A Proof of the main theorem

The proof of the main theorem is presented in this section. We start with a generic result on a family of AMP iterations including the (non Bayes-optimal) MLAMP one, using the framework of Gerbelot and Berthier [2021], from which we remind the required notions.

A.1 Notations and definitions

If $f : \mathbb{R}^{N\times q} \to \mathbb{R}^{N\times q}$ is an function and $i \in \{1, \ldots N\}$, we write $f_i : \mathbb{R}^{N\times q} \to \mathbb{R}^q$ the component of f generating the i-th row of its image, i.e., if $\mathbf{X} \in \mathbb{R}^{N\times q}$,

$$f(\mathbf{X}) = \begin{bmatrix} f_1(\mathbf{X}) \\ \vdots \\ f_N(\mathbf{X}) \end{bmatrix} \in \mathbb{R}^{N\times q}\,.$$

We write $\frac{\partial f_i}{\partial \mathbf{X}_i}$ the $q \times q$ Jacobian containing the derivatives of f_i with respect to (w.r.t.) the i-th row $\mathbf{X}_i \in \mathbb{R}^q$:

$$\frac{\partial f_i}{\partial \mathbf{X}_i} = \begin{bmatrix} \frac{\partial (f_i(\mathbf{X}))_1}{\partial \mathbf{X}_{i1}} & \cdots & \frac{\partial (f_i(\mathbf{X}))_1}{\partial \mathbf{X}_{iq}} \\ \vdots & & \vdots \\ \frac{\partial (f_i(\mathbf{X}))_q}{\partial \mathbf{X}_{i1}} & \cdots & \frac{\partial (f_i(\mathbf{X}))_q}{\partial \mathbf{X}_{iq}} \end{bmatrix} \in \mathbb{R}^{q\times q}\,. \tag{9}$$

For two sequences of random variables X_n, Y_n, we write $X_n \overset{\mathrm{P}}{\simeq} Y_n$ when their difference converges in probability to 0, i.e., $X_n - Y_n \xrightarrow{P} 0$. Oriented graphs with a set of vertices V and edges $\overrightarrow{E}$ are denoted $G = (V, \overrightarrow{E})$. The set of edges may be split into right-pointing and left-pointing edges, i.e., $\overrightarrow{E} = \{\overrightarrow{e}_1, ..., \overrightarrow{e}_L\}$, $\overleftarrow{E} = \{\overleftarrow{e}_1, ..., \overleftarrow{e}_L\}$.

Definition A.1 (pseudo-Lipschitz function)**.** For $k \in \mathbb{N}^*$ and any $N, m \in \mathbb{N}^*$, a function $\Phi : \mathbb{R}^{N\times q} \to \mathbb{R}^{m\times q}$ is said to be *pseudo-Lipschitz of order k* if there exists a constant L such that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{N\times q}$,

$$\frac{\|\Phi(\mathbf{x}) - \Phi(\mathbf{y})\|_F}{\sqrt{m}} \leqslant L\left(1 + \left(\frac{\|\mathbf{x}\|_F}{\sqrt{N}}\right)^{k-1} + \left(\frac{\|\mathbf{y}\|_F}{\sqrt{N}}\right)^{k-1}\right)\frac{\|\mathbf{x}-\mathbf{y}\|_F}{\sqrt{N}} \tag{10}$$

For a function $\varphi : \mathbb{R} \to \mathbb{R}$, the property becomes

$$\forall\, (x, y) \in \mathbb{R}^2,\ |\varphi(x) - \varphi(y)| \leqslant L(1 + |x|^{k-1} + |y|^{k-1})|x - y| \tag{11}$$

and a straightforward calculation shows that for any scalar pseudo-Lipshitz function of order 2, the function

$$\phi : \mathbb{R}^d \to \mathbb{R}, \tag{12}$$

$$\mathbf{x} \mapsto \frac{1}{d}\sum_{i=1}^{d} \varphi(x_i) \tag{13}$$

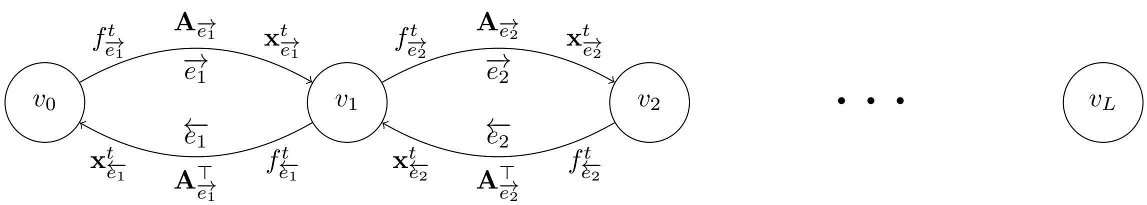
is pseudo-Lipschitz of order 2 according to the definition above. This definition is handy for proofs involving non-separable functions and leads to Gaussian concentration using the Gauss-Poincaré inequality (see Lemma C.8. from Berthier et al. [2020]), while in the separable case, a strong law of large number is proven for a class of distributions including sub-Gaussian ones in Lemma 5 of Bayati and Montanari [2011].

A.2 State evolution for generic multilayer AMP iterations with matrix valued variables and dense Gaussian matrices

In the notations of Gerbelot and Berthier [2021], consider the AMP iteration indexed by the following directed graph $G = (V, \overrightarrow{E})$, where the set of vertices is denoted $V = \{v_0, v_1, ..., v_L\}$, and the set of edges $\overrightarrow{E} = \{\overrightarrow{e}_1, ..., \overrightarrow{e}_l, \overleftarrow{e}_1, ..., \overleftarrow{e}_L\}$. For any edge $\overrightarrow{e}_l$, the corresponding matrix $\mathbf{A}_{\overrightarrow{e}_l}$ has dimensions $\mathbb{R}^{n_l\times n_{l-1}}$ with $\mathbf{A}_{\overleftarrow{e}_l} = \mathbf{A}^\top_{\overrightarrow{e}_l}$, and the variables $\mathbf{x}_{\overrightarrow{e}_l} \in \mathbb{R}^{n_l\times q}, \mathbf{x}_{\overleftarrow{e}_l} \in \mathbb{R}^{n_{l-1}\times q}$ for some finite $q \in \mathbb{N}$, with $N = \sum_{l=1}^{L} n_l$. Finally, we define the non-linearities of the iteration by specifying the variables they are acting on as follows:

- $f^t_{\overrightarrow{e}_1} : \mathbb{R}^{n_0 \times q} \to \mathbb{R}^{n_0 \times q}, \mathbf{x}^t_{\overleftarrow{e}_1} \mapsto f^t_{\overrightarrow{e}_1}\left(\mathbf{x}^t_{\overleftarrow{e_1}}\right)$,
- for any $2 \leqslant l \leqslant L$, $f^t_{\overrightarrow{e}_l} : (\mathbb{R}^{n_{l-1} \times q})^2 \to \mathbb{R}^{n_{l-1} \times q}$, $(\mathbf{x}^t_{\overrightarrow{e}_{l-1}}, \mathbf{x}^t_{\overleftarrow{e}_l}) \mapsto f^t_{\overrightarrow{e}_l}(\mathbf{x}^t_{\overrightarrow{e}_{l-1}}, \mathbf{x}^t_{\overleftarrow{e}_l})$,
- for any $1 \leqslant l \leqslant L-1$, $f^t_{\overleftarrow{e}_l} : (\mathbb{R}^{n_l \times q})^3 \to \mathbb{R}^{n_l \times q}$, $(\mathbf{x}^t_{\overrightarrow{e}_l}, \mathbf{x}^t_{\overleftarrow{e}_{l+1}}) \mapsto f^t_{\overleftarrow{e}_l}(\mathbf{A}_{\overrightarrow{e}_l}\mathbf{w}_{\overrightarrow{e}_l}, \mathbf{x}^t_{\overrightarrow{e}_l}, \mathbf{x}^t_{\overleftarrow{e}_{l+1}})$
- $f^t_{\overleftarrow{e}_L} : (\mathbb{R}^{n_L \times q})^2 \to \mathbb{R}^{n_L \times q}, \mathbf{x}^t_{\overleftarrow{e}_L} \to f^t_{\overleftarrow{e}_L}\left(\mathbf{A}_{\overrightarrow{e}_L}\mathbf{w}_{\overrightarrow{e}_L}, \mathbf{x}^t_{\overleftarrow{e_L}}\right)$

where $\mathbf{w}_{\overrightarrow{e}_1}, ..., \mathbf{w}_{\overrightarrow{e}_L}$ are low-rank matrices respectively in $\mathbb{R}^{n_0 \times q}, ..., \mathbb{R}^{n_{L-1} \times q}$, whose rows are sampled i.i.d. from subgaussian probability distributions in $\mathbb{R}^q$. The graph indexing the iteration then reads:



with the corresponding iteration:

$$
\begin{aligned}
\mathbf{x}^{t+1}_{\overrightarrow{e_1}} &= \mathbf{A}_{\overrightarrow{e_1}}\mathbf{m}^t_{\overrightarrow{e_1}} - \mathbf{m}^{t-1}_{\overleftarrow{e_1}}\left(\mathbf{b}^t_{\overrightarrow{e_1}}\right)^\top , \\
\mathbf{m}^t_{\overrightarrow{e_1}} &= f^t_{\overrightarrow{e}_1}\left(\mathbf{x}^t_{\overleftarrow{e_1}}\right) , \\
\mathbf{x}^{t+1}_{\overleftarrow{e_1}} &= \mathbf{A}^\top_{\overrightarrow{e_1}}\mathbf{m}^t_{\overleftarrow{e_1}} - \mathbf{m}^{t-1}_{\overrightarrow{e_1}}\left(\mathbf{b}^t_{\overleftarrow{e_1}}\right)^\top , \\
\mathbf{m}^t_{\overleftarrow{e_1}} &= f^t_{\overleftarrow{e_1}}\left(\mathbf{A}_{\overrightarrow{e}_1}\mathbf{w}_{\overrightarrow{e}_1}, \mathbf{x}^t_{\overrightarrow{e_1}}, \mathbf{x}^t_{\overleftarrow{e_2}}\right) , \\
\\
\mathbf{x}^{t+1}_{\overrightarrow{e_2}} &= \mathbf{A}_{\overrightarrow{e_2}}\mathbf{m}^t_{\overrightarrow{e_2}} - \mathbf{m}^{t-1}_{\overleftarrow{e_2}}\left(\mathbf{b}^t_{\overrightarrow{e_2}}\right)^\top , \\
\mathbf{m}^t_{\overrightarrow{e_2}} &= f^t_{\overrightarrow{e}_2}\left(\mathbf{x}^t_{\overrightarrow{e_1}}, \mathbf{x}^t_{\overleftarrow{e_2}}\right) , \\
\mathbf{x}^{t+1}_{\overleftarrow{e_2}} &= \mathbf{A}^\top_{\overrightarrow{e_2}}\mathbf{m}^t_{\overleftarrow{e_2}} - \mathbf{m}^{t-1}_{\overrightarrow{e_2}}(\mathbf{b}^t_{\overleftarrow{e_2}})^\top , \qquad (14) \\
\mathbf{m}^t_{\overleftarrow{e_2}} &= f^t_{\overleftarrow{e_2}}\left(\mathbf{A}_{\overrightarrow{e}_2}\mathbf{w}_{\overrightarrow{e}_2}, \mathbf{x}^t_{\overrightarrow{e_2}}, \mathbf{x}^t_{\overleftarrow{e_3}}\right) , \\
\\
&\vdots \\
\\
\mathbf{x}^{t+1}_{\overrightarrow{e_L}} &= \mathbf{A}_{\overrightarrow{e_L}}\mathbf{m}^t_{\overrightarrow{e_L}} - \mathbf{m}^{t-1}_{\overleftarrow{e_L}}\left(\mathbf{b}^t_{\overrightarrow{e_L}}\right)^\top , \\
\mathbf{m}^t_{\overrightarrow{e_L}} &= f^t_{\overrightarrow{e}_L}\left(\mathbf{x}^t_{\overrightarrow{e}_{L-1}}, \mathbf{x}^t_{\overleftarrow{e_L}}\right) , \\
\mathbf{x}^{t+1}_{\overleftarrow{e_L}} &= \mathbf{A}^\top_{\overrightarrow{e_L}}\mathbf{m}^t_{\overleftarrow{e_L}} - \mathbf{m}^{t-1}_{\overrightarrow{e_L}}(\mathbf{b}^t_{\overleftarrow{e_L}})^\top , \\
\mathbf{m}^t_{\overleftarrow{e_L}} &= f^t_{\overleftarrow{e_L}}\left(\mathbf{A}_{\overrightarrow{e}_L}\mathbf{w}_{\overrightarrow{e}_L}, \mathbf{x}^t_{\overrightarrow{e_L}}\right)
\end{aligned}
$$

and Onsager terms, for the right oriented edges

$$
\mathbf{b}^t_{\overrightarrow{e}_l} = \frac{1}{N}\sum_{i=1}^{n_{l-1}} \frac{\partial f^t_{\overrightarrow{e}_l,i}}{\partial \mathbf{x}_{\overleftarrow{e}_l,i}}\left(\left(\mathbf{x}^t_{\overrightarrow{e}'_l}\right)_{\overrightarrow{e}'_l : \overrightarrow{e}'_l \to \overrightarrow{e}_l}\right) \qquad \in \mathbb{R}^{q \times q} .
$$

and left oriented edges

$$\mathbf{b}^t_{\overleftarrow{e}_l} = \frac{1}{N}\sum_{i=1}^{n_l} \frac{\partial f^t_{\overleftarrow{e}_l,i}}{\partial \mathbf{x}_{\overrightarrow{e}_l,i}}\left(\mathbf{A}_{\overrightarrow{e}_l}\mathbf{w}_{\overrightarrow{e}_l}, \left(\mathbf{x}^t_{\overleftarrow{e}'_l}\right)_{\overleftarrow{e}'_l:\overleftarrow{e}'_l\to\overleftarrow{e}_l}\right) \qquad \in \mathbb{R}^{q\times q}.$$

We now make the following assumptions

- **(A1)** The matrices $(\mathbf{A}_{\overrightarrow{e}})_{\overrightarrow{e}\in\overrightarrow{E}}$ are random and independent, up to the symmetry condition $\mathbf{A}_{\overleftarrow{e}} = \mathbf{A}^\top_{\overrightarrow{e}}$. Moreover $\mathbf{A}_{\overrightarrow{e}}$ has independent centered Gaussian entries with variance $1/N$.
- **(A2)** For all $1 \leqslant l \leqslant L$, $n_l \to \infty$ and n_l/N converges to a well-defined limit $\delta_l \in [0,1]$. We denote by $n \to \infty$ the limit under this scaling.
- **(A3)** For all $t \in \mathbb{N}$ and $\overrightarrow{e} \in \overrightarrow{E}$, the non-linearity $f^t_{\overrightarrow{e}}$ is pseudo-Lipschitz of finite order, uniformly with respect to the problem dimensions $(n_l)_{0\leqslant l\leqslant L}$
- **(A4)** For all $\overrightarrow{e} \in E$, the rows of $\mathbf{x}^0_{\overrightarrow{e}}, \mathbf{w}_{\overrightarrow{e}}$ are sampled from subgaussian probability distributions in $\mathbf{R}^q$.
- **(A5)** For all $\overrightarrow{e} \in E$, the following limit exists and is finite:

$$\lim_{n\to\infty}\frac{1}{N}\left\langle f^0_{\overrightarrow{e}}\left(\left(\mathbf{x}^0_{\overrightarrow{e}'}\right)_{\overrightarrow{e}':\overrightarrow{e}'\to\overrightarrow{e}}\right), f^0_{\overrightarrow{e}}\left(\left(\mathbf{x}^0_{\overrightarrow{e}'}\right)_{\overrightarrow{e}':\overrightarrow{e}'\to\overrightarrow{e}}\right)\right\rangle$$

- **(A6)** Let $(\kappa_{\overrightarrow{e}})_{\overrightarrow{e}\in E}$ be an array of bounded non-negative reals and $\mathbf{Z}_{\overrightarrow{e}} \sim \mathbf{N}(0, \kappa_{\overrightarrow{e}}\mathbf{I}_{n_w})$ independent random variables for all $\overrightarrow{e}$. For all $\overrightarrow{e} \in E$, for any $t \in \mathbb{N}_{>0}$, the following limit exists and is finite:

$$\lim_{n\to\infty}\frac{1}{N}\mathbb{E}\left[\left\langle f^0_{\overrightarrow{e}}\left(\left(\mathbf{x}^0_{\overrightarrow{e}'}\right)_{\overrightarrow{e}':\overrightarrow{e}'\to\overrightarrow{e}}\right), f^t_{\overrightarrow{e}}\left(\left(\mathbf{Z}^t_{\overrightarrow{e}'}\right)_{\overrightarrow{e}':\overrightarrow{e}'\to\overrightarrow{e}}\right)\right\rangle\right].$$

- **(A7)** Consider any array of 2×2 positive definite matrices $(\boldsymbol{S}_{\overrightarrow{e}})_{\overrightarrow{e}\in E}$ and the collection of random variables $(\mathbf{Z}_{\overrightarrow{e}}, \mathbf{Z}'_{\overrightarrow{e}}) \sim \mathbf{N}(0, \boldsymbol{S}_{\overrightarrow{e}} \otimes \mathbf{I}_{n_w}))$ defined independently for each edge $\overrightarrow{e}$. Then for any $\overrightarrow{e} \in E$ and $s, t > 0$, the following limit exists and is finite:

$$\lim_{n\to\infty}\frac{1}{N}\mathbb{E}\left[\left\langle f^s_{\overrightarrow{e}}\left(\left(\mathbf{Z}^s_{\overrightarrow{e}'}\right)_{\overrightarrow{e}':\overrightarrow{e}'\to\overrightarrow{e}}\right), f^t_{\overrightarrow{e}}\left(\left(\tilde{\mathbf{Z}}^t_{\overrightarrow{e}'}\right)_{\overrightarrow{e}':\overrightarrow{e}'\to\overrightarrow{e}}\right)\right\rangle\right].$$

Under these assumptions, we define the following state evolution recursion:

- for $l = 1$:

$$\boldsymbol{\nu}^0_{\overrightarrow{e}_1} = \lim_{N\to\infty}\frac{1}{N}\mathbf{w}^\top_{\overrightarrow{e}_1} f^0_{\overrightarrow{e}_1}(\mathbf{x}^0_{\overleftarrow{e}_1}),\ \boldsymbol{\kappa}^{1,1}_{\overrightarrow{e}_1} = \lim_{N\to\infty}\frac{1}{N} f^0_{\overrightarrow{e}_1}(\mathbf{x}^0_{\overleftarrow{e}_1})^\top f^0_{\overrightarrow{e}_1}(\mathbf{x}^0_{\overleftarrow{e}_1}) \tag{15}$$

$$\boldsymbol{\nu}^{t+1}_{\overrightarrow{e}_1} = \lim_{N\to+\infty}\frac{1}{N}\mathbb{E}\left[\mathbf{w}^\top_{\overrightarrow{e}_1} f^t_{\overrightarrow{e}_1}\left(\mathbf{w}_{\overrightarrow{e}_1}\hat{\boldsymbol{\nu}}^t_{\overleftarrow{e}_1} + \mathbf{Z}^t_{\overleftarrow{e}_1}\right)\right] \tag{16}$$

$$\begin{aligned}\boldsymbol{\kappa}^{s+1,t+1}_{\overrightarrow{e}_1} = \boldsymbol{\kappa}^{t+1,s+1}_{\overrightarrow{e}_1} = \lim_{N\to+\infty}\frac{1}{N}\mathbb{E}\Bigg[&\left(f^s_{\overrightarrow{e}_1}\left(\mathbf{w}_{\overrightarrow{e}_1}\hat{\boldsymbol{\nu}}^s_{\overleftarrow{e}_1} + \mathbf{Z}^s_{\overleftarrow{e}_1}\right) - \mathbf{w}_{\overrightarrow{e}_1}\rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_1}}\boldsymbol{\nu}^{s+1}_{\overrightarrow{e}_1}\right)^\top \\ &\left(f^t_{\overrightarrow{e}_1}\left(\mathbf{w}_{\overrightarrow{e}_1}\hat{\boldsymbol{\nu}}^t_{\overleftarrow{e}_1} + \mathbf{Z}^t_{\overleftarrow{e}_1}\right) - \mathbf{w}_{\overrightarrow{e}_1}\rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_1}}\boldsymbol{\nu}^{t+1}_{\overrightarrow{e}_1}\right)\Bigg]\end{aligned} \tag{17}$$

$$\hat{\boldsymbol{\nu}}^0_{\overleftarrow{e}_1}, \boldsymbol{\kappa}^{1,1}_{\overleftarrow{e}_1} = \lim_{n\to\infty}\frac{1}{N} f^0_{\overleftarrow{e}_1}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_1}}, \mathbf{x}^0_{\overrightarrow{e}_1}, \mathbf{x}^0_{\overleftarrow{e}_2}\right)^\top f^0_{\overleftarrow{e}_1}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_1}}, \mathbf{x}^0_{\overrightarrow{e}_1}, \mathbf{x}^0_{\overleftarrow{e}_2}\right) \tag{18}$$

$$\hat{\boldsymbol{\nu}}^{t+1}_{\overleftarrow{e}_1} = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\left[\sum_{i=1}^N \frac{\partial f^t_{\overleftarrow{e}_1,i}}{\partial \mathbf{z}_{\mathbf{w}_{\overleftarrow{e}_1},i}, \phi_{\overleftarrow{e}_1}}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_1}}, \mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_1}}\rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_1}}\boldsymbol{\nu}^t_{\overrightarrow{e}_1} + \mathbf{Z}^t_{\overrightarrow{e}_1}, \mathbf{w}_{\overrightarrow{e}_2}\hat{\boldsymbol{\nu}}^t_{\overleftarrow{e}_2} + \mathbf{Z}^t_{\overleftarrow{e}_2}\right)\right] \tag{19}$$

$$\begin{aligned}\boldsymbol{\kappa}^{s+1,t+1}_{\overleftarrow{e}_1} = \lim_{n\to\infty}\frac{1}{N}\mathbb{E}\Bigg[&f^s_{\overleftarrow{e}_1}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_1}}, \mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_1}}\rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_1}}\boldsymbol{\nu}^s_{\overrightarrow{e}_1} + \mathbf{Z}^s_{\overrightarrow{e}_1}, \mathbf{w}_{\overrightarrow{e}_2}\hat{\boldsymbol{\nu}}^s_{\overleftarrow{e}_2} + \mathbf{Z}^s_{\overleftarrow{e}_2}\right)^\top \\ &f^t_{\overleftarrow{e}_1}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_1}}, \mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_1}}\rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_1}}\boldsymbol{\nu}^t_{\overrightarrow{e}_1} + \mathbf{Z}^t_{\overrightarrow{e}_1}, \mathbf{w}_{\overrightarrow{e}_2}\hat{\boldsymbol{\nu}}^t_{\overleftarrow{e}_2} + \mathbf{Z}^t_{\overleftarrow{e}_2}\right)\Bigg]\end{aligned} \tag{20}$$

- for any $2 \leqslant l \leqslant L-1$

$$\boldsymbol{\nu}^0_{\overrightarrow{e}_l} = \lim_{N\to\infty} \frac{1}{N} \mathbf{w}^\top_{\overrightarrow{e}_l} f^0_{\overrightarrow{e}_l}(\mathbf{x}^0_{\overleftarrow{e}_l}),\ \boldsymbol{\kappa}^{1,1}_{\overrightarrow{e}_l} = \lim_{N\to\infty} \frac{1}{N} f^0_{\overrightarrow{e}_l}(\mathbf{x}^0_{\overleftarrow{e}_l})^\top f^0_{\overrightarrow{e}_l}(\mathbf{x}^0_{\overleftarrow{e}_l}) \tag{21}$$

$$\boldsymbol{\nu}^{t+1}_{\overrightarrow{e}_l} = \lim_{N\to+\infty} \frac{1}{N} \mathbb{E}\left[\mathbf{w}^\top_{\overrightarrow{e}_l} f^t_{\overrightarrow{e}_l}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_{l-1}}} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_{l-1}}} \boldsymbol{\nu}^t_{\overrightarrow{e}_{l-1}} + \mathbf{Z}^t_{\overrightarrow{e}_{l-1}}, \mathbf{w}_{\overrightarrow{e}_l} \hat{\boldsymbol{\nu}}^t_{\overleftarrow{e}_l} + \mathbf{Z}^t_{\overleftarrow{e}_l}\right)\right] \tag{22}$$

$$\boldsymbol{\kappa}^{s+1,t+1}_{\overrightarrow{e}_l} = \boldsymbol{\kappa}^{t+1,s+1}_{\overrightarrow{e}_l} = \lim_{N\to+\infty} \tag{23}$$

$$\frac{1}{N}\mathbb{E}\Bigg[\left(f^s_{\overrightarrow{e}_l}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_{l-1}}} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_{l-1}}} \boldsymbol{\nu}^s_{\overrightarrow{e}_{l-1}} + \mathbf{Z}^s_{\overrightarrow{e}_{l-1}}, \mathbf{w}_{\overrightarrow{e}_l} \hat{\boldsymbol{\nu}}^s_{\overleftarrow{e}_l} + \mathbf{Z}^s_{\overleftarrow{e}_l}\right) - \mathbf{w}_{\overrightarrow{e}_l} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_l}} \boldsymbol{\nu}^{s+1}_{\overrightarrow{e}_l}\right)^\top$$

$$\left(f^t_{\overrightarrow{e}_l}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_{l-1}}} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_{l-1}}} \boldsymbol{\nu}^t_{\overrightarrow{e}_{l-1}} + \mathbf{Z}^t_{\overrightarrow{e}_{l-1}}, \mathbf{w}_{\overrightarrow{e}_l} \hat{\boldsymbol{\nu}}^t_{\overleftarrow{e}_l} + \mathbf{Z}^t_{\overleftarrow{e}_l}\right) - \mathbf{w}_{\overrightarrow{e}_l} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_l}} \boldsymbol{\nu}^{t+1}_{\overrightarrow{e}_l}\right)\Bigg] \tag{24}$$

$$\hat{\boldsymbol{\nu}}^0_{\overleftarrow{e}_l}, \boldsymbol{\kappa}^{1,1}_{\overleftarrow{e}_l} = \lim_{n\to\infty} \frac{1}{N} f^0_{\overleftarrow{e}_l}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_l}}, \mathbf{x}^0_{\overrightarrow{e}_l}, \mathbf{x}^0_{\overleftarrow{e}_{l+1}}\right)^\top f^0_{\overleftarrow{e}_l}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_l}}, \mathbf{x}^0_{\overrightarrow{e}_l}, \mathbf{x}^0_{\overleftarrow{e}_{l+1}}\right) \tag{25}$$

$$\hat{\boldsymbol{\nu}}^{t+1}_{\overleftarrow{e}_l} = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\Bigg[\sum_{i=1}^N \frac{\partial f^t_{\overleftarrow{e}_l,i}}{\partial \mathbf{z}_{\mathbf{w}_{\overleftarrow{e}_l},i}, \phi_{\overleftarrow{e}_l}}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_l}}, \mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_l}} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_l}} \boldsymbol{\nu}^t_{\overrightarrow{e}_l} + \mathbf{Z}^t_{\overrightarrow{e}_l}, \mathbf{w}_{\overrightarrow{e}_{l+1}} \hat{\boldsymbol{\nu}}^t_{\overrightarrow{e}_{l+1}} \mathbf{Z}^t_{\overleftarrow{e}_{l+1}}\right)\Bigg] \tag{26}$$

$$\boldsymbol{\kappa}^{s+1,t+1}_{\overleftarrow{e}_l} = \lim_{n\to\infty} \frac{1}{N}\mathbb{E}\Bigg[f^s_{\overleftarrow{e}_l}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_l}}, \mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_l}} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_l}} \boldsymbol{\nu}^s_{\overrightarrow{e}_l} + \mathbf{Z}^s_{\overrightarrow{e}_l}, \mathbf{w}_{\overrightarrow{e}_{l+1}} \hat{\boldsymbol{\nu}}^s_{\overrightarrow{e}_{l+1}} \mathbf{Z}^s_{\overleftarrow{e}_{l+1}}\right)^\top$$

$$f^t_{\overleftarrow{e}_l}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_l}}, \mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_l}} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_l}} \boldsymbol{\nu}^t_{\overrightarrow{e}_l} + \mathbf{Z}^t_{\overrightarrow{e}_l}, \mathbf{w}_{\overrightarrow{e}_{l+1}} \hat{\boldsymbol{\nu}}^t_{\overrightarrow{e}_{l+1}} \mathbf{Z}^t_{\overleftarrow{e}_{l+1}}\right)\Bigg] \tag{27}$$

- for l=L

$$\boldsymbol{\nu}^0_{\overrightarrow{e}_L} = \lim_{N\to\infty} \frac{1}{N} \mathbf{w}^\top_{\overrightarrow{e}_l} f^0_{\overrightarrow{e}_L}(\mathbf{x}^0_{\overleftarrow{e}_L}),\ \boldsymbol{\kappa}^{1,1}_{\overrightarrow{e}_L} = \lim_{N\to\infty} \frac{1}{N} f^0_{\overrightarrow{e}_L}(\mathbf{x}^0_{\overleftarrow{e}_L})^\top f^0_{\overrightarrow{e}_L}(\mathbf{x}^0_{\overleftarrow{e}_L}) \tag{28}$$

$$\boldsymbol{\nu}^{t+1}_{\overrightarrow{e}_L} = \lim_{N\to+\infty} \frac{1}{N} \mathbb{E}\left[\mathbf{w}^\top_{\overrightarrow{e}_L} f^t_{\overrightarrow{e}_L}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_{L-1}}} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_{L-1}}} \boldsymbol{\nu}^t_{\overrightarrow{e}_{L-1}} + \mathbf{Z}^t_{\overrightarrow{e}_{L-1}}, \mathbf{w}_{\overrightarrow{e}_L} \hat{\boldsymbol{\nu}}^t_{\overleftarrow{e}_L} + \mathbf{Z}^t_{\overleftarrow{e}_L}\right)\right] \tag{29}$$

$$\boldsymbol{\kappa}^{s+1,t+1}_{\overrightarrow{e}_L} = \boldsymbol{\kappa}^{t+1,s+1}_{\overrightarrow{e}_L} = \lim_{N\to+\infty} \tag{30}$$

$$\frac{1}{N}\mathbb{E}\Bigg[\left(f^s_{\overrightarrow{e}_L}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_{L-1}}} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_{L-1}}} \boldsymbol{\nu}^s_{\overrightarrow{e}_{L-1}} + \mathbf{Z}^s_{\overrightarrow{e}_{L-1}}, \mathbf{w}_{\overrightarrow{e}_L} \hat{\boldsymbol{\nu}}^s_{\overleftarrow{e}_L} + \mathbf{Z}^s_{\overleftarrow{e}_L}\right) - \mathbf{w}_{\overrightarrow{e}_L} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_L}} \boldsymbol{\nu}^{s+1}_{\overrightarrow{e}_L}\right)^\top$$

$$\left(f^t_{\overrightarrow{e}_L}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_{L-1}}} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_{L-1}}} \boldsymbol{\nu}^t_{\overrightarrow{e}_{L-1}} + \mathbf{Z}^t_{\overrightarrow{e}_{L-1}}, \mathbf{w}_{\overrightarrow{e}_L} \hat{\boldsymbol{\nu}}^t_{\overleftarrow{e}_L} + \mathbf{Z}^t_{\overleftarrow{e}_L}\right) - \mathbf{w}_{\overrightarrow{e}_L} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_L}} \boldsymbol{\nu}^{t+1}_{\overrightarrow{e}_L}\right)\Bigg] \tag{31}$$

$$\hat{\boldsymbol{\nu}}^0_{\overleftarrow{e}_L}, \boldsymbol{\kappa}^{1,1}_{\overleftarrow{e}_L} = \lim_{n\to\infty} \frac{1}{N} f^0_{\overleftarrow{e}_L}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_L}}, \mathbf{x}^0_{\overrightarrow{e}_L}\right)^\top f^0_{\overleftarrow{e}_L}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_L}}\right) \tag{32}$$

$$\hat{\boldsymbol{\nu}}^{t+1}_{\overleftarrow{e}_L} = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\Bigg[\sum_{i=1}^N \frac{\partial f^t_{\overleftarrow{e}_L,i}}{\partial \mathbf{z}_{\mathbf{w}_{\overleftarrow{e}_L},i}, \phi_{\overleftarrow{e}_L}}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_L}}, \mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_L}} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_L}} \boldsymbol{\nu}^t_{\overrightarrow{e}_L} + \mathbf{Z}^t_{\overrightarrow{e}_L}\right)\Bigg] \tag{33}$$

$$\boldsymbol{\kappa}^{s+1,t+1}_{\overleftarrow{e}_L} = \lim_{n\to\infty} \frac{1}{N}\mathbb{E}\Bigg[f^s_{\overleftarrow{e}_L}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_L}}, \mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_L}} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_L}} \boldsymbol{\nu}^s_{\overrightarrow{e}_L} + \mathbf{Z}^s_{\overrightarrow{e}_L}\right)^\top$$

$$f^t_{\overleftarrow{e}_L}\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_L}}, \mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_L}} \rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_L}} \boldsymbol{\nu}^t_{\overrightarrow{e}_L} + \mathbf{Z}^t_{\overrightarrow{e}_L}\right)\Bigg] \tag{34}$$

where, for any $1 \leqslant l \leqslant L$, the symbol $\partial \mathbf{z}_{\mathbf{w}_{\overrightarrow{e}},i}, \phi_{\overrightarrow{e}}$ denotes the partial derivative w.r.t. the argument of $\phi_{\overrightarrow{e}}$, $(\mathbf{Z}^1_{\overrightarrow{e}}, \dots, \mathbf{Z}^t_{\overrightarrow{e}})$ is a centered Gaussian random vector with covariance $(\boldsymbol{\kappa}^{r,s}_{\overrightarrow{e}})_{r,s\leq t} \otimes \mathbf{I}_{n_w}$ (and similarly for left-oriented edges), and $z_{\mathbf{w}_{\overrightarrow{e}}}$ is distributed according to $\mathbf{N}(0, \boldsymbol{\rho}_{\mathbf{w}_{\overrightarrow{e}}})$.

Theorem A.2. Assume (A1)-(A7). Define, as above, independently for each $\overrightarrow{e}_l$, $\mathbf{Z}^0_{\overrightarrow{e}_l} = \mathbf{x}^0_{\overrightarrow{e}_l}$ and $(\mathbf{Z}^1_{\overrightarrow{e}_l}, \dots, \mathbf{Z}^t_{\overrightarrow{e}_l})$ a centered Gaussian random vector of covariance $(\boldsymbol{\kappa}^{r,s}_{\overrightarrow{e}_l})_{r,s\leq t} \otimes \mathbf{I}_{n_{l-1}}$. Then for any sequence of uniformly (in n) pseudo-Lipschitz function $\Phi : (\mathbb{R}^{n_{l-1}\times(t+1)q})^2 \to \mathbb{R}$, for any $1 \leqslant l \leqslant L$

$$\Phi\left(\left(\mathbf{x}^s_{\overrightarrow{e}_l}\right)_{0\leq s\leq t}, \left(\mathbf{x}^s_{\overleftarrow{e}_{l-1}}\right)_{0\leq s\leq t}\right) \overset{\mathrm{P}}{\simeq}$$
$$\mathbb{E}\left[\Phi\left(\left(\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_l}}\rho^{-1}_{\mathbf{w}_{\overrightarrow{e}_l}}\boldsymbol{\nu}^s_{\overrightarrow{e}_l} + \mathbf{Z}^s_{\overrightarrow{e}_{l-1}}\right)_{0\leq s\leq t}, \left(\mathbf{w}_{\overrightarrow{e}_{l-1}}\hat{\boldsymbol{\nu}}^s_{\overleftarrow{e}_{l-1}} + \mathbf{Z}^s_{\overleftarrow{e}_{l-1}}\right)_{0\leq s\leq t}\right)\right]$$

In summary, at each time step, the variables associated with right oriented edges $\mathbf{x}_{\overrightarrow{e}_l}$ asymptotically behave as the sum of the ground truth $\mathbf{w}_{\overrightarrow{e}_l}$ reweighted by a $q\times q$ matrix coefficient $\hat{\boldsymbol{\nu}}_{\overleftarrow{e}_l}$ and a $n_{l-1}\times q$ random matrix with i.i.d. rows $\mathbf{Z}_{\overrightarrow{e}_l}$ with $q\times q$ covariance $\boldsymbol{\kappa}_{\overleftarrow{e}_l}$ determined by the function associated to the corresponding left-oriented arrow $f^t_{\overleftarrow{e}_l}$. Similarly, the variables associated with left oriented edges $\mathbf{x}_{\overleftarrow{e}_l}$ asymptotically behave as the sum of the linear response to the ground truth $\mathbf{z}_{\mathbf{w}_{\overrightarrow{e}_l}}$ (asymptotic equivalent of $\mathbf{A}_{\overrightarrow{e}_l}\mathbf{w}_{\overrightarrow{e}_l}$) reweighted by a $q\times q$ matrix coefficient $\boldsymbol{\nu}_{\overleftarrow{e}_l}$ and a $n_l\times q$ random matrix with i.i.d. rows $\mathbf{Z}_{\overleftarrow{e}_l}$ with $q\times q$ covariance $\boldsymbol{\kappa}_{\overrightarrow{e}_l}$ determined by the function associated to the corresponding right-oriented arrow $f^t_{\overrightarrow{e}_l}$.

Proof. This result is a special case of Lemma 2 from Gerbelot and Berthier [2021], with a perturbation where only the left-oriented edges involve an additional dependence on $\mathbf{A}_{\overrightarrow{e}}\mathbf{w}_{\overrightarrow{e}}$. The required conditions are the same as in Gerbelot and Berthier [2021], barring the subgaussian assumption (A3) which ensures the scaled norm of the $\mathbf{x}^0_{\overrightarrow{e}}, \mathbf{w}_{\overrightarrow{e}}$ are finite with high-probability as $n\to\infty$. □

A.3 State evolution for multilayer AMP iterations with random convolutional matrices

The following lemma proves the state evolution equations for a multilayer AMP iteration where the dense Gaussian matrices are replaced with random convolutional ones (MCC from Def.3.2) with variance $\frac{1}{N}$, with a vector valued variables, i.e. q=1, and separables non-linearities. We choose the variance as $\frac{1}{N}$ to follow the notations of Gerbelot and Berthier [2021] for more convenience, recovering the variances of iteration Eq.(4) is a straightforward rescaling as done in Berthier et al. [2020] and will be discussed in the next section. Assume $q = 1$ and that, for any $t\in\mathbb{N}$ and $1\leqslant l\leqslant L$, the functions $f^t_{\overrightarrow{e}_l}, f^t_{\overleftarrow{e}_l}$ are separable in all their arguments, i.e there exists scalar valued, pseudo-Lipschitz functions $\sigma^t_{\overrightarrow{e}_l}:\mathbb{R}^2\to\mathbb{R}, \sigma^t_{\overleftarrow{e}_l}:\mathbb{R}^3\to\mathbb{R}$ (where $\sigma^t_{\overrightarrow{e}_1}:\mathbb{R}\to\mathbb{R}, \sigma^t_{\overleftarrow{e}_L}:\mathbb{R}^2\to\mathbb{R}$) such that:

for $l=1$, for any $1\leqslant i\leqslant n_0$:
$$f^t_{\overleftarrow{e}_1}(\mathbf{x}^t_{\overleftarrow{e}_1})_i = \sigma^t_{\overleftarrow{e}_1}(x^t_{\overleftarrow{e}_1,i})$$
for any $1\leqslant l\leqslant L-1$, for any $1\leqslant i\leqslant n_l$:
$$f^t_{\overleftarrow{e}_l}\left(\mathbf{A}_{\overrightarrow{e}_l}\mathbf{w}_{\overrightarrow{e}_l}, \mathbf{x}^t_{\overrightarrow{e}_l}, \mathbf{x}^t_{\overleftarrow{e}_{l+1}}\right)_i = \sigma^t_{\overleftarrow{e}_l}\left((\mathbf{A}_{\overrightarrow{e}_l}\mathbf{w}_{\overrightarrow{e}_l})_i, x^t_{\overrightarrow{e}_l,i}, x^t_{\overleftarrow{e}_{l+1},i}\right)$$
for any $2\leqslant l\leqslant L$, $1\leqslant i\leqslant n_{l-1}$:
$$f^t_{\overrightarrow{e}_l}\left(\mathbf{x}^t_{\overrightarrow{e}_{l-1}}, \mathbf{x}^t_{\overleftarrow{e}_l}\right)_i = \sigma^t_{\overrightarrow{e}_l}\left(x^t_{\overrightarrow{e}_{l-1},i}, x^t_{\overleftarrow{e}_l,i}\right)$$
for l=L, any $1\leqslant i\leqslant n_L$:
$$f^t_{\overleftarrow{e}_L}(\mathbf{A}_{\overrightarrow{e}_L}\mathbf{w}_{\overleftarrow{e}_L}, \mathbf{x}^t_{\overrightarrow{e}_L})_i = \sigma^t_{\overleftarrow{e}_L}((\mathbf{A}_{\overrightarrow{e}_L}\mathbf{w}_{\overleftarrow{e}_L})_i, x^t_{\overrightarrow{e}_L,i})$$

Define the following scalar SE equations

- for $l = 1$:

$$\nu^0_{\overrightarrow{e}_1} = \delta_0 \mathbb{E}\left[w_{\overrightarrow{e}_1}\sigma^0_{\overrightarrow{e}_1}(x^0_{\overleftarrow{e}_1})\right],\ \kappa^{1,1}_{\overrightarrow{e}_1} = \delta_0 \mathbb{E}\left[\sigma^0_{\overrightarrow{e}_1}(x^0_{\overleftarrow{e}_1})\sigma^0_{\overrightarrow{e}_1}(x^0_{\overleftarrow{e}_1})\right] \tag{35}$$

$$\nu^{t+1}_{\overrightarrow{e}_1} = \delta_0 \mathbb{E}\left[w_{\overrightarrow{e}_1}\sigma^t_{\overrightarrow{e}_1}\left(w_{\overrightarrow{e}_1}\hat{\nu}^t_{\overleftarrow{e}_1} + Z^t_{\overleftarrow{e}_1}\right)\right] \tag{36}$$

$$\kappa^{s+1,t+1}_{\overrightarrow{e}_1} = \kappa^{t+1,s+1}_{\overrightarrow{e}_1} = \delta_0 \mathbb{E}\Big[\left(\sigma^s_{\overrightarrow{e}_1}\left(w_{\overrightarrow{e}_1}\hat{\nu}^s_{\overleftarrow{e}_1} + Z^s_{\overleftarrow{e}_1}\right) - w_{\overrightarrow{e}_1}\rho^{-1}_{w_{\overrightarrow{e}_1}}\nu^{s+1}_{\overrightarrow{e}_1}\right)$$

$$\left(\sigma^t_{\overrightarrow{e}_1}\left(w_{\overrightarrow{e}_1}\hat{\nu}^t_{\overleftarrow{e}_1} + Z^t_{\overleftarrow{e}_1}\right) - w_{\overrightarrow{e}_1}\rho^{-1}_{w_{\overrightarrow{e}_1}}\nu^{t+1}_{\overrightarrow{e}_1}\right)\Big] \tag{37}$$

$$\hat{\nu}^0_{\overleftarrow{e}_1}, \kappa^{1,1}_{\overleftarrow{e}_1} = \delta_1 \mathbb{E}\left[\sigma^0_{\overleftarrow{e}_1}\left(z_{w_{\overrightarrow{e}_1}}, x^0_{\overrightarrow{e}_1}, x^0_{\overleftarrow{e}_2}\right)\sigma^0_{\overleftarrow{e}_1}\left(z_{w_{\overrightarrow{e}_1}}, x^0_{\overrightarrow{e}_1}, x^0_{\overleftarrow{e}_2}\right)\right] \tag{38}$$

$$\hat{\nu}^{t+1}_{\overleftarrow{e}_1} = \delta_1 \mathbb{E}\left[\frac{\partial \sigma^t_{\overleftarrow{e}_1, i}}{\partial z_{w_{\overleftarrow{e}_1}, i}, \phi_{\overleftarrow{e}_1}}\left(z_{w_{\overrightarrow{e}_1}}, z_{w_{\overrightarrow{e}_1}}\rho^{-1}_{w_{\overrightarrow{e}_1}}\nu^t_{\overrightarrow{e}_1} + Z^t_{\overrightarrow{e}_1}, w_{\overrightarrow{e}_2}\hat{\nu}^t_{\overleftarrow{e}_2} + Z^t_{\overleftarrow{e}_2}\right)\right] \tag{39}$$

$$\kappa^{s+1,t+1}_{\overleftarrow{e}_1} = \delta_1 \mathbb{E}\Big[\sigma^s_{\overleftarrow{e}_1}\left(z_{w_{\overrightarrow{e}_1}}, z_{w_{\overrightarrow{e}_1}}\rho^{-1}_{w_{\overrightarrow{e}_1}}\nu^s_{\overrightarrow{e}_1} + Z^s_{\overrightarrow{e}_1}, w_{\overrightarrow{e}_2}\hat{\nu}^s_{\overleftarrow{e}_2} + Z^s_{\overleftarrow{e}_2}\right)$$

$$\sigma^t_{\overleftarrow{e}_1}\left(z_{w_{\overrightarrow{e}_1}}, z_{w_{\overrightarrow{e}_1}}\rho^{-1}_{w_{\overrightarrow{e}_1}}\nu^t_{\overrightarrow{e}_1} + Z^t_{\overrightarrow{e}_1}, w_{\overrightarrow{e}_2}\hat{\nu}^t_{\overleftarrow{e}_2} + Z^t_{\overleftarrow{e}_2}\right)\Big] \tag{40}$$

- for any $2 \leqslant l \leqslant L-1$

$$\nu^0_{\overrightarrow{e}_l} = \delta_{n_{l-1}} \mathbb{E}\left[w_{\overrightarrow{e}_l}\sigma^0_{\overrightarrow{e}_l}(x^0_{\overleftarrow{e}_l})\right],\ \kappa^{1,1}_{\overrightarrow{e}_l} = \delta_{n_{l-1}} \mathbb{E}\left[\sigma^0_{\overrightarrow{e}_l}(x^0_{\overleftarrow{e}_l})\sigma^0_{\overrightarrow{e}_l}(x^0_{\overleftarrow{e}_l})\right] \tag{41}$$

$$\nu^{t+1}_{\overrightarrow{e}_l} = \delta_{n_{l-1}} \mathbb{E}\left[w_{\overrightarrow{e}_l}\sigma^t_{\overrightarrow{e}_l}\left(z_{w_{\overrightarrow{e}_{l-1}}}\rho^{-1}_{w_{\overrightarrow{e}_{l-1}}}\nu^t_{\overrightarrow{e}_{l-1}} + Z^t_{\overrightarrow{e}_{l-1}}, w_{\overrightarrow{e}_l}\hat{\nu}^t_{\overleftarrow{e}_l} + Z^t_{\overleftarrow{e}_l}\right)\right] \tag{42}$$

$$\kappa^{s+1,t+1}_{\overrightarrow{e}_l} = \kappa^{t+1,s+1}_{\overrightarrow{e}_l} = \tag{43}$$

$$\delta_{n_{l-1}} \mathbb{E}\Big[\left(\sigma^s_{\overrightarrow{e}_l}\left(z_{w_{\overrightarrow{e}_{l-1}}}\rho^{-1}_{w_{\overrightarrow{e}_{l-1}}}\nu^s_{\overrightarrow{e}_{l-1}} + Z^s_{\overrightarrow{e}_{l-1}}, w_{\overrightarrow{e}_l}\hat{\nu}^s_{\overleftarrow{e}_l} + Z^s_{\overleftarrow{e}_l}\right) - w_{\overrightarrow{e}_l}\rho^{-1}_{w_{\overrightarrow{e}_l}}\nu^{s+1}_{\overrightarrow{e}_l}\right)$$

$$\left(\sigma^t_{\overrightarrow{e}_l}\left(z_{w_{\overrightarrow{e}_{l-1}}}\rho^{-1}_{w_{\overrightarrow{e}_{l-1}}}\nu^t_{\overrightarrow{e}_{l-1}} + Z^t_{\overrightarrow{e}_{l-1}}, w_{\overrightarrow{e}_l}\hat{\nu}^t_{\overleftarrow{e}_l} + Z^t_{\overleftarrow{e}_l}\right) - w_{\overrightarrow{e}_l}\rho^{-1}_{w_{\overrightarrow{e}_l}}\nu^{t+1}_{\overrightarrow{e}_l}\right)\Big] \tag{44}$$

$$\hat{\nu}^0_{\overleftarrow{e}_l}, \kappa^{1,1}_{\overleftarrow{e}_l} = \delta_{n_l} \mathbb{E}\left[\sigma^0_{\overleftarrow{e}_l}\left(z_{w_{\overrightarrow{e}_l}}, x^0_{\overrightarrow{e}_l}, x^0_{\overleftarrow{e}_{l+1}}\right)\sigma^0_{\overleftarrow{e}_l}\left(z_{w_{\overrightarrow{e}_l}}, x^0_{\overrightarrow{e}_l}, x^0_{\overleftarrow{e}_{l+1}}\right)\right] \tag{45}$$

$$\hat{\nu}^{t+1}_{\overleftarrow{e}_l} = \delta_{n_l} \mathbb{E}\left[\frac{\partial \sigma^t_{\overleftarrow{e}_l, i}}{\partial z_{w_{\overleftarrow{e}_l}, i}, \phi_{\overleftarrow{e}_l}}\left(z_{w_{\overrightarrow{e}_l}}, z_{w_{\overrightarrow{e}_l}}\rho^{-1}_{w_{\overrightarrow{e}_l}}\nu^t_{\overrightarrow{e}_l} + Z^t_{\overrightarrow{e}_l}, w_{\overrightarrow{e}_{l+1}}\hat{\nu}^t_{\overrightarrow{e}_{l+1}}Z^t_{\overleftarrow{e}_{l+1}}\right)\right] \tag{46}$$

$$\kappa^{s+1,t+1}_{\overleftarrow{e}_l} = \delta_{n_l} \mathbb{E}\Big[\sigma^s_{\overleftarrow{e}_l}\left(z_{w_{\overrightarrow{e}_l}}, z_{w_{\overrightarrow{e}_l}}\rho^{-1}_{w_{\overrightarrow{e}_l}}\nu^s_{\overrightarrow{e}_l} + Z^s_{\overrightarrow{e}_l}, w_{\overrightarrow{e}_{l+1}}\hat{\nu}^s_{\overrightarrow{e}_{l+1}}Z^s_{\overleftarrow{e}_{l+1}}\right)$$

$$\sigma^t_{\overleftarrow{e}_l}\left(z_{w_{\overrightarrow{e}_l}}, z_{w_{\overrightarrow{e}_l}}\rho^{-1}_{w_{\overrightarrow{e}_l}}\nu^t_{\overrightarrow{e}_l} + Z^t_{\overrightarrow{e}_l}, w_{\overrightarrow{e}_{l+1}}\hat{\nu}^t_{\overrightarrow{e}_{l+1}}Z^t_{\overleftarrow{e}_{l+1}}\right)\Big] \tag{47}$$

- for l=L

$$\nu^{0}_{\overrightarrow{e}_L} = \delta_{n_{L-1}}\mathbb{E}\Big[w_{\overrightarrow{e}_l}\sigma^{0}_{\overrightarrow{e}_L}(x^{0}_{\overleftarrow{e}_L})\Big],\ \kappa^{1,1}_{\overrightarrow{e}_L} = \delta_{n_{L-1}}\mathbb{E}\Big[\sigma^{0}_{\overrightarrow{e}_L}(x^{0}_{\overleftarrow{e}_L})\sigma^{0}_{\overrightarrow{e}_L}(x^{0}_{\overleftarrow{e}_L})\Big] \tag{48}$$

$$\nu^{t+1}_{\overrightarrow{e}_L} = \delta_{n_{L-1}}\mathbb{E}\left[w_{\overrightarrow{e}_L}\sigma^{t}_{\overrightarrow{e}_L}\left(z_{w_{\overrightarrow{e}_{L-1}}}\rho^{-1}_{w_{\overrightarrow{e}_{L-1}}}\nu^{t}_{\overrightarrow{e}_{L-1}} + Z^{t}_{\overrightarrow{e}_{L-1}}, w_{\overrightarrow{e}_L}\hat{\nu}^{t}_{\overleftarrow{e}_L} + Z^{t}_{\overleftarrow{e}_L}\right)\right] \tag{49}$$

$$\kappa^{s+1,t+1}_{\overrightarrow{e}_L} = \kappa^{t+1,s+1}_{\overrightarrow{e}_L} = \tag{50}$$

$$\delta_{n_{L-1}}\mathbb{E}\Bigg[\left(\sigma^{s}_{\overrightarrow{e}_L}\left(z_{w_{\overrightarrow{e}_{L-1}}}\rho^{-1}_{w_{\overrightarrow{e}_{L-1}}}\nu^{s}_{\overrightarrow{e}_{L-1}} + Z^{s}_{\overrightarrow{e}_{L-1}}, w_{\overrightarrow{e}_L}\hat{\nu}^{s}_{\overleftarrow{e}_L} + Z^{s}_{\overleftarrow{e}_L}\right) - w_{\overrightarrow{e}_L}\rho^{-1}_{w_{\overrightarrow{e}_L}}\nu^{s+1}_{\overrightarrow{e}_L}\right)$$
$$\left(\sigma^{t}_{\overrightarrow{e}_L}\left(z_{w_{\overrightarrow{e}_{L-1}}}\rho^{-1}_{w_{\overrightarrow{e}_{L-1}}}\nu^{t}_{\overrightarrow{e}_{L-1}} + Z^{t}_{\overrightarrow{e}_{L-1}}, w_{\overrightarrow{e}_L}\hat{\nu}^{t}_{\overleftarrow{e}_L} + Z^{t}_{\overleftarrow{e}_L}\right) - w_{\overrightarrow{e}_L}\rho^{-1}_{w_{\overrightarrow{e}_L}}\nu^{t+1}_{\overrightarrow{e}_L}\right)\Bigg] \tag{51}$$

$$\hat{\nu}^{0}_{\overleftarrow{e}_L}, \kappa^{1,1}_{\overleftarrow{e}_L} = \delta_{n_L}\mathbb{E}\Big[\sigma^{0}_{\overleftarrow{e}_L}\left(z_{w_{\overrightarrow{e}_L}}, x^{0}_{\overrightarrow{e}_L}\right)\sigma^{0}_{\overleftarrow{e}_L}\left(z_{w_{\overrightarrow{e}_L}}\right)\Big] \tag{52}$$

$$\hat{\nu}^{t+1}_{\overleftarrow{e}_L} = \delta_{n_L}\mathbb{E}\Bigg[\frac{\partial\sigma^{t}_{\overleftarrow{e}_L,i}}{\partial z_{w_{\overleftarrow{e}_L},i}, \phi_{\overleftarrow{e}_L}}\left(z_{w_{\overrightarrow{e}_L}}, z_{w_{\overrightarrow{e}_L}}\rho^{-1}_{w_{\overrightarrow{e}_L}}\nu^{t}_{\overrightarrow{e}_L} + Z^{t}_{\overrightarrow{e}_L}\right)\Bigg] \tag{53}$$

$$\kappa^{s+1,t+1}_{\overleftarrow{e}_L} = \delta_{n_L}\mathbb{E}\Big[\sigma^{s}_{\overleftarrow{e}_L}\left(z_{w_{\overrightarrow{e}_L}}, z_{w_{\overrightarrow{e}_L}}\rho^{-1}_{w_{\overrightarrow{e}_L}}\nu^{s}_{\overrightarrow{e}_L} + Z^{s}_{\overrightarrow{e}_L}\right)$$
$$\sigma^{t}_{\overleftarrow{e}_L}\left(z_{w_{\overrightarrow{e}_L}}, z_{w_{\overrightarrow{e}_L}}\rho^{-1}_{w_{\overrightarrow{e}_L}}\nu^{t}_{\overrightarrow{e}_L} + Z^{t}_{\overrightarrow{e}_L}\right)\Big] \tag{54}$$

Lemma A.3. Under the assumptions of section A.3, define, as above, independently for each $\overrightarrow{e}_l$, $Z^{0}_{\overrightarrow{e}_l} = x^{0}_{\overrightarrow{e}}$ and $(Z^{1}_{\overrightarrow{e}_l}, \dots, Z^{t}_{\overrightarrow{e}_l})$ a centered Gaussian random vector of covariance $(\boldsymbol{\kappa}^{r,s}_{\overrightarrow{e}_l})_{r,s\leq t}$ (and similarly for left-oriented edges). Then for any $1 \leqslant l \leqslant L$, for any sequence of uniformly (in n) pseudo-Lipschitz function $\Phi_l : (\mathbb{R}^{n_{l-1}\times(t+1)})^2 \to \mathbb{R}$

$$\Phi\left(\left(\mathbf{x}^{s}_{\overrightarrow{e}_l}\right)_{0\leq s\leq t}, \left(\mathbf{x}^{s}_{\overleftarrow{e}_l}\right)_{0\leq s\leq t, \overleftarrow{e}_{l-1}\in\overleftarrow{E}}\right) \overset{\mathrm{P}}{\simeq}$$
$$\mathbb{E}\left[\Phi\left(\left(z_{w_{\overrightarrow{e}_l}}\rho^{-1}_{w_{\overrightarrow{e}_l}}\nu^{s}_{\overrightarrow{e}_l} + Z^{s}_{\overrightarrow{e}_l}\right)_{0\leq s\leq t, \overleftarrow{e}_l\in\overleftarrow{E}}, \left(w_{\overrightarrow{e}_{l-1}}\hat{\nu}^{s}_{\overleftarrow{e}_{l-1}} + Z^{s}_{\overleftarrow{e}_{l-1}}\right)_{0\leq s\leq t}\right)\right]$$

Proof. Consider the following iteration, corresponding to the algorithm presented in the previous section Eq.(14) with $q = 1$ indexed on the same graph as above, but where the matrices $\mathbf{A}_{\overrightarrow{e}_l}$ are replaced with random convolutional ones, denoted $\hat{\mathbf{A}}_{\overrightarrow{e}_l}$ such that

$$\forall\ \overrightarrow{e} \in \overrightarrow{E}\ \hat{\mathbf{A}}_{\overrightarrow{e}_l} \sim \mathcal{M}(D_{\overrightarrow{e}_l}, P_{\overrightarrow{e}_l}, k_{\overrightarrow{e}_l}, q_{\overrightarrow{e}_l}) \tag{55}$$

where $\mathbf{A}_{\overrightarrow{e}_l} \in \mathbb{R}^{D_{\overrightarrow{e}_l}q_{\overrightarrow{e}_l}\times P_{\overrightarrow{e}_l}q_{\overrightarrow{e}_l}}$, and we remind that we chose variances of $1/N$. Since we assume that $q = 1$, thus the Onsager terms are scalars, which we denote with lowercase letters $b^{t}_{\overrightarrow{e}}$. The corresponding iteration then reads:

$$
\begin{aligned}
\mathbf{x}^{t+1}_{\overrightarrow{e_1}} &= \hat{\mathbf{A}}_{\overrightarrow{e_1}} \mathbf{m}^{t}_{\overrightarrow{e_1}} - b^{t}_{\overrightarrow{e_1}} \mathbf{m}^{t-1}_{\overleftarrow{e_1}} \,, \\
\mathbf{m}^{t}_{\overrightarrow{e_1}} &= f^{t}_{\overset{\rightleftarrows}{e}_1} \left(\mathbf{x}^{t}_{\overleftarrow{e_1}} \right) \,, \\
\mathbf{x}^{t+1}_{\overleftarrow{e_1}} &= \hat{\mathbf{A}}^{\top}_{\overrightarrow{e_1}} \mathbf{m}^{t}_{\overleftarrow{e_1}} - b^{t}_{\overleftarrow{e_1}} \mathbf{m}^{t-1}_{\overrightarrow{e_1}} \,, \\
\mathbf{m}^{t}_{\overleftarrow{e_1}} &= f^{t}_{\overleftarrow{e_1}} \left(\hat{\mathbf{A}}_{\overset{\rightleftarrows}{e}_1} \mathbf{w}_{\overset{\rightleftarrows}{e}_1}, \mathbf{x}^{t}_{\overrightarrow{e_1}}, \mathbf{x}^{t}_{\overleftarrow{e_2}} \right) \,, \\
\\
\mathbf{x}^{t+1}_{\overrightarrow{e_2}} &= \hat{\mathbf{A}}_{\overrightarrow{e_2}} \mathbf{m}^{t}_{\overrightarrow{e_2}} - b^{t}_{\overrightarrow{e_2}} \mathbf{m}^{t-1}_{\overleftarrow{e_2}} \,, \\
\mathbf{m}^{t}_{\overrightarrow{e_2}} &= f^{t}_{\overset{\rightleftarrows}{e}_2} \left(\mathbf{x}^{t}_{\overrightarrow{e_1}}, \mathbf{x}^{t}_{\overleftarrow{e_2}} \right) \,, \\
\mathbf{x}^{t+1}_{\overleftarrow{e_2}} &= \hat{\mathbf{A}}^{\top}_{\overrightarrow{e_2}} \mathbf{m}^{t}_{\overleftarrow{e_2}} - b^{t}_{\overleftarrow{e_2}} \mathbf{m}^{t-1}_{\overrightarrow{e_2}} \,, \\
\mathbf{m}^{t}_{\overleftarrow{e_2}} &= f^{t}_{\overleftarrow{e_2}} \left(\hat{\mathbf{A}}_{\overset{\rightleftarrows}{e}_2} \mathbf{w}_{\overset{\rightleftarrows}{e}_2}, \mathbf{x}^{t}_{\overrightarrow{e_2}}, \mathbf{x}^{t}_{\overleftarrow{e_3}} \right) \,, \\
\\
&\vdots \\
\\
\mathbf{x}^{t+1}_{\overrightarrow{e_L}} &= \hat{\mathbf{A}}_{\overrightarrow{e_L}} \mathbf{m}^{t}_{\overrightarrow{e_L}} - b^{t}_{\overrightarrow{e_L}} \mathbf{m}^{t-1}_{\overleftarrow{e_L}} \,, \\
\mathbf{m}^{t}_{\overrightarrow{e_L}} &= f^{t}_{\overset{\rightleftarrows}{e}_L} \left(\mathbf{x}^{t}_{\overset{\rightleftarrows}{e}_{L-1}}, \mathbf{x}^{t}_{\overleftarrow{e_L}} \right) \,, \\
\mathbf{x}^{t+1}_{\overleftarrow{e_L}} &= \hat{\mathbf{A}}^{\top}_{\overrightarrow{e_L}} \mathbf{m}^{t}_{\overleftarrow{e_L}} - b^{t}_{\overleftarrow{e_L}} \mathbf{m}^{t-1}_{\overrightarrow{e_L}} \,, \\
\mathbf{m}^{t}_{\overleftarrow{e_L}} &= f^{t}_{\overleftarrow{e_L}} \left(\hat{\mathbf{A}}_{\overset{\rightleftarrows}{e}_L} \mathbf{w}_{\overset{\rightleftarrows}{e}_L}, \mathbf{x}^{t}_{\overrightarrow{e_L}} \right)
\end{aligned}
\tag{56}
$$

Then, according to Lemma 4.3, for any $1 \leqslant l \leqslant L$, there exists a pair of orthogonal matrices $\mathbf{U}_{\overset{\rightleftarrows}{e}_l} \in \mathbb{R}^{D_{\overset{\rightleftarrows}{e}_l} q_{\overset{\rightleftarrows}{e}_l} \times D_{\overset{\rightleftarrows}{e}_l} q_{\overset{\rightleftarrows}{e}_l}}, \mathbf{V}_{\overset{\rightleftarrows}{e}_l} \in \mathbb{R}^{P_{\overset{\rightleftarrows}{e}_l} q_{\overset{\rightleftarrows}{e}_l} \times P_{\overset{\rightleftarrows}{e}_l} q_{\overset{\rightleftarrows}{e}_l}}$ such that $\hat{\mathbf{A}}_{\overset{\rightleftarrows}{e}_l} = \mathbf{U}_{\overset{\rightleftarrows}{e}_l} \tilde{\mathbf{A}}_{\overset{\rightleftarrows}{e}_l} \mathbf{V}^{\top}_{\overset{\rightleftarrows}{e}_l}$ and $\tilde{\mathbf{A}}_{\overset{\rightleftarrows}{e}_l} = \left[\left(\mathcal{P}_{P_{\overset{\rightleftarrows}{e}_l}, q_{\overset{\rightleftarrows}{e}_l}} \right)^{i-1} \mathbf{Q}_{\overset{\rightleftarrows}{e}_l} \right]_{i=1}^{q_{\overset{\rightleftarrows}{e}_l}}$, where $\mathbf{Q}_{\overset{\rightleftarrows}{e}_l} \in \mathbb{R}^{D_{\overset{\rightleftarrows}{e}_l} \times P_{\overset{\rightleftarrows}{e}_l} q_{\overset{\rightleftarrows}{e}_l}}$ is composed of $q_{\overset{\rightleftarrows}{e}_l}$ blocks of size $D_{\overset{\rightleftarrows}{e}_l} \times P_{\overset{\rightleftarrows}{e}_l}$, denoted $\mathbf{Q}^{j}_{\overset{\rightleftarrows}{e}_l}$, verifying

- for any $1 \leqslant j \leqslant k_{\overset{\rightleftarrows}{e}}$, $\mathbf{Q}^{j}_{\overset{\rightleftarrows}{e}}$ has i.i.d. $\mathcal{N}(0, \frac{1}{N})$ elements
- for any $k_{\overset{\rightleftarrows}{e}} < j \leqslant q_{\overset{\rightleftarrows}{e}}$, all elements of $\mathbf{Q}^{j}_{\overset{\rightleftarrows}{e}}$ are zero.

In the preceding definition of $\tilde{\mathbf{A}}_{\overset{\rightleftarrows}{e}_l}$, $\mathbf{Q}_{\overset{\rightleftarrows}{e}_l}$ is understood as a vector of size $\mathbb{R}^{P_{\overset{\rightleftarrows}{e}} q_{\overset{\rightleftarrows}{e}}}$ with elements in $\mathbb{R}^{D_{\overset{\rightleftarrows}{e}}}$, such that the permutation matrix $\mathcal{P}_{P_{\overset{\rightleftarrows}{e}}, q_{\overset{\rightleftarrows}{e}}}$ shifts blocks of size $D_{\overset{\rightleftarrows}{e}} \times P_{\overset{\rightleftarrows}{e}}$, yielding

$$
\tilde{\mathbf{A}}_{\overset{\rightleftarrows}{e}} = \begin{bmatrix}
\mathbf{Q}^{(1)}_{\overset{\rightleftarrows}{e}_l} & \mathbf{Q}^{(2)}_{\overset{\rightleftarrows}{e}_l} & \dots & \mathbf{Q}^{(k_{\overset{\rightleftarrows}{e}})}_{\overset{\rightleftarrows}{e}_l} & & & & \\
& \mathbf{Q}^{(1)}_{\overset{\rightleftarrows}{e}_l} & \mathbf{Q}^{(2)}_{\overset{\rightleftarrows}{e}_l} & \dots & \mathbf{Q}^{(k_{\overset{\rightleftarrows}{e}})}_{\overset{\rightleftarrows}{e}_l} & & & \\
& & \mathbf{Q}^{(1)}_{\overset{\rightleftarrows}{e}_l} & \mathbf{Q}^{(2)}_{\overset{\rightleftarrows}{e}_l} & \dots & \mathbf{Q}^{(k_{\overset{\rightleftarrows}{e}})}_{\overset{\rightleftarrows}{e}_l} & & \vdots \\
\vdots & \vdots & \ddots & & & & & \\
\mathbf{Q}^{(2)}_{\overset{\rightleftarrows}{e}_l} & \mathbf{Q}^{(3)}_{\overset{\rightleftarrows}{e}_l} & \dots & \mathbf{Q}^{(k_{\overset{\rightleftarrows}{e}})}_{\overset{\rightleftarrows}{e}_l} & & & & \mathbf{Q}^{(1)}_{\overset{\rightleftarrows}{e}_l}
\end{bmatrix}
\tag{57}
$$

The iteration then reads

$$
\begin{aligned}
\mathbf{x}^{t+1}_{\overrightarrow{e_1}} &= \mathbf{U}_{\overrightarrow{e}_1}\tilde{\mathbf{A}}_{\overrightarrow{e}_1}\mathbf{V}^{\top}_{\overrightarrow{e}_1}\mathbf{m}^{t}_{\overrightarrow{e_1}} - b^{t}_{\overrightarrow{e_1}}\mathbf{m}^{t-1}_{\overleftarrow{e_1}}\,, \\
\mathbf{m}^{t}_{\overrightarrow{e_1}} &= f^{t}_{\overrightarrow{e}_1}\left(\mathbf{x}^{t}_{\overleftarrow{e_1}}\right)\,, \\
\mathbf{x}^{t+1}_{\overleftarrow{e_1}} &= \mathbf{V}_{\overrightarrow{e}_1}\tilde{\mathbf{A}}^{\top}_{\overrightarrow{e}_1}\mathbf{U}^{\top}_{\overrightarrow{e}_1}\mathbf{m}^{t}_{\overleftarrow{e_1}} - b^{t}_{\overleftarrow{e_1}}\mathbf{m}^{t-1}_{\overrightarrow{e_1}}\,, \\
\mathbf{m}^{t}_{\overleftarrow{e_1}} &= f^{t}_{\overleftarrow{e_1}}\left(\mathbf{U}_{\overrightarrow{e}_1}\tilde{\mathbf{A}}_{\overrightarrow{e}_1}\mathbf{V}^{\top}_{\overrightarrow{e}_1}\mathbf{w}_{\overrightarrow{e}_1}, \mathbf{x}^{t}_{\overrightarrow{e_1}}, \mathbf{x}^{t}_{\overleftarrow{e_2}}\right)\,, \\
\\
\mathbf{x}^{t+1}_{\overrightarrow{e_2}} &= \mathbf{U}_{\overrightarrow{e}_2}\tilde{\mathbf{A}}_{\overrightarrow{e}_2}\mathbf{V}^{\top}_{\overrightarrow{e}_2}\mathbf{m}^{t}_{\overrightarrow{e_2}} - b^{t}_{\overrightarrow{e_2}}\mathbf{m}^{t-1}_{\overleftarrow{e_2}}\,, \\
\mathbf{m}^{t}_{\overrightarrow{e_2}} &= f^{t}_{\overrightarrow{e}_2}\left(\mathbf{x}^{t}_{\overrightarrow{e_1}}, \mathbf{x}^{t}_{\overleftarrow{e_2}}\right)\,, \\
\mathbf{x}^{t+1}_{\overleftarrow{e_2}} &= \mathbf{V}_{\overrightarrow{e}_2}\tilde{\mathbf{A}}^{\top}_{\overrightarrow{e}}\mathbf{U}^{\top}_{\overrightarrow{e}_2}\mathbf{m}^{t}_{\overleftarrow{e_2}} - b^{t}_{\overleftarrow{e_2}}\mathbf{m}^{t-1}_{\overrightarrow{e_2}}\,, \\
\mathbf{m}^{t}_{\overleftarrow{e_2}} &= f^{t}_{\overleftarrow{e_2}}\left(\mathbf{U}_{\overrightarrow{e}_2}\tilde{\mathbf{A}}_{\overrightarrow{e}_2}\mathbf{V}^{\top}_{\overrightarrow{e}_2}\mathbf{w}_{\overrightarrow{e}_2}, \mathbf{x}^{t}_{\overrightarrow{e_2}}, \mathbf{x}^{t}_{\overleftarrow{e_3}}\right)\,, \\
\\
&\vdots \\
\\
\mathbf{x}^{t+1}_{\overrightarrow{e_L}} &= \mathbf{U}_{\overrightarrow{e}_L}\tilde{\mathbf{A}}_{\overrightarrow{e}_L}\mathbf{V}^{\top}_{\overrightarrow{e}_L}\mathbf{m}^{t}_{\overrightarrow{e_L}} - b^{t}_{\overrightarrow{e_L}}\mathbf{m}^{t-1}_{\overleftarrow{e_L}}\,, \\
\mathbf{m}^{t}_{\overrightarrow{e_L}} &= f^{t}_{\overrightarrow{e}_L}\left(\mathbf{x}^{t}_{\overrightarrow{e}_{L-1}}, \mathbf{x}^{t}_{\overleftarrow{e_L}}\right)\,, \\
\mathbf{x}^{t+1}_{\overleftarrow{e_L}} &= \mathbf{V}_{\overrightarrow{e}_L}\tilde{\mathbf{A}}^{\top}_{\overrightarrow{e}_L}\mathbf{U}^{\top}_{\overrightarrow{e}_L}\mathbf{m}^{t}_{\overleftarrow{e_L}} - b^{t}_{\overleftarrow{e_L}}\mathbf{m}^{t-1}_{\overrightarrow{e_L}}\,, \\
\mathbf{m}^{t}_{\overleftarrow{e_L}} &= f^{t}_{\overleftarrow{e_L}}\left(\mathbf{U}_{\overrightarrow{e}_L}\tilde{\mathbf{A}}_{\overrightarrow{e}_L}\mathbf{V}^{\top}_{\overrightarrow{e}_L}\mathbf{w}_{\overrightarrow{e}_L}, \mathbf{x}^{t}_{\overrightarrow{e_L}}\right)
\end{aligned}
\tag{58}
$$

Since we will not be making any change of variable on the $\mathbf{w}_{\overrightarrow{e}_l}$, we will keep the $\hat{\mathbf{A}}_{\overrightarrow{e}_l}$ notation for the quantities related to the planted model. Define, for any $1 \leqslant l \leqslant L$ and any $t \in \mathbb{N}$:

$$
\begin{aligned}
&\tilde{\mathbf{x}}_{\overrightarrow{e}_l} = \mathbf{U}^{\top}_{\overrightarrow{e}_l}\mathbf{x}_{\overrightarrow{e}_l} \qquad \tilde{\mathbf{x}}_{\overleftarrow{e}_l} = \mathbf{V}^{\top}_{\overrightarrow{e}_l}\mathbf{x}_{\overleftarrow{e}_l} \\
&\tilde{\mathbf{m}}^{t}_{\overrightarrow{e}_l} = \mathbf{V}^{\top}_{\overrightarrow{e}_l}\mathbf{m}^{t}_{\overrightarrow{e}_l} \qquad \tilde{\mathbf{m}}^{t}_{\overleftarrow{e}_l} = \mathbf{U}^{\top}_{\overrightarrow{e}_l}\mathbf{m}^{t}_{\overleftarrow{e}_l} \\
&\tilde{f}^{t}_{\overrightarrow{e}_1}(\tilde{\mathbf{x}}^{t}_{\overleftarrow{e_1}}) = \mathbf{V}^{\top}_{\overrightarrow{e}_1}f^{t}_{\overrightarrow{e}_1}\left(\mathbf{V}_{\overrightarrow{e}_1}\tilde{\mathbf{x}}^{t}_{\overleftarrow{e_1}}\right) \\
&\tilde{f}^{t}_{\overleftarrow{e_1}}\left(\hat{\mathbf{A}}_{\overrightarrow{e}_1}\mathbf{w}_{\overrightarrow{e}_1}, \tilde{\mathbf{x}}^{t}_{\overrightarrow{e_1}}, \tilde{\mathbf{x}}^{t}_{\overleftarrow{e_2}}\right) = \mathbf{U}^{\top}_{\overrightarrow{e}_1}f^{t}_{\overleftarrow{e_1}}\left(\hat{\mathbf{A}}_{\overrightarrow{e}_1}\mathbf{w}_{\overrightarrow{e}_1}, \mathbf{U}_{\overrightarrow{e}_1}\tilde{\mathbf{x}}^{t}_{\overrightarrow{e_1}}, \mathbf{V}_{\overrightarrow{e}_2}\tilde{\mathbf{x}}^{t}_{\overleftarrow{e_2}}\right) \\
&\tilde{f}^{t}_{\overrightarrow{e}_2}\left(\tilde{\mathbf{x}}^{t}_{\overrightarrow{e_1}}, \tilde{\mathbf{x}}^{t}_{\overleftarrow{e_2}}\right) = \mathbf{V}^{\top}_{\overrightarrow{e}_2}f^{t}_{\overrightarrow{e}_2}\left(\mathbf{U}_{\overrightarrow{e}_1}\tilde{\mathbf{x}}^{t}_{\overrightarrow{e_1}}, \mathbf{V}_{\overrightarrow{e}_2}\tilde{\mathbf{x}}^{t}_{\overleftarrow{e_2}}\right) \\
&\tilde{f}^{t}_{\overleftarrow{e_2}}\left(\hat{\mathbf{A}}_{\overrightarrow{e}_2}\mathbf{w}_{\overrightarrow{e}_2}, \tilde{\mathbf{x}}^{t}_{\overrightarrow{e_2}}, \tilde{\mathbf{x}}^{t}_{\overleftarrow{e_3}}\right) = \mathbf{U}^{\top}_{\overrightarrow{e}_2}f^{t}_{\overleftarrow{e_2}}\left(\hat{\mathbf{A}}_{\overrightarrow{e}_2}\mathbf{w}_{\overrightarrow{e}_2}, \mathbf{U}_{\overrightarrow{e}_2}\tilde{\mathbf{x}}^{t}_{\overrightarrow{e_2}}, \mathbf{V}_{\overrightarrow{e}_3}\tilde{\mathbf{x}}^{t}_{\overleftarrow{e_3}}\right) \\
\\
&\quad\vdots \\
\\
&\tilde{f}^{t}_{\overrightarrow{e}_L}\left(\tilde{\mathbf{x}}^{t}_{\overrightarrow{e}_{L-1}}, \tilde{\mathbf{x}}^{t}_{\overleftarrow{e_L}}\right) = \mathbf{V}^{\top}_{\overrightarrow{e}_L}f^{t}_{\overrightarrow{e}_L}\left(\mathbf{U}_{\overrightarrow{e}_{L-1}}\tilde{\mathbf{x}}^{t}_{\overrightarrow{e}_{L-1}}, \mathbf{V}_{\overrightarrow{e}_L}\tilde{\mathbf{x}}^{t}_{\overleftarrow{e_L}}\right) \\
&\tilde{f}^{t}_{\overleftarrow{e_L}}\left(\hat{\mathbf{A}}_{\overrightarrow{e}_L}\mathbf{w}_{\overrightarrow{e}_L}, \tilde{\mathbf{x}}^{t}_{\overrightarrow{e_L}}\right) = \mathbf{U}^{\top}_{\overrightarrow{e}_L}f^{t}_{\overleftarrow{e_L}}\left(\mathbf{U}_{\overrightarrow{e}_L}\tilde{\mathbf{A}}_{\overrightarrow{e}_L}\mathbf{V}_{\overrightarrow{e}_L}\mathbf{w}_{\overrightarrow{e}_L}, \mathbf{U}_{\overrightarrow{e}_L}\tilde{\mathbf{x}}^{t}_{\overrightarrow{e_L}}\right)
\end{aligned}
$$

Using the orthogonality of the permutation matrices $\mathbf{U}_{\overrightarrow{e}}, \mathbf{V}_{\overrightarrow{e}}$, the iteration may be rewritten

$$
\begin{aligned}
\tilde{\mathbf{x}}^{t+1}_{\overrightarrow{e_1}} &= \tilde{\mathbf{A}}_{\overrightarrow{e}_1} \tilde{\mathbf{m}}^t_{\overrightarrow{e_1}} - b^t_{\overrightarrow{e_1}} \tilde{\mathbf{m}}^{t-1}_{\overleftarrow{e_1}} \,, \\
\tilde{\mathbf{m}}^t_{\overrightarrow{e_1}} &= \tilde{f}^t_{\overrightarrow{e}_1}(\tilde{\mathbf{x}}^t_{\overleftarrow{e_1}}) \,, \\
\tilde{\mathbf{x}}^{t+1}_{\overleftarrow{e_1}} &= \tilde{\mathbf{A}}^\top_{\overrightarrow{e}_1} \tilde{\mathbf{m}}^t_{\overleftarrow{e_1}} - b^t_{\overleftarrow{e_1}} \tilde{\mathbf{m}}^{t-1}_{\overrightarrow{e_1}} \,, \\
\tilde{\mathbf{m}}^t_{\overleftarrow{e_1}} &= \tilde{f}^t_{\overleftarrow{e_1}} \left(\hat{\mathbf{A}}_{\overrightarrow{e}_1} \mathbf{w}_{\overrightarrow{e}_1}, \tilde{\mathbf{x}}^t_{\overrightarrow{e_1}}, \tilde{\mathbf{x}}^t_{\overleftarrow{e_2}} \right) \,, \\
\\
\tilde{\mathbf{x}}^{t+1}_{\overrightarrow{e_2}} &= \tilde{\mathbf{A}}_{\overrightarrow{e}_2} \tilde{\mathbf{m}}^t_{\overrightarrow{e_2}} - b^t_{\overrightarrow{e_2}} \tilde{\mathbf{m}}^{t-1}_{\overleftarrow{e_2}} \,, \\
\tilde{\mathbf{m}}^t_{\overrightarrow{e_2}} &= \tilde{f}^t_{\overrightarrow{e}_2} \left(\tilde{\mathbf{x}}^t_{\overrightarrow{e_1}}, \tilde{\mathbf{x}}^t_{\overleftarrow{e_2}} \right) \,, \\
\tilde{\mathbf{x}}^{t+1}_{\overleftarrow{e_2}} &= \tilde{\mathbf{A}}^\top_{\overrightarrow{e}} \tilde{\mathbf{m}}^t_{\overleftarrow{e_2}} - b^t_{\overleftarrow{e_2}} \tilde{\mathbf{m}}^{t-1}_{\overrightarrow{e_2}} \,, \\
\tilde{\mathbf{m}}^t_{\overleftarrow{e_2}} &= \tilde{f}^t_{\overleftarrow{e_2}} \left(\hat{\mathbf{A}}_{\overrightarrow{e}_2} \mathbf{w}_{\overrightarrow{e}_2}, \tilde{\mathbf{x}}^t_{\overrightarrow{e_2}}, \tilde{\mathbf{x}}^t_{\overleftarrow{e_3}} \right) \,, \\
\\
&\vdots \\
\\
\tilde{\mathbf{x}}^{t+1}_{\overrightarrow{e_L}} &= \tilde{\mathbf{A}}_{\overrightarrow{e}_L} \tilde{\mathbf{m}}^t_{\overrightarrow{e_L}} - b^t_{\overrightarrow{e_L}} \tilde{\mathbf{m}}^{t-1}_{\overleftarrow{e_L}} \,, \\
\tilde{\mathbf{m}}^t_{\overrightarrow{e_L}} &= \tilde{f}^t_{\overrightarrow{e}_L} \left(\tilde{\mathbf{x}}^t_{\overrightarrow{e}_{L-1}}, \tilde{\mathbf{x}}^t_{\overleftarrow{e_L}} \right) \,, \\
\tilde{\mathbf{x}}^{t+1}_{\overleftarrow{e_L}} &= \tilde{\mathbf{A}}^\top_{\overrightarrow{e}_L} \tilde{\mathbf{m}}^t_{\overleftarrow{e_L}} - b^t_{\overleftarrow{e_L}} \tilde{\mathbf{m}}^{t-1}_{\overrightarrow{e_L}} \,, \\
\tilde{\mathbf{m}}^t_{\overleftarrow{e_L}} &= \tilde{f}^t_{\overleftarrow{e_L}} \left(\hat{\mathbf{A}}_{\overrightarrow{e}_L} \mathbf{w}_{\overrightarrow{e}_L}, \tilde{\mathbf{x}}^t_{\overrightarrow{e_L}} \right)
\end{aligned}
\tag{59}
$$

Recall, for any $1 \leqslant l \leqslant L$, the dimensions $\tilde{\mathbf{A}}_{\overrightarrow{e}_l} \in \mathbb{R}^{D_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l} \times P_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l}}$ and $\tilde{f}^t_{\overrightarrow{e}_l}(...) \in \mathbb{R}^{P_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l}}$. Consider then

$$
\tilde{f}^t_{\overrightarrow{e}_l}(...) = \begin{bmatrix} \left(\tilde{f}^t_{\overrightarrow{e}_l} \right)^{(1)} (...) \\ \vdots \\ \left(\tilde{f}^t_{\overrightarrow{e}_l} \right)^{(q_{\overrightarrow{e}_l})} (...) \end{bmatrix} \tag{60}
$$

where, for any $1 \leqslant k \leqslant q_{\overrightarrow{e}_l}$, $(\tilde{f}^t_{\overrightarrow{e}_l})^{(k)}(...) \in \mathbb{R}^{P_{\overrightarrow{e}}}$. The product $\tilde{\mathbf{A}}_{\overrightarrow{e}_l} \tilde{f}^t_{\overrightarrow{e}_l}(...) \in \mathbb{R}^{D_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l}}$ then reads, using the circulant structure of $\tilde{\mathbf{A}}_{\overrightarrow{e}_l}$

$$
\begin{aligned}
&\begin{bmatrix}
\mathbf{Q}^{(1)}_{\overrightarrow{e}_l} & \mathbf{Q}^{(2)}_{\overrightarrow{e}_l} & \dots & \mathbf{Q}^{(k_{\overrightarrow{e}})}_{\overrightarrow{e}_l} & & & & \\
& \mathbf{Q}^{(1)}_{\overrightarrow{e}_l} & \mathbf{Q}^{(2)}_{\overrightarrow{e}_l} & \dots & \mathbf{Q}^{(k_{\overrightarrow{e}})}_{\overrightarrow{e}_l} & & & \\
& & \mathbf{Q}^{(1)}_{\overrightarrow{e}_l} & \mathbf{Q}^{(2)}_{\overrightarrow{e}_l} & \dots & \mathbf{Q}^{(k_{\overrightarrow{e}})}_{\overrightarrow{e}_l} & & \vdots \\
\vdots & \vdots & \ddots & & & & & \\
\mathbf{Q}^{(2)}_{\overrightarrow{e}_l} & \mathbf{Q}^{(3)}_{\overrightarrow{e}_l} & \dots & \mathbf{Q}^{(k_{\overrightarrow{e}})}_{\overrightarrow{e}_l} & & & & \mathbf{Q}^{(1)}_{\overrightarrow{e}_l}
\end{bmatrix}
\begin{bmatrix} \left(\tilde{f}^t_{\overrightarrow{e}_l} \right)^{(1)} (...) \\ \vdots \\ \left(\tilde{f}^t_{\overrightarrow{e}_l} \right)^{(q_{\overrightarrow{e}_l})} (...) \end{bmatrix} && (61) \\
&= \left[\left(\left(\mathcal{P}_{P_{\overrightarrow{e}_l}, q_{\overrightarrow{e}_l}} \right)^{i-1} \mathbf{Q}_{\overrightarrow{e}_l} \right) \tilde{f}^t_{\overrightarrow{e}_l}(...) \right]_{i=1}^{q_{\overrightarrow{e}_l}} && (62) \\
&= \left[\sum_{j=1}^{k_{\overrightarrow{e}_l}} \mathbf{Q}^{(j)}_{\overrightarrow{e}_l} (\tilde{f}^t_{\overrightarrow{e}_l})^{(\lfloor j+n-2 \rfloor_{q_{\overrightarrow{e}_l}} + 1)} (...) \right]_{n=1}^{q_{\overrightarrow{e}_l}} && (63)
\end{aligned}
$$

where the notation $\lfloor . \rfloor_{q_{\overrightarrow{e}_l}}$ denotes the modulo $q_{\overrightarrow{e}_l}$, i.e. the remainder of the euclidian division by $q_{\overrightarrow{e}_l}$. Now define

$$\tilde{F}^t_{\overrightarrow{e}_l}(...) = \begin{bmatrix} \left[\left(\mathcal{P}_{P_{\overrightarrow{e}_l}, q_{\overrightarrow{e}_l}}\right)^{1-i} \left[(\tilde{f}^t_{\overrightarrow{e}_l})^{(1)} \dots (\tilde{f}^t_{\overrightarrow{e}_l})^{(q_{\overrightarrow{e}_l})}\right]\right]_{i=1}^{k_{\overrightarrow{e}_l}} \in \mathbb{R}^{P_{\overrightarrow{e}_l} k_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}} \\ \left[0_{P_{\overrightarrow{e}_l}} \dots 0_{P_{\overrightarrow{e}_l}}\right]_{j=1}^{q_{\overrightarrow{e}_l} - k_{\overrightarrow{e}_l}} \end{bmatrix} \in \mathbb{R}^{P_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}} \tag{64}$$

and the matrix $\tilde{\mathbf{Q}}_{\overrightarrow{e}_l} \in \mathbb{R}^{D_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l} \times P_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l}}$ is a dense Gaussian matrix with i.i.d. elements. Then

$$\tilde{\mathbf{Q}}_{\overrightarrow{e}_l} \tilde{F}^t_{\overrightarrow{e}_l}(...) = \begin{bmatrix} \sum_{j=1}^{k_{\overrightarrow{e}_l}} \mathbf{Q}^{(j)}_{\overrightarrow{e}_l} (\tilde{f}^t_{\overrightarrow{e}_l})^{\lfloor j-1 \rfloor_{q_{\overrightarrow{e}_l}} + 1}(...) & \dots & \sum_{j=1}^{k_{\overrightarrow{e}_l}} (\mathbf{Q}^{(j)}_{\overrightarrow{e}_l}) (\tilde{f}^t_{\overrightarrow{e}_l})^{\lfloor j + q_{\overrightarrow{e}_l} - 2 \rfloor_{q_{\overrightarrow{e}_l}} + 1}(...) \\ & \dots & \\ & \dots & \end{bmatrix}$$
$$\in \mathbb{R}^{D_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}}$$

where each $\dots$ is an identical copy of the first $D_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}$ block, for a total of $k_{\overrightarrow{e}_l}$ blocks. This means the $D_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l}$ output of the product $\tilde{\mathbf{A}}_{\overrightarrow{e}_l} f^t_{\overrightarrow{e}_l}(...)$ may be rewritten as a $D_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}$ matrix (copied $k_{\overrightarrow{e}_l}$ times) resulting from the product of a dense Gaussian matrix with i.i.d. elements and a matrix valued function $\tilde{F}^t_{\overrightarrow{e}_l}$ which verifies the same regularity conditions as $f^t_{\overrightarrow{e}_l}$. Note that, owing to the separability assumption, we may use any permutation of the $(\tilde{f}^t_{\overrightarrow{e}_l})^{(i)}, 1 \leqslant i \leqslant q_{\overrightarrow{e}_l}$ and will thus drop the permutations to write

$$\tilde{F}^t_{\overrightarrow{e}_l}(...) = \begin{bmatrix} \left[(\tilde{f}^t_{\overrightarrow{e}_l})^{(1)} \dots (\tilde{f}^t_{\overrightarrow{e}_l})^{(q_{\overrightarrow{e}_l})}\right]_{i=1}^{k_{\overrightarrow{e}_l}} \in \mathbb{R}^{P_{\overrightarrow{e}_l} k_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}} \\ \left[0_{P_{\overrightarrow{e}_l}} \dots 0_{P_{\overrightarrow{e}_l}}\right]_{j=1}^{q_{\overrightarrow{e}_l} - k_{\overrightarrow{e}_l}} \end{bmatrix} \in \mathbb{R}^{P_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}} \tag{65}$$

Similarly, for products of the form $\left(\tilde{\mathbf{A}}_{\overrightarrow{e}_l}\right)^\top \tilde{f}^t_{\overleftarrow{e}_l}(...) \in \mathbb{R}^{P_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l}}$, we may write:

$$\begin{bmatrix} \mathbf{Q}^{(1)}_{\overrightarrow{e}_l} & \mathbf{Q}^{(2)}_{\overrightarrow{e}_l} & \dots & \mathbf{Q}^{(k_{\overrightarrow{e}_l})}_{\overrightarrow{e}_l} & & & & \\ & \mathbf{Q}^{(1)}_{\overrightarrow{e}_l} & \mathbf{Q}^{(2)}_{\overrightarrow{e}_l} & \dots & \mathbf{Q}^{(k_{\overrightarrow{e}_l})}_{\overrightarrow{e}_l} & & & \\ & & \mathbf{Q}^{(1)}_{\overrightarrow{e}_l} & \mathbf{Q}^{(2)}_{\overrightarrow{e}_l} & \dots & \mathbf{Q}^{(k_{\overrightarrow{e}_l})}_{\overrightarrow{e}_l} & & \vdots \\ \vdots & \vdots & \ddots & & & & & \\ \mathbf{Q}^{(2)}_{\overrightarrow{e}_l} & \mathbf{Q}^{(3)}_{\overrightarrow{e}_l} & \dots & \mathbf{Q}^{(k_{\overrightarrow{e}_l})}_{\overrightarrow{e}_l} & & & & \mathbf{Q}^{(1)}_{\overrightarrow{e}_l} \end{bmatrix}^\top \begin{bmatrix} \left(\tilde{f}^t_{\overleftarrow{e}_l}\right)^{(1)}(...) \\ \vdots \\ \left(\tilde{f}^t_{\overleftarrow{e}_l}\right)^{(q_{\overrightarrow{e}_l})}(...) \end{bmatrix} \tag{66}$$

$$= \left[\left(\left(\mathcal{P}_{P_{\overrightarrow{e}_l}, q_{\overrightarrow{e}_l}}\right)^{i-1} \left[(\mathbf{Q}^{(1)}_{\overrightarrow{e}_l})^\top (0 \dots 0) (\mathbf{Q}^{(k_{\overrightarrow{e}_l})}_{\overrightarrow{e}_l})^\top \dots (\mathbf{Q}^{(2)}_{\overrightarrow{e}_l})^\top\right]\right) \tilde{f}^t_{\overleftarrow{e}_l}(...)\right]_{i=1}^{q_{\overrightarrow{e}_l}} \tag{67}$$

Then, using once again the separability assumption, we may define:

$$\tilde{F}^t_{\overleftarrow{e}_l}(...) = \begin{bmatrix} \left[(\tilde{f}^t_{\overleftarrow{e}_l})^{(1)} \dots (\tilde{f}^t_{\overleftarrow{e}_l})^{(q_{\overleftarrow{e}_l})}\right]_{i=1}^{k_{\overrightarrow{e}_l}} \in \mathbb{R}^{D_{\overrightarrow{e}_l} k_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}} \\ \left[0_{D_{\overrightarrow{e}_l}} \dots 0_{D_{\overrightarrow{e}_l}}\right] \end{bmatrix} \in \mathbb{R}^{D_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}} \tag{68}$$

such that the term $\tilde{\mathbf{Q}}^\top_{\overrightarrow{e}_l} \tilde{F}^t_{\overleftarrow{e}_l}(...)$ also contains $k_{\overrightarrow{e}_l}$ copies of a $P_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}$ block containing the $q_{\overrightarrow{e}_l}$ blocks of size $P_{\overrightarrow{e}_l}$ of the original $P_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l}$ vector $\tilde{\mathbf{A}}^\top_{\overrightarrow{e}_l} \tilde{f}_{\overleftarrow{e}_l}(...)$. The iterates of the sequences defined by Eq.(59) may then be rewritten as a subset of the rows of the following matrix valued iteration, i.e.:

$$
\begin{aligned}
\tilde{\mathbf{X}}^{t+1}_{\overrightarrow{e_1}} &= \tilde{\mathbf{Q}}_{\overrightarrow{e}_1}\tilde{\mathbf{M}}^{t}_{\overrightarrow{e_1}} - b^{t}_{\overrightarrow{e_1}}\tilde{\mathbf{M}}^{t-1}_{\overleftarrow{e_1}}\,,\\
\tilde{\mathbf{M}}^{t}_{\overrightarrow{e_1}} &= \tilde{F}^{t}_{\overrightarrow{e}_1}(\tilde{\mathbf{X}}^{t}_{\overleftarrow{e_1}})\,,\\
\tilde{\mathbf{X}}^{t+1}_{\overleftarrow{e_1}} &= \tilde{\mathbf{Q}}^{\top}_{\overrightarrow{e}_1}\tilde{\mathbf{M}}^{t}_{\overleftarrow{e_1}} - b^{t}_{\overleftarrow{e_1}}\tilde{\mathbf{M}}^{t-1}_{\overrightarrow{e_1}}\,,\\
\tilde{\mathbf{M}}^{t}_{\overleftarrow{e_1}} &= \tilde{F}^{t}_{\overleftarrow{e_1}}\left(\tilde{\mathbf{Q}}_{\overrightarrow{e}_1}\mathbf{W}_{\overrightarrow{e}_1}, \tilde{\mathbf{X}}^{t}_{\overrightarrow{e_1}}, \tilde{\mathbf{X}}^{t}_{\overleftarrow{e_2}}\right)\,,\\[1em]
\tilde{\mathbf{X}}^{t+1}_{\overrightarrow{e_2}} &= \tilde{\mathbf{Q}}_{\overrightarrow{e}_2}\tilde{\mathbf{M}}^{t}_{\overrightarrow{e_2}} - b^{t}_{\overrightarrow{e_2}}\tilde{\mathbf{M}}^{t-1}_{\overleftarrow{e_2}}\,,\\
\tilde{\mathbf{M}}^{t}_{\overrightarrow{e_2}} &= \tilde{F}^{t}_{\overrightarrow{e}_2}\left(\tilde{\mathbf{X}}^{t}_{\overrightarrow{e_1}}, \tilde{\mathbf{X}}^{t}_{\overleftarrow{e_2}}\right)\,,\\
\tilde{\mathbf{X}}^{t+1}_{\overleftarrow{e_2}} &= \tilde{\mathbf{Q}}^{\top}_{\overrightarrow{e}}\tilde{\mathbf{M}}^{t}_{\overleftarrow{e_2}} - b^{t}_{\overleftarrow{e_2}}\tilde{\mathbf{M}}^{t-1}_{\overrightarrow{e_2}}\,,\\
\tilde{\mathbf{M}}^{t}_{\overleftarrow{e_2}} &= \tilde{F}^{t}_{\overleftarrow{e_2}}\left(\tilde{\mathbf{Q}}_{\overrightarrow{e}_2}\mathbf{W}_{\overrightarrow{e}_2}, \tilde{\mathbf{X}}^{t}_{\overrightarrow{e_2}}, \tilde{\mathbf{X}}^{t}_{\overleftarrow{e_3}}\right)\,,\\[1em]
&\vdots\\[1em]
\tilde{\mathbf{X}}^{t+1}_{\overrightarrow{e_L}} &= \tilde{\mathbf{Q}}_{\overrightarrow{e}_L}\tilde{\mathbf{M}}^{t}_{\overrightarrow{e_L}} - b^{t}_{\overrightarrow{e_L}}\tilde{\mathbf{M}}^{t-1}_{\overleftarrow{e_L}}\,,\\
\tilde{\mathbf{M}}^{t}_{\overrightarrow{e_L}} &= \tilde{F}^{t}_{\overrightarrow{e}_L}\left(\tilde{\mathbf{X}}^{t}_{\overrightarrow{e}_{L-1}}, \tilde{\mathbf{X}}^{t}_{\overleftarrow{e_L}}\right)\,,\\
\tilde{\mathbf{X}}^{t+1}_{\overleftarrow{e_L}} &= \tilde{\mathbf{Q}}^{\top}_{\overrightarrow{e}_L}\tilde{\mathbf{M}}^{t}_{\overleftarrow{e_L}} - b^{t}_{\overleftarrow{e_L}}\tilde{\mathbf{M}}^{t-1}_{\overrightarrow{e_L}}\,,\\
\tilde{\mathbf{M}}^{t}_{\overleftarrow{e_L}} &= \tilde{F}^{t}_{\overleftarrow{e_L}}\left(\tilde{\mathbf{Q}}_{\overrightarrow{e}_L}\mathbf{W}_{\overrightarrow{e}_L}, \tilde{\mathbf{X}}^{t}_{\overrightarrow{e_L}}\right)
\end{aligned}
\tag{69}
$$

where each $\mathbf{W}_{\overrightarrow{e}_l}$ contains $k_{\overrightarrow{e}_l}$ copies of the initial $\mathbf{w}_{\overrightarrow{e}_l}$ reorganised into matrices as described above. The dimensions of the variables are Note that at this point we have almost reached an iteration verifying the structure of that appearing in Theorem A.2, except the Onsager term isn't, a priori, the correct one. Consider the following iteration, where we replaced the original, scalar Onsager terms with the correct, matrix-valued ones:

$$
\begin{aligned}
\tilde{\mathbf{X}}^{t+1}_{\overrightarrow{e_1}} &= \tilde{\mathbf{Q}}_{\overrightarrow{e}_1}\tilde{\mathbf{M}}^{t}_{\overrightarrow{e_1}} - \tilde{\mathbf{M}}^{t-1}_{\overleftarrow{e_1}}\left(\tilde{\mathbf{b}}^{t}_{\overrightarrow{e_1}}\right)^{\top}\,,\\
\tilde{\mathbf{M}}^{t}_{\overrightarrow{e_1}} &= \tilde{F}^{t}_{\overrightarrow{e}_1}(\tilde{\mathbf{X}}^{t}_{\overleftarrow{e_1}})\,,\\
\tilde{\mathbf{X}}^{t+1}_{\overleftarrow{e_1}} &= \tilde{\mathbf{Q}}^{\top}_{\overrightarrow{e}_1}\tilde{\mathbf{M}}^{t}_{\overleftarrow{e_1}} - \tilde{\mathbf{M}}^{t-1}_{\overrightarrow{e_1}}\left(\tilde{\mathbf{b}}^{t}_{\overleftarrow{e_1}}\right)^{\top}\,,\\
\tilde{\mathbf{M}}^{t}_{\overleftarrow{e_1}} &= \tilde{F}^{t}_{\overleftarrow{e_1}}\left(\tilde{\mathbf{Q}}_{\overrightarrow{e}_1}\mathbf{W}_{\overrightarrow{e}_1}, \tilde{\mathbf{X}}^{t}_{\overrightarrow{e_1}}, \tilde{\mathbf{X}}^{t}_{\overleftarrow{e_2}}\right)\,,\\[1em]
\tilde{\mathbf{X}}^{t+1}_{\overrightarrow{e_2}} &= \tilde{\mathbf{Q}}_{\overrightarrow{e}_2}\tilde{\mathbf{M}}^{t}_{\overrightarrow{e_2}} - \tilde{\mathbf{M}}^{t-1}_{\overleftarrow{e_2}}\left(\tilde{\mathbf{b}}^{t}_{\overrightarrow{e_2}}\right)^{\top}\,,\\
\tilde{\mathbf{M}}^{t}_{\overrightarrow{e_2}} &= \tilde{F}^{t}_{\overrightarrow{e}_2}\left(\tilde{\mathbf{X}}^{t}_{\overrightarrow{e_1}}, \tilde{\mathbf{X}}^{t}_{\overleftarrow{e_2}}\right)\,,\\
\tilde{\mathbf{X}}^{t+1}_{\overleftarrow{e_2}} &= \tilde{\mathbf{Q}}^{\top}_{\overrightarrow{e}}\tilde{\mathbf{M}}^{t}_{\overleftarrow{e_2}} - \tilde{\mathbf{M}}^{t-1}_{\overrightarrow{e_2}}\left(\tilde{\mathbf{b}}^{t}_{\overleftarrow{e_2}}\right)^{\top}\,,\\
\tilde{\mathbf{M}}^{t}_{\overleftarrow{e_2}} &= \tilde{F}^{t}_{\overleftarrow{e_2}}\left(\tilde{\mathbf{Q}}_{\overrightarrow{e}_2}\mathbf{W}_{\overrightarrow{e}_2}, \tilde{\mathbf{X}}^{t}_{\overrightarrow{e_2}}, \tilde{\mathbf{X}}^{t}_{\overleftarrow{e_3}}\right)
\end{aligned}
\tag{70}
$$

$$\vdots$$

$$\begin{aligned}\tilde{\mathbf{X}}^{t+1}_{\overrightarrow{e_L}} &= \tilde{\mathbf{Q}}_{\overrightarrow{e}_L}\tilde{\mathbf{M}}^{t}_{\overrightarrow{e_L}} - \tilde{\mathbf{M}}^{t-1}_{\overleftarrow{e_L}}\left(\tilde{\mathbf{b}}^{t}_{\overrightarrow{e_L}}\right)^{\top}, \\ \tilde{\mathbf{M}}^{t}_{\overrightarrow{e_L}} &= \tilde{F}^{t}_{\overrightarrow{e}_L}\left(\tilde{\mathbf{X}}^{t}_{\overrightarrow{e}_{L-1}}, \tilde{\mathbf{X}}^{t}_{\overleftarrow{e_L}}\right), \\ \tilde{\mathbf{X}}^{t+1}_{\overleftarrow{e_L}} &= \tilde{\mathbf{Q}}^{\top}_{\overrightarrow{e}_L}\tilde{\mathbf{M}}^{t}_{\overleftarrow{e_L}} - \tilde{\mathbf{M}}^{t-1}_{\overrightarrow{e_L}}\left(\tilde{\mathbf{b}}^{t}_{\overleftarrow{e_L}}\right)^{\top}, \\ \tilde{\mathbf{M}}^{t}_{\overleftarrow{e_L}} &= \tilde{F}^{t}_{\overleftarrow{e_L}}\left(\tilde{\mathbf{Q}}_{\overrightarrow{e}_L}\mathbf{W}_{\overrightarrow{e}_L}, \tilde{\mathbf{X}}^{t}_{\overrightarrow{e_L}}\right)\end{aligned} \tag{71}$$

where, for any $\overrightarrow{e} \in \overrightarrow{E}$ and any $t \in \mathbb{N}$ for the right oriented edges

$$\mathbf{b}^{t}_{\overrightarrow{e}_l} = \frac{1}{N}\sum_{i=1}^{n_{l-1}} \frac{\partial \tilde{F}^{t}_{\overrightarrow{e}_l,i}}{\partial \mathbf{X}_{\overleftarrow{e}_l,i}}\left(\left(\mathbf{X}^{t}_{\overrightarrow{e}'_l}\right)_{\overrightarrow{e}'_l : \overrightarrow{e}'_l \to \overrightarrow{e}_l}\right) \qquad \in \mathbb{R}^{q_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}}.$$

and left oriented edges

$$\mathbf{b}^{t}_{\overleftarrow{e}_l} = \frac{1}{N}\sum_{i=1}^{n_l} \frac{\partial \tilde{F}^{t}_{\overleftarrow{e}_l,i}}{\partial \mathbf{X}_{\overrightarrow{e}_l,i}}\left(\tilde{\mathbf{Q}}_{\overrightarrow{e}_l}\mathbf{W}_{\overrightarrow{e}_l}, \left(\mathbf{X}^{t}_{\overleftarrow{e}'_l}\right)_{\overleftarrow{e}'_l : \overleftarrow{e}'_l \to \overleftarrow{e}_l}\right) \qquad \in \mathbb{R}^{q_{\overleftarrow{e}_l} \times q_{\overleftarrow{e}_l}}.$$

Using the separability assumption, we can simplify this expression. To take a concrete example, consider $\tilde{F}^{t}_{\overrightarrow{e}_2}\left(\tilde{\mathbf{X}}^{t}_{\overrightarrow{e_1}}, \tilde{\mathbf{X}}^{t}_{\overleftarrow{e_2}}\right)$. Let's start with the dimensions. Recall

$$\tilde{f}^{t}_{\overrightarrow{e}_2}\left(\tilde{\mathbf{x}}^{t}_{\overrightarrow{e}_1}, \tilde{\mathbf{x}}^{t}_{\overleftarrow{e}_2}\right) \in \mathbb{R}^{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2}} = \mathbf{V}^{\top}_{\overrightarrow{e}_2} f^{t}_{\overrightarrow{e}_2}\left(\mathbf{U}_{\overrightarrow{e}_1}\tilde{\mathbf{x}}^{t}_{\overrightarrow{e_1}}, \mathbf{V}_{\overrightarrow{e}_2}\tilde{\mathbf{x}}^{t}_{\overleftarrow{e_2}}\right) \tag{72}$$

$$\text{where} \quad \tilde{\mathbf{x}}^{t}_{\overrightarrow{e}_1} \in \mathbb{R}^{D_{\overrightarrow{e}_1} q_{\overrightarrow{e}_1}} = \mathbb{R}^{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2}} \text{ and } \tilde{\mathbf{x}}^{t}_{\overleftarrow{e}_2} \in \mathbb{R}^{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2}} \tag{73}$$

using the separability assumption, we may write

$$\forall\, 1 \leqslant i \leqslant P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2} \tag{74}$$

$$\left(f^{t}_{\overrightarrow{e}_2}\left(\mathbf{U}_{\overrightarrow{e}_1}\tilde{\mathbf{x}}^{t}_{\overrightarrow{e_1}}, \mathbf{V}_{\overrightarrow{e}_2}\tilde{\mathbf{x}}^{t}_{\overleftarrow{e_2}}\right)\right)_i = \sigma^{t}_{\overrightarrow{e}_2}\left(\left(\mathbf{U}_{\overrightarrow{e}_1}\tilde{\mathbf{x}}^{t}_{\overrightarrow{e_1}}\right)_i, \left(\mathbf{V}_{\overrightarrow{e}_2}\tilde{\mathbf{x}}^{t}_{\overleftarrow{e_2}}\right)_i\right) \tag{75}$$

And

$$\tilde{F}^{t}_{\overrightarrow{e}_2}\left(\tilde{\mathbf{X}}^{t}_{\overrightarrow{e}_1}, \tilde{\mathbf{X}}^{t}_{\overleftarrow{e}_2}\right) \in \mathbb{R}^{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2} \times q_{\overrightarrow{e}_2}} \tag{76}$$

$$\text{where} \quad \tilde{\mathbf{X}}^{t}_{\overrightarrow{e}_1} \mathbb{R}^{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2} \times q_{\overrightarrow{e}_2}} \text{ and } \tilde{\mathbf{X}}^{t}_{\overleftarrow{e}_2} \in \mathbb{R}^{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2} \times q_{\overrightarrow{e}_2}} \tag{77}$$

$$\tilde{F}^{t}_{\overrightarrow{e}_2}\left(\tilde{\mathbf{X}}^{t}_{\overrightarrow{e}_1}, \tilde{\mathbf{X}}^{t}_{\overleftarrow{e}_2}\right) = \begin{bmatrix} \left[(\tilde{f}^{t}_{\overrightarrow{e}_l})^{(1)}(\tilde{\mathbf{x}}^{t,(1)}_{\overrightarrow{e}_1}, \tilde{\mathbf{x}}^{t,(1)}_{\overleftarrow{e}_2}) \ldots (\tilde{f}^{t}_{\overrightarrow{e}_l})^{(q_{\overrightarrow{e}_l})}(\tilde{\mathbf{x}}^{t,(q_{\overrightarrow{e}_l})}_{\overrightarrow{e}_1}, \tilde{\mathbf{x}}^{t,(q_{\overrightarrow{e}_l})}_{\overleftarrow{e}_2})\right]_{i=1}^{k_{\overrightarrow{e}_2}} \\ 0_{P_{\overrightarrow{e}_2}(q_{\overrightarrow{e}_2} - k_{\overrightarrow{e}_2}) \times q_{\overrightarrow{e}_2}} \end{bmatrix} \tag{78}$$

$$= \left[(\tilde{g}^{t}_{\overrightarrow{e}_l})^{(1)}(\tilde{\mathbf{x}}^{t,(1)}_{\overrightarrow{e}_1}, \tilde{\mathbf{x}}^{t,(1)}_{\overleftarrow{e}_2}) \ldots (\tilde{g}^{t}_{\overrightarrow{e}_l})^{(q_{\overrightarrow{e}_l})}(\tilde{\mathbf{x}}^{t,(q_{\overrightarrow{e}_l})}_{\overrightarrow{e}_1}, \tilde{\mathbf{x}}^{t,(q_{\overrightarrow{e}_l})}_{\overleftarrow{e}_2})\right]_{i=1}^{q_{\overrightarrow{e}_2}} \tag{79}$$

where each $\tilde{\mathbf{x}}^{t,(i)}_{\overleftarrow{e}_2} \in \mathbb{R}^{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2}}$. Recall that, for any $1 \leqslant i \leqslant Pk$, $\tilde{F}^t_{\overrightarrow{e}_2,i} : \mathbb{R}^{q_{\overrightarrow{e}_2}} \to \mathbb{R}^{q_{\overrightarrow{e}_2}}$. Then, for any $1 \leqslant k, l \leqslant q_{\overrightarrow{e}_2}$

$$\left(\tilde{\mathbf{b}}^t_{\overrightarrow{e}_2}\right)_{k,l} = \frac{1}{N} \sum_{i=1}^{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2}} \frac{\partial \tilde{F}^t_{\overrightarrow{e}_2,i,k}}{\partial \mathbf{X}_{\overleftarrow{e}_2,i,l}} \left(\tilde{\mathbf{X}}^t_{\overrightarrow{e}_1}, \tilde{\mathbf{X}}^t_{\overleftarrow{e}_2}\right) \tag{80}$$

$$= \frac{1}{N} \sum_{i=1}^{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2}} \frac{\partial (\tilde{g}^t_{\overrightarrow{e}_2,i})^{(k)}}{\partial \tilde{\mathbf{x}}^{(l)}_{\overleftarrow{e}_2,i}} (\tilde{\mathbf{x}}^{t,(k)}_{\overrightarrow{e}_1}, \tilde{\mathbf{x}}^{t,(k)}_{\overleftarrow{e}_2}) \tag{81}$$

$$= \frac{1}{N} \sum_{i=1}^{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2}} \frac{\partial}{\partial \tilde{\mathbf{x}}^{t,(l)}_{\overleftarrow{e}_2}} \mathbf{V}^\top_{\overrightarrow{e}_2} (g^t_{\overrightarrow{e}_2})^{(k)} \left(\mathbf{U}_{\overrightarrow{e}_1} \tilde{\mathbf{x}}^{t,(l)}_{\overrightarrow{e_1}}, \mathbf{V}_{\overrightarrow{e}_2} \tilde{\mathbf{x}}^{t,(l)}_{\overleftarrow{e_2}}\right) \tag{82}$$

$$= \frac{1}{N} \mathrm{Tr} \left(\mathbf{V}^\top_{\overrightarrow{e}_2} \mathcal{J}_{(g^t_{\overrightarrow{e}_2})^{(k)}} \left(\mathbf{U}_{\overrightarrow{e}_1} \tilde{\mathbf{x}}^{t,(l)}_{\overrightarrow{e_1}}, \mathbf{V}_{\overrightarrow{e}_2} \tilde{\mathbf{x}}^{t,(l)}_{\overleftarrow{e_2}}\right) \mathbf{V}_{\overrightarrow{e}_2}\right) \delta_{k,l} \tag{83}$$

$$= \frac{1}{N} \mathrm{Tr} \left(\mathcal{J}_{(g^t_{\overrightarrow{e}_2})} \left(\mathbf{U}_{\overrightarrow{e}_1} \tilde{\mathbf{x}}^t_{\overrightarrow{e_1}}, \mathbf{V}_{\overrightarrow{e}_2} \tilde{\mathbf{x}}^t_{\overleftarrow{e_2}}\right)\right) \delta_{k,l} \tag{84}$$

$$= \frac{1}{N} \sum_{i=1}^{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2}} (\sigma^t)'_{\overrightarrow{e}_2} \left(\left(\mathbf{U}_{\overrightarrow{e}_1} \tilde{\mathbf{x}}^t_{\overrightarrow{e_1}}\right)_i, \left(\mathbf{V}_{\overrightarrow{e}_2} \tilde{\mathbf{x}}^t_{\overleftarrow{e_2}}\right)_i\right) \delta_{k,l} \tag{85}$$

where we wrote $\mathcal{J}_{(g^t_{\overrightarrow{e}_2})^{(k)}}$ the $N \times N$ Jacobian matrix of the function $(g^t_{\overrightarrow{e}_2})^{(k)} : \mathbb{R}^N \to \mathbb{R}^N$. Using Berthier et al. [2020] corollary 2, the Onsager term can be replaced by any estimator based on the asymptotically Gaussian iterates converging, in the high-dimensional limit, to the correct expectation. The adaptation to the graph framework of Gerbelot and Berthier [2021] is immediate (see the proof in Berthier et al. [2020] and corresponding comment in Gerbelot and Berthier [2021]). Using the permutation invariance of the Gaussian distribution, we can therefore replace each element of the matrix the Onsager term with

$$\frac{1}{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2}} \sum_{i=1}^{P_{\overrightarrow{e}_2} q_{\overrightarrow{e}_2}} (\sigma^t)'_{\overrightarrow{e}_2} \left(\left(\tilde{\mathbf{x}}^t_{\overrightarrow{e_1}}\right)_i, \left(\tilde{\mathbf{x}}^t_{\overleftarrow{e_2}}\right)_i\right) \delta_{k,l} \tag{86}$$

which amounts to

$$\tilde{\mathbf{b}}^t_{\overrightarrow{e}_2} = b^t_{\overrightarrow{e}_2} \mathbf{I}_{q_{\overrightarrow{e}_2} \times q_{\overrightarrow{e}_2}} \tag{87}$$

We therefore obtain an exact reformulation of the initial MLAMP iteration with convolutional matrices in terms of a subset (first row of size $P_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}$ for right oriented edges and $D_{\overrightarrow{e}_l} \times q_{\overrightarrow{e}_l}$ for left-oriented variables) of the variables of a matrix-valued iteration with dense Gaussian matrices verifying the SE equations. Isolating the aforementioned first lines, recalling that the SE equations prescribes i.i.d. rows in the asymptotically Gaussian fields, we recover that, for any $1 \leqslant l \leqslant L$, the variable $\mathbf{x}_{\overrightarrow{e}_l} \in \mathbb{R}^{P_{\overrightarrow{e}_l} q_{\overrightarrow{e}_l}}$ is composed of $q_{\overrightarrow{e}_l}$ copies of block of size $P_{\overrightarrow{e}_l}$ with i.i.d. Gaussian elements distributed according to the SE equations (A.3). The distribution of the variables associated to left-oriented edges is obtained similarly. Note that, from a finite size point of view, the effect of $D_{\overrightarrow{e}_l}, P_{\overrightarrow{e}_l}$ is different from that of $q_{\overrightarrow{e}_l}$: the former results in subGaussian concentration i.e. exponential in the dimension, while the latter only represents copies (and not i.i.d. samples), and thus only has an averaging effect. This is observed in simulations.

□

A.4 Bayes-optimal MLAMP with random convolutional matrices

In this section, we specialize the equations obtained in the previous section to the Bayes-optimal MLAMP iteration of the main body of the paper. Several functions are reminded for convenience. Consider the MLAMP iteration outlined in section 3.3. The scalar updates described in Eq.(4) can be rewritten as vector-valued updates as follows, for any $t \in \mathbb{N}$, and any $0 \leqslant l \leqslant L$:

$$\boldsymbol{\omega}^{(l)}(t) = \mathbf{W}^{(l)} \hat{\mathbf{h}}^{(l)}(t) - V^{(l)}(t) \mathbf{g}^{(l)}(t-1) \tag{88}$$

$$\mathbf{B}^{(l)}(t) = \left(\mathbf{W}^{(l)}\right)^\top \mathbf{g}^{(l)}(t) - \hat{V}^{(l)}(t) \hat{\mathbf{h}}(t). \tag{89}$$

To define the update functions and terms $V^{(l)}, \hat{V}^{(l)}$, the following partition functions were introduced.

- for $l = 1$
$$\mathcal{Z}^{(1)}\left(y, V^{(1)}, \omega^{(1)}\right) = \frac{1}{\sqrt{2\pi V^{(1)}}} \int dz P^{(1)}_{out}(y|z) e^{-\frac{(z-\omega^{(1)})^2}{2V^{(1)}}} \tag{90}$$
- for any $2 \leqslant l \leqslant L-1$:
$$\mathcal{Z}^{(l)}\left(A^{(l-1)}, B^{(l-1)}, V^{(l)}, \omega^{(l)}\right) = \frac{1}{\sqrt{2\pi V^{(l)}}} \int dhdz P^{(l)}_{out}(h|z) e^{-\frac{1}{2}A^{(l-1)}h^2 + B^{(l-1)}h} e^{-\frac{(z-\omega^{(l)})^2}{2V^{(l)}}} \tag{91}$$
- for $l = L$
$$\mathcal{Z}^{(L)}(A^{(L)}, B^{(L)}) = \int dh P_X(h) e^{-\frac{1}{2}A^{(L)}h^2 + B^{(L)}h} \tag{92}$$

We then define the layer-dependent, time-dependent, scalar update functions $f^{(l),t}$, $\tilde{f}^{(l),t}$

$$\forall\ (B, \omega) \in \mathbb{R}^2$$
$$f^{(1),t}(\omega) = \partial_\omega \log \mathcal{Z}^{(1)}\left(y, V^{(1)}(t), \omega\right) \tag{93}$$
$$f^{(l),t}(B, \omega) = \partial_\omega \log \mathcal{Z}^{(l)}\left(A^{(l-1)}(t), B, V^{(l)}(t), \omega\right)\ 2 \leqslant l \leqslant L \tag{94}$$
$$\tilde{f}^{(l),t}(B, \omega) = \partial_B \log \mathcal{Z}^{(l+1)}\left(A^{(l)}(t-1), B, V^{(l+1)}(t-1), \omega\right)\ 1 \leqslant l \leqslant L-1 \tag{95}$$
$$\tilde{f}^{(L,t)}(B) = \partial_B \log \mathcal{Z}^{(L+1)}\left(A^{(L)}(t-1), B\right), \tag{96}$$

and their corresponding separable, vector valued counterparts $\mathbf{f}^{(l)}, \tilde{\mathbf{f}}^{(l)}$, which leads to the following iteration

$$\boldsymbol{\omega}^{(l)}(t) = \mathbf{W}^{(l)} \tilde{\mathbf{f}}^{(l),t}(\mathbf{B}^{(l),t-1}, \boldsymbol{\omega}^{(l+1),t-1}) - V^{(l)}(t) \mathbf{f}^{(l),t-1}(\mathbf{B}^{(l-1),t-1}, \boldsymbol{\omega}^{(l),t-1}) \tag{97}$$
$$\mathbf{B}^{(l)}(t) = \left(\mathbf{W}^{(l)}\right)^\top \mathbf{f}^{(l),t}(\mathbf{B}^{(l-1),t}, \boldsymbol{\omega}^{(l),t}) - \hat{V}^{(l)}(t) \tilde{\mathbf{f}}^{(l),t}(\mathbf{B}^{(l),t-1}, \boldsymbol{\omega}^{(l+1),t-1}), \tag{98}$$

where the Onsager terms $V^{(l),t}$ and $\hat{V}^{(l),t}$ reduce to, using the separability of the update functions,

$$V^{(l),t} = \frac{1}{n_l} \sum_{i=1}^{n_{l-1}} \partial_B \tilde{f}^{(l),t}(B_i^{(l),t-1}, \omega_i^{(l+1),t-1}) \tag{99}$$
$$\hat{V}^{(l),t} = \frac{1}{n_l} \sum_{j=1}^{n_l} \partial_\omega f^{(l),t}(B_j^{(l-1),t}, \omega_j^{(l),t}) = -A^{(l),t} \tag{100}$$

We now show that the update functions defined above are Lipschitz continuous and increasing, thus ensuring that the integrals are well defined through positivity of the parameters $V, \hat{V}$.

Lemma A.4. For any $1 \leqslant l \leqslant L$, and any $t \in \mathbb{N}$, the functions $f^{(l),t}$, $\tilde{f}^{(l),t}$ are Lipschitz continuous in B, ω. Furthermore, the functions $f^{(l),t}$, $\tilde{f}^{(l),t}$ are respectively decreasing in ω and increasing in B. As a consequence, the variance terms $A^{(l),t}$ and $V^{(l),t}$ are strictly positive.

Proof. Recall the partition function, omitting the layer index since all regularity assumptions are the same for all layers and time indices,

$$\mathcal{Z}(A, B, V, \omega) := \frac{1}{\sqrt{2\pi V}} \int P(h \mid z) \exp\left(Bh - \frac{1}{2}Ah^2 - \frac{(z-\omega)^2}{2V}\right) dh\, dz \tag{101}$$

recalling $p(h|z) = \int p(\xi)\delta(h - f_\xi(z))d\xi$, integrating in h yields

$$\mathcal{Z}(A, B, V, \omega) := \frac{1}{\sqrt{2\pi V}} \int P(\xi) \exp\left(Bf_\xi(z) - \frac{1}{2}Af_\xi(z)^2 - \frac{(z-\omega)^2}{2V}\right) d\xi\, dz \tag{102}$$

Starting with $\tilde{f}$, we can straightforwardly verify the conditions to apply the dominated convergence theorem and differentiate under the integral to obtain

$$\begin{aligned}&\partial_B \tilde{f}(B,\omega) = \partial_B^2 \log\left(\mathcal{Z}(A,B,V,\omega)\right) \\ &= \frac{1}{(\sqrt{2\pi V}\mathcal{Z}(A,B,V,\omega))^2}\left(\int P(\xi) f_\xi^2(z) \exp\left(B f_\xi(z) - \frac{1}{2}A f_\xi(z)^2 - \frac{(z-\omega)^2}{2V}\right) d\xi\, dz \times \right. \\ &\int P(\xi) \exp\left(B f_\xi(z) - \frac{1}{2}A f_\xi(z)^2 - \frac{(z-\omega)^2}{2V}\right) d\xi\, dz - \\ &\qquad \left.\left(\int P(\xi) f_\xi(z) \exp\left(B f_\xi(z) - \frac{1}{2}A f_\xi(z)^2 - \frac{(z-\omega)^2}{2V}\right) d\xi\, dz\right)^2\right) \geqslant 0\end{aligned} \tag{103}$$

where the positivity comes from the Cauchy-Schwarz inequality and positivity of the term $P(\xi)\exp\left(B f_\xi(z) - \frac{1}{2}A f_\xi(z)^2 - \frac{(z-\omega)^2}{2V}\right)$. Turning to f, we complete the square in the variable h to obtain

$$\mathcal{Z}(A,B,V,\omega) := \frac{\exp\left(\frac{B^2}{2A}\right)}{\sqrt{2\pi V}} \int P(\xi) \exp\left(-\frac{A}{2}\left(f_\xi(z) - \frac{B}{A}\right)^2\right) \exp\left(-\frac{(z-\omega)^2}{2V}\right) d\xi\, dz \tag{104}$$

and differentiating under the integral yields

$$f(B,\omega) = \partial_\omega \log\left(\mathcal{Z}(A,B,V,\omega)\right) \tag{105}$$

$$= \frac{1}{V}\left(\frac{\int P(\xi) z \exp\left(-\frac{A}{2}\left(f_\xi(z) - \frac{B}{A}\right)^2\right) \exp\left(-\frac{(z-\omega)^2}{2V}\right) d\xi\, dz}{\left(\int P(\xi) \exp\left(-\frac{A}{2}\left(f_\xi(z) - \frac{B}{A}\right)^2\right) \exp\left(-\frac{(z-\omega)^2}{2V}\right) d\xi\, dz\right)} - \omega\right) \tag{106}$$

where the term $\frac{\int P(\xi) z \exp\left(-\frac{A}{2}\left(f_\xi(z) - \frac{B}{A}\right)^2\right) \exp\left(-\frac{(z-\omega)^2}{2V}\right) d\xi\, dz}{\left(\int P(\xi) \exp\left(-\frac{A}{2}\left(f_\xi(z) - \frac{B}{A}\right)^2\right) \exp\left(-\frac{(z-\omega)^2}{2V}\right) d\xi\, dz\right)}$ is the conditional mean of the distribution with density $\frac{\int P(\xi) \exp\left(-\frac{A}{2}\left(f_\xi(z) - \frac{B}{A}\right)^2\right) \exp\left(-\frac{(z-\omega)^2}{2V}\right) d\xi}{\left(\int P(\xi) \exp\left(-\frac{A}{2}\left(f_\xi(z) - \frac{B}{A}\right)^2\right) \exp\left(-\frac{(z-\omega)^2}{2V}\right) d\xi\, dz\right)}$. The Lipschitz property is straightforward to verify using the polynomial bound assumption on the activation functions and the inverse exponential factors. □

In the Bayes-optimal MLAMP, see Manoel et al. [2017], the planted vectors $\mathbf{w}_{\overrightarrow{e}_l}$ are chosen as independently distributed as the asymptotic SE representation of the output of the previous layer, and are therefore Lipschitz transforms of subGaussian random variables, and thus are also subgaussian. Using the permuation invariance of the Gaussian distribution, the quantities $\mathbf{z}_{\overrightarrow{e}_l} = \hat{\mathbf{A}}_{\overrightarrow{e}_l}$ remain Gaussian. We can therefore apply the result of Lemma A.3 to this iteration and obtain that iterates of Eq.(4) verify the SE equations from Lemma A.3 with the corresponding update functions. Furthermore, in the Bayes optimal case, the Nishimori conditions, see e.g. Krzakala et al. [2012], allow to only keep the parameters $\nu_{\overrightarrow{e}_l}, \hat{\nu}_{\overleftarrow{e}_l}$ to describe the distribution of of the iterates, recovering the equations of Theorem 4.2. Finally, the rescaling of the variances to go from the factors δ_l to the β_l of the main can be done by rescaling each non-linearity $f^t_{\overrightarrow{e}_l}$ by $\sqrt{N/n_{l-1}}$ (and similary for the $f^t_{\overleftarrow{e}_l}$ with $\sqrt{N/n_l}$) as done in Javanmard and Montanari [2013], Berthier et al. [2020].

B Fast MCC-vector Products

Here is a simple sketch of an algorithm for multiplying $M \sim \mathrm{MCC}(D,P,q,k)$ with a vector $v \in \mathbb{R}^{Pq}$ that runs in time $O(DPq\log q)$. If $k \gg \log q$, this improves on the runtime required by a simple sparse matrix-vector product. We use Matlab index notation for matrix and vector coordinates, for example $M[i:j,k] = [M_{rk} : r = i \dots j]$, and we write shorthand M_{ij} for $M[i,j]$.

Data: matrix $M \sim \text{MCC}(D, P, q, k)$, vector $v \in \mathbb{R}^{Pq}$
Initialize $s \in \mathbb{R}^{Dq}$ the zero vector;
for $i = 1 \ldots D$ **do**
    for $j = 1 \ldots P$ **do**
        $C_{ij} \leftarrow M[qi : q(i+1),\, qj : q(j+1)]$;
        $\omega_{ij} = C_{ij}[0,\, 0 : k]$;
        $\hat{\omega}_{ij} = \text{FFT}(w_{ij})$;
        $\hat{v}_j = \text{FFT}(v[qj,\, q(j+1)]$;
        $\hat{s}_i = \hat{\omega}_{ij} * \hat{v}_j$;
        $s[qi : q(i+1)] = \text{IFFT}(\hat{s}_i)$;
    end
end

Algorithm 1: $O(DPq \log q)$ time algorithm for MCC-vector products

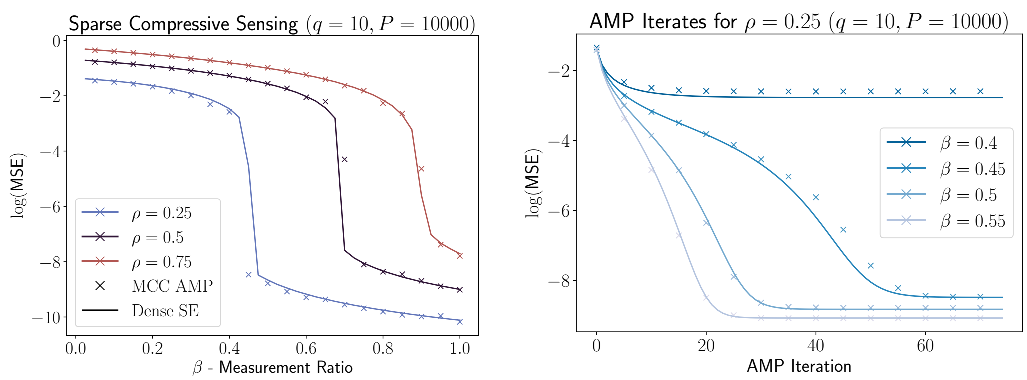


Figure 6: Replica of Figure 1 for $q = 10$ and $P = 10000$. (**left**) Compressive sensing $y_0 = Wx_0 + \zeta$ for noise $\zeta_i \sim \mathcal{N}(0, 10^{-4})$ and signal prior $x_0 \sim \rho\mathcal{N}(0, 1) + (1 - \rho)\delta(x)$, where $W \in \mathbb{R}^{Dq \times Pq}$ has varying aspect ratio $\beta = D/P$. Crosses correspond to AMP evaluations for $W \sim \text{MCC}(D, P, q, k)$ according to Definition 3.2, averaged over 10 independent trials. Lines show the state evolution predictions when $W_{ij} \sim \mathcal{N}(0, 1/Pq)$. The system size is $P = 10000$, $q = 10$, $k = 3$, where β and $D = \beta P$ vary. (**right**) AMP iterates at $\rho = 0.25$ and β near the recovery transition.

C Additional Experiments

C.1 Sparse Compressive Sensing

We observe empirically that in the sparse compressive sensing task of Figure 1, the relative sizes of (D, P) and q have little impact on the performance of the corresponding AMP iteration. In Figure 6, we show a replica of this figure with $q = 10$ and $P = 10000$. Despite a significant difference between the relative sizes of these parameters, the AMP iterations behave largely the same.

C.2 Empirical Results for Vector-AMP Algorithms

We observe that a similar equivalence property as Theorem 4.2 holds for algorithms based on the VAMP framework Schniter et al. [2016], Fletcher et al. [2018], Baker et al. [2020]. Previously, state evolution has been proven for such algorithms when their sensing matrices are drawn from a right-orthogonally-invariant ensemble. While the random MCC ensemble does not satisfy this property, we show in Figure 7 a comparison between empirical VAMP performance and the corresponding SE predictions for dense matrices, which are almost identical.

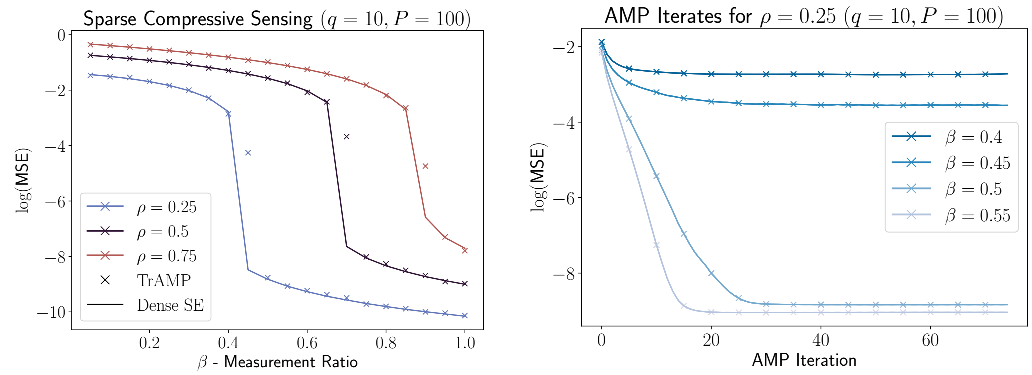


Figure 7: Replica of Figure 1 using Tree-AMP Baker et al. [2020], a compositional VAMP type algorithm, for $q = 10$ and $P = 100$. (**left**) Compressive sensing $y_0 = Wx_0 + \zeta$ for noise $\zeta_i \sim \mathcal{N}(0, 10^{-4})$ and signal prior $x_0 \sim \rho\mathcal{N}(0,1) + (1-\rho)\delta(x)$, where $W \in \mathbb{R}^{Dq \times Pq}$ has varying aspect ratio $\beta = D/P$. Crosses correspond to AMP evaluations for $W \sim \text{MCC}(D, P, q, k)$ according to Definition 3.2, averaged over 30 independent trials. Lines show the state evolution predictions when $W_{ij} \sim \mathcal{N}(0, 1/Pq)$. The system size is $P = 100$, $q = 10$, $k = 3$, where β and $D = \beta P$ vary. (**right**) AMP iterates at $\rho = 0.25$ and β near the recovery transition.

D Structured Convolutions and Non-separable Denoising

Our proof uses a relatively simple version of spatial coupling, leaving avenues for potential generalizations. Spatially coupled sensing matrices typically consist of a block structured matrix whose blocks are i.i.d. Gaussian with different variances, as in (for instance) Krzakala et al. [2012], Barbier et al. [2015]. As a model, consider $\tilde{M}_{\text{sp}}$ of the following form, with variances $\kappa \in \mathbb{R}_+^{q \times q}$,

$$\tilde{M}_{\text{sp}} = \begin{bmatrix} \kappa_{11}A_{11} & \kappa_{12}A_{12} & \dots & \kappa_{1q}A_{1q} \\ \kappa_{21}A_{21} & \ddots & & \vdots \\ \vdots & & & \\ \kappa_{1q}A_{1q} & \dots & & \kappa_{qq}A_{qq} \end{bmatrix}.$$

As a result of Lemma 4.3, a given MCC matrix M is equivalent to $\tilde{M}$ corresponding to the case where κ is a convolutional matrix according to Definition 3.1, with filter $\omega = [1\ 1\ \dots\ 1] \in \mathbb{R}^k$. One avenue to extend our results is to consider general $\tilde{M}$ where κ is any circulant matrix. Inverting the permutation lemma, this corresponds to MCC matrices whose convolutional blocks have filters with independent non-isotropic coordinates, as in the following definition, which may be viewed as a simple model for structured convolutional filters.

Definition D.1 (Independent Gaussian Random Convolutions)**.** Let $\vec{\kappa} = [\kappa_1, \dots, \kappa_k] \in \mathbb{R}_+^k$ and let $\Sigma = \text{diag}(\vec{\kappa})$. Let $q \geq k > 0$ be integers. The Gaussian convolutional ensemble $\mathcal{C}(q, k)$ contains random circulant matrices $C \in \mathbb{R}^{q \times q}$ whose first rows are given by $C_1 = \texttt{Zero-pad}_{q,k}[\omega]$ where $\omega \sim \mathcal{N}(0, \Sigma)$.

This model is a natural extension of our current setting, which is also amenable to proof techniques designed for spatial coupling. However, because the nonzero coordinates of the sensing matrix are no longer i.i.d., the Bayes-optimal denoising functions corresponding to this problem are non-separable. So, an equivalence theorem analogous to Theorem 4.2 is not expected to hold – in other words, state evolution in this convolutional model is not expected to reduce to that of a signal model with dense i.i.d. couplings. More generally, multilayer AMP iterations with non-separable non-linearities may be written to compute marginals of posterior distributions involving such functions, and will verify SE equations. However there will be no direct correspondance with the iteration and SE equations of the fully separable case.