+1}}{\beta})\omega_{k,1}\mathbb{E}[\|\overline{\boldsymbol{s}}^{(k)}-\hat{\boldsymbol{s}}^{(k)}\|^2]+2\sum_{k=0}^{K_m-1}(\gamma_{k+1}^2+\frac{\gamma_{k+1}}{\beta})\omega_{k,1}\mathbb{E}[\left\|\eta_{i_\ell}^{(k)}\right\|^2]\\ &+\sum_{k=0}^{K_m-1}4(\gamma_{k+1}^2+\frac{\gamma_{k+1}}{\beta})\omega_{k,1}\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^{n}\tilde{S}_i^{(\tau_i^k)}-\overline{\mathbf{s}}^{(k)}\|^2]\\ \leq&\sum_{k=0}^{K_m-1}\frac{4(\gamma_{k+1}^2+\frac{\gamma_{k+1}}{\beta})}{\Lambda_{(k+1)}}\mathbb{E}[\|\overline{\boldsymbol{s}}^{(k)}-\hat{\boldsymbol{s}}^{(k)}\|^2]+\sum_{k=0}^{K_m-1}\frac{2(\gamma_{k+1}^2+\frac{\gamma_{k+1}}{\beta})}{\Lambda_{(k+1)}}\mathbb{E}[\left\|\eta_{i_\ell}^{(k)}\right\|^2]\\ &+\sum_{k=0}^{K_m-1}\frac{4(\gamma_{k+1}^2+\frac{\gamma_{k+1}}{\beta})}{\Lambda_{(k+1)}}\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^{n}\tilde{S}_i^{(\tau_i^k)}-\overline{\mathbf{s}}^{(k)}\|^2]\,.\end{aligned}\tag{36}$$

We recall (35) where we have summed on both sides from $k=0$ to $k=K_m-1$:

$$\begin{aligned}&\sum_{k=0}^{K_m-1}\left(a_k-2\gamma_{k+1}^2\mathrm{L}_V\right)\mathbb{E}[\|\overline{\mathbf{s}}^{(k)}-\hat{\boldsymbol{s}}^{(k)}\|^2]\\ \leq&\mathbb{E}[V(\hat{\boldsymbol{s}}^{(0)})-V(\hat{\boldsymbol{s}}^{(K)})]+\sum_{k=0}^{K_m-1}\gamma_{k+1}\left(\frac{1}{2\beta}(1-1/n)+2\gamma_{k+1}\mathrm{L}_V\right)\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^{n}\tilde{S}_i^{(\tau_i^k)}-\overline{\mathbf{s}}^{(k)}\|^2]\\ &+\sum_{k=0}^{K_m-1}\gamma_{k+1}\left(\gamma_{k+1}\mathrm{L}_V+\frac{1}{2n}\right)\mathbb{E}[\|\eta_{i_k}^{(k)}\|^2]+\sum_{k=0}^{K_m-1}\frac{\gamma_{k+1}^2\mathrm{L}_V\mathrm{L}_{\mathbf{s}}^2}{n^2}\Delta^{(k)}\,.\end{aligned}\tag{37}$$

Plugging (36) into (37) results in:

$$\sum_{k=0}^{K_m-1}\tilde{\alpha}_k\mathbb{E}[\|\overline{\mathbf{s}}^{(k)}-\hat{\boldsymbol{s}}^{(k)}\|^2]+\sum_{k=0}^{K_m-1}\tilde{\beta}_k\mathbb{E}[\|\frac{1}{n}\sum_{i=1}^{n}\tilde{S}_i^{(\tau_i^k)}-\overline{\mathbf{s}}^{(k)}\|^2]\leq\mathbb{E}[V(\hat{\boldsymbol{s}}^{(0)})-V(\hat{\boldsymbol{s}}^{(K)})]+\sum_{k=0}^{K_m-1}\tilde{\Gamma}_k\mathbb{E}[\|\eta_{i_k}^{(k)}\|^2]\,,$$

where

$$\tilde{\alpha}_k = a_k - 2\gamma_{k+1}^2 \mathrm{L}_V - \frac{\gamma_{k+1}^2 \mathrm{L}_V \mathrm{L}_{\mathbf{s}}^2}{n^2} \frac{4(\gamma_{k+1}^2 + \frac{\gamma_{k+1}}{\beta})}{\Lambda_{(k+1)}} ,$$
$$\tilde{\beta}_k = \gamma_{k+1} \left(\frac{1}{2\beta}(1 - 1/n) + 2\gamma_{k+1} \mathrm{L}_V \right) - \frac{\gamma_{k+1}^2 \mathrm{L}_V \mathrm{L}_{\mathbf{s}}^2}{n^2} \frac{4(\gamma_{k+1}^2 + \frac{\gamma_{k+1}}{\beta})}{\Lambda_{(k+1)}} ,$$
$$\tilde{\Gamma}_k = \gamma_{k+1} \left(\gamma_{k+1} \mathrm{L}_V + \frac{1}{2n} \right) + \frac{\gamma_{k+1}^2 \mathrm{L}_V \mathrm{L}_{\mathbf{s}}^2}{n^2} \frac{2(\gamma_{k+1}^2 + \frac{\gamma_{k+1}}{\beta})}{\Lambda_{(k+1)}} ,$$

and

$$a_k = \gamma_{k+1} \left(\upsilon_{\min} - \upsilon_{\max}^2 \frac{\beta(n-1)+1}{2n} \right) ,$$
$$\Lambda_{(k+1)} = \frac{1}{n} - \gamma_{k+1}\beta - \frac{2\gamma_{k+1} \mathrm{L}_{\mathbf{s}}^2}{n^2}(\gamma_{k+1} + \frac{1}{\beta}) ,$$
$$c_1 = \upsilon_{\min}^{-1}, \alpha = \max\{8, 1 + 6\upsilon_{\min}\}, \overline{L} = \max\{\mathrm{L}_{\mathbf{s}}, \mathrm{L}_V\}, \gamma_{k+1} = \frac{1}{k\alpha c_1 \overline{L}}, \beta = \frac{c_1 \overline{L}}{n} .$$

When, for any $k > 0$, $\tilde{\alpha}_k \geq 0$, we have by Lemma 2 that:

$$\sum_{k=0}^{K_m} \tilde{\alpha}_k \mathbb{E}[\|\nabla V(\hat{\boldsymbol{s}}^{(k)})\|^2] \leq \upsilon_{\max}^2 \sum_{k=0}^{K_m} \tilde{\alpha}_k \mathbb{E}[\|\overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] ,$$

concluding the proof of the Theorem. □

B Proofs for the vrTTEM and the fiTTEM Algorithms

B.1 Additional Intermediary Results

We introduce additional Lemmas below before getting into the proofs of the desired results.

Lemma 9. *Consider the vrTTEM update (2) with $\rho_k = \rho$, it holds for all $k > 0$*

$$\begin{aligned}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - S_{tts}^{(k+1)}\|^2] \leq & 2\rho^2 \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \overline{\boldsymbol{s}}^{(k)}\|^2] + 2\rho^2 \, \mathrm{L}_{\mathbf{s}}^2 \, \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))}\|^2] \\ & + 2(1-\rho)^2 \mathbb{E}[\|\hat{\boldsymbol{s}}^{((k))} - S_{tts}^{(k)}\|^2] + 2\rho^2 \mathbb{E}[\|\eta_{i_k}^{(k+1)}\|^2] \,,\end{aligned}$$

where we recall that $\ell(k)$ is the first iteration number in the epoch that iteration k is in.

Proof. Beforehand, we provide an alternate expression of the quantity $\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}$ that will be useful throughout this proof:

$$\begin{aligned}\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)} = -\gamma_{k+1}(\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}) & = -\gamma_{k+1}(\hat{\boldsymbol{s}}^{(k)} - (1-\rho) S_{\text{tts}}^{(k)} - \rho \boldsymbol{\mathcal{S}}^{(k+1)}) \\ & = -\gamma_{k+1}\left((1-\rho)[\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k)}] + \rho[\hat{\boldsymbol{s}}^{(k)} - \boldsymbol{\mathcal{S}}^{(k+1)}] \right) \,.\end{aligned} \tag{38}$$

We observe, using the identity (38), that

$$\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}\|^2] \leq 2\rho^2 \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \overline{\boldsymbol{s}}^{(k)}\|^2] + 2\rho^2 \mathbb{E}[\|\overline{\boldsymbol{s}}^{(k)} - \boldsymbol{\mathcal{S}}^{(k+1)}\|^2] + 2(1-\rho)^2 \mathbb{E}[\|\hat{\boldsymbol{s}}^{((k))} - S_{\text{tts}}^{(k)}\|^2]. \tag{39}$$

For the latter term, we obtain its upper bound as

$$\begin{aligned}\mathbb{E}[\|\overline{\boldsymbol{s}}^{(k)} - \boldsymbol{\mathcal{S}}^{(k+1)}\|^2] = & \mathbb{E}[\|\frac{1}{n}\sum_{i=1}^n (\overline{\boldsymbol{s}}_i^{(k)} - \tilde{S}_i^{\ell(k)}) - (\overline{\boldsymbol{s}}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(\ell(k))})\|^2] \\ & \overset{(a)}{\leq} \mathbb{E}[\|\overline{\boldsymbol{s}}_{i_k}^{(k)} - \overline{\boldsymbol{s}}_{i_k}^{(\ell(k))}\|^2] + \mathbb{E}[\|\eta_{i_k}^{(k+1)}\|^2] \overset{(b)}{\leq} \mathrm{L}_{\mathbf{s}}^2 \, \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))}\|^2] + \mathbb{E}[\|\eta_{i_k}^{(k+1)}\|^2] \,,\end{aligned}$$

where (a) uses the variance inequality and (b) uses Lemma 1. Substituting into (39) proves the lemma. □

Lemma 10. *Consider the fiTTEM update (3) with $\rho_k = \rho$. It holds for all $k > 0$ that*

$$\begin{aligned}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - S_{tts}^{(k+1)}\|^2] \leq & 2\rho^2 \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \overline{\boldsymbol{s}}^{(k)}\|^2] + 2\rho^2 \frac{\mathrm{L}_{\mathbf{s}}^2}{n} \sum_{i=1}^n \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2] \\ & + 2(1-\rho)^2 \mathbb{E}[\|\hat{\boldsymbol{s}}^{((k))} - S_{tts}^{(k)}\|^2] + 2\rho^2 \mathbb{E}[\|\eta_{i_k}^{(k+1)}\|^2] \,,\end{aligned}$$

where $\mathrm{L}_{\mathbf{s}}$ is the smoothness constant defined in Lemma 1.

Proof. Beforehand, we provide a rewriting of the quantity $\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}$ that will be useful throughout this proof:

$$
\begin{aligned}
\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)} &= -\gamma_{k+1}(\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}) \\
&= -\gamma_{k+1}(\hat{\boldsymbol{s}}^{(k)} - (1-\rho)S_{\text{tts}}^{(k)} - \rho\boldsymbol{\mathcal{S}}^{(k+1)}) \\
&= -\gamma_{k+1}\left((1-\rho)[\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k)}] + \rho[\hat{\boldsymbol{s}}^{(k)} - \boldsymbol{\mathcal{S}}^{(k+1)}]\right) \\
&= -\gamma_{k+1}\left((1-\rho)[\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k)}] + \rho[\hat{\boldsymbol{s}}^{(k)} - \overline{\boldsymbol{\mathcal{S}}}^{(k)} - (\tilde{S}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)})]\right) \ .
\end{aligned}
\tag{40}
$$

We observe, using the identity (40), that

$$
\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}\|^2] \leq 2\rho^2\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \overline{\boldsymbol{s}}^{(k)}\|^2] + 2\rho^2\mathbb{E}[\|\overline{\boldsymbol{s}}^{(k)} - \boldsymbol{\mathcal{S}}^{(k+1)}\|^2] + 2(1-\rho)^2\mathbb{E}[\|\hat{\boldsymbol{s}}^{((k))} - S_{\text{tts}}^{(k)}\|^2] \ . \tag{41}
$$

For the latter term, we obtain its upper bound as

$$
\begin{aligned}
\mathbb{E}[\|\overline{\boldsymbol{s}}^{(k)} - \boldsymbol{\mathcal{S}}^{(k+1)}\|^2] &= \mathbb{E}[\|\frac{1}{n}\sum_{i=1}^{n}(\overline{\boldsymbol{s}}_i^{(k)} - \overline{\boldsymbol{\mathcal{S}}}_i^{(k)}) - (\tilde{S}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)})\|^2] \\
&\overset{(a)}{\leq} \mathbb{E}[\|\overline{\boldsymbol{s}}_{i_k}^{(k)} - \overline{\boldsymbol{s}}_{i_k}^{(\ell(k))}\|^2] + \mathbb{E}[\|\eta_{i_k}^{(k+1)}\|^2] \ ,
\end{aligned}
$$

where (a) uses the variance inequality. We can further bound the last expectation using Lemma 1:

$$
\mathbb{E}[\|\overline{\boldsymbol{s}}_{i_k}^{(k)} - \overline{\boldsymbol{s}}_{i_k}^{(t_{i_k}^k)}\|^2] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\|\overline{\boldsymbol{s}}_i^{(k)} - \overline{\boldsymbol{s}}_i^{(t_i^k)}\|^2] \overset{(a)}{\leq} \frac{\mathrm{L}_{\mathbf{s}}^2}{n}\sum_{i=1}^{n}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2] \ .
$$

Substituting the above into (41) proves the lemma. □

Lemma 11. *Considering a decreasing stepsize $\gamma_k \in (0,1)$ and a constant $\rho \in (0,1)$, we have*

$$
\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - S_{tts}^{(k)}\|^2] \leq \frac{\rho}{1-\rho}\sum_{\ell=0}^{k}(1-\gamma_\ell)^2(\boldsymbol{\mathcal{S}}^{(\ell)} - \tilde{S}^{(\ell)}) \ ,
$$

where $\boldsymbol{\mathcal{S}}^{(k)}$ is defined either by Line 2 (vrTTEM) or Line 3 (fiTTEM).

Proof. We begin by writing the two-timescale update:

$$
\begin{aligned}
S_{\text{tts}}^{(k+1)} &= S_{\text{tts}}^{(k)} + \rho(\boldsymbol{\mathcal{S}}^{(k+1)} - S_{\text{tts}}^{(k)}) \ , \\
\hat{\mathbf{s}}^{(k+1)} &= \hat{\mathbf{s}}^{(k)} + \gamma_{k+1}(S_{\text{tts}}^{(k+1)} - \hat{\mathbf{s}}^{(k)}) \ ,
\end{aligned}
\tag{42}
$$

where $\boldsymbol{\mathcal{S}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^n \tilde{S}_i^{(t_i^k)} + (\tilde{S}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)})$ according to (3). Denote $\delta^{(k+1)} = \hat{\boldsymbol{s}}^{(k+1)} - S_{\text{tts}}^{(k+1)}$. Then from (42), doing the subtraction of both equations yields:

$$\delta^{(k+1)} = (1-\gamma_{k+1})\delta^{(k)} + \frac{\rho}{1-\rho}(1-\gamma_{k+1})(\boldsymbol{\mathcal{S}}^{(k+1)} - S_{\text{tts}}^{(k+1)}) \ .$$

Using the telescoping sum and noting that $\delta^{(0)} = 0$, we have

$$\delta^{(k+1)} \leq \frac{\rho}{1-\rho}\sum_{\ell=0}^{k}(1-\gamma_{\ell+1})^2(\boldsymbol{\mathcal{S}}^{(\ell+1)} - \tilde{S}^{(\ell+1)}) \ .$$

□

B.2 Proofs of Auxiliary Lemmas (Lemma 6, Lemma 8 and Lemma 3)

Lemma. *At iteration $k+1$,the drift term of update (3), with $\rho_{k+1} = \rho$, is equivalent to the following :*

$$\begin{aligned}\hat{\boldsymbol{s}}^{(k)} - S_{tts}^{(k+1)} =& \rho(\hat{\boldsymbol{s}}^{(k)} - \overline{\mathbf{s}}^{(k)}) + \rho\eta_{i_k}^{(k+1)} + \rho[(\overline{\mathbf{s}}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)}) - \mathbb{E}[\overline{\mathbf{s}}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)}]] \\ &+ (1-\rho)\left(\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\right) \ ,\end{aligned}$$

where we recall that $\eta_{i_k}^{(k+1)}$, defined in (13), which is the gap between the MC approximation and the expected statistics.

Proof. Using the fiTTEM update $S_{\text{tts}}^{(k+1)} = (1-\rho)S_{\text{tts}}^{(k)} + \rho\boldsymbol{\mathcal{S}}^{(k+1)}$ where $\boldsymbol{\mathcal{S}}^{(k+1)} = \overline{\boldsymbol{\mathcal{S}}}^{(k)} + (\tilde{S}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)})$ leads to the following decomposition:

$$\begin{aligned}&S_{\text{tts}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)} \\ =&(1-\rho)S_{\text{tts}}^{(k)} + \rho\left(\overline{\boldsymbol{\mathcal{S}}}^{(k)} + (\tilde{S}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)})\right) - \hat{\boldsymbol{s}}^{(k)} + \rho\overline{\mathbf{s}}^{(k)} - \rho\overline{\mathbf{s}}^{(k)} \\ =&\rho(\overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}) + \rho(\tilde{S}_{i_k}^{(k)} - \overline{\mathbf{s}}_{i_k}^{(k)}) + (1-\rho)\left(S_{\text{tts}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\right) + \rho\left(\overline{\boldsymbol{\mathcal{S}}}^{(k)} - \overline{\mathbf{s}}^{(k)} + (\overline{\mathbf{s}}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)})\right) \\ =&\rho(\overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}) + \rho\eta_{i_k}^{(k+1)} - \rho[(\overline{\mathbf{s}}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)}) - \mathbb{E}[\overline{\mathbf{s}}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)}]] \\ &+(1-\rho)\left(S_{\text{tts}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\right) \ ,\end{aligned}$$

where we observe that $\mathbb{E}[\overline{\mathbf{s}}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)}] = \overline{\mathbf{s}}^{(k)} - \overline{\boldsymbol{\mathcal{S}}}^{(k)}$ and which concludes the proof.

Important Note: Note that $\overline{\mathbf{s}}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)}$ is not equal to $\eta_{i_k}^{(k+1)}$, defined in (13), which is the gap between the MC approximation and the expected statistics. Indeed $\tilde{S}_{i_k}^{(t_{i_k}^k)}$ is not computed under the same model as $\overline{\mathbf{s}}_{i_k}^{(k)}$. □

B.3 Proof of Theorem 2

Theorem. *Assume A1-A5. Consider the vrTTEM sequence* $\{\hat{\mathrm{s}}^{(k)}\}_{k>0} \in \mathcal{S}$ *for any* $k \leq K_m$ *where* K_m *is a positive integer. Let* $\{\gamma_{k+1} = 1/(k^a\overline{L})\}_{k>0}$*, where* $a \in (0,1)$*, be a sequence of stepsizes,* $\overline{L} = \max\{\mathrm{L}_{\mathbf{s}}, \mathrm{L}_V\}$*,* $\rho = \mu/(c_1\overline{L}n^{2/3})$*,* $m = nc_1^2/(2\mu^2 + \mu c_1^2)$ *and a constant* $\mu \in (0,1)$*. Then:*

$$\mathbb{E}[\|\nabla V(\hat{\boldsymbol{s}}^{(K)})\|^2] \leq \frac{2n^{2/3}\overline{L}}{\mu P_m \upsilon_{\min}^2 \upsilon_{\max}^2}\left(\mathbb{E}[\Delta V] + \sum_{k=0}^{K_m-1} \tilde{\eta}^{(k+1)} + \chi^{(k+1)}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2]\right) .$$

Proof. Using the smoothness of V and update (2), we obtain:

$$\begin{aligned} V(\hat{\boldsymbol{s}}^{(k+1)}) &\leq V(\hat{\boldsymbol{s}}^{(k)}) + \left\langle \hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)} \,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)}) \right\rangle + \frac{\mathrm{L}_V}{2}\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2 \\ &\leq V(\hat{\boldsymbol{s}}^{(k)}) - \gamma_{k+1}\left\langle \hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)} \,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)}) \right\rangle + \frac{\gamma_{k+1}^2 \mathrm{L}_V}{2}\|\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}\|^2 . \end{aligned} \tag{43}$$

Denote $\mathsf{H}_{k+1} := \hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}$ the drift term of the fiTTEM update in (8) and $\mathsf{h}_k = \hat{\boldsymbol{s}}^{(k)} - \overline{\mathrm{s}}^{(k)}$. Taking expectations on both sides show that

$$\begin{aligned} &\mathbb{E}[V(\hat{\boldsymbol{s}}^{(k+1)})] \\ &\stackrel{(a)}{\leq} \mathbb{E}[V(\hat{\boldsymbol{s}}^{(k)})] - \gamma_{k+1}(1-\rho)\mathbb{E}[\left\langle \hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k)} \,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)}) \right\rangle] \\ &\quad - \gamma_{k+1}\rho\mathbb{E}[\left\langle \hat{\boldsymbol{s}}^{(k)} - \boldsymbol{\mathcal{S}}^{(k+1)} \,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)}) \right\rangle] + \frac{\gamma_{k+1}^2 \mathrm{L}_V}{2}\mathbb{E}[\|\mathsf{H}_{k+1}\|^2] \\ &\stackrel{(b)}{\leq} \mathbb{E}[V(\hat{\boldsymbol{s}}^{(k)})] - \gamma_{k+1}\rho\mathbb{E}[\left\langle \mathsf{h}_k \,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)}) \right\rangle] - \gamma_{k+1}(1-\rho)\mathbb{E}[\left\langle \hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k)} \,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)}) \right\rangle] \\ &\quad - \gamma_{k+1}\rho\mathbb{E}[\left\langle \eta_{i_k}^{(k+1)} \,,\, \nabla V(\hat{\boldsymbol{s}}^{(k)}) \right\rangle] + \frac{\gamma_{k+1}^2 \mathrm{L}_V}{2}\mathbb{E}[\|\mathsf{H}_{k+1}\|^2] \\ &\stackrel{(c)}{\leq} \mathbb{E}[V(\hat{\boldsymbol{s}}^{(k)})] - \left(\gamma_{k+1}\rho\upsilon_{\min} + \gamma_{k+1}\upsilon_{\max}^2\right)\mathbb{E}[\|\mathsf{h}_k\|^2] + \frac{\gamma_{k+1}^2 \mathrm{L}_V}{2}\mathbb{E}[\|\mathsf{H}_{k+1}\|^2] \\ &\quad - \gamma_{k+1}\rho\mathbb{E}[\left\|\eta_{i_k}^{(k+1)}\right\|^2] - \gamma_{k+1}(1-\rho)\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] , \end{aligned} \tag{44}$$

where we have used (38) in (a) and $\mathbb{E}[\boldsymbol{\mathcal{S}}^{(k+1)}] = \overline{\mathrm{s}}^{(k)} + \mathbb{E}[\eta_{i_k}^{(k+1)}]$ in (b), the growth condition in Lemma 2 and Young's inequality with the constant equal to 1 in (c). Furthermore, for $k+1 \leq \ell(k) + m$ (*i.e.*, $k+1$

is in the same epoch as k), we have

$$\begin{aligned}
&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(\ell(k))}\|^2] = \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)} + \hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))}\|^2] \\
=&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))}\|^2 + \|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2 + 2\big\langle \hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))} \, , \, \hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)} \big\rangle] \\
=&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))}\|^2 + \gamma_{k+1}^2 \|\mathsf{H}_{k+1}\|^2 \\
&-2\gamma_{k+1}\big\langle \hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))} \, , \, \rho(\mathsf{h}_k - \eta_{i_k}^{(k+1)}) + (1-\rho)(\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k)}) \big\rangle] \\
\leq&\mathbb{E}[(1+\gamma_{k+1}\beta)\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))}\|^2 + \gamma_{k+1}^2 \|\mathsf{H}_{k+1}\|^2 + \frac{\gamma_{k+1}\rho}{\beta}\|\mathsf{h}_k\|^2 \\
&+\frac{\gamma_{k+1}\rho}{\beta}\|\eta_{i_k}^{(k+1)}\|^2 + \frac{\gamma_{k+1}(1-\rho)}{\beta}\|\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k)}\|^2] \; ,
\end{aligned}$$

where we first used (38) and the last inequality is due to Young's inequality. Consider the following sequence:

$$R_k := \mathbb{E}[V(\hat{\boldsymbol{s}}^{(k)}) + b_k \|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))}\|^2] \; ,$$

where $b_k := \overline{b}_{k \bmod m}$ is a periodic sequence where:

$$\overline{b}_i = \overline{b}_{i+1}(1 + \gamma_{k+1}\beta + 2\gamma_{k+1}^2 \rho^2 \, \mathrm{L}_{\mathbf{s}}^2) + \gamma_{k+1}^2 \rho^2 \, \mathrm{L}_V \, \mathrm{L}_{\mathbf{s}}^2, \;\; i = 0, 1, \ldots, m-1 \;\; \text{with} \;\; \overline{b}_m = 0 \; .$$

Note that $\overline{b}_i$ is decreasing with i and this implies

$$\overline{b}_i \leq \overline{b}_0 = \gamma_{k+1}^2 \rho^2 \, \mathrm{L}_V \, \mathrm{L}_{\mathbf{s}}^2 \, \frac{(1 + \gamma_{k+1}\beta + 2\gamma_{k+1}^2 \rho^2 \, \mathrm{L}_{\mathbf{s}}^2)^m - 1}{\gamma_{k+1}\beta + 2\gamma_{k+1}^2 \rho^2 \, \mathrm{L}_{\mathbf{s}}^2}, \; i = 1, 2, \ldots, m \; .$$

For $k+1 \leq \ell(k) + m$, we have the following inequality

$$\begin{aligned}
R_{k+1} \leq & \, \mathbb{E}[V(\hat{\boldsymbol{s}}^{(k)}) - \big(\gamma_{k+1}\rho\upsilon_{\min} + \gamma_{k+1}\upsilon_{\max}^2\big) \|\mathsf{h}_k\|^2 + \frac{\gamma_{k+1}^2 \, \mathrm{L}_V}{2}\|\mathsf{H}_{k+1}\|^2] \\
& + \gamma_{k+1}\mathbb{E}[\rho \left\|\eta_{i_k}^{(k+1)}\right\|^2 - (1-\rho)\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] \\
& + b_{k+1}\mathbb{E}[(1+\gamma_{k+1}\beta)\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))}\|^2 + \gamma_{k+1}^2\|\mathsf{H}_{k+1}\|^2 + \frac{\gamma_{k+1}\rho}{\beta}\|\mathsf{h}_k\|^2] \\
& + b_{k+1}\mathbb{E}[\frac{\gamma_{k+1}\rho}{\beta}\|\eta_{i_k}^{(k+1)}\|^2 + \frac{\gamma_{k+1}(1-\rho)}{\beta}\|\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k)}\|^2] \; .
\end{aligned}$$

And using Lemma 6 we obtain:

$$\begin{aligned}
& R_{k+1} \\
\leq & \mathbb{E}[V(\hat{\boldsymbol{s}}^{(k)}) - \big(\gamma_{k+1}\rho\upsilon_{\min} + \gamma_{k+1}\upsilon_{\max}^2 - \gamma_{k+1}^2\rho^2 \, \mathrm{L}_V\big) \|\mathsf{h}_k\|^2 + \gamma_{k+1}^2 \rho^2 \, \mathrm{L}_V \, \mathrm{L}_{\mathbf{s}}^2 \, \|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))}\|^2] \\
& + b_{k+1}\mathbb{E}[(1 + \gamma_{k+1}\beta + 2\gamma_{k+1}^2\rho^2 \, \mathrm{L}_{\mathbf{s}}^2)\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))}\|^2 + (\frac{\gamma_{k+1}\rho}{\beta} + 2\gamma_{k+1}^2\rho^2)\|\mathsf{h}_k\|^2] \\
& + \gamma_{k+1}\mathbb{E}[(\rho + \rho^2\gamma_{k+1} \, \mathrm{L}_V) \left\|\eta_{i_k}^{(k+1)}\right\|^2 - (1 - \rho - (1-\rho)^2\gamma_{k+1} \, \mathrm{L}_V)\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] \\
& + b_{k+1}\mathbb{E}[(\frac{\gamma_{k+1}\rho}{\beta} + 2\gamma_{k+1}^2\rho^2)\|\eta_{i_k}^{(k+1)}\|^2 + (\frac{\gamma_{k+1}(1-\rho)}{\beta} + 2\gamma_{k+1}^2(1-\rho)^2)\|\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k)}\|^2] \; .
\end{aligned}$$

Rearranging the terms yields:

$$R_{k+1} \leq \mathbb{E}[V(\hat{\boldsymbol{s}}^{(k)})] - \gamma_{k+1}(\rho \upsilon_{\min} + \upsilon_{\max}^2 - \gamma_{k+1}\rho^2 \mathrm{L}_V - b_{k+1}(\frac{\rho}{\beta} + 2\gamma_{k+1}\rho^2))\mathbb{E}[\|\mathsf{h}_k\|^2]$$
$$+ (\underbrace{b_{k+1}(1 + \gamma\beta + 2\gamma^2\rho^2 \mathrm{L}_{\mathbf{s}}^2) + \gamma^2\rho^2 \mathrm{L}_V \mathrm{L}_{\mathbf{s}}^2}_{=b_k \quad \text{since } k+1 \leq \ell(k)+m})\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(\ell(k))}\|^2] + \tilde{\eta}^{(k+1)} + \tilde{\chi}^{(k+1)} ,$$

where

$$\tilde{\eta}^{(k+1)} = \left(\gamma_{k+1}(\rho + \rho^2\gamma_{k+1}\mathrm{L}_V) + b_{k+1}(\frac{\gamma_{k+1}\rho}{\beta} + 2\gamma_{k+1}^2\rho^2)\right)\mathbb{E}[\left\|\eta_{i_k}^{(k+1)}\right\|^2]$$
$$\chi^{(k+1)} = \left(b_{k+1}(\frac{\gamma_{k+1}(1-\rho)}{\beta} + 2\gamma_{k+1}^2(1-\rho)^2) - \gamma_{k+1}(1 - \rho - (1-\rho)^2\gamma_{k+1}\mathrm{L}_V)\right)$$
$$\tilde{\chi}^{(k+1)} = \chi^{(k+1)}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k)}\|^2] .$$

This leads, using Lemma 2, that for any γ_{k+1}, ρ and β such that $\rho \upsilon_{\min} + \upsilon_{\max}^2 - \gamma_{k+1}\rho^2 \mathrm{L}_V - b_{k+1}(\frac{\rho}{\beta} + 2\gamma_{k+1}\rho^2) > 0$,

$$\upsilon_{\max}^2\mathbb{E}[\|\nabla V(\hat{\boldsymbol{s}}^{(k)})\|^2] \leq \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \overline{\boldsymbol{s}}^{(k)}\|^2] \leq \frac{R_k - R_{k+1}}{\gamma_{k+1}(\rho \upsilon_{\min} + \upsilon_{\max}^2 - \gamma_{k+1}\rho^2 \mathrm{L}_V - b_{k+1}(\frac{\rho}{\beta} + 2\gamma_{k+1}\rho^2))}$$
$$+ \frac{\tilde{\eta}^{(k+1)} + \tilde{\chi}^{(k+1)}}{\gamma_{k+1}(\rho \upsilon_{\min} + \upsilon_{\max}^2 - \gamma_{k+1}\rho^2 \mathrm{L}_V - b_{k+1}(\frac{\rho}{\beta} + 2\gamma_{k+1}\rho^2))} .$$

We first remark that

$$\gamma_{k+1}(\rho \upsilon_{\min} + \upsilon_{\max}^2 - \gamma_{k+1}\rho^2 \mathrm{L}_V - b_{k+1}(\frac{\rho}{\beta} + 2\gamma_{k+1}\rho^2)) \geq \frac{\gamma_{k+1}\rho}{c_1}(1 - \gamma_{k+1}c_1\rho \mathrm{L}_V - b_{k+1}(\frac{c_1}{\beta} + 2\gamma_{k+1}\rho c_1)) ,$$

where $c_1 = \upsilon_{\min}^{-1}$. By setting $\overline{L} = \max\{\mathrm{L}_{\mathbf{s}}, \mathrm{L}_V\}$, $\beta = \frac{c_1\overline{L}}{n^{1/3}}$, $\rho = \frac{\mu}{c_1\overline{L}n^{2/3}}$, $m = \frac{nc_1^2}{2\mu^2 + \mu c_1^2}$ and $\{\gamma_{k+1}\}$ any sequence of decreasing stepsizes in $(0,1)$, it can be shown that there exists $\mu \in (0,1)$, such that the following lower bound holds

$$\begin{aligned}
&1 - \gamma_{k+1}c_1\rho \mathrm{L}_V - b_{k+1}(\frac{c_1}{\beta} + 2\gamma_{k+1}\rho c_1) \\
\geq& 1 - \frac{\mu}{n^{\frac{2}{3}}} - \overline{b}_0(\frac{n^{\frac{1}{3}}}{\overline{L}} + \frac{2\mu}{\overline{L}n^{\frac{2}{3}}}) \\
\geq& 1 - \frac{\mu}{n^{\frac{2}{3}}} - \frac{\mathrm{L}_V \mu^2}{c_1^2 n^{\frac{4}{3}}}\frac{(1 + \gamma\beta + 2\gamma^2 \mathrm{L}_{\mathbf{s}}^2)^m - 1}{\gamma\beta + 2\gamma^2 \mathrm{L}_{\mathbf{s}}^2}(\frac{n^{\frac{1}{3}}}{\overline{L}} + \frac{2\mu}{\overline{L}n^{\frac{2}{3}}}) \\
\overset{(a)}{\geq}& 1 - \frac{\mu}{n^{\frac{2}{3}}} - \frac{\mu}{c_1^2}(\mathrm{e} - 1)(1 + \frac{2\mu}{n}) \geq 1 - \mu - \mu(1 + 2\mu)\frac{\mathrm{e} - 1}{c_1^2} \overset{(b)}{\geq} \frac{1}{2} ,
\end{aligned}$$

where the simplification in (a) is due to

$$\frac{\mu}{n} \leq \gamma\beta + 2\gamma^2 \mathrm{L}_{\mathbf{s}}^2 \leq \frac{\mu}{n} + \frac{2\mu^2}{c_1^2 n^{\frac{4}{3}}} \leq \frac{\mu c_1^2 + 2\mu^2}{c_1^2}\frac{1}{n} \quad \text{and} \quad (1 + \gamma\beta + 2\gamma^2 \mathrm{L}_{\mathbf{s}}^2)^m \leq \mathrm{e} - 1 ,$$

and the required μ in (b) can be found by solving the quadratic equation. Finally, these results yield:

$$\upsilon_{\max}^2 \sum_{k=0}^{K_m-1} \gamma_{k+1} \mathbb{E}[\|\nabla V(\hat{\boldsymbol{s}}^{(k)})\|^2] \leq \frac{2(R_0 - R_{K_m})}{\upsilon_{\min}\rho} + 2 \sum_{k=0}^{K_m-1} \frac{\tilde{\eta}^{(k+1)} + \tilde{\chi}^{(k+1)}}{\upsilon_{\min}\rho} .$$

Note that $R_0 = \mathbb{E}[V(\hat{\boldsymbol{s}}^{(0)})]$ and if K_m is a multiple of m, then $R_{max} = \mathbb{E}[V(\hat{\boldsymbol{s}}^{(K_m)})]$. Under the latter condition, we have

$$\sum_{k=0}^{K_m-1} \gamma_{k+1} \mathbb{E}[\|\nabla V(\hat{\boldsymbol{s}}^{(k)})\|^2] \leq \frac{2n^{2/3}\overline{L}}{\mu\upsilon_{\min}^2\upsilon_{\max}^2} \mathbb{E}[V(\hat{\boldsymbol{s}}^{(0)}) - V(\hat{\boldsymbol{s}}^{(K_m)})] + \frac{2n^{2/3}\overline{L}}{\mu\upsilon_{\min}^2\upsilon_{\max}^2} \sum_{k=0}^{K_m-1} [\tilde{\eta}^{(k+1)} + \tilde{\chi}^{(k+1)}] .$$

This concludes our proof.

$\square$

B.4 Proof of Theorem 3

Theorem. *Assume A1-A5. Consider the fiTTEM sequence $\{\hat{\mathrm{s}}^{(k)}\}_{k>0} \in \mathcal{S}$ for any $k \leq K_m$ where K_m be a positive integer. Let $\{\gamma_{k+1} = 1/(k^a \alpha c_1 \overline{L})\}_{k>0}$, where $a \in (0,1)$, be a sequence of positive stepsizes, $\alpha = \max\{2, 1+2\upsilon_{\min}\}$, $\overline{L} = \max\{\mathrm{L}_{\mathbf{s}}, \mathrm{L}_V\}$, $\beta = 1/(\alpha n)$, $\rho = 1/(\alpha c_1 \overline{L} n^{2/3})$ and $c_1(k\alpha - 1) \geq c_1(\alpha - 1) \geq 2$. Then:*

$$
\mathbb{E}[\|\nabla V(\hat{\boldsymbol{s}}^{(K)})\|^2] \leq \frac{4\alpha \overline{L} n^{2/3}}{P_m \upsilon_{\min}^2 \upsilon_{\max}^2} \left(\mathbb{E}[\Delta V] + \sum_{k=0}^{K_m - 1} \Xi^{(k+1)} + \Gamma^{(k+1)} \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] \right) .
$$

Proof. Using the smoothness of V and update (3), we obtain:

$$
\begin{aligned}
V(\hat{\boldsymbol{s}}^{(k+1)}) &\leq V(\hat{\boldsymbol{s}}^{(k)}) + \left\langle \hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)} \, , \, \nabla V(\hat{\boldsymbol{s}}^{(k)}) \right\rangle + \frac{\mathrm{L}_V}{2} \|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2 \\
&\leq V(\hat{\boldsymbol{s}}^{(k)}) - \gamma_{k+1} \left\langle \hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)} \, , \, \nabla V(\hat{\boldsymbol{s}}^{(k)}) \right\rangle + \frac{\gamma_{k+1}^2 \mathrm{L}_V}{2} \|\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}\|^2 .
\end{aligned} \tag{45}
$$

Denote $\mathsf{H}_{k+1} := \hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}$ the drift term of the fiTTEM update in (8) and $\mathsf{h}_k = \hat{\boldsymbol{s}}^{(k)} - \overline{\mathbf{s}}^{(k)}$. Using Lemma 7 and the additional following identity:

$$
\mathbb{E}[(\overline{\mathbf{s}}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)}) - \mathbb{E}[\overline{\mathbf{s}}_{i_k}^{(k)} - \tilde{S}_{i_k}^{(t_{i_k}^k)}]] = 0 \, , \tag{46}
$$

we have

$$
\begin{aligned}
\mathbb{E}[V(\hat{\boldsymbol{s}}^{(k+1)})] \leq & \mathbb{E}[V(\hat{\boldsymbol{s}}^{(k)})] - \gamma_{k+1} \rho \mathbb{E}[\left\langle \mathsf{h}_k \, , \, \nabla V(\hat{\boldsymbol{s}}^{(k)}) \right\rangle] \\
& - \gamma_{k+1} \mathbb{E}[\left\langle \rho \mathbb{E}[\eta_{i_k}^{(k+1)} | \mathcal{F}_k] + (1-\rho) \mathbb{E}[\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}] \, , \, \nabla V(\hat{\boldsymbol{s}}^{(k)}) \right\rangle] + \frac{\gamma_{k+1}^2 \mathrm{L}_V}{2} \|\mathsf{H}_{k+1}\|^2 \\
\overset{(a)}{\leq} & - \upsilon_{\min} \gamma_{k+1} \rho \mathbb{E}[\|\mathsf{h}_k\|^2] - \gamma_{k+1} \mathbb{E}[\left\|\nabla V(\hat{\boldsymbol{s}}^{(k)})\right\|^2] \\
& - \frac{\gamma_{k+1} \rho^2}{2} \xi^{(k+1)} - \frac{\gamma_{k+1} (1-\rho)^2}{2} \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] + \frac{\gamma_{k+1}^2 \mathrm{L}_V}{2} \|\mathsf{H}_{k+1}\|^2 \\
\overset{(b)}{\leq} & - (\upsilon_{\min} \gamma_{k+1} \rho + \gamma_{k+1} \upsilon_{\max}^2) \mathbb{E}[\|\mathsf{h}_k\|^2] - \frac{\gamma_{k+1} \rho^2}{2} \xi^{(k+1)} - \frac{\gamma_{k+1} (1-\rho)^2}{2} \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] \\
& + \frac{\gamma_{k+1}^2 \mathrm{L}_V}{2} \|\mathsf{H}_{k+1}\|^2 \, ,
\end{aligned}
$$

where $\xi^{(k+1)} := \mathbb{E}[\|\mathbb{E}[\eta_{i_k}^{(k+1)} | \mathcal{F}_k]\|^2]$. Next, we bound the quantity $\mathbb{E}[\|\mathsf{H}_{k+1}\|^2]$. Using Lemma 8, we obtain

$$
\begin{aligned}
& \gamma_{k+1} (\upsilon_{\min} \rho + \upsilon_{\max}^2 - \gamma_{k+1} \rho^2 \mathrm{L}_V) \mathbb{E}[\|\mathsf{h}_k\|^2] \\
\leq & \mathbb{E}[V(\hat{\boldsymbol{s}}^{(k)}) - V(\hat{\boldsymbol{s}}^{(k+1)})] + \tilde{\xi}^{(k+1)} + \left((1-\rho)^2 \gamma_{k+1}^2 \mathrm{L}_V - \frac{\gamma_{k+1} (1-\rho)^2}{2} \right) \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] \\
& + \frac{\gamma_{k+1}^2 \mathrm{L}_V \rho^2 \mathrm{L}_{\mathbf{s}}^2}{n} \sum_{i=1}^{n} \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2] \, ,
\end{aligned} \tag{47}
$$

where $\tilde{\xi}^{(k+1)} = \gamma_{k+1}^2 \rho^2 \, \mathrm{L}_V \, \mathbb{E}[\|\eta_{i_k}^{(k+1)}\|^2] - \frac{\gamma_{k+1}\rho^2}{2}\xi^{(k+1)}$. Next, we observe that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(t_i^{k+1})}\|^2] = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{n}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2] + \frac{n-1}{n}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2]) \,, \tag{48}$$

where the equality holds as i_k and j_k are drawn independently. Then,

$$\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2] = \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2 + \|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2 + 2\langle \hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)} \,,\, \hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\rangle] \,.$$

Note that $\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)} = -\gamma_{k+1}(\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}) = -\gamma_{k+1}\mathsf{H}_{k+1}$ and that in expectation we recall that $\mathbb{E}[\mathsf{H}_{k+1}|\mathcal{F}_k] = \rho \mathsf{h}_k + \rho\mathbb{E}[\eta_{i_k}^{(k+1)}|\mathcal{F}_k] + (1-\rho)\mathbb{E}[S_{\text{tts}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}]$ where $\mathsf{h}_k = \hat{\boldsymbol{s}}^{(k)} - \overline{\mathsf{s}}^{(k)}$. Thus, for any $\beta > 0$, it holds

$$\begin{aligned}
&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2] = \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2 + \|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2 + 2\langle \hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)} \,,\, \hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\rangle] \\
\leq &\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2 + (1+\gamma_{k+1}\beta)\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2 + \frac{\gamma_{k+1}\rho^2}{\beta}\|\mathsf{h}_k\|^2 + \frac{\gamma_{k+1}\rho^2}{\beta}\mathbb{E}[\left\|\eta_{i_k}^{(k+1)}\right\|^2] \\
&+ \frac{\gamma_{k+1}(1-\rho)^2}{\beta}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2]] \,,
\end{aligned}$$

where the last inequality is due to Young's inequality. Plugging this into (48) yields:

$$\begin{aligned}
&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2] \\
=&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2 + \|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2 + 2\langle \hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)} \,,\, \hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\rangle] \\
\leq&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2 + (1+\gamma_{k+1}\beta)\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2 + \frac{\gamma_{k+1}\rho^2}{\beta}\|\mathsf{h}_k\|^2 + \frac{\gamma_{k+1}\rho^2}{\beta}\mathbb{E}[\left\|\eta_{i_k}^{(k+1)}\right\|^2] \\
&+ \frac{\gamma_{k+1}(1-\rho)^2}{\beta}\mathbb{E}[\left\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\right\|^2]] \,.
\end{aligned}$$

Subsequently, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(t_i^{k+1})}\|^2] \\
\leq&\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2] + \frac{n-1}{n^2}\sum_{i=1}^{n}\mathbb{E}[(1+\gamma_{k+1}\beta)\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2 + \frac{\gamma_{k+1}\rho^2}{\beta}\|\mathsf{h}_k\|^2 \\
&+ \frac{\gamma_{k+1}\rho^2}{\beta}\mathbb{E}[\|\eta_{i_k}^{(k+1)}\|^2] + \frac{\gamma_{k+1}(1-\rho)^2}{\beta}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}|^2]] \,.
\end{aligned}$$

We now use Lemma 8 on $\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(k)}\|^2 = \gamma_{k+1}^2 \|\hat{\boldsymbol{s}}^{(k)} - S_{\text{tts}}^{(k+1)}\|^2$ and obtain:

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k+1)} - \hat{\boldsymbol{s}}^{(t_i^{k+1})}\|^2] \\
\leq &\left(2\gamma_{k+1}^2\rho^2 + \frac{\gamma_{k+1}\rho^2}{\beta}\right) \mathbb{E}[\|\overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] \\
&+ \sum_{i=1}^{n} \left(\frac{\gamma_{k+1}^2 \rho^2 \, \mathrm{L}_{\mathbf{s}}^2}{n} + \frac{(n-1)(1+\gamma_{k+1}\beta)}{n^2}\right) \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2] \\
&+ \gamma_{k+1}(1-\rho)^2 \left(2\gamma_{k+1} + \frac{1}{\beta}\right) \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] + \left(2\gamma_{k+1}^2 + \frac{\gamma_{k+1}\rho^2}{\beta}\right) \mathbb{E}[\left\|\eta_{i_k}^{(k+1)}\right\|^2] \\
\leq &\left(2\gamma_{k+1}^2\rho^2 + \frac{\gamma_{k+1}\rho^2}{\beta}\right) \mathbb{E}[\|\overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] \\
&+ \sum_{i=1}^{n} \left(\frac{1 - \frac{1}{n} + \gamma_{k+1}\beta + \gamma_{k+1}^2\rho^2 \, \mathrm{L}_{\mathbf{s}}^2}{n} \right) \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2] \\
&+ \gamma_{k+1}(1-\rho)^2 \left(2\gamma_{k+1} + \frac{1}{\beta}\right) \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] + \left(2\gamma_{k+1}^2 + \frac{\gamma_{k+1}\rho^2}{\beta}\right) \mathbb{E}[\left\|\eta_{i_k}^{(k+1)}\right\|^2] \,.
\end{aligned}
$$

Let us define

$$
\Delta^{(k)} := \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \hat{\boldsymbol{s}}^{(t_i^k)}\|^2] \,.
$$

From the above, we obtain

$$
\begin{aligned}
\Delta^{(k+1)} \leq &\left(1 - \frac{1}{n} + \gamma_{k+1}\beta + \gamma_{k+1}^2\rho^2 \, \mathrm{L}_{\mathbf{s}}^2\right)\Delta^{(k)} + \left(2\gamma_{k+1}^2\rho^2 + \frac{\gamma_{k+1}\rho^2}{\beta}\right) \mathbb{E}[\|\overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] \\
&+ \gamma_{k+1}(1-\rho)^2 \left(2\gamma_{k+1} + \frac{1}{\beta}\right) \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] + \gamma_{k+1}\left(2\gamma_{k+1} + \frac{\rho^2}{\beta}\right) \mathbb{E}[\left\|\eta_{i_k}^{(k+1)}\right\|^2] \,.
\end{aligned}
$$

Setting $c_1 = \upsilon_{\min}^{-1}$, $\alpha = \max\{2, 1 + 2\upsilon_{\min}\}$, $\overline{L} = \max\{\mathrm{L}_{\mathbf{s}}, \mathrm{L}_V\}$, $\gamma_{k+1} = \frac{1}{k}$, $\beta = \frac{1}{\alpha n}$, $\rho = \frac{1}{\alpha c_1 \overline{L} n^{2/3}}$, then we have that $c_1(k\alpha - 1) \geq c_1(\alpha - 1) = \max\{\frac{1}{\upsilon_{\min}}, 2\} \geq 2$. Hence, we observe that

$$
1 - \frac{1}{n} + \gamma_{k+1}\beta + \gamma_{k+1}^2\rho^2 \, \mathrm{L}_{\mathbf{s}}^2 \leq 1 - \frac{1}{n} + \frac{1}{\alpha k n} + \frac{1}{\alpha^2 c_1^2 k^2 n^{\frac{4}{3}}} \leq 1 - \frac{c_1(k\alpha - 1) - 1}{k\alpha n c_1} \leq 1 - \frac{1}{k\alpha n c_1},
$$

which shows that $1 - \frac{1}{n} + \gamma_{k+1}\beta + \gamma_{k+1}^2\rho^2 \, \mathrm{L}_{\mathbf{s}}^2 \in (0,1)$ for any $k > 0$. Denote $\Lambda_{(k+1)} = \frac{1}{n} - \gamma_{k+1}\beta - \gamma_{k+1}^2\rho^2 \, \mathrm{L}_{\mathbf{s}}^2$ and note that $\Delta^{(0)} = 0$, thus the telescoping sum yields

$$
\begin{aligned}
\Delta^{(k+1)} \leq &\sum_{\ell=0}^{k} \omega_{k,\ell} \left(2\gamma_{\ell+1}^2\rho^2 + \frac{\gamma_{\ell+1}^2\rho^2}{\beta}\right) \mathbb{E}[\left\|\overline{\mathbf{s}}^{(\ell)} - \hat{\boldsymbol{s}}^{(\ell)}\right\|^2] \\
&+ \sum_{\ell=0}^{k} \omega_{k,\ell}\gamma_{\ell+1}(1-\rho)^2 \left(2\gamma_{\ell+1} + \frac{1}{\beta}\right) \mathbb{E}[\left\|\tilde{S}^{(\ell)} - \hat{\boldsymbol{s}}^{(\ell)}\right\|^2] + \sum_{\ell=0}^{k} \omega_{k,\ell}\gamma_{\ell+1}\tilde{\epsilon}^{(\ell+1)} \,,
\end{aligned}
$$

where $\omega_{k,\ell} = \prod_{j=\ell+1}^{k}(1-\Lambda_{(j)})$ and $\tilde{\epsilon}^{(\ell+1)} = \left(2\gamma_{k+1} + \frac{\rho^2}{\beta}\right)\mathbb{E}[\left\|\eta_{i_k}^{(k+1)}\right\|^2]$.

Summing on both sides over $k=0$ to $k=K_m-1$ yields:

$$\sum_{k=0}^{K_m-1} \Delta^{(k+1)} \leq \sum_{k=0}^{K_m-1} \frac{2\gamma_{k+1}^2\rho^2 + \frac{\gamma_{k+1}\rho^2}{\beta}}{\Lambda_{(k+1)}}\mathbb{E}[\|\overline{\mathbf{s}}^{(k)} - \hat{\boldsymbol{s}}^{(k)}\|^2] + \sum_{k=0}^{K_m-1} \frac{\gamma_{k+1}(1-\rho)^2\left(2\gamma_{k+1}+\frac{1}{\beta}\right)}{\Lambda_{(k+1)}}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] + \sum_{k=0}^{K_m-1} \frac{\gamma_{k+1}}{\Lambda_{(k+1)}}\tilde{\epsilon}^{(k+1)} .$$

We recall (47) where we have summed on both sides from $k=0$ to $k=K_m-1$:

$$\begin{aligned}
&\mathbb{E}[V(\hat{\mathbf{s}}^{(K_m)}) - V(\hat{\mathbf{s}}^{(0)})] \\
&\leq \sum_{k=0}^{K_m-1}\left\{\gamma_{k+1}(-(\upsilon_{\min}\rho + \upsilon_{\max}^2) + \gamma_{k+1}\rho^2 \,\mathrm{L}_V)\mathbb{E}[\|\mathsf{h}_k\|^2] + \gamma^2 \,\mathrm{L}_V\, \rho^2 \,\mathrm{L}_{\mathbf{s}}^2\, \Delta^{(k)}\right\} \\
&\quad + \sum_{k=0}^{K_m-1}\left\{\tilde{\xi}^{(k+1)} + \left((1-\rho)^2\gamma_{k+1}^2\,\mathrm{L}_V - \frac{\gamma_{k+1}(1-\rho)^2}{2}\right)\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2]\right\} \\
&\leq \sum_{k=0}^{K_m-1}\left\{\gamma_{k+1}\left[-(\upsilon_{\min}\rho + \upsilon_{\max}^2) + \gamma_{k+1}\rho^2\,\mathrm{L}_V + \frac{\rho^2\gamma_{k+1}\,\mathrm{L}_V\,\mathrm{L}_{\mathbf{s}}^2\left(2\gamma_{k+1}^2\rho^2 + \frac{\gamma_{k+1}\rho^2}{\beta}\right)}{\Lambda_{(k+1)}}\right]\mathbb{E}[\|\mathsf{h}_k\|^2]\right\} \\
&\quad + \sum_{k=0}^{K_m-1}\Xi^{(k+1)} + \sum_{k=0}^{K_m-1}\Gamma^{(k+1)}\mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] ,
\end{aligned} \tag{49}$$

where

$$\Xi^{(k+1)} = \tilde{\xi}^{(k+1)} + \frac{\gamma_{k+1}^3\,\mathrm{L}_V\,\rho^2\,\mathrm{L}_{\mathbf{s}}^2}{\Lambda_{(k+1)}}\tilde{\epsilon}^{(k+1)}$$

and

$$\Gamma^{(k+1)} = \left((1-\rho)^2\gamma_{k+1}^2\,\mathrm{L}_V - \frac{\gamma_{k+1}(1-\rho)^2}{2}\right) + \frac{\gamma_{k+1}^3\,\mathrm{L}_V\,\rho^2\,\mathrm{L}_{\mathbf{s}}^2(1-\rho)^2\left(2\gamma_{k+1}+\frac{1}{\beta}\right)}{\Lambda_{(k+1)}} .$$

Furthermore, given the values set for c_1, α, $\overline{L}$, γ_{k+1}, β and ρ, then

$$\begin{aligned}
&\gamma_{k+1}\rho^2\,\mathrm{L}_V + \frac{\rho^2\gamma_{k+1}\,\mathrm{L}_V\,\mathrm{L}_{\mathbf{s}}^2\left(2\gamma_{k+1}^2\rho^2 + \frac{\gamma_{k+1}\rho^2}{\beta}\right)}{\frac{1}{n} - \gamma_{k+1}\beta - \gamma_{k+1}^2\rho^2\,\mathrm{L}_{\mathbf{s}}^2} \\
\leq & \frac{1}{k\alpha^2c_1^2\overline{L}n^{4/3}} + \frac{\overline{L}(k\alpha^2c_1^2n^{4/3})^{-1}\left(\frac{2}{k^2\alpha^2c_1^2\overline{L}^2n^{4/3}} + \frac{1}{k\alpha c_1^2\overline{L}^2n^{1/3}}\right)}{\frac{1}{n} - \frac{1}{k\alpha n} - \frac{1}{k^2\alpha^2c_1^2n^{4/3}}} \\
= & \frac{1}{k\alpha^2c_1^2\overline{L}n^{4/3}} + \frac{\overline{L}\left(\frac{2}{k^2\alpha^2c_1^2\overline{L}^2n^{4/3}} + \frac{1}{k\alpha c_1^2\overline{L}^2n^{1/3}}\right)}{(k\alpha c_1 n^{1/3})(k\alpha-1)c_1 - 1} \\
\overset{(a)}{\leq} & \frac{1}{k\alpha^2c_1^2\overline{L}n^{4/3}} + \frac{\frac{1}{k\alpha c_1^2\overline{L}n^{1/3}}\left(\frac{2}{k\alpha n}+1\right)}{2(\alpha c_1 n^{1/3}) - 1} \leq \frac{1}{k^2\alpha c_1^2\overline{L}n^{4/3}} + \frac{1}{4k\alpha^2c_1^3\overline{L}n^{2/3}} \leq \frac{3/4}{\alpha c_1^2\overline{L}n^{2/3}} ,
\end{aligned} \tag{50}$$

where (a) is due to $c_1(k\alpha - 1) \geq c_1(\alpha - 1) \geq 2$ and $k\alpha c_1 n^{1/3} \geq 1$. Note also that

$$-(\upsilon_{\min}\rho + \upsilon_{\max}^2) \leq -\rho\upsilon_{\min} = -\frac{1}{\alpha c_1^2 \overline{L} n^{2/3}} \ ,$$

which yields that

$$\left[-(\upsilon_{\min}\rho + \upsilon_{\max}^2) + \gamma_{k+1}\rho^2 \, \mathrm{L}_V + \frac{\rho^2 \gamma_{k+1} \, \mathrm{L}_V \, \mathrm{L}_{\mathbf{s}}^2 \left(2\gamma_{k+1}^2 \rho^2 + \frac{\gamma_{k+1}\rho^2}{\beta} \right)}{\Lambda_{(k+1)}} \right] \leq -\frac{1/4}{\alpha c_1^2 \overline{L} n^{2/3}} \ .$$

Using the Lemma 2, we know that $\upsilon_{\max}^2 \|\nabla V(\hat{\boldsymbol{s}}^{(k)})\|^2 \leq \|\hat{\boldsymbol{s}}^{(k)} - \overline{\boldsymbol{s}}^{(k)}\|^2$ and using (50) on (49) yields:

$$\begin{aligned} \upsilon_{\max}^2 \sum_{k=0}^{K_m - 1} \gamma_{k+1} \mathbb{E}[\|\nabla V(\hat{\boldsymbol{s}}^{(k)})\|^2] \leq & \frac{4\alpha \overline{L} n^{2/3}}{\upsilon_{\min}^2} [V(\hat{\mathbf{s}}^{(0)}) - V(\hat{\mathbf{s}}^{(K_m)})] \\ & + \frac{4\alpha \overline{L} n^{2/3}}{\upsilon_{\min}^2} \sum_{k=0}^{K_m - 1} \Xi^{(k+1)} + \sum_{k=0}^{K_m - 1} \Gamma^{(k+1)} \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] \ , \end{aligned}$$

proving the bound on the second order moment of the gradient of the Lyapunov function:

$$\begin{aligned} \sum_{k=0}^{K_m - 1} \gamma_{k+1} \mathbb{E}[\|\nabla V(\hat{\boldsymbol{s}}^{(k)})\|^2] \leq & \frac{4\alpha \overline{L} n^{2/3}}{\upsilon_{\min}^2 \upsilon_{\max}^2} [V(\hat{\mathbf{s}}^{(0)}) - V(\hat{\mathbf{s}}^{(K_m)})] \\ & + \frac{4\alpha \overline{L} n^{2/3}}{\upsilon_{\min}^2 \upsilon_{\max}^2} \sum_{k=0}^{K_m - 1} \Xi^{(k+1)} + \sum_{k=0}^{K_m - 1} \Gamma^{(k+1)} \mathbb{E}[\|\hat{\boldsymbol{s}}^{(k)} - \tilde{S}^{(k)}\|^2] \ . \end{aligned}$$

□