Supplemental Material: Reconstructing Video from Radio Interferometric Measurements of Time-Varying Sources

Katherine L. Bouman[1,2], Michael D. Johnson[2], Adrian V. Dalca[1,3], Andrew Chael[2], Freek Roelofs[4], Sheperd S. Doeleman[2], William T. Freeman[1,5]

[1]Massachusetts Institute of Technology, CSAIL [2]Harvard-Smithsonian Center for Astrophysics
[3]Massachusetts General Hospital, HMS [4]Radboud University [5]Google Research

Contents

1 Data Products

1.1 Approximate Gaussian Noise on the Bispectrum

In this section we present how we approximate the noise on each bispectrum term in $\boldsymbol{y}$ when systematic gain error is negligible.

A measured visibility, Γ, is corrupted by thermal and atmospheric phase error in the form

$$\Gamma_{jk} = \hat{\Gamma}_{jk} \exp\left[i(\phi_j - \phi_k)\right] + \epsilon_{jk}, \tag{1}$$

for uniformly distributed phase errors, ϕ_j and ϕ_k, and Gaussian distributed thermal noise, ϵ_{jk} on an ideal visibility $\hat{\Gamma}_{jk}$ [2]. Although the phase error due to atmospheric inhomogeneity cancels when using the bispectrum, residual error exists due to thermal noise [3]. Since the bispectrum is the product of three visibilities, its noise distribution is not Gaussian; nonetheless, it can be approximated using a Gaussian which characterizes its first-order noise terms. We determine how to approximate the bispectrum's noise by first expanding the noise terms in the triple product:

$$\begin{aligned}
\Gamma_{12}\Gamma_{23}\Gamma_{31} &= \left(\hat{\Gamma}_{12} \exp\left[i(\phi_1 - \phi_2)\right] + \epsilon_{12}\right)\left(\hat{\Gamma}_{23} \exp\left[i(\phi_2 - \phi_3)\right] + \epsilon_{23}\right)\left(\hat{\Gamma}_{31} \exp\left[i(\phi_3 - \phi_1)\right] + \epsilon_{31}\right) \\
&= \hat{\Gamma}_{12}\hat{\Gamma}_{23}\hat{\Gamma}_{31} + \epsilon_{12}\hat{\Gamma}_{23}\hat{\Gamma}_{31} \exp\left[i(\phi_2 - \phi_1)\right] + \epsilon_{23}\hat{\Gamma}_{12}\hat{\Gamma}_{31} \exp\left[i(\phi_3 - \phi_2)\right] \ldots \\
&+ \epsilon_{31}\hat{\Gamma}_{23}\hat{\Gamma}_{12} \exp\left[i(\phi_1 - \phi_3)\right] + \epsilon_{12}\epsilon_{23}\hat{\Gamma}_{31} \exp\left[i(\phi_3 - \phi_1)\right] \ldots \\
&+ \epsilon_{23}\epsilon_{31}\hat{\Gamma}_{12} \exp\left[i(\phi_1 - \phi_2)\right] + \epsilon_{12}\epsilon_{31}\hat{\Gamma}_{23} \exp\left[i(\phi_2 - \phi_3)\right] + \epsilon_{12}\epsilon_{23}\epsilon_{31}
\end{aligned} \tag{2}$$

As expected, the resulting expression contains a term, $\hat{\Gamma}_{12}\hat{\Gamma}_{23}\hat{\Gamma}_{31}$, which no longer is affected by the introduced phase errors. The first order noise terms,

$$\epsilon_{12}\hat{\Gamma}_{23}\hat{\Gamma}_{31} \exp\left[i(\phi_2 - \phi_1)\right] + \epsilon_{23}\hat{\Gamma}_{12}\hat{\Gamma}_{31} \exp\left[i(\phi_3 - \phi_2)\right] + \epsilon_{31}\hat{\Gamma}_{23}\hat{\Gamma}_{12} \exp\left[i(\phi_1 - \phi_3)\right], \tag{3}$$

contribute Gaussian noise to the bispectrum. However, they are each scaled by a value related to the ideal visibilities. Although we are unable to measure these ideal values, we approximate them by using the corresponding measured visibilities. As we expect noise on the bispectrum to be isotropic, we approximate noise on the real and imaginary component of $\Gamma_{12}\Gamma_{23}\Gamma_{31}$ as Gaussian with variance

$$\sigma_{12}^2|\Gamma_{23}\Gamma_{31}|^2 + \sigma_{23}^2|\Gamma_{12}\Gamma_{31}|^2 + \sigma_{31}^2|\Gamma_{12}\Gamma_{23}|^2. \tag{4}$$

By comparing our approximation to distributions obtained through sampling we have seen that it accurately models the true noise for SNR values greater than 1. As we expect values of SNR greater than 1 for imaging, a Gaussian noise model is a reasonable approximation for the bispectrum.

1.2 Data Product Derivatives

In this section $\boldsymbol{F}^R$ and $\boldsymbol{F}^I$ indicate the real and imaginary portions of a complex-valued matrix $\boldsymbol{F}$, respectively.

1.2.1 Bispectrum

$$\begin{aligned}
\boldsymbol{F}_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}_{i_{1,3}}\boldsymbol{x} &= (\boldsymbol{F}^R_{i_{1,2}} + i\boldsymbol{F}^I_{i_{1,2}})\boldsymbol{x}(\boldsymbol{F}^R_{i_{2,3}} + i\boldsymbol{F}^I_{i_{2,3}})\boldsymbol{x}(\boldsymbol{F}^R_{i_{1,3}} + i\boldsymbol{F}^I_{i_{1,3}})\boldsymbol{x} \\
&= (\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x} + i\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x})(\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x} + i\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x})(\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} + i\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x}) \\
&= \boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} - \boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} - \boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} - \boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} + ... \\
&\quad i\left(\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} - \boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x}\right) \\
&= \xi^R_i(\boldsymbol{x}) + i\xi^I_i(\boldsymbol{x})
\end{aligned} \tag{5}$$

We must find the derivative of ξ^R_i with respect to $\boldsymbol{x}$

$$\begin{aligned}
\frac{d}{d\boldsymbol{x}}\xi^R_i &= \frac{d}{d\boldsymbol{x}}\left(\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} - \boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} - \boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} - \boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x}\right) \\
&= \frac{d}{d\boldsymbol{x}}(\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x}) - \frac{d}{d\boldsymbol{x}}(\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x}) - \frac{d}{d\boldsymbol{x}}(\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x}) - \frac{d}{d\boldsymbol{x}}(\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x}) \\
&= \boldsymbol{F}^R_{i_{1,2}}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^R_{i_{2,3}}\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^R_{i_{1,3}}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}... \\
&- (\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^I_{i_{2,3}}\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^I_{i_{1,3}}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x})... \\
&- (\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^R_{i_{2,3}}\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^I_{i_{1,3}}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x})... \\
&- (\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^I_{i_{2,3}}\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^R_{i_{1,3}}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x})
\end{aligned} \tag{6}$$

Similarly, the derivative of ξ^I_i with respect to $\boldsymbol{x}$ is

$$\begin{aligned}
\frac{d}{d\boldsymbol{x}}\xi^I_i &= \frac{d}{d\boldsymbol{x}}\left(\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} - \boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x}\right) \\
&= \boldsymbol{F}^R_{i_{1,2}}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^R_{i_{2,3}}\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^I_{i_{1,3}}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}... \\
&+ \boldsymbol{F}^R_{i_{1,2}}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^I_{i_{2,3}}\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^R_{i_{1,3}}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,2}}\boldsymbol{x}... \\
&+ \boldsymbol{F}^I_{i_{1,2}}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^R_{i_{2,3}}\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^R_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^R_{i_{1,3}}\boldsymbol{F}^R_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}... \\
&- (\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^I_{i_{2,3}}\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,3}}\boldsymbol{x} + \boldsymbol{F}^I_{i_{1,3}}\boldsymbol{F}^I_{i_{2,3}}\boldsymbol{x}\boldsymbol{F}^I_{i_{1,2}}\boldsymbol{x})
\end{aligned} \tag{7}$$

1.2.2 Visibility Amplitudes

$$\frac{d}{d\boldsymbol{x}}\sqrt{(\boldsymbol{F}^R\boldsymbol{x})^2 + (\boldsymbol{F}^I\boldsymbol{x})^2} = \frac{d}{d\boldsymbol{x}}\sqrt{\boldsymbol{F}^R\boldsymbol{x}\boldsymbol{F}^R\boldsymbol{x} + \boldsymbol{F}^I\boldsymbol{x}\boldsymbol{F}^I\boldsymbol{x}} = \frac{(\boldsymbol{F}^{RT}(\boldsymbol{F}^R\boldsymbol{x}) + \boldsymbol{F}^{IT}(\boldsymbol{F}^I\boldsymbol{x}))}{\sqrt{(\boldsymbol{F}^R\boldsymbol{x})^2 + (\boldsymbol{F}^I\boldsymbol{x})^2}} \tag{8}$$

2 StarWarps Derivations

Given a set of observations and model parameters θ, we would like to estimate the marginal distribution for each of the N latent images. For ease of notation, we define:

$$\boldsymbol{y}_{a:b} = \{\boldsymbol{y}_a, ..., \boldsymbol{y}_b\} \tag{9}$$

$$\boldsymbol{y}_{a:b\setminus t} = \{\boldsymbol{y}_a, ..., \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t+1}, ..., \boldsymbol{y}_b\} \tag{10}$$

According to Equation 24 of the main text, likelihood of the the full joint model is written as:

$$p(\boldsymbol{x}_1, ..., \boldsymbol{x}_N | \boldsymbol{y}_1, ..., \boldsymbol{y}_N) \propto \prod_{t=1}^{N} \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \prod_{t=1}^{N} \mathcal{N}_{\boldsymbol{y}_t}(\boldsymbol{F}_t\boldsymbol{x}_t, \boldsymbol{R}_t) \prod_{t=2}^{N} \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q}) \tag{11}$$

2.1 Solving for the $E[\boldsymbol{x}_t]$ and $E[\boldsymbol{x}_t\boldsymbol{x}_t^T]$ Sufficient Statistics

We can compute $p(\boldsymbol{x}_t|\boldsymbol{y}_1, ..., \boldsymbol{y}_N)$ by marginalizing out the other latent images:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:N}) \propto \int_{\boldsymbol{x}_{1:N\setminus t}} p(\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N}) \quad d\boldsymbol{x}_1, ..., d\boldsymbol{x}_{1:N\setminus t} \tag{12}$$

$$\propto \int_{\boldsymbol{x}_{1:N\setminus t}} \prod_{t=1}^{N} \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \prod_{t=1}^{N} \mathcal{N}_{\boldsymbol{y}_t}(\boldsymbol{F}_t\boldsymbol{x}_t, \boldsymbol{R}_t) \prod_{t=2}^{N} \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q}) \quad d\boldsymbol{x}_{1:N\setminus t} \tag{13}$$

Using the Elimination Algorithm we can solve for this probability efficiently.

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:N}) \propto p(\boldsymbol{x}_t, \boldsymbol{y}_{1:N}) \tag{14}$$

$$\propto \phi_t(\boldsymbol{x}_t) m_{t-1\to t}(\boldsymbol{x}_t) m_{t+1\to t}(\boldsymbol{x}_t) \tag{15}$$

We define the unary and joint potential functions as follows:

$$\phi_t(\boldsymbol{x}_t) = p(\boldsymbol{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t) \tag{16}$$

$$= \mathcal{N}_{\boldsymbol{y}_t}(\boldsymbol{F}_t\boldsymbol{x}_t, \boldsymbol{R}_t)\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \tag{17}$$

Using Lemma 4.4

$$= \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \tag{18}$$

$$\Psi_{t-1,t}(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) \tag{19}$$

$$= \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q}) \tag{20}$$

Using these potential functions, the resulting sum-product messages are:

$$m_{1\to 2}(\boldsymbol{x}_2) = \int_{\boldsymbol{x}_1} \phi_1(\boldsymbol{x}_1)\Psi_{1,2}(\boldsymbol{x}_1, \boldsymbol{x}_2)d\boldsymbol{x}_1 \tag{21}$$

$$m_{t\to t+1}(\boldsymbol{x}_{t+1}) = \int_{\boldsymbol{x}_t} \phi_t(\boldsymbol{x}_t)\Psi_{t,t+1}(\boldsymbol{x}_t, \boldsymbol{x}_{t+1})m_{t-1\to t}(\boldsymbol{x}_t)d\boldsymbol{x}_t \qquad 2 \le t \le N \tag{22}$$

$$= \mathcal{N}_{\boldsymbol{x}_{t+1}}(\mathbf{z}^{\alpha}_{t+1|t}, \mathbf{P}^{\alpha}_{t+1|t}) = \int_{\boldsymbol{x}_t} \Psi_{t,t+1}(\boldsymbol{x}_t, \boldsymbol{x}_{t+1})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha}_{t|t}, \mathbf{P}^{\alpha}_{t|t})d\boldsymbol{x}_t \tag{23}$$

$$m_{N\to N-1}(\boldsymbol{x}_2) = \int_{\boldsymbol{x}_N} \phi_N(\boldsymbol{x}_N)\Psi_{N-1,N}(\boldsymbol{x}_{N-1}, \boldsymbol{x}_N)d\boldsymbol{x}_N \tag{24}$$

$$m_{t\to t-1}(\boldsymbol{x}_{t-1}) = \int_{\boldsymbol{x}_t} \phi_t(\boldsymbol{x}_t)\Psi_{t-1,t}(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t)m_{t+1\to t}(\boldsymbol{x}_t)d\boldsymbol{x}_t \qquad 1 \le t \le N-1 \tag{25}$$

$$= \mathcal{N}_{\boldsymbol{x}_{t-1}}(\mathbf{z}^{\beta}_{t-1|t}, \mathbf{P}^{\beta}_{t-1|t}) = \int_{\boldsymbol{x}_t} \Psi_{t-1,t}(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t)\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\beta}_{t|t}, \mathbf{P}^{\beta}_{t|t})d\boldsymbol{x}_t \tag{26}$$

2.1.1 Forward Messages

$$m_{t\to t+1}(\boldsymbol{x}_{t+1}) = \int_{\boldsymbol{x}_t} \Psi_{t,t+1}(\boldsymbol{x}_t, \boldsymbol{x}_{t+1})\phi_t(\boldsymbol{x}_t)m_{t-1\to t}(\boldsymbol{x}_t)d\boldsymbol{x}_t \tag{27}$$

$$= \int_{\boldsymbol{x}_t} N_{\boldsymbol{x}_{t+1}}(\boldsymbol{A}\boldsymbol{x}_t, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda})m_{t-1\to t}(\boldsymbol{x}_t)d\boldsymbol{x}_t \tag{28}$$

$$= \int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{x}_{t+1}}(\boldsymbol{A}\boldsymbol{x}_t, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha}_{t|t-1}, \mathbf{P}^{\alpha}_{t|t-1})d\boldsymbol{x}_t \tag{29}$$

Using Lemma 4.1: $\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha*}_{t|t-1}, \mathbf{P}^{\alpha*}_{t|t-1}) \propto \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha}_{t|t-1}, \mathbf{P}^{\alpha}_{t|t-1})$

where $\mathbf{z}^{\alpha*}_{1|0} = \boldsymbol{\mu}$, $\mathbf{P}^{\alpha*}_{1|0} = \boldsymbol{\Lambda}$

$$\propto \int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{x}_{t+1}}(\boldsymbol{A}\boldsymbol{x}_t, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha*}_{t|t-1}, \mathbf{P}^{\alpha*}_{t|t-1})d\boldsymbol{x}_t \tag{30}$$

Using Lemma 4.2: $\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha}_{t|t}, \mathbf{P}^{\alpha}_{t|t}) \propto \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha*}_{t|t-1}, \mathbf{P}^{\alpha*}_{t|t-1})$

$$\propto \int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}^{-1}\boldsymbol{x}_{t+1}, (\boldsymbol{A}^T\boldsymbol{Q}^{-1}\boldsymbol{A})^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha}_{t|t}, \mathbf{P}^{\alpha}_{t|t})d\boldsymbol{x}_t \tag{31}$$

$$= \mathcal{N}_{\boldsymbol{A}^{-1}\boldsymbol{x}_{t+1}}(\mathbf{z}^{\alpha}_{t|t}, (\boldsymbol{A}^T\boldsymbol{Q}^{-1}\boldsymbol{A})^{-1} + \mathbf{P}^{\alpha}_{t|t}) \int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{x}_t}(.,.)d\boldsymbol{x}_t \tag{32}$$

$$= \mathcal{N}_{\boldsymbol{A}^{-1}\boldsymbol{x}_{t+1}}(\mathbf{z}^{\alpha}_{t|t}, (\boldsymbol{A}^T\boldsymbol{Q}^{-1}\boldsymbol{A})^{-1} + \mathbf{P}^{\alpha}_{t|t}) \tag{33}$$

Using Lemma 4.6

$$= \mathcal{N}_{\boldsymbol{x}_{t+1}}(\boldsymbol{A}\mathbf{z}^{\alpha}_{t|t}, \boldsymbol{Q} + \boldsymbol{A}\mathbf{P}^{\alpha}_{t|t}\boldsymbol{A}^T) \tag{34}$$

$$= \mathcal{N}_{\boldsymbol{x}_{t+1}}(\mathbf{z}^{\alpha}_{t+1|t}, \mathbf{P}^{\alpha}_{t+1|t}) \tag{35}$$

2.1.2 Backward Messages

$$m_{t\to t-1}(\boldsymbol{x}_{t-1}) = \int_{\boldsymbol{x}_t} \Psi_{t-1,t}(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t)\phi_t(\boldsymbol{x}_t)m_{t+1\to t}(\boldsymbol{x}_t)d\boldsymbol{x}_t \tag{36}$$

$$= \int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda})m_{t+1\to t}(\boldsymbol{x}_{t-1})d\boldsymbol{x}_t \tag{37}$$

$$= \int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\beta}_{t|t+1}, \mathbf{P}^{\beta}_{t|t+1})d\boldsymbol{x}_t \tag{38}$$

Using Lemma 4.3: $\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\beta*}_{t|t+1}, \mathbf{P}^{\beta*}_{t|t+1}) \propto \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\beta}_{t|t+1}, \mathbf{P}^{\beta}_{t|t+1})$

where $\mathbf{z}^{\beta*}_{N|N+1} = \boldsymbol{\mu}$, $\mathbf{P}^{\beta*}_{N|N+1} = \boldsymbol{\Lambda}$

$$\propto \int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\beta*}_{t|t+1}, \mathbf{P}^{\beta*}_{t|t+1})d\boldsymbol{x}_t \tag{39}$$

Using Lemma 4.2: $\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\beta}_{t|t}, \mathbf{P}^{\beta}_{t|t}) \propto \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\beta*}_{t|t+1}, \mathbf{P}^{\beta*}_{t|t+1})$

$$\propto \int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\beta}_{t|t}, \mathbf{P}^{\beta}_{t|t})d\boldsymbol{x}_t \tag{40}$$

$$= \int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{A}\boldsymbol{x}_{t-1}}(\mathbf{z}^{\beta}_{t|t}, \boldsymbol{Q} + \mathbf{P}^{\beta}_{t|t})\mathcal{N}_{\boldsymbol{x}_t}(.,.)d\boldsymbol{x}_t \tag{41}$$

$$= C\exp\left[\frac{-1}{2}(\boldsymbol{A}\boldsymbol{x}_{t-1} - \mathbf{z}^{\beta}_{t|t})^T(\boldsymbol{Q} + \mathbf{P}^{\beta}_{t|t})^{-1}(\boldsymbol{A}\boldsymbol{x}_{t-1} - \mathbf{z}^{\beta}_{t|t})\right]\int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{x}_t}(.,.)d\boldsymbol{x}_t \tag{42}$$

$$= C\exp\left[\frac{-1}{2}(\boldsymbol{x}_{t-1} - \boldsymbol{A}^{-1}\mathbf{z}^{\beta}_{t|t})^T\boldsymbol{A}^T(\boldsymbol{Q} + \mathbf{P}^{\beta}_{t|t})^{-1}\boldsymbol{A}(\boldsymbol{x}_{t+1} - \boldsymbol{A}^{-1}\mathbf{z}^{\beta}_{t|t})\right] \tag{43}$$

$$= \mathcal{N}_{\boldsymbol{x}_{t-1}}(\boldsymbol{A}^{-1}\mathbf{z}^{\beta}_{t|t}, (\boldsymbol{A}^T(\boldsymbol{Q} + \mathbf{P}^{\beta}_{t|t})^{-1}\boldsymbol{A})^{-1}) \tag{44}$$

$$= \mathcal{N}_{\boldsymbol{x}_{t-1}}(\mathbf{z}^{\beta}_{t-1|t}, \mathbf{P}^{\beta}_{t-1|t}) \tag{45}$$

2.1.3 Putting it Together

$$
\begin{aligned}
p(\boldsymbol{x}_t|\boldsymbol{y}_{1:N}) &\propto p(\boldsymbol{x}_t, \boldsymbol{y}_{1:N}) && (46)\\
&\propto m_{t-1\to t}(\boldsymbol{x}_t) m_{t+1\to t}(\boldsymbol{x}_t)\phi_t(\boldsymbol{x}_t) && (47)\\
&= m_{t-1\to t}(\boldsymbol{x}_t) m_{t+1\to t}(\boldsymbol{x}_t)\mathcal{N}_{\boldsymbol{y}_t}(\boldsymbol{F}_t\boldsymbol{x}_t, \boldsymbol{R}_t)\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) && (48)\\
&= m_{t-1\to t}(\boldsymbol{x}_t) m_{t+1\to t}(\boldsymbol{x}_t)\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) && (49)\\
&= m_{t-1\to t}(\boldsymbol{x}_t)\left[\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) m_{t+1\to t}(\boldsymbol{x}_t)\right] && (50)\\
&\propto \mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha}_{t|t-1}, \mathbf{P}^{\alpha}_{t|t-1})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\beta}_{t|t}, \mathbf{P}^{\beta}_{t|t})) && (51)\\
&\propto \mathcal{N}_{\boldsymbol{x}_t}(\hat{\boldsymbol{x}}_t, \mathbf{C}_t) && (52)
\end{aligned}
$$

where using Lemma 4.1

$$
\begin{aligned}
\hat{\boldsymbol{x}}_t &= \mathbf{P}^{\beta}_{t|t}(\mathbf{P}^{\alpha}_{t|t-1} + \mathbf{P}^{\beta}_{t|t})^{-1}\mathbf{z}^{\alpha}_{t|t-1} + \mathbf{P}^{\alpha}_{t|t-1}(\mathbf{P}^{\alpha}_{t|t-1} + \mathbf{P}^{\beta}_{t|t})^{-1}\mathbf{z}^{\beta}_{t|t} && (53)\\
\mathbf{C}_t &= \mathbf{P}^{\alpha}_{t|t-1}(\mathbf{P}^{\alpha}_{t|t-1} + \mathbf{P}^{\beta}_{t|t})^{-1}\mathbf{P}^{\beta}_{t|t} && (54)
\end{aligned}
$$

Note that for $t = 1$ $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:N}) \propto \mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\beta}_{t|t}, \mathbf{P}^{\beta}_{t|t}))$. Note that $E[\boldsymbol{x}_t] = \mathbf{z}$. The statistic $E[\boldsymbol{x}_t\boldsymbol{x}_t^T]$ can be also solved using these terms:

$$
\begin{aligned}
E_{\boldsymbol{x}_t|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[\boldsymbol{x}_t\boldsymbol{x}_t^T|\boldsymbol{y}_{1:N}\right] &= \int_{\boldsymbol{x}_{1:N}} \boldsymbol{x}_t\boldsymbol{x}_t^T p(\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N})d\boldsymbol{x}_{1:N} && (55)\\
&= \int_{\boldsymbol{x}_t} \boldsymbol{x}_t\boldsymbol{x}_t^T \int_{\boldsymbol{x}_{1:N\setminus t}} p(\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N})d\boldsymbol{x}_{1:N\setminus t}d\boldsymbol{x}_t && (56)\\
&= \int_{\boldsymbol{x}_t} \boldsymbol{x}_t\boldsymbol{x}_t^T p(\boldsymbol{x}_t|\boldsymbol{y}_{1:N})d\boldsymbol{x}_t && (57)\\
&= \int_{\boldsymbol{x}_t} \boldsymbol{x}_t\boldsymbol{x}_t^T \mathcal{N}_{\boldsymbol{x}_t}(\hat{\boldsymbol{x}}_t, \mathbf{C}_t)d\boldsymbol{x}_t && (58)\\
&= \hat{\boldsymbol{x}}_t\hat{\boldsymbol{x}}_t^T + \mathbf{C}_t && (59)
\end{aligned}
$$

2.2 Solving for the $E[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T]$ Sufficient Statistic

$$
\begin{aligned}
p(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:N}) &\propto \int_{\boldsymbol{x}_{1:N\setminus t,t-1}} p(\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N}) \quad d\boldsymbol{x}_{1:N\setminus t,t-1} && (60)\\
&\propto \int_{\boldsymbol{x}_{1:N\setminus t,t-1}} \prod_{t=1}^{N}\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda})\prod_{t=1}^{N}\mathcal{N}_{\boldsymbol{y}_t}(\boldsymbol{F}_t\boldsymbol{x}_t, \boldsymbol{R}_t)\prod_{t=2}^{N}\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q}) \quad d\boldsymbol{x}_{1:N\setminus t,t-1} && (61)
\end{aligned}
$$

Using the Elimination Algorithm we can solve for this probability efficiently.

$$
\begin{aligned}
p(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:N}) \propto\, & p(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \boldsymbol{y}_{1:N}) && (62)\\
&\propto m_{t-2\to t-1}\phi_{t-1}(\boldsymbol{x}_{t-1})\Psi_{t-1,t}(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t)\phi_t(\boldsymbol{x}_t)(\boldsymbol{x}_{t-1})m_{t+1\to t}(\boldsymbol{x}_t) && (63)\\
&= m_{t-2\to t-1}\mathcal{N}_{\boldsymbol{x}_{t-1}}(\boldsymbol{F}_{t-1}^{-1}\boldsymbol{y}_{t-1}, (\boldsymbol{F}_{t-1}^T\boldsymbol{R}_{t-1}^{-1}\boldsymbol{F}_{t-1})^{-1})\mathcal{N}_{\boldsymbol{x}_{t-1}}(\boldsymbol{\mu}, \boldsymbol{\Lambda})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q})\\
&\qquad \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda})(\boldsymbol{x}_{t-1})m_{t+1\to t}(\boldsymbol{x}_t) && (64)\\
&= \mathcal{N}_{\boldsymbol{x}_{t-1}}(\mathbf{z}^\alpha_{t-1|t-1}, \mathbf{P}^\alpha_{t-1|t-1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^\beta_{t|t}, \mathbf{P}^\beta_{t|t}) && (65)\\
&= \mathcal{N}_{\boldsymbol{x}_{t-1}}(\mathbf{z}^\alpha_{t-1|t-1}, \mathbf{P}^\alpha_{t-1|t-1})\mathcal{N}_{\boldsymbol{A}\boldsymbol{x}_{t-1}}(\mathbf{z}^\beta_{t|t}, \boldsymbol{Q} + \mathbf{P}^\beta_{t|t}) && (66)\\
&\quad \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t} + \mathbf{P}^\beta_{t|t}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t}) && (67)\\
&= \mathcal{N}_{\boldsymbol{x}_{t-1}}(\mathbf{z}^\alpha_{t-1|t-1}, \mathbf{P}^\alpha_{t-1|t-1})\mathcal{N}_{\boldsymbol{x}_{t-1}}(\boldsymbol{A}^{-1}\mathbf{z}^\beta_{t|t}, (\boldsymbol{A}^T(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\boldsymbol{A})^{-1}) && (68)\\
&\quad \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t} + \mathbf{P}^\beta_{t|t}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t}) && (69)\\
&= K_t\mathcal{N}_{\boldsymbol{x}_{t-1}}(m_{t-1}, C_{t-1})\\
&\quad \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t} + \mathbf{P}^\beta_{t|t}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t}) && (70)\\
&= K_t\mathcal{N}_{\boldsymbol{x}_{t-1}}(m_{t-1}, C_{t-1})\\
&\quad \mathcal{N}_{\mathbf{P}^\beta_{t|t}(\boldsymbol{Q}+\mathbf{P}^\beta_{t|t})^{-1}\boldsymbol{A}\boldsymbol{x}_{t-1}}(\boldsymbol{x}_t - \boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t}, \boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t}) && (71)\\
&= K_t\mathcal{N}_{\boldsymbol{x}_{t-1}}(m_{t-1}, C_{t-1})\\
&\quad \mathcal{N}_{\boldsymbol{x}_{t-1}}(M^{-1}(\boldsymbol{x}_t - \boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t}), ((M^T(\boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t})^{-1}M)^{-1}) && (72)\\
&= K_t\mathcal{N}_{\boldsymbol{x}_{t-1}}(m_{t-1} + C_{t-1}M^T(\boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t} + MC_{t-1}M^T)^{-1}\\
&\qquad (\boldsymbol{x}_t - \boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t} - Mm_{t-1}), && (73)\\
&\qquad C_{t-1} - C_{t-1}M^T(\boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t} + MC_{t-1}M^T)^{-1}MC_{t-1}) && (74)\\
&= N_{\boldsymbol{x}_{t-1}}(D\boldsymbol{x}_t + G, \boldsymbol{F}) && (75)
\end{aligned}
$$

$$
\begin{aligned}
M =& \mathbf{P}^\beta_{t|t}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\boldsymbol{A} && (76)\\
m_{t-1} =& \mathbf{z}^\alpha_{t-1|t-1} + \mathbf{P}^\alpha_{t-1|t-1}\boldsymbol{A}^T(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t} + \boldsymbol{A}\mathbf{P}^\alpha_{t-1|t-1}\boldsymbol{A}^T)^{-1}(\mathbf{z}^\beta_{t|t} - \boldsymbol{A}\mathbf{z}^\alpha_{t-1|t-1}) && (77)\\
C_{t-1} =& \mathbf{P}^\alpha_{t-1|t-1} - \mathbf{P}^\alpha_{t-1|t-1}\boldsymbol{A}^T(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t} + \boldsymbol{A}\mathbf{P}^\alpha_{t-1|t-1})\boldsymbol{A}^T)^{-1}\boldsymbol{A}\mathbf{P}^\alpha_{t-1|t-1} && (78)\\
D =& C_{t-1}M^T(\boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t} + MC_{t-1}M^T)^{-1} && (79)\\
G =& m_{t-1} + D(-\boldsymbol{Q}(\boldsymbol{Q} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t} - Mm_{t-1}) && (80)\\
F =& C_{t-1} - DMC_{t-1} && (81)
\end{aligned}
$$

Using the following relations we can then compute $E[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T]$:

$$
E[y(A\boldsymbol{x} + B)^T] = E[yx^TA^T + yB^T] = E[yx^T]A^T + E[y]B^T \qquad (82)
$$

$$
\begin{aligned}
F + E[\boldsymbol{x}_{t-1}]E[D\boldsymbol{x}_t + G]^T =& E[\boldsymbol{x}_{t-1}(D\boldsymbol{x}_t + G)^T] = E[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T]D^T + E[\boldsymbol{x}_{t-1}]G^T && (83)\\
E[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T]D^T =& (F + E[\boldsymbol{x}_{t-1}]E[D\boldsymbol{x}_t + G]^T - E[\boldsymbol{x}_{t-1}]G^T) && (84)\\
E[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T] =& (F + E[\boldsymbol{x}_{t-1}]E[D\boldsymbol{x}_t + G]^T - E[\boldsymbol{x}_{t-1}]G^T)D^{T-1} && (85)\\
E[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T] =& (F + E[\boldsymbol{x}_{t-1}]E[\boldsymbol{x}_t]^T D^T + E[\boldsymbol{x}_{t-1}]G^T - E[\boldsymbol{x}_{t-1}]G^T)D^{T-1} && (86)\\
E[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T] =& (F + E[\boldsymbol{x}_{t-1}]E[\boldsymbol{x}_t]^T D^T)D^{T-1} && (87)\\
E[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T] =& FD^{T-1} + E[\boldsymbol{x}_{t-1}]E[\boldsymbol{x}_t]^T && (88)
\end{aligned}
$$

2.3 Likelihood

$$
\begin{aligned}
p(\boldsymbol{y}_{1:N}) &= \int_{\boldsymbol{x}_{1:N}} p(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N}) \quad d\boldsymbol{x}_{1:N} && (89)\\
&= \int_{\boldsymbol{x}_{1:N}} \prod_{t=1}^{N} \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \prod_{t=1}^{N} \mathcal{N}_{\boldsymbol{y}_t}(\boldsymbol{F}_t\boldsymbol{x}_t, \boldsymbol{R}_t) \prod_{t=2}^{N} \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q}) \quad d\boldsymbol{x}_{1:N} && (90)
\end{aligned}
$$

Note that this can be solved with message passing. This is very similar to the calculations done for the forward message passing. The only difference is that it also keeps tracks of scaling coefficients.

$$
m_{t\to t+1}(\boldsymbol{x}_{t+1}) = \int_{\boldsymbol{x}_t} \Psi_{t,t+1}(\boldsymbol{x}_t, \boldsymbol{x}_{t+1})\phi_t(\boldsymbol{x}_t)m_{t-1\to t}(\boldsymbol{x}_t)d\boldsymbol{x}_t \tag{91}
$$

$$
= \int_{\boldsymbol{x}_t} N_{\boldsymbol{x}_{t+1}}(\boldsymbol{A}\boldsymbol{x}_t, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda})m_{t-1\to t}(\boldsymbol{x}_t)d\boldsymbol{x}_t \tag{92}
$$

$$
= \int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{x}_{t+1}}(\boldsymbol{A}\boldsymbol{x}_t, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda})\ell_t\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha}_{t|t-1}, \mathbf{P}^{\alpha}_{t|t-1})d\boldsymbol{x}_t \tag{93}
$$

Using Lemma 4.1 and Matrix Cookbook [1]:

$\mathcal{N}_{\boldsymbol{\mu}}(\mathbf{z}^{\alpha}_{t|t-1}, \boldsymbol{\Lambda} + \mathbf{P}^{\alpha}_{t|t-1})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha*}_{t|t-1}, \mathbf{P}^{\alpha*}_{t|t-1}) = \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha}_{t|t-1}, \mathbf{P}^{\alpha}_{t|t-1})$

where $\mathbf{z}^{\alpha*}_{1|0} = \boldsymbol{\mu}$, $\mathbf{P}^{\alpha*}_{1|0} = \boldsymbol{\Lambda}$

$$
= \ell_t\mathcal{N}_{\boldsymbol{\mu}}(\mathbf{z}^{\alpha}_{t|t-1}, \boldsymbol{\Lambda} + \mathbf{P}^{\alpha}_{t|t-1}) \int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{x}_{t+1}}(\boldsymbol{A}\boldsymbol{x}_t, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha*}_{t|t-1}, \mathbf{P}^{\alpha*}_{t|t-1})d\boldsymbol{x}_t \tag{94}
$$

Using Matrix Cookbook [1], Lemma 4.2, and Lemma 4.6:

$$
\begin{aligned}
&\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{F}_t^{-1}\boldsymbol{y}_t, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha*}_{t|t-1}, \mathbf{P}^{\alpha*}_{t|t-1})\\
&= \mathcal{N}_{\boldsymbol{F}_t^{-1}\boldsymbol{y}_t}(\mathbf{z}^{\alpha*}_{t|t-1}, (\boldsymbol{F}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{F}_t)^{-1} + \mathbf{P}^{\alpha*}_{t|t-1})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha}_{t|t}, \mathbf{P}^{\alpha}_{t|t})\\
&= \mathcal{N}_{\boldsymbol{y}_t}(\boldsymbol{F}_t\mathbf{z}^{\alpha*}_{t|t-1}, \boldsymbol{R}_t + \boldsymbol{F}_t\mathbf{P}^{\alpha*}_{t|t-1}\boldsymbol{F}_t^T)\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha}_{t|t}, \mathbf{P}^{\alpha}_{t|t})
\end{aligned}
$$

$$
= \ell_t\mathcal{N}_{\boldsymbol{\mu}}(\mathbf{z}^{\alpha}_{t|t-1}, \boldsymbol{\Lambda} + \mathbf{P}^{\alpha}_{t|t-1})\mathcal{N}_{\boldsymbol{y}_t}(\boldsymbol{F}_t\mathbf{z}^{\alpha*}_{t|t-1}, \boldsymbol{R}_t + \boldsymbol{F}_t\mathbf{P}^{\alpha}_{t|t-1}\boldsymbol{F}_t^T) \int_{\boldsymbol{x}_t} \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}^{-1}\boldsymbol{x}_{t+1}, (\boldsymbol{A}^T\boldsymbol{Q}^{-1}\boldsymbol{A})^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^{\alpha}_{t|t}, \mathbf{P}^{\alpha}_{t|t})d\boldsymbol{x}_t \tag{95}
$$

$$
= \ell_t\mathcal{N}_{\boldsymbol{\mu}}(\mathbf{z}^{\alpha}_{t|t-1}, \boldsymbol{\Lambda} + \mathbf{P}^{\alpha}_{t|t-1})\mathcal{N}_{\boldsymbol{y}_t}(\boldsymbol{F}\mathbf{z}^{\alpha*}_{t|t-1}, \boldsymbol{R}_t + \boldsymbol{F}_t\mathbf{P}^{\alpha}_{t|t-1}\boldsymbol{F}^T)\mathcal{N}_{\boldsymbol{x}_{t+1}}(\mathbf{z}^{\alpha}_{t+1|t}, \mathbf{P}^{\alpha}_{t+1|t}) \tag{96}
$$

$$
= \ell_{t+1}\mathcal{N}_{\boldsymbol{x}_{t+1}}(\mathbf{z}^{\alpha}_{t+1|t}, \mathbf{P}^{\alpha}_{t+1|t}) \tag{97}
$$

2.4 Non-linear Observation Model

Now, let's say that our observation model is no longer linear, but instead is a non-linear function of x.

$$
\boldsymbol{y}_t \sim \mathcal{N}(f(\boldsymbol{x}_t), \boldsymbol{R}) \tag{100}
$$

Algorithm 1: Computing Likelihood: $t = 1 \to 2 \to ... \to N$
$$\ell_{t+1} = \ell_t \mathcal{N}_{\boldsymbol{\mu}}(\mathbf{z}^{\alpha}_{t\|t-1}, \boldsymbol{\Lambda} + \mathbf{P}^{\alpha}_{t\|t-1}) \mathcal{N}_{\boldsymbol{y}_t}(\boldsymbol{F}_t \mathbf{z}^{\alpha *}_{t\|t-1}, \boldsymbol{R}_t + \boldsymbol{F}_t \mathbf{P}^{\alpha *}_{t\|t-1} \boldsymbol{F}_t^T) \tag{98}$$ **Initialization:** $$\ell_0 = 1 \ , \ \mathcal{N}_{\boldsymbol{\mu}}(\mathbf{z}^{\alpha}_{1\|0}, \boldsymbol{\Lambda} + \mathbf{P}^{\alpha}_{1\|0}) = 1 \tag{99}$$

Using a first order Taylor Series Expansion Approximation, we can approximate $\mathcal{N}_{\boldsymbol{y}}(f(\boldsymbol{x}), \boldsymbol{R})$ as

$$\begin{aligned}
\mathcal{N}_{\boldsymbol{y}}(f(\boldsymbol{x}), \boldsymbol{R}) &\propto \exp\left[\frac{-1}{2}(\boldsymbol{y} - f(\boldsymbol{x}))^T \boldsymbol{R}^{-1} (\boldsymbol{y} - f(\boldsymbol{x}))\right] \qquad (101)\\
&\approx \exp\left[\frac{-1}{2}(\boldsymbol{y} - (f(\tilde{\boldsymbol{x}}) + \dot{f}(\tilde{\boldsymbol{x}})(\boldsymbol{x} - \tilde{\boldsymbol{x}})))^T \boldsymbol{R}^{-1} (\boldsymbol{y} - (f(\tilde{\boldsymbol{x}}) + \dot{f}(\tilde{\boldsymbol{x}})(\boldsymbol{x} - \tilde{\boldsymbol{x}})))\right] \\
&= \exp\left[\frac{-1}{2}(\boldsymbol{y} - f(\tilde{\boldsymbol{x}}) - \dot{f}(\tilde{\boldsymbol{x}})\boldsymbol{x} + \dot{f}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}})^T \boldsymbol{R}^{-1} (\boldsymbol{y} - f(\tilde{\boldsymbol{x}}) - \dot{f}(\tilde{\boldsymbol{x}})\boldsymbol{x} + \dot{f}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}})\right] \\
&= \exp\left[\frac{-1}{2}(-1 \times (\dot{f}(\tilde{\boldsymbol{x}})\boldsymbol{x} - (-f(\tilde{\boldsymbol{x}}) + \dot{f}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}} + \boldsymbol{y})))^T \boldsymbol{R}^{-1} (-1 \times (\dot{f}(\tilde{\boldsymbol{x}})\boldsymbol{x} - (-f(\tilde{\boldsymbol{x}}) + \dot{f}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}} + \boldsymbol{y})))\right] \\
&= \exp\left[\frac{-1}{2}(\dot{f}(\tilde{\boldsymbol{x}})\boldsymbol{x} - (-f(\tilde{\boldsymbol{x}}) + \dot{f}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}} + \boldsymbol{y}))^T \boldsymbol{R}^{-1} (\dot{f}(\tilde{\boldsymbol{x}})\boldsymbol{x} - (-f(\tilde{\boldsymbol{x}}) + \dot{f}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}} + \boldsymbol{y}))\right] \\
&= \exp\left[\frac{-1}{2}(\boldsymbol{x} - \dot{f}(\tilde{\boldsymbol{x}})^{-1}(-f(\tilde{\boldsymbol{x}}) + \dot{f}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}} + \boldsymbol{y}))^T \dot{f}(\tilde{\boldsymbol{x}})^T \boldsymbol{R}^{-1} \dot{f}(\tilde{\boldsymbol{x}}) (\boldsymbol{x} - \dot{f}(\tilde{\boldsymbol{x}})^{-1}(-f(\tilde{\boldsymbol{x}}) + \dot{f}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}} + \boldsymbol{y}))\right] \\
&\propto \mathcal{N}_{\boldsymbol{x}}(\dot{f}(\tilde{\boldsymbol{x}})^{-1}(\boldsymbol{y} + \dot{f}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}} - f(\tilde{\boldsymbol{x}})), (\dot{f}(\tilde{\boldsymbol{x}})^T \boldsymbol{R}^{-1} \dot{f}(\tilde{\boldsymbol{x}}))^{-1}) \qquad (102)
\end{aligned}$$

Updates are very similar to in the linear model. We simply need to replace

$$\boldsymbol{F}_t \to \dot{f}_t(\tilde{\boldsymbol{x}}_t) \tag{103}$$

$$\boldsymbol{y}_t \to \boldsymbol{y}_t + \dot{f}_t(\tilde{\boldsymbol{x}}_t)\tilde{\boldsymbol{x}}_t - f_t(\tilde{\boldsymbol{x}}_t) \tag{104}$$

where $\tilde{\boldsymbol{x}}_t$ is the current estimate of $\boldsymbol{x}_t$

2.5 Estimating the Evolution Parameters

Calculate the expected value of $\log \Pi_t P(\boldsymbol{y}_t|\boldsymbol{x}_t;\theta)$ with respect to the distribution of x

$$\begin{aligned}
Q(\theta|\theta^{(i)}) &= E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[\log P(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N};\theta)\right] \qquad (105)\\
&= E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[\frac{-1}{2}\sum_{t=2}^{N}\left[(\boldsymbol{x}_t - \boldsymbol{A}_{\theta^{(i)}}\boldsymbol{x}_{t-1})^T \boldsymbol{Q}^{-1}(\boldsymbol{x}_t - \boldsymbol{A}_{\theta^{(i)}}\boldsymbol{x}_{t-1})\right] + \mathcal{G}(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N})\right] \\
&= E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[\frac{-1}{2}\sum_{t=2}^{N}\left[\boldsymbol{x}_t^T \boldsymbol{Q}^{-1}\boldsymbol{x}_t - \boldsymbol{x}_t^T \boldsymbol{Q}^{-1}\boldsymbol{A}_{\theta^{(i)}}\boldsymbol{x}_{t-1} - \boldsymbol{x}_{t-1}^T \boldsymbol{A}_{\theta^{(i)}}^T \boldsymbol{Q}^{-1}\boldsymbol{x}_t + \boldsymbol{x}_{t-1}^T \boldsymbol{A}_{\theta^{(i)}}^T \boldsymbol{Q}^{-1}\boldsymbol{A}_{\theta^{(i)}}\boldsymbol{x}_{t-1}\right] + \mathcal{G}(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N})\right] \\
&= \frac{-1}{2}\sum_{t=2}^{N}\Big[E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[\boldsymbol{x}_t^T \boldsymbol{Q}^{-1}\boldsymbol{x}_t\right] - E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[\boldsymbol{x}_t^T \boldsymbol{Q}^{-1}\boldsymbol{A}_{\theta^{(i)}}\boldsymbol{x}_{t-1}\right] - E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[\boldsymbol{x}_{t-1}^T \boldsymbol{A}_{\theta^{(i)}}^T \boldsymbol{Q}^{-1}\boldsymbol{x}_t\right] \\
&\qquad + E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[\boldsymbol{x}_{t-1}^T \boldsymbol{A}_{\theta^{(i)}}^T \boldsymbol{Q}^{-1}\boldsymbol{A}_{\theta^{(i)}}\boldsymbol{x}_{t-1}\right] + E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[\mathcal{G}(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N})\right]\Big] \qquad (106)
\end{aligned}$$

Where $\mathcal{G}(\boldsymbol{x}_{1:N}, \boldsymbol{y}_{1:N})$ is a term that incorporates the prior for $\boldsymbol{x}$ as well as the terms to normalize each of the distributions. Next, we must find the $\boldsymbol{A}$ that maximizes $Q(\theta|\theta^{(i)})$. Using the Matrix Cookbook (70) and (88) [1]

$$\begin{aligned}
\frac{d}{d\boldsymbol{A}_{\theta^{(i)}}} Q(\theta|\theta^{(i)}) &= \frac{1}{2}\sum_{t=2}^{N}\Big[E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[(\boldsymbol{x}_t^T\boldsymbol{Q}^{-1})^T\boldsymbol{x}_{t-1}^T\right] + E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[(\boldsymbol{x}_{t-1}^T\boldsymbol{Q}^{-1})^T\boldsymbol{x}_t^T\right] \\
&\qquad - E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[2\boldsymbol{Q}^{-1}\boldsymbol{A}_{\theta^{(i)}}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}^T\right]\Big] \\
&= \frac{1}{2}\sum_{t=2}^{N}\Big[\boldsymbol{Q}^{-1}E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[\boldsymbol{x}_t\boldsymbol{x}_{t-1}^T\right] + \boldsymbol{Q}^{-1}E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T\right] - 2\boldsymbol{Q}^{-1}\boldsymbol{A}_{\theta^{(i)}}E_{\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\theta^{(i)}}\left[\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}^T\right]\Big] \\
&= \sum_{t=2}^{N}\left[\boldsymbol{Q}^{-1}\left(\frac{1}{2}M_{t,t-1} + \frac{1}{2}M_{t-1,t}\right) - \boldsymbol{Q}^{-1}\boldsymbol{A}_{\theta^{(i)}}M_{t-1}\right]
\end{aligned} \tag{107}$$

$$\boldsymbol{A}_{\theta^{(i+1)}} = \left[\sum_{t=2}^{N}\frac{1}{2}M_{t,t-1} + \frac{1}{2}M_{t-1,t} - Bz_t^T\right]\left[\sum_{t=2}^{N}M_{t-1}\right]^{-1} \tag{108}$$

Now Let's say that $\boldsymbol{A}$ is a function of θ and we would like to solve for the best θ. To do this we use the Chain rule. From the Matrix Cookbook [1] we know

$$\frac{dQ}{d\theta_j} = \sum_p\sum_q \frac{dQ}{d\boldsymbol{A}_{p,q}}\frac{d\boldsymbol{A}_{p,q}}{d\theta_j} \tag{109}$$

$$\frac{d}{d\theta_j^{(i)}}Q(\theta|\theta^{(i)}) = \sum_p\sum_q \frac{d}{d\boldsymbol{A}_{p,q}}Q(\theta|\theta^{(i)})\frac{d\boldsymbol{A}_{p,q}}{d\theta_j^{(i)}} \tag{110}$$

$$= \sum_p\sum_q\left[\sum_{t=2}^{N}\left[\boldsymbol{Q}^{-1}\left(\frac{1}{2}M_{t,t-1} + \frac{1}{2}M_{t-1,t}\right) - \boldsymbol{Q}^{-1}\boldsymbol{A}_{\theta^{(i)}}M_{t-1} - \boldsymbol{Q}^{-1}Bz_t^T\right]\right]_{p,q}\left[\frac{d\boldsymbol{A}(\theta^{(i)})}{d\theta_j^{(i)}}\right]_{p,q} \tag{111}$$

3 Additional Figures

3.1 Time Variability in Data Products

VLBI data taken from an evolving source is significantly different from that of a static source. For example, Figure S.1 shows a simulated visibility amplitude and closure phase expected for a source evolving over time, and compares it to the data products expected if the same source were static. As the data can no longer be explained by a single image with similar structure, most static imaging methods break down on this data and are unable to produce accurate results.

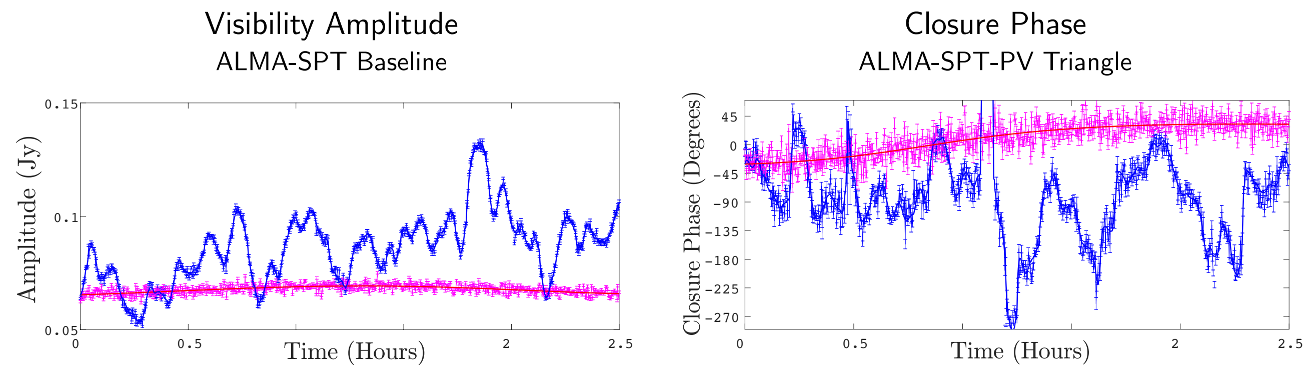

Figure S.1: **Simulated data under a static vs. varying source:** Contrasting of data observed from a static emission region (magenta) to that of a varying emission region (blue) over the course of 2.5 hours. Although both sequences start with the same image, the visibility amplitude and closure phase both begin to deviate from the static image very quickly. The ideal observation for the static and time-varying source is shown by the solid red and blue lines, respectively. We also show sample measurements with their respective error bars in the same colors. This data is simulated at a sampling interval of 11 seconds using the EHT2017 array from the frames in Video 3 presented in the paper. This figure shows 1 of the $\binom{K}{2}$ possible visibility amplitudes and 1 of $\binom{K}{3}$ possible closure phases (phase of the bispectrum).

3.2 Static Imaging under a Gaussian Prior

We compare results of the image reconstruction method presented in Section IV of the paper to other state-of-the-art methods for a static source in Figure S.2. The figure demonstrates that, although the presented approach of using a Gaussian model does not outperform other state-of-the-art static imaging methods, reasonable results are achieved despite a simpler image regularizer and optimization procedure.

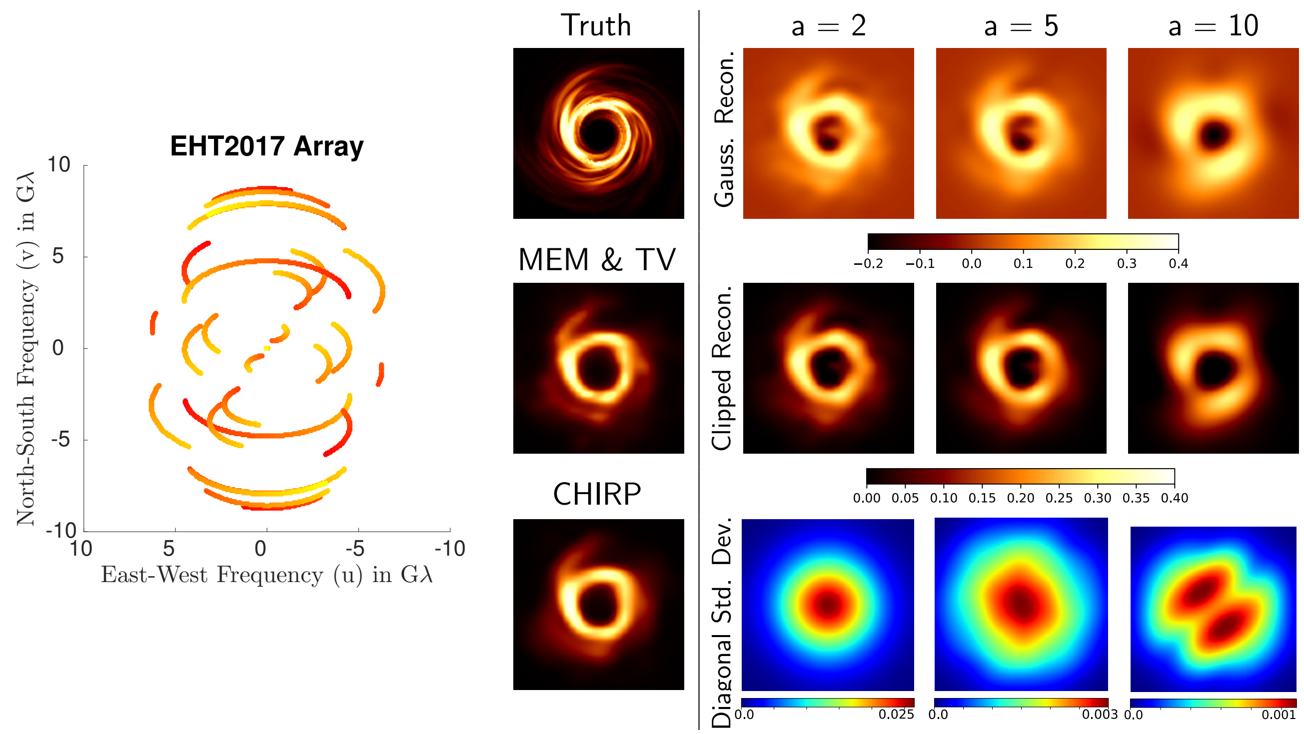

Figure S.2: **Static Imaging Comparison:** Results of static imaging using a multivariate Gaussian prior ($\mathsf{a} = 2, 5, 10$) compared to state-of-the-art reconstruction methods using MEM & TV regularizers[4] as well as patch-based regularizers (CHIRP)[5]. All images are shown with a field of view of 160 μ-arcseconds. Data is generated using a static image with the uv-coverage of the EHT2017 array shown on the left. The uv-coverage is colored by time, as indicated by the colorbar in the paper's uv-coverage plots. Note however that in this static imaging case the time of measurements is not relevant. Although the previous algorithms (MEM & TV and CHIRP) both produce better results, the Gaussian reconstruction is able to correctly get the broad structure of the underlying image. Since we do not impose positivity, negative values are reconstructed. However, by clipping the resulting image we can see that the result aligns well with the true static image. The Gaussian prior model also allows us to easily estimate our reconstructed image uncertainty. We visualize the diagonal entries of the posterior covariance matrix as the reshaped standard deviation image. Note that as the smoothness parameter a is increased, the per-pixel standard deviation becomes smaller, but the structure of the standard deviation deviates from what was specified in the prior (recall $\boldsymbol{\Lambda}$ is scaled by $\boldsymbol{\mu}$, which we have specified as a 2D Gaussian in this work). For large a the uncertainty is shown to be primarily in the diagonal north-west to south-east direction, due to the lack of spatial frequencies sampled by the telescope array in this direction. To avoid approximations and best show the recovered posterior covariance matrices, atmospheric error has not been included in the data used to recover these images. The scaling of the colormaps is in mili-Jansky per squared μ-arcsecond.

3.3 StarWarps Reconstructions

Figures S.3 and S.4 compare results obtained when we assume no global motion ($A = \mathbb{1}$) to those when we allow the method to search for a persistent warp field. Results are shown in two settings: when data is generated using the EHT2017+ array assuming no atmospheric phase error (VIS), as well as when phase errors are introduced (AMP & BISP) into the measurements. At each time, only a small number of measurements are observed (indicated by the corresponding uv-coverage). However, by propagating information across the video we are able to reconstruct good quality images at each time step. In the case of large global motion, most of the motion is suppressed in the resulting reconstruction when we assume $A = \mathbb{1}$. However, by solving for the low dimensional parameters of the warp field, θ, we can estimate the warp field and produce higher quality videos, while also inferring the underlying dynamics of the source. In Video 1 the true underlying motion of the emission region perfectly fits the affine model we assume. This allows us to freely recover a very similar warp field. However, in the "hot spot" video (Video 2), although this is no longer the case, we still recover an accurate estimate indicating the direction of motion.

Results of our method are compared to that of a simple baseline method that we refer to as 'snapshot imaging'. In snapshot imaging each frame of the video is independently reconstructed using only the small number of measurements taken at that time step. In particular, we use the MEM & TV method shown in Figure S.2 to reconstruct each snapshot. In the case of using complex visibilities, both our method and snapshot imaging produce meaningful results. Although distinctive features of the true underlying image are recovered by both methods, the quality of our StarWarps reconstructions is higher. However, in the case of data containing atmospheric phase errors our method shows substantial improvement over snapshot imaging. As the closure phase and bispectrum are invariant to the absolute position of the source, each snapshot reconstruction produces an image that is shifted by a different amount. This makes it challenging to align the snapshot frames to pull out meaningful structure in the reconstructed video when there is sparse uv coverage. For this reason our method substantially outperforms snapshot imaging.

In Video 1 we can meaningfully compare the recovered low-dimensional warp field to the true underlying motion. Refer to Table S.1. All results were initialized with the warp parameters $[1, 0, 0, 1]$, which correspond to static evolution (No Warp).

	$\theta[1]$	$\theta[2]$	$\theta[3]$	$\theta[4]$	NRMSE
Truth Values	0.9945	-0.1045	0.1045	0.9945	0
Initialization	1.0000	0.0000	0.0000	1.0000	0.1047
Recovered Under No Atmospheric Error	0.9979	-0.0792	0.0799	0.9984	0.0253
Recovered Under Atmospheric Error	1.0018	-0.0169	0.0256	1.0006	0.0836

Table S.1: True and recovered 4-dimensional affine warp parameters, θ, for Video 1. We also specify the normalized root mean squared error (NRMSE) of the recovered flow field with respect to the true field on a 30×30 pixel image.

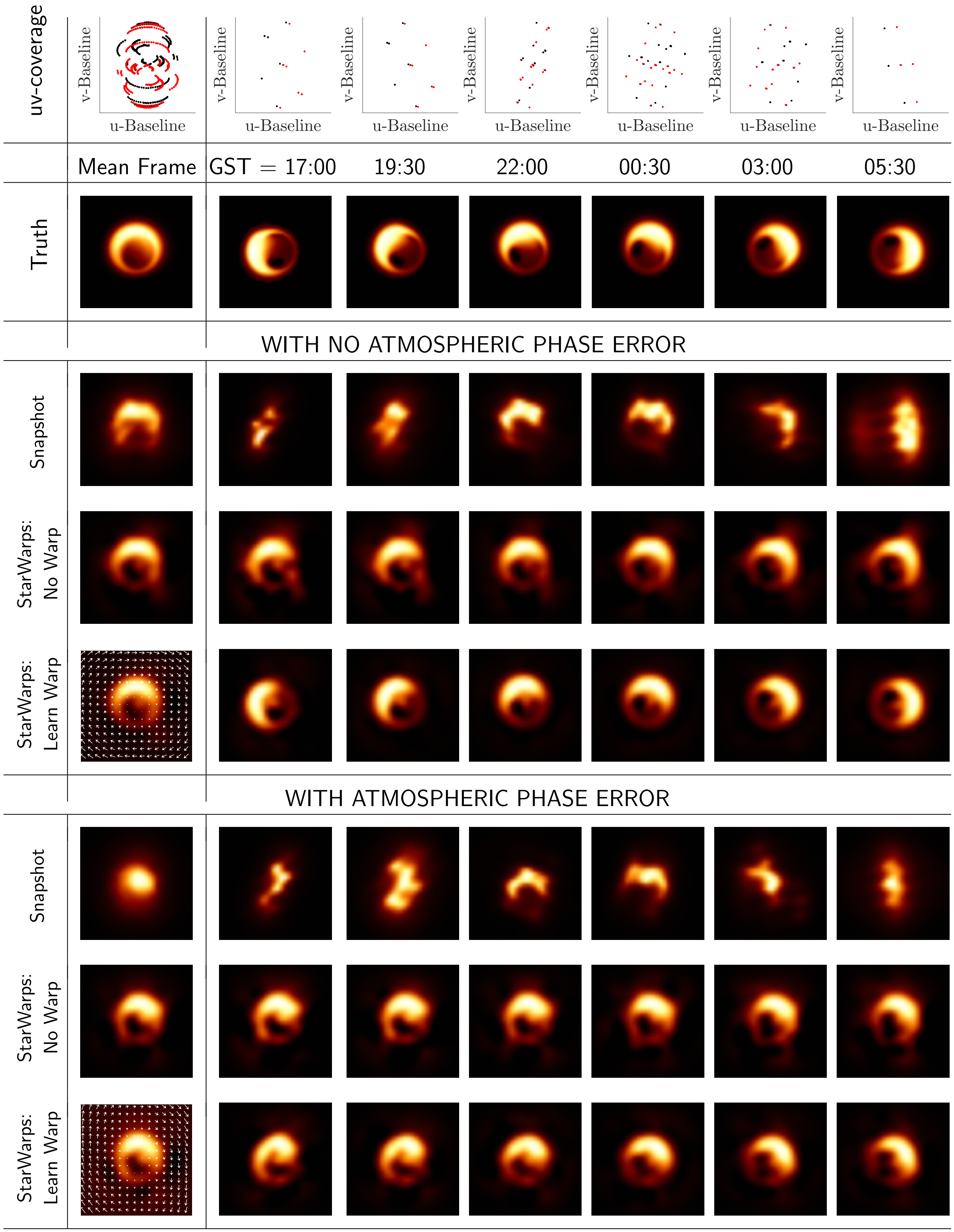

Figure S.3: **Time-resolved reconstruction of Video 1:** Video 1 contains an image rotating clockwise by 180° over the course of the observation. At each time, the interferometric telescope array measures values related to 2D spatial frequencies of the current underlying image, shown in the row labeled 'Truth'. These are indicated by the dots on the uv-coverage plots. As light emitted from the source is real-valued we obtain two values on opposite sides of the frequency plane – each independent set of measurements displayed as either black or red. We present results obtained when using calibrated data with no atmospheric error, as well as when there is atmospheric phase error still present and we must use data products invariant to its effects. Below the true images, we show a subset of images from the baseline 'snapshot imaging' method and compare it to our StarWarps reconstructed video obtained when we assume a static warp field or an inferred warp field. The mean image for each sequence is shown in the leftmost column. In the case that we simultaneously estimate a warp field, we indicate the resulting field as arrows on the mean image. Our method substantially improves results over the snapshot method, especially in the case of atmospheric error when the absolute position of the source cannot be recovered. Additionally, our proposed method can estimate a warp field that gives a sense of the underlying motion of the emission region. This can help to improve results, most notably for the calibrated data with no remaining atmospheric error.

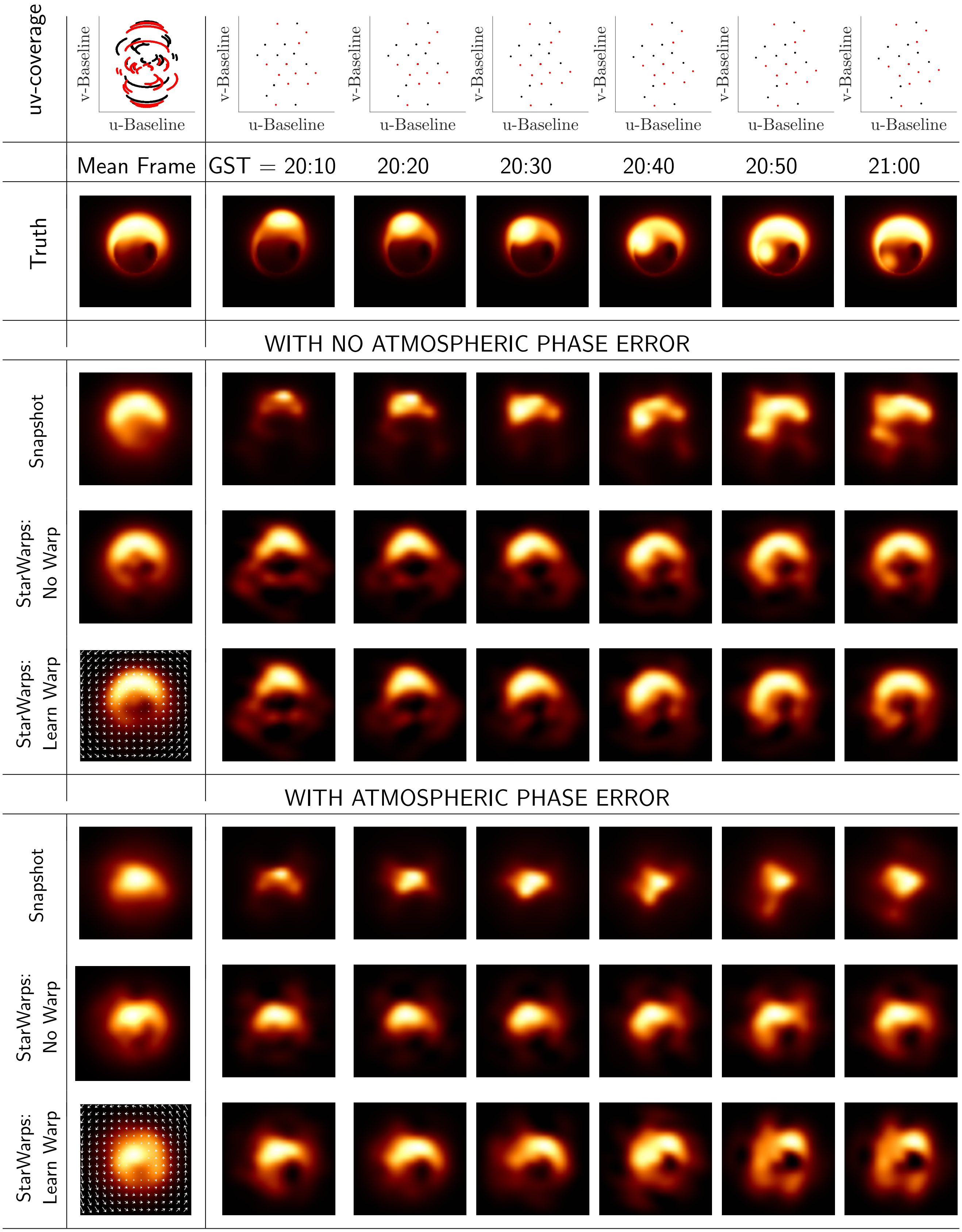

Figure S.4: **Time-resolved reconstruction of Video 2:** Video 2 contains a sequence of a hotspot orbiting counter-clockwise around a black hole. We present time-resolved results obtained using data derived from this sequence. Below the true images, we show a subset of images from the baseline 'snapshot imaging' method and compare it to our StarWarps reconstructed video obtained assuming a static warp field or an inferred warp field. The mean image for each sequence is shown in the leftmost column. If we simultaneously estimate a warp field, we indicate the resulting field as arrows on the mean image. Our method substantially improve results over the snapshot method, especially in the case of atmospheric error when the absolute position of the source cannot be recovered. Additionally, despite the fact that this hotspot video does not match our assumed motion model, using our proposed approach we were able to estimate a warp field that provides the direction of the source's true underlying motion. See the caption of Figure S.3 for more detail.

4 Lemmas

4.1 Lemma 1

$$\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^*_{t|t-1}, \mathbf{P}^*_{t|t-1}) = \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}_{t|t-1}, \mathbf{P}_{t|t-1}) \propto \mathcal{N}_{\boldsymbol{x}_t}(m_t, C_t) \tag{112}$$

where from the Matrix Cookbook (8.1.8) and (3.2.5) [1] we know that

$$\begin{aligned} C_t &= (\boldsymbol{\Lambda}^{-1} + \mathbf{P}^{-1}_{t|t-1})^{-1} && (113)\\ &= \boldsymbol{\Lambda}(\boldsymbol{\Lambda} + \mathbf{P}_{t|t-1})^{-1}\mathbf{P}_{t|t-1} && (114)\\ m_t &= (\boldsymbol{\Lambda}^{-1} + \mathbf{P}^{-1}_{t|t-1})^{-1}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\mu} + \mathbf{P}^{-1}_{t|t-1}\mathbf{z}_{t|t-1}) && (115)\\ &= \boldsymbol{\Lambda}(\boldsymbol{\Lambda} + \mathbf{P}_{t|t-1})^{-1}\mathbf{z}_{t|t-1} + \mathbf{P}_{t|t-1}(\boldsymbol{\Lambda} + \mathbf{P}_{t|t-1})^{-1}\boldsymbol{\mu} && (116)\\ & && (117) \end{aligned}$$

4.2 Lemma 2

$$\mathcal{N}_x(\boldsymbol{F}^{-1}\boldsymbol{y}, (\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F})^{-1})\mathcal{N}_x(\boldsymbol{\mu}, \boldsymbol{Q}) \propto \mathcal{N}_x(m, C) \tag{118}$$

where from the Matrix Cookbook (8.1.8) [1] and the Woodbury matrix identity we know that

$$\begin{aligned} C_t &= (\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F} + \boldsymbol{Q}^{-1})^{-1} && (119)\\ &= \boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{F}^T(\boldsymbol{R} + \boldsymbol{F}\boldsymbol{Q}\boldsymbol{F}^T)^{-1}\boldsymbol{F}\boldsymbol{Q} && (120)\\ m_t &= C_t(\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F}\boldsymbol{F}^{-1}\boldsymbol{y} + \boldsymbol{Q}^{-1}\boldsymbol{\mu}) && (121)\\ &= (\boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{F}^T(\boldsymbol{R} + \boldsymbol{F}\boldsymbol{Q}\boldsymbol{F}^T)^{-1}\boldsymbol{F}\boldsymbol{Q})(\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{y} + \boldsymbol{Q}^{-1}\boldsymbol{\mu}) && (122)\\ &= \boldsymbol{Q}\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{y} + \boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{\mu} - \boldsymbol{Q}\boldsymbol{F}^T(\boldsymbol{R} + \boldsymbol{F}\boldsymbol{Q}\boldsymbol{F}^T)^{-1}\boldsymbol{F}\boldsymbol{Q}\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{F}^T(\boldsymbol{R} + \boldsymbol{F}\boldsymbol{Q}\boldsymbol{F}^T)^{-1}\boldsymbol{F}\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{\mu}) \\ & && (123)\\ &= \boldsymbol{\mu} + \boldsymbol{Q}\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{F}^T(\boldsymbol{R} + \boldsymbol{F}\boldsymbol{Q}\boldsymbol{F}^T)^{-1}\boldsymbol{F}\boldsymbol{Q}\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{F}^T(\boldsymbol{R} + \boldsymbol{F}\boldsymbol{Q}\boldsymbol{F}^T)^{-1}\boldsymbol{F}\boldsymbol{\mu}) && (124)\\ &= \boldsymbol{\mu} + \boldsymbol{Q}\boldsymbol{F}^T(\boldsymbol{R} + \boldsymbol{F}\boldsymbol{Q}\boldsymbol{F}^T)^{-1}((\boldsymbol{R} + \boldsymbol{F}\boldsymbol{Q}\boldsymbol{F}^T)\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{F}\boldsymbol{Q}\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\mu}) && (125)\\ &= \boldsymbol{\mu} + \boldsymbol{Q}\boldsymbol{F}^T(\boldsymbol{R} + \boldsymbol{F}\boldsymbol{Q}\boldsymbol{F}^T)^{-1}(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\mu}) && (126) \end{aligned}$$

4.3 Lemma 3

$$\begin{aligned} \mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}^*_{t|t+1}, \mathbf{P}^*_{t|t+1}) &= \mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}_{t|t+1}, \mathbf{P}_{t|t+1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) && (127)\\ &= \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}^{-1}\mathbf{z}_{t+1|t+1}, (\boldsymbol{A}^T(\boldsymbol{Q} + \mathbf{P}_{t+1|t+1})^{-1}\boldsymbol{A})^{-1})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) && (128)\\ &\propto \mathcal{N}_{\boldsymbol{x}_t}(m_t, C_t) && (129) \end{aligned}$$

Use Lemma 4.2 where $\boldsymbol{R} = \boldsymbol{Q} + \mathbf{P}_{t+1|t+1}$ and $\boldsymbol{y} = \mathbf{z}_{t+1|t+1}$

4.4 Lemma 4

$$
\begin{aligned}
\mathcal{N}_{\boldsymbol{y}}(\boldsymbol{F}\boldsymbol{x}+B,\boldsymbol{R}) &= C\exp\left[\frac{-1}{2}(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{x}-B)^T\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{x}-B)\right] && (130)\\
&= C\exp\left[\frac{-1}{2}(\boldsymbol{F}(\boldsymbol{F}^{-1}(\boldsymbol{y}-B)-\boldsymbol{x}))^T\boldsymbol{R}^{-1}\boldsymbol{F}(\boldsymbol{F}^{-1}(\boldsymbol{y}-B)-\boldsymbol{x})\right] && (131)\\
&= C\exp\left[\frac{-1}{2}(\boldsymbol{F}^{-1}(\boldsymbol{y}-B)-\boldsymbol{x})^T\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F}(\boldsymbol{F}^{-1}(\boldsymbol{y}-B)-\boldsymbol{x})\right] && (132)\\
&= \mathcal{N}_{\boldsymbol{x}}(\boldsymbol{F}^{-1}(\boldsymbol{y}-B),(\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}) && (133)
\end{aligned}
$$

4.5 Lemma 5

Using Lemma 4.2 we can solve for

$$
N_x(\dot{\boldsymbol{F}}(\hat{x})^{-1}(\boldsymbol{y}+\dot{\boldsymbol{F}}(\hat{\boldsymbol{x}})\hat{\boldsymbol{x}}-\boldsymbol{F}(\hat{\boldsymbol{x}})),(\dot{\boldsymbol{F}}(\hat{\boldsymbol{x}})^T\boldsymbol{R}^{-1}\dot{\boldsymbol{F}}(\hat{\boldsymbol{x}}))^{-1})\mathcal{N}_{\boldsymbol{x}}(\hat{\boldsymbol{x}},\boldsymbol{Q}) \propto \mathcal{N}_{\boldsymbol{x}}(m,C) \quad (134)
$$

$$
\begin{aligned}
C_t &= \boldsymbol{Q}-\boldsymbol{Q}\dot{\boldsymbol{F}}(\hat{\boldsymbol{x}})^T(\boldsymbol{R}+\dot{\boldsymbol{F}}(\hat{x})\boldsymbol{Q}\dot{\boldsymbol{F}}(\hat{\boldsymbol{x}})^T)^{-1}\dot{\boldsymbol{F}}(\hat{x})\boldsymbol{Q} && (135)\\
m_t &= \hat{\boldsymbol{x}}+\boldsymbol{Q}\dot{\boldsymbol{F}}(\hat{x})^T(\boldsymbol{R}+\dot{\boldsymbol{F}}(\hat{x})\boldsymbol{Q}\dot{\boldsymbol{F}}(\hat{\boldsymbol{x}})^T)^{-1}(\boldsymbol{y}+\dot{\boldsymbol{F}}(\hat{x})\hat{\boldsymbol{x}}-\boldsymbol{F}(\hat{\boldsymbol{x}})-\dot{\boldsymbol{F}}(\hat{\boldsymbol{x}})\hat{\boldsymbol{x}}) && (136)\\
&= \hat{\boldsymbol{x}}+\boldsymbol{Q}\dot{\boldsymbol{F}}(\hat{x})^T(\boldsymbol{R}+\dot{\boldsymbol{F}}(\hat{x})\boldsymbol{Q}\dot{\boldsymbol{F}}(\hat{\boldsymbol{x}})^T)^{-1}(\boldsymbol{y}-\boldsymbol{F}(\hat{\boldsymbol{x}})) && (137)
\end{aligned}
$$

4.6 Lemma 6

$$
\begin{aligned}
&\mathcal{N}_{\boldsymbol{A}^{-1}(\boldsymbol{x}_{t+1}-B)}(\mathbf{z},(\boldsymbol{A}^T\boldsymbol{Q}^{-1}\boldsymbol{A})^{-1}+\mathbf{P})\\
&= C\exp\left[\frac{-1}{2}(\boldsymbol{A}^{-1}(\boldsymbol{x}_{t+1}-B)-\mathbf{z})^T((\boldsymbol{A}^T\boldsymbol{Q}^{-1}\boldsymbol{A})^{-1}+\mathbf{P})^{-1}(\boldsymbol{A}^{-1}(\boldsymbol{x}_{t+1}-B)-\mathbf{z})\right] && (138)\\
&= C\exp\left[\frac{-1}{2}(\boldsymbol{x}_{t+1}-B-\boldsymbol{A}\mathbf{z})^T\boldsymbol{A}^{-1T}((\boldsymbol{A}^T\boldsymbol{Q}^{-1}\boldsymbol{A})^{-1}+\mathbf{P})^{-1}\boldsymbol{A}^{-1}(\boldsymbol{x}_{t+1}-B-\boldsymbol{A}\mathbf{z})\right] && (139)\\
&= \mathcal{N}_{\boldsymbol{x}_{t+1}}(\boldsymbol{A}\mathbf{z}+B,(\boldsymbol{A}^{-1T}((\boldsymbol{A}^T\boldsymbol{Q}^{-1}\boldsymbol{A})^{-1}+\mathbf{P})^{-1}\boldsymbol{A}^{-1})^{-1}) && (140)\\
&= \mathcal{N}_{\boldsymbol{x}_{t+1}}(\boldsymbol{A}\mathbf{z}+B,\boldsymbol{A}((\boldsymbol{A}^T\boldsymbol{Q}^{-1}\boldsymbol{A})^{-1}+\mathbf{P})\boldsymbol{A}^T) && (141)\\
&= \mathcal{N}_{\boldsymbol{x}_{t+1}}(\boldsymbol{A}\mathbf{z}+B,\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{Q}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^T+\boldsymbol{A}\mathbf{P}\boldsymbol{A}^T) && (142)\\
&= \mathcal{N}_{\boldsymbol{x}_{t+1}}(\boldsymbol{A}\mathbf{z}+B,\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{Q}\boldsymbol{A}^{-1T}\boldsymbol{A}^T+\boldsymbol{A}\mathbf{P}\boldsymbol{A}^T) && (143)\\
&= \mathcal{N}_{\boldsymbol{x}_{t+1}}(\boldsymbol{A}\mathbf{z}+B,\boldsymbol{Q}+\boldsymbol{A}\mathbf{P}\boldsymbol{A}^T) && (144)\\
& && (145)
\end{aligned}
$$

References

[1] Petersen, K.B., Pedersen, M.S., et al.: The matrix cookbook. Technical University of Denmark **7** (2008) 15

[2] Cornwell, T., Wilkinson, P.: A new method for making maps with unstable radio interferometers. Monthly Notices of the Royal Astronomical Society **196**(4) (1981) 1067–1086

[3] Thompson, A.R., Moran, J.M., Swenson, Jr., G.W.: Interferometry and Synthesis in Radio Astronomy, 3rd Edition. (2017)