A Detailed Proofs in Sections 2 and 3

A.1 Proof of Lemma 1

Proof. We only need to show

$$\mathbb{E}\left(\mathbf{v}^\top \boldsymbol{Y}\right)^4 = 3 + (\psi - 3)\sum_{i=1}^d v_i^4. \tag{A.1}$$

Note due to the following well-known expansion [9]

$$\left(\sum x_i\right)^4 = \sum x_i^4 + 4\sum x_i^3 x_j + 6\sum x_i^2 x_j^2 + 12\sum x_{i_1}^2 x_{i_2} x_{i_3} + 24\sum x_{i_1} x_{i_2} x_{i_3} x_{i_4}.$$

where the summations above iterate through all monomial terms. Plugging in $x_i = v_i Y_i$ and taking expectations, we conclude that under Assumption 1

$$\begin{aligned}\mathbb{E}\left(\mathbf{v}^\top \boldsymbol{Y}\right)^4 &= \sum_{i=1}^d v_i^4 \mathbb{E}\left(Y_i^4\right) + 6\sum_{1\le i<j\le d} v_i^2 v_j^2 \mathbb{E}\left(Y_i^2\right)\mathbb{E}\left(Y_j^2\right)\\ &= \psi\sum_{i=1}^d v_i^4 + 6\sum_{1\le i<j\le d} v_i^2 v_j^2.\end{aligned} \tag{A.2}$$

Note that from the constraint of our optimization problem Eq. (2.2), we have

$$1 = \|\mathbf{v}\|^4 = \left(\sum_{i=1}^d v_i^2\right)^2 = \sum_{i=1}^d v_i^4 + 2\sum_{1\le i<j\le d} v_i^2 v_j^2. \tag{A.3}$$

Combining both Eqs. (A.2) and (A.3) we conclude Eq. (A.1) and hence the lemma. ■

A.2 Proof of Proposition 1

Proof. Let $\mathcal{F}_n = \sigma(\mathbf{u}^{(n')} : n' \le n)$ be the σ-field filtration generated by the iteration $\mathbf{u}^{(n)}$, viewed as a stochastic process. From the recursion equation in Eq. (2.3) we have a Markov transition kernel $p(\mathbf{u}, \mathcal{S})$ such that for each Borel set $\mathcal{A} \subseteq \mathcal{S}^{d-1}$

$$\mathbb{P}\left(\mathbf{u}^{(n)} \in \mathcal{A} \mid \mathcal{F}_{n-1}\right) = p(\mathbf{u}^{(n-1)}, \mathcal{A}).$$

Therefore it is a time-homogeneous Markov chain. The strong Markov property holds directly from Markov property, see [16] as a reference. This proves Proposition 1.

■

A.3 Proof of Theorem 1

We first use the standard one-step analysis and conclude the following proposition, whose proof is deferred to Subsection C.1.

Proposition 3. For brevity let $\mathbf{v} = \mathbf{v}^{(0)}$ and $\boldsymbol{Y} = \boldsymbol{Y}^{(1)}$, separately. Under Assumption 1, when

$$B^2\beta \le 2/3, \tag{A.4}$$

for each $k = 1, 2, \ldots, d$ and $n \ge 0$ we have the following:

(i) There exists a random variable R_k that depends solely on $\mathbf{v}, \boldsymbol{Y}$ with $|R_k| \le 9B^4\beta^2$ almost surely, such that the increment $v_k^{(1)} - v_k^{(0)}$ can be represented as

$$v_k^{(1)} - v_k^{(0)} = \beta\left(\left(\mathbf{v}^\top \mathbf{Y}\right)^3 Y_k - v_k\left(\mathbf{v}^\top \mathbf{Y}\right)^4\right) + R_k; \tag{A.5}$$

(ii) The increment of v_k on coordinate k has the following bound

$$\left|v_k^{(1)} - v_k^{(0)}\right| \le 8B^2\beta; \tag{A.6}$$

(iii) There exists a deterministic function $E_k(\mathbf{v})$ with $\sup_{\mathbf{v}\in\mathcal{S}^{d-1}} |E_k(\mathbf{v})| \le 9B^4\beta^2$, such that the conditional expectation of the increment $v_k^{(1)} - v_k^{(0)}$ is

$$\mathbb{E}\left[v_k^{(1)} - v_k^{(0)} \mid \mathbf{v}^{(0)} = \mathbf{v}\right] = \beta\left|\psi - 3\right| v_k\left(v_k^2 - \sum_{i=1}^d v_i^4\right) + E_k(\mathbf{v}). \tag{A.7}$$

In Proposition 3, (i) characterizes the relationship between the increment on v_k and the online sample, and (ii) bounds such increment. From (iii) we can compute the infinitesimal mean and variance for SGD for tensor method and conclude that as the stepsize $\beta \to 0^+$, the iterates generated by Eq. (2.3), under the time scaling that speeds up the algorithm by a factor β^{-1}, can be globally approximated by the solution to the following ODE system in Eq. (3.2) as

$$\frac{\mathrm{d}V_k}{\mathrm{d}t} = |\psi - 3|\, V_k \left(V_k^2 - \sum_{i=1}^{d} V_i^4 \right), \qquad k = 1, \ldots, d.$$

To characterize such approximation we use theory of weak convergence to diffusions [17, 40]. We remind the readers of the definition of weak convergence $Z^\beta \Rightarrow Z$ in stochastic processes: for any $0 \le t_1 < t_2 < \cdots < t_n$ the following convergence in distribution occurs as $\beta \to 0^+$

$$\left(Z^\beta(t_1), Z^\beta(t_2), \ldots, Z^\beta(t_n)\right) \xrightarrow{d} \left(Z(t_1), Z(t_2), \ldots, Z(t_n)\right).$$

To highlight the dependence on β we add it in the superscipts of iterates $\mathbf{v}^{\beta,(n)} = \mathbf{v}^{(n)}$.

Proof of Theorem 1. Let $V_k^\beta(t) = v_k^{\beta,(\lfloor t\beta^{-1}\rfloor)}$. Proposition 3 implies for coordinate k $V_k^\beta(t)$ satisfies

$$V_k^\beta(\beta) - V_k^\beta(0) = \beta\left((\mathbf{v}^\top \boldsymbol{Y})^3 Y_k - v_k(\mathbf{v}^\top \boldsymbol{Y})^4\right) + R_k,$$

where $|R_k| \le 9B^4\beta^2$. Eq. (A.7) implies that if the infinitesimal mean is [17]

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}V_k^\beta(t)\bigg|_{t=0} &= \beta^{-1}\mathbb{E}\left[V_k^\beta(\beta) - v_k \,\middle|\, \boldsymbol{V}^\beta(0) = \mathbf{v}\right] \\ &= |\psi - 3|\, v_k \left(v_k^2 - \sum_{i=1}^{d} v_i^4\right) + \mathcal{O}(B^4\beta).\end{aligned}$$

Using Eq. (A.6) we have the infinitesimal variance

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}(V_k^\beta(t) - v_k)^2\bigg|_{t=0} &= \beta^{-1}\mathbb{E}\left[(V_k^\beta(\beta) - v_k)^2 \,\middle|\, V_k^\beta(0) = \mathbf{v}\right] \\ &\le \beta^{-1} \cdot C^2 B^4 \beta^2,\end{aligned}$$

which tends to 0 as $\beta \to 0^+$. Let $V_k(t)$ be the solution to ODE system Eq. (3.2) with initial values $V_k(0) = v_k^{\beta,(0)}$. Applying standard infinitesimal generator argument [17, Corollary 4.2 in Sec. 7.4] one can conclude that as $\beta \to 0^+$, the Markov process $V_k^\beta(t)$ converges weakly to $V_k(t)$.

■

A.4 Proof of Proposition 2

For simplicity we denote in the proofs that the initial value $V_k(0) = V_k, k = 1, \ldots, d$. Also, throughout this subsection we assume without loss of generality that V_1^2 is maximal among V_k^2, $k = 1, \ldots, d$, and furthermore

$$V_1^2 \ge 2 \max_{k>1} V_k^2. \tag{A.8}$$

Lemma 2. For $\boldsymbol{V} \in \mathcal{S}^{d-1}$ that satisfies Eq. (A.8), then we have for all $t \ge 0$

$$(V_1(t))^2 \ge 2 \max_{k>1} (V_k(t))^2. \tag{A.9}$$

Proof. We compare the coordinate between two distinct coordinates i, j and have by calculus that for all $k > 1$

$$\frac{\mathrm{d}}{\mathrm{d}t} \log\left(\frac{V_k(t)}{V_1(t)}\right)^2 = 2\,|\psi - 3|\left(V_k^2(t) - V_1^2(t)\right). \tag{A.10}$$

So if initially Eq. (A.8) is valid then $\log\left(V_k^2(t)/V_1^2(t)\right)$ is nondecreasing, which indicates for all $t > 0$

$$\log\left(\frac{V_k(t)}{V_1(t)}\right)^2 \le \log\left(\frac{V_k}{V_1}\right)^2 \le \log\frac{1}{2}.$$

Rearranging the above display and taking maximum over $k = 2, \ldots, d$ gives Eq. (A.9).

■

We then establish a lemma that gives the lower bound of drift term related to V_1. To bound the bracket term on the right hand of ODE, one has

$$V_1^2 - \sum_{k=1}^{d} V_k^4 = V_1^2 - V_1^4 - \sum_{k=2}^{d} V_k^4 \leq V_1^2(1 - V_1^2). \tag{A.11}$$

which gives us an upper bound. To obtain a lower bound estimate we first state a lemma stating that the gap between the first and all other coordinates is nondecreasing.

Lemma 3. For $\boldsymbol{V} \in \mathcal{S}^{d-1}$ that satisfies Eq. (A.8) we have

$$V_1^2(1 - V_1^2) \geq V_1^2 - \sum_{k=1}^{d} V_k^4 \geq \frac{V_1^2}{2}(1 - V_1^2). \tag{A.12}$$

Proof. Note Hölder's inequality gives

$$\sum_{k>1} V_k^4 \leq \left(\max_{k>1} V_k^2\right)\left(\sum_{k>1} V_k^2\right), \tag{A.13}$$

where the equality in the above display holds when $V_2^2 = \cdots = V_d^2$. Using Eq. (A.8) and (A.13) one has

$$\begin{aligned} V_1^2 - \sum_{k=1}^{d} V_k^4 &\geq V_1^2 - V_1^4 - \left(\max_{k>1} V_k^2\right)\left(1 - V_1^2\right) \\ &\geq V_1^2 - V_1^4 - \frac{V_1^2}{2}\left(1 - V_1^2\right) = \frac{V_1^2}{2}\left(1 - V_1^2\right). \end{aligned} \tag{A.14}$$

This completes the proof. ■

Lemma 4. For the ODE in Eq. (3.4) which is

$$\frac{\mathrm{d}y}{\mathrm{d}t} = y^2\left(1 - y\right), \tag{A.15}$$

with $y(0) = 2/(d+1)$. By letting T_0 be such that $y(T_0) = 1 - \delta$, we have

$$T_0 \leq d - 3 + 4\log(2\delta)^{-1}. \tag{A.16}$$

Proof. Let T_1 be the traverse time from $2/(d+1)$ to $1/2$, and T_2 be from $1/2$ to $1-\delta$. We have for $y \in [0, 1/2]$

$$\frac{1}{2}y^2 \leq \frac{\mathrm{d}y}{\mathrm{d}t} \leq y^2.$$

Therefore by comparison theorem of ODE [23], $T_1^* \leq T_1 \leq 2T_1^*$ where $y_1(t) = \frac{y_0}{1-y_0 t}$ solves $\mathrm{d}y_1/\mathrm{d}t = y_1^2$, $y_1(0) = 2/(d+1)$. Letting $y_1(T_1^*) = 1/2$ we obtain $T_1^* = (d-3)/2$. For T_2 we note for $y \in [1/2, 1]$

$$\frac{1}{4}(1 - y) \leq \frac{\mathrm{d}y}{\mathrm{d}t} \leq 1 - y.$$

Comparing with $y_2(t) = 1 - (1/2)e^{-t}$ which solves the ODE $\mathrm{d}y_2/\mathrm{d}t = 1 - y_2$ with $y_2(0) = 1/2$, we have $T_2^* = \log(2\delta)^{-1}$ such that $y_2(T_2^*) = 1 - \delta$. To summarize we have

$$T_0 \leq 2T_1^* + 4T_2^* = d - 3 + 4\log(2\delta)^{-1}.$$

■

Proof of Proposition 2. From the ODE in Eq. (3.2) we have

$$\frac{\mathrm{d}V_1^2}{\mathrm{d}t} = 2\left|\psi - 3\right| V_1^2\left(V_1^2 - \sum_{i=1}^{d} V_i^4\right).$$

Combining both Lemmas 2 and 3 we have

$$2\left|\psi - 3\right| V_1^4\left(1 - V_1^2\right) \geq \frac{\mathrm{d}V_1^2}{\mathrm{d}t} \geq \left|\psi - 3\right| V_1^4\left(1 - V_1^2\right).$$

If the starting value of algorithm has $V_1^2 \geq 2\max_{k>1} V_k^2$ then $V_1^2 \geq 2/(d+1)$. By comparison theorem in ODE [23] we know $V_1^2(t)$ runs the auxiliary ODE Eq. (3.4) at a nonconstant rate within $[|\psi-3|, 2|\psi-3|]$. Therefore the time

$$\frac{1}{2}|\psi-3|^{-1}T_0 \leq T \leq |\psi-3|^{-1}T_0.$$

Combining with Lemma 4 we are done.

■

B Detailed Proofs in Section 4

B.1 Proof of Theorem 2

Proof. Proposition 3 implies for $U_k^\beta(t) = \beta^{-1/2} v_k^{\beta,(\lfloor t\beta^{-1}\rfloor)}$, under the conditions in Theorem 2 the one-step increment on coordinate k is

$$U_k^\beta(\beta) - U_k^\beta(0) = \beta^{-1/2}\left(v_k^{\beta,(1)} - v_k^{\beta,(0)}\right) = \beta^{-1/2}\beta\left(\left(\mathbf{v}^\top \boldsymbol{Y}\right)^3 Y_k - v_k\left(\mathbf{v}^\top \boldsymbol{Y}\right)^4\right) + \beta^{-1/2}R_k.$$

Eq. (A.7) implies that the infinitesimal mean is

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}U_k^\beta(t)\bigg|_{t=0} &= \beta^{-1}\mathbb{E}\left[U_k^\beta(\beta) - U_k^\beta(0) \mid \boldsymbol{V}^\beta(0) = \mathbf{v}, \boldsymbol{U}^\beta(0) = \mathbf{u}\right] \\ &= \beta^{-1}\beta^{-1/2}\cdot\beta\,|\psi-3|\, v_k\left(v_k^2 - \sum_{i=1}^d v_i^4\right) + \beta^{-1}\beta^{-1/2}\cdot E_k(\mathbf{v}) \\ &= -\,|\psi-3|\, u_k + o(1).\end{aligned}$$

Using Eq. (A.6) we have the infinitesimal variance

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}(U_k^\beta(t) - U_k^\beta(0))^2\bigg|_{t=0} &= \beta^{-1}\mathbb{E}\left[(U_k^\beta(\beta) - U_k^\beta(0))^2 \mid \boldsymbol{V}^\beta(0) = \mathbf{v}\right] \\ &= \beta^{-2}\mathbb{E}\left[\left(v_k^{\beta,(1)} - v_k^{\beta,(0)}\right)^2 \mid \boldsymbol{V}^\beta(0) = \mathbf{v}\right] \\ &= \mathbb{E}(Y_1^3 Y_k)^2 + o(1) = \psi_6 + o(1).\end{aligned}$$

In addition $\left|U_k^\beta(t) - U_k^\beta(0)\right| \leq CB^2\beta$. Applying standard infinitesimal generator argument [17, Sec. 7.4] one can conclude that as $\beta \to 0^+$, the Markov process $U_k^\beta(t)$ *converges weakly* to $U_k(t)$ the solution to Eq. (4.1).

■

B.2 Proof of Theorem 3

We first prove an auxillary lemma on moment calculations. Proof is deferred to Subsection C.2

Lemma B.1. We have for each $k = 1, \ldots, d$ the following moment expressions:

$$\mathbb{E}\left(\sum_{i=1}^d Y_i\right)^6 Y_k^2 = \psi_8 + 16(d-1)\psi_6 + 15(d-1)\psi_4^2 + 60(d-1)(d-2)\psi_4 + 30(d-1)(d-2)(d-3),$$

and

$$\mathbb{E}\left(\sum_{i=1}^d Y_i\right)^8 = d\psi_8 + 28d(d-1)\psi_6 + 35d(d-1)(1+12(d-1)(d-2))\psi_4 + 105d(d-1)(d-2)(d-3).$$

Proof of Theorem 3. Note from the definition in Eq. (4.3) we have for distinct coordinate pair k, k',

$$\beta^{1/2}W_{kk'} = \log\left(v_k^2\right) - \log\left(v_{k'}^2\right). \tag{B.1}$$

By symmetry we without loss of generality that $v_k^{(0)}, v_{k'}^{(0)} > 0$ and hence

$$W_{kk'}^\beta(\beta) - W_{kk'}^\beta(0) = 2\beta^{-1/2}\log\left(\frac{v_k^{(1)}}{v_k^{(0)}}\right) - 2\beta^{-1/2}\log\left(\frac{v_{k'}^{(1)}}{v_{k'}^{(0)}}\right).$$

However Proposition 3 indicates that

$$\log\left(\frac{v_k^{(1)}}{v_k^{(0)}}\right) = \frac{v_k^{(1)} - v_k^{(0)}}{v_k^{(0)}} + \mathcal{O}(\beta^2) = \beta\left(\mathbf{v}^\top \boldsymbol{Y}\right)^3 \frac{Y_k}{v_k} - \beta(\mathbf{v}^\top \boldsymbol{Y})^4 + \mathcal{O}(\beta^2),$$

and analogously for k'. For infinitesimal mean

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}(W_{kk'}^\beta(t) - W_{kk'}^\beta(0))\Big|_{t=0} &= \beta^{-1}\mathbb{E}\left[W_{kk'}^\beta(\beta) - W_{kk'}^\beta(0)\,\middle|\, W_{kk'}^\beta(0) = W_{kk'}\right]\\ &= \beta^{-1}\cdot 2\beta^{-1/2}\mathbb{E}\left[\beta\left(\mathbf{v}^\top \boldsymbol{Y}\right)^3\left(\frac{Y_k}{v_k} - \frac{Y_{k'}}{v_{k'}}\right) + \mathcal{O}(\beta^2)\right]\end{aligned}$$

Since $\mathbb{E}\left[\left(\mathbf{v}^\top \boldsymbol{Y}\right)^3 (Y_k/v_k)\right] = 3 + (\psi - 3)v_k^2$, and analogously for k', and also $(v_k^{\beta,(0)})^2 \to 1$, we conclude from Eq. (B.1) that

$$\begin{aligned}2(\psi-3)\beta^{-1/2}\left((v_k^{\beta,(0)})^2 - (v_{k'}^{\beta,(0)})^2\right) &= 2(\psi-3)\beta^{-1/2}\cdot\frac{1}{d}\cdot\log\left(\frac{v_k^{\beta,(0)}}{v_{k'}^{\beta,(0)}}\right)^2 + \mathcal{O}(\beta)\\ &\to \frac{2(\psi-3)}{d}W_{kk'}.\end{aligned}$$

For infinitesimal variance

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}(W_{kk'}^\beta(t) - W_{kk'}^\beta(0))^2\Big|_{t=0} &= \beta^{-1}\mathbb{E}\left[\left(W_{kk'}^\beta(\beta) - W_{kk'}^\beta(0)\right)^2\,\middle|\, W_{kk'}^\beta(0) = W_{kk'}\right]\\ &= 4\beta^{-2}\mathbb{E}\left[\left(\log\left(\frac{v_k^{(1)}}{v_k^{(0)}}\right) - \log\left(\frac{v_{k'}^{(1)}}{v_{k'}^{(0)}}\right)\right)^2\,\middle|\, W_{kk'}^\beta(0) = W_{kk'}\right]\\ &= 4\mathbb{E}\left[\left(\mathbf{v}^{*\top}\boldsymbol{Y}\right)^3\frac{Y_k}{v_k} - \left(\mathbf{v}^{*\top}\boldsymbol{Y}\right)^3\frac{Y_{k'}}{v_{k'}}\right]^2 + \mathcal{O}(\beta).\end{aligned}$$

Note the second-order term

$$\mathbb{E}\left[\left(\mathbf{v}^\top \boldsymbol{Y}\right)^6\left(\frac{Y_k}{v_k}\right)^2\,\middle|\, W_{kk'}^\beta(0) = W_{kk'}\right] = 4d^{-2}\cdot\mathbb{E}\left(\sum_{i=1}^d Y_i\right)^6 Y_k^2 \equiv 4d^{-2}Q_1,$$

and similarly for index k'. For the cross term in the expectation we have

$$\mathbb{E}\left[\left(\mathbf{v}^\top \boldsymbol{Y}\right)^6\frac{Y_k}{v_k}\cdot\frac{Y_{k'}}{v_{k'}}\,\middle|\, W_{kk'}^\beta(0) = W_{kk'}\right] = 4d^{-2}\cdot\mathbb{E}\left(\sum_{i=1}^d Y_i\right)^6 Y_kY_{k'} \equiv 4d^{-2}Q_2.$$

From standard polynomial manipulations we have

$$dQ_1 + d(d-1)Q_2 = d\psi_8 + 28d(d-1)\psi_6 + 35d(d-1)(1+12(d-1)(d-2))\psi_4 + 105d(d-1)(d-2)(d-3),$$

and

$$Q_1 = \psi_8 + 16(d-1)\psi_6 + 15(d-1)\psi_4^2 + 60(d-1)(d-2)\psi_4 + 30(d-1)(d-2)(d-3).$$

Therefore

$$\begin{aligned}Q_1 - Q_2 &= \frac{d^2Q_1 - dQ_1 - d(d-1)Q_2}{d(d-1)}\\ &= \psi_8 + (16d-28)\psi_6 + 15d\psi_4^2 - 5(72d^2 - 228d + 175)\psi_4 + 15(2d-7)(d-2)(d-3).\end{aligned}$$

Summarize the above calculations we obtain as $\beta \to 0^+$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}(W_{kk'}^\beta(t) - W_{kk'}^\beta(0))^2\Big|_{t=0} &= 4\mathbb{E}\left[\left(\mathbf{v}^{*\top}\boldsymbol{Y}\right)^3\frac{Y_k}{v_k} - \left(\mathbf{v}^{*\top}\boldsymbol{Y}\right)^3\frac{Y_{k'}}{v_{k'}}\right]^2 + \mathcal{O}(\beta)\\ &= 8d^{-2}(Q_1 - Q_2) + \mathcal{O}(\beta).\end{aligned}$$

Combining the last two displays concludes the theorem.

■

C Proof of Auxillary Results

C.1 Proof of Proposition 3

For $\mathbf{v}^{(0)} = \mathbf{v} \in \mathcal{S}^{d-1}$ the update equation becomes

$$\mathbf{v}^{(1)} = \left\| \mathbf{v} + \beta \left(\mathbf{v}^\top \boldsymbol{Y} \right)^3 \boldsymbol{Y} \right\|^{-1} \left(\mathbf{v} + \beta \left(\mathbf{v}^\top \boldsymbol{Y} \right)^3 \boldsymbol{Y} \right).$$

For the simplicity for discussion we prove under the condition $\psi > 3$ (the case of $\psi < 3$ is analogous). To prove Proposition 3 in the case of $\psi > 3$, we first introduce

Lemma 5. For $x \in [0, 1)$ we have

$$\left| (1+x)^{-1/2} - 1 + \frac{x}{2} \right| \le 2 \left(\frac{x}{2} \right)^2. \tag{C.1}$$

Proof. Taylor expansion suggests for $|x| < 1$

$$(1+x)^{-1/2} = \sum_{n=0}^{\infty} \binom{-\frac{1}{2}}{n} x^n = 1 - \frac{1}{2}x + \frac{3}{8}x^2 - \frac{5}{16}x^3 + \cdots$$

which is an alternating series for $x \in [0, 1)$, whereas the absolute terms approach to 0 monotonically

$$\left| \binom{-\frac{1}{2}}{n+1} x^{n+1} \right| \le \left| \binom{-\frac{1}{2}}{n} x^n \right|.$$

This indicates that for $x \in [0, 1)$

$$\left| (1+x)^{-1/2} - 1 + \frac{1}{2}x \right| \le \frac{3}{8}x^2 \le \frac{1}{2}x^2,$$

which completes the proof of Lemma 5.

■

Proof of Proposition 3. When Eq. (A.4) is satisfied, and noting $|\mathbf{v}^\top \boldsymbol{Y}|^2 \le \|\boldsymbol{Y}\|^2 \le B$, we have from Eq. (A.4)

$$\beta(\mathbf{v}^\top \boldsymbol{Y})^4 + \frac{1}{2}\beta^2 (\mathbf{v}^\top \boldsymbol{Y})^6 \|\boldsymbol{Y}\|^2 \le B^2\beta + \frac{1}{2}B^4\beta^2 \le \frac{4}{3}B^2\beta < 1,$$

and hence from Eq. (C.1) in Lemma 5 there exists a $Q_1(\mathbf{v}, \boldsymbol{Y})$ with

$$|Q_1(\mathbf{v}, \boldsymbol{Y})| \le 2 \left(\beta(\mathbf{v}^\top \boldsymbol{Y})^4 + \frac{1}{2}\beta^2 (\mathbf{v}^\top \boldsymbol{Y})^6 \|\boldsymbol{Y}\|^2 \right)^2 \le \frac{32}{9} B^4 \beta^2,$$

such that, with $Q_2(\mathbf{v}, \boldsymbol{Y}) = -\frac{1}{2}\beta^2 (\mathbf{v}^\top \boldsymbol{Y})^6 \|\boldsymbol{Y}\|^2 + Q_1(\mathbf{v}, \boldsymbol{Y})$, we have

$$\begin{aligned}
\|\mathbf{v} + \beta(\mathbf{v}^\top \boldsymbol{Y})^3 \mathbf{Y}\|^{-1} &= \left(1 + 2\beta(\mathbf{v}^\top \boldsymbol{Y})^4 + \beta^2 (\mathbf{v}^\top \boldsymbol{Y})^6 \|\boldsymbol{Y}\|^2\right)^{-1/2} \\
&= 1 - \beta(\mathbf{v}^\top \boldsymbol{Y})^4 - \frac{1}{2}\beta^2 (\mathbf{v}^\top \boldsymbol{Y})^6 \|\boldsymbol{Y}\|^2 + Q_1(\mathbf{v}, \boldsymbol{Y}) \\
&= 1 - \beta(\mathbf{v}^\top \boldsymbol{Y})^4 + Q_2(\mathbf{v}, \boldsymbol{Y}), \qquad \text{(C.2)}
\end{aligned}$$

where

$$|Q_2(\mathbf{v}, \boldsymbol{Y})| \le \frac{1}{2}B^4\beta^2 + \frac{32}{9}B^4\beta^2 = \frac{73}{18}B^4\beta^2. \tag{C.3}$$

Using Eqs. (C.2) and (C.3) we have

$$\begin{aligned}
\widehat{v}_k - v_k &= \left\| \mathbf{v} + \beta(\mathbf{v}^\top \boldsymbol{Y})^3 \mathbf{Y} \right\|^{-1} \left(v_k + \beta \left(\mathbf{v}^\top \boldsymbol{Y} \right)^3 Y_k \right) - v_k \\
&= \left(1 - \beta(\mathbf{v}^\top \boldsymbol{Y})^4 + Q_2(\mathbf{v}, \boldsymbol{Y})\right) \left(v_k + \beta \left(\mathbf{v}^\top \boldsymbol{Y} \right)^3 Y_k \right) - v_k \\
&= \beta \left((\mathbf{v}^\top \boldsymbol{Y})^3 Y_k - v_k (\mathbf{v}^\top \boldsymbol{Y})^4 \right) + Q_3(\mathbf{v}, \boldsymbol{Y}), \qquad \text{(C.4)}
\end{aligned}$$

where

$$Q_3(\mathbf{v}, \boldsymbol{Y}) = \left(v_k + \beta \left(\mathbf{v}^\top \boldsymbol{Y} \right)^3 Y_k \right) Q_2(\mathbf{v}, \boldsymbol{Y}) - \beta^2 (\mathbf{v}^\top \boldsymbol{Y})^7 Y_k \tag{C.5}$$

which has the following estimate

$$\begin{aligned} |Q_3(\mathbf{v},\boldsymbol{Y})| &\le \left|v_k + \beta\big(\mathbf{v}^\top \mathbf{Y}\big)^3 Y_k\right| |Q_2(\mathbf{v},\boldsymbol{Y})| + \beta^2 \left|\big(\mathbf{v}^\top \mathbf{Y}\big)^7 Y_k\right| \\ &\le \left(1 + B^2\beta\right) \frac{73}{18} B^4\beta^2 + B^4\beta^2 \le 9B^4\beta^2. \end{aligned} \tag{C.6}$$

Denoting $Q_3(\mathbf{v},\boldsymbol{Y})$ by the random variable R_k, Eqs. (C.4), (C.5), (C.6) together concludes (i) of Prop. 3.

For (ii), note Eq. (A.5) gives

$$\left|v_k^{(1)} - v_k^{(n)}\right| \le \beta\left(\|\boldsymbol{Y}\|^2 + \|\boldsymbol{Y}\|^2\right) + 9B^4\beta^2 \le 8B^2\beta,$$

so it is concluded.

For (iii), we set $E_k(\mathbf{v}) = \mathbb{E}\left[R_k \mid \mathbf{v}^{(0)} = \mathbf{v}\right]$. Under Assumption 1 we take conditional expectation on $\mathbf{v}^{(n)} = \mathbf{v}$ on both sides of Eq. (A.5) to obtain

$$\begin{aligned} \mathbb{E}\left[v_k^{(1)} - v_k^{(0)} \,\middle|\, \mathbf{v}^{(0)} = \mathbf{v}\right] &= \beta\mathbb{E}\left[\left(\mathbf{v}^\top \mathbf{Y}\right)^3 Y_k - v_k\left(\mathbf{v}^\top \mathbf{Y}\right)^4 \,\middle|\, \mathbf{v}^{(0)} = \mathbf{v}\right] + \mathbb{E}\left[R_k \mid \mathbf{v}^{(0)} = \mathbf{v}\right] \\ &= \beta(\psi - 3)v_k\left(v_k^2 - \sum_{i=1}^d v_i^4\right) + E_k(\mathbf{v}). \end{aligned} \tag{C.7}$$

Similar to the proof of Lemma 1 in Subsection A.1 we quote another polynomial expansion [9]

$$\left(\sum x_i\right)^3 = \sum x_i^3 + 3\sum x_i^2 x_j + 6\sum x_{i_1}x_{i_2}x_{i_3}.$$

where the summations above iterate through all monomial terms. Plugging in $x_i = v_i Y_i$ and taking conditional expectations, we conclude that under Assumption 1

$$\begin{aligned} \mathbb{E}\left[\left(\mathbf{v}^\top \mathbf{Y}\right)^3 Y_k \,\middle|\, \mathbf{v}^{(0)} = \mathbf{v}\right] &= v_k^3\mathbb{E}\left(Y_i^4\right) + 3\sum_{i:i\neq k} v_i^2 v_k \mathbb{E}\left(Y_i^2\right)\mathbb{E}\left(Y_k^2\right) \\ &= \psi v_k^3 + 3(1 - v_k^2)v_k = 3v_k + (\psi - 3)v_k^3. \end{aligned} \tag{C.8}$$

In Eq. (A.1) we have

$$\begin{aligned} \mathbb{E}\left[\left(\mathbf{v}^\top \mathbf{Y}\right)^3 Y_k - v_k\left(\mathbf{v}^\top \mathbf{Y}\right)^4 \,\middle|\, \mathbf{v}^{(0)} = \mathbf{v}\right] &= 3v_k + (\psi - 3)v_k^3 - v_k\left(3 + (\psi - 3)\sum_{i=1}^d v_i^4\right) \\ &= (\psi - 3)v_k\left(v_k^2 - \sum_{i=1}^d v_i^4\right). \end{aligned} \tag{C.9}$$

Combining Eqs. (C.7) and (C.9) completes the proof.

■

C.2 Proof of Lemma B.1

Proof. As in proof of Lemmas 1 and Proposition 3, we have the final polynomial expansions [9] that

$$\left(\sum x_i\right)^6 = \sum x_i^6 + 15\sum x_i^4 x_j^2 + 90\sum x_i^2 x_j^2 x_k^2 + \text{terms that has odd-order factors},$$

and using some combinatorics counting we have

$$\begin{aligned} \left(\sum x_i\right)^8 = \sum x_i^8 + 28\sum x_i^6 x_j^2 + 70\sum x_i^4 x_j^4 + 420\sum x_i^4 x_j^2 x_k^2 \\ + 2520\sum x_i^2 x_j^2 x_k^2 x_l^2 + \text{terms that has odd-order factors}. \end{aligned}$$

Therefore to show the first equality, note from Assumption 1 we can assume WLOG that $k = 1$. Thus

$$\begin{aligned} \mathbb{E}\left(\sum_{i=1}^d Y_i\right)^6 Y_1^2 = \sum_{i=1}^d \mathbb{E}Y_i^6 Y_1^2 + 15\sum_{1\le i<j\le d} \mathbb{E}Y_i^4 Y_j^2 Y_1^2 + 15\sum_{1\le i<j\le d} \mathbb{E}Y_j^4 Y_i^2 Y_1^2 \\ + 90\sum_{1\le i<j<k\le d} \mathbb{E}Y_i^2 Y_j^2 Y_k^2 Y_1^2 \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}Y_1^8 + 15 \sum_{2\le j\le d} \mathbb{E}Y_1^6 Y_j^2 + 15 \sum_{2\le j\le d} \mathbb{E}Y_1^4 Y_j^4 + 90 \sum_{2\le j<k\le d} \mathbb{E}Y_1^4 Y_j^2 Y_k^2 + \sum_{i=2}^{d} \mathbb{E}Y_i^6 Y_1^2 \\
&\qquad + 15 \sum_{2\le i<j\le d} \mathbb{E}Y_i^4 Y_j^2 Y_1^2 + 15 \sum_{2\le i<j\le d} \mathbb{E}Y_j^4 Y_i^2 Y_1^2 + 90 \sum_{2\le i<j<k\le d} \mathbb{E}Y_i^2 Y_j^2 Y_k^2 Y_1^2 \\
&= \psi_8 + 15(d-1)\psi_6 + 15(d-1)\psi_4^2 + 90\binom{d-1}{2}\psi_4 + (d-1)\psi_6 \\
&\qquad + 15\binom{d-1}{2}\psi_4 + 15\binom{d-1}{2}\psi_4 + 90\binom{d-1}{3} \\
&= \psi_8 + 16(d-1)\psi_6 + 15(d-1)\psi_4^2 + 60(d-1)(d-2)\psi_4 + 30(d-1)(d-2)(d-3).
\end{aligned}
$$

Also

$$
\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{d} Y_i\right)^8 &= \sum_{i=1}^{d} \mathbb{E}Y_i^8 + 28 \sum_{1\le i<j\le d} \mathbb{E}Y_i^6 \mathbb{E}Y_j^2 + 28 \sum_{1\le j<i\le d} \mathbb{E}Y_i^6 \mathbb{E}Y_j^2 + 70 \sum_{1\le i<j\le d} \mathbb{E}Y_i^4 \mathbb{E}Y_j^4 \\
&\qquad + 420 \sum_{i<j<k} \mathbb{E}Y_i^4 \mathbb{E}Y_j^2 \mathbb{E}Y_k^2 + 420 \sum_{j<i<k} \mathbb{E}Y_i^4 \mathbb{E}Y_j^2 \mathbb{E}Y_k^2 \\
&\qquad + 420 \sum_{j<k<i} \mathbb{E}Y_i^4 \mathbb{E}Y_j^2 \mathbb{E}Y_k^2 + 2520 \sum_{i<j<k<l} \mathbb{E}Y_i^2 \mathbb{E}Y_j^2 \mathbb{E}Y_k^2 \mathbb{E}Y_l^2,
\end{aligned}
$$

which is equal to

$$
\begin{aligned}
&d\psi_8 + 28\binom{d}{2}\psi_6 + 28\binom{d}{2}\psi_6 + 70\binom{d}{2}\psi_4 \\
&\qquad + 420(d-1)\binom{d}{3}\psi_4 + 420(d-1)\binom{d}{3}\psi_4 + 420(d-1)\binom{d}{3}\psi_4 + 2520\binom{d}{4} \\
&\quad = d\psi_8 + 28d(d-1)\psi_6 + 35d(d-1)(1 + 12(d-1)(d-2))\psi_4 + 105(d-1)(d-2)(d-3).
\end{aligned}
$$

This completes the proof.

■