Filtered finite state projection method for the analysis and estimation of stochastic biochemical reaction networks: *Supplementary Material*

Elena D'Ambrosio, Zhou Fang, Ankit Gupta, Mustafa Khammash

S.1 Verification of the filtering equation

In this section, we prove the first theorem in our paper that under the non-explosivity condition

$$\sup_{t\in[0,T]} \sum_{j=1}^{M} \mathbb{E}\left[a_j^2(\mathbf{Z}(t))\right] < \infty \qquad\qquad \forall T > 0, \tag{1}$$

the conditional distribution $\pi(t,x)$ uniquely solves the filtering equation up to indistinguishability. When the state space is finite (e.g., the chemical reaction system admits positive conservation laws), the result has already been shown in [4]; however, the result in the case of infinite state spaces was still open before this work. In Section S.1.1, we first discuss the well-posedness of the conditional distribution $\pi(t,x)$. Then in Section S.1.2, we rewrite the filtering equation into a form that is easier to analyze. Finally, we verify the filtering equation in Section S.1.3 using the innovation method [2] and a Picard iteration [3, 1].

S.1.1 Well-posedness of $\pi(t,x)$

In this paper, we are interested in calculating the conditional distribution of hidden state $X(t)$ given the observations up to time t, i.e., $P\{\mathbf{X}(t) = x|\mathbf{Y}(s), 0 \leq s \leq t\}$. Here, we need to emphasize a few points. First, for every fixed t, this conditional distribution is random (rather than deterministic) and is measurable with respect to the filtration generated by $\mathbf{Y}(t)$, which we denote by $\mathcal{Y}_t$. In other words, the value of this conditional distribution fully depends on the random trajectory of the observation process. Second, for every fixed t, this conditional distribution is not uniquely defined. Mathematically, any $\mathcal{Y}_t$-adaptive random distribution $\mathcal{P}(x)$ can be viewed as a version of this conditional distribution as long as it satisfies

$$\mathbb{E}\left[\mathcal{P}(x)A\right] = \mathbb{E}\left[\mathbb{1}(\mathbf{X}(t) = x)A\right] \quad \forall \mathcal{Y}_t\text{-measurable random variable A with finite expectation and } \forall x \in \mathbb{Z}_{\geq 0}^{n_1}.$$

Consequently, $P\{\mathbf{X}(t) = x|\mathbf{Y}(s), 0 \leq s \leq t\}$ viewed as a continuous-time process is not uniquely defined either.

Among all these choices of conditional distributions, there is a càdlàg version, denoted by $\pi(t,x)$, satisfying $\lim_{s\to t^+} \pi(s,x) = \pi(t,x)$ and permitting the existence of $\lim_{s\to t^-} \pi(s,x)$ [2, Theorem 2.24]. In this paper, we specifically investigate such a conditional distribution. Moreover, the non-explosivity condition (1) implies that the propensities $a_j(\mathbf{Z}(t))$ (for each $j = 1,\dots,M$) are uniformly integrable in any finite time interval $[0,T]$. Therefore, in addition to the càdlàg property of $\pi(t,x)$, the conditional expectation $\pi(t,a_j) \triangleq \sum_{x'} a_j(x',\mathbf{Y}(t))\pi(t,x')$ (for each $j = 1,\dots M$) is also càdlàg under the non-explosiviy condition (1) [2, Remark 2.27].

S.1.2 Rewrite the filtering equation

Here, we rewrite the filtering equation in a form that is easier to analyze. We recall that ν_i'', $i \in \mathcal{O}$, are the sub-stoichiometry vectors regulating the net change of the observation process $\mathbf{Y}(t)$ once the reaction i has fired. Such sub-vectors may not be all distinct. Therefore, in addition to the notations introduced in the main text, we further term

- $\{\mu_1, \dots, \mu_{m_1}\}$ as the set of non-zero and distinguishable ν_i'', $i \in \mathcal{O}$
- $\mathcal{O}_{\mu_k} \triangleq \{j | \nu_j'' = \mu_k\}$ $(k = 1, \dots, m_1)$ as the set in which non-zero ν_j'', $j \in \mathcal{O}$, are identical to μ_k,
- $\tilde{R}_{\mu_k}(t) \triangleq \sum_{j \in \mathcal{O}_{\mu_k}} R_j \left(\int_0^t a_j(\mathbf{X}(s), \mathbf{Y}(s))\, ds \right)$ as the total firing number of the reactions in $\mathcal{O}_{\mu_k}$ up to time t,
- $a^{\mathcal{O}_{\mu_k}}(\mathbf{X}(s), \mathbf{Y}(s)) \triangleq \sum_{j \in \mathcal{O}_{\mu_k}} a_j(\mathbf{X}(s), \mathbf{Y}(s))$ as the rate of the process $\tilde{R}_{\mu_k}(t)$.

Then, the filtering equation can be rewritten as

$$\begin{aligned} \pi(t,x) =& \pi(0,x) + \int_0^t \sum_{j \in \mathcal{U}} a_j(x - \nu_j', \mathbf{Y}(s)) \pi(s, x - \nu_j') - \sum_{j \in \mathcal{U}} a_j(x, \mathbf{Y}(s)) \pi(s,x) ds \\ & - \int_0^t \pi(s,x) \left(a^{\mathcal{O}}(x, \mathbf{Y}(s)) - \sum_{\tilde{x}} a^{\mathcal{O}}(\tilde{x}, \mathbf{Y}(s)) \pi(s, \tilde{x}) \right) ds \\ & + \sum_{k=1}^{m_1} \int_0^t \left(\frac{\sum_{j \in \mathcal{O}_{\mu_k}} a_j(x - \nu_j', \mathbf{Y}(s^-)) \pi(s^-, x - \nu_j')}{\sum_{\tilde{x}} a^{\mathcal{O}_{\mu_k}}(\tilde{x}, \mathbf{Y}(s^-)) \pi(s^-, \tilde{x})} - \pi(s^-, x) \right) d\tilde{R}_{\mu_k}(s) \\ & \qquad \forall t \geq 0 \text{ and } \forall x \in \mathbb{Z}_{\geq 0}^{n_1} \text{ almost surely.} \end{aligned} \tag{2}$$

S.1.3 Proof of Theorem 1

We first show that the conditional distribution $\pi(t,x)$ satisfies the filtering equation (2).

Theorem 1 (Validity of the filtering equation)**.** *Under the non-explosivity condition* (1)*, the conditional probability $\pi(t,x)$ (for any $x \in \mathbb{Z}_+^{n_1}$) satisfies* (2).

Proof. Here, we prove the result using the innovation method. First, the non-explosivity condition (1) suggests that the Chemical Master Equation (CME) precisely characterises the probability of the reaction process $\mathbf{Z}(t)$ [1]. Then, we exploit the CME to find an adequate martingale to represent the conditional distribution as a continuous-time process through the martingale representation theorem. Along these lines, by conditioning the CME w.r.t the observation process $\mathbf{Y}(t)$, we can easily prove that the process

$$\mathcal{M}(t,x) = \pi(t,x) - \left[\pi(0,x) + \int_0^t \left(\sum_{j=1}^{M} a_j(x - \nu_j', \mathbf{Y}(s)) \pi(s, x - \nu_j') - \sum_{j=1}^{M} a_j(x, \mathbf{Y}(s)) \pi(s,x) \right) ds \right] \tag{3}$$

is a square integrable martingale (due to (1)) adapted to $\mathcal{Y}_t$. By the martingale representation theorem [5], we can find $\mathcal{Y}_t$-predictable processes $\phi_1(t), \dots, \phi_{m_1}(t)$ such that this martingale can almost surely be expressed by:

$$\mathcal{M}(t) = \sum_{k=1}^{m_1} \int_0^t \phi_k(s) \left(d\tilde{R}_{\mu_k}(s) - \pi(s^-, a^{\mathcal{O}_{\mu_k}}) ds \right) \tag{4}$$

where $\pi(s^-, a^{\mathcal{O}_{\mu_k}})$ is the left limit of $\pi(s, a^{\mathcal{O}_{\mu_k}})$, and $\tilde{R}_{\mu_k}(t) - \int_0^t \pi(s^-, a^{\mathcal{O}_{\mu_k}})ds$ are also $\mathcal{Y}_t$-adaptive martingales. The square integrability of (3) guarantees the integrability of $\int_0^t (\phi_k(s))^2 \pi(s^-, a^{\mathcal{O}_{\mu_k}})ds$ which ensures that (4) holds globally. To finalise the proof, we only need to identify $\phi_1(t), \dots, \phi_{m_1}(t)$.

To this end, we can exploit the process $\pi(t,x)e^{-\tilde{R}_{\mu_k}(t)}$ and write its dynamics by

$$
\begin{aligned}
&\pi(t,x)e^{-\tilde{R}_{\mu_k}(t)} \\
&= \pi(0,x)e^{-\tilde{R}_{\mu_k}(0)} + \int_0^t e^{-\tilde{R}_{\mu_k}(s)} \left[\sum_{j=1}^{M} a_j(x-\nu_j', \mathbf{Y}(s))\pi(s, x-\nu_j') - \sum_{j=1}^{M} a_j(x, \mathbf{Y}(s))\pi(s,x) \right] ds \\
&\quad + \int_0^t e^{-\tilde{R}_{\mu_k}(s^-)} \left[\sum_{b=1}^{m_1} \phi_b(s) \left(d\tilde{R}_{\mu_b}(s) - \pi(s^-, a^{\mathcal{O}_{\mu_b}}) \right) \right] ds \\
&\quad + \int_0^t \pi(s^-, x)(e^{-1}-1)e^{-\tilde{R}_{\mu_k}(s^-)} d\tilde{R}_{\mu_k}(s) + \int_0^t \phi_k(s)(e^{-1}-1)e^{-\tilde{R}_{\mu_k}(s^-)} d[\tilde{R}_{\mu_k}]_s
\end{aligned} \tag{5}
$$

where $[\tilde{R}_{\mu_k}]_t$ is the quadratic variation of $\tilde{R}_{\mu_k}(t)$. Considering the process $\mathbb{1}(\mathbf{X}(t) = x)e^{-\tilde{R}_{\mu_k}(t)}$, we can write:

$$
\begin{aligned}
&\mathbb{1}(\mathbf{X}(t) = x)e^{-\tilde{R}_{\mu_k}(t)} \\
&= \mathbb{1}(\mathbf{X}(0) = x)e^{-\tilde{R}_{\mu_k}(0)} + \sum_{j=1}^{M} \int_0^t \Big[\mathbb{1}\left(\mathbf{X}(s^-) + \nu_j' = x\right) \left(\mathbb{1}(j \notin \mathcal{O}_{\mu_k})e^{-\tilde{R}_{\mu_k}(s)} + \mathbb{1}(j \in \mathcal{O}_{\mu_k})e^{-\tilde{R}_{\mu_k}(s)-1} \right) \\
&\qquad - \mathbb{1}(\mathbf{X}(s^-) = x)e^{-\tilde{R}_{\mu_k}(s^-)} \Big] d\tilde{R}_j(s)
\end{aligned} \tag{6}
$$

where $\tilde{R}_j(t) = R_j\left(\int_0^t a_j(\mathbf{X}(s), \mathbf{Y}(s))ds\right)$.

Then, by denoting $b(s) = \left(e^{-1}-1\right) e^{-\tilde{R}_{\mu_k}(s^-)}$ and applying (5) and (6) to the relation $\mathbb{E}\left[\pi(t,x)e^{-\tilde{R}_{\mu_k}(t)}\right] = \mathbb{E}\left[\mathbb{1}(\mathbf{X}(t) = x)e^{-\tilde{R}_{\mu_k}(t)}\right]$, we have

$$
\begin{aligned}
&\mathbb{E}\left[-\int_0^t b(s) \sum_{j \in \mathcal{O}_{\mu_k}} a_j(x-\nu', \mathbf{Y}(s))ds + \int_0^t b(s)\pi(s^-,x)\pi(s^-, a^{\mathcal{O}_{\mu_k}})ds + \int_0^t b(s)\phi_k(s)\pi(s^-, a^{\mathcal{O}_{\mu_k}})ds \right] \\
&= 0
\end{aligned}
$$

for $k = 1, \dots, m_1$. Similarly, we can show that the term in this bracket is a martingale by comparing $\mathbb{E}\left[\pi(t,x)e^{-\tilde{R}_{\mu_k}(t)} \middle| \mathcal{F}_s\right] = \mathbb{E}\left[\mathbb{1}\left(\mathbf{X}(t) = x\right) e^{-\tilde{R}_{\mu_k}(t)} \middle| \mathcal{F}_s\right]$ $0 \le s \le t$, where $\mathcal{F}_t$ is the natural filtration generated by the whole process $\mathbf{Z}(t)$. To summarise, this term is a continuous martingale starting at zero with finite total variation; it then follows that this process is almost surely zero [6]. Therefore, for every $k \in \{1, \dots, m_1\}$, every state $x \in \mathbb{Z}_+^{n_1}$, and almost every $t > 0$, given that the process $b(t)$ is non-zero almost surely, it almost surely holds the relation:

$$
\pi(t^-, a^{\mathcal{O}_{\mu_k}})\phi_k(t) = \sum_{j \in \mathcal{O}_{\mu_k}} a_j(x-\nu_j', \mathbf{Y}(t^-))\pi(t^-, x-\nu_j') - \pi(t^-,x)\pi(t^-, a^{\mathcal{O}_{\mu_k}}).
$$

Finally, by inserting the above equality while exploiting the càdlàg property of $\pi(t,x)$ and $\pi(t, a^{\mathcal{O}_{\mu_k}})$ in the martingale representation (4), we prove the result. □

We then prove the uniqueness of the solution of the filtering equation.

Theorem 2. *The filtering equation* (2) *has a unique non-negative solution (denoted by $\tilde{p}(t,x)$) up to indistinguishability that starts from $\pi(0,\cdot)$ and satisfies*

$$\int_0^t \sum_{x\in\mathbb{Z}_+^{n_1}} \sum_{j=1}^{M} a_j(x,\mathbf{Y}(s))\tilde{p}(s,x)ds < \infty \text{ almost surely}, \quad \forall t \geq 0. \tag{7}$$

Proof. The existence of the solution is proven in Theorem 1; here, we only prove the uniqueness of the solution. Let $t_1, t_2, \ldots$ be the jumping times of $\mathbf{Y}(t)$ and $t_0 = 0$. Note that if the solution of (2) is unique between any interval $[t_k, t_{k+1})$ ($k = 0, 1, \ldots$) given the initial condition at time t_k (i.e., $\pi(t_k,\cdot)$), then the result naturally holds. In the following, we will use the Picard iteration to prove the result.

Between any time interval $[t_k, t_{k+1})$ and any fixed state, we can view (2) as a linear ordinary differential equation of the form $\dot{y}(t) = -\alpha y(t) + \beta(t)$, with $y(t) = \pi(t,x)$, $\alpha = \sum_{j=1}^{M} a_j(x,\mathbf{Y}(t_k))$, and

$$\beta(t) = \sum_{j\in\mathcal{U}} a_j(x-\nu'_j,\mathbf{Y}(t_k))\pi(t,x-\nu'_j) + \pi(t,x)\left(\sum_{x'\in\mathbb{Z}_+^{n_1}} a^{\mathcal{O}}(x',\mathbf{Y}(t_k))\pi(t,x')\right)$$

Therefore, between any time interval $[t_k, t_{k+1})$, the filtering equation (2) can be equivalently rewritten by

$$\begin{aligned}\pi(t,x) =& \pi(t_k,x)\exp(-\bar{a}(x)t) + \int_{t_k}^{t} \sum_{j\in\mathcal{U}} \exp(-\bar{a}(x)(t-s))a_j(x-\nu'_j,\mathbf{Y}(t_k))\pi(s,x-\nu'_j)ds \\ &+ \int_{t_k}^{t} \exp(-\bar{a}(x)(t-s))\pi(s,x)\left(\sum_{x'\in\mathbb{Z}_+^{n_1}} a^{\mathcal{O}}(x',\mathbf{Y}(t_k))\pi(s,x')\right)ds \end{aligned}\tag{8}$$

for any $x \in \mathbb{Z}_{\geq 0}^{n_1}$, where $\bar{a}(x) = \sum_{j=1}^{M} a_j(x,\mathbf{Y}(t_k))$.

Now, we can construct the Picardi iteration through which a solution of (8) is attained by setting $p^{(0)}(t,z) \equiv 0$ for $t \in [t_k, t_{k+1})$ and

$$\begin{aligned}p^{(\ell+1)}(t,x) =& \pi(t_k,x)\exp(-\bar{a}(x)t) + \int_{t_k}^{t} \sum_{j\in\mathcal{U}} \exp(-\bar{a}(x)(t-s))a_j(x-\nu'_j,\mathbf{Y}(t_k))p^{(\ell)}(s,x-\nu'_j)ds \\ &+ \int_{t_k}^{t} \exp(-\bar{a}(x)(t-s))p^{(\ell)}(s,x)\left(\sum_{x'\in\mathbb{Z}_+^{n_1}} a^{\mathcal{O}}(x',\mathbf{Y}(t_k))p^{(\ell)}(s,x')\right)ds\end{aligned}$$

With mathematical induction, we can check that, for any $t \in [t_k, t_{k+1})$ and $x \in \mathbb{Z}_+^{n_1}$, the succession $\left\{p^{(\ell)}(t,x)\right\}_{\ell\in\mathbb{N}}$ is monotonic with respect to ℓ and each term of the succession is less than or equal to any non-negative solution of (8) satisfying (7). Therefore, as ℓ grows to infinity, $p^{(\ell)}(t,x)$ almost surely converges to a non-negative random variable, denoted by $p^{(\infty)}(t,x)$. Moreover, $p^{(\infty)}(t,\cdot)$ is almost surely no greater than any non-negative $\tilde{p}(t,\cdot)$ solving (8) with initial condition $\pi(t_k,\cdot)$ and satisfying (7), i.e.,

$$p^{(\infty)}(t,x) \leq \tilde{p}(t,x) \quad \text{for any } t \in [t_k, t_{k+1}) \text{ and any } x \in \mathbb{Z}_+^{n_1}, \text{ almost surely.} \tag{9}$$

Furthermore, by the convergence of $\left\{p^{(\ell)}(t,x)\right\}_{\ell\in\mathbb{N}}$, we can verify that $p^{(\infty)}(t,\cdot)$ also solves (8) in the interval (t_k, t_{k+1}) with initial condition $\pi(t_k,\cdot)$. Also, from (9), we can show verify that $p^{(\infty)}(t,x)$ also satisfies (7) within the integral region $[t_k, t_{k+1})$.

By the dominant convergence theorem, we can prove that any non-negative solution $\tilde{p}(t,\cdot)$ of (8) satisfying (7) has a conserved total mass, i.e.,

$$\sum_{x\in\mathbb{Z}_+^{n_1}} p^{(\infty)}(t,x) = \sum_{x\in\mathbb{Z}_+^{n_1}} \pi(t_k,x) = \sum_{x\in\mathbb{Z}_+^{n_1}} \tilde{p}(t,x) \qquad \forall t\in[t_k,t_{k+1}) \text{ almost surely.} \tag{10}$$

Thanks to (9) and (10), all non-negative solutions $\tilde{p}(t,\cdot)$ of (8) with initial condition $\pi(t_k,\cdot)$ are the same in the time interval $[t_k,t_{k+1})$ almost surely, which proves the result. □

By combining these two theorems, we prove the existence and the uniqueness of the solution of the filtering equation.

S.2 Error analysis of the filtered finite state projection (FFSP) algorithms

In this section, we provide more details about the error analysis for the FFSP algorithms. In Section S.2.1, we present the proof of the extended FSP theorem. Then, in Section S.2.2 and Section S.2.3, we show more details about the error analyses for the first and second FFSP algorithms, respectively.

Also, note that the results are given under the conditions that

$$\text{there exists a function } \bar{a}^{\mathcal{O}}(y) \text{ such that } \sup_{x\in\mathbb{Z}_{\geq 0}^{n_1}} a^{\mathcal{O}}(x,y)\leq \bar{a}^{\mathcal{O}}(y), \tag{11}$$

and that

$$\begin{aligned}&\text{if } \mathbb{E}\left[\left\|\tilde{\mathbf{X}}_y(0)\right\|_1^q\right]<+\infty \text{ for all } q\geq 0,\\ &\text{then } \textstyle\sum_{j\in\mathcal{U}}\mathbb{E}\left[\int_0^t a_j\left(\tilde{\mathbf{X}}_y(s),y\right)ds\right]<\infty,\ \forall t\geq 0 \text{ and } \forall y\in\mathbb{Z}_{\geq 0}^{n_2}.\end{aligned} \tag{12}$$

S.2.1 proof of the extended FSP theorem

Proof of the extended FSP theorem. The framework of this proof is as follows. First, we introduce an auxiliary process $\hat{p}(t,\mathcal{X})$ which has the same time-evolution dynamics as $\bar{p}(t,\mathcal{X})$ but starts at $p_{\mathrm{FSP}}(0,\mathcal{X})$. Then, the error between $\bar{p}(t,\mathcal{X})$ and $p_{\mathrm{FSP}}(t,\mathcal{X})$ can be bounded by

$$\|\bar{p}(t,\mathcal{X})-p_{\mathrm{FSP}}(t,\mathcal{X})\|_1 \leq \|\bar{p}(t,\mathcal{X})-\hat{p}(t,\mathcal{X})\|_1 + \|\hat{p}(t,\mathcal{X})-p_{\mathrm{FSP}}(t,\mathcal{X})\|_1 \tag{13}$$

Finally, we prove the result by investigating the two errors on the right hand side of this inequality.

Construct $\hat{p}(t,\mathcal{X})$: Note that $p_{\mathrm{FSP}}(0,\mathcal{X})$ has a compact support; therefore, the stochastic process $\tilde{\mathbf{X}}_y(t)$ with the initial distribution $\frac{p_{\mathrm{FSP}}(0,\cdot)}{\sum_{\tilde{x}\in\mathcal{X}} p_{\mathrm{FSP}}(0,\tilde{x})}$ is non-explosive due to (12). By [7, Lemma 1], the process

$$\tilde{p}(t,x)\triangleq\mathbb{E}\left[\mathbb{1}\left(\tilde{\mathbf{X}}_y(t)=x\right)\exp\left(-\int_0^t a^{\mathcal{O}}\left(\tilde{\mathbf{X}}_y(s),y\right)ds\right)\right]$$

evolves according to the same dynamics as $\bar{p}(t,\mathcal{X})$ but starts from $\frac{p_{\mathrm{FSP}}(0,x)}{\sum_{\tilde{x}\in\mathcal{X}} p_{\mathrm{FSP}}(0,\tilde{x})}$. Then, the process $\hat{p}(t,\mathcal{X})=\tilde{p}(t,\mathcal{X})\left(\sum_{\tilde{x}\in\mathcal{X}} p_{\mathrm{FSP}}(0,\tilde{x})\right)$ satisfies

$$\frac{d}{dt}\hat{p}(t,\mathcal{X})=\mathbb{A}(y)\hat{p}(t,\mathcal{X}) \quad \forall t\geq 0 \qquad \text{and} \qquad \hat{p}(0,\mathcal{X})=p_{\mathrm{FSP}}(0,\mathcal{X}).$$

Moreover, by this construction, we can easily check that $\hat{p}(t,\mathcal{X})$ is non-negative component-wise and has a finite L_1 norm at every time point; also, by (12), there holds $\int_0^t \sum_i |\mathbb{A}_{ii}(y)|\,\hat{p}(s,x_i)ds < \infty$.

Estimate of $\|\bar{p}(t,\mathcal{X}) - \hat{p}(t,\mathcal{X})\|_1$: Let us denote $e_1(t,x_i) = \bar{p}(t,x_i) - \hat{p}(t,x_i)$ for all $x_i \in \mathcal{X}$. Then, by linearity, we have $\dot{e}_1(t,x_i) = \sum_{j=1}^{\infty} \mathbb{A}_{ij}(y)e_1(t,x_j)$ for all $x_i \in \mathcal{X}$ and all $t \geq 0$; furthermore, we also have

$$
\begin{aligned}
&\frac{d}{dt^+}|e_1(t,x_i)| \\
&\triangleq \lim_{dt\to 0^+} \frac{|e_1(t+dt,x_i)| - |e_1(t,x_i)|}{dt} \\
&= \lim_{dt\to 0^+} \frac{\left|e_1(t,x_i) + \sum_{j=1}^{\infty} \mathbb{A}_{ij}(y)e_1(t,x_j)dt + o(dt)\right| - |e_1(t,x_i)|}{dt} \\
&\leq \lim_{dt\to 0^+} \frac{\left|e_1(t,x_i) + \mathbb{A}_{ii}(y)e_1(t,x_i)dt\right| + \left|\sum_{j\neq i} \mathbb{A}_{ij}(y)e_1(t,x_j)dt + o(dt)\right| - |e_1(t,x_i)|}{dt} \\
&\leq \lim_{dt\to 0^+} \frac{\left|e_1(t,x_i)\Big(1 + \mathbb{A}_{ii}(y)dt\Big)\right| - |e_1(t,x_i)|}{dt} + \left|\sum_{j\neq i} \mathbb{A}_{ij}(y)e_1(t,x_j)\right| \\
&= \mathbb{A}_{ii}(y)\,|e_1(t,x_i)| + \left|\sum_{j\neq i} \mathbb{A}_{ij}(y)e_1(t,x_j)\right| \\
&\leq \sum_{j=1}^{\infty} \mathbb{A}_{ij}(y)\,|e_1(t,x_j)| \qquad\qquad \forall i \in \mathbb{Z}_{>0} \text{ and } \forall t \geq 0,
\end{aligned}
$$

where $\frac{d}{dt^+}$ indicates the right derivative. Thus, we have the expression

$$
\|\bar{p}(t,\mathcal{X}) - \hat{p}(t,\mathcal{X})\|_1 = \sum_{i=1}^{\infty} |e_1(t,x_i)| \leq \sum_{i=1}^{\infty} |e_1(0,x_i)| + \sum_{i=1}^{\infty} \int_0^t \sum_{j=1}^{\infty} \mathbb{A}_{ij}(y)|e_1(s,x_j)|ds \qquad \forall t \geq 0.
$$

Note that the matrix $\mathbb{A}(y)$ is diagonally dominant, so the finiteness of $\int_0^t \sum_i |\mathbb{A}_{ii}(y)|\,\bar{p}(s,x_i)ds$ and $\int_0^t \sum_i |\mathbb{A}_{ii}(y)|\,\hat{p}(s,x_i)ds$ implies the finiteness of the terms $\int_0^t \sum_i \sum_j |\mathbb{A}_{ij}(y)|\,\bar{p}(s,x_j)ds$ and $\int_0^t \sum_i \sum_j |\mathbb{A}_{ij}(y)|\,\hat{p}(s,x_j)ds$. Therefore, by Fubini's theorem, we can further express the error by

$$
\begin{aligned}
\|\bar{p}(t,\mathcal{X}) - \hat{p}(t,\mathcal{X})\|_1 &\leq \sum_{i=1}^{\infty} |e_1(0,x_i)| + \int_0^t \sum_{j=1}^{\infty} |e_1(s,x_j)| \underbrace{\left(\sum_{i=1}^{\infty} \mathbb{A}_{ij}(y)\right)}_{\leq 0 \text{ component-wise}} ds \\
&\leq \|\bar{p}(0,\mathcal{X}) - \hat{p}(0,\mathcal{X})\|_1 \qquad (14)
\end{aligned}
$$

where the last line follows from the diagonal dominance of $\mathbb{A}(y)$.

Estimate of $\|\hat{p}(t,\mathcal{X}) - p_{FSP}(t,\mathcal{X})\|_1$: The analysis in this part is similar to the proof in the classical FSP theorem in [2]. Let us denote $\hat{p}(t,\mathcal{X}_J) \triangleq (\hat{p}(t,x_1),\ldots,\hat{p}(t,x_J))^\top$ and $\hat{p}(t,\mathcal{X}_{J'}) \triangleq (\hat{p}(t,x_{J+1}),\hat{p}(t,x_{J+2}),\ldots)^\top$. Then, we can rewrite the dynamics of $\hat{p}(t,\mathcal{X})$ by

$$
\begin{bmatrix} \dot{\hat{p}}(t,\mathcal{X}_J) \\ \dot{\hat{p}}(t,\mathcal{X}_{J'}) \end{bmatrix} = \begin{bmatrix} \mathbb{A}_J(y) & \mathbb{A}_{JJ'}(y) \\ \mathbb{A}_{J'J}(y) & \mathbb{A}_{J'J'}(y) \end{bmatrix} \begin{bmatrix} \hat{p}(t,\mathcal{X}_J) \\ \hat{p}(t,\mathcal{X}_{J'}) \end{bmatrix} \qquad \forall t \geq 0,
$$

where $\mathbb{A}_{JJ}(y)$, $\mathbb{A}_{JJ'}(y)$, $\mathbb{A}_{JJ'}(y)$, and $\mathbb{A}_{J'J'}(y)$ are sub-matrices of $\mathbb{A}(y)$ with proper size; therefore, we can obtain

$$\hat{p}(t,\mathcal{X}_J) = \underbrace{e^{\mathbb{A}_J(y)t}\, p_{\mathrm{FSP}}(0,\mathcal{X}_J)}_{=p_{\mathrm{FSP}}(t,\mathcal{X}_J)} + \int_0^t e^{\mathbb{A}_J(y)\,(t-\tau)}\, \mathbb{A}_{JJ'}(y)\, \hat{p}(\tau,\mathcal{X}_{J'})\, d\tau \qquad \forall t\geq 0$$

where $p_{\mathrm{FSP}}(t,\mathcal{X}_J) = \left(\bar{p}_{\mathrm{FSP}}(t,x_1),\dots,\bar{p}_{\mathrm{FSP}}(t,x_J)\right)^\top$. Note that $e^{\mathbb{A}_J(y)\,(t-\tau)}$ and $\mathbb{A}_{JJ'}(y)$ are component-wise positive due to the Metzler matrix $\mathbb{A}(y)$; also $\hat{p}(\tau,\mathcal{X}_{J'})$ is component-wise positive due to the discussion in the first part of this proof. Consequently, we can conclude that

$$\hat{p}(t,X_J)\geq p_{\mathrm{FSP}}(t,\mathcal{X}_J) \quad \text{component-wise} \qquad \forall t\geq 0. \tag{15}$$

Now, we denote the total mass of $\hat{p}(t,X)$ at time t by $c(t)\triangleq \left\|\hat{p}(t,X)\right\|_1$. By the Fubini's theorem, we can express $c(t)$ by

$$\begin{aligned} c(t) &= c(0) + \int_0^t \sum_{j=1}^{\infty} \hat{p}(s,x_j) \underbrace{\left(\sum_{i=1}^{\infty}\mathbb{A}_{i,j}(y)\right)}_{\leq 0 \text{ component-wise}} ds \\ &\leq \sum_{x\in\mathcal{X}} p_{\mathrm{FSP}}(0,x) + \int_0^t \mathbf{1}^\top \mathbb{A}(y)\, p_{\mathrm{FSP}}(s,\mathcal{X})ds \qquad \forall t\geq 0 \end{aligned}$$

where the last line follows from (15) and the fact that $p_{\mathrm{FSP}}(t,x)=0$ for all $x\in\mathcal{X}_{J'}$. Finally, combining this with (15), we have that

$$\begin{aligned} \left\|\hat{p}(t,\mathcal{X}) - p_{\mathrm{FSP}}(t,\mathcal{X})\right\|_1 &= \sum_{x\in\mathcal{X}} \left[\hat{p}(t,x) - p_{\mathrm{FSP}}(t,x)\right] \\ &= c(t) - \|p_{FSP}(t,\mathcal{X}_J)\|_1 \\ &\leq \sum_{x\in\mathcal{X}} p_{\mathrm{FSP}}(0,x) + \int_0^t \mathbf{1}^\top\mathbb{A}(y)\, p_{\mathrm{FSP}}(s,\mathcal{X})ds - \|p_{FSP}(t,\mathcal{X})\|_1 \qquad \forall t\geq 0. \end{aligned} \tag{16}$$

By combining (13), (14), and (16), we prove the result. □

S.2.2 Error analysis for the first FFSP algorithm

Proof of Theorem 2. Here, we prove the result by math induction. Obviously, we have $\|\pi_{\mathrm{FFSP}}(t_0,\mathcal{X}) - \pi(t_0,\mathcal{X})\|_1 = \epsilon(t_0) \triangleq \sum_{x\notin\mathcal{X}_J}\pi(t_0,x)$, suggesting that the result holds at time t_0. Then, we show that once the result holds at time t_k ($\forall k\in\mathbb{Z}_{>0}$), it also holds in the time interval $(t_k,t_{k+1}]$.

By the extended FSP theorem (whose assumption almost surely holds due to (1)), we almost surely have

$$\left\|\rho_{\mathrm{FFSP}}(t,\mathcal{X}) - \rho(t,\mathcal{X})\right\|_1 \leq \epsilon(t_k) + \tilde{\epsilon}(t) \qquad \forall t\in(t_k,t_{k+1}) \tag{17}$$

where $\rho_{\mathrm{FFSP}}(t,\mathcal{X}) = \left(\rho_{\mathrm{FFSP}}(t,x_1),\rho_{\mathrm{FFSP}}(t,x_2),\dots\right)^\top$, $\rho(t,\mathcal{X}) = \left(\rho(t,x_1),\rho(t,x_2),\dots\right)^\top$, and

$$\tilde{\epsilon}(t) = \|\rho_{\mathrm{FFSP}}(t_k,\mathcal{X})\|_1 + \int_0^t \mathbf{1}^\top\mathbb{A}\left(\mathbf{Y}(t_k)\right)\rho_{\mathrm{FFSP}}(s,\mathcal{X})\, ds - \|\rho_{\mathrm{FFSP}}(t,\mathcal{X})\|_1.$$

Then, for the normalized filter, we almost surely have that

$$
\begin{aligned}
&\|\pi_{\text{FFSP}}(t,\mathcal{X})-\pi(t,\mathcal{X})\|_1\\
\leq &\left\|\frac{\rho_{\text{FFSP}}(t,\mathcal{X})}{\|\rho_{\text{FFSP}}(t,\mathcal{X})\|_1}-\frac{\rho(t,\mathcal{X})}{\|\rho_{\text{FFSP}}(t,\mathcal{X})\|_1}\right\|_1+\left\|\frac{\rho(t,\mathcal{X})}{\|\rho_{\text{FFSP}}(t,\mathcal{X})\|_1}-\frac{\rho(t,\mathcal{X})}{\|\rho(t,\mathcal{X})\|_1}\right\|_1\\
= &\frac{\|\rho_{\text{FFSP}}(t,\mathcal{X})-\rho(t,\mathcal{X})\|_1}{\|\rho_{\text{FFSP}}(t,\mathcal{X})\|_1}+\frac{\left|\|\rho_{\text{FFSP}}(t,\mathcal{X})\|_1-\|\rho(t,\mathcal{X})\|_1\right|}{\|\rho_{\text{FFSP}}(t,\mathcal{X})\|_1}\\
\leq &\frac{2\big[\epsilon(t_k)+\tilde{\epsilon}(t)\big]}{\|\rho_{\text{FFSP}}(t,\mathcal{X})\|_1} \qquad \forall t\in(t_k,t_{k+1})
\end{aligned}
\tag{18}
$$

where the second line follows from the triangle inequality, and the last line follows from (17) and the variant triangle inequalities $\|\alpha\|_1-\|\beta\|_1\leq\|\alpha-\beta\|_1$ and $\|\beta\|_1-\|\alpha\|_1\leq\|\alpha-\beta\|_1$.

Now, we analyze the error at the next jump time t_{k+1}. We denote

$$\tilde{\rho}(t_{k+1},x)=\sum_{j\in\mathcal{O}_{k+1}}a_j\left(x-\nu'_j,\mathbf{Y}(t_k)\right)\rho\left(t^-_{k+1},x-\nu'_j\right) \qquad \forall x\in\mathbb{Z}^{n_1}_{\geq 0},$$

and

$$\tilde{\rho}_{\text{FFSP}}(t_{k+1},x)=\sum_{j\in\mathcal{O}_{k+1}}a_j\left(x-\nu'_j,\mathbf{Y}(t_k)\right)\rho_{\text{FFSP}}\left(t^-_{k+1},x-\nu'_j\right) \qquad \forall x\in\mathbb{Z}^{n_1}_{\geq 0}.$$

As before, we denote infinite dimensional vectors $\tilde{\rho}(t_{k+1},\mathcal{X})\triangleq(\tilde{\rho}(t_{k+1},x_1),\tilde{\rho}(t_{k+1},x_2),\dots)^\top$, and $\tilde{\rho}_{\text{FFSP}}(t_{k+1},\mathcal{X})\triangleq(\tilde{\rho}_{\text{FFSP}}(t_{k+1},x_1),\tilde{\rho}_{\text{FFSP}}(t_{k+1},x_2),\dots)^\top$. Then, by (17), we almost surely have

$$\|\tilde{\rho}_{\text{FFSP}}(t_{k+1},\mathcal{X})-\tilde{\rho}(t_{k+1},\mathcal{X})\|\leq\bar{a}^{\mathcal{O}_{k+1}}\left[\epsilon(t_k)+\tilde{\epsilon}(t)\right]$$

where $\bar{a}^{\mathcal{O}_{k+1}}\triangleq\sup_{x\in\mathbb{Z}^{n_1}_{\geq 0}}\left\{\sum_{j\in\mathcal{O}_{k+1}}a_j(x,\mathbf{Y}(t_k))\right\}$. Then, similar to the analysis in (18), the error between the normalized filters is almost surely given by

$$
\begin{aligned}
&\|\pi_{\text{FFSP}}(t_{k+1},\mathcal{X})-\pi(t_{k+1},\mathcal{X})\|_1\\
\leq &\frac{\|\tilde{\rho}_{\text{FFSP}}(t_{k+1},\mathcal{X})-\tilde{\rho}(t_{k+1},\mathcal{X})\|_1}{\|\tilde{\rho}_{\text{FFSP}}(t_{k+1},\mathcal{X})\|_1}+\frac{\left|\|\tilde{\rho}_{\text{FFSP}}(t_{k+1},\mathcal{X})\|_1-\|\tilde{\rho}(t_{k+1},\mathcal{X})\|_1\right|}{\|\tilde{\rho}_{\text{FFSP}}(t_{k+1},\mathcal{X})\|_1}\\
\leq &\frac{2\bar{a}^{\mathcal{O}_{k+1}}\left[\epsilon(t_k)+\tilde{\epsilon}(t^-_{k+1})\right]}{\|\tilde{\rho}_{\text{FFSP}}(t_{k+1},\mathcal{X})\|_1}
\end{aligned}
\tag{19}
$$

By combining (18), (19), and the fact that the L_1 distance of two probability distribution cannot exceed 2, we prove the result. □

S.2.3 Error analysis for the second FFSP algorithm

Proof of theorem 4. The second part of this theorem is a direct consequence of the first one, because with the dominance property, it is very easy to show the result. Therefore, we only prove the first part of this theorem, and we prove it using math induction.

Obviously, the first part of this theorem holds at time t_0. Then, we prove that once this result holds at time t_k ($k = 1, 2, \dots$), it also holds in the time interval $t \in (t_k, t_{k+1}]$. By definition, we almost surely have

$$\begin{aligned}
\rho(t, \mathcal{X}_J) &= e^{\mathbb{A}_J(\mathbf{Y}(t_k))\,(t-t_k)}\,\pi(t_k, \mathcal{X}_J) + \int_{t_k}^{t} e^{\mathbb{A}_J(\mathbf{Y}(t_k))\,(t-s)}\,\mathbb{A}_{JJ'}(\mathbf{Y}(t_k))\,\rho(s, \mathcal{X}_{J'})\,ds \\
&\geq e^{\mathbb{A}_J(\mathbf{Y}(t_k))\,(t-t_k)}\,\pi(t_k, \mathcal{X}_J) \\
&\geq e^{\mathbb{A}_J(\mathbf{Y}(t_k))\,(t-t_k)}\,\pi_{FSP}(t_k, \mathcal{X}_J) \qquad \Big(= \rho_{\mathrm{FFSP}}(t, \mathcal{X}_J)\Big) \qquad \forall t \in (t_k, t_{k+1}).
\end{aligned} \tag{20}$$

where the inequalities holds element-wisely, and the last two lines follow from the non-negativity of all the terms above. This gives the dominance property between the unnormalized filter and its FFSP counterpart. Then, we estimate the normalization factors. For every $t \in (t_k, t_{k+1})$, there almost surely holds

$$\begin{aligned}
\|\rho(t, \mathcal{X})\|_1 &= \|\pi(t_k, \mathcal{X})\|_1 + \int_{t_k}^{t} \mathbf{1}^\top \mathbb{A}(\mathbf{Y}(t_k))\rho(s, \mathcal{X})ds \qquad \text{(by definition \& Fubini's theorem)} \\
&\leq 1 + \int_{t_k}^{t} \mathbf{1}^\top \mathbb{A}(\mathbf{Y}(t_k))\rho_{\mathrm{FFSP}}(s, \mathcal{X})ds \\
&= c_{\mathrm{FFSP}}(t)
\end{aligned} \tag{21}$$

where the second line follows from (20) and the non-positivity of $1^\top \mathbb{A}(\mathbf{Y}(t_k))$ (c.f. its definition). Also, we have that

$$\begin{aligned}
&\sum_{x \in \mathbb{Z}_{\geq 0}^{n_1}} \sum_{j \in \mathcal{O}_{k+1}} a_j(x, \mathbf{Y}(t_k))\rho(t_{k+1}^-, x) \\
&= \sum_{x \in \mathbb{Z}_{\geq 0}^{n_1}} \sum_{j \in \mathcal{O}_{k+1}} a_j(x, \mathbf{Y}(t_k))\rho_{\mathrm{FFSP}}(t_{k+1}^-, x) \\
&\quad + \sum_{x \in \mathbb{Z}_{\geq 0}^{n_1}} \left[\sum_{j \in \mathcal{O}_{k+1}} a_j(x, \mathbf{Y}(t_k))\right] \left[\rho(t_{k+1}^-, x) - \rho_{\mathrm{FFSP}}(t_{k+1}^-, x)\right] \\
&\leq \sum_{x \in \mathcal{X}_J} \sum_{j \in \mathcal{O}_{k+1}} a_j(x, \mathbf{Y}(t_k))\rho_{\mathrm{FFSP}}(t_{k+1}^-, x) + a^{\mathcal{O}_{i+1}} \left[c_{\mathrm{FFSP}}(t_{k+1}^-) - \left\|\rho_{\mathrm{FFSP}}(t_{k+1}^-, \mathcal{X})\right\|_1\right] \\
&= c_{\mathrm{FFSP}}(t_{k+1})
\end{aligned} \tag{22}$$

where the inequality follows from (20). Finally, the normalized filter almost surely satisfies

$$\pi(t, x) = \frac{\rho(t, x)}{\|\rho(t, \mathcal{X})\|_1} \geq \frac{\rho_{\mathrm{FFSP}}(t, x)}{c_{\mathrm{FFSP}}(t)} = \pi_{\mathrm{FFSP}}(t, x) \qquad \forall t \in (t_k, t_{k+1}), \qquad \text{(due to (20) and (21))}$$

and

$$\pi(t_{k+1}, x) = \frac{\rho(t_{k+1}, x)}{\sum_{x \in \mathbb{Z}_{\geq 0}^{n_1}} \sum_{j \in \mathcal{O}_{k+1}} a_j(x, \mathbf{Y}(t_k))\rho(t_{k+1}^-, x)} \geq \frac{\rho_{\mathrm{FFSP}}(t_{k+1}, x)}{c_{\mathrm{FFSP}}(t_{k+1})} = \pi_{\mathrm{FFSP}}(t_{k+1}, x), \qquad \text{(due to (20) and (22))}$$

which suggests that the first part of the theorem holds in $(t_k, t_{k+1}]$. Therefore, we prove this theorem. □

S.3 Derivation of the Kalman filter for the simple transcription-translation network

Now, we derive the Kalman filter for the simple transcription-translation network. This filter is built on top of the linear noise approximation, which is introduced as follows. We first denote scaled version of the hidden species (the mRNA) and the observed species (the protein) as the following:

$$X^{\Omega}(t) \triangleq \frac{\mathbf{X}(t)}{\Omega} \qquad \text{and} \qquad Y^{\Omega}(t) \triangleq \frac{\mathbf{Y}(t)}{\Omega},$$

where Ω is the scaling factor. When the scaling factor Ω is large, the dynamics of the scaled state can be approximated by [8]

$$X^{\Omega}(t) \approx x(t) + \frac{L_1(t)}{\sqrt{\Omega}} \qquad \text{and} \qquad Y^{\Omega}(t) \approx y(t) + \frac{L_2(t)}{\sqrt{\Omega}} \tag{23}$$

where $(x(t), y(t))$ satisfies the deterministic dynamics of the chemical reacting system

$$\begin{cases} \dot{x}(t) = \frac{c_2}{\Omega} - c_3 x(t) \\ \dot{y}(t) = c_1 x(t) \end{cases}, \tag{24}$$

and $L_1(t)$ and $L_2(t)$ satisfy stochastic differential equations

$$\begin{cases} dL_1(t) = -c_3 L_1(t) + \sqrt{\frac{c_2}{\Omega}} dW_1(t) - \sqrt{c_3 x(t)} dW_2(t) \\ dL_2(t) = c_1 L_1(t) + \sqrt{c_1 x(t)} dW_3(t) \end{cases}. \tag{25}$$

This approximation is called the linear noise approximation because the stochastic differential equations (25) is linear. This approximation is accurate when the scaling factor Ω is large, and $\frac{c_2}{\Omega}$ is of constant order.

Now, we build the Kalman filter based on this linear noise approximation. Note that $x(t)$ and $y(t)$ are deterministic, so we can build the filter purely by looking at the stochastic parts $L_1(t)$ and $L_2(t)$. Then, by the Kalman-Bucy filter [2], the conditional probability distribution of $L_1(t)$ given the trajectory of $L_2(t)$ (up to time t) is Gaussian; its conditional mean $\mu(t)$ and variance $\Sigma(t)$ evolve according to

$$d\mu(t) = -c_3\mu(t)dt + \Sigma(t)x^{-1}(t)\left[dL_2(t) - c_1\mu(t)dt\right] = -c_3\mu(t)dt + \Sigma(t)x^{-1}(t)\left[d(Y^{\Omega}(t) - y(t)) - c_1\mu(t)dt\right]$$
$$\dot{\Sigma}(t) = -2c_3\Sigma(t) + \frac{c_2}{\Omega} + c_3 x(t) - \Sigma^2(t)c_1 x^{-1}(t).$$

Consequently, by (23), we can estimate the mean and variance of $X(t)$ by

$$\mathbb{E}\left[X(t) \middle| Y(s), 0 \le s \le t\right] \approx x(t) + \frac{\mu(t)}{\sqrt{\Omega}} \qquad \text{and} \qquad \mathrm{Var}\left(X(t) \middle| Y(s), 0 \le s \le t\right) \approx \frac{\Sigma(t)}{\Omega},$$

which is the Kalman filter for the transcription-translation model.

References

[1] D. F. Anderson and T. G. Kurtz, *Stochastic analysis of biochemical systems*, vol. 674, Springer, 2015.

[2] A. Bain and D. Crisan, *Fundamentals of stochastic filtering*, vol. 3, Springer, 2009.

[3] E. A. Coddington and N. Levinson, *Theory of ordinary differential equations*, Tata McGraw-Hill Education, 1955.

[4] F. Confortola and M. Fuhrman, *Filtering of continuous-time markov chains with noise-free observation and applications*, Stochastics An International Journal of Probability and Stochastic Processes, 85 (2013), pp. 216–251.

[5] M. H. Davis, *The representation of martingales of jump processes*, SIAM Journal on control and optimization, 14 (1976), pp. 623–638.

[6] R. L. Karandikar and B. Rao, *On quadratic variation of martingales*, Proceedings-Mathematical Sciences, 124 (2014), pp. 457–469.

[7] M. Rathinam and M. Yu, *State and parameter estimation from exact partial state observation in stochastic reaction networks*, The Journal of Chemical Physics, 154 (2021), p. 034103.

[8] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, vol. 1, Elsevier, 1992.