A. Remarks

In this section we give a number of remarks relating to content within the main body of the paper.

Remark 1 (Sketching and Communication Savings) *We highlight that the Random Feature framework considered also incorporates a number of sketching techniques. For instance, when $\psi(x,\omega) = x^\top \omega$ where $\omega \sim \mathcal{N}(0, I)$ and the associated kernel is simply linear as $\mathbf{E}[\psi(x,\omega)\psi(x',\omega)]\mathbf{E}[x^\top \omega\omega^\top x] = x^\top \mathbf{E}[\omega\omega^\top]x = x^\top x'$. The case $M < D$ then represents a simple setting in which communication savings can be achieved, as agents in this case would only need to communicate an M dimensional vector instead of D. A natural future direction would be to investigate whether there exists particular sketches/Random Features tailored to the objective of communication savings, in a similar manner to Orthogonal Random Features (Yu et al., 2016), Fast Food (Le et al., 2013) or Low-precision Random Features (Zhang et al., 2019). Although, as noted in (Carratino et al., 2018), some of these methods sample the features in a correlated manner, and thus, do not fit within the assumptions of this work.*

Remark 2 (Previous Literature Decentralised Kernel Methods) *This remark highlights two previous works for Decentralised Kernel Methods. The work (Forero et al., 2010) considers decentralised Support Vector Machines with potentially high-dimensional finite feature spaces that could approximate a non-linear kernel. They develop a variant of the Alternating Direction Method of Multiplers (ADMM) to target the augmented optimisation problem. In this case, the high-dimensional constraints across the agents are approximated so the agents local estimated functions are equal on a subset of chosen points. Meanwhile (Koppel et al., 2018) consider online stochastic optimisation with penalisation between neighbouring agents. The penalisation introduced is an expectation with respect to a newly sampled data point and not in the norm of the Reproducing Kernel Hilbert Space. In both of these cases, the original optimisation problem is altered to facilitate a decentralised algorithm, but no guarantee is given on how these approximation impact statistical performance.*

Remark 3 (Concurrent Work) *The concurrent work (Xu et al., 2020) consider the homogeneous setting where a network of agents have data from the same distribution and wish to learn a function within a RKHS that performs well on unseen data. The consensus optimisation formulation of the single machine explicitly penalised kernel learning problem is considered, and the challenges of decentralised kernel learning (as described in Section 2.1 in the main body of the manuscript) are overcome by utilising Random Fourier Features. An ADMM method is developed to solve the consensus optimisation problem, and, provided hyper-parameters are tuned appropriately, optimisation guarantees are given. Due to considering the consensus optimisation formulation of a single machine penalised problem, the Generalisation Error is decoupled from the Optimisation Error. Therefore, while optimisation results for ADMM applied to consensus optimisation objectives (Shi et al., 2014) are applied, the statistical setting is not leveraged to achieve speed-ups. It is then not clear how the network connectivity, number of samples held by agents and finer statistical assumptions (source and capacity) impacts either generalisation or optimisation performance. This is in contrast to our work, where we directly study the Generalisation Error of Distributed Gradient Descent with Implicit Regularisation, and show how the number of samples held by agents, network topology, step size and number of iterations can impact Generalisation Error.*

B. Analysis Setup

This section provides the setup for the analysis. We adopt the notation of (Carratino et al., 2018), which is included here for completeness. Section B.1 introduces additional auxiliary quantities required for the analysis. Section B.2 introduces notation for the operators required for the analysis. Section B.3 introduces the error decomposition.

B.1. Additional Auxiliary Sequences

We begin by introducing some auxiliary sequences that will be useful in the analysis. Begin by defining $\{v_t\}_{t\geq 1}$ initialised at $v_1 = 0$ and updated for $t \geq 1$ and updated

$$v_{t+1} = v_t - \eta \int_X \big(\langle v_t, \phi_M(x)\rangle - f_{\mathcal{H}}(x)\big)\phi_M(x) d\rho_X(x)$$

Further for $\lambda > 0$ let

$$\widetilde{u}_\lambda = \operatorname{argmin}_{u \in \mathbb{R}^M} \int_X \big(\langle u, \phi_M(x)\rangle - f_{\mathcal{H}}(x)\big)^2 d\rho_X(x) + \lambda \|u\|^2,$$
$$u_\lambda = \operatorname{argmin}_{u \in \mathcal{F}} \int_X (\langle u, \phi(x)\rangle - y)^2 d\rho(x,y) + \lambda \|u\|^2,$$

where $(\mathcal{F}, \phi)$ are feature space and feature map associated to the kernel k. As described previously, it will be useful to work with functions in $L^2(X, \rho_X)$, therefore define the functions

$$g_t = \langle v_t, \phi_M(\cdot)\rangle, \quad \widetilde{g}_\lambda = \langle \widetilde{u}_\lambda, \phi_M(\cdot)\rangle, \quad g_\lambda = \langle u_\lambda, \phi(\cdot)\rangle.$$

The quantities introduced here in this section will be useful in analysing the *Statistical Error* term.

B.2. Notation

Let $\mathcal{F}$ be the feature space corresponding to the kernel k given by Assumption 2.

Given $\phi : X \to \mathcal{F}$ (feature map), we define the operator $S : \mathcal{F} \to L^2(X, \rho_X)$ as

$$(S\omega)(\cdot) = \langle \omega, \phi(\cdot)\rangle_{\mathcal{F}}, \quad \forall \omega \in \mathcal{F}.$$

If $S^\star$ is the adjoint operator of S, we let $C : \mathcal{F} \to \mathcal{F}$ be the linear operator $C = S^\star S$, which can be written as

$$C = \int_X \phi(x) \otimes \phi(x) d\rho_X(x).$$

We also define the linear operator $L : L^2(X, \rho_X) \to L^2(X, \rho_X)$ such that $L = SS^\star$, that can be represented as

$$(Lf)(\cdot) = \int_X \langle \phi(x), \phi(\cdot)\rangle_{\mathcal{F}} f(x) d\rho_X(x), \quad \forall f \in L^2(X, \rho_X).$$

We now define the analog of the previous operators where we use the feature map ϕ_M instead of ϕ. We have $S_M : \mathbb{R}^M \to L^2(X, \rho_X)$ defined as

$$(S_M v)(\cdot) = \langle v, \phi_M(\cdot)\rangle_{\mathbb{R}^M}, \quad \forall v \in \mathbb{R}^M$$

together with $C_M : \mathbb{R}^M \to \mathbb{R}^M$ and $L_M : L^2(X, \rho_X) \to L^2(X, \rho_X)$ defined as $C_M = S_M^\star S_M$ and $L_M = S_M S_M^\star$ respectively. For $v \in \mathbb{R}^M$ note we have the equality

$$\begin{aligned}
\|S_M v\|_\rho^2 &= \int_X \langle v, \phi_M(x)\rangle^2 d\rho_X(x) \\
&= \int_X v^\top \phi_M(x) \otimes \phi_M(x) v d\rho_X(x) \\
&= v^\top C_M v \\
&= \|C_M^{1/2} v\|^2
\end{aligned} \tag{7}$$

where we have denoted the standard Euclidean norm as $\|\cdot\|$. Define the empirical counterpart of the previous operators for each agent. For each agent $v \in V$ define the operator $\widehat{S}_M^{(v)} : \mathbb{R}^M \to \mathbb{R}^m$ as

$$\widehat{S}_M^{(v)\top} = \frac{1}{\sqrt{m}} (\phi_M(x_{1,v}), \dots, \phi_M(x_{m,v})),$$

and with $\widehat{C}_M^{(v)} : \mathbb{R}^M \to \mathbb{R}^M$ and $\widehat{L}_M^{(v)} : \mathbb{R}^m \to \mathbb{R}^m$ are defined as $\widehat{C}_M^{(v)} = \widehat{S}_M^{(v)\top} \widehat{S}_M^{(v)}$ and $\widehat{L}_M^{(v)} = \widehat{S}_M^{(v)} \widehat{S}_M^{(v)\top}$ respectively. Moreover, define the empirical operators associated to all of the samples held by agents in the network. To do so index the agents in V between 1 and n, so $x_{i,j}$ is the ith data point held by agent j. Then, define the operator $\widehat{S}_M : \mathbb{R}^M \to \mathbb{R}^{nm}$ as

$$\begin{aligned}
\widehat{S}_M^\top &= \frac{1}{\sqrt{nm}} (\phi_M(x_{1,1}), \dots, \phi_M(x_{m,1}), \phi_M(x_{1,2}), \dots, \phi_M(x_{m,2}), \dots, \phi_M(x_{1,n}), \dots, \phi_M(x_{m,n})) \\
&= \frac{1}{\sqrt{n}} (\widehat{S}_M^{(1)\top}, \dots, \widehat{S}_M^{(n)\top})
\end{aligned}$$

and with $\widehat{C}_M : \mathbb{R}^M \to \mathbb{R}^M$ and $\widehat{L}_M : \mathbb{R}^{nm} \to \mathbb{R}^{nm}$ are defined as $\widehat{C}_M = \widehat{S}_M^\top \widehat{S}_M$ and $\widehat{L}_M = \widehat{S}_M \widehat{S}_M^\top$ respectively. From the above it is clear that we have $\widehat{C}_M = \frac{1}{n}\sum_{w \in V} \widehat{S}_M^{(w)\top} \widehat{S}_M^{(w)} = \frac{1}{n}\sum_{w\in V} C_M^{(w)}$. For some number $\lambda > 0$ we let the operator plus the identity times λ be denoted $L_\lambda = L + \lambda I$, and similarly for $\widehat{L}_\lambda$, as well as $C_{M,\lambda} = C_M + \lambda I$ and $\widehat{C}_{M,\lambda}$.

Remark 4 *Let $P : L^2(X, \rho_X) \to L^2(X, \rho_X)$ be the projection operator whose range is the closure of the range of L. Let $f_\rho : X \to \mathbb{R}$ be defined as*

$$f_\rho(x) = \int y d\rho(y|x).$$

If there exists $f_{\mathcal{H}} \in \mathcal{H}$ such that

$$\inf_{f \in \mathcal{H}} \mathcal{E}(f) = \mathcal{E}(f_{\mathcal{H}})$$

then

$$Pf_\rho = Sf_{\mathcal{H}}.$$

or equivalently, there exists $g \in L^2(X, \rho_X)$ such that

$$Pf_\rho = L^{1/2} g.$$

In particular, we have $R := \|f_{\mathcal{H}}\|_{\mathcal{H}} = \|g\|_{L^2(X,\rho_X)}$. The above condition is commonly relaxed in approximation theory as

$$Pf_\rho = L^r g$$

with $1/2 \leq r \leq 1$.

With the operators introduced above and the above remark, we can rewrite the auxiliary objects respectively as

$$\begin{aligned} \widehat{v}_1 = 0; \quad & \widehat{v}_{t+1} = (I - \eta \widehat{C}_M)\widehat{v}_t + \eta \widehat{S}_M^\top \widehat{y} \\ \widetilde{v}_1 = 0; \quad & \widetilde{v}_{t+1} = (I - \eta C_M)\widetilde{v}_t + \eta S_M^\star f_\rho \\ v_1 = 0; \quad & v_{t+1} = (I - \eta C_M) v_t + \eta S_M^\star P f_\rho \end{aligned}$$

where the vector of all nm responses are $\widehat{y}^\top = (nm)^{-1/2}(y_{1,1}, \dots, y_{1,m}, y_{2,m}, \dots, y_{n,m}) = (n)^{-1/2}(\widehat{y}_1, \dots, \widehat{y}_n)$, and each agents responses are, for $i = 1, \dots, n$, denoted $\widehat{y}_v = (m)^{-1/2}(y_{i,1} \dots, y_{i,m})$. We then denote

$$\begin{aligned} \widetilde{u}_\lambda &= S_M^\star L_{M,\lambda}^{-1} P f_\rho \\ u_\lambda &= S^\star L_\lambda^{-1} P f_\rho. \end{aligned}$$

Inductively the three sequences can be written as

$$\widehat{v}_{t+1} = \sum_{k=1}^{t} \eta (I - \eta \widehat{C}_M)^{t-k} \widehat{S}_M^\top \widehat{y}$$

$$\widetilde{v}_{t+1} = \sum_{k=1}^{t} \eta (I - \eta C_M)^{t-k} S_M^\star f_\rho$$

$$v_{t+1} = \sum_{k=1}^{t} \eta (I - \eta C_M)^{t-k} S_M^\star P f_\rho$$

B.3. Error Decomposition

We can now write the deviation $\widehat{f}_{t+1,v} - f_{\mathcal{H}}$ using the operators

$$\widehat{f}_{t+1,v} - f_{\mathcal{H}} = \underbrace{S_M \widehat{\omega}_{t+1,v} - S_M \widehat{v}_t}_{\text{Network Error}} + \underbrace{S_M \widehat{v}_t - P f_\rho}_{\text{Statistical Error}} \tag{8}$$

where the first term aligns with the network error and the second with the statistical error. Each of these will be analysed in it own section.

C. Statistical Error

In this section we summarise the analysis for the Statistical Error which has been conducted within (Carratino et al., 2018). Here we provided the proof for completeness. Firstly, we further decompose the statistical error into the following terms

$$\|S_M\widehat{v}_{t+1} - Pf_\rho\|_\rho \leq \underbrace{\|S_M\widehat{v}_{t+1} - S_M\widetilde{v}_{t+1} + S_M\widetilde{v}_{t+1} - S_M v_t\|_\rho}_{\text{Sample Error}} + \underbrace{\|S_M v_{t+1} - L_M L_{M,\lambda}^{-1} Pf_\rho\|_\rho}_{\text{Gradient Descent and Ridge Regression}} \tag{9}$$

$$+ \underbrace{\|L_M L_{M,\lambda}^{-1} Pf_\rho - LL_\lambda^{-1} Pf_\rho\|_\rho}_{\text{Random Features Error}} + \underbrace{\|LL_\lambda^{-1}Pf_\rho - Pf_\rho\|_\rho}_{\text{Bias}}$$

Each of the terms have been labelled to help clarity. The first term, *sample error* includes the difference between the empirical iterations with sampled data $\widehat{v}_t$, as well as iterates under the population measure v_t. The second term *Gradient Descent and Ridge Regression* is the difference between the population variants of the Gradient Descent v_t and ridge regression $L_M L_{M,\lambda}^{-1} Pf_\rho$ solutions. The third term *Random Feature Error* accounts for the error introduced from using Random Features. Finally the *Bias* term accounts for the bias introduced due to the regularisation. Each of these terms will be bounded within their own sub-section, except the *Bias* term which will be bounded when bounds for all of the terms are brought together.

The remainder of this section is then as follows. Section C.1, C.2 and C.3 give the analysis for the Sample Error, Gradient Descent and Ridge Regression and Random Feature Error error respectively. Section C.4 bounds the Bias and combines bounds for the previous terms.

C.1. Sample Error

The bound for this term is summarised within the following Lemma which itself comes from Lemma 1 and 6 in (Carratino et al., 2018).

Lemma 1 (Sample Error) *Under assumptions 2, 4 and 3 , let $\delta \in (0,1)$, $\eta \in (0, \kappa^{-2})$. When*

$$M \geq \big(4 + 18\eta t\big) \log \frac{12\eta t}{\delta}$$

for all $t \geq 1$ with probability atleast $1 - 3\delta$

$$\|S_M\widehat{v}_t - S_M\widetilde{v}_t + S_M\widetilde{v}_t - S_M v_t\|_\rho \leq 4\Big(R\kappa^{2r}\Big(1 + \sqrt{\frac{9}{M}\log\frac{M}{\delta}}\big(\sqrt{\eta t} \vee 1\big)\Big) + \sqrt{B}\Big)$$
$$\times \big(12 + 4\log(t) + \sqrt{2}\eta\big)\Big(\frac{\sqrt{\eta t}}{nm} + \frac{\sqrt{2\sqrt{p}q_0\mathcal{N}(\frac{\kappa^2}{\eta t})}}{\sqrt{nm}}\Big)\log\frac{4}{\delta}$$

where $q_0 = \max\big(2.55, \frac{2\kappa^2}{\|L\|}\big)$

Proof 1 *Apply Lemma 1 in (Carratino et al., 2018) to say $\|S_M\widetilde{v}_t - S_M v_t\|_\rho = 0$, meanwhile Lemma 6 in the same work to bound $\|S_M\widehat{v}_t - S_M\widetilde{v}_t\|$ with $\theta = 0$ and $T = t$.*

C.2. Gradient Descent and Ridge Regression

This term is controlled by Lemma 9 in (Carratino et al., 2018).

Lemma 2 (Gradient Descent and Ridge Regression) *Under Assumption 3 the following holds with probability $1 - \delta$ for $\lambda = \frac{1}{\eta t}$ for $t \geq 1$*

$$\|S_M v_{t+1} - L_M L_{M,\lambda}^{-1} Pf_\rho\|_\rho \leq 8R\kappa^{2r}\Big(\frac{\log\frac{2}{\delta}}{M^r} + \sqrt{\frac{\mathcal{N}\big(\frac{1}{\eta t}\big)^{2r-1}\log\frac{2}{\delta}}{M(\eta t)^{2r-1}}}\Big)\log^{1-r}\big(11\kappa^2\eta t\big) + \frac{2R}{(\eta t)^r}$$

when

$$M \geq (4 + 18\eta t)\log\big(\frac{8\kappa^2\eta t}{\delta}\big)$$

C.3. Random Features Error

The following Lemma is from Lemma 8 of (Rudi & Rosasco, 2017; Carratino et al., 2018).

Lemma 3 *Under assumption 2 and 3 for any $\lambda > 0$, $\delta \in (0, 1/2]$, when*

$$M \geq \big(4 + \frac{18\kappa^2}{\lambda}\big) \log \frac{8\kappa^2}{\lambda\delta}$$

the following holds with probability at least $1 - 2\delta$

$$\|L_M L_{M,\lambda}^{-1} P f_\rho - L L_\lambda^{-1} P f_\rho\|_\rho \leq 4R\kappa^{2r} \left(\frac{\log \frac{2}{\delta}}{M^r} + \sqrt{\frac{\lambda^{2r-1}\mathcal{N}(\lambda)^{2r-1} \log \frac{2}{\delta}}{M}} \right) q^{1-r}$$

where $q = \log \frac{11\kappa^2}{\lambda}$

C.4. Combined Error Bound

The following Lemma combines the error bounds.

Lemma 4 *Under assumption 1 to 4, let $\delta \in (0, 1)$ and $\eta \in (0, \kappa^{-2})$ when*

$$M \geq (4 + 18\eta t \kappa^2) \log \frac{60\kappa^2 \eta t}{\delta}$$

the following holds with probability greater than $1 - \delta$

$$\|S_M \widehat{v}_{t+1} - P f_\rho\|_\rho^2 \leq c_1^2 \Big(1 \vee \frac{(\eta t \vee 1) \log \frac{3M}{\delta}}{M}\Big)\Big(\frac{\eta t}{(nm)^2} \vee \frac{\mathcal{N}(\frac{1}{\eta t})}{nm}\Big) \log^2(t) \log^2 \frac{12}{\delta}$$
$$+ c_2^2 \Big(\frac{1}{M^{2r}} \vee \frac{\mathcal{N}(\frac{1}{\eta t})^{2r-1}}{M(\eta t)^{2r-1}}\Big) \log^{2(1-r)}(11\kappa^2 \eta t) \log^2 \big(\frac{6}{\delta}\big) + \frac{c_3^2}{(\eta t)^{2r}}$$

where the constants

$$c_1 = 8 \times 12 \times 15\big(\sqrt{B} \vee (R\kappa^{2r})\big)(1 \vee \sqrt{2\sqrt{p} q_0})$$
$$c_2 = 24R\kappa^{2r}$$
$$c_3 = 3R$$

Proof 2 (Lemma 1) *Begin fixing $\lambda = \frac{1}{\eta t}$ and bounding the bias from Lemma 5 of (Rudi & Rosasco, 2017) as*

$$\|L L_\lambda^{-1} P f_\rho - P f_\rho\|_\rho \leq R\lambda^r.$$

Now use Lemma 1 to bound the Sample Error, Lemma 2 for the Gradient Descent and Ridge Regression Term, and 3 for the Random Features Error. With a union bound, note that the conditions on M for each of these Lemmas is satisfied by $M \geq (4 + 18\eta t\kappa^2) \log \frac{60\kappa^2 \eta t}{\delta}$. Cleaning up constants and squaring then yields the bound.

D. Network Error

In this section we the proof of the following bound on the network error, which improves upon (Richards & Rebeschini, 2019). This section is then structured as follows. Section D.1 provides the error decomposition for the Network Error. Section D.2 introduces a number of prelimary lemmas utilised within the analysis. Section D.3, D.4, D.5, D.6 and D.7 then provides bounds for each of the error terms that arise within the decomposition.

D.1. Error Decomposition

Recall the vector of observations associated to agent $v \in V$ is denoted $\widehat{y}_v = \frac{1}{\sqrt{m}}(y_{1,v}, \dots, y_{m,v})$. Using the previously introduced notation note that we can write the Distributed Gradient Descent iterates as for $t \geq 1$ and $v \in V$

$$\widehat{\omega}_{t+1,v} = \sum_{w \in V} P_{vw}\Big(\widehat{\omega}_{t,w} - \eta \widehat{C}_M^{(w)} \widehat{\omega}_{t,w} + \eta \widehat{S}_M^{(w)\top} \widehat{y}_w\Big)$$

Centering the iterates around the population sequence $\widetilde{v}_t$ we have from the doubly stochastic property of P

$$\begin{aligned}
\widehat{\omega}_{t+1,v} - \widetilde{v}_{t+1} &= \sum_{w \in V} P_{vw}\Big(\widehat{\omega}_{t,w} - \widetilde{v}_t + \eta\big\{(C_M \widetilde{v}_t - S_M^{\star} f_\rho) - (\widehat{C}_M^{(w)} \widehat{\omega}_{t,w} + \widehat{S}_M^{(w)\top} \widehat{y}_w)\big\}\Big) \\
&= \sum_{w \in V} P_{vw}\Big((I - \widehat{C}_M^{(w)})(\widehat{\omega}_{t,w} - \widetilde{v}_t) + \eta \underbrace{\big\{(C_M \widetilde{v}_t - S_M^{\star} f_\rho) - (\widehat{C}_M^{(w)} \widetilde{v}_t + \widehat{S}_M^{(w)\top} \widehat{y}_w)\big\}}_{N_{t,w}}\Big) \\
&= \sum_{w \in V} P_{vw}\Big((I - \widehat{C}_M^{(w)})(\widehat{\omega}_{t,w} - \widetilde{v}_t) + \eta N_{t,w}\Big)
\end{aligned}$$

where we have defined the error term

$$N_{t,w} := (C_M \widetilde{v}_t - S_M^{\star} f_\rho) - (\widehat{C}_M^{(w)} \widetilde{v}_t + \widehat{S}_M^{(w)\star} \widehat{y}_w) \quad \forall s \geq 1\, w \in V.$$

Note that a similar set of calculation can be performed for the iterates $\widehat{v}_t$ leading to the recursion for $v \in V$ initialised at $\widehat{v}_{1,v} = 0$ and updated for $t \geq 1$

$$\widehat{v}_{t+1,v} - \widetilde{v}_{t+1} = \sum_{w \in V} \frac{1}{n}\Big((I - \widehat{C}_M^{(w)})(\widehat{v}_{t,w} - \widetilde{v}_t) + \eta N_{t,w}\Big)$$

For a path indexed from time step t to k such that $1 \leq k \leq t$ as $w_{t:k} = (w_t, w_{t-1}, \dots, w_k) \in V^{t-k+1}$, let the product of operators be denoted

$$\Pi(w_{t:k}) = (I - \widehat{C}_M^{(w_t)})(I - \widehat{C}_M^{(w_{t-1})}) \dots (I - \widehat{C}_M^{(w_k)}) \tag{10}$$

Meanwhile for $k > t$ we say $\Pi(w_{t:k}) = I$. Unravelling the sequences $\widehat{\omega}_{t+1,v} - \widetilde{v}_{t+1}$ and $\widehat{v}_{t+1} - \widetilde{v}_{t+1}$ with the above notation and taking the difference we then have

$$\begin{aligned}
\widehat{\omega}_{t+1,v} - \widehat{v}_{t+1} &= \sum_{k=1}^{t} \eta \sum_{w_{t:k} \in V^{t-k+1}} \big(P_{vw_{t:k}} - \frac{1}{n^{t-k+1}}\big)\Pi(w_{t:k+1}) N_{k,w_k} \\
&= \sum_{k=1}^{t} \eta \sum_{w_{t:k} \in V^{t-k+1}} \Delta(w_{t:k}) \Pi(w_{t:k+1}) N_{k,w_k}
\end{aligned}$$

where we have introduced the notation where we have denoted $\big(P_{vw_{t:k}} - \frac{1}{n^{t-k+1}}\big) = \Delta(w_{t:k}) \in \mathbb{R}$. Introduce notation for the difference between the product of operators indexed by the paths and the population equivalent

$$\Pi^{\Delta}(w_{t:k+1}) := \Pi(w_{t:k+1}) - (I - \eta C_M)^{t-k}.$$

Fixing some $t^\star \in \mathbb{N}$ and supposing that $t > 2t^\star \geq 2$, observe that we can then write, for $k \leq t - t^\star - 1$,

$$\begin{aligned}
&\Pi^{\Delta}(w_{t:k+1}) \\
&= \Pi(w_{t:k+1}) - \Pi(w_{t:k+t^\star+1})(I - \eta C_M)^{t^\star} + \Pi(w_{t:k+t^\star+1})(I - \eta C_M)^{t^\star} - (I - \eta C_M)^{t-k} \\
&= \Pi(w_{t:k+t^\star+1})\Pi^{\Delta}(w_{k+t^\star:k+1}) + \Pi^{\Delta}(w_{t:k+t^\star+1})(I - \eta C_M)^{t^\star}
\end{aligned}$$

where we have replaced the first $t^\star$ operators in $\Pi(w_{t:k})$ with the population variant $(I-\eta C_M)$. Plugging this in then yields

$$\begin{aligned}\widehat{\omega}_{t+1,v}-\widehat{v}_{t+1} &= \sum_{k=1}^{t}\eta \sum_{w_{t:k}\in V^{t-k+1}} \Delta(w_{t:k})(I-\eta C_M)^{t-k}N_{k,w_k} \\ &+ \sum_{k=t-2t^\star}^{t}\eta\sum_{w_{t:k}\in V^{t-k+1}}\Delta(w_{t:k})\Pi^{\Delta}(w_{t:k+1})N_{k,w_k} \\ &+ \sum_{k=1}^{t-2t^\star-1}\eta\sum_{w_{t:k}\in V^{t-k+1}}\Delta(w_{t:k})\Pi(w_{t:k+t^\star+1})\Pi^{\Delta}(w_{k+t^\star:k+1})N_{k,w_k} \\ &+ \sum_{k=1}^{t-2t^\star-1}\eta\sum_{w_{t:k}\in V^{t-k+1}}\Delta(w_{t:k})\Pi^{\Delta}(w_{t:k+t^\star+1})(I-\eta C_M)^{t^\star}N_{k,w_k}\end{aligned}$$

where we split the series off for paths shorter than $2t^\star$. Note for the first and last term above, elements in the series can be simplified by summing over the nodes in the path. Defining for $s\geq 1$ and $v,w\in V$ the difference $\Delta^s(v,w)=P^s_{vw}-\frac{1}{n}$, we get for the first term when $k<t$

$$\begin{aligned}\sum_{w_{t:k}\in V^{t-k+1}}\Delta(w_{t:k})(I-\eta C_M)^{t-k}N_{k,w_k} &= \sum_{w_k\in V}\Big(\sum_{w_{t:k+1}\in V^{t-k}}\Delta(w_{t:k})\Big)(I-\eta C_M)^{t-k}N_{k,w_k} \\ &= \sum_{w\in V}\Delta^{t-k}(v,w)(I-\eta C_M)^{t-k}N_{k,w}\end{aligned}$$

where $\sum_{w_{t:k+1}\in V^{t-k}}\Delta(w_{t:k})=\sum_{w_{t:k+1}\in V^{t-k}}P_{vw_{t:k}}-\sum_{w_{t:k+1}\in V^{t-k}}\frac{1}{n^{t-k+1}}=P^{t-k}_{vw}-\frac{1}{n}=\Delta^{t-k}(v,w)$. Meanwhile for the last term we can sum over the last $t^\star$ nodes in the path $w_{t:k}$, that is with

$$\begin{aligned}\sum_{w_{k+t^\star:k+1}\in V^{t^\star}}\Delta(w_{t:k}) &= \sum_{w_{k+t^\star:k+1}\in V^{t^\star}}P_{vw_{t:k}}-\frac{1}{n^{t-k+1}} \\ &= P_{vw_{t:k+t^\star+1}}\sum_{w_{k+t^\star:k+1}\in V^{t^\star}}P_{w_{k+t^\star+1:k}}-\sum_{w_{k+t^\star:k+1}\in V^{t^\star}}\frac{1}{n^{t-k+1}} \\ &= P_{vw_{t:k+t^\star+1}}(P^{t^\star})_{w_{k+t^\star+1}w_k}-\frac{1}{n^{t-t^\star-k+1}} \\ &= P_{vw_{t:k+t^\star+1}}((P^{t^\star})_{w_{k+t^\star+1}w_k}-\frac{1}{n})+\frac{1}{n}(P_{vw_{t:k+t^\star+1}}-\frac{1}{n^{t-k-t^\star}}) \\ &= P_{vw_{t:k+t^\star+1}}\Delta^{t^\star}(w_{k+t^\star+1},w_k)+\frac{1}{n}\Delta(w_{t:k+t^\star+1})\end{aligned}$$

Plugging this in we get for $1\leq k\leq t-2t^\star-1$

$$\begin{aligned}&\sum_{w_{t:k}\in V^{t-k+1}}\Delta(w_{t:k})\Pi^{\Delta}(w_{t:k+t^\star+1})(I-\eta C_M)^{t^\star}N_{k,w_k} \\ &= \sum_{w_k\in V}\sum_{w_{t:k+t^\star+1}\in V^{t-t^\star-k}}\Big(\sum_{w_{k+t^\star:k+1}\in V^{t^\star}}\Delta(w_{t:k})\Big)\Pi^{\Delta}(w_{t:k+t^\star+1})(I-\eta C_M)^{t^\star}N_{k,w_k} \\ &= \sum_{w_k\in V}\sum_{w_{t:k+t^\star+1}\in V^{t-t^\star-k}}P_{vw_{t:k+t^\star+1}}\Delta^{t^\star}(w_{k+t^\star+1},w_k)\Pi^{\Delta}(w_{t:k+t^\star+1})(I-\eta C_M)^{t^\star}N_{k,w_k} \\ &\quad+\frac{1}{n}\sum_{w_k\in V}\sum_{w_{t:k+t^\star+1}\in V^{t-t^\star-k}}\Delta(w_{t:k+t^\star+1})\Pi^{\Delta}(w_{t:k+t^\star+1})(I-\eta C_M)^{t^\star}N_{k,w_k} \\ &= \sum_{w_k\in V}\sum_{w_{t:k+t^\star+1}\in V^{t-t^\star-k}}P_{vw_{t:k+t^\star+1}}\Delta^{t^\star}(w_{k+t^\star+1},w_k)\Pi^{\Delta}(w_{t:k+t^\star+1})(I-\eta C_M)^{t^\star}N_{k,w_k} \\ &\quad+\sum_{w_{t:k+t^\star+1}\in V^{t-t^\star-k}}\Delta(w_{t:k+t^\star+1})\Pi^{\Delta}(w_{t:k+t^\star+1})(I-\eta C_M)^{t^\star}N_k\end{aligned}$$

where at the end for the second term we have

$$\frac{1}{n}\sum_{w_k \in v} N_{k,w_k} = N_k = (C_M \widetilde{v}_t - \mathcal{S}_M^\star f_\rho) - (\widehat{C}_M \widetilde{v}_t + \widehat{S}_M^\top \widehat{y}) \quad \forall k \geq 1.$$

Plugging the above in, using the isometry property (7) and triangle inequality we get

$$\begin{aligned}
\|S_M(\widehat{\omega}_{t+1,v} - \widehat{v}_{t+1})\|_\rho &\leq \sum_{k=1}^{t} \eta \sum_{w \in V} |\Delta^{t-k}(v,w)| \|C_M^{1/2}(I - \eta C_M)^{t-k} N_{k,w}\| \\
&+ \sum_{k=t-2t^\star}^{t} \eta \sum_{w_{t:k} \in V^{t-k+1}} |\Delta(w_{t:k})| \|C_M^{1/2} \Pi^\Delta(w_{t:k+1}) N_{k,w_k}\| \\
&+ \sum_{k=1}^{t-2t^\star-1} \eta \sum_{w_{t:k} \in V^{t-k+1}} |\Delta(w_{t:k})| \|C_M^{1/2} \Pi(w_{t:k+t^\star+1}) \Pi^\Delta(w_{k+t^\star:k+1}) N_{k,w_k}\| \\
&+ \sum_{k=1}^{t-2t^\star-1} \eta \sum_{w_k \in V} \sum_{w_{t:k+t^\star+1} \in V^{t-t^\star-k}} |P_{v w_{t:k+t^\star+1}} \Delta^{t^\star}(w_{k+t^\star+1}, w_k)| \\
&\qquad \times \|C_M^{1/2} \Pi^\Delta(w_{t:k+t^\star+1})(I - \eta C_M)^{t^\star} N_{k,w_k}\| \\
&+ \sum_{k=1}^{t-2t^\star-1} \eta \Big\| \sum_{w_{t:k+t^\star+1} \in V^{t-t^\star-k}} \Delta(w_{t:k+t^\star+1}) C_M^{1/2} \Pi^\Delta(w_{t:k+t^\star+1})(I - \eta C_M)^{t^\star} N_k \Big\| \\
&= \mathbf{E_1} + \mathbf{E_2} + \mathbf{E_3} + \mathbf{E_4} + \mathbf{E_5}
\end{aligned} \tag{11}$$

where we have respectively labelled the error terms $\mathbf{E_i}$ for $i = 1, \dots, 5$. We will aim to construct high probability bounds for each of these error terms within the following sections. This will rely on utilising the mixing properties of P to control the deviations $\Delta^s(v,w)$ for some $s \geq 1$ and $v, w \in V$, the contractive property of operators $C_M^{1/2}(I - \eta C_M)^k$ for some $k \in \mathbb{N}_+$ as well as concentration of the error terms $N_{k,w}$ and N_k for $k \geq 1$ and $w \in V$. These are summarised within the following section.

D.2. Preliminary Lemmas

In this section we provide some Lemmas that will be useful for later. We begin with the following that bounds the deviation $\Delta^s(v,w)$ in terms of the second largest eigenvalue in absolute value of P.

Lemma 5 (Spectral Bound) *Let $s \geq 1$, $v \in V$. Then the following holds*

$$\sum_{w \in V} |\Delta^s(v,w)| \leq 2(\sqrt{n}\sigma_2^s \wedge 1)$$

Proof 3 (Lemma 5) *Let $e_v \in \mathbb{R}^n$ denoting the standard basis with a 1 in the place associated to agent v. Observe that we can write the deviation in terms of the ℓ_1 norm $\sum_{w \in V} |\Delta^s(v,w)| = \|e_v^\top P^s - \frac{1}{n}\mathbf{1}\|_1$. We immediately have an upper bound from triangle inequality that $\sum_{w \in V} |\Delta^s(v,w)| \leq \| \|e_v^\top P^s\|_1 + \|\frac{1}{n}\mathbf{1}\|_1 = 2$. Meanwhile, we can also go to the ℓ_2 norm and bound*

$$\|e_v^\top P^s - \frac{1}{n}\mathbf{1}\|_1 \leq \sqrt{n}\|e_v^\top P^s - \frac{1}{n}\mathbf{1}\|_2 \leq \sqrt{n}\sigma_2^s.$$

The bound is arrived at by taking the maximum between the two upper bounds.

The following Lemma bonds the norm of contractions

Lemma 6 (Contraction) *Let $\mathcal{L}$ be a compact, positive operator on a separable Hilbert Space H. Assume that $\eta\|\mathcal{L}\| \leq 1$. For $t \in \mathbb{N}$, $a > 0$ and any non-negative integer $k \leq t-1$ we have*

$$\|(I - \eta\mathcal{L})^{t-k}\mathcal{L}^a\| \leq \left(\frac{1}{\eta(t-k)}\right)^a.$$

Proof 4 (Lemma 6) *The proof in Lemma 15 of (Lin & Rosasco, 2017) considers this result with $a = r$. The proof for more general $a > 0$ follows the same steps.*

The following remark will summarise how the above Lemma is applied to control series of contractions.

Remark 5 (Lemma 6) *Lemma 6 will be applied to control series of the form $\eta \sum_{k=1}^{t} \|(I - \eta\mathcal{L})^{t-k}\mathcal{L}^a\|$ for some $t \geq 3$, most notably with powers $a = 1, 1/2$. In the case $a = 1$ we immediately have the bound*

$$
\begin{aligned}
\eta \sum_{k=1}^{t} \|(I - \eta\mathcal{L})^{t-k}\mathcal{L}\| &= \eta \sum_{k=1}^{t-1} \|(I - \eta\mathcal{L})^{t-k}\mathcal{L}\| + \eta\|\mathcal{L}\| \\
&\leq \eta \sum_{k=1}^{t-1} \frac{1}{\eta(t-k)} + \eta\|\mathcal{L}\| \\
&\leq 5\log(t)
\end{aligned}
$$

where we have bounded the series $\sum_{k=1}^{t-1} \frac{1}{t-k} \leq 4\log(t)$ and $\eta\|\mathcal{L}\| \leq 1$. Similarly for $a = 1/2$ we have

$$
\begin{aligned}
\eta \sum_{k=1}^{t} \|(I - \eta\mathcal{L})^{t-k}\mathcal{L}^{1/2}\| &\leq \eta \sum_{k=1}^{t-1} \frac{1}{\sqrt{\eta(t-k)}} + \eta\|\mathcal{L}^{1/2}\| \\
&\leq 3\sqrt{\eta t} + \sqrt{\eta} \\
&\leq 5\sqrt{\eta t}
\end{aligned}
$$

where we have bounded the series $\sum_{k=1}^{t-1} \frac{1}{\sqrt{(t-k)}} \leq 4\sqrt{t}$, see for instance Lemma 23 in (Richards & Rebeschini, 2019) with $q = 0$, as well as the bound that $\sqrt{\eta}\|L^{1/2}\| \leq 1$.

Now for $\lambda > 0$ define the effective dimension associated the feature map ϕ_M, that is

$$
\mathcal{N}_M(\lambda) := \mathrm{Tr}\big((L_M + \lambda I)^{-1} L_M\big).
$$

Given this, the following Lemma summarises the concentration results used within our analysis.

Lemma 7 (Concentration of Error) *Let $\delta \in (0, 1]$, $n, m, M \in \mathbb{N}_+$, $\lambda > 0$ and $\eta\kappa^2 \leq 1$. Under assumption 2,3 and 4 we have with probability greater than $1 - \delta$ for $1 \leq k \leq t$*

$$
\begin{aligned}
&\max_{w \in V} \|C_{M,\lambda}^{-1/2}(C_M - \widehat{C}_M^{(w)})\| \leq 2\kappa\Big(\frac{2\kappa}{m\sqrt{\lambda}} + \sqrt{\frac{\mathcal{N}_M(\lambda)}{m}}\Big)\log\frac{6n}{\delta} \\
&\max_{w \in V} \|C_{M,\lambda}^{-1/2} N_{k,w}\| \leq 2\sqrt{B}\Big(\frac{\kappa}{\sqrt{\lambda}m} + \sqrt{\frac{2\sqrt{p}\mathcal{N}_M(\lambda)}{m}}\Big)\log\frac{6n}{\delta} \\
&\quad + 4\kappa\Big(\frac{2\kappa}{m\sqrt{\lambda}} + \sqrt{\frac{\mathcal{N}_M(\lambda)}{m}}\Big)\Big(1 + \sqrt{\frac{9}{M}\log\frac{3Mn}{\delta}}\big(\sqrt{\eta t}\kappa \vee 1\big)\Big)\log\frac{6n}{\delta}
\end{aligned}
$$

Meanwhile, under the same assumptions with probability greater than $1 - \delta$ for $k \geq 1$

$$
\begin{aligned}
&\|C_{M,\lambda}^{-1/2}(C_M - \widehat{C}_M)\| \leq 2\kappa\Big(\frac{2\kappa}{nm\sqrt{\lambda}} + \sqrt{\frac{\mathcal{N}_M(\lambda)}{nm}}\Big)\log\frac{2}{\delta} \\
&\|C_{M,\lambda}^{-1/2} N_k\| \leq 2\sqrt{B}\Big(\frac{\kappa}{\sqrt{\lambda}nm} + \sqrt{\frac{2\sqrt{p}\mathcal{N}_M(\lambda)}{nm}}\Big)\log\frac{6}{\delta} \\
&\quad + 4\kappa\Big(\frac{2\kappa}{nm\sqrt{\lambda}} + \sqrt{\frac{\mathcal{N}_M(\lambda)}{nm}}\Big)\Big(1 + \sqrt{\frac{9}{M}\log\frac{3M}{\delta}}\big(\sqrt{\eta t}\kappa \vee 1\big)\Big)\log\frac{6}{\delta}
\end{aligned}
$$

The proof for this result is given in Section F.1. Lemma 7 will be used extensively within the following analysis. To save on the burden of notation we define the following two functions for $\lambda > 0$, $K \in \mathbb{N}_+$ and $\delta \in (0,1]$

$$
\begin{aligned}
g(\lambda, K) &= 2\kappa\Big(\frac{2\kappa}{K\sqrt{\lambda}} + \sqrt{\frac{\mathcal{N}_M(\lambda)}{K}}\Big) \\
f(\lambda, K, \delta) &= 2\sqrt{B}\Big(\frac{\kappa}{\sqrt{\lambda}K} + \sqrt{\frac{2\sqrt{p}\mathcal{N}_M(\lambda)}{K}}\Big) \\
&\quad + 4\kappa\Big(\frac{2\kappa}{K\sqrt{\lambda}} + \sqrt{\frac{\mathcal{N}_M(\lambda)}{K}}\Big)\Big(1 + \sqrt{\frac{9}{M}\log\frac{3M}{\delta}}\big(\sqrt{\eta t\kappa} \vee 1\big)\Big).
\end{aligned}
$$

Looking to Lemma 7 we note the function g is associated to the high probability bound on the difference between the covariance operators, for instance $C_{M,\lambda}^{-1/2}(C_M - \widehat{C}_M)$, meanwhile f is associated to the bound on the error terms, for instance $C_{M,\lambda}^{-1/2}N_k$.

D.3. Bounding $\mathbf{E_1}$

The bound for $\mathbf{E_1}$ is then summarised within the following Lemma.

Lemma 8 (Bounding $\mathbf{E_1}$) *Let $\delta \in (0,1]$, $n, m, M \in \mathbb{N}_+$ and $\eta\kappa^2 \leq 1$ and $t \geq 2t^\star \geq 2$ and $\lambda, \lambda' > 0$. Under assumption 2,3 and 4 we have with probability greater than $1-\delta$*

$$
\boldsymbol{E}_1 \leq \Big(\|C_{M,\lambda'}^{1/2}\|\sigma_2^{t^\star} t\kappa^{-1} f(\lambda', m, \delta/(2n)) + 20\log(t^\star)(1 \vee \sqrt{\lambda\eta t^\star}) f(\lambda, m, \delta/(2n))\Big)\log\frac{12n}{\delta}
$$

Proof 5 (Lemma 8) *Splitting the series at $1 \leq k \leq t - t^\star$ we have the following*

$$
\begin{aligned}
\boldsymbol{E}_1 \leq &\Big(\max_{1\leq k\leq t, w\in V}\|N_{k,w}\|\Big)\underbrace{\sum_{k=1}^{t-t^\star}\eta\sum_{w\in V}|\Delta^{t-k}(v,w)|\|C_M^{1/2}(I-\eta C_M)^{t-k}\|}_{\boldsymbol{E}_{11}} \\
&+ \Big(\max_{1\leq k\leq t, w\in V}\|C_{M,\lambda}^{-1/2}N_{k,w}\|\Big)\underbrace{\sum_{k=t-t^\star+1}^{t}\eta\sum_{w\in V}|\Delta^{t-k}(v,w)|\|C_M^{1/2}(I-\eta C_M)^{t-k}C_{M,\lambda}^{1/2}\|}_{\boldsymbol{E}_{12}}
\end{aligned}
$$

To bound $\boldsymbol{E}_{11}$ utilise the mixing properties of the matrix P through Lemma 5. With $\eta\kappa^2 \leq 1$ ensuring that $\eta\|C_M^{1/2}(I - \eta C_M)^{t-k}\| \leq \eta\|C_M^{1/2}\| \leq \sqrt{\eta} \leq \kappa^{-1}$, we arrive at the bound

$$
\boldsymbol{E}_{11} \leq \kappa^{-1}\sum_{k=1}^{t-t^\star}\sigma_2^{t-k} \leq \sigma_2^{t^\star} t\kappa^{-1}.
$$

Meanwhile to bound $\boldsymbol{E}_{12}$ utilise the contraction of the gradients, that is Lemma 6 remark with $a = 1/2$ and $\mathcal{L} = C_M$. With $\sum_{w\in V}|\Delta^{t-k}(v,w)| \leq 2$ this allows us to say

$$
\begin{aligned}
\boldsymbol{E}_{12} &\leq 2\eta\sum_{k=t-t^\star+1}^{t}\|C_M(I-\eta C_M)^{t-k}\| + 2\eta\sqrt{\lambda}\sum_{k=t-t^\star+1}^{t}\|C_M^{1/2}(I-\eta C_M)^{t-k}\| \\
&\leq 20\log(t^\star)(1 \vee \sqrt{\lambda\eta t^\star}).
\end{aligned}
$$

Bounding $\max_{1\leq k\leq t, w\in V}\|N_{k,w}\| \leq \|C_{M,\lambda'}^{1/2}\|\max_{1\leq k\leq t,w\in V}\|C_{M,\lambda'}^{-1/2}N_{k,w}\|$ and plugging in high probability bounds for both $\max_{1\leq k\leq t,w\in V}\|C_{M,\lambda'}^{-1/2}N_{k,w}\|$ and $\max_{1\leq k\leq t,w\in V}\|C_{M,\lambda}^{-1/2}N_{k,w}\|$ from Lemma 7 yields the result.

D.4. Bounding $\mathbf{E}_2$

The bound for this term utilises the following Lemma to bound operator $\|C_M^{1/2}\Pi^{\Delta}(w_{t:k})\|$. To save on national burden, we define the following random quantity for $\lambda > 0$

$$\Delta_\lambda := \max_{v \in V} \|C_M^{-1/2}(C_M - \widehat{C}_M^{(v)})\|.$$

We begin with the following Lemma which rewrites the norm of $\Pi^{\Delta}(w_{t:1})$ for any path $w_{t:1}$ as a series of contractions.

Lemma 9 *Let $N \in \mathbb{R}^M$ and $w_{t:1} \in V^t$ and $\eta\kappa^2 \leq 1$. Then for $u \in [0, 1/2]$*

$$\|C_M^{1/2-u}\Pi^{\Delta}(w_{t:1})N\| \leq 2\eta\Delta_\lambda\|N\| \sum_{\ell=1}^{t} \|C_M^{1/2-u}(I - \eta C_M)^{t-\ell}C_{M,\lambda}^{1/2}\|$$

Given this Lemma we present the high probability bound for $\mathbf{E}_2$.

Lemma 10 (Bounding $\mathbf{E}_2$) *Let $\delta \in (0, 1]$, $n, m, M \in \mathbb{N}_+$ and $\eta\kappa^2 \leq 1$ and $t \geq 2t^\star \geq 2$ and $\lambda, \lambda' > 0$. Under assumption 2,3 and 4 we have with probability greater than $1 - \delta$*

$$\boldsymbol{E}_2 \leq 40\kappa\|C_{M,\lambda'}^{1/2}\|\eta t^\star \log(t)(1 \vee \sqrt{\lambda\eta t}) \log^2 \frac{12n}{\delta} g(\lambda, m) f(\lambda', m, \delta/(2n))$$

Proof 6 (Lemma 10) *Using Lemma 9 with $u = 0$ we have for any $t \geq k \geq t - 2t^\star$ and $w_{t:k} \in V^{t-k+1}$*

$$\begin{aligned}
\|C_M^{1/2}\Pi^{\Delta}(w_{t:k+1})N_{k,w_k}\| &\leq 2\eta\Delta_\lambda\|N_{k,w_k}\| \sum_{\ell=1}^{t-k} \|C_M^{1/2}(I - \eta C_M)^{t-k-\ell}C_{M,\lambda}^{1/2}\| \\
&\leq 2\eta\Delta_\lambda\|N_{k,w_k}\|\Big(\sum_{\ell=1}^{t-k}\|C_M(I - \eta C_M)^{t-k-\ell}\| + \sqrt{\lambda}\sum_{\ell=1}^{t-k}\|C_M^{1/2}(I - \eta C_M)^{t-k-\ell}\|\Big) \\
&\leq 20\eta\Delta_\lambda\|N_{k,w_k}\| \log(t)(1 \vee \sqrt{\lambda\eta t})
\end{aligned}$$

where we applied Lemma 6 remark 5 to the bound the series of contractions. The case $k = t$ the above quantity is zero. With $\sum_{w_{t:k} \in V^{t-k+1}} |\Delta(w_{t:k})| \leq 2$ this leads to the error term being bounded

$$\boldsymbol{E}_2 \leq 40\Delta_\lambda \log(t)(1 \vee \sqrt{\lambda\eta t})\eta t^\star\Big(\max_{1 \leq k \leq t, w \in V} \|N_{k,w}\|\Big).$$

The final bound is arrived at by bounding for $\lambda' > 0$ the error term in the brackets as $\max_{1\leq k\leq t, w\in V}\|N_{k,w}\| \leq \|C_{M,\lambda'}^{1/2}\| \max_{1\leq k\leq t,w\in V} \|C_{M,\lambda'}^{-1/2}N_{k,w}\|$, and plugging in high probability bounds for $\max_{1\leq k\leq t,w\in V}\|C_{M,\lambda'}^{-1/2}N_{k,w}\|$ and Δ_λ from Lemma 7, with a union bound.

D.5. Bounding $\mathbf{E}_3$

The bound for this error term is similar to $\mathbf{E}_2$ and will be presented within the following Lemma.

Lemma 11 (Bounding $\mathbf{E}_3$) *Let $\delta \in (0, 1]$, $n, m, M \in \mathbb{N}_+$ and $\eta\kappa^2 \leq 1$ and $t \geq 2t^\star \geq 2$ and $\lambda, \lambda' > 0$. Under assumption 2,3 and 4 we have with probability greater than $1 - \delta$*

$$\boldsymbol{E}_3 \leq 24\|C_M^{1/2}\|\|C_{M,\lambda'}^{1/2}\|(\eta t)\sqrt{\eta t^\star}\big(1 \vee \sqrt{\lambda\eta t^\star}\big) \log^2 \frac{12n}{\delta} g(\lambda, m) f(\lambda', m, \delta/(2n))$$

Proof 7 (Lemma 11) *For $1 \leq k \leq t - 2t^\star - 1$ and $w_{t:k} \in V^{t-k+1}$ use Lemma 9 with $u = 1/2$ as well as $\eta\kappa^2 \leq 1$ to* *bound with* $\lambda > 0$

$$
\begin{aligned}
&\|C_M^{1/2}\Pi(w_{t:k+t^\star+1})\Pi^\Delta(w_{k+t^\star:k+1})N_{k,w_k}\| \\
&\leq \|C_M^{1/2}\|\|\Pi^\Delta(w_{k+t^\star:k+1})N_{k,w_k}\| \\
&\leq 2\eta\|C_M^{1/2}\|\Delta_\lambda\|N_{k,w_k}\|\sum_{\ell=1}^{t^\star}\|(I-\eta C_M)^{t^\star-\ell}C_{M,\lambda}^{1/2}\| \\
&\leq 2\|C_M^{1/2}\|\Delta_\lambda\|N_{k,w_k}\|\Big(\eta\sum_{\ell=1}^{t^\star}\|(I-\eta C_M)^{t^\star-\ell}C_M^{1/2}\|+\sqrt{\lambda}\eta t^\star\Big) \\
&\leq 12\|C_M^{1/2}\|\Delta_\lambda\|N_{k,w_k}\|\sqrt{\eta t^\star}(1\vee\sqrt{\lambda\eta t^\star})
\end{aligned}
$$

where we have bounded the series of contractions using Lemma 6 remark 5 once again. With $\sum_{w_{t:k}\in V^{t-k+1}}|\Delta(w_{t:k})|\leq 2$, *plugging in the above yields the bound for* $\boldsymbol{E}_3$

$$
\boldsymbol{E}_3 \leq 24\|C_M^{1/2}\|(\eta t)\sqrt{\eta t^\star}\Delta_\lambda(1\vee\sqrt{\lambda\eta t^\star})\Big(\max_{1\leq k\leq t,w\in V}\|N_{k,w}\|\Big).
$$

The final bound is arrived at by bounding Δ_λ *and* $\Big(\max_{1\leq k\leq t,w\in V}\|N_{k,w}\|\Big)$ *in an identical manner to Lemma 10 for error term* $\boldsymbol{E}_2$.

D.6. Bounding $\mathbf{E}_4$

This term will be controlled through the convergence of $P^{t^\star}$ to the stationary distribution. It is summarised within the following Lemma.

Lemma 12 (Bounding $\mathbf{E}_4$) *Let* $\delta\in(0,1]$, $n,m,M\in\mathbb{N}_+$ *and* $\eta\kappa^2\leq 1$ *and* $t\geq 2t^\star\geq 2$ *and* $\lambda>0$. *Under assumption 2,3 and 4 we have with probability greater than* $1-\delta$

$$
\boldsymbol{E}_4 \leq 4\|C_{M,\lambda}^{1/2}\|\big(\sqrt{n}\sigma_2^{t^\star}\wedge 1\big)(\eta t)\log\frac{6n}{\delta}f(\lambda,m,\delta/n)
$$

Proof 8 (Lemma 12) *Begin by bounding for* $t-2t^\star-1\geq k\geq 1$, $w_k\in V$ *and* $w_{t:k+t^\star+1}\in V^{t-t^\star-k}$ *the following*

$$
\|C_M^{1/2}\Pi^\Delta(w_{t:k+t^\star+1})(I-\eta C_M)^{t^\star}N_{k,w_k}\|\leq 2\|C_M^{1/2}\|\|N_{k,w_k}\|.
$$

Furthermore, we can bound the summation over paths by the deviation of the form $\sum_{w\in V}|\Delta^{t^\star}(v,w)|$ *and use Lemma 5 thereafter to arrive at*

$$
\begin{aligned}
\sum_{w_k\in V}\sum_{w_{t:k+t^\star+1}\in V^{t-t^\star-k}}|P_{vw_{t:k+t^\star+1}}\Delta^{t^\star}(w_{k+t^\star+1},w_k)| &= \sum_{w_{t:k+t^\star+1}\in V^{t-t^\star-k}}|P_{vw_{t:k+t^\star+1}}|\Big(\sum_{w_k\in V}|\Delta^{t^\star}(w_{k+t^\star+1},w_k)|\Big) \\
&\leq \max_{u\in V}\Big(\sum_{w\in V}|\Delta^{t^\star}(u,w)|\Big)\Big(\sum_{w_{t:k+t^\star+1}\in V^{t-t^\star-k}}|P_{vw_{t:k+t^\star+1}}|\Big) \\
&= \max_{u\in V}\Big(\sum_{w\in V}|\Delta^{t^\star}(u,w)|\Big) \\
&\leq 2\big(\sqrt{n}\sigma_2^{t^\star}\wedge 1\big).
\end{aligned}
$$

Bringing everything together yields the following bound for $\boldsymbol{E}_4$

$$
\boldsymbol{E}_4 \leq 2\big(\sqrt{n}\sigma_2^{t^\star}\wedge 1\big)(\eta t)\Big(\max_{1\leq k\leq t,w\in V}\|N_{k,w}\|\Big) \tag{12}
$$

Plugging in high probability bounds for $\max_{1\leq k\leq t,w\in V}\|N_{k,w}\|$ *following Lemma 10 for error term* $\boldsymbol{E}_2$ *then yields the bound.*

D.7. Bounding $\mathbf{E}_5$

The summation over paths in this case is decoupled from the error. This allows for a more sophisticated bound to be applied, which considers the deviation of the iterates from the average. The following Lemma effectively bounds the norm of $\sum_{w_{t:1} \in V^t} \Delta(w_{t:1})\Pi^{\Delta}(w_{t:1})$, which involves a sum over the paths $w_{t:1}$.

Lemma 13 *Let $N \in \mathbb{R}^M$, $w_{t:1} \in V^t$ and $\lambda_i \geq 0$ for $i \in \{1,2,3\}$. Then,*

$$\|\sum_{w_{t:1}\in V^t} \Delta(w_{t:1}) C_M^{1/2} \Pi^{\Delta}(w_{t:1}) N\| \leq 4\eta \Delta_{\lambda_1} \|N\| \sum_{k=1}^{t} \|C_M^{1/2}(I - \eta \widehat{C}_M)^{t-k} C_{M,\lambda_1}^{1/2}\| (\sigma_2^{t-k+1} \wedge 1)$$
$$+ 8\eta^2 \Delta_{\lambda_2}\Delta_{\lambda_3}\|N\| \sum_{k=2}^{t}\sum_{\ell=1}^{k-1} \|C_M^{1/2}(I - \eta \widehat{C}_M)^{t-k} C_{M,\lambda_2}^{1/2}\| \|(I - \eta \widehat{C}_M)^{k-1-\ell} C_{M,\lambda_3}^{1/2}\| (\sigma_2^{k-\ell} \wedge 1)$$

The bound for this error term is then summarised within the following Lemma.

Lemma 14 (Bounding $\mathbf{E}_5$) *Let $\delta \in (0,1]$, $n, m, M \in \mathbb{N}_+$ and $\eta\kappa^2 \leq 1$ and $t \geq 2t^\star \geq 2$ and $\lambda', \lambda_i > 0$ for $i = 1, \ldots, 3$. Under assumption 2,3 and 4 and if $\frac{9\kappa^2}{M}\log\frac{M}{\delta} \leq \lambda_i$ for $i = 1, 2$ then with probability greater than $1 - 8\delta$*

$$\boldsymbol{E}_5 \leq \boldsymbol{E}_{51} + \boldsymbol{E}_{52}$$

where

$$\boldsymbol{E}_{51} \leq 84\|C_M^{1/2}C_{M,\lambda_1}^{1/2}\|\|C_{M,\lambda'}^{1/2}\|\eta t(1 \vee \sigma_2^{t^\star}\eta t \vee \lambda_1 \eta t^\star) \times g(\lambda_1, m) f(\lambda', nm, \delta)\log(t)\log^2\frac{6n}{\delta}$$
$$\boldsymbol{E}_{52} \leq 160\|C_{M,\lambda'}^{1/2}\|\|C_{M,\lambda_3}^{1/2}\|(\eta t)(1 \vee \lambda_2 \eta t)(\sigma_2^{t^\star}\eta t \vee \eta t^\star) \times g(\lambda_2, m) g(\lambda_3, m) f(\lambda', nm, \delta)\log(t)\log^3\frac{6n}{\delta}$$

Proof 9 (Lemma 14) *Applying for $1 \leq k \leq t - 2t^\star - 1$ Lemma 13 with $N = (I - \eta C_M)^{t^\star} N_k = N'_k$, and $w_{t:k+t^\star+1} \in V^{t-t^\star-k}$ to elements within the series of $\boldsymbol{E}_5$ we arrive at*

$$\begin{aligned}
\boldsymbol{E}_5 &\leq 4 \sum_{k=1}^{t-2t^\star-1} \eta^2 \Delta_{\lambda_1}\|N'_k\| \sum_{\ell=1}^{t-t^\star-k} \|C_M^{1/2}(I - \eta\widehat{C}_M)^{t-t^\star-k-\ell} C_{M,\lambda_1}^{1/2}\| \big(\sigma_2^{t-t^\star-k-\ell+1} \wedge 1\big) \\
&\quad + 8 \sum_{k=1}^{t-2t^\star-1} \eta^3 \Delta_{\lambda_2}\Delta_{\lambda_3}\|N'_k\| \sum_{\ell=2}^{t-t^\star-k}\sum_{j=1}^{\ell-1} \|C_M^{1/2}(I - \eta\widehat{C}_M)^{t-t^\star-k-\ell} C_{M,\lambda_2}^{1/2}\| \\
&\qquad\qquad \times \|(I - \eta\widehat{C}_M)^{\ell-j-1} C_{M,\lambda_3}^{1/2}\| (\sigma_2^{\ell-j} \wedge 1) \\
&= \boldsymbol{E}_{51} + \boldsymbol{E}_{52}
\end{aligned}$$

where we have labelled the remaining error terms $\boldsymbol{E}_{51}, \boldsymbol{E}_{52}$. Each of these terms are now bounded.

To bound the first term $\boldsymbol{E}_{51}$, begin by for $1 \leq k \leq t - 2t^\star - 1$ splitting the series at $1 \leq \ell \leq t - 2t^\star - k$ to arrive at

$$\begin{aligned}
&\eta \sum_{\ell=1}^{t-t^\star-k} \|C_M^{1/2}(I - \eta\widehat{C}_M)^{t-t^\star-k-\ell} C_{M,\lambda_1}^{1/2}\| \big(\sigma_2^{t-t^\star-k-\ell+1} \wedge 1\big) \\
&\leq \|C_M^{1/2}C_{M,\lambda_1}^{1/2}\|\eta \sum_{\ell=1}^{t-2t^\star-k} \big(\sigma_2^{t-t^\star-k-\ell+1} \wedge 1\big) + \eta \sum_{\ell=t-2t^\star-k}^{t-t^\star-k} \|C_M^{1/2}(I - \eta\widehat{C}_M)^{t-t^\star-k-\ell} C_{M,\lambda_1}^{1/2}\| \\
&\leq \|C_M^{1/2}C_{M,\lambda_1}^{1/2}\|\eta \sum_{\ell=1}^{t-2t^\star-k} \big(\sigma_2^{t-t^\star-k-\ell+1} \wedge 1\big) + \eta\|C_M^{1/2}\widehat{C}_{M,\lambda_1}^{-1/2}\|\|\widehat{C}_{M,\lambda_1}^{-1/2}C_{M,\lambda_1}^{1/2}\| \sum_{\ell=t-2t^\star-k}^{t-t^\star-k} \|\widehat{C}_{M,\lambda_1}^{1/2}(I - \eta\widehat{C}_M)^{t-t^\star-k-\ell}\widehat{C}_{M,\lambda_1}^{1/2}\| \\
&\leq \|C_M^{1/2}C_{M,\lambda_1}^{1/2}\|\sigma_2^{t^\star}\eta t + 10\|C_M^{1/2}\widehat{C}_{M,\lambda_1}^{-1/2}\|\|\widehat{C}_{M,\lambda_1}^{-1/2}C_{M,\lambda_1}^{1/2}\|\log(t)(1 \vee \lambda_1\eta t^\star)
\end{aligned}$$

where for the first series used that $\sigma_2^{t-t^\star-k-\ell+1} \leq \sigma_2^{t^\star}$ from $\ell \leq t-2t^\star-k$ meanwhile for the second series

$$\eta \sum_{\ell=t-2t^\star-k}^{t-t^\star-k} \|\widehat{C}_{M,\lambda_1}^{1/2}(I-\eta\widehat{C}_M)^{t-t^\star-k-\ell}\widehat{C}_{M,\lambda_1}^{1/2}\| \leq \eta \sum_{\ell=t-2t^\star-k}^{t-t^\star-k} \|\widehat{C}_M(I-\eta\widehat{C}_M)^{t-t^\star-k-\ell}\| + \eta\lambda_1 \sum_{\ell=t-2t^\star-k}^{t-t^\star-k} \|(I-\eta\widehat{C}_M)^{t-t^\star-k-\ell}\|$$
$$\leq 5\log(t) + 5\lambda_1\eta t^\star$$

to which we applied Lemma 6 remark 5 to bound the series of contractions. This leads to the bound for $\boldsymbol{E}_{51}$

$$\boldsymbol{E}_{51} \leq 4\Delta_{\lambda_1}\eta t\Big(\|C_M^{1/2}C_{M,\lambda_1}^{1/2}\|\sigma_2^{t^\star}\eta t + 10\|C_M^{1/2}\widehat{C}_{M,\lambda_1}^{-1/2}\|\|\widehat{C}_{M,\lambda_1}^{-1/2}C_{M,\lambda_1}^{1/2}\|\log(t)(1\vee\lambda_1\eta t^\star)\Big)\big(\max_{1\leq k\leq t}\|N_k'\|\big).$$

Provided $\frac{9\kappa^2}{M}\log\frac{M}{\delta} \leq \lambda_1$ we have from Lemma 3 in (Carratino et al., 2018) that with probability greater than $1-\delta$

$$\|C_M^{1/2}\widehat{C}_{M,\lambda_1}^{-1/2}\|\|\widehat{C}_{M,\lambda_1}^{-1/2}C_{M,\lambda_1}^{1/2}\| \leq \|\widehat{C}_{M,\lambda_1}^{-1/2}C_{M,\lambda_1}^{1/2}\|^2 \leq 2.$$

Meanwhile for $\lambda' > 0$, we can bound $\max_{1\leq k\leq t}\|N_k'\| \leq \|C_{M,\lambda'}^{1/2}\|\max_{1\leq k\leq t}\|C_{M,\lambda'}^{-1/2}N_k'\| \leq \|C_{M,\lambda'}^{1/2}\|\max_{1\leq k\leq t}\|C_{M,\lambda'}^{-1/2}N_k\|$. The bound is arrived at by also plugging in high probability bounds for $\|C_{M,\lambda'}^{-1/2}N_k\|$ and Δ_{λ_1} from Lemma 7.

Finally to bound $\boldsymbol{E}_{52}$. Begin by bounding for $1 \leq k \leq t-2t^\star-1$ as well as $2 \leq \ell \leq t^\star$ the series as

$$\sum_{j=1}^{\ell-1}\|(I-\eta\widehat{C}_M)^{\ell-j}C_{M,\lambda_3}^{1/2}\|(\sigma_2^{\ell-j}\wedge 1) \leq \|C_{M,\lambda_3}^{1/2}\|t^\star.$$

Meanwhile for $t^\star+1 \leq \ell \leq t-t^\star-k$ we can split the series as $1 \leq j \leq \ell-t^\star$

$$\sum_{j=1}^{\ell-1}\|(I-\eta\widehat{C}_M)^{\ell-j}C_{M,\lambda_3}^{1/2}\|(\sigma_2^{\ell-j}\wedge 1)$$
$$\leq \|C_{M,\lambda_3}^{1/2}\|\sum_{j=1}^{\ell-t^\star}(\sigma_2^{\ell-j}\wedge 1) + \sum_{j=\ell-t^\star+1}^{\ell-1}\|(I-\eta\widehat{C}_M)^{\ell-j}C_{M,\lambda_3}^{1/2}\|$$
$$\leq \|C_{M,\lambda_3}^{1/2}\|(\sigma_2^{t^\star}t + t^\star)$$

where for the first series we applied $j \leq \ell-t^\star$ to say $\sigma_2^{\ell-j} \leq \sigma_2^{t^\star}$, and for the second simply summed up the $t^\star$ terms after bounding $\|(I-\eta\widehat{C}_M)^{\ell-j}C_{M,\lambda_3}^{1/2}\| \leq \|C_{M,\lambda_3}^{1/2}\|$. Plugging in the above bound for all $2 \leq \ell \leq t-t^\star-k$ we arrive at the following bound for $\boldsymbol{E}_{52}$

$$\boldsymbol{E}_{52} \leq 8\Delta_{\lambda_2}\Delta_{\lambda_3}\big(\max_{1\leq k\leq t}\|N_k'\|\big)\|C_{M,\lambda_3}^{1/2}\|(\sigma_2^{t^\star}\eta t + \eta t^\star)\sum_{k=1}^{t-2t^\star-1}\eta^2\sum_{\ell=2}^{t-t^\star-k}\|C_M^{1/2}(I-\eta\widehat{C}_M)^{t-t^\star-k-\ell}C_{M,\lambda_2}^{1/2}\|$$

For $1 \leq k \leq t-2t^\star-1$ the series of contractions over ℓ can be bounded using Lemma 6 remark 5 in a similar manner to previously as

$$\eta\sum_{\ell=2}^{t-t^\star-k}\|C_M^{1/2}(I-\eta\widehat{C}_M)^{t-t^\star-k-\ell}C_{M,\lambda_2}^{1/2}\| \leq \|C_M^{1/2}\widehat{C}_{M,\lambda_2}^{-1/2}\|\|\widehat{C}_{M,\lambda_2}^{-1/2}C_{M,\lambda_2}^{1/2}\|10\log(t)(1\vee\lambda_2\eta t).$$

Summing up the remaining series for over k, using that $\|C_M^{1/2}\widehat{C}_{M,\lambda_2}^{-1/2}\|\|\widehat{C}_{M,\lambda_2}^{-1/2}C_{M,\lambda_2}^{1/2}\| \leq 2$ from $\frac{9\kappa^2}{M}\log\frac{M}{\delta} \leq \lambda_2$, plugging in high probability bounds for $\max_{1\leq k\leq t}\|N_k'\|$ from the the error term $\boldsymbol{E}_{51}$, as well as high probability bounds for $\Delta_{\lambda_2}, \Delta_{\lambda_3}$ from Lemma 7 yields the bound.

E. Final bounds

In this section we bring together the high probability bounds for the Statistical Error and Distributed Error. This section is then as follows. Section E.1 provides the proof for Theorem 1. Section E.2 gives the proof for Theorem 1.

E.1. Refined Bound (Theorem 2)

In this section we give conditions under which we obtained a refined bound.

Proof 10 (Theorem 2) *Fixing $\delta \in (0,1]$ and a constant $c_{union} > 1$, assume that*

$$\eta t = (nm)^{\frac{1}{2r+\gamma}}$$
$$M \geq \Big((nm)^{\frac{1+\gamma(2r-1)}{2r+\gamma}}\Big) \vee \Big(\eta t \log \frac{60 n\kappa^2(\eta t \vee M)c_{union}}{\delta}\Big)$$
$$t^\star \geq 2\frac{\log(nmt)}{1-\sigma_2}$$
$$m \geq \Big((1 \vee (\eta t^\star))^{2r+\gamma} n^{2r/\gamma}\Big) \vee \Big((1 \vee (\eta t^\star))^2 n\Big) \vee \Big((1 \vee \eta t^\star)^{\frac{(1+\gamma)(2r+\gamma)}{2(r+\gamma-1)}} n^{\frac{(r+1)}{(r+\gamma-1)}}\Big)$$

Now, consider the error decomposition given (8), to arrive at the bound

$$\mathcal{E}(f_{t+1,v}) - \mathcal{E}(f_{\mathcal{H}}) \leq 2\underbrace{\|S_M\widehat{\omega}_{t+1,v} - S_M\widehat{v}_t\|^2_\rho}_{(\textit{Network Error})^2} + 2\underbrace{\|S_M\widehat{v}_t - Pf_\rho\|^2_\rho}_{(\textit{Statistical Error})^2}.$$

Begin by bounding the statistical error by using Lemma 4. Using Assumption 5 to bound $\mathcal{N}(\frac{1}{\eta t}) \leq Q^2(\eta t)^\gamma$, and noting that $M \geq (4+18\eta t\kappa^2)\log\frac{60\kappa^2\eta t}{\delta}$ is satisfied, allows us to upper bound with probability greater than $1-\delta$

$$\|S_M\widehat{v}_t - Pf_\rho\|^2_\rho \leq (nm)^{-2r/(2r+\gamma)}\Big(c_1^2\Big(1 \vee \frac{(\eta t)\log\frac{3M}{\delta}}{M}\Big)(1 \vee Q^2)\log^2(t)\log^2\Big(\frac{12}{\delta}\Big) + c_3^2\Big)$$
$$+ c_2^2\Big(\frac{1}{M^{2r}} \vee \frac{Q^2}{M(nm)^{(1-\gamma)(2r-1)/(2r+\gamma)}}\Big)\log^{2(1-r)}(11\kappa^2\eta t)\log^2\big(\frac{6}{\delta}\big)$$

The quantity within the brackets for second term is then upper bounded $\frac{1}{M(nm)^{(1-\gamma)(2r-1)/(2r+\gamma)}} \leq (nm)^{-2r/(2r+\gamma)}$ provided $M \geq (nm)^{\frac{1+\gamma(2r-1)}{2r+\gamma}}$, which is satisfied as an assumption in the Theorem. This results in an upper bound on the statistical error that is, up to log factors, decreasing as $(nm)^{-2r/(2r+\gamma)}$ in high probability.

We now proceed to bound the Network Error Term. Begin by considering error decomposition given in (11) into the terms $\boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3, \boldsymbol{E}_4, \boldsymbol{E}_5$, in particular by applying the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ multiple times we get

$$\|S_M\widehat{\omega}_{t+1,v} - S_M\widehat{v}_t\|^2_\rho \leq 2\boldsymbol{E}_1^2 + 4\boldsymbol{E}_2^2 + 8\boldsymbol{E}_3^2 + 16\boldsymbol{E}_4^2 + 32\boldsymbol{E}_5^2,$$

and thus it is sufficient to show each of these terms is decreasing as $(nm)^{-2r/(2r+\gamma)}$ in high probability. Before doing so we note Lemma 4 in (Carratino et al., 2018) states for any $\lambda > 0$ that if

$$M \geq \big(4 + \frac{18\kappa^2}{\lambda}\big)\log\frac{12\kappa^2}{\lambda\delta}$$

then with probability greater than $1-\delta$ we have $\mathcal{N}_M(\lambda) \leq q\mathcal{N}(\lambda)$ where $q = \max\big(2.55, \frac{2\kappa^2}{\|L\|}\big)$. We note this is satisfied with both $\lambda = (\eta t)^{-1}, (1 \vee (\eta t^\star))^{-1}$ by the assumptions within the Theorem, and as such, we can interchange from $\mathcal{N}_M(\lambda)$ to $\mathcal{N}(\lambda)$ with at most a constant cost of q.

We begin by bounding $\boldsymbol{E}_1^2$ by considering Lemma 8 with $\lambda' = \kappa^2$ and $\lambda = (1 \vee \eta t^\star)^{-1}$, which leads to with probability greater than $1-\delta$

$$\boldsymbol{E}_1^2 \leq \Big(2\|C^{1/2}_{M,\lambda'}\|^2\sigma_2^{2t^\star}t^2\kappa^{-2}(f(\lambda', m, \delta/(2n)))^2 + 40\log^2(t^\star)(f(\lambda, m, \delta/(2n)))^2\Big)\log^2\frac{12n}{\delta}$$

Now due to $t^\star \geq \frac{2\log(nmt)}{1-\sigma_2} \geq \frac{2\log(nmt)}{-\log(\sigma_2)}$ (the second inequality arising from $\log(x) \geq 1 - x^{-1}$ for $x \geq 0$) we have $\sigma_2^{t^\star} \leq (tnm)^{-2}$. As such with the fact that $f(\kappa^2, m, \delta/(2n)) \lesssim m^{-1/2}$ in high probability, the first term above is decreasing, upto logarithmic factors, as $(nm)^{-2r/(2r+\gamma)}$. Meanwhile for the second term we have that

$$f((1 \vee \eta t^\star)^{-1}, m, \delta/2)^2 \leq a_1^2\big(\frac{(1 \vee \eta t^\star)}{m^2} \vee \frac{(1 \vee \eta t^\star)^\gamma}{m}\big)\big(1 \vee \frac{3(\eta t\kappa \vee 1)}{M}\log\frac{6Mn}{\delta}\big)$$

for the constant $a_1 = 64\Big(\sqrt{B}(\kappa \vee \sqrt{\sqrt{p}q})\Big) \vee \big(\kappa \vee \sqrt{q}\big)$. *For* $\boldsymbol{E}_1^2$ *to be decreasing at the rate* $(nm)^{-2r/(2r+\gamma)}$, *up to logarithmic factors, we then require* $\frac{(1\vee \eta t^\star)^\gamma}{m} \leq (nm)^{-2r/(2r+\gamma)}$ *which is satisfied when* $m \geq (1 \vee (\eta t^\star))^{2r+\gamma} n^{2r/\gamma}$.

Proceed to bound $\boldsymbol{E}_2^2$ *by considering Lemma 10 with* $\lambda = 1/(\eta t)$ *and* $\lambda' = \kappa^2$ *to arrive at with probability greater than* $1-\delta$

$$\boldsymbol{E}_2^2 \leq 40^2 \kappa^2 \|C_{M,\lambda'}^{1/2}\|^2 \log^2(t)(\eta t^\star)^2 (g(\lambda,m))^2 (f(\lambda', m, \delta/(2n)))^2 \log^4 \frac{12n}{\delta}$$

As discussed previously, we have with high probability that $(f(\kappa^2, m, \delta/(2n)))^2 \lesssim 1/m$, *meanwhile*

$$g((\eta t)^{-1}, m)^2 \leq a_2^2 \big(\frac{\eta t}{m^2} \vee \frac{(\eta t)^\gamma}{m}\big)$$

where $a_2 = 8\kappa(\kappa \vee \sqrt{q})$. *As such for* $\boldsymbol{E}_2^2$ *to be decreasing at the rate* $(nm)^{-2r/(2r+\gamma)}$ *we require* $\frac{(\eta t)^\gamma (1\vee \eta t^\star)^2}{m^2} \leq (nm)^{-2r/(2r+\gamma)}$ *which, plugging in* $\eta t = (nm)^{1/(2r+\gamma)}$ *is satisfied when* $m \geq (1 \vee \eta t^\star)^2 n$.

Bounding $\boldsymbol{E}_3$ *using Lemma 11 with* $\lambda = (1 \vee (\eta t^\star))^{-1}$ *and* $\lambda' = \kappa^2$ *we have with probability greater than* $1-\delta$

$$\boldsymbol{E}_3^2 \leq 24^2 \|C_M^{1/2}\|^2 \|C_{M,\lambda'}^{1/2}\|^2 (\eta t)^2 (\eta t^\star)(g(\lambda,m))^2 (f(\lambda', m, \delta/(2n)))^2 \log^4 \frac{12n}{\delta}.$$

Following the steps for $\boldsymbol{E}_2$, *we have with high probability that* $f(\kappa^2, m, \delta/(2n))^2 \lesssim 1/m$, *meanwhile* $g((1 \vee (\eta t^\star))^{-1}, m) \lesssim (1 \vee \eta t^\star)^\gamma / m$. *As such for* $\boldsymbol{E}_3^2$ *to be decreasing with the rate* $(nm)^{-2r/(2r+\gamma)}$ *we require* $\frac{(\eta t)^2 (1 \vee \eta t^\star)^{1+\gamma}}{m^2} \leq (nm)^{-2r/(2r+\gamma)}$, *which is satisfied when* $r+\gamma > 1$ *and* $m \geq (1 \vee \eta t^\star)^{\frac{(1+\gamma)(2r+\gamma)}{2(r+\gamma-1)}} n^{\frac{(r+1)}{(r+\gamma-1)}}$.

Now to bound $\boldsymbol{E}_4$ *we consider Lemma 12 with* $\lambda = \kappa^2$ *to arrive at with probability greater than* $1-\delta$

$$\boldsymbol{E}_4^2 \leq 16 \|C_{M,\lambda}^{1/2}\|^2 (n\sigma_2^{2t^\star} \wedge 1)(\eta t)^2 \log^2 \Big(\frac{6n}{\delta}\Big) (f(\lambda, m, \delta/n))^2.$$

Following the previous analysis we know with high probability $(f(\lambda, m, \delta/n))^2 = \widetilde{O}(1/m)$ *and that* $t^\star$ *is such that* $\sigma_2^{t^\star} \leq (tnm)^{-2}$. *Combining these two facts we have that* $\boldsymbol{E}_4^2$ *is of the order* $(nm)^{-2r/(2r+\gamma)}$ *with high probability.*

The bound for $\boldsymbol{E}_5^2$ *is naturally split across the terms* $\boldsymbol{E}_{51}, \boldsymbol{E}_{52}$ *from Lemma 14. In particular we have that*

$$\boldsymbol{E}_5^2 \leq 2\boldsymbol{E}_{51}^2 + 2\boldsymbol{E}_{52}^2$$

The remainder of the proof then shows each of the terms above are decreasing at the rate $(nm)^{-2r/(2r+\gamma)}$ *in high probability by using the bounds provided within Lemma 14. We note the condition* $\frac{9\kappa^2}{M} \log \frac{M9\kappa^2}{\delta} \leq \lambda_i$ *for* $i = 1, 2$ *is satisfied for* $\lambda_1 = (1 \vee (\eta t^\star))^{-1}$ *and* $\lambda_2 = (\eta t)^{-1}$ *by the assumptions.*

Consider the bound for $\boldsymbol{E}_{51}$ *with* $\lambda_1 = (1 \vee (\eta t^\star))^{-1}$ *and* $\lambda' = \kappa^2$, *so we have with probability greater than* $1-\delta$

$$\boldsymbol{E}_{51}^2 \leq 84^2 \|C_M^{1/2} C_{M,\lambda_1}^{1/2}\|^2 \|C_{M,\lambda'}^{1/2}\|^2 (\eta t)^2 (1 \vee \sigma_2^{2t^\star}(\eta t)^2)(g(\lambda_1, m))^2 (f(\lambda', nm, \delta/8))^2 \log^2(t) \log^4 \frac{48n}{\delta}.$$

From previously we have that $t^\star$ *so that* $\sigma_2^{t^\star} \leq (tnm)^{-2}$ *and thus* $\sigma_2^{t^\star} \eta t \leq 1$. *Meanwhile following steps from previously we have* $(g(\lambda_1, m))^2 \lesssim (1 \vee (\eta t^\star))^\gamma / m$ *as well as with high probability* $(f(\lambda', nm, \delta))^2 \lesssim (nm)^{-1}$. *As such we require* $\frac{(\eta t)^2 (1 \vee (\eta t^\star))^\gamma}{m(nm)} \leq (nm)^{-2r/(2r+\gamma)}$ *which is satisfied when* $r+\gamma > 1$ *and* $m \geq n^{\frac{2-\gamma}{2(r+\gamma-1)}} (1 \vee (\eta t^\star))^{\frac{\gamma(2r+\gamma)}{2(r+\gamma-1)}}$. *This is then implied by the assumption that* $m \geq (1 \vee \eta t^\star)^{\frac{(1+\gamma)(2r+\gamma)}{2(r+\gamma-1)}} n^{\frac{(r+1)}{(r+\gamma-1)}}$ *and* $r + \gamma \geq 1$.

Finally to bound $\boldsymbol{E}_{52}$ *consider the bound given with* $\lambda_2 = (\eta t)^{-1}$, *and* $\lambda_3 = \lambda' = \kappa^2$ *to arrive at with probability greater than* $1-\delta$

$$\boldsymbol{E}_{52}^2 \leq 160^2 \|C_{M,\lambda'}^{1/2}\|^2 \|C_{M,\lambda_3}^{1/2}\|^2 (\eta t)^2 \big(\sigma_2^{t^\star} \eta t \vee (\eta t^\star)^2\big) g(\lambda_2, m) g(\lambda_3, m) f(\lambda', nm, \delta/8) \log^2(t) \log^6 \frac{48n}{\delta}.$$

Once again $\sigma_2^{t^\star} \leq (tnm)^{-2}$ *ensures* $\sigma_2^{t^\star} \eta t \leq (1 \vee \eta t^\star)$. *Meanwhile we have* $(g(\lambda_2, m))^2 \lesssim (\eta t)^\gamma / m$, $(g(\lambda_3, m))^2 \lesssim 1/m$ *and with high probability* $(f(\lambda', nm, \delta/8))^2 \lesssim 1/(nm)$. *As such to ensure this term is sufficiently small we require*

$\frac{(\eta t)^{2+\gamma}(1 \vee \eta t^\star)^2}{m^2(nm)} \leq (nm)^{-2r/(2r+\gamma)}$, *which satisfied if* $m \geq n^{\frac{1}{2r+\gamma}}(1 \vee (\eta t^\star))^{\frac{2r+\gamma}{2r+\gamma-1}}$. *This then being implied by* $m \geq (1 \vee \eta t^\star)^{\frac{(1+\gamma)(2r+\gamma)}{2(r+\gamma-1)}} n^{\frac{(r+1)}{(r+\gamma-1)}}$ *since* $\frac{r+1}{r+\gamma-1} \geq \frac{1}{2r+\gamma}$ *and* $\frac{(1+\gamma)(2r+\gamma)}{2(r+\gamma-1)} \geq \frac{2r+\gamma}{2r+\gamma-1}$. *The second inequality arising from the observation that* $\frac{1}{2(r+\gamma-1)} \geq \frac{1}{2(r+\gamma-1)+1-\gamma} = \frac{1}{2r+\gamma-1}$.

Each of the bounds for $\mathbf{E}_i^2$ *for* $i = 1, \dots, 5$ *hold in high probability, and as such, can be combined with a union bound. This incurs at most a logarithmic factor in the bound, with the number of unions applied being upper bounded by the constant* $c_{union} > 1$ *chosen at the start.*

E.2. Worst Case (Theorem 1)

Consider the refined bound in Theorem 2 with $r = 1/2$ and $\gamma = 1$.

E.3. Leading Order Error Terms (Theorem 3)

Follow the proof of Theorem 2, where the error is decomposed into the following terms

$$\mathcal{E}(f_{t+1,v}) - \mathcal{E}(f_{\mathcal{H}}) \leq (\text{Network Error})^2 + (\text{Statistical Error})^2.$$

The statistical error follows (Carratino et al., 2018) and, in our work, is summarised within Lemma 4 to be upto logarithmic factors in high-probability

$$(\text{Statistical Error})^2 \lesssim \underbrace{\Big(1 \vee \frac{\eta t}{M}\Big)\frac{(\eta t)^\gamma}{nm}}_{\text{Sample Variance}} + \underbrace{\frac{1}{M(\eta t)^{(1-\gamma)(2r-1)}}}_{\text{Random Fourier Error}} + \underbrace{\frac{1}{(\eta t)^{2r}}}_{\text{Bias}}.$$

Meanwhile the network error is bounded into terms

$$(\text{Network Error})^2 \lesssim \mathbf{E}_1^2 + \mathbf{E}_2^2 + \mathbf{E}_3^2 + \mathbf{E}_4^2 + \mathbf{E}_5^2$$

where high-probability bounds from Section D are used. In particular, the bounds each term are, up to logarithmic factors, in high probability

$$\mathbf{E}_1^2 \lesssim \frac{(\eta t^\star)^\gamma}{m}$$
$$\mathbf{E}_2^2 \lesssim \frac{(\eta t^\star)^2(\eta t)^\gamma}{m^2}$$
$$\mathbf{E}_3^2 \lesssim \frac{(\eta t)^2(\eta t^\star)^{1+\gamma}}{m^2}$$
$$\mathbf{E}_4^2 \lesssim \frac{n\sigma_2^{2t^\star}(\eta t)^2}{m}$$
$$\mathbf{E}_5^2 \lesssim \frac{(\eta t)^2(1 \vee (\eta t^\star))^\gamma}{m(nm)} + \frac{(\eta t)^{2+\gamma}(1 \vee \eta t^\star)^2}{m^2(nm)}$$

The leading order terms are then defined as $\mathbf{E}_1^2$ and $\mathbf{E}_3^2$.

F. Proofs of Auxiliary Lemmas

In this section we provide the proofs of the auxiliary lemmas. This section is then as follows. Section F.1 provides the proof for Lemma 7. Section F.2 provides the proof of Lemma 9. Section F.3 provides the proof of Lemma 13.

F.1. Concentration of Error terms (Lemma 7)

Proof 11 (Lemma 7) *Fix* $w \in V$. *We begin by collecting the necessary concentration results. Following Lemma 18 in (Lin & Cevher, 2018) with* $\mathcal{T}_\rho, \mathcal{T}_{\mathbf{x}}$ *swapped for* $C_M, \widehat{C}_M^{(w)}$ *respectively (or Proposition 5 in (Rudi & Rosasco, 2017)) we have with probability greater than* $1 - \delta$

$$\|C_{M,\lambda}^{-1/2}(C_M - \widehat{C}_M^{(w)})\| \leq 2\kappa\Big(\frac{2\kappa}{m\sqrt{\lambda}} + \sqrt{\frac{\mathcal{N}_M(\lambda)}{m}}\Big)\log\frac{2}{\delta}$$

From Lemma 2 in (Carratino et al., 2018) under assumptions 2 and 3 we have with probability greater than $1-\delta$ for all $t \geq 1$

$$\|\widetilde{v}_{t+1}\| \leq 2R\kappa^{2r-1}\Big(1+\sqrt{\frac{9\kappa^2}{M}\log\frac{M}{\delta}}\max\Big(\sqrt{\eta t},\kappa^{-1}\Big)\Big).$$

Meanwhile from Lemma 6 in (Rudi & Rosasco, 2017) under assumption 2 and 4 we have with probability greater than $1-\delta$

$$\|C_{M,\lambda}^{-1/2}(\widehat{S}_M^{(w)\top}\widehat{y} - S_M^\star f_\rho)\| \leq 2\sqrt{B}\Big(\frac{\kappa}{\sqrt{\lambda}m} + \sqrt{\frac{2\sqrt{p}\mathcal{N}_M(\lambda)}{m}}\Big)\log\frac{2}{\delta}$$

Considering $\|C_{M,\lambda}^{-1/2}N_{k,w}\|$, using triangle inequality and plugging the above bounds with a union bound, we have with probability greater than $1-\delta$

$$\begin{aligned}
\|C_{M,\lambda}^{-1/2}N_{k,w}\| &\leq \|C_{M,\lambda}^{-1/2}(C_M - \widehat{C}_M^{(w)})\|\|\widetilde{v}_{t+1}\| + \|C_{M,\lambda}^{-1/2}(\widehat{S}_M^{(w)\star}\widehat{y} - S_M^\star f_\rho)\| \\
&\leq 2\kappa\Big(\frac{2\kappa}{m\sqrt{\lambda}} + \sqrt{\frac{\mathcal{N}_M(\lambda)}{m}}\Big)\log\frac{6}{\delta}\Big(1+\sqrt{\frac{9\kappa^2}{M}\log\frac{3M}{\delta}}\max\Big(\sqrt{\eta t},\kappa^{-1}\Big)\Big) \\
&\quad + 2\sqrt{B}\Big(\frac{\kappa}{\sqrt{\lambda}m} + \sqrt{\frac{2\sqrt{p}\mathcal{N}_M(\lambda)}{m}}\Big)\log\frac{6}{\delta}.
\end{aligned}$$

Now a bound over the maximum $\max_{w\in V}\|C_{M,\lambda}^{-1/2}N_{k,w}\|$ is obtained by taking a union bound over $w \in V$. Meanwhile, an identical set of steps with $\widehat{C}_M^{(w)}, \widehat{S}_M^{(w),\top}$ swapped for $\widehat{C}_M, \widehat{S}_M$ yields the bound for $\|C_{M,\lambda}^{-1/2}N_k\|$ and $\|C_{M,\lambda}^{-1/2}(C_M - \widehat{C}_M)\|$.

F.2. Difference between Product of Empirical and Population Operators (Lemma 9)

In this section we provide the proof for Lemma 9.

Proof 12 (Lemma 9) *Begin by writing the quantity $\Pi^\Delta(w_{t:1})N$ using two auxiliary sequences. Initialized at $\gamma_1 = \gamma_1' = N$ and updated for $t \geq s \geq 1$ we have*

$$\begin{aligned}
\gamma_{s+1}' &= (I - \eta\widehat{C}_M^{(w_s)})\gamma_s' = \Pi(w_{s:1})N \\
\gamma_{s+1} &= (I - \eta C_M)\gamma_s = (I-\eta C_M)^s N
\end{aligned}$$

We can then write the difference as between these two sequences as the recursion

$$\begin{aligned}
\gamma_{s+1}' - \gamma_{s+1} &= (I-\eta C_M)(\gamma_s' - \gamma_s) + \eta\{C_M - \widehat{C}_M^{(w_s)}\}\gamma_s' \\
&= (I-\eta C_M)^s(\gamma_1' - \gamma_1) + \sum_{\ell=1}^{s}\eta(I-\eta C_M)^{s-\ell}\{C_M - \widehat{C}_M^{(w_\ell)}\}\gamma_\ell' \\
&= \sum_{\ell=1}^{s}\eta(I-\eta C_M)^{s-\ell}\{C_M - \widehat{C}_M^{(w_\ell)}\}\gamma_\ell'.
\end{aligned}$$

We then have

$$\begin{aligned}
\|C_M^{1/2-u}\Pi^\Delta(w_{t:1})N\| &= \|C_M^{1/2-u}(\gamma_{t+1}' - \gamma_{t+1})\| \\
&= \|\sum_{\ell=1}^{t}\eta C_M^{1/2-u}(I-\eta C_M)^{t-\ell}\{C_M - \widehat{C}_M^{(w_\ell)}\}\gamma_\ell'\| \\
&\leq \sum_{\ell=1}^{t}\eta\|C_M^{1/2-u}(I-\eta C_M)^{t-\ell}C_{M,\lambda}^{1/2}\|\|C_{M,\lambda}^{-1/2}(C_M - \widehat{C}_M^{(w_\ell)})\|\|\gamma_\ell'\| \\
&\leq \Delta_\lambda\|N\|\sum_{\ell=1}^{t}\eta\|C_M^{1/2-u}(I-\eta C_M)^{t-\ell}C_{M,\lambda}^{1/2}\|
\end{aligned}$$

where we have taken out the maximum over the $w_\ell \in V$ for $\|C_{M,\lambda}^{-1/2}(C_{M,\lambda} - \widehat{C}_M^{(w_\ell)})\|$ and simply bounded $\|\gamma_\ell'\| = \|(I-\eta\widehat{C}_M^{(w_{\ell-1})})\gamma_{\ell-1}'\| \leq \|\gamma_{\ell-1}'\| \leq \|N\|$ from $\eta\kappa^2 \leq 1$.

F.3. Convolution of Difference between Product of Empirical and Population Operators (Lemma 13)

This section provides the proof of Lemma 13.

Proof 13 (Lemma 13) *Begin by observing that this quantity can be written as*

$$\begin{aligned}\sum_{w_{t:1}\in V^t}\Delta(w_{t:1})\Pi^{\Delta}(w_{t:1})N &= \sum_{w_{t:1}\in V^t}\Delta(w_{t:1})\Pi(w_{t:1})N - \sum_{w_{t:1}\in V^t}\Delta(w_{t:1})(I-\eta C_M)^t N\\ &= \sum_{w_{t:1}\in V^t}\Delta(w_{t:1})\Pi(w_{t:1})N\end{aligned}$$

since $\sum_{w_{t:1}\in V^t}\Delta(w_{t:1}) = 0$*. Now introduce the following auxiliary variables. Initialized as* $\gamma_{1,w} = \gamma'_{1,w} = N$ *for all* $w \in V$ *we update the sequences for* $t \geq s \geq 1$

$$\begin{aligned}\gamma_{s+1,v} &= \sum_{w\in V} P_{vw}(I-\eta\widehat{C}_M^{(w)})\gamma_{s,w} = \sum_{w_{s:1}\in V^s} P_{vw_{s:1}}\Pi(w_{s:1})N \\ \gamma'_{s+1,v} &= \sum_{w\in V}\frac{1}{n}(I-\eta\widehat{C}_M^{(w)})\gamma'_{s,w} = \sum_{w_{s:1}\in V^s}\frac{1}{n^s}\Pi(w_{s:1})N.\end{aligned} \tag{13}$$

The quantity bounded within Lemma 13 can then be seen as the difference

$$\|C_M^{1/2}(\gamma_{t+1,v}-\gamma'_{t+1,v})\| = \Big\|\sum_{w_{t:1}\in V^t}\Delta(w_{t:1})C_M^{1/2}\Pi(w_{t:1})N\Big\|.$$

Introducing the auxiliary sequence $\{\gamma'_s\}_{s\geq 1}$ *independent of the agents. Also initialised* $\gamma'_{1,w} = N =: \gamma'_1$ *for all* $w \in V$ *we have due to averaging over all of the agents uniformly* $\gamma'_{2,w} = \gamma'_2 = (I-\eta\widehat{C}_M)N$ *for all* $w \in V$*. Applying this recursively we have for* $s \geq 1$ *and* $v \in V$

$$\gamma'_{s+1,v} = \gamma'_{s+1} = (I-\eta\widehat{C}_M)^s N.$$

Combined with the fact that the iterates $\{\gamma_{s,v}\}_{s\in[t],v\in V}$ *can be written and unravelled*

$$\begin{aligned}\gamma_{t+1,v} &= \sum_{w\in V}P_{vw}\big((I-\eta\widehat{C}_M)\gamma_{t,w} + \eta\{\widehat{C}_M - \widehat{C}_M^{(w)}\}\gamma_{t,w}\big)\\ &= (I-\eta\widehat{C}_M)^t N + \eta\sum_{k=1}^{t}\sum_{w\in V}(P^{t-k+1})_{vw}(I-\eta\widehat{C}_M)^{t-k}\{\widehat{C}_M-\widehat{C}_M^{(w)}\}\gamma_{k,w},\end{aligned}$$

means the difference is written as

$$\gamma_{t+1,v}-\gamma'_{t+1,v} = \eta\sum_{k=1}^{t}\sum_{w\in V}(P^{t-k+1})_{vw}(I-\eta\widehat{C}_M)^{t-k}\{\widehat{C}_M-\widehat{C}_M^{(w)}\}\gamma_{k,w}.$$

To analyse the difference $\gamma_{t+1,v}-\gamma'_{t+1,v}$ *we then consider the following decomposition where we denote the network averaged iterates* $\overline{\gamma}_t = \frac{1}{n}\sum_{w\in V}\gamma_{t,w}$

$$\|C_M^{1/2}(\gamma_{t+1,v}-\gamma'_{t+1,v})\| \leq \underbrace{\|C_M^{1/2}(\gamma_{t+1,v}-\overline{\gamma}_{t+1})\|}_{\textit{Term 1}} + \underbrace{\|C_M^{1/2}(\overline{\gamma}_{t+1}-\gamma'_{t+1})\|}_{\textit{Term 2}} \tag{14}$$

It is clear the network average can be written using the fact that the communication matrix P *is doubly stochastic i.e.* $\sum_{v\in V}P_{vw}^{t-k+1} = 1$ *as follows*

$$\overline{\gamma}_{t+1}-\gamma'_{t+1} = \frac{1}{n}\sum_{v\in V}\gamma_{t+1,v}-\gamma'_{t+1} = \eta\sum_{k=1}^{t}\frac{1}{n}\sum_{w\in V}(I-\eta\widehat{C}_M)^{t-k}\{\widehat{C}_M-\widehat{C}_M^{(w)}\}\gamma_{k,w}.$$

When taking the difference we then arrive at

$$\gamma_{t+1,v} - \gamma'_{t+1} - (\overline{\gamma}_{t+1} - \gamma'_{t+1}) = \eta \sum_{k=1}^{t} \sum_{w \in V} ((P^{t-k+1})_{vw} - \frac{1}{n})(I - \eta \widehat{C}_M)^{t-k} \{ \widehat{C}_M - \widehat{C}_M^{(w)} \} \gamma_{k,w}$$

*We can then bound **Term 1** with* $\lambda_1 > 0$

$$\begin{aligned}
&\|C_M^{1/2}(\gamma_{t+1,v} - \overline{\gamma}_{t+1})\| \\
&\leq \eta \sum_{k=1}^{t} \sum_{w \in V} |(P^{t-k+1})_{vw} - \frac{1}{n}| \|C_M^{1/2}(I - \eta \widehat{C}_M)^{t-k} C_{M,\lambda_1}^{1/2}\| \|C_{M,\lambda_1}^{-1/2}\{\widehat{C}_M - \widehat{C}_M^{(w)}\}\| \|\gamma_{k,w}\| \\
&\leq 2\eta \Delta_{\lambda_1} \|N\| \sum_{k=1}^{t} \|C_M^{1/2}(I - \eta \widehat{C}_M^{1/2})^{t-k} C_{M,\lambda_1}^{1/2}\| \big(\sum_{w \in V} |(P^{t-k+1})_{vw} - \frac{1}{n}| \big) \\
&\leq 4\eta \Delta_{\lambda_1} \|N\| \sum_{k=1}^{t} \|C_M^{1/2}(I - \eta \widehat{C}_M)^{t-k} C_{M,\lambda_1}^{1/2}\| (\sigma_2^{t-k+1} \wedge 1)
\end{aligned}$$

where we have used that $\|\gamma_{s+1,v}\| \leq \sum_{w \in V} P_{vw} \|(I - \eta \widehat{C}_M^{(w)})\gamma_{s,w}\| \leq \sum_{w \in V} P_{vw} \|\gamma_{s,w}\| \leq \|N\|$ *as well as*

$$\begin{aligned}
\|C_{M,\lambda_1}^{-1/2}(\widehat{C}_M - \widehat{C}_M^{(w)})\| &\leq \|C_{M,\lambda_1}^{-1/2}(\widehat{C}_M - C_M)\| + \|C_{M,\lambda_1}^{-1/2}(C_M - \widehat{C}_M^{(w)})\| \\
&\leq \frac{1}{n} \sum_{v \in V} \|C_{M,\lambda_1}^{-1/2}(C_M - \widehat{C}_M^{(v)})\| + \|C_{M,\lambda_1}^{-1/2}(C_M - \widehat{C}_M^{(w)})\| \\
&\leq 2\Delta_{\lambda_1}
\end{aligned}$$

in addition to Lemma 5 to bound $\sum_{w \in V} |(P^{t-k+1})_{vw} - \frac{1}{n}| = \sum_{w \in V} |\Delta^{t-k+1}(v,w)|$.

*To bound **Term 2** we note that we can rewrite*

$$\overline{\gamma}_{t+1} - \gamma'_{t+1} = \eta \sum_{k=2}^{t} \frac{1}{n} \sum_{w \in V} (I - \eta \widehat{C}_M)^{t-k} \{ \widehat{C}_M - \widehat{C}_M^{(w)} \} (\gamma_{k,w} - \overline{\gamma}_k).$$

where $\frac{1}{n} \sum_{w \in V} (I - \eta \widehat{C}_M)^{t-k} \{ \widehat{C}_M - \widehat{C}_M^{(w)} \} \overline{\gamma}_k = 0$ *for* $k \geq 1$. *Applying triangle inequality as well as similar step to previously, we get with* $\lambda_2, \lambda_3 \geq 0$

$$\begin{aligned}
\|C_M^{1/2}(\overline{\gamma}_{t+1} - \gamma'_{t+1})\| &\leq \eta \sum_{k=2}^{t} \|C_M^{1/2}(I - \eta \widehat{C}_M)^{t-k} C_{M,\lambda_2}^{1/2}\| \frac{1}{n} \sum_{w \in V} \|C_{M,\lambda_2}^{-1/2}(\widehat{C}_M - \widehat{C}_M^{(w)})\| \|\gamma_{k,w} - \overline{\gamma}_k\| \\
&\leq 8\eta^2 \Delta_{\lambda_2} \Delta_{\lambda_3} \|N\| \sum_{k=2}^{t} \sum_{\ell=1}^{k-1} \|C_M^{1/2}(I - \eta \widehat{C}_M)^{t-k} C_{M,\lambda_2}^{1/2}\| \|(I - \eta \widehat{C}_M)^{k-1-\ell} C_{M,\lambda_3}^{1/2}\| (\sigma_2^{k-\ell} \wedge 1)
\end{aligned}$$

*where we plugged in the bound from **Term 1** for the deviation* $\|\gamma_{k,w} - \overline{\gamma}_k\|$ *for* $k \geq 2$.