APPENDIX II
PROOFS FROM SECTION V

A. *Proof for Lemma 1*

Proof. Let $\mathbf{x}^*$ be such that $\mathbf{v}(\mathbf{x}^*) = 0$. Then, for all $t \geq 0$,

$$\mathbf{x}_{t+1} - \mathbf{x}^* = \mathbf{x}_t - \mathbf{x}^* + \beta_t(\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*) + \mathbf{M}_{t+1}).$$

Now, using the standard expansion, we get

$$\begin{aligned}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2 &= \|\mathbf{x}_t - \mathbf{x}^* + \beta_t(\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*) + \mathbf{M}_{t+1})\|^2 \\ &= \|\mathbf{x}_t - \mathbf{x}^*\|^2 + \beta_t^2\|\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*) + \mathbf{M}_{t+1}\|^2 \qquad (6a)\\ &\quad + 2\beta_t\langle \mathbf{x}_t - \mathbf{x}^*, \mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*) + \mathbf{M}_{t+1}\rangle. \qquad (6b)\end{aligned}$$

For the term (6a), we note that

$$\|\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*) + \mathbf{M}_{t+1}\|^2 = \|\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*)\|^2 + \|\mathbf{M}_{t+1}\|^2 + 2\langle \mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*), \mathbf{M}_{t+1}\rangle.$$

For the term (6b), since $-\mathbf{v}(\cdot)$ is λ-co-coercive, we have

$$\langle \mathbf{x}_t - \mathbf{x}^*, \mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*)\rangle \leq -\lambda\|\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*)\|^2.$$

Hence,

$$2\beta_t\langle \mathbf{x}_t - \mathbf{x}^*, \mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*) + \mathbf{M}_{t+1}\rangle \leq -2\lambda\beta_t\|\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*)\|^2 + 2\beta_t\langle \mathbf{x}_t - \mathbf{x}^*, \mathbf{M}_{t+1}\rangle.$$

Returning to (6), we get

$$\begin{aligned}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2 &\leq \|\mathbf{x}_t - \mathbf{x}^*\|^2 + \beta_t^2\|\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*)\|^2 + \beta_t^2\|\mathbf{M}_{t+1}\|^2 \\ &\quad + 2\beta_t^2\langle \mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*), \mathbf{M}_{t+1}\rangle - 2\lambda\beta_t\|\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*)\|^2 \\ &\quad + 2\beta_t\langle \mathbf{x}_t - \mathbf{x}^*, \mathbf{M}_{t+1}\rangle \\ &= \|\mathbf{x}_t - \mathbf{x}^*\|^2 - \beta_t(2\lambda - \beta_t)\|\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*)\|^2 + \beta_t^2\|\mathbf{M}_{t+1}\|^2 \\ &\quad + 2\beta_t\langle \mathbf{x}_t - \mathbf{x}^*, \mathbf{M}_{t+1}\rangle + 2\beta_t^2\langle \mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*), \mathbf{M}_{t+1}\rangle.\end{aligned}$$

Since $T_0 \geq 1/\lambda$, $\beta_t \leq \lambda$ for all t, then $2\lambda - \beta_t \geq \lambda$, and so

$$\begin{aligned}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2 &\leq \|\mathbf{x}_t - \mathbf{x}^*\|^2 - \lambda\beta_t\|\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*)\|^2 + \beta_t^2\|\mathbf{M}_{t+1}\|^2 \\ &\quad + 2\beta_t\langle \mathbf{x}_t - \mathbf{x}^*, \mathbf{M}_{t+1}\rangle + 2\beta_t^2\langle \mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*), \mathbf{M}_{t+1}\rangle.\end{aligned}$$

On taking expectation conditioned on $\mathcal{F}_t$, we have

$$\mathbb{E}[\langle \mathbf{x}_t - \mathbf{x}^*, \mathbf{M}_{t+1}\rangle \mid \mathcal{F}_t] = 0, \qquad \mathbb{E}[\langle \mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*), \mathbf{M}_{t+1}\rangle \mid \mathcal{F}_t] = 0,$$

since $\{\mathbf{M}_{t+1}\}$ is a martingale difference sequence with respect to $\{\mathcal{F}_t\}$, and both $\mathbf{x}_t$ and $\mathbf{v}(\mathbf{x}_t)$ are $\mathcal{F}_t$-measurable.

Therefore,

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2 \mid \mathcal{F}_t\right] &\leq \|\mathbf{x}_t - \mathbf{x}^*\|^2 + \beta_t^2\mathbb{E}\left[\|\mathbf{M}_{t+1}\|^2 \mid \mathcal{F}_t\right] - \lambda\beta_t\|\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{x}^*)\|^2 \\ &\leq \|\mathbf{x}_t - \mathbf{x}^*\|^2 + \beta_t^2\sigma^2(1 + \|\mathbf{x}_t\|^2) - \lambda\beta_t\|\mathbf{v}(\mathbf{x}_t)\|^2 \\ &\leq \|\mathbf{x}_t - \mathbf{x}^*\|^2 + 2\beta_t^2\sigma^2\|\mathbf{x}_t - \mathbf{x}^*\|^2 + \beta_t^2\sigma^2(1 + 2\|\mathbf{x}^*\|^2) - \lambda\beta_t\|\mathbf{v}(\mathbf{x}_t)\|^2,\end{aligned}$$

where we used $\mathbf{v}(\mathbf{x}^*) = 0$ for the second inequality. This completes the proof for Lemma 1.

□

B. *Proof for Lemma 2*

Proof. Taking expectation and dropping the negative residual term in Lemma 1 gives, for all $t \geq 0$,

$$\mathbb{E}\left[\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2\right] \leq \mathbb{E}\left[\|\mathbf{x}_t - \mathbf{x}^*\|^2\right] + \beta_t^2\sigma^2(1 + 2\|\mathbf{x}^*\|^2) + 2\beta_t^2\sigma^2\mathbb{E}\left[\|\mathbf{x}_t - \mathbf{x}^*\|^2\right].$$

Summing from $i = 0$ to $t - 1$, we get

$$\mathbb{E}\left[\|\mathbf{x}_t - \mathbf{x}^*\|^2\right] \leq \mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{x}^*\|^2\right] + \sigma^2\left(1 + 2\|\mathbf{x}^*\|^2\right)\sum_{i=0}^{t-1}\beta_i^2 + 2\sigma^2\sum_{i=0}^{t-1}\beta_i^2\mathbb{E}\left[\|\mathbf{x}_i - \mathbf{x}^*\|^2\right]$$

Set

$$B = \mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{x}^*\|^2\right] + \sigma^2\big(1 + 2\|\mathbf{x}^*\|^2\big)\sum_{i=0}^{\infty}\beta_i^2.$$

Then

$$\mathbb{E}\left[\|\mathbf{x}_t - \mathbf{x}^*\|^2\right] \le B + 2\sigma^2\sum_{i=0}^{t-1}\beta_i^2\mathbb{E}\left[\|\mathbf{x}_i - \mathbf{x}^*\|^2\right].$$

By the discrete Gronwall inequality it follows that

$$\mathbb{E}\left[\|\mathbf{x}_t - \mathbf{x}^*\|^2\right] \le B\exp\Big(2\sigma^2\sum_{i=0}^{t-1}\beta_i^2\Big),$$

Since $\sum_{i=0}^{\infty}\beta_i^2 < \infty$, the right-hand side of the above bound is itself bounded by some constant $\Gamma_1 := B\exp\Big(2\sigma^2\sum_{i=0}^{\infty}\beta_i^2\Big) > 0$. Hence there exists $\Gamma_1 > 0$ such that

$$\sup_{t\ge 0}\mathbb{E}\left[\|\mathbf{x}_t - \mathbf{x}^*\|^2\right] \le \Gamma_1.$$

□

C. *Proof for Lemma 3*

Proof. Taking expectation in Lemma 1, and rearranging terms, we get

$$\lambda\beta_t\mathbb{E}\left[\|\mathbf{v}(\mathbf{x}_t)\|^2\right] \le \mathbb{E}\left[\|\mathbf{x}_t - \mathbf{x}^*\|^2\right] - \mathbb{E}\left[\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2\right] + \beta_t^2\sigma^2\big(1 + 2\|\mathbf{x}^*\|^2\big) + 2\beta_t^2\sigma^2\mathbb{E}\left[\|\mathbf{x}_t - \mathbf{x}^*\|^2\right].$$

Summing from $i = 0$ to t gives

$$\lambda\sum_{i=0}^{t}\beta_i\mathbb{E}\left[\|\mathbf{v}(\mathbf{x}_i)\|^2\right] \le \mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{x}^*\|^2\right] - \mathbb{E}\left[\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2\right] + \sigma^2\big(1 + 2\|\mathbf{x}^*\|^2\big)\sum_{i=0}^{t}\beta_i^2 + 2\sigma^2\sum_{i=0}^{t}\beta_i^2\mathbb{E}\left[\|\mathbf{x}_i - \mathbf{x}^*\|^2\right].$$

Dropping the nonnegative term $\mathbb{E}[\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2]$, we obtain

$$\lambda\sum_{i=0}^{t}\beta_i\mathbb{E}\left[\|\mathbf{v}(\mathbf{x}_i)\|^2\right] \le \mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{x}^*\|^2\right] + \sigma^2\big(1 + 2\|\mathbf{x}^*\|^2\big)\sum_{i=0}^{t}\beta_i^2 + 2\sigma^2\sum_{i=0}^{t}\beta_i^2\mathbb{E}\left[\|\mathbf{x}_i - \mathbf{x}^*\|^2\right].$$

Now using

$$\mathbb{E}\left[\|\mathbf{x}_i - \mathbf{x}^*\|^2\right] \le \Gamma_1,$$

we get

$$\lambda\sum_{i=0}^{t}\beta_i\mathbb{E}\left[\|\mathbf{v}(\mathbf{x}_i)\|^2\right] \le \mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{x}^*\|^2\right] + \sigma^2\big(1 + 2\|\mathbf{x}^*\|^2\big)\sum_{i=0}^{t}\beta_i^2 + 2\sigma^2\Gamma_1\sum_{i=0}^{t}\beta_i^2.$$

Since $\sum_{i=0}^{\infty}\beta_i^2 < \infty$, we have

$$\lambda\sum_{i=0}^{t}\beta_i\mathbb{E}\left[\|\mathbf{v}(\mathbf{x}_i)\|^2\right] \le \mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{x}^*\|^2\right] + \Big(\sigma^2\big(1 + 2\|\mathbf{x}^*\|^2\big) + 2\sigma^2\Gamma_1\Big)\sum_{i=0}^{\infty}\beta_i^2.$$

Define

$$\Gamma_2 := \frac{1}{\lambda}\left(\mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{x}^*\|^2\right] + \Big(\sigma^2\big(1 + 2\|\mathbf{x}^*\|^2\big) + 2\sigma^2\Gamma_1\Big)\sum_{i=0}^{\infty}\beta_i^2\right).$$

Then, for all $t \ge 0$,

$$\sum_{i=0}^{t}\beta_i\mathbb{E}\left[\|\mathbf{v}(\mathbf{x}_i)\|^2\right] \le \Gamma_2.$$

This completes the proof. □

D. Proof for Lemma 4

Proof. We first note that co-coercivity of $-\mathbf{v}$ implies that $\mathbf{v}$ is $(1/\lambda)$-Lipschitz. Indeed, for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^{Nd}$,

$$\lambda \|\mathbf{v}(\mathbf{x}) - \mathbf{v}(\mathbf{x}')\|^2 \le -\langle \mathbf{x} - \mathbf{x}', \mathbf{v}(\mathbf{x}) - \mathbf{v}(\mathbf{x}') \rangle \le \|\mathbf{x} - \mathbf{x}'\| \, \|\mathbf{v}(\mathbf{x}) - \mathbf{v}(\mathbf{x}')\|,$$

and hence

$$\|\mathbf{v}(\mathbf{x}) - \mathbf{v}(\mathbf{x}')\| \le \frac{1}{\lambda} \|\mathbf{x} - \mathbf{x}'\|.$$

(a) Bound for $\mathbb{E}\|\mathbf{U}_k\|^2$.

Expanding and taking conditional expectation yields

$$\begin{aligned} \|\mathbf{U}_{k+1}\|^2 &= \|(1-\beta_k)\mathbf{U}_k + \beta_k \mathbf{M}_{k+1}\|_2^2 \\ &= (1-\beta_k)^2 \|\mathbf{U}_k\|^2 + 2(1-\beta_k)\beta_k \langle \mathbf{U}_k, \mathbf{M}_{k+1} \rangle + \beta_k^2 \|\mathbf{M}_{k+1}\|^2, \\ \mathbb{E}[\|\mathbf{U}_{k+1}\|^2 \mid \mathcal{F}_k] &= (1-\beta_k)^2 \|\mathbf{U}_k\|^2 + \beta_k^2 \, \mathbb{E}[\|\mathbf{M}_{k+1}\|^2 \mid \mathcal{F}_k], \end{aligned}$$

since $\mathbb{E}[\mathbf{M}_{k+1} \mid \mathcal{F}_k] = 0$ implies

$$\mathbb{E}[\langle \mathbf{U}_k, \mathbf{M}_{k+1} \rangle \mid \mathcal{F}_k] = \langle \mathbf{U}_k, \mathbb{E}[\mathbf{M}_{k+1} \mid \mathcal{F}_k] \rangle = 0.$$

Taking total expectation and using

$$\mathbb{E}\|\mathbf{M}_{k+1}\|^2 \le \sigma^2 \big(1 + \mathbb{E}\|\mathbf{x}_k\|^2\big) \le \sigma^2 \big(1 + 2\mathbb{E}\|\mathbf{x}_k - \mathbf{x}^*\|^2 + 2\|\mathbf{x}^*\|^2\big) \le \sigma^2 \big(1 + 2\Gamma_1 + 2\|\mathbf{x}^*\|^2\big) =: A,$$

we get

$$\mathbb{E}\|\mathbf{U}_{k+1}\|^2 \le (1-\beta_k)^2 \mathbb{E}\|\mathbf{U}_k\|^2 + A\beta_k^2.$$

Since $\beta_k \in (0,1]$, we have $(1-\beta_k)^2 \le (1-\beta_k)$, hence

$$\mathbb{E}\|\mathbf{U}_{k+1}\|^2 \le (1-\beta_k) \mathbb{E}\|\mathbf{U}_k\|^2 + A\beta_k^2. \tag{7}$$

Now we know,

$$\beta_k = \frac{1}{(k+T_0)^{\mathfrak{b}}}, \qquad 0.5 < \mathfrak{b} < 1.$$

Hence

$$\beta_k - \beta_{k+1} = \frac{1}{(k+T_0)^{\mathfrak{b}}} - \frac{1}{(k+1+T_0)^{\mathfrak{b}}} = \frac{1}{(k+T_0)^{\mathfrak{b}}} \Big(1 - \Big(1 + \frac{1}{k+T_0}\Big)^{-\mathfrak{b}}\Big).$$

Using the inequality $1 - (1+t)^{-\mathfrak{b}} \le \mathfrak{b} t$ for $t \ge 0$, we obtain

$$\beta_k - \beta_{k+1} \le \frac{1}{(k+T_0)^{\mathfrak{b}}} \cdot \frac{\mathfrak{b}}{k+T_0} = \frac{\mathfrak{b}}{(k+T_0)^{\mathfrak{b}+1}}.$$

Also,

$$\beta_k^2 = \frac{1}{(k+T_0)^{2\mathfrak{b}}}.$$

Therefore

$$\frac{\beta_k - \beta_{k+1}}{\beta_k^2} \le \frac{\mathfrak{b}}{(k+T_0)^{1-\mathfrak{b}}} \to 0.$$

Since $\mathfrak{b} < 1$, there exists k_0 such that

$$\beta_k - \beta_{k+1} \le \frac{1}{2}\beta_k^2, \qquad \forall k \ge k_0.$$

Now define

$$\Gamma_3 := \max\Big\{2A, \ \max_{0 \le k \le k_0} \frac{\mathbb{E}\|\mathbf{U}_k\|^2}{\beta_k}\Big\}.$$

Then, by definition,

$$\mathbb{E}\|\mathbf{U}_k\|^2 \le \Gamma_3 \beta_k, \qquad \forall\, 0 \le k \le k_0.$$

Now fix $k \ge k_0$ and assume inductively that

$$\mathbb{E}\|\mathbf{U}_k\|^2 \le \Gamma_3 \beta_k.$$

Then by (7),

$$\mathbb{E}\|\mathbf{U}_{k+1}\|^2 \le (1-\beta_k)\Gamma_3 \beta_k + A\beta_k^2 = \Gamma_3 \beta_k - (\Gamma_3 - A)\beta_k^2.$$

Since $\Gamma_3 \geq 2A$, we have $\Gamma_3 - A \geq \Gamma_3/2$, and therefore

$$\mathbb{E}\|\mathbf{U}_{k+1}\|^2 \leq \Gamma_3\beta_k - \frac{\Gamma_3}{2}\beta_k^2.$$

Also, since $k \geq k_0$,

$$\beta_k - \beta_{k+1} \leq \frac{1}{2}\beta_k^2,$$

which implies

$$\Gamma_3(\beta_k - \beta_{k+1}) \leq \frac{\Gamma_3}{2}\beta_k^2.$$

Hence

$$\mathbb{E}\|\mathbf{U}_{k+1}\|^2 \leq \Gamma_3\beta_k - \Gamma_3(\beta_k - \beta_{k+1}) = \Gamma_3\beta_{k+1}.$$

Thus, by induction,

$$\mathbb{E}\|\mathbf{U}_k\|^2 \leq \Gamma_3\beta_k, \qquad \forall k \geq 0.$$

This completes the proof for part (a).

(b) Bound on the summation of $\beta_i\mathbb{E}[\|\mathbf{v}(\mathbf{z}_i)\|^2]$.

Using the Lipschitz property of $\mathbf{v}$,

$$\|\mathbf{v}(\mathbf{z}_t)\| \leq \|\mathbf{v}(\mathbf{x}_t)\| + \|\mathbf{v}(\mathbf{z}_t) - \mathbf{v}(\mathbf{x}_t)\| \leq \|\mathbf{v}(\mathbf{x}_t)\| + \frac{1}{\lambda}\|\mathbf{U}_t\|.$$

Hence

$$\|\mathbf{v}(\mathbf{z}_t)\|^2 \leq 2\|\mathbf{v}(\mathbf{x}_t)\|^2 + \frac{2}{\lambda^2}\|\mathbf{U}_t\|^2.$$

Multiplying by β_t, taking expectation, and summing from $i = 0$ to t, we get

$$\begin{aligned}\sum_{i=0}^{t}\beta_i\mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_i)\|^2\right] &\leq 2\sum_{i=0}^{t}\beta_i\mathbb{E}\left[\|\mathbf{v}(\mathbf{x}_i)\|^2\right] + \frac{2}{\lambda^2}\sum_{i=0}^{t}\beta_i\mathbb{E}\left[\|\mathbf{U}_i\|^2\right] \\ &\leq 2\Gamma_2 + \frac{2\Gamma_3}{\lambda^2}\sum_{i=0}^{\infty}\beta_i^2.\end{aligned}$$

Define

$$\Gamma_4 := 2\Gamma_2 + \frac{2\Gamma_3}{\lambda^2}\sum_{i=0}^{\infty}\beta_i^2.$$

Then

$$\sum_{i=0}^{t}\beta_i\mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_i)\|^2\right] \leq \Gamma_4, \qquad \forall t \geq 0.$$

(c) Bound on $\mathbb{E}[\|\mathbf{v}(\mathbf{z}_i)\|^2]$ in terms of $\mathbb{E}[\|\mathbf{v}(\mathbf{z}_t)\|^2]$.

Subtracting the recursion for $\mathbf{U}_t$ from the recursion

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \beta_t\big(\mathbf{v}(\mathbf{x}_t) + \mathbf{M}_{t+1}\big)$$

gives

$$\begin{aligned}\mathbf{z}_{t+1} &= \mathbf{x}_{t+1} - \mathbf{U}_{t+1} \\ &= \mathbf{x}_t + \beta_t\big(\mathbf{v}(\mathbf{x}_t) + \mathbf{M}_{t+1}\big) - \big((1-\beta_t)\mathbf{U}_t + \beta_t\mathbf{M}_{t+1}\big) \\ &= \mathbf{z}_t + \beta_t\big(\mathbf{v}(\mathbf{x}_t) + \mathbf{U}_t\big) \\ &= \mathbf{z}_t + \beta_t\big(\mathbf{v}(\mathbf{z}_t) + \mathbf{e}_t\big),\end{aligned}$$

where

$$\mathbf{e}_t := \mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{z}_t) + \mathbf{U}_t.$$

By the $(1/\lambda)$-Lipschitz property of $\mathbf{v}$,

$$\|\mathbf{e}_t\| \leq \|\mathbf{v}(\mathbf{x}_t) - \mathbf{v}(\mathbf{z}_t)\| + \|\mathbf{U}_t\| \leq \Big(1 + \frac{1}{\lambda}\Big)\|\mathbf{U}_t\|.$$

Hence, using part (a),

$$\mathbb{E}\left[\|\mathbf{e}_t\|^2\right] \le \Big(1+\frac{1}{\lambda}\Big)^2 \mathbb{E}\left[\|\mathbf{U}_t\|^2\right] \le \Big(1+\frac{1}{\lambda}\Big)^2 \Gamma_3 \beta_t.$$

Since

$$\mathbf{z}_{t+1}-\mathbf{z}_t = \beta_t\big(\mathbf{v}(\mathbf{z}_t)+\mathbf{e}_t\big),$$

we have

$$\mathbf{v}(\mathbf{z}_t) = \frac{\mathbf{z}_{t+1}-\mathbf{z}_t-\beta_t\mathbf{e}_t}{\beta_t}.$$

Therefore

$$\begin{aligned}\|\mathbf{v}(\mathbf{z}_{t+1})\|^2 &= \|\mathbf{v}(\mathbf{z}_t)\|^2 + \|\mathbf{v}(\mathbf{z}_{t+1})-\mathbf{v}(\mathbf{z}_t)\|^2 + 2\langle \mathbf{v}(\mathbf{z}_{t+1})-\mathbf{v}(\mathbf{z}_t), \mathbf{v}(\mathbf{z}_t)\rangle \\ &= \|\mathbf{v}(\mathbf{z}_t)\|^2 + \|\mathbf{v}(\mathbf{z}_{t+1})-\mathbf{v}(\mathbf{z}_t)\|^2 \\ &\quad + \frac{2}{\beta_t}\Big\langle \mathbf{v}(\mathbf{z}_{t+1})-\mathbf{v}(\mathbf{z}_t), \mathbf{z}_{t+1}-\mathbf{z}_t-\beta_t\mathbf{e}_t\Big\rangle.\end{aligned}$$

Since $-\mathbf{v}$ is λ-co-coercive,

$$\langle \mathbf{z}_{t+1}-\mathbf{z}_t, \mathbf{v}(\mathbf{z}_{t+1})-\mathbf{v}(\mathbf{z}_t)\rangle \le -\lambda\|\mathbf{v}(\mathbf{z}_{t+1})-\mathbf{v}(\mathbf{z}_t)\|^2.$$

Substituting this above yields

$$\|\mathbf{v}(\mathbf{z}_{t+1})\|^2 \le \|\mathbf{v}(\mathbf{z}_t)\|^2 - \Big(\frac{2\lambda}{\beta_t}-1\Big)\|\mathbf{v}(\mathbf{z}_{t+1})-\mathbf{v}(\mathbf{z}_t)\|^2 - 2\langle \mathbf{v}(\mathbf{z}_{t+1})-\mathbf{v}(\mathbf{z}_t), \mathbf{e}_t\rangle.$$

Now note that for any $c>0$,

$$0 \le \Big\|\sqrt{c}\,u + \frac{1}{\sqrt{c}}\,w\Big\|^2 = c\|u\|^2 + 2\langle u, w\rangle + \frac{1}{c}\|w\|^2,$$

which implies

$$-c\|u\|^2 - 2\langle u, w\rangle \le \frac{1}{c}\|w\|^2.$$

Applying this with

$$c = \frac{2\lambda}{\beta_t}-1, \qquad u = \mathbf{v}(\mathbf{z}_{t+1})-\mathbf{v}(\mathbf{z}_t), \qquad w = \mathbf{e}_t,$$

we get

$$\|\mathbf{v}(\mathbf{z}_{t+1})\|^2 \le \|\mathbf{v}(\mathbf{z}_t)\|^2 + \frac{1}{c}\|\mathbf{e}_t\|^2.$$

Since $T_0 \ge 1/\lambda$, $\beta_t \le \lambda$ for all t, we have

$$c = \frac{2\lambda}{\beta_t}-1 \ge \frac{\lambda}{\beta_t},$$

and hence

$$\|\mathbf{v}(\mathbf{z}_{t+1})\|^2 \le \|\mathbf{v}(\mathbf{z}_t)\|^2 + \frac{\beta_t}{\lambda}\|\mathbf{e}_t\|^2.$$

Taking expectation and using the bound on $\mathbb{E}[\|\mathbf{e}_t\|^2]$, we obtain

$$\mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_{t+1})\|^2\right] \le \mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_t)\|^2\right] + \frac{(1+\lambda^{-1})^2\Gamma_3}{\lambda}\beta_t^2.$$

Define

$$\Gamma_5 := \frac{(1+\lambda^{-1})^2\Gamma_3}{\lambda}.$$

Then

$$\mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_{t+1})\|^2\right] \le \mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_t)\|^2\right] + \Gamma_5\beta_t^2.$$

Iterating this inequality from $j=i$ to $t-1$ gives

$$\mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_i)\|^2\right] \ge \mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_t)\|^2\right] - \Gamma_5\sum_{j=i}^{t-1}\beta_j^2,$$

which proves part (c). □

E. Proof for Lemma 5

Proof. For each $i \leq t-1$, Lemma 4(c) implies

$$\mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_t)\|^2\right] \leq \mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_i)\|^2\right] + \Gamma_5 \sum_{j=i}^{t-1} \beta_j^2.$$

Multiplying by β_i and summing from $i = 0$ to $t-1$, we get

$$\mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_t)\|^2\right] \sum_{i=0}^{t-1} \beta_i \leq \sum_{i=0}^{t-1} \beta_i \mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_i)\|^2\right] + \Gamma_5 \sum_{i=0}^{t-1} \beta_i \sum_{j=i}^{t-1} \beta_j^2.$$

Adding the nonnegative term $\beta_t \mathbb{E}[\|\mathbf{v}(\mathbf{z}_t)\|^2]$ to both sides and using Lemma 4(b) yields

$$\mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_t)\|^2\right] \sum_{i=0}^{t} \beta_i \leq \Gamma_4 + \Gamma_5 \sum_{i=0}^{t-1} \beta_i \sum_{j=i}^{t-1} \beta_j^2. \tag{8}$$

Hence

$$\mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_t)\|^2\right] \leq \frac{\Gamma_4}{\sum_{i=0}^{t} \beta_i} + \Gamma_5 \frac{\sum_{i=0}^{t-1} \beta_i \sum_{j=i}^{t-1} \beta_j^2}{\sum_{i=0}^{t} \beta_i}.$$

Since $T_0 \geq 1$, for every $i \geq 0$,

$$i + 1 \leq i + T_0 \leq T_0(i+1),$$

and hence

$$T_0^{-\mathfrak{b}}(i+1)^{-\mathfrak{b}} \leq \beta_i \leq (i+1)^{-\mathfrak{b}}.$$

Therefore,

$$\sum_{i=0}^{t} \beta_i \geq (t+1)\beta_t = (t+1)(t+T_0)^{-\mathfrak{b}} \geq T_0^{-\mathfrak{b}}(t+1)^{1-\mathfrak{b}},$$

so

$$\frac{\Gamma_4}{\sum_{i=0}^{t} \beta_i} \leq \Gamma_4 T_0^{\mathfrak{b}} (t+1)^{-(1-\mathfrak{b})}.$$

Also,

$$\sum_{i=0}^{j} \beta_i \leq \sum_{i=0}^{j} (i+1)^{-\mathfrak{b}} \leq 1 + \int_1^{j+1} x^{-\mathfrak{b}}\, dx \leq \frac{(j+1)^{1-\mathfrak{b}}}{1-\mathfrak{b}}.$$

Hence

$$\sum_{i=0}^{t-1} \beta_i \sum_{j=i}^{t-1} \beta_j^2 = \sum_{j=0}^{t-1} \Big(\sum_{i=0}^{j} \beta_i \Big) \beta_j^2 \leq \frac{1}{1-\mathfrak{b}} \sum_{j=0}^{t-1} (j+1)^{1-3\mathfrak{b}}.$$

If $b \in (\frac{1}{2}, \frac{2}{3})$, then

$$\sum_{j=0}^{t-1} (j+1)^{1-3\mathfrak{b}} \leq 1 + \int_1^{t+1} x^{1-3\mathfrak{b}}\, dx \leq \frac{(t+1)^{2-3\mathfrak{b}}}{2-3\mathfrak{b}}.$$

Therefore,

$$\Gamma_5 \frac{\sum_{i=0}^{t-1} \beta_i \sum_{j=i}^{t-1} \beta_j^2}{\sum_{i=0}^{t} \beta_i} \leq \frac{\Gamma_5 T_0^{\mathfrak{b}}}{(1-\mathfrak{b})(2-3\mathfrak{b})} (t+1)^{-(2\mathfrak{b}-1)}.$$

Combining this with the bound on the first term, and using

$$(t+1)^{-(1-\mathfrak{b})} \leq (t+1)^{-(2\mathfrak{b}-1)},$$

we get

$$\mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_t)\|^2\right] \leq \left(\Gamma_4 T_0^{\mathfrak{b}} + \frac{\Gamma_5 T_0^{\mathfrak{b}}}{(1-\mathfrak{b})(2-3\mathfrak{b})} \right) (t+1)^{-(2\mathfrak{b}-1)}.$$

If $\mathfrak{b} = \frac{2}{3}$, then

$$\sum_{j=0}^{t-1} (j+1)^{-1} \leq 1 + \int_1^{t+1} \frac{dx}{x} \leq 2\log(t+2).$$

Therefore,

$$\Gamma_5 \frac{\sum_{i=0}^{t-1} \beta_i \sum_{j=i}^{t-1} \beta_j^2}{\sum_{i=0}^{t} \beta_i} \leq 6\Gamma_5 T_0^{2/3} \frac{\log(t+2)}{(t+1)^{1/3}}.$$

Combining this with the bound on the first term, and using

$$1 \leq \frac{\log(t+2)}{\log 2},$$

we get

$$\mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_t)\|^2\right] \leq \left(\frac{\Gamma_4 T_0^{2/3}}{\log 2} + 6\Gamma_5 T_0^{2/3} \right) \frac{\log(t+2)}{(t+1)^{1/3}}.$$

If $\mathfrak{b} \in (\frac{2}{3}, 1)$, then

$$\sum_{j=0}^{t-1} (j+1)^{1-3\mathfrak{b}} \leq \sum_{j=0}^{\infty} (j+1)^{1-3\mathfrak{b}}.$$

Therefore,

$$\Gamma_5 \frac{\sum_{i=0}^{t-1} \beta_i \sum_{j=i}^{t-1} \beta_j^2}{\sum_{i=0}^{t} \beta_i} \leq \\ \frac{\Gamma_5 T_0^{\mathfrak{b}}}{1-\mathfrak{b}} \left(\sum_{j=0}^{\infty} (j+1)^{1-3\mathfrak{b}} \right) (t+1)^{-(1-\mathfrak{b})}.$$

Combining this with the bound on the first term, we get

$$\mathbb{E}\left[\|\mathbf{v}(\mathbf{z}_t)\|^2\right] \leq \\ \left(\Gamma_4 T_0^{\mathfrak{b}} + \frac{\Gamma_5 T_0^{\mathfrak{b}}}{1-\mathfrak{b}} \sum_{j=0}^{\infty} (j+1)^{1-3\mathfrak{b}} \right) (t+1)^{-(1-\mathfrak{b})}.$$

$\square$