Bayesian Filtering in Spiking Neural Networks: Noise, Adaptation, and Multisensory Integration
Online Appendix

Omer Bobrowski, Ron Meir and Yonina C. Eldar
Department of Electrical Engineering
Technion, Israel

January 2008

As mentioned in the paper, the full derivation of the filtering equation (2.4) is presented in [1]. The derivation in [1] is more general than the context of the paper, and uses very sophisticated mathematical tools. In this appendix we present a simplified outline of this derivation. We are aware that in our particular case of interest, the same results can be derived from discrete-time approximate. However, we believe that rigorous continuous-time treatment is always preferred, to avoid approximation errors. In addition we hope that this appendix will help making the profound literature of point-process filtering more accessible to the neuroscience community.

A Change of Measure

The concept known as *change of measure* is a mathematical technique borrowed from the theory of measure and probability. Since this method is at the heart of all the derivations presented throughout this appendix, we start by giving a brief and superficial introduction to this concept. We refer the reader to [1] for a more rigorous treatment.

A.1 The main concept

Let X, Y be two random variables, having a joint probability density function (pdf) $f_{X,Y}(x,y)$. Our goal is to estimate the value of some arbitrary function $h(X)$ given the observation Y. It is well known in the theory of estimation, that the minimum mean-squared error (MSE) estimator is the conditional expectation

$$\hat{X}_{opt} = \mathbb{E}\left[h(X)\,|Y=y\right]. \tag{A.1}$$

There are cases, however, in which the statistical relationship between X and Y (dictated by $f_{X,Y}(x,y)$) is complex, and calculation of this optimal estimator becomes too difficult.

Sometimes, the evaluation of (A.1) can be simplified by treating X and Y as having an alternative pdf $g_{X,Y}(x,y)$. This new pdf introduces simpler statistics for X, Y (most commonly X and Y will be independent). The possible values of X and Y need to be consistent between the two distributions, and therefore we require that,

$$\forall x,y:\ f_{X,Y}(x,y) > 0 \Longrightarrow g_{X,Y}(x,y) > 0. \tag{A.2}$$

This means that every possible observation of X and Y can be interpreted by $g_{X,Y}$. We further denote by $\mathbb{P}_g, \mathbb{E}_g$ the probability and expectation with respect to the alternative distribution $g_{X,Y}$.

The following theorem states that (A.1) can be evaluated using the simpler pdf $g_{X,Y}$.

Theorem A.1. *Let X, Y be two random variables with the joint pdf $f_{X,Y}$, and let $g_{X,Y}$ be an alternative pdf for those variables, satisfying* (A.2)*. Then for every $y \in \mathbb{R}$ such that $f_Y(y) \neq 0$,*

$$\mathbb{E}\left[h(X)\,|Y=y\right] = \frac{\mathbb{E}_g\left[h(X)L\,|Y=y\right]}{\mathbb{E}_g\left[L\,|Y=y\right]}, \tag{A.3}$$

where

$$L = \begin{cases} \frac{f_{X,Y}(X,Y)}{g_{X,Y}(X,Y)} & \text{if } g_{X,Y}(X,Y) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Note that L, is a random variable, since it is a function of X and Y.

Proof. Evaluating the numerator on the right-hand side of (A.3) yields

$$\begin{aligned}\mathbb{E}_g\left[h(X)L\,|Y=y\right] &= \int h(x)L(x,y)g_{X|Y}(x|y)dx \\ &= \int h(x)L(x,y)\frac{g_{X,Y}(x,y)}{g_Y(y)}dx \\ &= \frac{1}{g_Y(y)}\int h(x)f_{X,Y}(x,y)dx.\end{aligned}$$

Note that $g_Y(y) \neq 0$ according to (A.2) since we assume that $f_Y(y) \neq 0$. The denominator of (A.3) is equal to

$$\mathbb{E}_g\left[L\,|Y=y\right] = \int L(x,y)g_{X|Y}(x|y)dx = \frac{1}{g_Y(y)}\int f_{X,Y}(x,y)dx = \frac{f_Y(y)}{g_Y(y)}$$

Therefore, we can write the right-hand side of (A.3) as

$$\begin{aligned}\frac{\mathbb{E}_g\left[h(X)L\,|Y=y\right]}{\mathbb{E}_g\left[L\,|Y=y\right]} &= \frac{1}{f_Y(y)}\int h(x)f_{X,Y}(x,y)dx \\ &= \int h(x)f_{X|Y}(x|y)dx = \mathbb{E}\left[h(X)\,|Y=y\right],\end{aligned}$$

which completes the proof. □

The main contribution of this theorem is that it establishes a way to convert expectation calculations under the more complicated probability law f into calculations under a simpler law g. This method is referred to in the mathematical literature as *change of measure*. The special random variable L is used as a 'translator' between the two probability laws and is named the *likelihood-ratio* or the *Radon-Nikodym derivative*. Note that here we assumed X and Y to be continuous random variables; however, similar results can be derived in a more general setting, e.g. any number of random variables (continuous or discrete), random vectors, random processes, etc. The only requirement is to take L to be the appropriate likelihood-ratio of the two probability laws.

A.1.1 Non-normalized probabilities

We now show how the change of measure method can be utilized to simplify posterior probability calculations.

Assume that we have a hidden state X that can take on a discrete set of values $\{s_i\}_{i\in\mathbb{Z}}$. Given a partial noisy observation Y of X, our goal is to determine the conditional probability $\mathrm{p}_i = \mathbb{P}(X = s_i|Y = y)$. Using the indicator function $\mathbb{1}_i(x) = 1\{x = s_i\}$, we can write

$$\mathrm{p}_i = \mathbb{E}_g\left[\mathbb{1}_i(X)\,|Y = y\right].$$

Applying Theorem A.1 yields,

$$\mathrm{p}_i = \frac{\mathbb{E}_g\left[\mathbb{1}_i(X)L\,|Y = y\right]}{\mathbb{E}_g\left[L\,|Y = y\right]}.$$

Since the denominator does not depend on the value of i, it can be regarded as a normalization factor. Consequently, in order to calculate the value of p_i we can settle for the calculation of the values

$$\rho_i = \mathbb{E}_g\left[\mathbb{1}_i(X)L\,|Y = y\right],$$

which we call the *non-normalized probabilities*. The value of p_i can then be retrieved by normalization

$$\mathrm{p}_i = \frac{\rho_i}{\sum_j \rho_j}.$$

The fact that we only need to calculate the values of non-normalized probabilities ρ_i will substantially simplify the results in the sequel.

B Filtering a Markov Process from Poisson Measurements

Consider a dynamic process X_t representing the state of the world (e.g., the location of an object, its shape, orientation, velocity, etc.). We assume that X_t is a continuous time finite state Markov process, with a finite state-space $\mathcal{S} = \{s_1, \ldots, s_N\}$. The state X_t is not directly observed, but is processed through a set of M sensory cells, each of which produces a spike train, associated with a counting process $N_t^{(m)}$. At this point we take the spikes to be generated by a Poisson process, where the process rate depends on the current environmental state. We denote the rate of the process generated by the m-th

cell by $\lambda_m(X_t)$, where $\lambda_m(\cdot)$ represents the *tuning curve* of the m-th cell. The firing events of the different sensory cells are assumed to be independent given the state. Our goal is to compute the posterior probabilities

$$\mathrm{p}_i(t) = \mathbb{P}\left(X_t = s_i \,\middle|\, N^{(1)}_{[0,t]}, \ldots, N^{(M)}_{[0,t]}\right),$$

where $N^{(m)}_{[0,t]} = \left\{N^{(m)}_s\right\}_{s=0}^{t}$ is the full history of the process N_t. More specifically, we are looking for an online computation method that can be carried out by a neural network.

B.1 Preliminaries

The main assumptions of the model under consideration are the following:

1. **World state** - A continuous-time finite-state Markov process;
2. **Sensory activity** - Doubly-stochastic Poisson processes.

In this section we briefly introduce the main properties of these processes, and refer the reader to [3, 2] for a more comprehensive overview.

B.1.1 Continuous-time finite-state Markov process

A continuous-time finite-state Markov process (CFMP) is any process $X_t \in \{s_1, \ldots, s_N\}$, that satisfies the Markov property:

$$\mathbb{P}\left(X_t = s_j \,|X_{t_1} = s_{i_1}, \ldots, X_{t_n} = s_{i_n}, t_1 < \cdots < t_n < t\right) = \mathbb{P}\left(X_t = s_j \,|X_{t_n} = s_{i_n}\right).$$

This implies that given the process' path in the past, the distribution of the future states depends only on the most recent measurement. Thus, the evolution of the process statistics is determined by the elements in the process' *transition matrix* defined as $\mathbf{P}^{(\tau)}_{ij} = \mathbb{P}\left(X_{t+\tau} = s_j \,|X_t = s_i\right)$, where it is commonly assumed (e.g. [2]) that

$$\mathbf{P}^{(\tau)}_{ij} = \begin{cases} q_{ij}\tau + o(\tau) & i \neq j \\ 1 + q_{ii}\tau + o(\tau) & i = j, \end{cases} \tag{B.1}$$

with $q_{ii} = -\sum_{j \neq i} q_{ij} < 0$. The representation (B.1) implies that as $\tau \to 0$, it becomes less likely to have a transition between states. Clearly

$$q_{ij} = \left. \frac{d\mathbf{P}_{ij}^{(\tau)}}{d\tau} \right|_{\tau=0} . \tag{B.2}$$

Defining,

$$\mathrm{Q} = \{q_{ij}\} \quad ; \quad \mathbf{P}^{(\tau)} = \left\{p_{ij}^{(\tau)}\right\} \quad ; \quad \boldsymbol{\pi}(t) = (\Pr(X_t = s_1), \ldots, \Pr(X_t = s_N))^\top \tag{B.3}$$

it can be shown (e.g. [2]) that

$$\begin{aligned} \boldsymbol{\pi}(t+\tau) &= \boldsymbol{\pi}(t)\mathbf{P}^{(\tau)}, & \text{(B.4)} \\ \dot{\boldsymbol{\pi}}(t) &= \boldsymbol{\pi}(t)\mathrm{Q}, & \text{(B.5)} \\ \frac{d\mathbf{P}^{(\tau)}}{d\tau} &= \mathbf{P}^{(\tau)}\mathrm{Q}. & \text{(B.6)} \end{aligned}$$

The matrix $\mathbf{P}^{(\tau)}$ is called the *transition matrix* of the process, and Q is the *infinitesimal generator matrix*. Evidently, Q indicates the change in the transition probabilities in an infinitesimal-length interval. Note that the row-sums of $\mathbf{P}^{(\tau)}$ and Q satisfy

$$\sum_{j=1}^{N} q_{ij} = 0 \qquad ; \qquad \sum_{j=1}^{N} \mathbf{P}_{ij}^{(\tau)} = 1.$$

B.1.2 Doubly-stochastic Poisson process

Let X_t be an arbitrary continuous-time random process. A counting process $N_t,\ t \geq 0$, is called a *doubly-stochastic Poisson process* (DSPP) with rate function $\lambda(t, X_t)$ [4] if, given any realization of X_t,

1. $N_0 = 0$
2. $N_t \sim \text{Poisson}\left(\int_0^t \lambda(s, X_s) ds\right)$
3. For every $s < t$, N_s and $N_t - N_s$ are independent (independent increments).

In other words, *given* $\{X_t, t \geq 0\}$, N_t is an inhomogeneous Poisson process, with rate function $\lambda(t, X_t)$.

B.2 Filtering Equations

In this section we present the full derivation of the main filtering equation (2.4) based on the concepts introduced in the previous section. We start by considering the case of a single observation process (sensory spike-train), and then extend the result to the case of multiple observations.

Consider a CFMP $X_t \in \{s_1, \ldots, s_N\}$ observed via a DSPP N_t, characterized by a random rate function $\lambda(X_t)$. We assume that the rate function has no explicit time dependence (i.e.,$\lambda(t, X_t) \triangleq \lambda(X_t)$). The joint probability law of X_t and N_t will be denoted by f. Our goal is to derive recursive equations calculating the posterior probabilities $\mathrm{p}_i(t) = \mathbb{P}\left(X_t = s_i \left| N_{[0,t]} \right.\right)$, where $N_{[0,t]} = \{N_s; 0 \leq s \leq t\}$ is the full history of the process N_t. Using the results from the previous section, we can settle for calculating the values

$$\rho_i(t) = \mathbb{E}_g\left[\mathbb{1}_i(X_t) L_t \left| N_{[0,t]} \right.\right],$$

where $\mathbb{E}_g$ is the mean calculated assuming a different probability law g (which will be defined later), and L_t is the likelihood-ratio needed to translate the instances of X_t, N_t in $[0,t]$ from f to g.

The joint likelihood of X_t and N_t realizations in the interval $[0,t]$ is (see [3])

$$\begin{aligned} f_{X,N}\left(X_{[0,t]}, N_{[0,t]}\right) &= f_X\left(X_{[0,t]}\right) f_{N|X}\left(N_{[0,t]} | X_{[0,t]}\right) \\ &= f_X\left(X_{[0,t]}\right) \left(\prod_{t_n \leq t} \lambda\left(X_{t_n}\right)\right) \exp\left(-\int_0^t \lambda\left(X_s\right) ds\right), \end{aligned} \tag{B.7}$$

where $\{t_n\}_{n\in\mathbb{N}}$ are N_t's event arrival times. In order to simplify the calculation of $\rho_i(t)$ as much as possible, the alternative probability law g is chosen so that X_t's statistics remains the same, while N_t is assumed to be a homogeneous Poisson process, with a constant rate of 1, *independent* of X_t. In this alternative probability law the likelihood function is given by

$$g_{X,N}\left(N_{[0,t]}, X_{[0,t]}\right) = g_X\left(X_{[0,t]}\right) g_N\left(N_{[0,t]}\right) = g_X\left(X_{[0,t]}\right) e^{-t} = f_X\left(X_{[0,t]}\right) e^{-t}.$$

Therefore, the likelihood-ratio in the interval $[0,t]$ is

$$L_t = \frac{f_{X,N}\left(X_{[0,t]}, N_{[0,t]}\right)}{g_{X,N}\left(X_{[0,t]}, N_{[0,t]}\right)} = \left(\prod_{t_n \leq t} \lambda\left(X_{t_n}\right)\right) \exp\left(\int_0^t \left(1 - \lambda\left(X_s\right)\right) ds\right). \tag{B.8}$$

As mentioned previously, our goal is to derive a set of recursive equations for computing $\rho_i(t) = \mathbb{E}_g\left[\mathbb{1}_i(X_t)L_t \left| N_{[0,t]} \right.\right], i = 1, \ldots, N$. In order to do so, we first derive a recursive equation for L_t.

B.2.1 A Recursive equation for L_t

To compute L_t recursively, we begin by examining L_t's behavior in the interval $[t_n, t_{n+1})$. In this interval the product on the right-hand side of (B.8) remains constant (since no new events arrive). Therefore, we can write

$$L_t = c \exp\left(\int_0^t \left(1 - \lambda\left(X_s\right)\right) ds\right), \quad t \in [t_n, t_{n+1}). \tag{B.9}$$

Differentiating (B.9) yields

$$\frac{dL_t}{dt} = c\left(1 - \lambda\left(X_t\right)\right) \exp\left(\int_0^t \left(1 - \lambda\left(X_s\right)\right) ds\right) = \left(1 - \lambda\left(X_t\right)\right) L_t,$$

which can be written in integral form as

$$L_t = L_{t_n} + \int_{t_n}^t L_s \left(1 - \lambda\left(X_s\right)\right) ds, \quad t \in [t_n, t_{n+1}). \tag{B.10}$$

Next, suppose a new event arrives at time t_{n+1}. The effect on L_t is that a new element $\lambda\left(X_{t_{n+1}}\right)$, is added to the product on the right-hand side of (B.8), so that

$$L_{t_{n+1}} = \lambda\left(X_{t_{n+1}}\right) L_{t_{n+1}^-},$$

which can be rewritten as

$$L_{t_{n+1}} = L_{t_{n+1}^-} + \left(\lambda\left(X_{t_{n+1}}\right) - 1\right) L_{t_{n+1}^-}. \tag{B.11}$$

Combining (B.10) and (B.11) and using the fact that $L_0 = 1$ yields

$$L_t = 1 + \int_0^t L_s \left(1 - \lambda\left(X_s\right)\right) ds + \sum_{t_n \le t} \left(\lambda\left(X_{t_n}\right) - 1\right) L_{t_n^-}. \tag{B.12}$$

By introducing the notation

$$\int_0^t h(s)dN_s = \sum_{t_n \leq t} h(t_n),$$

we can write (B.12) in the form of a stochastic integral equation as

$$L_t = 1 + \int_0^t L_{s^-} \left(\lambda\left(X_s\right) - 1\right)\left(dN_s - ds\right).$$

B.2.2 A Recursive equation for $\rho_i(t)$

Applying L_t's integral equation to the expression defining $\rho_i(t)$ yields

$$\begin{aligned}\rho_i(t) &= \mathbb{E}_g\left[\mathbb{1}_i\left(X_t\right) L_t \left| N_{[0,t]}\right.\right] \\ &= \mathbb{E}_g\left[\mathbb{1}_i\left(X_t\right) \left| N_{[0,t]}\right.\right] + \int_0^t \mathbb{E}_g\left[\mathbb{1}_i\left(X_t\right) L_{s^-}\left(\lambda\left(X_s\right) - 1\right) \left| N_{[0,t]}\right.\right]\left(dN_s - ds\right)\end{aligned} \tag{B.13}$$

Since the probability law g is chosen so that X_t and N_t are independent, we can write

$$\mathbb{E}_g\left[\mathbb{1}_i\left(X_t\right) \left| N_{[0,t]}\right.\right] = \mathbb{E}_g\left[\mathbb{1}_i\left(X_t\right)\right] = \mathbb{P}_g\left(X_t = s_i\right) = \pi_i(t), \tag{B.14}$$

which is X_t's prior distribution. The rest of this section will be dedicated to simplifying the integral term appearing in (B.13).

By applying the law of total expectation (smoothing theorem) we get

$$\begin{aligned}&\mathbb{E}_g\left[\mathbb{1}_i(X_t) L_{s-}(\lambda(X_s) - 1) \left| N_{[0,t]}\right.\right] \\ &= \mathbb{E}_g\left[\mathbb{E}_g\left[\mathbb{1}_i(X_t) L_{s^-}(\lambda(X_s) - 1) \left| N_{[0,t]}, L_{s^-}, \lambda(X_s)\right.\right] \left| N_{[0,t]}\right.\right] \\ &= \mathbb{E}_g\left[L_{s^-}(\lambda(X_s) - 1)\mathbb{E}_g\left[\mathbb{1}_i(X_t) \left| N_{[0,t]}, L_{s^-}, \lambda(X_s)\right.\right] \left| N_{[0,t]}\right.\right].\end{aligned} \tag{B.15}$$

The interior expectation can be written as

$$\begin{aligned}&\mathbb{E}_g\left[\mathbb{1}_i(X_t) \left| N_{[0,t]}, L_{s^-}, \lambda(X_s)\right.\right] = \mathbb{P}_g\left(X_t = s_i \left| N_{[0,t]}, L_{s^-}, \lambda(X_s)\right.\right) \\ &= \sum_j \mathbb{P}_g\left(X_t = s_i \left| X_s = s_j, N_{[0,t]}, L_{s^-}, \lambda(X_s)\right.\right) \mathbb{P}_g\left(X_s = s_j \left| N_{[0,t]}, L_{s^-}, \lambda(X_s)\right.\right) \\ &\overset{(*)}{=} \sum_j \mathbf{P}_{ji}^{(t-s)} \mathbb{E}_g\left[\mathbb{1}_j(X_s) \left| N_{[0,t]}, L_{s^-}, \lambda(X_s)\right.\right],\end{aligned} \tag{B.16}$$

where $\mathbf{P}^{(\tau)}$ is the transition matrix of the process X_t (defined in (B.3)). Equality $(*)$ is derived from the fact that under the probability law g, the processes X_t and N_t are independent. Combining (B.15),(B.16) yields

$$\begin{aligned}
&\mathbb{E}_g\left[\mathbb{1}_i(X_t)L_{s^-}(\lambda(X_s)-1)\left|N_{[0,t]}\right.\right] \\
&= \mathbb{E}_g\left[L_{s^-}(\lambda(X_s)-1)\sum_j \mathbf{P}_{ji}^{(t-s)}\mathbb{E}_g\left[\mathbb{1}_j(X_s)\left|N_{[0,t]}, L_{s^-}, \lambda(X_s)\right.\right]\left|N_{[0,t]}\right.\right] \\
&= \sum_j \mathbf{P}_{ji}^{(t-s)}\mathbb{E}_g\left[\mathbb{1}_j(X_s)L_{s^-}(\lambda(X_s)-1)\left|N_{[0,t]}\right.\right] \\
&\overset{(\sharp)}{=} \sum_j \mathbf{P}_{ji}^{(t-s)}(\lambda(s_j)-1)\mathbb{E}_g\left[\mathbb{1}_j(X_s)L_{s-}\left|N_{[0,s]}\right.\right] \\
&= \sum_j \mathbf{P}_{ji}^{(t-s)}(\lambda(s_j)-1)\rho_j(s^-),
\end{aligned} \tag{B.17}$$

where $(\sharp)$ is based on the fact that if $X_s \neq s_j$ then the expression inside the conditional expectation is equal to zero. Substituting (B.14) and (B.17) into (B.13) yields the stochastic integral equation

$$\rho_i(t) = \pi_i(t) + \sum_j (\lambda(s_j)-1)\int_0^t \mathbf{P}_{ji}^{(t-s)}\rho_j\left(s^-\right)(dN_s - ds). \tag{B.18}$$

In order to further simplify (B.18), we examine its behavior in two different cases.

Between two consecutive events of N_t:

In this case $dN_s = 0$ and therefore $\rho_i(t)$ follows the differential equation

$$\dot{\rho}_i(t) = \frac{d}{dt}\pi_i(t) + \sum_j (1-\lambda(s_j))\int_0^t \frac{d}{dt}\mathbf{P}_{ji}^{(t-s)}\rho_j\left(s^-\right)ds + \sum_j (1-\lambda(s_j))\,\mathbf{P}_{ji}^{(0)}\rho_j\left(t^-\right) \tag{B.19}$$

Using the differential equations (B.5),(B.6), we get

$$\begin{aligned}
\frac{d}{dt}\pi_i(t) &= \sum_k q_{ki}\pi_k(t) \\
\frac{d}{dt}\mathbf{P}_{ji}^{(t-s)} &= \sum_k q_{ki}\mathbf{P}_{jk}^{(t-s)}.
\end{aligned} \tag{B.20}$$

Subtituting (B.20), along with the fact that $\mathbf{P}_{ji}^{(0)} = \delta_{ji}$ (no transition is possible within

a zero-length interval) into (B.19) yields

$$\dot{\rho}_i(t) = \sum_k q_{ki}\rho_k(t) + (1 - \lambda(s_i))\,\rho_i(t). \tag{B.21}$$

At an event of N_t:

When a new event arrives, N_t is increased by one, causing an infinitesimal-time change to the stochastic integral in (B.18),

$$\rho_i(t) - \rho_i\left(t^-\right) = \sum_j (\lambda(s_j) - 1)\,\mathbf{P}^{(0)}_{ji}\rho_j\left(t^-\right) = (\lambda(s_i) - 1)\,\rho_i\left(t^-\right). \tag{B.22}$$

Using (B.21) and (B.22) leads to the following representation of (B.18):

$$\rho_i(t) = \rho_i(0) + \int_0^t \sum_k q_{ki}\rho_k(s)ds + (\lambda(s_i) - 1)\int_0^t \rho_i\left(s^-\right)(dN_s - ds),$$

Though not mathematically rigorous, we can differentiate this equation with regard to t, and use the fact that $dN_t/dt = \sum_n \delta(t - t_n)$ (where $\delta(t)$ is the Dirac delta function) to get the final result

$$\begin{aligned}\dot{\rho}_i(t) &= \sum_k q_{ki}\rho_k(t) + \left[(\lambda(s_i) - 1)\left(\sum_n \delta(t - t_n) - 1\right)\right]\rho_i(t) \\ &= \sum_k q_{ki}\rho_k(t) + (\lambda(s_i) - 1)(\nu(t) - 1)\rho_i(t),\end{aligned} \tag{B.23}$$

where $\nu(t) = \sum_n \delta(t - t_n)$ represents the input activity.

B.2.3 Multiple observations

The results of the previous section can be easily extended to the case where we are observing the process X_t via M Poisson processes $\left\{N_t^{(1)}, \ldots, N_t^{(M)}\right\}$, each with a different rate function $\{\lambda_1(X_t), \ldots, \lambda_M(X_t)\}$. The analog of equation (B.23) in this case is (see [1])

$$\dot{\rho}_i(t) = \sum_k q_{ki}\rho_k(t) + \left(\sum_{m=1}^M (\lambda_m(s_i) - 1)\,\nu_m(t)\right)\rho_i(t) - \left(\sum_{m=1}^M (\lambda_m(s_i) - 1)\right)\rho_i(t),$$

where $\nu_m(t) = \sum_n \delta\left(t - t_n^{(m)}\right)$, and $t_n^{(m)}$ is the arrival time of the n-th event from the m-th input process. Note that the constant subtraction in the last term serves merely

as an exponential factor that affects all states the same way, thus we can omit it from the equation and get

$$\dot{\rho}_i(t) = \sum_k q_{ki}\rho_k(t) + \left(\sum_{m=1}^{M} (\lambda_m(s_i) - 1)\,\nu_m(t)\right)\rho_i(t) - \lambda(s_i)\rho_i(t), \tag{B.24}$$

where $\lambda(s_i) = \sum_{m=1}^{M} \lambda_m(s_i)$.

We can present (B.24) in vector form, by defining

$$\begin{aligned} \boldsymbol{\rho}(t) &= (\rho_1(t), \ldots, \rho_N(t))^\top \\ \Lambda_m &= \mathrm{diag}\,(\lambda_m(s_1), \ldots, \lambda_m(s_N)) \\ \Lambda &= \sum_{m=1}^{M} \Lambda_m \end{aligned}$$

Using these definition, (B.24) becomes

$$\dot{\boldsymbol{\rho}}(t) = \mathrm{Q}^\top \boldsymbol{\rho}(t) + \left(\sum_{m=1}^{M} (\Lambda_m - \mathbf{I})\,\nu_m(t)\right)\boldsymbol{\rho}(t) - \Lambda\boldsymbol{\rho}(t), \tag{B.25}$$

where $\mathbf{I}$ is the identity matrix.

References

[1] P. Brémaud. *Point Processes and Queues: Martingale Dynamics*. Springer, New York, 1981.

[2] G.R. Grimmett and D.R. Stirzaker. *Probability and Random Processes*. Oxford University Press, third edition, 2001.

[3] D.L. Snyder. Filtering and detection for doubly stochastic Poisson processes. *IEEE Transactions on Information Theory*, IT-18:91–102, 1972.

[4] D.L. Snyder and M.I. Miller. *Random Point Processes in Time and Space*. Springer, New York, second edition, 1991.