List of proofs

Proof of Lemma 1. First, note that
\[ \mathbb{E} \left[ \|GX\|^2 \right] = \mathbb{E} \left[ \left( \sum_{i=1}^n \|g_i\| \|\hat{g}_i X\| \right)^2 \right], \]
where \( g_i \) is the \( i \)th column of \( G \) and \( \hat{g}_i = g_i/\|g_i\|_\infty \). Using the condition \( \|G\|_\infty \leq 1 \) and Jensen’s inequality, we find that
\[ \mathbb{E} \left[ \|GX\|^2 \right] \leq \sum_{i=1}^n \|g_i\| \mathbb{E} \left[ \|\hat{g}_i X\|^2 \right] \leq \|X\|^2. \]
The other inequality follows by noting that for any \( f \) inequality, we find that
\[ \text{Proof of Theorem 1.} \]
\[ \|f\|_\infty \leq 1 \]
\[ \text{dimensions. Suppose that the result holds in } n - 1 \text{ dimensions. We will show that it must also therefore hold in } n \text{ dimensions and conclude, by induction, that the result holds in any dimension.} \]
Let \( \bar{\Phi} \) be the \((n-1) \times (n-1)\) principle submatrix of an \( n \times n \) matrix \( A \). For any vector \( f \in \mathbb{C}^n \) we can write
\[ f^H Af = \sum_{i=1}^n |f_i|^2 A_{ii} + 2 \mathbb{R} \left( \sum_{i=1}^n \sum_{j=1}^{i-1} f_i A_{ij} \hat{f}_j \right) \]
\[ = f^H \bar{A} f + |f_n|^2 A_{nn} + 2 \mathbb{R} \left( f_n \sum_{j=1}^{n-1} A_{nj} \hat{f}_j \right), \]
where \( \hat{f} \in \mathbb{C}^{n-1} \) has entries equal to the first \( n - 1 \) entries of \( f \).
By the induction hypothesis, we can choose the first \( n - 1 \) entries of \( f \) (i.e., \( \hat{f} \)) so that the right-hand side of the last display is not less than
\[ \sum_{i=1}^{n-1} A_{ii} + |f_n|^2 A_{nn} + 2 \mathbb{R} \left( f_n \sum_{j=1}^{n-1} A_{nj} \hat{f}_j \right). \]
If, for this choice of \( \hat{f} \), \( \sum_{j=1}^{n-1} A_{nj} \hat{f}_j \) is nonzero, then choose \( f_n \) as
\[ f_n = \frac{\sum_{j=1}^{n-1} A_{nj} \hat{f}_j}{\sum_{j=1}^{n-1} A_{nj} \hat{f}_j}. \]
Otherwise set \( f_n = 1 \). With the resulting choice of \( f_n \),
\[ |f_n|^2 A_{nn} + 2 \mathbb{R} \left( f_n \sum_{j=1}^{n-1} A_{nj} \hat{f}_j \right) \geq A_{nn}. \]
We have therefore shown that
\[ \sup_{\|f\|_\infty \leq 1} f^H Af \geq \sum_{i=1}^n A_{ii}. \]

Proof of Theorem 2. Let \( V_t^m \) be generated by (11). Let \( Y_t^m = \Phi_t^m(V_t^m) \) and notice that
\[ \mathcal{U}(V_t^m) = \mathcal{U}(\mathcal{M}(Y_{t-1}^m)) \]
\[ \leq R + \alpha \mathcal{U}(V_{t-1}^m) + \alpha \left( \mathcal{U}(Y_{t-1}^m) - \mathcal{U}(V_{t-1}^m) \right). \]
Using the fact that \( \mathcal{U} \) is twice differentiable with bounded second derivative, this last expression is bounded above by
\[ \mathcal{U}(V_t^m) \leq R + \alpha \mathcal{U}(V_{t-1}^m) + \alpha \nabla \mathcal{U}(V_{t-1}^m) (Y_{t-1}^m - V_{t-1}^m) + \frac{\alpha \sigma}{2} \|G (Y_{t-1}^m - V_{t-1}^m)\|^2. \]
Taking the expectation and using (30) yields
\[ \mathbb{E} \left[ \mathcal{U}(V_t^m) \right] \leq R + \alpha \mathbb{E} \left[ \mathcal{U}(V_{t-1}^m) \right] + \frac{\alpha \sigma}{2} \mathbb{E} \left[ \|G (Y_{t-1}^m - V_{t-1}^m)\|^2 \right]. \]
An application of Lemma 1 reveals that
\[ \mathbb{E} [ ||G (Y_{t-1}^{m} - V_{t-1}^{m})||^2 ] \leq ||Y_{t-1}^{m} - V_{t-1}^{m}||^2. \]
As a consequence, noting (28), we arrive at the upper bound
\[
\begin{align*}
\mathbb{E} [ \mathcal{U}(V_{t}^{m}) ] & \leq R + \alpha \mathbb{E} [ \mathcal{U}(V_{t-1}^{m}) ] + \frac{\alpha \gamma^2 \sigma^2}{2m} \mathbb{E} [ ||V_{t-1}^{m}||^2 ] \\
& \leq R + \alpha \left( 1 + \frac{\beta \gamma^2 \sigma^2}{2m} \right) \mathbb{E} [ \mathcal{U}(V_{t-1}^{m}) ],
\end{align*}
\]
from which we can conclude that
\[ \mathbb{E} [ ||V_{t}^{m}||^2 ] \leq \beta \mathbb{E} [ \mathcal{U}(V_{t}^{m}) ] \leq \beta R \left[ 1 - \alpha t \left( 1 + \frac{\beta \gamma^2 \sigma^2}{2m} \right) \right] + \beta \alpha t \left( 1 + \frac{\beta \gamma^2 \sigma^2}{2m} \right)^t \mathcal{U}(V_0^{m}). \]

\textbf{Proof of Theorem 2.} We begin with a standard expansion of the scheme’s error.
\[ ||V_{t}^{m} - v_t|| = ||V_{t}^{m} - \mathcal{M}_{t}^{m}(v_0)|| \]
\[ = \left| \sum_{r=0}^{t-1} \mathcal{M}_{r}^{m}(V_{r+1}^{m}) - \mathcal{M}_{t}^{m}(V_{r}^{m}) \right|. \]
Now notice that if we define \( Y_{r}^{m} = \Phi_{r}^{m}(V_{r}^{m}) \), then \( V_{r}^{m} = \mathcal{M}(Y_{r}^{m}) \) and the last equation becomes
\[ ||V_{t}^{m} - v_t|| = \left| \sum_{r=0}^{t-1} \mathcal{M}_{r}^{m}(Y_{r}) - \mathcal{M}_{t}^{m}(Y_{r}^{m}) \right|. \]
The right-hand side of the last equation is bounded above by
\[ \left| \sum_{r=0}^{t-1} \mathcal{M}_{r}^{m}(Y_{r}) - \mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m}] \right| + \sum_{r=0}^{t-1} ||\mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m}] - \mathcal{M}_{r}^{m}(V_{r}^{m})||. \]
Considering the first term in the last display, note that, for any fixed \( f \in \mathbb{C}^{\mathbb{R}} \),
\[
\begin{align*}
\mathbb{E} [ |f^{H} \sum_{r=0}^{t-1} (\mathcal{M}_{r}^{m}(Y_{r}) - \mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m}])|^{2} ] &= \sum_{r=0}^{t-1} \mathbb{E} [ |f^{H}(\mathcal{M}_{r}^{m}(Y_{r}) - \mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m})]|^{2} ] \\
& + 2 \sum_{s=0}^{t-1} \sum_{r=s+1}^{t-1} \mathbb{R} \{ \mathbb{E} [ (f^{H}(\mathcal{M}_{r}^{m}(Y_{r}) - \mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m})]) \times (f^{H}(\mathcal{M}_{s}^{m}(Y_{s}) - \mathbb{E}[\mathcal{M}_{s}^{m}(Y_{s}) | V_{s}^{m})]) ] \}.
\end{align*}
\]
Letting \( \mathcal{F}_{r} \) denote the \( \sigma \)-algebra generated by \( \{Y_{s}^{m}\}_{s=0}^{r-1} \) and \( \{Y_{r}^{m}\}_{r=0}^{t-1} \), for \( s < r \) we can write
\[
\begin{align*}
\mathbb{E} [ (f^{H}(\mathcal{M}_{r}^{m}(Y_{r}) - \mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m})]) \times (f^{H}(\mathcal{M}_{s}^{m}(Y_{s}) - \mathbb{E}[\mathcal{M}_{s}^{m}(Y_{s}) | V_{s}^{m})]) ]
& = \mathbb{E} [ (f^{H}(\mathcal{M}_{r}^{m}(Y_{r}) - \mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m})]) \times (f^{H}(\mathcal{M}_{s}^{m}(Y_{s}) - \mathbb{E}[\mathcal{M}_{s}^{m}(Y_{s}) | V_{s}^{m})]) ] \\
& \quad \times (f^{H}(\mathcal{M}_{s}^{m}(Y_{s}) - \mathbb{E}[\mathcal{M}_{s}^{m}(Y_{s}) | V_{s}^{m})])].
\end{align*}
\]
Because, conditioned on \( V_{r}^{m} \), \( Y_{r}^{m} \) is independent of \( \mathcal{F}_{r} \), the expression above vanishes exactly.

Supreming over the choice of \( f \), we have shown that
\[ \|V_{t}^{m} - v_t\| \leq \left( \sum_{r=0}^{t-1} ||\mathcal{M}_{r}^{m}(Y_{r}) - \mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m}]|| \right)^{1/2} + \sum_{r=0}^{t-1} ||\mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m}] - \mathcal{M}_{r}^{m}(V_{r}^{m})||. \]
Expanding the term inside of the square root, we find that
\[ \|V_{t}^{m} - v_t\| \leq \left( \sum_{r=0}^{t-1} \left( ||\mathcal{M}_{r}^{m}(Y_{r}) - \mathcal{M}_{r}^{m}(V_{r}^{m})|| + ||\mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m}] - \mathcal{M}_{r}^{m}(V_{r}^{m})|| \right)^{2} \right)^{1/2} \\
+ \sum_{r=0}^{t-1} \left( ||\mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m}] - \mathcal{M}_{r}^{m}(V_{r}^{m})|| \right)^{2} \]
\[ \leq \left( \sum_{r=0}^{t-1} \left( ||\mathcal{M}_{r}^{m}(Y_{r}) - \mathcal{M}_{r}^{m}(V_{r}^{m})|| \right)^{2} \right)^{1/2} + \left( \sum_{r=0}^{t-1} \left( ||\mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m}] - \mathcal{M}_{r}^{m}(V_{r}^{m})|| \right)^{2} \right)^{1/2} \]
\[ + \sum_{r=0}^{t-1} \left( ||\mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m}] - \mathcal{M}_{r}^{m}(V_{r}^{m})|| \right)^{2}, \]
where, in the second inequality, we have used the triangle inequality for the \( \ell^2 \)-norm in \( \mathbb{R}^{t} \). Noting that \( \mathbb{E} [A(V_{r}^{m})(Y_{r} - V_{r}^{m}) | V_{r}^{m}] = 0 \) yields
\[ \mathbb{E}[\mathcal{M}_{r}^{m}(Y_{r}) | V_{r}^{m}] - \mathcal{M}_{r}^{m}(V_{r}^{m}) = \mathbb{E} [ (\mathcal{M}_{r}^{m} - A_{r})(Y_{r}) | V_{r}^{m}] - (\mathcal{M}_{r}^{m} - A_{r})(V_{r}^{m}), \]
As a consequence, applying our assumptions (31) and (32), we obtain the upper bound
\[ \|V_{t}^{m} - v_t\| \leq (L_{1} + L_{2}) \left( \sum_{r=0}^{t-1} \alpha^{2(t-r)} || \Phi_{r}^{m}(V_{r}^{m}) - V_{r}^{m} ||^{2} \right)^{1/2} + L_{2} \sum_{r=0}^{t-1} \alpha^{r-t} || \Phi_{r}^{m}(V_{r}^{m}) - V_{r}^{m} ||^{2}. \]
Bounding the error from the random compressions, we arrive at the error bound
\[ \|V_{t}^{m} - v_t\| \leq \frac{\gamma(L_{1} + L_{2})}{\sqrt{m}} \left( \sum_{r=0}^{t-1} \alpha^{2(t-r)} \mathbb{E} [ ||V_{r}^{m}||^2 ] \right)^{1/2} + \frac{\gamma L_{2}}{m} \sum_{r=0}^{t} \alpha^{r-t} \mathbb{E} [ ||V_{r}^{m}||^2 ]. \]

\[ \square \]
Proof of Corollary 1. We have already seen that when \( \mathcal{M}(v) = Kv \) we can take \( \alpha = \|K\|_1 \) in the statement of Theorem 2 to verify conditions (31) and (32). We have also commented above that when \( K \) is nonnegative, the quantities \( E\left[\|V^m_r\|_1^2\right] \) can be bounded independently of \( n \).

When \( \mathcal{M}(v) = Kv/\|Kv\|_1 \), bounding the size of the iterates is not an issue, but it becomes slightly more difficult to verify (31) and (32). That \( K \) is aperiodic and irreducible implies that the dominant left and right eigenvectors, \( v_L \) and \( v_R \), of \( K \) are unique and have all positive entries. Because power iteration is invariant to scalar multiples of \( K \) we can assume that the dominant eigenvalue of \( K \) is 1. We will assume that \( v_r \) is normalized so that \( \|v_r\|_\infty = 1 \) and that \( v_R \) is normalized so that \( v_L^T v_R = 1 \). Let \( D \) be the diagonal matrix with \( D_{ii} = (v_L)_i \) (i.e., \( D1 = v_L \)).

Our matrix \( K \) can be written \( K = D^{-1}SPD \) where \( S \) is an aperiodic, irreducible, column-stochastic matrix. Let

\[
\tilde{K} = K - v_Rv_L^T = D^{-1}SPD,
\]

where we have defined the projection \( P = I - Dv_R1^T \). Note that \( \|P\|_1 \leq 2 \) and that \( PSP = SP \) so that for any positive integer \( r \), \( \tilde{K}^r = D^{-1}S^rPD \). Letting

\[
C = \frac{1}{\min_j \{(v_L)_j\}} \geq 1
\]

we find that, for any positive integer \( r \),

\[
\|\tilde{K}^r\|_1 \leq \|D^{-1}\|_1 \|D\|_1 \|S^rP\|_1 \leq 2C \sup_{\|v\|_1 = 1} \|S^r v\|_1 \leq 2C \alpha^r
\]

where

\[
\alpha = \sup_{\|v\|_1 = 1} \|Sv\|_1
\]

Aperiodicity and irreducibility of \( S \) implies that \( \alpha < 1 \). We also have that

\[
\sup_{v^*_Lv = 1} \|K^rv\|_1 \leq C \quad \text{and} \quad \inf_{v^*_Lv \geq 0} \|K^rv\|_1 \geq 1.
\]

Now let \( u \) and \( v \) be any two non-negative vectors normalized so that \( v^*_Lv = v^*_Lv = 1 \) and, for \( \theta \in [0, 1] \), define \( w_\theta = (1 - \theta)u + \theta v \). Note that \( w_\theta \) also has non-negative entries and that \( v^*_Lw_\theta = 1 \). For any fixed \( f \in \mathbb{R}^n \) with \( \|f\|_\infty \leq 1 \), define the function

\[
\varphi_r(u, v; \theta) = \frac{f^T K^r w_\theta}{\|K^r w_\theta\|_1} - \frac{f^T K^r u}{\|K^r u\|_1}.
\]

Our goal is to establish bounds on

\[
\varphi_r(u, v; 1) = \frac{f^T K^r v}{\|K^r v\|_1} - \frac{f^T K^r u}{\|K^r u\|_1}.
\]

To that end note that

\[
\frac{d}{d\theta} \varphi_r(u, v; \theta) = \frac{f^T K^r (v - u)}{\|K^r w_\theta\|_1} - \frac{(f^T K^r w_\theta)(1^T K^r (v - u))}{\|K^r w_\theta\|_1^2}
\]

and

\[
\frac{d^2}{d\theta^2} \varphi_r(u, v; \theta) = -2 \frac{(f^T K^r (v - u))(1^T K^r (v - u))}{\|K^r w_\theta\|_1^2} + 2 \frac{(f^T K^r w_\theta)(1^T K^r (v - u))^2}{\|K^r w_\theta\|_1^3}.
\]

Observing that \( K^r (v - u) = \tilde{K}^r (v - u) \), and applying our bounds we find that

\[
|\varphi_r(u, v; 1)| \leq \max_{\theta} \left| \frac{d}{d\theta} \varphi_r(u, v; \theta) \right|
\]

\[
\leq \left| f^T \tilde{K}^r (v - u) \right| + C \left| 1^T \tilde{K}^r (v - u) \right|
\]

\[
\leq 4C^2 \alpha^r \|G(v - u)\|_1
\]

(52)
where $G \in \mathbb{R}^{n \times n}$ is the matrix with first row equal to $f^T \tilde{K}^r/\|f^T \tilde{K}^r\|_\infty$, second row equal to $1^T \tilde{K}^r/\|1^T \tilde{K}^r\|_\infty$, and all other entries equal to 0.

Defining the matrix valued function
\[
A_r(u) = \frac{1}{\|K^ru\|_1} \left[ I - \frac{K^ru1^T}{\|K^ru\|_1} \right] K^r
\]
we observe that
\[
\frac{d}{d\theta} \varphi_r(u, v; \theta) = f^T A_r(u)(v - u)
\]
so that
\[
|\varphi_r(u, v; 1) - f^T A_r(u)(v - u)| \leq \frac{1}{2} \max_\theta \left| \frac{d^2}{d\theta^2} \varphi_r(u, v; \theta) \right|
\leq |f^T \tilde{K}^r(v - u)||1^T \tilde{K}^r(v - u)| + C|1^T \tilde{K}^r(v - u)|^2
\leq 16C^3 \alpha^{2r} \|G(v - u)\|^2
\]
Expressions (52) and (53) verify the stability conditions in the statement of Theorem 2 with $L_1$ and $L_2$ dependent only on $C$ yielding the first term on the right-hand side of (33). The second term follows similarly when one observes that (31) implies
\[
\sup_{v, \tilde{v} \in \mathcal{X}} \left| \mathcal{M}_s^r(v) - \mathcal{M}_s^r(\tilde{v}) \right|_1 \leq L_1 \alpha^{r-s}.
\]

Proof of Lemma 3. If $Y^m_t = \Phi^m_t(V^m_t)$, then
\[
\mathbb{E} \left[ |f^t \Phi^m_t(V^m_t) - f^t V^m_t|^2 \mid Y^m_{t-1} \right] = \mathbb{E} \left[ |f^t \Phi^m_t (Y^m_{t-1} + \varepsilon b(Y^m_{t-1})) - f^t (Y^m_{t-1} + \varepsilon b(Y^m_{t-1}))|^2 \mid Y^m_{t-1} \right]
\leq \gamma_p \frac{\varepsilon}{m} \|b(Y^m_{t-1})\|_1 \|V^m_t\|_1
\]
for some constant $C$. Our assumed bound on the growth of $b$ along with (29) implies that
\[
\mathbb{E} \left[ \|b(Y^m_{t-1})\|^2 \right] \leq C' (1 + \mathbb{E} \left[ \|V^m_{t-1}\|^2 \right])
\]
for some constant $C'$. From these bounds it follows that for some constant $\hat{\gamma}$,
\[
\|\Phi^m_t(V^m_t) - V^m_t\|^2 \leq \hat{\gamma}^2 \frac{\varepsilon}{m} \mathbb{E} \left[ \|V^m_t\|^2 \right] \sqrt{1 + \mathbb{E} \left[ \|V^m_{t-1}\|^2 \right]}.
\]

Proof of Theorem 5. By exactly the same arguments used in the proof of Theorem 2 we arrive at the bound
\[
\|V^m_t - v_t\| \leq (L_1 + L_2) \left( \sum_{r=0}^{t-1} e^{-2\beta(t-r)\varepsilon} \|\Phi^m_t(V^m_r) - V^m_r\|^2 \right)^{1/2} + L_2 \sum_{r=0}^{t-1} e^{-\beta(t-r)\varepsilon} \|\Phi^m_t(V^m_r) - V^m_r\|^2.
\]

Bounding the error from the random compressions, we arrive at the error bound
\[
\|V^m_t - v_t\| \leq \frac{\hat{\gamma}(L_1 + L_2)}{\sqrt{m}} \left( e^{-2\beta\varepsilon} \mathbb{E} \left[ \|V^m_0\|^2 \right] + \varepsilon \sum_{r=0}^{t-1} e^{-2\beta(t-r)\varepsilon} \mathbb{E} \left[ \|V^m_r\|^2 \right] \sqrt{1 + \mathbb{E} \left[ \|V^m_{t-1}\|^2 \right]} \right)^{1/2}
+ \frac{\hat{\gamma}^2 L_2}{m} \sum_{r=0}^{t-1} e^{-\beta(t-r)\varepsilon} \mathbb{E} \left[ \|V^m_r\|^2 \right] \sqrt{1 + \mathbb{E} \left[ \|V^m_{t-1}\|^2 \right]}.
\]

Proof of Theorem 6. By an argument very similar to that in the proof of Theorem 2, we arrive at the bound
\[
\|V^m_t - v_t\| \leq \left( \sum_{r=0}^{t-1} \mathbb{E} \left[ \mathcal{M}_{t+1}^r(V^m_r + \varepsilon b(Y^m_r)) - \mathcal{M}_{t+1}^r(V^m_r + \varepsilon b(V^m_r)) \right]^2 \right)^{1/2}
+ \left( \sum_{r=0}^{t-1} \mathbb{E} \left[ \mathcal{M}_{t+1}^r(V^m_r + \varepsilon b(Y^m_r)) \mid V^m_r \right] - \mathcal{M}_{t+1}^r(V^m_r) \right)^2 \right)^{1/2}
+ \sum_{r=0}^{t-1} \mathbb{E} \left[ \mathcal{M}_{t+1}^r(V^m_r + \varepsilon b(Y^m_r)) \mid V^m_r \right] - \mathcal{M}_{t+1}^r(V^m_r + \varepsilon b(V^m_r)) \right].
which, also as in that proof, is bounded above by
\[ \|V_t^m - v_t\| \leq (L_1 + L_2) \left( \varepsilon^2 \sum_{r=0}^{t-1} \alpha^{2(t-r-1)} \|b(Y_r^m) - b(V_r^m)\|^2 \right)^{1/2} \]
\[ + L_2 \varepsilon^2 \sum_{r=0}^{t-1} \alpha^{t-r} \|b(Y_r^m) - b(V_r^m)\|^2. \]

From (37) and Lemma 1 we find that
\[ \|b(Y_r^m) - b(V_r^m)\| \leq L_1 \|Y_r^m - V_r^m\|. \]

The rest of the argument proceeds exactly as in the proof of Theorem 2.

\[ \square \]

**Proof of Lemma 4.** Observe that if \( \tau_v^m > 0 \), then condition
\[ \sum_{j=\ell+1}^{n} |v_{\sigma_j}| \leq \frac{m - \ell}{m} \|v\|_1 \]
holds for \( \ell = 0 \). Assume that
\[ \sum_{j=\ell}^{n} |v_{\sigma_j}| \leq \frac{m - \ell + 1}{m} \|v\|_1 \]
for some \( \ell \leq \tau_v^m \). From the definition of \( \tau_v^m \) and the fact that \( \ell \leq \tau_v^m \), we must also have that
\[ \frac{1}{m - \ell} \sum_{j=\ell+1}^{n} |v_{\sigma_j}| < |v_{\sigma_{\ell+1}}|. \]

Combining the last two inequalities yields
\[ \sum_{j=\ell+1}^{n} |v_{\sigma_j}| \leq \frac{m - \ell}{m} \|v\|_1. \]

\[ \square \]

**Proof of Lemma 5.** First we assume that, for all \( j \), \( |v_j + w_j| \leq \|v + w\|/m \). We will remove this assumption later. With this assumption in place, \( N_j \in \{0,1\} \) and the while loop in Algorithm 1 is inactive so that
\[ f^H \Phi_t(v + w) = \sum_{j=1}^{n} \tilde{f}_j \frac{v_j + w_j}{|v_j + w_j|} \frac{\|v + w\|}{m} N_j, \]
\[ \mathbb{E}[|f^H \Phi_t(v + w) - f^H(v + w)|^2] = \frac{\|v + w\|^2}{m^2} \mathbb{E} \left[ \sum_{j=1}^{n} \tilde{f}_j \left( \frac{v_j + w_j}{|v_j + w_j|} \right)^2 \left( N_j - \frac{m|v_j - w_j|}{\|v + w\|_1} \right)^2 \right]. \]

The random variables in the sum are independent, so the last expression becomes
\[ \mathbb{E}[|f^H \Phi_t(v + w) - f^H(v + w)|^2] \leq \frac{\|v + w\|^2}{m^2} \sum_{j=1}^{n} \tilde{f}_j \mathbb{E} \left[ \left( N_j - \frac{m|v_j - w_j|}{\|v + w\|_1} \right)^2 \right]. \]

Since \( N_j \in \{0,1\} \), the expression for the variance of \( N_j \) becomes
\[ \text{var} [N_j] = \mathbb{E} [N_j] (1 - \mathbb{E} [N_j]) = \frac{m|v_j + w_j|}{\|v + w\|_1} \left( 1 - \frac{m|v_j + w_j|}{\|v + w\|_1} \right), \]
so that
\[ \mathbb{E}[|f^H \Phi_t(v + w) - f^H(v + w)|^2] \leq \frac{\|v + w\|^2}{m^2} \left[ m - \left( \frac{m}{\|v + w\|_1} \right)^2 \|v + w\|_2^2 \right]. \]

Because this scheme does not depend on the ordering of the entries of \( v + w \) we can assume that the entries have been ordered so that \( v_j = 0 \) for \( j > m \). In this case we can write
\[ \|v + w\|_2^2 = \sum_{j=1}^{m} |v_j + w_j|^2 + \sum_{j=m+1}^{n} |w_j|^2 \geq \frac{1}{m} \left( \sum_{j=1}^{m} |v_j + w_j| \right)^2, \]
which then implies that

\[
E \left[ |f^H \Phi_t(v + w) - f^H(v + w)|^2 \right] \leq \frac{\|v + w\|^2}{m} \left( 1 - \frac{1}{\|v + w\|^2} \left( \|v + w\|_1 - \sum_{j=m+1}^n |w_j| \right)^2 \right)
\]

\[
\leq \frac{2\|v + w\|_1}{m} \|v + w\|_1.
\]

We now remove the assumption that \(|v_j + w_j| \leq \|v + w\|/m\). Let \(\sigma\) be a permutation of the indices of \(v + w\) resulting in a vector \(v_{\sigma} + w_{\sigma}\) with entries of nonincreasing magnitude. Since Algorithm 1 preserves the largest \(\tau_{v+w}^{m}\) entries of \(v + w\) and the remaining entries, \(v_{\sigma} + w_{\sigma}\) for \(j > \tau_{v+w}^{m}\), satisfy

\[
|v_{\sigma_j} + w_{\sigma_j}| \leq \frac{1}{m - \tau_{v+w}^{m}} \sum_{k=\tau_{v+w}^{m}}^n |v_{\sigma_k} + w_{\sigma_k}|,
\]

we can apply the sampling error bound just proved to find that

\[
\|\Phi_t(v + w) - v - w\| \leq \frac{\sqrt{2} \left( \sum_{j=\tau_{v+w}^{m}}^n |w_j| \right) \frac{1}{2} \left( \sum_{j=\tau_{v+w}^{m}}^n |v_j + w_j| \right) \frac{1}{2}}{\sqrt{m - \tau_{v+w}^{m}}}
\]

An application of Lemma 4 then yields (43).

In bounding the size of \(\Phi_t(v + w)\) we will again assume that \(\tau_{v+w}^{m} = 0\) and that the entries have been ordered so that \(v_j = 0\) for \(j > m\). The size of the resampled vector can be bounded by first noting that, since the \(N_j\) are independent and are in \(\{0, 1\}, \)

\[
E \left[ \left( \sum_{j=1}^n N_j \right)^2 \right] = \sum_{j=1}^n \frac{m|v_j + w_j|}{\|v + w\|_1} + 2 \sum_{i=1}^n \sum_{j=i+1}^n \frac{m|v_i + w_i| m|v_j + w_j|}{\|v + w\|_1}
\]

\[
= \sum_{j=1}^n \left( \frac{m|v_j + w_j|}{\|v + w\|_1} \right)^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n \frac{m|v_i + w_i| m|v_j + w_j|}{\|v + w\|_1}
\]

\[
+ \sum_{j=1}^n \frac{m|v_j + w_j|}{\|v + w\|_1} - \left( \frac{m|v_j + w_j|}{\|v + w\|_1} \right) \frac{2}{2}
\]

\[
= m^2 + \sum_{j=1}^n \frac{m|v_j + w_j|}{\|v + w\|_1} - \left( \frac{m|v_j + w_j|}{\|v + w\|_1} \right)^2.
\]

Breaking up the last sum in this expression, we find that

\[
\sum_{j=1}^m \frac{m|v_j + w_j|}{\|v + w\|_1} - \left( \frac{m|v_j + w_j|}{\|v + w\|_1} \right)^2 \leq m \sum_{j=1}^m \frac{|v_j + w_j|}{\|v + w\|_1} - m \left( \sum_{j=1}^m \frac{|v_j + w_j|}{\|v + w\|_1} \right)^2
\]

\[
\leq m \left( 1 - \sum_{j=1}^m \frac{|v_j + w_j|}{\|v + w\|_1} \right) \leq \frac{m\|w\|_1}{\|v + w\|_1}
\]

and that

\[
\sum_{j=m+1}^n \frac{m|w_j|}{\|v + w\|_1} - \left( \frac{m|w_j|}{\|v + w\|_1} \right)^2 \leq \frac{m\|w\|_1}{\|v + w\|_1},
\]

so that

\[
E \left[ \left( \sum_{j=1}^n N_j \right)^2 \right] \leq m^2 + 2 \frac{m\|w\|_1}{\|v + w\|_1}.
\]

It follows then that (at least when \(\tau_{v+w}^{m} = 0\))

\[
E \left[ \|\Phi_t^m(v + w)\|^2 \right] \leq \|v + w\|_1 + 2 \frac{\|v + w\|_1\|w\|_1}{m}
\]

Writing the corresponding formula for \(\tau_{v+w}^{m} > 0\) and applying Lemma 4 gives the bound in the statement of the lemma.
Finally we consider the probability of the event \( \{ \Phi^{m}_i(v + w) = 0 \} \). If \( \tau^m_{v+w} = 0 \), then \( N_j \in \{0, 1\} \), so that \( P[N_j = 0] = 1 - m|v_j + w_j|/\|v + w\|_1 \), and, since the \( N_j \) are independent,

\[
P[N_j = 0 \text{ for all } j] = \prod_{j=1}^{n} \left( 1 - \frac{m|v_j + w_j|}{\|v + w\|_1} \right) \leq \prod_{j \leq n, v_j \neq 0} \left( 1 - \frac{m|v_j + w_j|}{\|v + w\|_1} \right).
\]

The first product in the last display is easily seen to be bounded above by \( e^{-m} \). The second product is maximized subject to the constraint

\[
\sum_{j \leq n, v_j \neq 0} \left( 1 - \frac{m|v_j + w_j|}{\|v + w\|_1} \right) \leq \frac{m\|w\|_1}{\|v + w\|_1}
\]

when the terms in the product are all equal, in which case we get

\[
P[N_j = 0 \text{ for all } j] \leq \left( \frac{\|w\|_1}{\|v + w\|_1} \right)^m.
\]

\( \square \)