

Supporting Information for Reconstructing dynamic molecular states from single-cell time series

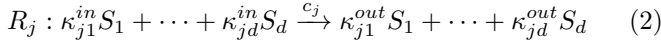
Lirong Huang, Loic Pauleve, Christoph Zechner, Michael Unger, Anders S. Hansen, Heinz Koeppl

S.1 Continuous-time Markov chains

Let us consider a well-stirred chemical reaction system in thermal equilibrium, where the volume of the spatial domain is constant. The state vector of the system, which involves d different species of particles S_1, \dots, S_d ,

$$X(t) = (X_1(t) \ \cdots \ X_d(t))^T \quad (1)$$

of particle numbers changes according to r reaction channels R_1, \dots, R_r , which represent events such as, e.g., a reactive collision between two particles from different species that produces a particle of another type, the decomposition of a particle into two particles, the conversion of a particle from one species to another one, the inflow of new particles into the systems from an exterior source, the "death" of particles and so on. Each reaction channel is described by



with nonnegative integers $\kappa_{jk}^{in}, \kappa_{jk}^{out}$ and $j \in \{1, 2, \dots, r\}$. If R_j fires at time t , then the particle numbers jump from the old state $X(t-)$ to the new state $X(t) = X(t-) + \nu_j$ where

$$\nu_j = (\nu_{j1} \ \cdots \ \nu_{jd})^T = (\kappa_{j1}^{out} - \kappa_{j1}^{in} \ \cdots \ \kappa_{jd}^{out} - \kappa_{jd}^{in})^T$$

is called the stoichiometric vector of R_j . The time of the next reaction event and the index of the reaction channel that fires are both random, and depend only on the current state of the system such that $X(t)$ is a realization of continuous-time Markov chain (CTMC)

$$X(t) = X(0) + \sum_{j=1}^r \xi_j \left(\int_0^t \alpha_j(X(s)) ds \right) \nu_j \quad (3)$$

with (density-dependent) Poisson processes ξ_j and initial state $X(0) \in \mathcal{X}$, where $\mathcal{X} \subseteq \mathbb{N}^d$, $\mathbb{N} = \{0, 1, 2, \dots\}$, is the countable set of states of CTMC (3). The function α_j is called the propensity of reaction R_j and is typically defined as

$$\alpha_j(x) = c_j \Pi_{q=1}^d \binom{x_q}{\kappa_{jq}^{in}} \quad (4)$$

with a reaction constant $c_j > 0$ dependent on physical properties such as the temperature or the volume of the domain,

where

$$\binom{m}{q} = \begin{cases} \frac{1}{q!} \Pi_{j=0}^{q-1} (m-j), & m \geq q \\ 0, & m < q \\ 1, & q = 0 \end{cases} \quad (5)$$

denotes the generalized binomial coefficient.

As a typical interpretation of the model (3), one may consider a chemical reaction system where the particles of the species correspond to the molecules of the chemical substances. But the same setting can be used to model discrete stochastic predator-prey systems, viral kinetics, gene regulatory networks and so on, see, e.g., [24, 15, 18] and the references therein.

S.2 Chemical master equation

The probability distribution of the states of CTMC (3) can be described by a system of differential equations. The probability distribution $p(x, t) = \mathbb{P}(X(t) = x)$ of state $X(t)$ at time t is the solution of chemical master equation (CME)

$$\frac{dp(x, t)}{dt} = \sum_{j=1}^r [p(x - \nu_j, t) \alpha_j(x - \nu_j) - p(x, t) \alpha_j(x)] \quad (6)$$

for $x \in \mathcal{X}$, $t \geq 0$ and initial condition $p(x, 0) = \mathbb{P}(X(0) = x)$ (see [11]) with the convention that $p(x, t) = 0$ for all $x \notin \mathcal{X}$.

Typically, the CME is too high-dimensional to deal with computationally. The stochastic simulation algorithm (SSA) gets around this issue by computing single realisations of the state vector rather than an entire probability distribution, see [9, 10].

Denote the transition probabilities of $X(t)$ by

$$p_{zx}(h) = \mathbb{P}(X(t+h) = x \mid X(t) = z). \quad (7)$$

Therefore,

$$p_{zx}(0) = \delta(x - z) = \begin{cases} 1, & x = z \\ 0, & x \neq z. \end{cases} \quad (8)$$

According to the CTMC (3), the transition probabilities (7) obey, as $h \rightarrow 0$,

$$p_{zx}(h) = \begin{cases} 1 - \sum_{j=1}^r \alpha_j(z) h + o(h), & x = z \\ \alpha_j(z) h + o(h), & x = z + \nu_j, 1 \leq j \leq r \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

To avoid triviality, assume that $p(x, 0)$ is not concentrated on any absorbing state.

Let Φ be the transition rate matrix of the CTMC (3) whose entries are given by

$$\Phi_{zx} = \begin{cases} -\sum_{j=1}^r \alpha_j(z), & x = z \\ \alpha_j(z) & x = z + \nu_j, 1 \leq j \leq r \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

Write $p(t)$ as a row vector with components $\{p(x, t)\}_{x \in \mathcal{X}}$. Then the CME (6) can be written in a compact form

$$\frac{dp(t)}{dt} = p(t)\Phi, \quad t \geq 0 \quad (11)$$

with initial condition $p(0)$. Note that Φ is infinite dimensional when the state space \mathcal{X} is unbounded, or even when \mathcal{X} is bounded, Φ is usually a big and sparse matrix.

S.3 Wonham filter

For the CTMC (3) of a chemical reaction system, suppose that the observation process $\{Y(t), t \geq 0\}$ is given by

$$dY(t) = g(X(t))dt + B(t)dW(t) \quad (12)$$

for $t \geq 0$ with $Y(0) = 0$, where $y \in \mathbb{R}^l$ (usually $1 \leq l \leq d$), $g : \mathbb{R}^d \rightarrow \mathbb{R}^l$ is a continuous function, $B(t) \in \mathbb{R}^{l \times l}$ is a full-rank matrix with bounded norms of $B(t)$ and $B^{-1}(t)$ for all $t \geq 0$ and $W(t)$ is the standard l -dimensional Brownian motion independent of process $X(t)$. The posterior probabilities

$$\pi(x, t) = \mathbb{P}(X(t) = x \mid y(s), s \in [0, t]), \quad x \in \mathcal{X}. \quad (13)$$

Each $\pi(x, t)$ is described with a stochastic differential equation (SDE). The Wonham filter gives the following evaluation of $\pi(x, t)$, which is obtained with few modification from [43]

$$d\pi(x, t) = \sum_{j=1}^r [\pi(x - \nu_j, t)\alpha_j(x - \nu_j) - \pi(x, t)\alpha_j(x)] dt + \pi(x, t)[g(x) - \bar{g}(t)]^T [B(t)B^T(t)]^{-1} [dy(t) - \bar{g}(t)dt] \quad (14)$$

for $x \in \mathcal{X}$ with initial condition $\pi(x, 0) = p(x, 0) = \mathbb{P}\{X(0) = x\}$, where

$$\bar{g}(t) = \sum_{x \in \mathcal{X}} g(x)\pi(x, t). \quad (15)$$

This is a general form of the main result Eq.(21) in [43], which is a special case of (14) with $l = 1$ and $g(a) = a$. Due to the expectation (15), it is cumbersome to calculate (14). One often resorts to the unnormalized version of $\pi(x, t)$ and performs normalization after numerical integration. The unnormalized filter $V(x, t)$ is the unnormalized version of the Wonham filter (14), which is given as follows

$$dV(x, t) = \sum_{j=1}^r [V(x - \nu_j, t)\alpha_j(x - \nu_j) - V(x, t)\alpha_j(x)] dt + V(x, t)g^T(x)[B(t)B^T(t)]^{-1}dy(t), \quad x \in \mathcal{X} \quad (16)$$

with initial condition $V(x, 0) = p(x, 0) = \mathbb{P}\{X(0) = x\}$.

S.4 CTMC with discrete-time observations

In the previous section, the observations (12) are described by a continuous-time process, which means that information about the system is available continuously. But this does not fit well with many practical situations. The dynamics of chemical reaction systems are naturally continuous, and, however, observations are usually taken at discrete time instants. In this section, we consider filtering and smoothing problems of CTMC (3) with discrete-time observations.

S.4.1 Filter for CTMC with discrete-time observations

Suppose that observations at discrete times $y_1 = y(t_1), \dots, y_N = y(t_N)$ with $0 < t_1 < \dots < t_N = T$ admit the conditional distribution of the form

$$p(y_i | X(s), s \in [0, t_i]) = p(y_i | X(t_i)) \quad (17)$$

for $1 \leq i \leq N$. The posterior probabilities are denoted by

$$\pi(x, t) = \mathbb{P}(X(t) = x \mid y_1, \dots, y_n) \quad (18)$$

with $n = \max\{n \in \mathbb{N} \mid t_n \leq t\}$ for all $x \in \mathcal{X}$ and $t \geq 0$.

Write $\pi(t)$, $t \geq 0$, as a row vector with components $\{\pi(x, t)\}_{x \in \mathcal{X}}$. We now seek a system of differential equations for $\pi(t)$.

Theorem 1: The evaluation of $\pi(t)$ is given by recursive CME

$$\frac{d\pi(x, t)}{dt} = \sum_{j=1}^r [\pi(x - \nu_j, t)\alpha_j(x - \nu_j) - \pi(x, t)\alpha_j(x)] \quad (19)$$

for $x \in \mathcal{X}$ with $\pi(x, 0) = \mathbb{P}(X(0) = x)$ and reset conditions

$$\pi(x, t_i) = C_i^{-1} \pi(x, t_i^-) p(y_i | x) \quad (20)$$

at the observation times $t_i \leq t_n$, where C_i is a normalizing constant

$$C_i = \sum_{z \in \mathcal{X}} \pi(z, t_i^-) p(y_i | z). \quad (21)$$

Proof: Note that Eq. (19) is exactly the same as the CME (6) except the reset conditions (20) at the observation times. Notice that derivation of the CME (6) can be found in many references, e.g., [11]. Therefore, we just need to show the reset conditions (20) at the observation times. According to

Bayesian formula, the reset condition (19) at the observation time t_i is

$$\begin{aligned}
 \pi(x, t_i) &= \mathbb{P}(X(t_i) = x \mid y_1, \dots, y_i) \\
 &= \frac{\mathbb{P}(\{y_1, \dots, y_i\} \cap X(t_i) = x)}{\mathbb{P}(\{y_1, \dots, y_i\})} \\
 &= \frac{\mathbb{P}(y_i \mid \{y_1, \dots, y_{i-1}\} \cap X(t_i) = x)}{\mathbb{P}(y_i \mid \{y_1, \dots, y_{i-1}\})} \\
 &\quad \cdot \frac{\mathbb{P}(\{y_1, \dots, y_{i-1}\} \cap X(t_{q+1}) = x)}{\mathbb{P}(\{y_1, \dots, y_{i-1}\})} \\
 &= \frac{p(y_i|x)\pi(x, t_i^-)}{\sum_{z \in \mathcal{X}} p(y_i|z)\pi(z, t_i^-)} \\
 &= C_i^{-1} \pi(x, t_i^-) p(y_i|x)
 \end{aligned} \tag{22}$$

with $C_i = \sum_{z \in \mathcal{X}} \pi(z, t_i^-) p(y_i|z)$, which is exactly the reset condition (20). This completes the proof. \square

Similarly, the system of ODEs (19) can be written in a compact form

$$\frac{d\pi(t)}{dt} = \pi(t)\Phi \tag{23}$$

with $\pi(0) = p(0)$ and reset conditions

$$\pi(t_i) = C_i^{-1} \pi(t_i^-) p(y_i|x), \tag{24}$$

where Φ is the transition rate matrix given by (10).

S.4.2 Smoother for CTMC with discrete-time observations

Suppose that we have obtained the posterior probability $\pi(x, t)$ for $t \in [0, T]$ with $T = t_N$. Let us now study the smoothing (probability) density denoted by

$$\tilde{\pi}(x, t) = \mathbb{P}(X(t) = x \mid y_1, \dots, y_N) \tag{25}$$

for all $x \in \mathcal{X}$ and $t \in [0, T]$. By (18) and (25), we have $\tilde{\pi}(x, T) = \pi(x, T)$ for all $x \in \mathcal{X}$.

For all $x \in \mathcal{X}$, define $\beta(x, t)$ by

$$\beta(x, t) = \frac{\mathbb{P}(y_k, \dots, y_N \mid X(t) = x)}{\mathbb{P}(y_k, \dots, y_N \mid y_1, \dots, y_{k-1})} \tag{26}$$

for all $t \in [0, T]$, where $k = \min\{k \in \mathbb{N} \mid t_k > t\}$. Obviously,

$$\beta(x, T) = 1, \quad \forall x \in \mathcal{X}. \tag{27}$$

We have the following result for smoothing of CTMC (3) with observations $\{y_1, \dots, y_N\}$.

Theorem 2: The evaluation of smoothing density (25) is given by

$$\tilde{\pi}(x, t) = \beta(x, t)\pi(x, t) \tag{28}$$

for $t \in [0, T]$, where $\pi(x, t)$ is given by (18) above. Moreover, $\beta(x, t)$ satisfies the backward equation

$$\frac{d\beta(x, t)}{dt} = \sum_{j=1}^r [\beta(x, t) - \beta(x + \nu_j, t)] \alpha_j(x) \tag{29}$$

with reset conditions

$$\beta(x, t_i^-) = C_i^{-1} \beta(x, t_i) p(y_i|x) \tag{30}$$

at the observation times $t_i \leq T$ and terminal conditions (27), where normalizing constant C_i is given by (21).

Proof: For $t \in [0, T]$, by definitions (18), (25) and (26), we have

$$\begin{aligned}
 \tilde{\pi}(x, t) &= \frac{\mathbb{P}(\{y_1, \dots, y_N\} \cap X(t) = x)}{\mathbb{P}(\{y_1, \dots, y_N\})} \\
 &= \frac{\mathbb{P}(\{y_k, \dots, y_N\} \cap \{y_1, \dots, y_{k-1}\} \cap X(t) = x)}{\mathbb{P}(\{y_k, \dots, y_N\} \cap \{y_1, \dots, y_{k-1}\})} \\
 &= \frac{\mathbb{P}(\{y_k, \dots, y_N\} \mid \{y_1, \dots, y_{k-1}\} \cap X(t) = x)}{\mathbb{P}(\{y_k, \dots, y_N\} \mid \{y_1, \dots, y_{k-1}\})} \\
 &\quad \cdot \frac{\mathbb{P}(\{y_1, \dots, y_{k-1}\} \cap X(t) = x)}{\mathbb{P}(\{y_1, \dots, y_{k-1}\})} \\
 &= \frac{\mathbb{P}(\{y_k, \dots, y_N\} \mid X(t) = x)}{\mathbb{P}(\{y_k, \dots, y_N\} \mid \{y_1, \dots, y_{k-1}\})} \pi(x, t) \\
 &= \beta(x, t)\pi(x, t).
 \end{aligned} \tag{31}$$

That is, (28) holds for all $t \in [0, T]$.

Let us proceed to show that $\beta(x, t)$ satisfies the backward equation (29) with reset conditions (30) and terminal conditions (27). Since, by definition (26), terminal conditions (27) hold, we consider reset conditions (30) at observation time t_i when, without loss of generality, $\beta(x, t_i)$, $x \in \mathcal{X}$, are available. For small $h > 0$, we have

$$\begin{aligned}
 \beta(x, t_i - h) &= \frac{\mathbb{P}(\{y_i, \dots, y_N\} \mid X(t_i - h) = x)}{\mathbb{P}(\{y_i, \dots, y_N\} \mid \{y_1, \dots, y_{i-1}\})} \\
 &= \frac{\sum_{u \in \mathcal{X}} \mathbb{P}(\{y_i, \dots, y_N\} \mid X(t_i) = u \cap X(t_i - h) = x) p_{xu}(h)}{\mathbb{P}(\{y_{i+1}, \dots, y_N\} \mid \{y_1, \dots, y_i\}) \mathbb{P}(y_i \mid \{y_1, \dots, y_{i-1}\})} \\
 &= \frac{\sum_{u \in \mathcal{X}} \mathbb{P}(\{y_i, \dots, y_N\} \mid X(t_i) = u) p_{xu}(h)}{\mathbb{P}(\{y_{i+1}, \dots, y_N\} \mid \{y_1, \dots, y_i\}) \mathbb{P}(y_i \mid \{y_1, \dots, y_{i-1}\})}.
 \end{aligned}$$

Note that, by (22), $\mathbb{P}(y_i \mid \{y_1, \dots, y_{i-1}\}) = C_i$. And

$$\begin{aligned}
 \mathbb{P}(\{y_i, \dots, y_N\} \mid X(t_i) = u) \\
 = \mathbb{P}(\{y_{i+1}, \dots, y_N\} \mid X(t_i) = u) p(y_i|u).
 \end{aligned}$$

Therefore, by (26), we have

$$\begin{aligned}
 \beta(x, t_i - h) &= C_i^{-1} \\
 &\quad \cdot \sum_{u \in \mathcal{X}} \frac{\mathbb{P}(\{y_{i+1}, \dots, y_N\} \mid X(t_i) = u)}{\mathbb{P}(\{y_{i+1}, \dots, y_N\} \mid \{y_1, \dots, y_i\})} p_{xu}(h) p(y_i|u) \\
 &= C_i^{-1} \sum_{u \in \mathcal{X}} \beta(u, t_i) p_{xu}(h) p(y_i|u)
 \end{aligned} \tag{32}$$

Letting $h \rightarrow 0$ in (32), by (8), yields

$$\beta(x, t_i^-) = C_i^{-1} \beta(x, t_i) p(y_i|x)$$

which is the reset conditions (30).

To show (29), assume that, without loss of generality, $t < t_i$ and there is no observation time on the interval $[t-h, t]$ for some small $h > 0$. Then we have

$$\begin{aligned}
 \beta(x, t-h) &= \frac{\mathbb{P}(\{y_i, \dots, y_N\} \mid X(t-h) = x)}{\mathbb{P}(\{y_i, \dots, y_N\} \mid \{y_1, \dots, y_{i-1}\})} \\
 &= \frac{\mathbb{P}(\{y_i, \dots, y_N\} \cap X(t-h) = x)}{\mathbb{P}(\{y_i, \dots, y_N\} \cap X(t-h) = x)} \\
 &= \frac{\mathbb{P}(\{y_i, \dots, y_N\} \mid \{y_1, \dots, y_{i-1}\}) \mathbb{P}(X(t-h) = x)}{\mathbb{P}(\{y_i, \dots, y_N\} \mid \{y_1, \dots, y_{i-1}\}) \mathbb{P}(X(t-h) = x)} \\
 &= \sum_{u \in \mathcal{X}} \frac{\mathbb{P}(\{y_i, \dots, y_N\} \mid X(t) = u \cap X(t-h) = x)}{\mathbb{P}(\{y_i, \dots, y_N\} \mid \{y_1, \dots, y_{i-1}\})} \\
 &\quad \cdot \frac{\mathbb{P}\{X(t) = u \cap X(t-h) = x\}}{\mathbb{P}\{X(t-h) = x\}} \\
 &= \sum_{u \in \mathcal{X}} \frac{\mathbb{P}(\{y_i, \dots, y_N\} \mid X(t) = u)}{\mathbb{P}(\{y_i, \dots, y_N\} \mid \{y_1, \dots, y_{i-1}\})} p_{xu}(h) \\
 &= \sum_{u \in \mathcal{X}} \beta(u, t) p_{xu}(h) \tag{33}
 \end{aligned}$$

and therefore

$$\begin{aligned}
 \beta(x, t) - \beta(x, t-h) &= \beta(x, t)[1 - p_{xx}(h)] - \sum_{u \in \mathcal{X}, u \neq x} \beta(u, t) p_{xu}(h). \tag{34}
 \end{aligned}$$

This with (9) yields

$$\frac{d\beta(x, t)}{dt} = \sum_{j=1}^r [\beta(x, t) - \beta(x + \nu_j, t)] \alpha_j(x), \tag{35}$$

which is (29). The proof is complete. \square

Write $\beta(t)$, $t \in [0, T]$, as a row vector with components $\{\beta(x, t)\}_{x \in \mathcal{X}}$. Similarly, the backward ODE (29) with reset conditions (30) can be written in a compact form

$$\frac{d\beta(t)}{dt} = \beta(t) \Phi^T \tag{36}$$

with reset conditions at observation time t_i

$$\beta(t_i^-) = C_i^{-1} \beta(t_i) p(y_i | x), \tag{37}$$

where Φ is the transition matrix given by (10).

Note that the conditional process is also a Markovian process with time-dependent transition probability. Denote by $\tilde{\alpha}_j(x, t)$, $j = 1, \dots, r$, the posterior transition function of the conditional process. By (19) and (29), differentiating

$\tilde{\pi}(x, t)$ with respect to t yields the smoothing CME

$$\begin{aligned}
 \frac{d}{dt} \tilde{\pi}(x, t) &= \frac{d}{dt} [\beta(x, t) \pi(x, t)] \\
 &= \left[\frac{d}{dt} \beta(x, t) \right] \pi(x, t) + \beta(x, t) \left[\frac{d}{dt} \pi(x, t) \right] \\
 &= \sum_{j=1}^r [\beta(x, t) - \beta(x + \nu_j, t)] \alpha_j(x) \pi(x, t) \\
 &\quad + \sum_{j=1}^r \beta(x, t) [\pi(x - \nu_j, t) \alpha_j(x - \nu_j) - \pi(x, t) \alpha_j(x)] \\
 &= \sum_{j=1}^r [\pi(x - \nu_j, t) \alpha_j(x - \nu_j) \beta(x, t) \\
 &\quad - \pi(x, t) \alpha_j(x) \beta(x + \nu_j, t)] \\
 &= \sum_{j=1}^r \left[\tilde{\pi}(x - \nu_j, t) \alpha_j(x - \nu_j) \frac{\beta(x, t)}{\beta(x - \nu_j, t)} \right. \\
 &\quad \left. - \tilde{\pi}(x, t) \alpha_j(x) \frac{\beta(x + \nu_j, t)}{\beta(x, t)} \right] \\
 &= \sum_{j=1}^r [\tilde{\pi}(x - \nu_j, t) \tilde{\alpha}_j(x - \nu_j, t) - \tilde{\pi}(x, t) \tilde{\alpha}_j(x, t)] \tag{38}
 \end{aligned}$$

with initial value $\tilde{\pi}(x, 0) = \mathbb{P}(X(0) = x)$ for all $x \in \mathcal{X}$, where

$$\tilde{\alpha}_j(x, t) = \alpha_j(x) \frac{\beta(x + \nu_j, t)}{\beta(x, t)}. \tag{39}$$

S.5 Moments of CTMC with discrete-time observations

S.5.1 Prior moment equations

Let $|\gamma| = \gamma_1 + \dots + \gamma_d$ for $\gamma \in \mathbb{N}^d$. The multivariate moment generating function for $X(t)$ is defined as, see [8, 35],

$$\begin{aligned}
 M_G(\theta, t) &\equiv \sum_{x \in \mathcal{X}} p(x, t) e^{\theta_1 x_1 + \dots + \theta_d x_d} \\
 &= \sum_{x \in \mathcal{X}} p(x, t) e^{\theta^T x} = \sum_{|\gamma|=0}^{\infty} \frac{\hat{M}^\gamma(t) \theta^\gamma}{\gamma!} \tag{40}
 \end{aligned}$$

for $t \geq 0$, where $\gamma! = \gamma_1! \dots \gamma_d!$, $\theta^\gamma = \theta_1^{\gamma_1} \dots \theta_d^{\gamma_d}$ and

$$\begin{aligned}
 \hat{M}^\gamma(t) &= \mathbb{E}_p[X_1^{\gamma_1}(t) \dots X_d^{\gamma_d}(t)] \\
 &= \mathbb{E}_p[X^\gamma(t)] = \sum_{x \in \mathcal{X}} p(x, t) x^\gamma. \tag{41}
 \end{aligned}$$

Clearly, $\hat{M}^\gamma(t) \equiv 1$ for all $t \geq 0$ if $|\gamma| = 0$. For convenience, denote the mean and the covariance by $\hat{M}(t) = \hat{M}^1(t)$ and $\hat{M}^2(t)$, that is, $\hat{M}(t) = \hat{M}^1(t) = \mathbb{E}_p[X(t)]$ and $\hat{M}^2(t) =$

$\mathbb{E}_p[X(t)X^T(t)]$, respectively. The moments of $X(t)$ are given by (see, e.g., [35])

$$\hat{M}^\gamma(t) = d^\gamma M_G(\theta, t)|_{\theta=0^d} = \frac{\partial^{|\gamma|} M_G(\theta, t)}{\partial \theta_1^{\gamma_1} \dots \partial \theta_d^{\gamma_d}} \Big|_{\theta=0^d}. \quad (42)$$

The propensity function $\alpha_j(x)$ of reaction R_j can be written of the form, see, e.g., [8],

$$\alpha_j(x) = \sum_{|\gamma| \geq 0} c_{j,\gamma} x^\gamma \quad (43)$$

with $c_{j,\gamma} \geq 0$ for all j and γ . Let $\eta \in \mathbb{N}^d$ and $\zeta \in \mathbb{N}^d$, then $\zeta \geq \eta$ (resp. $\zeta < \eta$) means that $\zeta_j \leq \eta_j$ (resp. $\zeta_j < \eta_j$) for all $1 \leq j \leq d$. It is observed that, see, e.g., [8],

$$\frac{\partial M_G(\theta, t)}{\partial t} = \sum_{|\eta| > 0} \frac{\theta^\eta}{\eta!} \sum_{j=1}^r \sum_{|\gamma| \geq 0} c_{j,\gamma} \sum_{|\zeta| > 0, \zeta \leq \eta} \nu_j^\zeta \binom{\eta}{\zeta} M_p^{\eta-\zeta+\gamma}(t)$$

where

$$\nu_j^\zeta \binom{\eta}{\zeta} = \nu_{j1}^{\zeta_1} \binom{\eta_1}{\zeta_1} \times \dots \times \nu_{jd}^{\zeta_d} \binom{\eta_d}{\zeta_d}.$$

This with (41) gives

$$\frac{d\hat{M}^\eta(t)}{dt} = \sum_{j=1}^r \sum_{|\gamma| \geq 0} c_{j,\gamma} \sum_{|\zeta| > 0, \zeta \leq \eta} \nu_j^\zeta \binom{\eta}{\zeta} \hat{M}^{\eta-\zeta+\gamma}(t) \quad (44)$$

with initial value $\hat{M}^\eta(0) = \mathbb{E}_p[X^\eta(0)]$. Particularly, let $\eta = e_i$ (i.e., $\hat{M}^{e_i}(t) = \mathbb{E}_p[X_i(t)]$), where $e_i \in \mathbb{N}^d$ with the i th element being 1 and the others 0. Then one has

$$\begin{aligned} \frac{d\hat{M}^{e_i}(t)}{dt} &= \sum_{j=1}^r \sum_{|\gamma| \geq 0} c_{j,\gamma} \sum_{|\zeta| > 0, \zeta \leq e_i} \nu_j^\zeta \binom{e_i}{\zeta} \hat{M}^{e_i-\zeta+\gamma}(t) \\ &= \sum_{j=1}^r \sum_{|\gamma| \geq 0} c_{j,\gamma} \nu_{ji} \hat{M}^\gamma(t) \end{aligned} \quad (45)$$

with initial value $\hat{M}^{e_i}(0) = \mathbb{E}_p[X_i(0)]$ for all $1 \leq k \leq d$. This is the estimation of expectation in the references, e.g., [25],

$$\frac{d}{dt} \mathbb{E}_p[X(t)] = \sum_{j=1}^r \nu_j \mathbb{E}_p[\alpha_j(X(t))], \quad t \geq 0 \quad (46)$$

with initial value $\mathbb{E}_p[X(0)]$.

Particularly, for the first order chemical reaction networks, see [16],

$$\alpha_j(x) = \sum_{|\gamma|=0}^1 c_{j,\gamma} x^\gamma, \quad j = 1, \dots, r \quad (47)$$

given any $\bar{L} \geq 1$, the moment dynamics (44) yields a system of ODEs for the first order reaction network (47)

$$\frac{d\hat{M}^{\bar{L}}(t)}{dt} = \hat{C}_{\bar{L}} \hat{M}^{\bar{L}}(t) + b_{\bar{L}}, \quad t \geq 0 \quad (48)$$

with initial value $\hat{M}^{\bar{L}}(0)$, where $\hat{M}^{\bar{L}}(t) \in \mathbb{R}^{\hat{d}}$ with components $\{M^\gamma(t) : |\gamma| = 1, \dots, \bar{L}\}$ and \hat{d} being the cardinality of set $\{\gamma : |\gamma| = 1, \dots, \bar{L}\}$, $\hat{M}^{\bar{L}}(0) \in \mathbb{R}^{\hat{d}}$ with components $\{\mathbb{E}_p[X^\gamma(0)] : |\gamma| = 1, \dots, \bar{L}\}$, $\hat{C}_{\bar{L}} \in \mathbb{R}^{\hat{d} \times \hat{d}}$ and $b_{\bar{L}} \in \mathbb{R}^{\hat{d}}$ with entries given by (44). An example for (48) will be specified below. As is known, the solution of linear system (48) is given by

$$\hat{M}^{\bar{L}}(t) = e^{t\hat{C}_{\bar{L}}} \hat{M}^{\bar{L}}(0) + \int_0^t e^{(t-\tau)\hat{C}_{\bar{L}}} b_{\bar{L}} d\tau, \quad t \geq 0. \quad (49)$$

S.5.2 Posterior moment equations

The posterior moments are given as

$$M^\eta(t) = \mathbb{E}_\pi[X^\eta(t)] = \sum_{x \in \mathcal{X}} \pi(x, t) x^\eta \quad (50)$$

for all $\eta \in \mathbb{N}^d$ and $t \geq 0$. For convenience, denote the mean and the covariance by $M(t) = M^1(t)$ and $M^2(t)$, that is, $M(t) = M^1(t) = \mathbb{E}_\pi[X(t)]$ and $\hat{M}^2(t) = \mathbb{E}_\pi[X(t)X^T(t)]$, respectively.

According to Theorem 1, the evolution of posterior moments (50) are described by recursive moment equations

$$\frac{dM^\eta(t)}{dt} = \sum_{j=1}^r \sum_{|\gamma| \geq 0} c_{j,\gamma} \sum_{|\zeta| > 0, \zeta \leq \eta} \nu_j^\zeta \binom{\eta}{\zeta} M^{\eta-\zeta+\gamma}(t) \quad (51)$$

with initial condition $M^\eta(0) = M_p^\eta(0)$ and reset conditions

$$M^\eta(t_i) = C_i^{-1} \sum_{x \in \mathcal{X}} \pi(x, t_i^-) x^\eta p(y_i|x) \quad (52)$$

at observation time $t_i \leq t_n$ with C_i given by (21).

To calculate the value $M^\eta(t_i)$, one should know the conditional probability distribution function $p(y_i|x)$. In a very special case when $p(y_i|x) = \delta(y_i - x)$, it is easy to find $M^\eta(t_i) = y_i^\eta$. Generally, measurements are corrupted with noise. In fact, the update formula (52) for the moments derived from the filtering results (20) are usually intractable. Assume that $p(y_i|x)$ can be expressed as

$$p(y_i|x) = \sum_{|\gamma|=1}^\Gamma \sum_{q \geq 0} b_{\gamma,q}(y_i) x^{\gamma q}. \quad (53)$$

For instance, $y_i = X_1(t_i) + w_i$ with $w_i \sim \mathcal{N}(0, \sigma_i^2)$ independent of $X(t)$. In this case, $\Gamma = 1$ and

$$\begin{aligned} p(y_i|x) &= \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(y_i-x_1)^2}{2\sigma_i^2}} = \frac{1}{\sqrt{2\pi}\sigma_i} \sum_{j \geq 0} \frac{(-1)^j (y_i - x_1)^{2j}}{2^j \sigma_i^{2j} j!} \\ &= \frac{1}{\sqrt{2\pi}\sigma_i} \sum_{j \geq 0} \frac{(-1)^j}{2^j \sigma_i^{2j} j!} \sum_{q=0}^{2j} \binom{2j}{q} y_i^{2j-q} (-x_1)^q \\ &= \frac{1}{\sqrt{2\pi}\sigma_i} \sum_{q \geq 0} x_1^q \sum_{j \geq \lceil q/2 \rceil} \binom{2j}{q} \frac{(-1)^{j+q} y_i^{2j-q}}{2^j \sigma_i^{2j} j!} \end{aligned}$$

where $[a] = \min\{j : |a| \leq j, j \in \mathbb{N}\}$ for $a \in \mathbb{R}$, which is a specific case of (53) with

$$\gamma = e_1 \quad \text{and} \quad b_{e_1, q}(y_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \sum_{j \geq [q/2]} \binom{2j}{q} \frac{(-1)^{j+q} y_i^{2j-q}}{2^j \sigma_i^{2j} j!}.$$

Substitution of (53) into (52) and (21) yields

$$\begin{aligned} \sum_{x \in \mathcal{X}} \pi(x, t_i^-) x^\eta p(y_i | x) &= \sum_{|\gamma| > 0} \sum_{q \geq 0} b_{\gamma, q}(y_i) M^{\eta + \gamma q}(t_i^-), \\ C_i &= \sum_{z \in \mathcal{X}} \pi(z, t_i^-) p(y_i | z) = \sum_{|\gamma| > 0} \sum_{q \geq 0} b_{\gamma, q}(y_i) M^{\gamma q}(t_i^-) \end{aligned}$$

and hence

$$M^\eta(t_i) = \frac{\sum_{|\gamma| > 0} \sum_{q \geq 0} b_{\gamma, q}(y_i) M^{\eta + \gamma q}(t_i^-)}{\sum_{|\gamma| > 0} \sum_{l \geq 0} b_{\gamma, q}(y_i) M^{\gamma q}(t_i^-)}.$$

S.6 A novel approximate approach to filtering

The update formula (52) of moments derived from filtering theory is usually intractable. The challenging problem is to find some appropriate approximation to the update formula (52). In this section, we propose an approximate approach to the update formula by exploiting the fundamental relationship

$$\begin{aligned} \mathbb{E}[s(x(t_i))u(y_i) | \{y_1, \dots, y_{i-1}\}] \\ = \mathbb{E}[\mathbb{E}[s(x(t_i)) | \{y_1, \dots, y_i\}] u(y_i) | \{y_1, \dots, y_{i-1}\}] \end{aligned} \quad (54)$$

where $s(\cdot)$ and $u(\cdot)$ are integrable functions. This is a special form of tower property for conditional expectation, see, e.g., [19, 41].

In this section, we first present our approximate to filtering for univariate systems ($d = 1$) and then generalize it to multivariate systems ($d \geq 2$) in next section.

For simplicity, let us first consider a univariate system (1) ($d = 1$, i.e., $X(t)$ is a scalar) with measurements

$$y_i = X(t_i) + w_i \quad (55)$$

where $\{w_i\}$, independent of $\{X(t)\}$, is an independent sequence with $\mathbb{E}[|w_i|^q] < \infty$ for some $q \geq 4$. Given $\eta \geq 1$, assume that

$$\mathbb{E}[w_i^\eta | \{y_1, \dots, y_i\}] = \mathbb{E}[w_i^\eta | y_i] = \sum_{q=0}^{\eta_i} h_{\eta, q} y_i^q \quad (56)$$

for some $\eta_i \geq \eta$, where $h_{\eta, q}$ are constants at time t_i^- . Recall that $\{w_i\}$ is an independent sequence and independent of $\{X(t)\}$ and, therefore, moments $M^q(t_i^-) = \mathbb{E}[X^q(t_i^-) | \{y_1, \dots, y_{i-1}\}]$, $q \geq 1$, contain no information of w_i and can be treated as constants at time t_i^- . Let $s(w_i) = w_i^\eta$ and $u(y_i) = 1$ in (54), then substitution of (56) into (54) gives

$$\mathbb{E}[w_i^\eta] = \sum_{q=0}^{\eta_i} \hat{h}_{\eta, q} \mathbb{E}[w_i^q], \quad (57)$$

where $\hat{h}_{\eta, q}$ are constants at time t_i^- . To avoid involving the higher order moments in the expression of $\mathbb{E}[w_i^\eta]$, let $\eta_i = \eta$ in (56) and (57). This with

$$\begin{aligned} \mathbb{E}[X^\eta(t_i) | \{y_1, \dots, y_i\}] &= \mathbb{E}[(y_i - w_i)^\eta | \{y_1, \dots, y_i\}] \\ &= \sum_{q=0}^{\eta} \binom{\eta}{q} (-1)^q y_i^{\eta-q} \mathbb{E}[w_i^q | y_i] \end{aligned}$$

yields

$$M^\eta(t_i) = \sum_{q=0}^{\eta} \vartheta_{\eta, q} y_i^q \quad (58)$$

with $\eta \in \mathbb{N}$, where $m_{\eta, q}$ are constants at time t_i^- .

We choose functions

$$s(x(t_i)) = x^\eta(t_i) \quad \text{and} \quad u(y_i) = y_i^j \quad (59)$$

with $j = 0, 1, \dots, \eta$ in (54) and then have

$$\begin{aligned} \mathbb{E}[x^\eta(t_i) y_i^j | \{y_1, \dots, y_{i-1}\}] \\ = \mathbb{E}[x^\eta(t_i) (x(t_i) + w_i)^j | \{y_1, \dots, y_{i-1}\}] \\ = \mathbb{E}\left[\sum_{b=0}^j \binom{j}{b} x^{\eta+b}(t_i) w_i^{j-b} | \{y_1, \dots, y_{i-1}\}\right] \\ = \sum_{b=0}^j \binom{j}{b} M^{\eta+b}(t_i^-) \mathbb{E}[w_i^{j-b}], \end{aligned} \quad (60)$$

$$\begin{aligned} \mathbb{E}[\mathbb{E}[x^\eta(t_i) | \{y_1, \dots, y_i\}] y_i^j | \{y_1, \dots, y_{i-1}\}] \\ = \mathbb{E}\left[\sum_{q=0}^{\eta} \vartheta_{\eta, q} y_i^q y_i^j | \{y_1, \dots, y_{i-1}\}\right] \\ = \mathbb{E}\left[\sum_{q=0}^{\eta} \vartheta_{\eta, q} (x(t_i) + w_i)^{q+j} | \{y_1, \dots, y_{i-1}\}\right] \\ = \sum_{q=0}^{\eta} \vartheta_{\eta, q} \sum_{b=0}^{q+j} \binom{q+j}{b} M^b(t_i^-) \mathbb{E}[w_i^{q+j-b}]. \end{aligned} \quad (61)$$

Then (54) implies

$$\begin{aligned} \sum_{q=0}^{\eta} \vartheta_{\eta, q} \sum_{b=0}^{q+j} \binom{q+j}{b} M^b(t_i^-) \mathbb{E}[w_i^{q+j-b}] \\ = \sum_{b=0}^j \binom{j}{b} M^{\eta+b}(t_i^-) \mathbb{E}[w_i^{j-b}] \end{aligned} \quad (62)$$

with $j = 0, 1, \dots, \eta$, which determines coefficients $m_{\eta, q}$, $q = 0, 1, \dots, \eta$. For example, solving (62) for $\eta = 1$ gives

$$M^1(t_i) = \vartheta_{1,0} + \vartheta_{1,1} y_i \quad (63)$$

where

$$\begin{aligned}\vartheta_{1,0} &= \frac{\sigma_i^2 M^1(t_i^-)}{M^2(t_i^-) - [M^1(t_i^-)]^2 + \sigma_i^2}, \\ \vartheta_{1,1} &= \frac{M^2(t_i^-) - [M^1(t_i^-)]^2}{M^2(t_i^-) - [M^1(t_i^-)]^2 + \sigma_i^2}.\end{aligned}\quad (64)$$

As is well known, the larger measurement noise σ_i^2 is, the less information observation y_i gives (see, e.g., [43]). Note that (63) implies

$$\vartheta_{1,0} \rightarrow 0, \quad \vartheta_{1,1} \rightarrow 1, \quad M^1(t_i) \rightarrow y_i \quad \text{as } \sigma_i^2 \rightarrow 0$$

and

$$\begin{aligned}m_{1,0} &\rightarrow M^1(t_i^-), \quad \vartheta_{1,1} \rightarrow 0, \quad M^1(t_i) \rightarrow M^1(t_i^-) \\ &\quad \text{as } \sigma_i^2 \rightarrow \infty,\end{aligned}$$

which match very well the cases with noise-free ($\sigma_i^2 = 0$) and extremely-large-noise ($\sigma_i^2 \rightarrow \infty$) measurements, respectively. This justifies the use of approximate formula (63) at the measurement points.

It is observed that the update formula (62) for the η th moment at measurement point t_i involves $M^q(t_i^-)$ with q up to 2η . To cope with this problem, one can employ some moment closure techniques (see [38, 28, 8, 23, 37, 25, 13] and the references therein). And one may choose a moment closure based on the characteristics of the signals. As application examples, the normal [38], the lognormal [36] and the modified normal [37] approximations will be studied as follows.

For simplicity, assume that measurement noise w_i has a symmetric probability density function such that

$$\mathbb{E}[w_i] = \mathbb{E}[w_i^3] = 0, \quad \mathbb{E}[w_i^2] = \sigma_i^2, \quad \mathbb{E}[w_i^4] = h\sigma_i^4, \quad (65)$$

e.g., $h = 3$ when $w_i \sim \mathcal{N}(0, \sigma_i^2)$, and $h = 6$ when $w_i \sim \mathcal{L}(0, \sigma_i)$, where $\mathcal{L}(0, \sigma_i)$ denotes the Laplace distribution with zero mean and variance $\sigma_i^2 > 0$, which is also widely used in science and engineering, see [22, 6, 32]. Let us consider the update approximate for the second moment

$$M^2(t_i) = \vartheta_{2,0} + \vartheta_{2,1}y_i + \vartheta_{2,2}y_i^2. \quad (66)$$

So, for $\eta = 2$, (62) gives a set of linear equations with respect to $m_{2,0}$, $m_{2,1}$ and $m_{2,2}$:

$$\vartheta_{2,0} + \vartheta_{2,1}M^1(t_i^-) + \vartheta_{2,2}[\sigma_i^2 + M^2(t_i^-)] = M^2(t_i^-), \quad (67)$$

$$\begin{aligned}\vartheta_{2,0}M^1(t_i^-) + \vartheta_{2,1}[\sigma_i^2 + M^2(t_i^-)] \\ + \vartheta_{2,2}[3\sigma_i^2 M^1(t_i^-) + M^3(t_i^-)] = M^3(t_i^-),\end{aligned}\quad (68)$$

$$\begin{aligned}\vartheta_{2,0}[\sigma_i^2 + M^2(t_i^-)] + \vartheta_{2,1}[3\sigma_i^2 M^1(t_i^-) + M^3(t_i^-)] \\ + \vartheta_{2,2}[h\sigma_i^4 + 6\sigma_i^2 M^2(t_i^-) + M^4(t_i^-)] \\ = \sigma_i^2 M^2(t_i^-) + M^4(t_i^-).\end{aligned}\quad (69)$$

Solving (67)-(69) will gives the update approximate (66). As observed, (67)-(69) involve $M^3(t_i^-)$ and $M^4(t_i^-)$ and can be solved with different moment closure techniques. As application examples of our proposed method, we apply the normal, the lognormal and the modified normal moment closures below, respectively.

To approximate some higher order moment, e.g. $\eta = 3$, one would obtain (62) involving $M^j(t_i^-)$, $1 \leq j \leq 6$. Application of some moment closure techniques, e.g., setting the cumulant neglect level $\eta = 3$, that is, $\bar{\kappa}_j(t_i^-) = 0$ for $4 \leq j \leq 6$, see [28], would give moment closure functions for $M^j(t_i^-)$, $4 \leq j \leq 6$, in terms of $\{M^q(t_i^-)\}_{q=1}^3$. Substitution of these expressions into (62) would determine the coefficients $m_{j,q}$ for $1 \leq j \leq 3$ and $0 \leq q \leq j$.

S.6.1 Normal moment closure

Let the cumulant neglect level be set to 2, that is, $\bar{\kappa}_j(t_i^-) = 0$ for $j \geq 3$, where $\bar{\kappa}_j(t_i^-)$ is the j th cumulant of $X(t_i^-)$. This yields

$$\begin{cases} M^3(t_i^-) = 3M^2(t_i^-)M^1(t_i^-) - 2[M^1(t_i^-)]^3 \\ M^4(t_i^-) = 4M^3(t_i^-)M^1(t_i^-) + 3[M^2(t_i^-)]^2 \\ \quad - 12M^2(t_i^-)[M^1(t_i^-)]^2 + 6[M^1(t_i^-)]^4, \end{cases} \quad (70)$$

and therefore

$$\begin{cases} M^3(t_i^-) = 3M^2(t_i^-)M^1(t_i^-) - 2[M^1(t_i^-)]^3 \\ M^4(t_i^-) = 3[M^2(t_i^-)]^2 - 2[M^1(t_i^-)]^4. \end{cases} \quad (71)$$

Solving (67)-(69) with (71) yields

$$\begin{aligned}\vartheta_{2,2} &= \frac{[\Sigma(t_i^-)]^2}{[\Sigma(t_i^-) + \sigma_i^2]^2 + \frac{1}{2}(h-3)\sigma_i^4}, \\ \vartheta_{2,1} &= 2 \left[\frac{\Sigma(t_i^-)}{\Sigma(t_i^-) + \sigma_i^2} - \vartheta_{2,2} \right] M^1(t_i^-), \\ \vartheta_{2,0} &= M^2(t_i^-) - \vartheta_{2,1}M^1(t_i^-) - \vartheta_{2,2}[M^2(t_i^-) + \sigma_i^2],\end{aligned}\quad (72)$$

It is observed in (72) that

$$\vartheta_{2,2} \rightarrow 1, \quad \vartheta_{2,1} \rightarrow 0, \quad \vartheta_{2,0} \rightarrow 0, \quad M^2(t_i) \rightarrow y_i^2 \quad \text{as } \sigma_i^2 \rightarrow 0$$

and

$$\begin{aligned}\vartheta_{2,2} \rightarrow 0, \quad \vartheta_{2,1} \rightarrow 0, \quad \vartheta_{2,0} \rightarrow M^2(t_i^-), \quad M^2(t_i) \rightarrow M^2(t_i^-) \\ \text{as } \sigma_i^2 \rightarrow \infty,\end{aligned}$$

which correspond to the cases with noise-free and extremely-large-noise measurements, respectively. This shows that it is reasonable to use the proposed update approximation (66) for the second moment. It is also noticed that, if $w_i \sim \mathcal{L}(0, \sigma_i)$, i.e., $h = 6$,

$$\begin{aligned}\vartheta_{2,2} &= \frac{[\Sigma(t_i^-)]^2}{[\Sigma(t_i^-) + \sigma_i^2]^2 + 1.5\sigma_i^4} \approx \frac{[\Sigma(t_i^-)]^2}{[\Sigma(t_i^-) + \sigma_i^2]^2}, \\ \vartheta_{2,1} &\approx \frac{2\sigma_i^2 \Sigma(t_i^-)}{[\Sigma(t_i^-) + \sigma_i^2]^2} M^1(t_i^-)\end{aligned}$$

when signal-to-noise ratio

$$SNR := \frac{\Sigma(t_i^-)}{\sigma_i^2} \quad (73)$$

is large. This implies that scheme (66) with $h = 3$ can be also used to approximate the second moment when $w_i \sim \mathcal{L}(0, \sigma_i)$ with small variance σ_i^2 .

It should be pointed out that setting the cumulant $\bar{\kappa}_j(t_i^-) = 0$ for $j \geq 3$ implies that the approximation is based on the Gaussian assumption of the states, see [38] and [28]. If $w_i \sim \mathcal{N}(0, \sigma_i^2)$, then the Gaussian assumption is imposed on both the states and the measurements.

Remark 1: It is observed that formula (63) and (72) with $h = 3$ are exactly the update formula of a Kalman filter, see, e.g., [4], which is a special case of our proposed approach when the normal moment closure is employed for measurements with Gaussian noise. This helps justify our proposed approach.

S.6.2 Lognormal moment closure

We can employ some nonzero cumulant moment closure techniques instead of (71). The following formula are based on the assumption of lognormal, instead of normal, moment dynamics, see [20, 21, 26, 36],

$$M^3(t_i^-) = \frac{[M^2(t_i^-)]^3}{[M^1(t_i^-)]^3}, \quad M^4(t_i^-) = \frac{[M^2(t_i^-)]^6}{[M^1(t_i^-)]^8}. \quad (74)$$

Solving (67)-(69) with (74) yields

$$\begin{aligned} \vartheta_{2,2} &= [\sigma_i^2 A(t_i^-) + B(t_i^-)] \left\{ (h-1)\sigma_i^6 + (h+3)\sigma_i^4 (M^2(t_i^-) \right. \\ &\quad \left. - [M^1(t_i^-)]^2) + \sigma_i^2 C(t_i^-) + B(t_i^-) \right\}^{-1}, \\ \vartheta_{2,1} &= \left\{ (1 - m_{2,2}) \left[\frac{[M^2(t_i^-)]^3}{[M^1(t_i^-)]^3} - M^2(t_i^-) M^1(t_i^-) \right] \right. \\ &\quad \left. - 2\vartheta_{2,2}\sigma_i^2 M^1(t_i^-) \right\} \left\{ M^2(t_i^-) - [M^1(t_i^-)]^2 + \sigma_i^2 \right\}^{-1}, \\ \vartheta_{2,0} &= M^2(t_i^-) - \vartheta_{2,1} M^1(t_i^-) - \vartheta_{2,2} [M^2(t_i^-) + \sigma_i^2], \end{aligned} \quad (75)$$

where

$$\begin{aligned} A(t_i^-) &= \frac{[M^2(t_i^-)]^6}{[M^1(t_i^-)]^8} - 2 \frac{[M^2(t_i^-)]^3}{[M^1(t_i^-)]^2} - [M^2(t_i^-)]^2 \\ &\quad + 2[M^1(t_i^-)]^2 M^2(t_i^-), \\ B(t_i^-) &= \frac{[M^2(t_i^-)]^7}{[M^1(t_i^-)]^8} - 2 \frac{[M^2(t_i^-)]^6}{[M^1(t_i^-)]^6} + 2 \frac{[M^2(t_i^-)]^4}{[M^1(t_i^-)]^2} \\ &\quad - [M^2(t_i^-)]^3, \\ C(t_i^-) &= \frac{[M^2(t_i^-)]^6}{[M^1(t_i^-)]^8} - 4 \frac{[M^2(t_i^-)]^3}{[M^1(t_i^-)]^2} + 3[M^2(t_i^-)]^2. \end{aligned}$$

As above, it is observed in (75) that

$$\vartheta_{2,2} \rightarrow 1, \quad \vartheta_{2,1} \rightarrow 0, \quad \vartheta_{2,0} \rightarrow 0, \quad M^2(t_i) \rightarrow y_i^2 \quad \text{as } \sigma_i^2 \rightarrow 0$$

and

$$\begin{aligned} \vartheta_{2,2} \rightarrow 0, \quad \vartheta_{2,1} \rightarrow 0, \quad \vartheta_{2,0} \rightarrow M^2(t_i^-), \quad M^2(t_i) \rightarrow M^2(t_i^-) \\ \text{as } \sigma_i^2 \rightarrow \infty, \end{aligned}$$

which correspond to the cases with noise-free and extremely-large-noise measurements, respectively.

S.6.3 Modified normal moment closure

A slight modification of the normal moment closure functions (70) is presented in [37], which is given as follows, see Eq.(30) in [37],

$$\begin{cases} M^3(t_i^-) = 3M^2(t_i^-)M^1(t_i^-) - 2[M^1(t_i^-)]^3 \\ M^4(t_i^-) = 4M^3(t_i^-)M^1(t_i^-) + 3[M^2(t_i^-)]^2 \\ \quad - 12M^2(t_i^-)[M^1(t_i^-)]^2 + 6[M^1(t_i^-)]^4 \\ \quad + M_{df}[M^2(t_i^-) - [M^1(t_i^-)]^2] \end{cases} \quad (76)$$

with modification indicator $M_{df} = 1$. This leads to

$$\begin{cases} M^3(t_i^-) = 3M^2(t_i^-)M^1(t_i^-) - 2[M^1(t_i^-)]^3 \\ M^4(t_i^-) = 3[M^2(t_i^-)]^2 - 2[M^1(t_i^-)]^4 \\ \quad + M_{df}[M^2(t_i^-) - [M^1(t_i^-)]^2]. \end{cases} \quad (77)$$

Solving (67)-(69) with (77) yields

$$\begin{aligned} \vartheta_{2,2} &= \frac{[\Sigma(t_i^-)]^2 + \frac{M_{df}}{2}\Sigma(t_i^-)}{[\Sigma(t_i^-) + \sigma_i^2]^2 + \frac{1}{2}(h-3)\sigma_i^4 + \frac{M_{df}}{2}\Sigma(t_i^-)}, \\ \vartheta_{2,1} &= 2 \left[\frac{\Sigma(t_i^-)}{\Sigma(t_i^-) + \sigma_i^2} - \vartheta_{2,2} \right] M^1(t_i^-), \\ \vartheta_{2,0} &= M^2(t_i^-) - \vartheta_{2,1} M^1(t_i^-) - \vartheta_{2,2} [M^2(t_i^-) + \sigma_i^2]. \end{aligned} \quad (78)$$

As above, one sees

$$\vartheta_{2,2} \rightarrow 1, \quad \vartheta_{2,1} \rightarrow 0, \quad \vartheta_{2,0} \rightarrow 0, \quad M^2(t_i) \rightarrow y_i^2 \quad \text{as } \sigma_i^2 \rightarrow 0$$

and

$$\begin{aligned} \vartheta_{2,2} \rightarrow 0, \quad \vartheta_{2,1} \rightarrow 0, \quad \vartheta_{2,0} \rightarrow M^2(t_i^-), \quad M^2(t_i) \rightarrow M^2(t_i^-) \\ \text{as } \sigma_i^2 \rightarrow \infty, \end{aligned}$$

which correspond to the cases with noise-free and extremely-large-noise measurements, respectively. Moreover, it is noted that modified normal moment closures (76), (77) and (78) become normal moment closures (70), (71) and (72), respectively, if the modification indicator M_{df} is set $M_{df} = 0$.

S.6.4 Approximations of central moments

In the previous section, we have studied the approximations for the moments of $X(t)$. Alternatively, we can approximate the central moments of $X(t)$ by exploiting some of their properties, e.g., the symmetry of a covariance matrix. Similarly to (50) above, denote the central moments by

$$C_\pi^\eta(t) = \mathbb{E}_\pi[(X_1(t) - \mathbb{E}_\pi[X_1(t)])^\eta \cdots (X_d(t) - \mathbb{E}_\pi[X_d(t)])^\eta] \\ = \sum_{x \in \mathcal{X}} \pi(x, t) [(x_1 - \mathbb{E}_\pi[X_1(t)])^\eta \cdots (x_d - \mathbb{E}_\pi[X_d(t)])^\eta]$$

for all $\eta \in \mathbb{N}^d$ with $|\eta| > 0$ and $t \geq t_0$. For convenience, also denote by $\Sigma(t)$ the variance for a univariate system and the covariance matrix for a multivariate system.

In this section, we consider a univariate system (i.e., $d = 1$). It is natural to assume that the update step, which incorporates the information of a measurement, leads to a variance $\Sigma(t_i)$ smaller than $\Sigma(t_i^-)$. That is, at measurement point t_i ,

$$\Sigma(t_i) = M^2(t_i) - [M^1(t_i)]^2 = m\Sigma(t_i^-) \quad (79)$$

and hence

$$M^2(t_i) = [M^1(t_i)]^2 + m\Sigma(t_i^-), \quad (80)$$

where m is a coefficient to be determined. Then (58) with (63) and (80) gives

$$m_{2,0} = \vartheta_{1,0}^2 + m\Sigma(t_i^-), \quad \vartheta_{2,1} = 2\vartheta_{1,0}\vartheta_{1,1}, \quad \vartheta_{2,2} = \vartheta_{1,1}^2.$$

Letting $s(x(t_i)) = x^2(t_i)$, $u(y_i) = 1$ in (54) and using (66) yield

$$m = \frac{\sigma_i^2}{\Sigma(t_i^-) + \sigma_i^2}, \quad (81)$$

where $m_{1,0}$ and $m_{1,1}$ are given by (64). Substitution of (64) and (81) into (81) leads to (72) with $h = 3$.

Generally, by (58), one has

$$\Sigma(t_i) = M^2(t_i) - [M^1(t_i)]^2 = v_{2,0} + v_{2,1}y_i + v_{2,2}y_i^2, \quad (82)$$

and hence

$$v_{2,0} = \vartheta_{2,0} - \vartheta_{1,0}^2, \quad v_{2,1} = \vartheta_{2,1} - 2\vartheta_{1,0}\vartheta_{1,1}, \quad v_{2,2} = \vartheta_{2,2} - \vartheta_{1,1}^2,$$

where $m_{1,0}$ and $m_{1,1}$ are given by (64) while $m_{2,0}$, $m_{2,1}$ and $m_{2,2}$ are given by, e.g., (78). It is observed that $v_{2,0} = m\Sigma(t_i^-)$ and $v_{2,1} = v_{2,2} = 0$ when $h = 3$ and $M_{df} = 0$ in (78).

Let us consider the birth-death reaction model



where $X \in \mathbb{N}$ denotes the quantity of species S , $r = 2$, $\nu_1 = -\nu_2 = 1$, $\alpha_1(x) = c_1$ and $\alpha_2(x) = c_2x$. In this case, $\mathcal{X} = \{x \in \mathbb{N}\}$. The approximates by normal, log-normal, modified normal moment closures for birth-death model (83) and measurements (55) (with reaction rates $c_1 = 5 \text{ molec} \cdot \text{sec}^{-1}$, $c_2 = 0.1 \text{ sec}^{-1}$, measurement noise standard deviation $\sigma = 10 \text{ molec}$, time interval $t_i - t_{i-1} = 25 \text{ sec}$) are given in Fig. 1, Fig. 2, Fig. 3 below, respectively.

S.7 Extension to multivariate systems

In this section, we extend the results in the previous section to multivariate systems. As an example, we shall apply our proposed results to the two-state gene expression model, see (92)-(96) below, in this work. Since, in the two-state gene expression model, the population of either activated DNA or deactivated DNA is always zero, some techniques of lognormal distribution are not applicable to this case, see [14, 13]. To cope with this problem, one may employ the hybrid framework (see, e.g., [7, 12, 30]), which is a future work. Therefore, we focus on the extension of normal and modified normal approximations in this work.

For a system (1) with $d \geq 2$ (i.e., $X(t)$ is a d -dimensional vector), denote the first and second posterior moments by

$$M(t) = \mathbb{E}_\pi[X(t)] \in \mathbb{R}^d, \quad M^2(t) = \mathbb{E}_\pi[X(t)X^T(t)] \in \mathbb{R}^{d \times d}, \quad (84)$$

and then posterior covariance

$$\Sigma(t) = M^2(t) - M(t)M^T(t) \in \mathbb{R}^{d \times d}$$

for all $t \geq t_0$, where $\Sigma^{(j,q)}(t)$ is the (j, q) th element of matrix $\Sigma(t)$. Since $M^2(t)$ is symmetric, the second posterior moment can be alternatively represented by

$$M^{2v}(t) = \mathbb{E}_\pi[\tilde{X}(t)] \in \mathbb{R}^{d(d+1)/2}, \quad (85)$$

where $\tilde{X}(t) = \text{vech}(X(t)X^T(t))$ and $\text{vech}(\cdot)$ is the half-vectorization operator, see, e.g., [27], that is,

$$\text{vech}(xx^T) = [x_1x_1 \quad x_2x_1 \quad \cdots \quad x_dx_1 \quad x_2x_2 \quad \cdots \quad x_dx_2 \quad x_3x_3 \quad \cdots \quad x_dx_d]^T$$

for $x = [x_1 \quad \cdots \quad x_d]^T$.

Before and between the measurement points, the propagation of these moments is described by (51). Here, our main task is to find an update approximate for (52) at the measurement points. In this section, we extend our proposed formula (63) and (66) to a multivariate system (1) ($d \geq 2$) with measurements

$$y_i = FX(t_i) + w_i, \quad (86)$$

where $y_i \in \mathbb{R}^l$, $l \leq d$, $F \in \mathbb{R}^{l \times d}$, and $\{w_i\}$ with $w_i \in \mathbb{R}^l$, independent of $\{X(t)\}$, is an independent sequence with $\mathbb{E}[w_i] = 0 \in \mathbb{R}^l$, $\mathbb{E}[w_iw_i^T] = R_w \in \mathbb{R}^{l \times l}$ and

$$\mathbb{E}[|w_i^\eta|] = \mathbb{E}[|w_{i,1}^{\eta_1} \cdots w_{i,l}^{\eta_l}|] < \infty$$

with $|\eta| \leq q$ for some $q \geq 4$, e.g., $w_i \sim \mathcal{N}(0, R_w)$ is a special case.

Let

$$M(t_i) = \Theta_{1,0} + \Theta_{1,1}y_i \quad (87)$$

where $L_{1,0} \in \mathbb{R}^d$ and $L_{1,1} \in \mathbb{R}^{d \times q}$. With $s(x(t_i)) = x(t_i)$ and $u(y_i) = 1$, (54) becomes

$$M(t_i^-) = \Theta_{1,0} + \Theta_{1,1}FM(t_i^-). \quad (88)$$

Setting $s(x(t_i)) = x(t_i)$ and $u(y_i) = y_i^T$ in (54), we have

$$\begin{aligned} \mathbb{E}\{x(t_i)y_i^T | Y_1^{n-1}\} &= \mathbb{E}\{x(t_i)[x^T(t_i)F^T + w_i^T] | Y_1^{n-1}\} \\ &= M^2(t_i^-)F^T, \\ \mathbb{E}\{x(t_i)y_i^T | Y_1^{n-1}\} &= \mathbb{E}\{(\Theta_{1,0} + \Theta_{1,1}y_i)y_i^T | Y_1^{n-1}\} \\ &= \mathbb{E}\{\Theta_{1,0}[x^T(t_i)F^T + w_i^T] | Y_1^{n-1}\} \\ &\quad + \mathbb{E}\{\Theta_{1,1}[Fx(t_i) + w_i][x^T(t_i)F^T + w_i^T] | Y_1^{n-1}\} \\ &= \Theta_{1,0}M^T(t_i^-)F^T + \Theta_{1,1}[FM^2(t_i^-)F^T + R_w] \end{aligned}$$

and therefore

$$\begin{aligned} M^2(t_i^-)F^T &= \Theta_{1,0}M^T(t_i^-)F^T \\ &\quad + \Theta_{1,1}[FM^2(t_i^-)F^T + R_w]. \end{aligned} \quad (89)$$

Combining (85), (88) and (89), we obtain

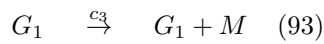
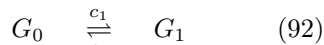
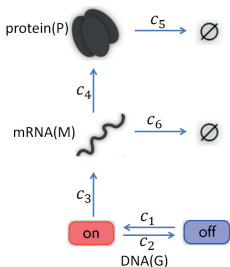
$$\begin{aligned} \Theta_{1,1} &= [M^2(t) - M(t)M^T(t)]F^T \\ &\quad \{F[M^2(t) - M(t)M^T(t)]F^T + R_w\}^{-1} \\ &= \Sigma(t_i^-)F^T[F\Sigma(t_i^-)F^T + R_w]^{-1}. \end{aligned} \quad (90)$$

And then, by (88), we have

$$\Theta_{1,0} = M(t_i^-) - \Theta_{1,1}FM(t_i^-). \quad (91)$$

Let us proceed to study the update formula, the counterpart of (66), for the second moments of the multivariate system with normal and modified normal approximations, respectively. We will illustrate our proposed method with the two-state gene expression model.

The two-state gene expression model consists of six reactions that involve four different species G_0 (for deactivated DNA), G_1 (for activated DNA), M (for mRNA) and P (for protein)



Since $G_0 + G_1 = 1$, the state of the system can be represented by $X = (X_1 \ X_2 \ X_3)^T = (G_1 \ M \ P)^T \in \mathbb{N}^3$. The observation is

$$y_i = X_3(t_i) + w_i \quad (97)$$

with w_i given by (65).

S.7.1 Normal approximation

A multivariate version of normal approximation (66) with (72) can be obtained by the approach (67)-(69) as illustrated with an example below. Here, we extend the normal approximation (81)-(81) for a multivariate system by the approach (79). Assume that, at measurement point t_i , the use of measurement y_i leads to a smaller covariance

$$\Sigma(t_i) = [\Sigma(t_i^-)]^{1/2}Q[\Sigma(t_i^-)]^{1/2} \quad (98)$$

where $Q \in \mathbb{R}^{d \times d}$ is a symmetric matrix to be determined.

Letting $s(x(t_i)) = x(t)x^T(t)$, $u(y_i) = 1$ in (54) and using (87), (98), we obtain

$$\begin{aligned} M^2(t_i^-) &= \mathbb{E}\{\Sigma(t_i) + M(t_i)M^T(t_i)\} \\ &= [\Sigma(t_i^-)]^{1/2}Q[\Sigma(t_i^-)]^{1/2} + \Theta_{1,0}\Theta_{1,0}^T \\ &\quad + \Theta_{1,0}M^T(t_i^-)F^T\Theta_{1,1}^T + \Theta_{1,1}FM(t_i^-)\Theta_{1,0}^T \\ &\quad + \Theta_{1,1}[FM^2(t_i^-)F^T + R_w]\Theta_{1,1}^T \end{aligned} \quad (99)$$

and hence

$$\begin{aligned} Q &= [\Sigma(t_i^-)]^{-1/2} \\ &\quad \left\{ M^2(t_i^-) - \Theta_{1,0}\Theta_{1,0}^T - \Theta_{1,0}M^T(t_i^-)F^T\Theta_{1,1}^T \right. \\ &\quad \left. - \Theta_{1,1}FM(t_i^-)\Theta_{1,0}^T - \Theta_{1,1}[FM^2(t_i^-)F^T + R_w]\Theta_{1,1}^T \right\} \\ &\quad [\Sigma(t_i^-)]^{-1/2}. \end{aligned} \quad (100)$$

It is observed that (91), (90) and (100) are the multivariate versions of (64) and (81), respectively. Substitution of (100) into (98) leads to

$$\begin{aligned} \Sigma(t_i) &= M^2(t_i^-) - \Theta_{1,0}\Theta_{1,0}^T - \Theta_{1,0}M^T(t_i^-)F^T\Theta_{1,1}^T \\ &\quad - \Theta_{1,1}FM(t_i^-)\Theta_{1,0}^T - \Theta_{1,1}[FM^2(t_i^-)F^T + R_w]\Theta_{1,1}^T \end{aligned} \quad (101)$$

and then the second moment

$$\begin{aligned} M^2(t_i) &= \Sigma(t_i) + M(t_i)M^T(t_i) \\ &= M^2(t_i^-) - \Theta_{1,0}M^T(t_i^-)F^T\Theta_{1,1}^T \\ &\quad - \Theta_{1,1}FM(t_i^-)\Theta_{1,0}^T - \Theta_{1,1}[FM^2(t_i^-)F^T + R_w]\Theta_{1,1}^T \\ &\quad + \Theta_{1,0}y_i^T\Theta_{1,1}^T + \Theta_{1,1}y_i\Theta_{1,0} + \Theta_{1,1}y_iy_i^T\Theta_{1,1}^T, \end{aligned} \quad (102)$$

where $\Theta_{1,0}$ and $\Theta_{1,1}$ are given by (91) and (90), respectively.

S.7.2 Modified normal approximation

The approximate for the second moment is given by

$$M^{2v}(t_i) = \Theta_{2,0} + \Theta_{2,1}y_i + \Theta_{2,2}\tilde{y}_i \quad (103)$$

where $L_{2,0} \in \mathbb{R}^{d(d+1)/2}$, $L_{2,1} \in \mathbb{R}^{d(d+1)/2 \times q}$, $L_{2,2} \in \mathbb{R}^{d(d+1)/2 \times q(q+1)/2}$ and $\tilde{y}_i = \text{vech}(y_i y_i^T) \in \mathbb{R}^{q(q+1)/2}$. Coefficients $L_{2,0}$, $L_{2,1}$ and $L_{2,2}$, which are comprised of

$$\begin{aligned} \frac{d(d+1)}{2} + \frac{d(d+1)}{2}l + \frac{d(d+1)}{2} \frac{l(l+1)}{2} \\ = \frac{d(d+1)}{2} \frac{(l+1)(l+2)}{2} \end{aligned}$$

unknown scalar entries, are to be determined by (104)-(106) below. But (54) with $s(x(t_i)) = \tilde{X}(t_i)$ and $u(y_i) = 1, y_i^T, \tilde{y}_i^T$ give

$$M^{2v}(t_i^-) = \Theta_{2,0} + \Theta_{2,1}FM(t_i^-) + \Theta_{2,2}\mathbb{E}[\tilde{y}_i|Y_1^{n-1}], \quad (104)$$

$$\begin{aligned} \mathbb{E}[\tilde{X}(t_i)X^T(t_i)|Y_1^{n-1}]F^T = \Theta_{2,0}M^T(t_i^-)F^T \\ + \Theta_{2,1}[FM^2(t_i^-)F^T + R_w] + \Theta_{2,2}\mathbb{E}[\tilde{y}_i y_i^T|Y_1^{n-1}], \end{aligned} \quad (105)$$

$$\begin{aligned} \mathbb{E}[\tilde{X}(t_i)\tilde{y}_i^T|Y_1^{n-1}] = \Theta_{2,0}\mathbb{E}[\tilde{y}_i^T|Y_1^{n-1}] \\ + \Theta_{2,1}\mathbb{E}[y_i \tilde{y}_i^T|Y_1^{n-1}] + \Theta_{2,2}\mathbb{E}[\tilde{y}_i \tilde{y}_i^T|Y_1^{n-1}], \end{aligned} \quad (106)$$

respectively, which are a set of $\frac{d(d+1)}{2} \frac{(l+1)(l+2)}{2}$ scalar equations as well.

Note that (104)-(106) involve higher order moments $M^\eta(t_i^-)$ with $|\eta| = 3, 4$. But, with some moment closure

$$M^\eta(t_i^-) = \varphi^{(\eta)}(M(t_i^-), M^{2v}(t_i^-)) \quad (107)$$

for $|\eta| = 3, 4$, one can obtain the coefficients $L_{2,0}$, $L_{2,1}$ and $L_{2,2}$ by solving equations (104)-(106). For example, one can obtain $L_{2,0}$, $L_{2,1}$ and $L_{2,2}$ in (103) by solving the equations with a multivariate version of modified normal moment closure (76), which is given as

$$\begin{aligned} \mathbb{E}_\pi[\bar{X}\bar{Y}\bar{Z}] &= \mathbb{E}_\pi[\bar{X}\bar{Y}]\mathbb{E}_\pi[\bar{Z}] + \mathbb{E}_\pi[\bar{X}\bar{Z}]\mathbb{E}_\pi[\bar{Y}] + \mathbb{E}_\pi[\bar{Y}\bar{Z}]\mathbb{E}_\pi[\bar{X}] \\ &\quad - 2\mathbb{E}_\pi[\bar{X}]\mathbb{E}_\pi[\bar{Y}]\mathbb{E}_\pi[\bar{Z}] \\ \mathbb{E}_\pi[\bar{X}\bar{Y}\bar{Z}\bar{W}] &= \mathbb{E}_\pi[\bar{X}\bar{Y}\bar{Z}]\mathbb{E}_\pi[\bar{W}] + \mathbb{E}_\pi[\bar{X}\bar{Y}\bar{W}]\mathbb{E}_\pi[\bar{Z}] \\ &\quad + \mathbb{E}_\pi[\bar{X}\bar{Z}\bar{W}]\mathbb{E}_\pi[\bar{Y}] + \mathbb{E}_\pi[\bar{Y}\bar{Z}\bar{W}]\mathbb{E}_\pi[\bar{X}] \\ &\quad + \mathbb{E}_\pi[\bar{X}\bar{Y}]\mathbb{E}_\pi[\bar{Z}\bar{W}] + \mathbb{E}_\pi[\bar{X}\bar{Z}]\mathbb{E}_\pi[\bar{Y}\bar{W}] + \mathbb{E}_\pi[\bar{X}\bar{W}]\mathbb{E}_\pi[\bar{Y}\bar{Z}] \\ &\quad - 2\left\{ \mathbb{E}_\pi[\bar{X}\bar{Y}]\mathbb{E}_\pi[\bar{Z}]\mathbb{E}_\pi[\bar{W}] + \mathbb{E}_\pi[\bar{X}\bar{Z}]\mathbb{E}_\pi[\bar{Y}]\mathbb{E}_\pi[\bar{W}] \right. \\ &\quad + \mathbb{E}_\pi[\bar{X}\bar{W}]\mathbb{E}_\pi[\bar{Y}]\mathbb{E}_\pi[\bar{Z}] + \mathbb{E}_\pi[\bar{Y}\bar{Z}]\mathbb{E}_\pi[\bar{X}]\mathbb{E}_\pi[\bar{W}] \\ &\quad + \mathbb{E}_\pi[\bar{Y}\bar{W}]\mathbb{E}_\pi[\bar{X}]\mathbb{E}_\pi[\bar{Z}] + \mathbb{E}_\pi[\bar{Z}\bar{W}]\mathbb{E}_\pi[\bar{X}]\mathbb{E}_\pi[\bar{Y}] \left. \right\} \\ &\quad + 6\mathbb{E}_\pi[\bar{X}]\mathbb{E}_\pi[\bar{Y}]\mathbb{E}_\pi[\bar{Z}]\mathbb{E}_\pi[\bar{W}] + \frac{M_{df}}{6} \left\{ \Sigma[\bar{X}\bar{Y}] \right. \\ &\quad + \Sigma[\bar{X}\bar{Z}] + \Sigma[\bar{X}\bar{W}] + \Sigma[\bar{Y}\bar{Z}] \\ &\quad \left. + \Sigma[\bar{Y}\bar{W}] + \Sigma[\bar{Z}\bar{W}] \right\}, \end{aligned} \quad (108)$$

where $\bar{X}, \bar{Y}, \bar{Z}, \bar{W} \in \{X_1, X_2, \dots, X_d\}$, $\Sigma[\bar{X}\bar{Y}] = \mathbb{E}_\pi[\bar{X}\bar{Y}] - \mathbb{E}_\pi[\bar{X}]\mathbb{E}_\pi[\bar{Y}]$ and modification indicator $M_{df} = 1$. This will be

illustrated with the example of the two-state gene expression model (92)-(97).

In this example, $d = 3$, $l = 1$, $L_{2,0} \in \mathbb{R}^6$, $L_{2,1} \in \mathbb{R}^6$, $L_{2,2} \in \mathbb{R}^6$,

$$M^{2v}(t) = \mathbb{E}_\pi[\tilde{X}(t)] = \begin{pmatrix} M^{(2,0,0)}(t) \\ M^{(1,1,0)}(t) \\ M^{(1,0,1)}(t) \\ M^{(0,2,0)}(t) \\ M^{(0,1,1)}(t) \\ M^{(0,0,2)}(t) \end{pmatrix} \in \mathbb{R}^6, \quad (109)$$

and (104)-(106) with (65) become

$$\begin{aligned} M^{2v}(t_i^-) &= \Theta_{2,0} + \Theta_{2,1}M^{(0,0,1)}(t_i^-) \\ &\quad + \Theta_{2,2}[M^{(0,0,2)}(t_i^-) + \sigma_i^2], \end{aligned} \quad (110)$$

$$\begin{aligned} \mathbb{E}_\pi[\tilde{X}(t_i^-)X_3(t_i^-)] &= \Theta_{2,0}M^{(0,0,1)}(t_i^-) \\ &\quad + \Theta_{2,1}[M^{(0,0,2)}(t_i^-) + \sigma_i^2] \\ &\quad + \Theta_{2,2}[M^{(0,0,3)}(t_i^-) + 3\sigma_i^2 M^{(0,0,1)}(t_i^-)], \end{aligned} \quad (111)$$

$$\begin{aligned} \mathbb{E}_\pi[\tilde{X}(t_i^-)X_3^2(t_i^-)] &= \sigma_i^2 M^{2v}(t_i^-) \\ &= \Theta_{2,2}[M^{(0,0,4)}(t_i^-) + 6\sigma_i^2 M^{(0,0,2)}(t_i^-) + h\sigma_i^4] \\ &\quad + \Theta_{2,1}[M^{(0,0,3)}(t_i^-) + 3\sigma_i^2 M^{(0,0,1)}(t_i^-)] \\ &\quad + \Theta_{2,0}[M^{(0,0,2)}(t_i^-) + \sigma_i^2] \end{aligned} \quad (112)$$

with

$$\begin{aligned} \mathbb{E}_\pi[\tilde{X}(t_i^-)X_3(t_i^-)] &= \begin{pmatrix} M^{(2,0,1)}(t_i^-) \\ M^{(1,1,1)}(t_i^-) \\ M^{(1,0,2)}(t_i^-) \\ M^{(0,2,1)}(t_i^-) \\ M^{(0,1,2)}(t_i^-) \\ M^{(0,0,3)}(t_i^-) \end{pmatrix}, \\ \mathbb{E}_\pi[\tilde{X}(t_i^-)X_3^2(t_i^-)] &= \begin{pmatrix} M^{(2,0,2)}(t_i^-) \\ M^{(1,1,2)}(t_i^-) \\ M^{(1,0,3)}(t_i^-) \\ M^{(0,2,2)}(t_i^-) \\ M^{(0,1,3)}(t_i^-) \\ M^{(0,0,4)}(t_i^-) \end{pmatrix}, \end{aligned} \quad (113)$$

where $M^\eta(t_i^-)$, $|\eta| = 3, 4$, are given by the multivariate moment closure functions (108) as follows

$$\begin{aligned} M^{(2,0,1)}(t) &= M^{(2,0,0)}(t)M^{(0,0,1)}(t) + 2M^{(1,0,1)}(t)M^{(1,0,0)}(t) \\ &\quad - 2[M^{(1,0,0)}(t)]^2 M^{(0,0,1)}(t), \\ M^{(1,1,1)}(t) &= M^{(1,1,0)}(t)M^{(0,0,1)}(t) + M^{(1,0,1)}(t)M^{(0,1,0)}(t) \\ &\quad + M^{(0,1,1)}(t)M^{(1,0,0)}(t) - 2M^{(1,0,0)}(t)M^{(0,1,0)}(t)M^{(0,0,1)}(t), \\ M^{(1,0,2)}(t) &= 2M^{(1,0,1)}(t)M^{(0,0,1)}(t) + M^{(0,0,2)}(t)M^{(1,0,0)}(t) \\ &\quad - 2M^{(1,0,0)}(t)[M^{(0,0,1)}(t)]^2, \\ M^{(0,2,1)}(t) &= M^{(0,2,0)}(t)M^{(0,0,1)}(t) + 2M^{(0,1,1)}(t)M^{(0,1,0)}(t) \\ &\quad - 2[M^{(0,1,0)}(t)]^2 M^{(0,0,1)}(t), \\ M^{(0,1,2)}(t) &= 2M^{(0,1,1)}(t)M^{(0,0,1)}(t) + M^{(0,0,2)}(t)M^{(0,1,0)}(t) \\ &\quad - 2M^{(0,1,0)}(t)[M^{(0,0,1)}(t)]^2, \end{aligned}$$

$$\begin{aligned}
 M^{(0,0,3)}(t) &= 3M^{(0,0,2)}(t)M^{(0,0,1)}(t) - 2[M^{(0,0,1)}(t)]^3, \\
 M^{(2,0,2)}(t) &= 2M^{(2,0,1)}(t)M^{(0,0,1)}(t) + 2M^{(1,0,2)}(t)M^{(1,0,0)}(t) \\
 &\quad + M^{(2,0,0)}(t)M^{(0,0,2)}(t) + 2[M^{(1,0,1)}(t)]^2 \\
 &\quad - 2\left\{M^{(2,0,0)}(t)[M^{(0,0,1)}(t)]^2 + M^{(0,0,2)}(t)[M^{(1,0,0)}(t)]^2\right. \\
 &\quad \left.+ 4M^{(1,0,1)}(t)M^{(1,0,0)}(t)M^{(0,0,1)}(t)\right\} \\
 &\quad + 6[M^{(1,0,0)}(t)M^{(0,0,1)}(t)]^2 \\
 &\quad + \frac{M_{df}}{6}\left\{\Sigma^{(1,1)}(t) + 4\Sigma^{(1,3)}(t) + \Sigma^{(3,3)}(t)\right\}, \\
 M^{(1,1,2)}(t) &= 2M^{(1,1,1)}(t)M^{(0,0,1)}(t) + M^{(1,0,2)}(t)M^{(0,1,0)}(t) \\
 &\quad + M^{(0,1,2)}(t)M^{(1,0,0)}(t) + M^{(1,1,0)}(t)M^{(0,0,2)}(t) \\
 &\quad + 2M^{(1,0,1)}(t)M^{(0,1,1)}(t) - 2\left\{M^{(1,1,0)}(t)[M^{(0,0,1)}(t)]^2\right. \\
 &\quad \left.+ 2M^{(1,0,1)}(t)M^{(0,1,0)}(t)M^{(0,0,1)}(t)\right. \\
 &\quad \left.+ 2M^{(0,1,1)}(t)M^{(1,0,0)}(t)M^{(0,0,1)}(t)\right. \\
 &\quad \left.+ M^{(0,0,2)}(t)M^{(1,0,0)}(t)M^{(0,1,0)}(t)\right\} \\
 &\quad + 6M^{(1,0,0)}(t)M^{(0,1,0)}(t)[M^{(0,0,1)}(t)]^2 + \frac{M_{df}}{6}\left\{\Sigma^{(1,2)}(t)\right. \\
 &\quad \left.+ 2\Sigma^{(1,3)}(t) + 2\Sigma^{(2,3)}(t) + \Sigma^{(3,3)}(t)\right\}, \\
 M^{(1,0,3)}(t) &= 3M^{(1,0,2)}(t)M^{(0,0,1)}(t) + M^{(0,0,3)}(t)M^{(1,0,0)}(t) \\
 &\quad + 3M^{(1,0,1)}(t)M^{(0,0,2)}(t) - 6\left\{M^{(1,0,1)}(t)[M^{(0,0,1)}(t)]^2\right. \\
 &\quad \left.+ M^{(0,0,2)}(t)M^{(1,0,0)}(t)M^{(0,0,1)}(t)\right\} \\
 &\quad + 6M^{(1,0,0)}(t)[M^{(0,0,1)}(t)]^3 \\
 &\quad + \frac{M_{df}}{2}\left\{\Sigma^{(1,3)}(t) + \Sigma^{(3,3)}(t)\right\}, \\
 M^{(0,2,2)}(t) &= 2M^{(0,2,1)}(t)M^{(0,0,1)}(t) + 2M^{(0,1,2)}(t)M^{(0,1,0)}(t) \\
 &\quad + M^{(0,2,0)}(t)M^{(0,0,2)}(t) + 2[M^{(0,1,1)}(t)]^2 \\
 &\quad - 2\left\{M^{(0,2,0)}(t)[M^{(0,0,1)}(t)]^2\right. \\
 &\quad \left.+ 4M^{(0,1,1)}(t)M^{(0,1,0)}(t)M^{(0,0,1)}(t)\right. \\
 &\quad \left.+ M^{(0,0,2)}(t)[M^{(0,1,0)}(t)]^2\right\} + 6[M^{(0,1,0)}(t)M^{(0,0,1)}(t)]^2 \\
 &\quad + \frac{M_{df}}{6}\left\{\Sigma^{(2,2)}(t) + 4\Sigma^{(2,3)}(t) + \Sigma^{(3,3)}(t)\right\}, \\
 M^{(0,1,3)}(t) &= 3M^{(0,1,2)}(t)M^{(0,0,1)}(t) + M^{(0,0,3)}(t)M^{(0,1,0)}(t) \\
 &\quad + 3M^{(0,1,1)}(t)M^{(0,0,2)}(t) - 6\left\{M^{(0,1,1)}(t)[M^{(0,0,1)}(t)]^2\right. \\
 &\quad \left.+ M^{(0,0,2)}(t)M^{(0,1,0)}(t)M^{(0,0,1)}(t)\right\} \\
 &\quad + 6M^{(0,1,0)}(t)[M^{(0,0,1)}(t)]^3 \\
 &\quad + \frac{M_{df}}{2}\left\{\Sigma^{(2,3)}(t) + \Sigma^{(3,3)}(t)\right\}, \\
 M^{(0,0,4)}(t) &= 4M^{(0,0,3)}(t)M^{(0,0,1)}(t) + 3[M^{(0,0,2)}(t)]^2 \\
 &\quad - 12M^{(0,0,2)}(t)[M^{(0,0,1)}(t)]^2 \\
 &\quad + 6[M^{(0,0,1)}(t)]^4 + M_{df}\Sigma^{(3,3)}(t).
 \end{aligned}$$

Moreover, (110)-(112) yield

$$\bar{b}_1(t_i^-)\Theta_{2,1} + \bar{b}_2(t_i^-)\Theta_{2,2} = \bar{A}_v(t_i^-), \quad (114)$$

$$\bar{b}_2(t_i^-)\Theta_{2,1} + \bar{b}_3(t_i^-)\Theta_{2,2} = \bar{B}_v(t_i^-), \quad (115)$$

where vectors

$$\bar{A}_v(t_i^-) = \mathbb{E}_\pi[\tilde{X}(t_i^-)X_3(t_i^-)] - M^{(0,0,1)}(t_i^-)M^{2v}(t_i^-),$$

$$\bar{B}_v(t_i^-) = \mathbb{E}_\pi[\tilde{X}(t_i^-)X_3^2(t_i^-)] - M^{(0,0,2)}(t_i^-)M^{2v}(t_i^-),$$

scalars

$$\bar{b}_1(t_i^-) = M^{(0,0,2)}(t_i^-) - [M^{(0,0,1)}(t_i^-)]^2 + \sigma_i^2,$$

$$\bar{b}_2(t_i^-) = M^{(0,0,3)}(t_i^-) - M^{(0,0,1)}(t_i^-)M^{(0,0,2)}(t_i^-) + 2\sigma_i^2M^{(0,0,1)}(t_i^-),$$

$$\bar{b}_3(t_i^-) = M^{(0,0,4)}(t_i^-) - [M^{(0,0,2)}(t_i^-)]^2 + 4\sigma_i^2M^{(0,0,2)}(t_i^-) + (h-1)\sigma_i^4,$$

and $\mathbb{E}_\pi[\tilde{X}(t_i^-)X_3(t_i^-)]$ and $\mathbb{E}_\pi[\tilde{X}(t_i^-)X_3^2(t_i^-)]$ are given by (113).

Solving (114)-(115) yields

$$L_{2,1} = \frac{\bar{b}_3(t_i^-)\bar{A}_v(t_i^-) - \bar{b}_2(t_i^-)\bar{B}_v(t_i^-)}{\bar{b}_1(t_i^-)\bar{b}_3(t_i^-) - \bar{b}_2^2(t_i^-)}, \quad (116)$$

$$L_{2,2} = \frac{\bar{b}_1(t_i^-)\bar{B}_v(t_i^-) - \bar{b}_2(t_i^-)\bar{A}_v(t_i^-)}{\bar{b}_1(t_i^-)\bar{b}_3(t_i^-) - \bar{b}_2^2(t_i^-)} \quad (117)$$

and (110) gives

$$\begin{aligned} \Theta_{2,0} &= M^{2v}(t_i^-) - M^{(0,0,1)}(t_i^-)\Theta_{2,1} \\ &\quad - [M^{(0,0,2)}(t_i^-) + \sigma_i^2]\Theta_{2,2}. \end{aligned} \quad (118)$$

Substituting (116)-(118) into (103), one obtains the update approximate for the second moment.

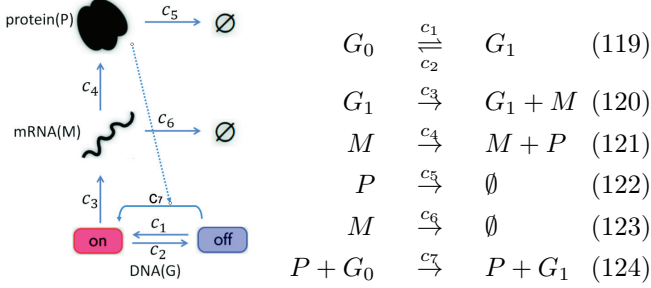
Clearly, (108) and (116)-(118) are multivariate versions of the modified normal moment closure (76) and (78), respectively. It should be pointed out that (108) and (116)-(118) are extensions of the normal moment closure (70) and (72), respectively, when the modification indicator M_{df} is set $M_{df} = 0$.

The normal and modified normal approximations to filtering of gene expression model (92)-(96) and measurements (97) are given in Fig. 4 and Fig. 5, respectively, where reaction rates $c_1 = 0.06 \text{ sec}^{-1}$, $c_2 = 0.03 \text{ sec}^{-1}$, $c_3 = 1 \text{ sec}^{-1}$, $c_4 = 0.05 \text{ sec}^{-1}$, $c_5 = 0.005 \text{ sec}^{-1}$, $c_6 = 0.03 \text{ sec}^{-1}$, measurement noise standard deviation $\sigma = 20 \text{ molec}$, time interval $t_i - t_{i-1} = 50 \text{ sec}$.

S.7.3 An application for nonlinear systems

The gene expression model (92)-(96) is a linear system so that one can easily compute the moment equations and then update the moments by our proposed approximations at the measurement time, recursively. In this case, we can focus on the application of our proposed approximate approach to update formula and do not need to deal with the nonlinearity in the moment equations. Our proposed approach can be applied to nonlinear systems, where some higher order moments

are involved in the computation of moment equations between the measurement times and, like many works (see, e.g., [8, 23, 37, 25, 13] and the references therein), one can apply moment closure techniques to cope with the problem as well. We will illustrate the application of our proposed approach to nonlinear systems with the gene auto regulation model (119)-(124). This nonlinear system describes the transcription and translation of a transcription factor which increases its own synthesis via a positive regulatory feedback, a well known motif occurring in many gene regulatory networks (see, e.g., [12]).



As above, the state of the system can be represented by $X = (X_1 \ X_2 \ X_3)^T = (G_1 \ M \ P)^T \in \mathbb{N}^3$ and the observation is given by (97). The moment equations of this nonlinear system

$$\begin{aligned}
 dM^{(1,0,0)}/dt &= c_1 - c_1 M^{(1,0,0)} - c_2 M^{(0,1,0)} + c_7 M^{(0,0,1)} \\
 &\quad - c_7 M^{(1,0,1)} \\
 dM^{(0,1,0)}/dt &= c_3 M^{(1,0,0)} - c_6 M^{(0,1,0)} \\
 dM^{(0,0,1)}/dt &= c_4 M^{(0,1,0)} - c_5 M^{(0,0,1)} \\
 dM^{(2,0,0)}/dt &= c_1 + c_1 M^{(0,0,1)} + c_2 M^{(0,1,0)} - 2c_1 M^{(2,0,0)} \\
 &\quad - 2c_2 M^{(2,0,0)} + c_7 M^{(0,0,1)} + c_7 M^{(1,0,1)} \\
 &\quad - 2c_7 M^{(2,0,1)} \\
 dM^{(1,1,0)}/dt &= c_1 M^{(0,1,0)} - c_1 M^{(1,1,0)} - c_2 M^{(1,1,0)} \\
 &\quad + c_3 M^{(2,0,0)} - c_6 M^{(1,1,0)} + c_7 M^{(0,1,1)} \\
 &\quad - c_7 M^{(1,1,1)} \\
 dM^{(1,0,1)}/dt &= c_1 M^{(0,0,1)} - c_1 M^{(1,0,1)} - c_2 M^{(1,0,1)} \\
 &\quad + c_4 M^{(1,1,0)} - c_5 M^{(1,0,1)} + c_7 M^{(0,0,2)} \\
 &\quad - c_7 M^{(1,0,2)} \\
 dM^{(0,2,0)}/dt &= c_3 M^{(1,0,0)} + c_6 M^{(0,1,0)} + 2c_3 M^{(1,1,0)} \\
 &\quad - 2c_6 M^{(0,2,0)} \\
 dM^{(0,1,1)}/dt &= c_3 M^{(1,0,1)} - c_6 M^{(0,1,1)} + c_4 M^{(0,2,0)} \\
 &\quad - c_5 M^{(0,1,1)} \\
 dM^{(0,0,2)}/dt &= c_4 M^{(0,1,0)} + c_5 M^{(0,0,1)} + 2c_4 M^{(0,1,1)} \\
 &\quad - 2c_5 M^{(0,0,2)}
 \end{aligned}$$

involve the higher order moments $M^{(2,0,1)}$, $M^{(1,1,1)}$ and $M^{(1,0,2)}$.

The bimodality of this nonlinear system is verified with stochastic simulation algorithm over $T =$

500 hours in Fig. 6 and Fig. 7, where reaction rates $c_1 = 0.002 \text{ sec}^{-1}$, $c_2 = 0.08 \text{ sec}^{-1}$, $c_3 = 1.1 \text{ sec}^{-1}$, $c_4 = 0.16 \text{ sec}^{-1}$, $c_5 = 0.03 \text{ sec}^{-1}$, $c_6 = 0.03 \text{ sec}^{-1}$, $c_7 = 0.001 \text{ molec}^{-1} \cdot \text{sec}^{-1}$. Applying moment closures

$$\begin{aligned}
 M^{(2,0,1)}(t) &= M^{(2,0,0)}(t)M^{(0,0,1)}(t) + 2M^{(1,0,1)}(t)M^{(1,0,0)}(t) \\
 &\quad - 2[M^{(1,0,0)}(t)]^2 M^{(0,0,1)}(t), \\
 M^{(1,1,1)}(t) &= M^{(1,1,0)}(t)M^{(0,0,1)}(t) + M^{(1,0,1)}(t)M^{(0,1,0)}(t) \\
 &\quad + M^{(0,1,1)}(t)M^{(1,0,0)}(t) - 2M^{(1,0,0)}(t)M^{(0,1,0)}(t)M^{(0,0,1)}(t), \\
 M^{(1,0,2)}(t) &= 2M^{(1,0,1)}(t)M^{(0,0,1)}(t) + M^{(0,0,2)}(t)M^{(1,0,0)}(t)
 \end{aligned}$$

and update formula (87) and (101)-(102) gave the approximates to filtering in Fig. 8 and Fig. 9, where measurement noise standard deviation $\sigma = 15 \text{ molec}$, time interval $t_i - t_{i-1} = 50 \text{ sec}$.

S.8 An RTS approximation to smoothing for the first order reaction networks

There are smoothing techniques such as RTS and M B-F smoothers in the literature, e.g., [4]. Here we consider the RTS smoother since the RTS smoother has some desired properties (e.g., stability) and some other smoothers are based upon it, see [1, 4]. Similarly, for a system (1) with $d \geq 1$, denote the first and the second smoothing moments by

$$\begin{aligned}
 \tilde{M}(t) &= \tilde{M}^1(t) = \mathbb{E}_{\tilde{\pi}}[X(t)] \in \mathbb{R}^d, \\
 \tilde{M}^2(t) &= \mathbb{E}_{\tilde{\pi}}[X(t)X^T(t)] \in \mathbb{R}^{d \times d}, \quad (125)
 \end{aligned}$$

and then posterior covariance

$$\begin{aligned}
 \tilde{\Sigma}(t) &= \text{Cov}[X(t) | \{y_1, \dots, y_N\}] \\
 &= \tilde{M}^2(t) - \tilde{M}(t)\tilde{M}^T(t) \in \mathbb{R}^{d \times d}. \quad (126)
 \end{aligned}$$

Due to the symmetry, the second smoothing moment can be alternatively represented by

$$\tilde{M}^{2v}(t) = \mathbb{E}_{\tilde{\pi}}[\tilde{X}(t)] \in \mathbb{R}^{d(d+1)/2}. \quad (127)$$

Clearly, by Theorem 2, $\tilde{M}(T) = M(T)$, $\tilde{M}^2(T) = M^2(T)$, $\Sigma_{\tilde{\pi}}(T) = \Sigma(T)$ and $\tilde{M}^{2v}(T) = M^{2v}(T)$. Moreover, by (38), one can describe dynamics of smoothing mean with the ODE

$$\frac{d}{dt} \mathbb{E}_{\tilde{\pi}}[X(t)] = \sum_{j=1}^r \nu_j \mathbb{E}_{\tilde{\pi}}[\tilde{\alpha}_j(X(t))] \quad (128)$$

with $\mathbb{E}_{\tilde{\pi}}[X(0)] = \mathbb{E}_p[X(0)]$.

The approximates for $M(t)$, $M^2(t)$ and $\Sigma_{\pi}(t)$, for $t \in [0, T]$ (with $T = t_N$), can be obtained by the approach presented above. In this section, we consider the smoother for the first order reaction networks (47). Let $\Psi(\bar{t}, \tau)$ be the state transition matrix corresponding to \hat{C}_1 given by (49) with $\bar{L} = 1$, that is,

$$\frac{d}{d\tau} \Psi(\bar{t}, \tau) = -\Psi(\bar{t}, \tau) \hat{C}_1; \quad \Psi(\bar{t}, \tau)|_{\tau=\bar{t}} = I \quad (129)$$

which yields $\Psi(\bar{t}, \tau) = e^{(\bar{t}-\tau)\hat{C}_1}$ for all $\bar{t} \geq \tau$.

Our computation exploits the RTS discrete smoother equations, see [33, 4],

$$\tilde{M}(t_i) = M(t_i) - K_{sm}(t_i) [M(t_{i+1}^-) - \tilde{M}(t_{i+1})] \quad (130)$$

$$\begin{aligned} \tilde{\Sigma}(t_i) &= \Sigma(t_i) \\ &\quad - K_{sm}(t_i) [\Sigma_\pi(t_{i+1}^-) - \tilde{\Sigma}(t_{i+1})] K_{sm}^T(t_i) \end{aligned} \quad (131)$$

at observation time $t_i < T$, where smoother gain $K_{sm}(t_i) = \Sigma(t_i) \Psi^T(t_{i+1}, t_i) [\Sigma(t_{i+1}^-)]^{-1}$.

As pointed out in [33, 1], the equations (130)-(131) hold for $t \in [0, T)$, that is,

$$\tilde{M}(t) = M(t) - K_{sm}(t) [M(t_k^-) - \tilde{M}(t_k)] \quad (132)$$

$$\begin{aligned} \tilde{\Sigma}(t) &= \Sigma(t) \\ &\quad - K_{sm}(t) [\Sigma(t_k^-) - \tilde{\Sigma}(t_k)] K_{sm}^T(t) \end{aligned} \quad (133)$$

with smoother gain

$$K_{sm}(t) = \Sigma(t) \Psi^T(t_k, t) [\Sigma(t_k^-)]^{-1} \quad (134)$$

and $K_{sm}(t_k^-) = I$, where, as above, $k = \min(k \in \mathbb{N} \mid t_k > t)$, and the fact that no measurements occur within the interval between observations is used. A systematic application of these equations will be used to derive differential equations describing the approximate smoothing process. For $T > t \geq 0$, (132)-(134) gives the backward ODEs for the approximate smoothing

$$\begin{aligned} \frac{d}{dt} \tilde{M}(t) &= \frac{d}{dt} M(t) \\ &\quad - \frac{d}{dt} K_{sm}(t) [M(t_k^-) - \tilde{M}(t_k)] \end{aligned} \quad (135)$$

$$\begin{aligned} \frac{d}{dt} \tilde{\Sigma}(t) &= \frac{d}{dt} \Sigma(t) - \left[\frac{d}{dt} K_{sm}(t) \right] \\ &\quad \cdot [\Sigma_\pi(t_k^-) - \tilde{\Sigma}(t_k)] K_{sm}^T(t) - K_{sm}(t) \\ &\quad \cdot [\Sigma_\pi(t_k^-) - \tilde{\Sigma}(t_k)] \left[\frac{d}{dt} K_{sm}^T(t) \right] \end{aligned} \quad (136)$$

with terminal conditions $\tilde{M}(T) = M(T)$ and $\tilde{\Sigma}(T) = \Sigma_\pi(T) = M^2(T) - M(T)M^T(T)$ obtained in the previous section, where

$$\begin{aligned} \frac{d}{dt} K_{sm}(t) &= \left[\frac{d}{dt} \Sigma_\pi(t) - \Sigma_\pi(t) C_1^T \right] \\ &\quad \Psi^T(t_k, t) [\Sigma(t_k^-)]^{-1}, \end{aligned} \quad (137)$$

$K_{sm}(t_k^-) = I$. Note that $\frac{d}{dt} M(t)$ and $\frac{d}{dt} \Sigma_\pi(t) = \frac{d}{dt} [M^2(t) - M(t)M^T(t)]$ can be immediately inferred by (51).

As an example for applications, we consider the three stage gene expression model (92)-(96). Note that the state of

system is represented by $X = (X_1 \ X_2 \ X_3)^T = (G_1 \ M \ P)^T$. Therefore, we have \hat{C}_L and b_L in (48) with $\bar{L} = 1$

$$\hat{C}_1 = \begin{pmatrix} -(c_1 + c_2) & 0 & 0 \\ c_3 & -c_6 & 0 \\ 0 & c_4 & -c_5 \end{pmatrix}, \quad b_1 = \begin{pmatrix} c_1 \\ 0 \\ 0 \end{pmatrix} \quad (138)$$

and the state transition matrix $\Psi(\cdot, \cdot)$ in (129)

$$\Psi(t_k, t) = e^{(t_k - t)\hat{C}_1} \quad (139)$$

for all $t \in [0, T)$. The normal and modified normal RTS approximates to smoothing of gene expression model (92)-(96) and measurements (97) (with reaction rates $c_1 = 0.06 \text{ sec}^{-1}$, $c_2 = 0.03 \text{ sec}^{-1}$, $c_3 = 1 \text{ sec}^{-1}$, $c_4 = 0.05 \text{ sec}^{-1}$, $c_5 = 0.005 \text{ sec}^{-1}$, $c_6 = 0.03 \text{ sec}^{-1}$, measurement noise standard deviation $\sigma = 20 \text{ molec}$, time interval $t_i - t_{i-1} = 50 \text{ sec}$) are given in Fig. 10 and Fig. 11, which are based on the approximations to filtering in Fig. 4 and Fig. 5, respectively.

Remark 2: The RTS smoothers are usually applied to linear systems, see [1, 4]. For the nonlinear system, the gene auto regulation model (119)-(124), it would be appropriate to apply the RTS approximation to smoothing when our proposed approach is extended to a hybrid framework because the subsystems in the hybrid setting are linear, see, e.g., [12, 30].

S.9 Application to Lognormal measurement models

In the previous sections, approximate approaches have been presented for CTMC with discrete-time observations, where the observations are of the form with additive noise. These approximate approaches can be applied to some other observation models. In this section, we extend the approaches to the cases when the observations of CTMC are Lognormal measurement models. As above, we first obtain a result for univariate systems and generalize it to multivariate systems. Moreover, as an application example, we only consider the modified normal approximation in this section. Based on the filtering approximates, we also consider the RTS approximates to smoothing for the first order reaction networks.

S.9.1 Filtering approximates for univariate systems

Consider a univariate system (1) with lognormal measurements of the form

$$y_i = e^{w_i} X(t_i) \quad (140)$$

where $\{w_i\}$ is an independent sequence with $w_i \sim \mathcal{N}(0, \sigma_i^2)$, which is independent of $\{X(t)\}$. Recall that the update approximate formula are given by

$$M^\eta(t_i) = \sum_{q=0}^{\eta} \vartheta_{\eta,q} y_i^q \quad (141)$$

for $\eta \in \mathbb{N}$ and observation time t_i , where $m_{\eta,q}$ are constants at time t_i^- . As above, let $s(x(t_i)) = x^\eta(t_i)$ and $u(y_i) = y_i^j$ with $j = 0, 1, \dots, \eta$ in (54). Then

$$\begin{aligned} & \mathbb{E} \left[x^\eta(t_i) y_i^j \mid \{y_1, \dots, y_{i-1}\} \right] \\ &= \mathbb{E} \left[x^{\eta+j}(t_i) e^{j w_i} \mid \{y_1, \dots, y_{i-1}\} \right] \\ &= \mathbb{E} \left[e^{j w_i} \right] M^{\eta+j}(t_i^-) = e^{\frac{1}{2} j^2 \sigma_i^2} M^{\eta+j}(t_i^-), \end{aligned} \quad (142)$$

$$\begin{aligned} & \mathbb{E} \left[\mathbb{E} [x^\eta(t_i) \mid \{y_1, \dots, y_i\}] y_i^j \mid \{y_1, \dots, y_{i-1}\} \right] \\ &= \mathbb{E} \left[\sum_{q=0}^{\eta} \vartheta_{\eta,q} y_i^q y_i^j \mid \{y_1, \dots, y_{i-1}\} \right] \\ &= \mathbb{E} \left[\sum_{q=0}^{\eta} \vartheta_{\eta,q} e^{(q+j) w_i} x^{q+j}(t_i) \mid Y_1^{n-1} \right] \\ &= \sum_{q=0}^{\eta} \vartheta_{\eta,q} e^{\frac{1}{2} (q+j)^2 \sigma_i^2} M^{q+j}(t_i^-). \end{aligned} \quad (143)$$

And (54) leads to

$$\sum_{q=0}^{\eta} \vartheta_{\eta,q} e^{\frac{1}{2} (q+j)^2 \sigma_i^2} M^{q+j}(t_i^-) = e^{\frac{1}{2} j^2 \sigma_i^2} M^{\eta+j}(t_i^-) \quad (144)$$

with $j = 0, 1, \dots, \eta$, which determines coefficients $m_{\eta,q}$, $q = 0, 1, \dots, \eta$.

Solving (144) for $\eta = 1$ yields

$$M^1(t_i) = \vartheta_{1,0} + \vartheta_{1,1} y_i \quad (145)$$

where

$$\begin{aligned} \vartheta_{1,0} &= \frac{(e^{\sigma_i^2} - 1) M^2(t_i^-) M^1(t_i^-)}{(e^{\sigma_i^2} - 1) M^2(t_i^-) + \Sigma(t_i^-)}, \\ \vartheta_{1,1} &= \frac{e^{-\sigma_i^2/2} \Sigma(t_i^-)}{(e^{\sigma_i^2} - 1) M^2(t_i^-) + \Sigma(t_i^-)}. \end{aligned} \quad (146)$$

This gives

$$\vartheta_{1,0} \rightarrow 0, \quad \vartheta_{1,1} \rightarrow 1, \quad M^1(t_i) \rightarrow y_i \quad \text{as } e^{\sigma_i^2} \rightarrow 1$$

and

$$m_{1,0} \rightarrow M^1(t_i^-), \quad \vartheta_{1,1} \rightarrow 0, \quad M^1(t_i) \rightarrow M^1(t_i^-) \quad \text{as } e^{\sigma_i^2} \rightarrow \infty$$

for the cases with noise-free ($\sigma_i^2 = 0$) and extremely-large-noise ($\sigma_i^2 \rightarrow \infty$) measurements, respectively.

The update of approximation to the second moment can be obtained by the approach (79) and (80). At the measurement time t_i , let

$$\Sigma(t_i) = m \Sigma(t_i^-) \quad (147)$$

and hence $M^2(t_i) = [M^1(t_i)]^2 + m \Sigma(t_i^-)$, where m is a coefficient to be determined. Then (141) with (145) and (147) gives

$$\begin{aligned} \vartheta_{2,0} &= \vartheta_{1,0}^2 + m \Sigma(t_i^-), \\ \vartheta_{2,1} &= 2\vartheta_{1,0}\vartheta_{1,1}, \quad \vartheta_{2,2} = \vartheta_{1,1}^2. \end{aligned} \quad (148)$$

Letting $s(x(t_i)) = x^2(t_i)$, $u(y_i) = 1$ with $j = 0, 1, \dots, \eta$ in (54) and using (141) with $\vartheta_{2,0} = \vartheta_{1,0}^2 + m \Sigma(t_i^-)$, one has

$$\begin{aligned} m &= \frac{1}{\Sigma(t_i^-)} \left[M^2(t_i^-) - \vartheta_{2,2} e^{2\sigma_i^2} M^2(t_i^-) \right. \\ &\quad \left. - \vartheta_{2,1} e^{\sigma_i^2/2} M^1(t_i^-) - m_{1,0}^2 \right]. \end{aligned} \quad (149)$$

Substitution of (146) and (148) into (149) yields

$$m = \frac{(e^{\sigma_i^2} - 1) M^2(t_i^-)}{(e^{\sigma_i^2} - 1) M^2(t_i^-) + \Sigma(t_i^-)}. \quad (150)$$

Then (146), (148) and (150) imply

$$\begin{aligned} \vartheta_{2,2} &\rightarrow 1, \quad \vartheta_{2,1} \rightarrow 0, \quad \vartheta_{2,0} \rightarrow 0, \quad m \rightarrow 0, \quad \Sigma(t_i) \rightarrow 0, \\ M^2(t_i) &\rightarrow y_i^2 \quad \text{as } e^{\sigma_i^2} \rightarrow 1 \end{aligned}$$

and

$$\begin{aligned} m_{2,2} &\rightarrow 0, \quad \vartheta_{2,1} \rightarrow 0, \quad \vartheta_{2,0} \rightarrow M^2(t_i^-), \quad m \rightarrow 1, \\ \Sigma(t_i) &\rightarrow \Sigma(t_i^-), \quad M^2(t_i) \rightarrow M^2(t_i^-) \quad \text{as } e^{\sigma_i^2} \rightarrow \infty \end{aligned}$$

for the cases with noise-free and extremely-large-noise measurements, respectively.

Let us consider the birth-death model (83) and lognormal measurement (140) (with reaction rates $c_1 = 10 \text{ molec} \cdot \text{sec}^{-1}$, $c_2 = 0.1 \text{ sec}^{-1}$, measurement noise standard deviation $\sigma = 0.2$, time interval $t_i - t_{i-1} = 25 \text{ sec}$), for which a typical realization of the above schemes are given in Fig. 12.

S.9.2 Filtering approximates for multivariate systems

In this section, we extend our proposed formula (145) and (141) to a multivariate system with lognormal measurements

$$y_i = e^{W_i} F X(t_i), \quad (151)$$

where $y_i \in \mathbb{R}^l$, $l \leq d$, $F \in \mathbb{R}^{l \times d}$ with full rank l , $\{W_i\}$ with $W_i = \text{diag}\{w_{i,1}, \dots, w_{i,l}\} \in \mathbb{R}^{l \times l}$ and $\{w_i\}$ is an independent sequence of random variables. For simplicity, assume $W_i \sim \mathcal{N}(0, \sigma_i^2 I_l)$ and I_l is the identity matrix of order l . Our proposed method can be extended to more complicated cases such as $W_i \sim \mathcal{N}(0, R_w)$. Given matrix $A = (a_{jq})_{l \times l} \in \mathbb{R}^{l \times l}$, let A_{Diag} be

$$A_{\text{Diag}} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{qq} \end{pmatrix}. \quad (152)$$

Recall that

$$M(t_i) = \Theta_{1,0} + \Theta_{1,1}y_i \quad (153)$$

where $L_{1,0} \in \mathbb{R}^d$ and $L_{1,1} \in \mathbb{R}^{d \times q}$. Let $s(x(t_i)) = x(t_i)$ and $u(y_i) = 1$, then (54) gives

$$M(t_i^-) = \Theta_{1,0} + e^{\frac{1}{2}\sigma_i^2}\Theta_{1,1}FM(t_i^-). \quad (154)$$

Setting $s(x(t_i)) = x(t_i)$ and $u(y_i) = y_i^T$ in (54), we have

$$\begin{aligned} \mathbb{E}\{x(t_i)y_i^T|Y_1^{n-1}\} &= \mathbb{E}\{x(t_i)x^T(t_i)F^Te^{W_i}|Y_1^{n-1}\} \\ &= M^2(t_i^-)F^Te^{\frac{1}{2}\sigma_i^2I_q}, \quad (155) \\ \mathbb{E}\{\mathbb{E}[x(t_i)|Y_1^n]y_i^T|Y_1^{n-1}\} &= \mathbb{E}\{(\Theta_{1,0} + \Theta_{1,1}y_i)y_i^T|Y_1^{n-1}\} \\ &= \mathbb{E}\{\Theta_{1,0}[x^T(t_i)F^Te^{W_i}]|Y_1^{n-1}\} \\ &\quad + \mathbb{E}\{\Theta_{1,1}[e^{W_i}Fx(t_i)][x^T(t_i)F^Te^{W_i}]|Y_1^{n-1}\} \\ &= e^{\frac{1}{2}\sigma_i^2}\Theta_{1,0}M^T(t_i^-)F^T + e^{\sigma_i^2}\Theta_{1,1}[FM^2(t_i^-)F^T \\ &\quad + (e^{\sigma_i^2} - 1)(FM^2(t_i^-)F^T)_{Diag}] \quad (156) \end{aligned}$$

and therefore

$$\begin{aligned} M^2(t_i^-)F^T &= \Theta_{1,0}M^T(t_i^-)F^T + e^{\frac{1}{2}\sigma_i^2}\Theta_{1,1} \\ &\cdot [FM^2(t_i^-)F^T + (e^{\sigma_i^2} - 1)(FM^2(t_i^-)F^T)_{Diag}]. \quad (157) \end{aligned}$$

Combining (85), (154) and (157), we obtain

$$\begin{aligned} \Theta_{1,1} &= e^{-\sigma_i^2/2}\Sigma(t_i^-)F^T \left\{ F\Sigma(t_i^-)F^T \right. \\ &\quad \left. + (e^{\sigma_i^2} - 1)(FM^2(t_i^-)F^T)_{Diag} \right\}^{-1}. \quad (158) \end{aligned}$$

And then, by (154), we have

$$\Theta_{1,0} = M(t_i^-) - e^{\sigma_i^2/2}\Theta_{1,1}FM(t_i^-). \quad (159)$$

For the update approximates of the second moments, a multivariate version of approximation (148) and (150) can be obtained by the approach (147). At measurement time t_i , let

$$\Sigma(t_i) = [\Sigma(t_i^-)]^{1/2}Q[\Sigma(t_i^-)]^{1/2}, \quad (160)$$

where $Q \in \mathbb{R}^{d \times d}$ is a symmetric matrix to be determined.

Letting $s(x(t_i)) = x(t_i)x^T(t_i)$, $u(y_i) = 1$ in (54) and using (153) and (160), we have

$$\begin{aligned} M^2(t_i^-) &= \mathbb{E}[\Sigma(t_i) + M(t_i)M^T(t_i)|Y_1^{n-1}] \\ &= [\Sigma(t_i^-)]^{1/2}Q[\Sigma(t_i^-)]^{1/2} + \Theta_{1,0}\Theta_{1,0}^T \\ &\quad + e^{\frac{1}{2}\sigma_i^2}\Theta_{1,0}M^T(t_i^-)F^T\Theta_{1,1}^T + e^{\frac{1}{2}\sigma_i^2}\Theta_{1,1}FM(t_i^-)\Theta_{1,0}^T \\ &\quad + e^{\sigma_i^2}\Theta_{1,1} \left[FM^2(t_i^-)F^T + (e^{\sigma_i^2} - 1)(FM^2(t_i^-)F^T)_{Diag} \right] \Theta_{1,1}^T \end{aligned}$$

and hence

$$\begin{aligned} Q &= [\Sigma(t_i^-)]^{-1/2} \left\{ M^2(t_i^-) - \Theta_{1,0}\Theta_{1,0}^T - e^{\frac{1}{2}\sigma_i^2}\Theta_{1,0} \right. \\ &\quad \cdot M^T(t_i^-)F^T\Theta_{1,1}^T - e^{\sigma_i^2}\Theta_{1,1} \left[FM^2(t_i^-)F^T + (e^{\sigma_i^2} - 1) \right. \\ &\quad \cdot (FM^2(t_i^-)F^T)_{Diag} \left. \right] \Theta_{1,1}^T \left. \right\} [\Sigma(t_i^-)]^{-1/2}. \quad (161) \end{aligned}$$

It is observed that that (153) and (161) are the multivariate versions of (145) and (149), respectively. Substitution of (161) into (160) gives

$$\begin{aligned} \Sigma(t_i) &= M^2(t_i^-) - \Theta_{1,0}\Theta_{1,0}^T - e^{\frac{1}{2}\sigma_i^2}\Theta_{1,0}M^T(t_i^-)F^T\Theta_{1,1}^T \\ &\quad - e^{\frac{1}{2}\sigma_i^2}\Theta_{1,1}FM(t_i^-)\Theta_{1,0} - e^{\sigma_i^2}\Theta_{1,1} \left[FM^2(t_i^-)F^T \right. \\ &\quad \left. + (e^{\sigma_i^2} - 1)(FM^2(t_i^-)F^T)_{Diag} \right] \Theta_{1,1}^T \quad (162) \end{aligned}$$

and hence

$$\begin{aligned} M^2(t_i) &= \Sigma(t_i) + M(t_i)M^T(t_i) \\ &= M^2(t_i^-) - \Theta_{1,0}\Theta_{1,0}^T - e^{\frac{1}{2}\sigma_i^2}\Theta_{1,0}M^T(t_i^-)F^T\Theta_{1,1}^T \\ &\quad - e^{\frac{1}{2}\sigma_i^2}\Theta_{1,1}FM(t_i^-)\Theta_{1,0} - e^{\sigma_i^2}\Theta_{1,1} \left[FM^2(t_i^-)F^T \right. \\ &\quad \left. + (e^{\sigma_i^2} - 1)(FM^2(t_i^-)F^T)_{Diag} \right] \Theta_{1,1}^T \\ &\quad + \Theta_{1,0}y_i^T\Theta_{1,1}^T + \Theta_{1,1}y_i\Theta_{1,0} + \Theta_{1,1}y_iy_i^T\Theta_{1,1}^T, \quad (163) \end{aligned}$$

where $L_{1,0}$ and $L_{1,1}$ are given by (159) and (158), respectively.

When the approximation to filtering is obtained, one can also compute the approximation to smoothing such as the RTS approximation.

As an illustrative example, we consider the gene expression model (92)-(96) with lognormal measurement

$$y_i = e^{w_i}X_3(t_i) \quad (164)$$

and $w_i \sim \mathcal{N}(0, \sigma_i^2)$. The approximations to filtering and smoothing are given by Fig. 13 and Fig. 14, respectively, where reaction rates $c_1 = 0.06 \text{ sec}^{-1}$, $c_2 = 0.03 \text{ sec}^{-1}$, $c_3 = 1 \text{ sec}^{-1}$, $c_4 = 0.05 \text{ sec}^{-1}$, $c_5 = 0.005 \text{ sec}^{-1}$, $c_6 = 0.03 \text{ sec}^{-1}$, measurement noise standard deviation $\sigma = 0.2$, time interval $t_i - t_{i-1} = 50 \text{ sec}$.

Clearly, our proposed method can be also applied to time-varying systems such as gene expression model (92)-(96) with time-varying parameters, e.g., $c_1 = c_1(t)$. That we compare the results between our proposed method and the exact computation method with silico and wet lab experimental data sets show that our proposed method works well in practice.

References

- [1] Bierman GJ. 1973 Fixed interval smoothing with discrete measurements, *Int. J. Control*, 18: 65-75.
- [2] Borisov AV. 2011 The Wonham filter under uncertainty: a game-theoretic approach, *Automatica*, 47: 1015-1019.
- [3] Budhiraja A, Chen L, Lee C. 2007 A survey of numerical methods for nonlinear filtering problems, *Physica D*, 230: 27-36.

- [4] Crassidis JL, Junkins JL. 2012 Optimal estimation of dynamic systems (second edition). U.S.: Chapman & Hall/CRC.
- [5] Davis MHA. 1977 Linear estimation and stochastic control. London, U.K.: Chapman & Hall.
- [6] Eltoft T, Kim T, Lee T-W. 2006 On the multivariate Laplace distribution, *IEEE Sign. Proc. Letter*, 13: 300-303.
- [7] Ganguly A, Altintan D, Koepl H. 2015 Jump-diffusion approximation of stochastic reaction dynamics: error bounds and algorithms, *Multiscale Model. Simul.* 13: 1390-1419.
- [8] Gillespie CS. 2009 Moment-closure approximations for mass-action models, *IET Syst. Biol.*, 3: 52-58.
- [9] Gillespie DT. 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *J. Comput. Phys.*, 22: 403-434.
- [10] Gillespie DT. 1977 Exact stochastic simulation of coupled chemical reactions, *J. Chem. Phys.*, 81: 2340-2361.
- [11] Gillespie DT. 1992 A rigorous derivation of the chemical master equation, *Physica A*, 188: 404-425.
- [12] Hasenauer J, Wolf V, Kazeroonian A, Theis FJ. 2014 Method of conditional moments (MCM) for the Chemical Master Equation: a unified framework for the method of moments and hybrid stochastic-deterministic models. *J Math Biol.* 69: 687-735.
- [13] Hausken K, J.F. Moines JF. 2011 Systematization of a set of closure techniques, *Theoretical Population Biology*, 80: 175-184.
- [14] Hespanha J. 2008 Moment closure for biochemical networks, *Proc. 3rd Int. Symp. on Control, Communications and Signal Processing*.
- [15] Higham DJ. 2008 Modeling and simulating chemical reactions, *SIAM Review*, 50: pp.347-368.
- [16] Higham DJ, R. Khanin R. 2008 Chemical master versus chemical Langevin for first-order reaction networks, *The Open Applied Mathematics Journal*, 2: 59-79.
- [17] Jahnke T. 2011 On Reduced models for the chemical master equation, *SIAM Multiscale Model. Simul.*, 9: 1646-1676.
- [18] Jahnke T, Kreim M. 2012 Error bound for piecewise deterministic processes modeling stochastic reaction systems, *SIAM Multiscale Model. Simul.*, 10: 1119-1147.
- [19] Jazwinski AH. 1970 Stochastic processes and filtering theory. New York, U.S.: Academic Press.
- [20] Keeling MJ. 2000 Multiplicative moments and measures of persistence in ecology, *J. Theor. Biol.*, 205: 269-281.
- [21] Koch AL. 1966 The logarithm in biology 1. Mechanisms generating the log-normal distribution exactly, *J. Theor. Biol.*, 12: 276-290.
- [22] Kotz S, Kozubowski T, Podgorski K. 2001 The Laplace distribution and generalizations. Boston, U.S.: Birkhäuser.
- [23] Krishnarajah I, Cook A, Marion G, Bibson G. 2005 Novel moment closure approximations in stochastic epidemics, *Bull. Math. Biol.*, 67: 855-873.
- [24] Kurtz TG. 1972 The relationship between stochastic and deterministic models for chemical reactions, *J. Chem. Phys.*, 57: 2976-2978.
- [25] Lee CH, Kim K-H, Kim P. 2009 A moment closure method for stochastic reaction networks, *J. Chem. Phys.*, 130: 134107.
- [26] Limpert E, Stahel WA, Abbt M. 2001 Log-normal Distributions across the Sciences: keys and clues, *BioScience*, 51:341-352.
- [27] Lütkepohl H. 1996 Handbook of Matrices. Chichester, U.K.: Wiley.
- [28] Matis T, Guardiola I. 2010 Achieving moment closure through cumulant neglect, *The Mathematica Journal* (doi:10.3888/tmj.12|2)
- [29] McReynolds SR. 1990 Fixed interval smoothing: revisited, *Guidance*, 13: 913-921.
- [30] Menz S, Latorre JC, Schuütte C, Huisinga W. 2012 Hybrid stochastic deterministic solution of the chemical master equation. *SIAM J Multiscale Model Simul.* 10: 1232-1262.
- [31] Miller KS. 1964 Multidimensional Gaussian distributions. New York, U.S.: John Wiley & Sons.
- [32] Purdom E, Holmes SP. 2005 Error distribution for gene expression data, *Stat. Appl. Gene. Mol. Biol.*, 4.
- [33] Rauch HE, Tung F, Striebel C. 1965 Maximum likelihood estimates of linear dynamic systems, *AIAA Journal*, 3: 1445-1450.
- [34] Ruttor A, Oppen M. 2009 Efficient statistical inference for stochastic reaction processes. *Phys. Rev. Lett.*, 103(23): 230601.
- [35] Sasvári Z. 2013 Multivariate characteristic and correlation functions. Germany: Walter de Gruyter GmbH.

- [36] Singh A, Hespanha JP. 2006 Lognormal moment closures for biochemical reactions, *Proc. 45th IEEE Conf. Decision and Control*.
- [37] Singh A, Hespanha JP. 2007 A derivative matching approach to moment-closure approximations, *Bull. Math. Biol.*, 69:1909-1925.
- [38] Whittle P. 1957 On the use of the normal approximation in the treatment of stochastic processes, *J. Roy. Stat. Societ.*, 19: 268–281.
- [39] Wilhelm T. 2009 The smallest chemical reaction system with bistability, *BMC Systems Biology*, 3:90
- [40] Wilkinson D. Stochastic modelling for systems biology. U.S.: Chapman & Hall/CRC.
- [41] Williams D. 1991 Probability with martingales. U.K.: Cambridge Univ. Press.
- [42] Wolf V, Geol R, M. Mateescu M, Henzinger TA. 2010 Solving the chemical master equation using sliding windows, *BMC Systems Biology*.
- [43] Wonham WM. 1965 Some applications of stochastic differential equations to optimal nonlinear filtering, *SIAM J. Control*, 2: 347-369.
- [44] Yin G, Zhang Q, Liu YJ. 2004 Discrete-time approximation of Wonham filters, *J. Control Theory Appl.*, 2: 1-10.
- [45] Zechner C, Unger M, Pelet S, Peter M, Koeppl H. 2014 Scalable inference of heterogeneous reaction kinetics for pooled single-cell recordings. *Nature Methods* 11: 197-202.
- [46] Zhang Q, Yin G, Moore JB. 2007 Two-time-scale approximation for Wonham filters, *IEEE Trans. Information Theory*, 53: 1706-1715.

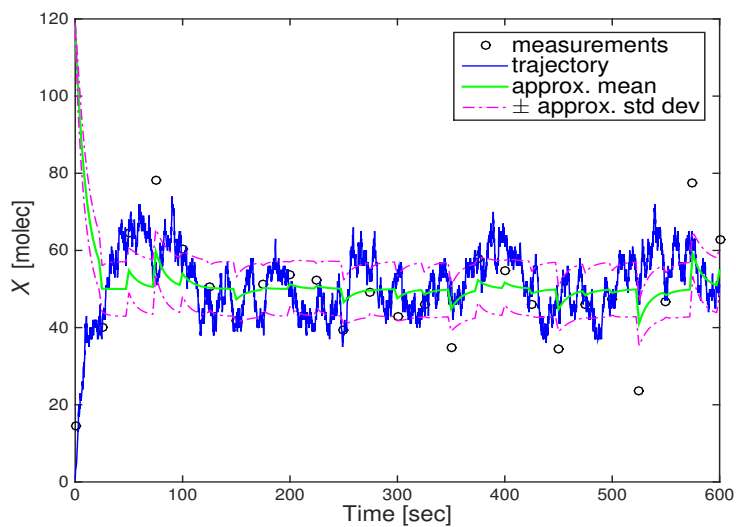


Fig. 1: Normal approximate for the birth-death model.

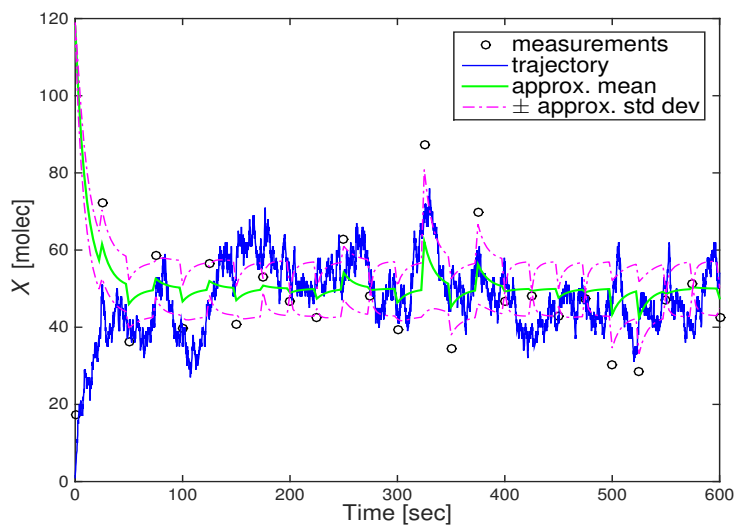


Fig. 2: Lognormal approximate for the birth-death model.

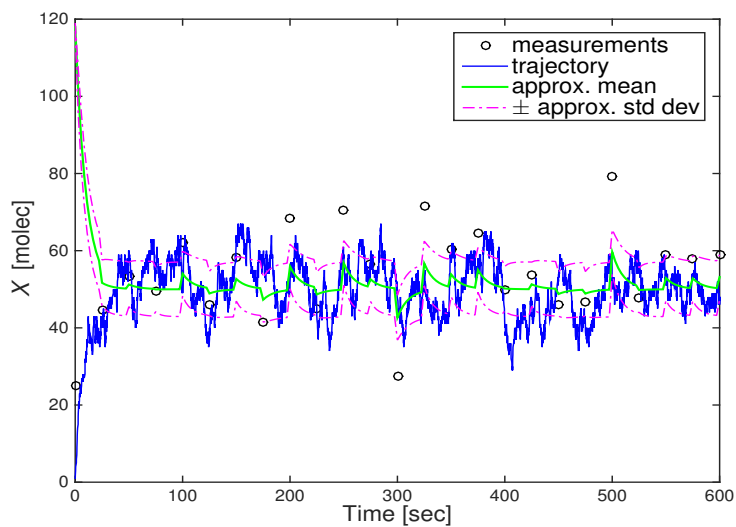


Fig. 3: Modified Normal approximate for the birth-death model.

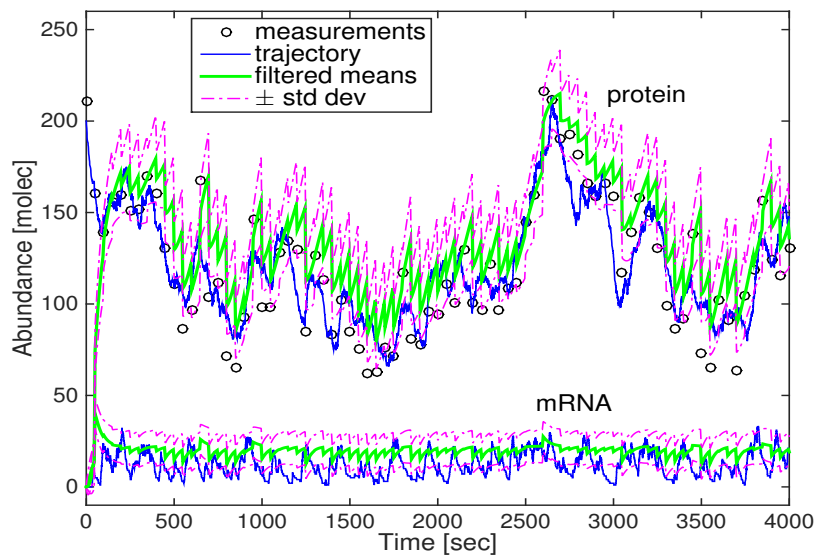


Fig. 4: Normal approximate to filtering for the gene expression model.

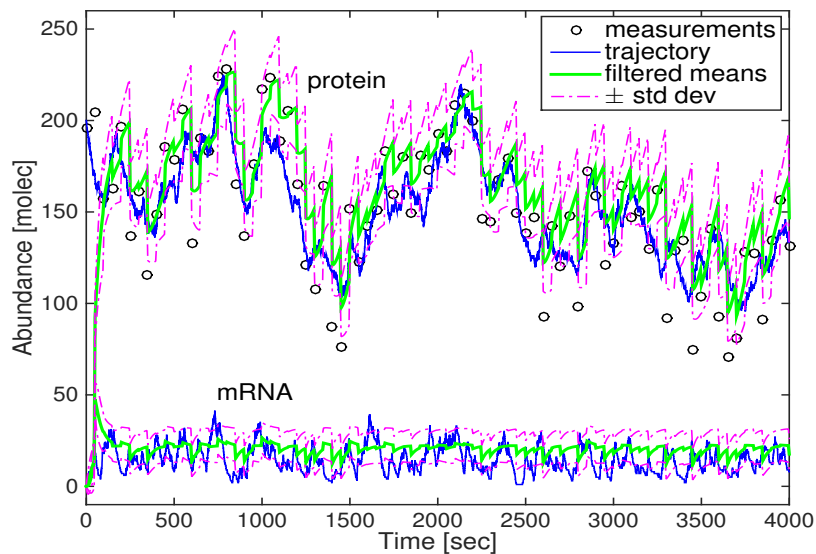


Fig. 5: Modified Normal approximate to filtering for the gene expression model.

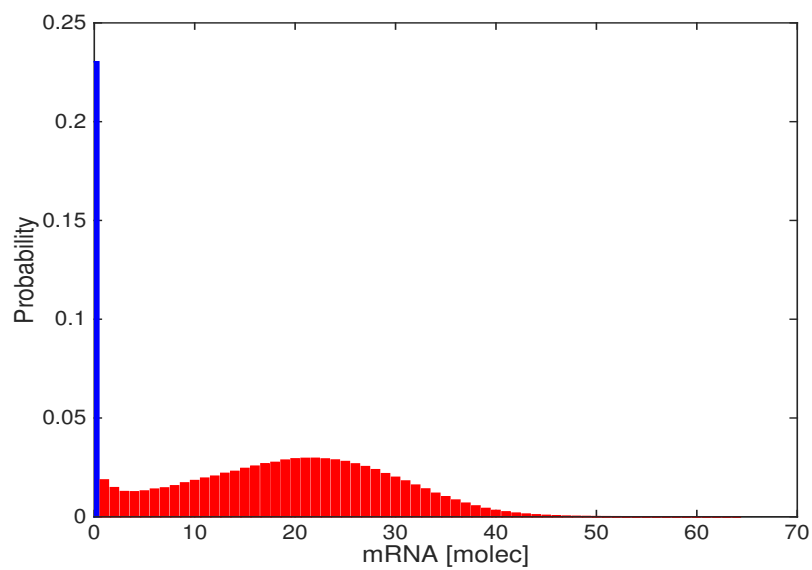


Fig. 6: Steady distribution of mRNA for the gene auto regulation model.

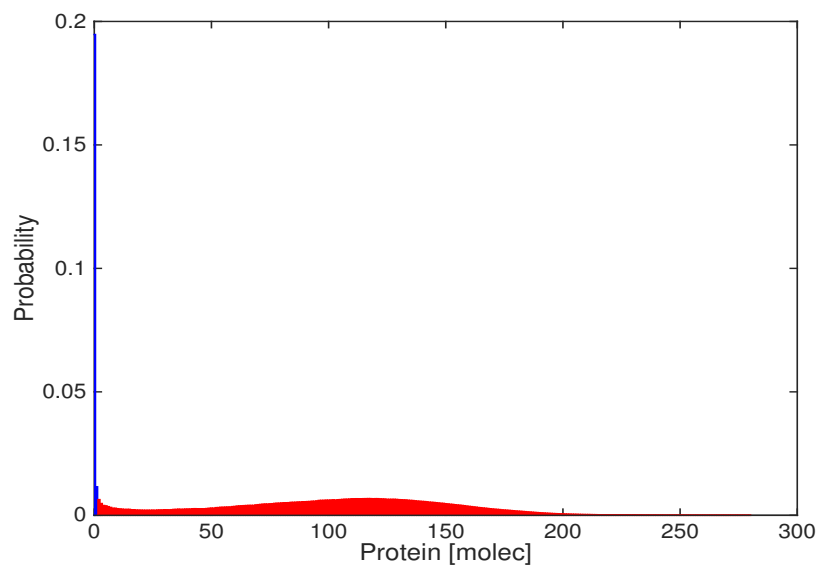


Fig. 7: Steady distribution of protein for the gene auto regulation model.

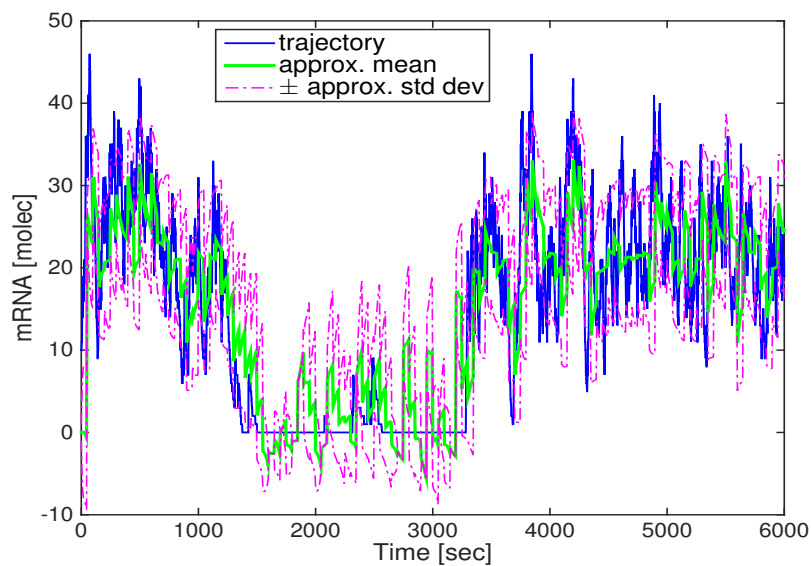


Fig. 8: Approximations to filtering of mRNA for the gene auto regulation model.

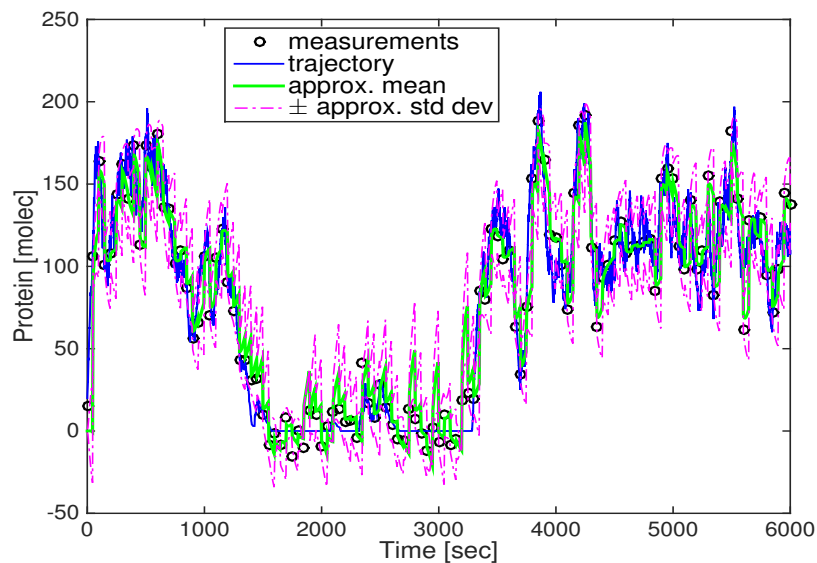


Fig. 9: Approximations to filtering of protein for the gene auto regulation model.

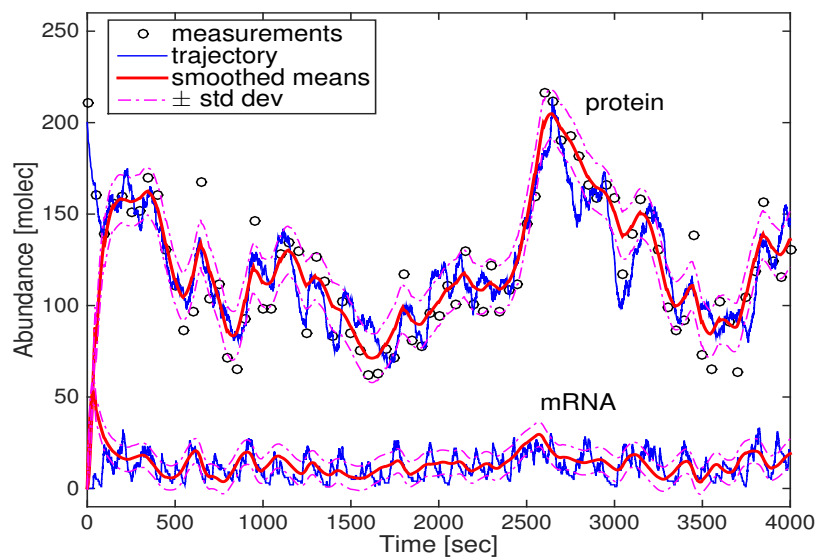


Fig. 10: Normal-based RTS approximate to smoothing for the gene expression model.

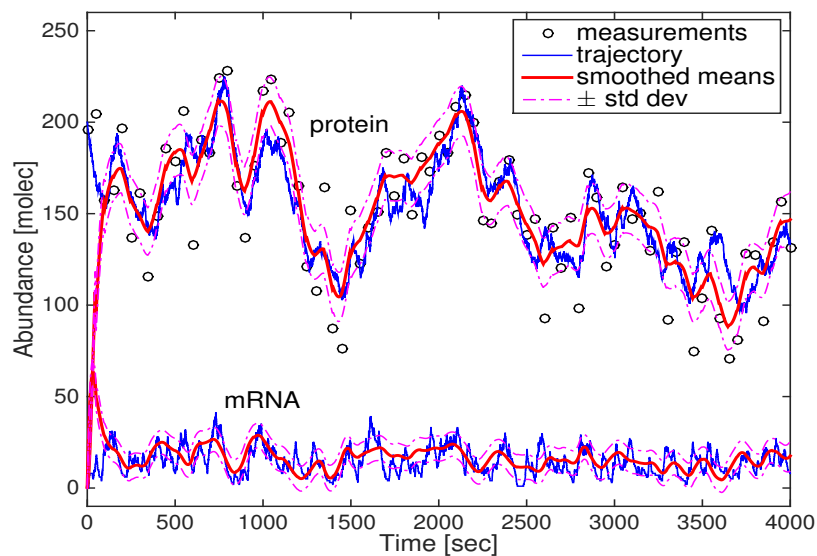


Fig. 11: Modified Normal-based RTS approximate to smoothing for the gene expression model.

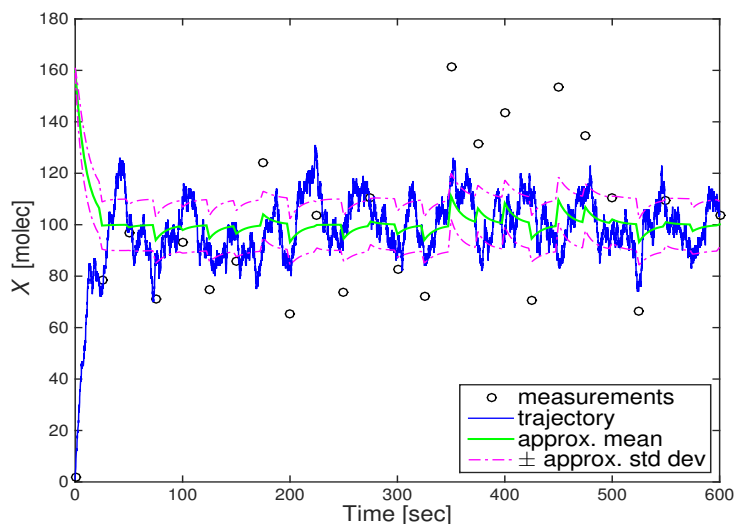


Fig. 12: Approximates to filtering for the birth-death model with lognormal measurements.

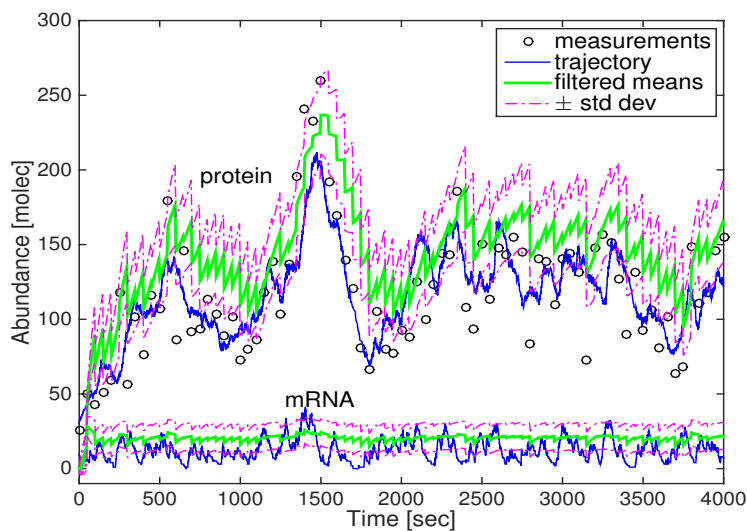


Fig. 13: Approximates to filtering for the gene expression model with lognormal measurements.

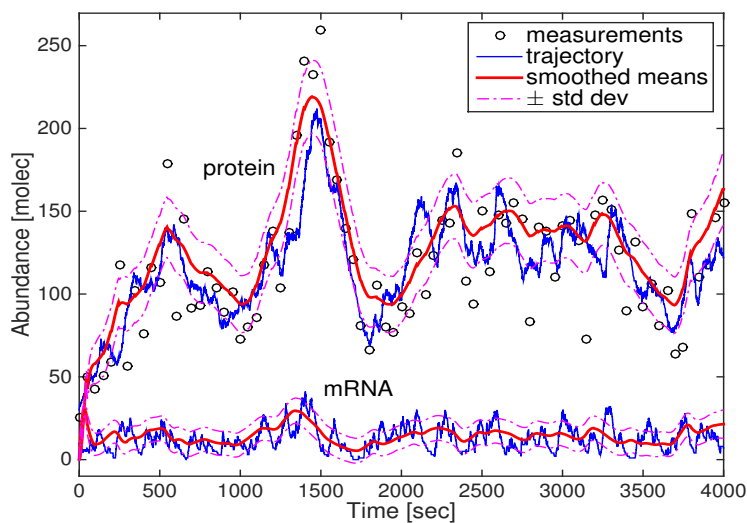


Fig. 14: RTS approximate to smoothing for the gene expression model with lognormal measurements.