We may obtain a more accurate expression for $\{\alpha_k\}$ as follows, as long as $|k|$ is not too large. The Taylor expansion of sinc(t) about the origin is

$$\text{sinc}(t) = \sum_{k=0}^{n-1} \frac{(-1)^k \pi^{2k}}{(2k+1)!} t^{2k} + \frac{\text{sinc}^{(n)}(\theta t)}{n!} t^n, \quad 0 < \theta < 1. \tag{4.9}$$

For a series expansion with an infinite number of terms, the radius of convergence includes all finite t. Thus

$$\begin{aligned}\alpha(t) = 2^{L/2} \sum_{k=0}^{n-1} \frac{(-1)^k \pi^{2k}}{(2k+1)!} \int_{-\infty}^{\infty} x^{2k} \phi(2^L x - t)\, dx \\ + 2^{L/2} \int_{-\infty}^{\infty} \frac{\text{sinc}^{(n)}(\theta x)}{n!} x^n \phi(2^L x - t)\, dx.\end{aligned} \tag{4.10}$$

The nth approximant to $\alpha(t)$ is the first term of (4.10) and is

$$\alpha_n(t) = 2^{L/2} \sum_{k=0}^{n-1} \frac{(-1)^k \pi^{2k}}{(2k+1)!} \int_{-\infty}^{\infty} x^{2k} \phi(2^L x - t)\, dx \tag{4.11}$$

for $n \geq 1$. It can therefore be readily shown that

$$\alpha_n(t) = 2^{-L/2} \sum_{k=0}^{n-1} \frac{(-1)^k}{(2k+1)!} \left(\frac{\pi}{2^L}\right)^{2k} \left\{ \sum_{j=0}^{2k} \binom{2k}{j} m_j t^{2k-j} \right\}. \tag{4.12}$$

Thus, $\alpha_k \approx \alpha_n(k)$, provided that n is big enough for the range of k of interest. In other words, the approximation is good if we use enough moments of $\phi(t)$. The error term in (4.10) is the second term and may be rewritten as

$$\begin{aligned}\epsilon_n(t) = \frac{1}{n!} h^{n+\frac{1}{2}} \sum_{r=0}^{n} \binom{n}{r} \\ \times \left[\int_{-\infty}^{\infty} \text{sinc}^{(n)}[h\theta(\tau + t)] \tau^r \phi(\tau)\, d\tau \right] t^{n-r}\end{aligned} \tag{4.13}$$

where $h = 2^{-L}$. We see that the useful range of t for which $\alpha_n(t)$ is a good approximation to $\alpha(t)$ is limited but that it increases when n and/or L increase.

References

[1] M. V. Wickerhauser, "Lectures on wavelet packet algorithms," Nov. 18, 1991; Available [Online] http://www.mathsoft.com/wavelets.html.

[2] C. Zarowski, "Notes on orthogonal wavelets and wavelet packets," Nov. 3, 1995: Available [Online] http://www.mathsoft.com/wavelets.html.

[3] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[4] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley, MA: Wellesley-Cambridge, 1996.

[5] X.-P. Zhang, L.-S. Tian, and Y.-N. Peng, "From the wavelet series to the discrete wavelet transform—The initialization," *IEEE Trans. Signal Processing*, vol. 44, pp. 129–133, Jan. 1996.

[6] X.-G. Xia, C.-C. J. Kuo, and Z. Zhang, "Wavelet coefficient computation with optimal prefiltering," *IEEE Trans. Signal Processing*, vol. 42, pp. 2191–2197, Aug. 1994.

[7] C. J. Zarowski, "An approach to initializing the wavelet packet transform," *IEEE Signal Processing Lett.*, vol. 4, pp. 132–134, May 1997.

[8] W. Sweldens and R. Piessens, "Quadrature formulae and asymptotic error expansions for wavelet approximations of smooth functions," *SIAM J. Numer. Anal.*, vol. 31, pp. 1240–1264, Aug. 1994.

Estimating Rate Constants in Hidden Markov Models by the EM Algorithm

Steffen Michalek and Jens Timmer

***Abstract*— The EM algorithm, e.g., the Baum–Welch re-estimation, is an important tool for parameter estimation in discrete-time hidden Markov models. We present a direct re-estimation of rate constants for applications in which the underlying Markov process is continuous in time. Previous estimation of discrete-time transition probabilities is not necessary.**

***Index Terms*—Continuous-time hidden Markov model, EM algorithm, maximum likelihood estimate, parameter estimation, parameterized hidden Markov model.**

I. Introduction

Hidden Markov models (HMM's) were successfully applied in various fields of time series analysis, e.g., in speech recognition [1] or ion channel analysis [2]–[4]. For discrete-time HMM's, the EM algorithm for maximum likelihood parameter estimation is well known [1], [5]. In some fields, however, the formulation as a discrete-time process does not appear to be completely adequate for the dynamics to be described. In the analysis of ion channel recordings, for example, the underlying biophysical process is the molecular dynamics of proteins in the cell membrane. Their behavior is described by a process of transitions between a few classes of conformations called states. Since this process is continuous in time, the transitions are more adequately described by transition rates then by discrete-time probabilities for state changes [6].

For many applications, a parameterization of the rate constants for the system's state transitions is necessary since certain relations between rate constants like identical values, distinct ratios, or functional dependence from common underlying variables are needed to express the dynamical behavior of a system, for example, to consider physical mechanisms and combinatorics [7], [8], agonist concentration [9], [10], multiple conductance sublevels [11], or for multichannel recordings [12]–[14]. Sometimes, it is necessary to set certain rates to zero, defining a special topology of the directed graph of transitions. In either case, it is hardly possible to formulate the recommended constraints in terms of discrete transition probabilities since they arise from the nonlinear matrix exponentiation operation [13].

Up to now, the EM algorithm as main tool in HMM estimation [1], [5] is not formulated to deal directly with the parameters of the underlying continuous-time process. This problem has been stated frequently [3], [11], [15].

We present equations to perform EM re-estimation iterations for such HMM's in which the Markov process is described in terms of rate constants and that are observed by discrete sampling. The approach to this formulation is a generalization of parameterized discrete-time hidden Markov models like that proposed in [14].

Manuscript received June 19, 1997; revised April 20, 1998. This work was supported by the Deutsche Forschungsgemeinschaft (DFG Ho 496/4-1). The associate editor coordinating the review of this paper and approving it for publication was Dr. Ali H. Sayed.

The authors are with the Center for Data Analysis and Modeling (FDM), University of Freiburg, Freiburg, Germany (e-mail: Jens.Timmer@fdm.uni-freiburg.de).

Publisher Item Identifier S 1053-587X(99)00153-1.

This correspondence is organized as follows. Section II will outline the EM algorithm for parameterized discrete-time hidden Markov models. In Section III, the equations for direct re-estimation of rate constants are derived. The Appendix gives further details on calculation as well as re-estimation formulae for other parameters in parameterized HMM's.

II. EM Re-estimation in Discrete-Time HMM's

Consider a parameterized discrete-time hidden Markov model $\mathcal{M}(\boldsymbol{\vartheta}) = (\mathbf{A}, \boldsymbol{\pi}, \mathbf{h})$ and the two related stochastic processes $\mathbf{X} = (X_t)_{1\le t\le N}$ and $\mathbf{Y} = (Y_t)_{1\le t\le N}$. The background process $\mathbf{X}$ of discrete-time transitions between the m background states is a homogeneous Markov process that is governed by the $m \times m$ stochastic matrix $\mathbf{A}$ of transition probabilities $a_{ij} = P(X_t = j \mid X_{t-1} = i)$. Its initial distribution is given by $\pi_i = P(X_1 = i)$. The real-valued output $\mathbf{Y}$ is determined by the conditional distributions $p_{\mathbf{h}}(Y_t = y_t \mid X_t = i)$[1]. The vector $\mathbf{h}$ may, for example, contain the means and variances of m normal distributions. Assume the model being parameterized by a parameter vector $\boldsymbol{\vartheta}$ containing disjoint subsets $\boldsymbol{\vartheta}^{\mathbf{A}}$, $\boldsymbol{\vartheta}^{\pi}$, and $\boldsymbol{\vartheta}^{\mathbf{h}}$, which parameterize $\mathbf{A}$, $\boldsymbol{\pi}$, and $\mathbf{h}$, respectively.

In order to obtain maximum likelihood estimates for $\boldsymbol{\vartheta}$ given data $\mathbf{y}$, the idea of the iterative EM algorithm is to maximize the expectation value of the log-likelihood with respect to the (new) parameter set $\boldsymbol{\vartheta}'$, given the background paths $\mathbf{x}$ distributed following the (old) parameters $\boldsymbol{\vartheta}$. The information inequality [16] ensures that the likelihood always increases with a step $\boldsymbol{\vartheta} \to \boldsymbol{\vartheta}'$ [17], [18]. The basic EM equations as well as the reestimation of the parameters for the output distributions and the initial distribution of parameterized HMM's are given in the Appendix.

For re-estimation of the parameters $\boldsymbol{\vartheta}^{\mathbf{A}}$ of the background dynamics, define

$$\Psi_{ij}(t, \mathbf{y}, \boldsymbol{\vartheta}) := P_{\boldsymbol{\vartheta}}(X_t = i, X_{t+1} = j \mid \mathbf{y}) \quad \text{for } 1 \le t < T$$

and

$$\Psi_{ij}^{\mathbf{T}}(\mathbf{y}, \boldsymbol{\vartheta}) := \sum_{t=1}^{T-1} \Psi_{ij}(t, \mathbf{y}, \boldsymbol{\vartheta})$$

which is computable efficiently using the so-called forward–backward algorithm [1], [5]. The normalization conditions $\sum_{j=1}^{m} a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}}) = 1$ for $1 \le i \le m$ are considered by Lagrange multipliers.

Then, with ϑ'_α a component of $\boldsymbol{\vartheta}^{\mathbf{A}'}$, the system of equations

$$\begin{aligned} 0 &= \sum_{i,j=1}^{m} \frac{\partial a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}'})}{\partial \vartheta'_\alpha} \left(\frac{\Psi_{ij}^{\mathbf{T}}(\mathbf{y}, \boldsymbol{\vartheta})}{a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}'})} + \lambda_i \right) \\ 0 &= \sum_{j=1}^{m} a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}'}) - 1 \end{aligned} \tag{1}$$

has to be solved for $\boldsymbol{\vartheta}^{\mathbf{A}'}$ with $1 \le \alpha \le \dim \boldsymbol{\vartheta}^{\mathbf{A}}$, $1 \le i \le m$.

In the special case of $\boldsymbol{\vartheta}^{\mathbf{A}}$ directly containing the transition probabilities a_{ij} (trivial parameterization), the original Baum–Welch re-estimation formulae [5] arise.

Without loss of generality, we assume a parameterization $a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}})$ of the nondiagonal elements a_{ij} for $i \ne j$, and obtain the diagonal elements from the normalization $a_{ii}(\boldsymbol{\vartheta}^{\mathbf{A}}) := 1 - \sum_{j\ne i} a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}})$ ensuring that $\mathbf{A}$ is a stochastic matrix.

[1] Small letters related to random variables denote realizations, i.e., elements of their image set.

With this convention and keeping in mind that $\frac{\partial a_{ii}(\boldsymbol{\vartheta}^{\mathbf{A}'})}{\partial \vartheta'_\alpha} = \sum_{j\ne i} (-\frac{\partial a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}'})}{\partial \vartheta'_\alpha})$, the number of equations for $\boldsymbol{\vartheta}^{\mathbf{A}'}$ reduces to $\dim \boldsymbol{\vartheta}^{\mathbf{A}}$

$$0 = \sum_{\substack{i,j=1 \\ i\ne j}}^{m} \frac{\partial a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}'})}{\partial \vartheta'_\alpha} \left(\frac{\Psi_{ij}^{\mathbf{T}}(\mathbf{y}, \boldsymbol{\vartheta})}{a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}'})} - \frac{\Psi_{ii}^{\mathbf{T}}(\mathbf{y}, \boldsymbol{\vartheta})}{a_{ii}(\boldsymbol{\vartheta}^{\mathbf{A}'})} \right). \tag{2}$$

In general, it is not possible to give an analytical solution for the re-estimated $\boldsymbol{\vartheta}^{\mathbf{A}'}$ since the $a_{ij}(\cdot)$ are almost arbitrary functions.

III. Re-estimation of Rate Constants

Up to now, we dealt with the discrete-time transition probabilities only. However, it is possible to describe the dynamics by rate constants in order to reflect that $\mathbf{X}$ is the sampled version of a continuous-time process with infinitesimal generator $\mathbf{Q}$. We consider the relation between the matrix $\mathbf{Q}$ of rate constants and the stochastic matrix $\mathbf{A}$

$$\mathbf{A}(\mathbf{Q}) = \exp(\mathbf{Q}\tau)$$

with $\mathbf{A} = \{a_{ij}\}$, $\mathbf{Q} = \{q_{ij}\}$, $f = \frac{1}{\tau}$ as the sampling rate. This relation is now interpreted as a special parameterization of the a_{ij}. Furthermore, if the rate matrix $\mathbf{Q}$ is parameterized itself, then

$$\mathbf{A}(\boldsymbol{\vartheta}^{\mathbf{A}}) = \{a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}})\} = \exp(\mathbf{Q}(\boldsymbol{\vartheta}^{\mathbf{A}})\tau)$$

whereas by analog convention, $\boldsymbol{\vartheta}^{\mathbf{A}}$ parameterizes the off-diagonal elements of $\mathbf{Q}$, ensuring that $\mathbf{Q}$ is a generator matrix by normalization $q_{ii}(\boldsymbol{\vartheta}^{\mathbf{A}}) := -\sum_{j\ne i} q_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}})$. This leads to

$$\begin{aligned} \frac{\partial a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}})}{\partial \vartheta'_\alpha} &= \sum_{r,s=1}^{m} \frac{\partial a_{ij}(\mathbf{Q}(\boldsymbol{\vartheta}^{\mathbf{A}}))}{\partial q_{rs}} \frac{\partial q_{rs}(\boldsymbol{\vartheta}^{\mathbf{A}})}{\partial \vartheta'_\alpha} \\ &= \sum_{\substack{r,s=1 \\ r\ne s}}^{m} \frac{\partial q_{rs}(\boldsymbol{\vartheta}^{\mathbf{A}})}{\partial \vartheta'_\alpha} \left(\frac{\partial a_{ij}(\mathbf{Q}(\boldsymbol{\vartheta}^{\mathbf{A}}))}{\partial q_{rs}} - \frac{\partial a_{ij}(\mathbf{Q}(\boldsymbol{\vartheta}^{\mathbf{A}}))}{\partial q_{rr}} \right). \end{aligned}$$

The resulting system of equations for $\boldsymbol{\vartheta}^{\mathbf{A}'}$ reads, with $1 \le \alpha \le \dim \boldsymbol{\vartheta}^{\mathbf{A}}$

$$\begin{aligned} 0 = \sum_{\substack{i,j=1 \\ i\ne j}}^{m} & \left(\sum_{\substack{r,s=1 \\ r\ne s}}^{m} \frac{\partial q_{rs}(\boldsymbol{\vartheta}^{\mathbf{A}})}{\partial \vartheta'_\alpha} \left(\underbrace{\frac{\partial a_{ij}(\mathbf{Q}(\boldsymbol{\vartheta}^{\mathbf{A}}))}{\partial q_{rs}} - \frac{\partial a_{ij}(\mathbf{Q}(\boldsymbol{\vartheta}^{\mathbf{A}}))}{\partial q_{rr}}}_{(*)} \right) \right) \\ & \cdot \left(\frac{\Psi_{ij}^{\mathbf{T}}(\mathbf{y}, \boldsymbol{\vartheta})}{a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}'})} - \frac{\Psi_{ii}^{\mathbf{T}}(\mathbf{y}, \boldsymbol{\vartheta})}{a_{ii}(\boldsymbol{\vartheta}^{\mathbf{A}'})} \right). \end{aligned} \tag{3}$$

Since (3) cannot be expected to be solved analytically for $\boldsymbol{\vartheta}^{\mathbf{A}'}$, simultaneous root finding must be applied numerically.

An efficient method for calculating the marked terms $(*)$ in (3) is described in [19]. Assuming the detailed balance property of the generator matrix $\mathbf{Q}$, they can be determined analytically [20]. Both methods are fast compared with the calculations necessary to obtain the $\Psi_{ij}^{\mathbf{T}}$.

Note that it is also possible within this framework to define constant entries of $\mathbf{Q}$. Thus, with $q_{ij} = 0$ for some transitions, different topologies for the directed graph of the permitted state transitions can be defined easily.

IV. Discussion

The parameterization of discrete-time transition probabilities by rate constants is an adequate approach to deal with the discrete sampled version of a process that is naturally described to be continuous in time. The explicit continuous-time formulation with continuous processes for both background and output [21] is not applicable in the present case since the usual methods of discretization of stochastic integrals cannot be carried out with the recursion equations [21] for *a priori* given sampling intervals.

Equations are given to generalize the EM algorithm for our model. The generalization does not affect the structure of the EM formalism. The resulting equations include the existing EM re-estimation as a special case in which they can be solved analytically. The same re-estimation method is applicable for the rate constants themselves as well as for arbitrarily parameterized generator matrices $\mathbf{Q}$. Analogously, the re-estimation equations for any modified or generalized discrete-time HMM, e.g., to deal with time correlated output [22], could be changed to consider rate constants.

The presented method performs direct re-estimation of parameterized rate constants without previous estimation of the $\mathbf{A}$ matrix. Thus, inference of the $\mathbf{Q}$ matrix from the discrete transition probabilities, which is known to be problematical [3], [11], [15], is not necessary.

Appendix
Re-Estimating the Other Parameters of a Parameterized Model

The EM iteration consists in finding the maximum of the following function with respect to $\boldsymbol{\vartheta}'$

$$\begin{aligned}\mathcal{Q}(\boldsymbol{\vartheta},\boldsymbol{\vartheta}') &:= \sum_{\mathbf{x}\in\mathcal{X}} \ln(P_{\boldsymbol{\vartheta}'}(\mathbf{x})p_{\boldsymbol{\vartheta}'}(\mathbf{y}\mid\mathbf{x}))P_{\boldsymbol{\vartheta}}(\mathbf{x}\mid\mathbf{y})\\ &= \sum_{i=1}^{m}\Psi_i(1,\mathbf{y},\boldsymbol{\vartheta})\ln\pi_i(\boldsymbol{\vartheta}^{\pi'})\\ &\quad+\sum_{i,j=1}^{m}\sum_{t=1}^{T-1}\Psi_{ij}(t,\mathbf{y},\boldsymbol{\vartheta})\ln a_{ij}(\boldsymbol{\vartheta}^{\mathbf{A}'})\\ &\quad+\sum_{i=1}^{m}\sum_{t=1}^{T}\Psi_i(t,\mathbf{y},\boldsymbol{\vartheta})\ln p_{\boldsymbol{\vartheta}\mathbf{h}'}(y_t\mid X_t=i)\end{aligned}$$

with $\mathcal{X}$ the set of all m^N possible background paths, and $\Psi_i(t,\mathbf{y},\boldsymbol{\vartheta}) = P_{\boldsymbol{\vartheta}}(X_t=i\mid\mathbf{y})$ for $1\le t\le T$, obtainable from the forward–backward algorithm.

A. Estimating Gaussian Output Parameters

Assume disjunct subsets $\boldsymbol{\vartheta}^{\mathbf{M}}$ and $\boldsymbol{\vartheta}^{\mathbf{S}}$ of $\boldsymbol{\vartheta}^{\mathbf{h}}$ parameterizing the means μ_i and standard deviations σ_i of Gaussian output for $1\le i\le m$.

Then, with ϑ'_β a component of $\boldsymbol{\vartheta}^{\mathbf{M}'}$, ϑ'_γ a component of $\boldsymbol{\vartheta}^{\mathbf{S}'}$, and with $\Psi_i^{\mathbf{T}}(\mathbf{y},\boldsymbol{\vartheta}) := \sum_{t=1}^{T}\Psi_i(t,\mathbf{y},\boldsymbol{\vartheta})$, $Y_i^{1\mathbf{T}}(\mathbf{y},\boldsymbol{\vartheta}) := \sum_{t=1}^{T} y_t\Psi_i(t,\mathbf{y},\boldsymbol{\vartheta})$, and $Y_i^{2\mathbf{T}}(\mathbf{y},\boldsymbol{\vartheta}) := \sum_{t=1}^{T} y_t^2\Psi_i(t,\mathbf{y},\boldsymbol{\vartheta})$, respectively, the re-estimation formulae are, with $1\le\beta\le\dim\boldsymbol{\vartheta}^{\mathbf{M}}$, $1\le\gamma\le\dim\boldsymbol{\vartheta}^{\mathbf{S}}$

$$0=\sum_{i=1}^{m}\frac{\partial\mu_i(\boldsymbol{\vartheta}^{\mathbf{M}'})}{\partial\vartheta'_\beta}\frac{1}{\sigma_i^2(\boldsymbol{\vartheta}^{\mathbf{S}'})}\left(Y_i^{1\mathbf{T}}(\mathbf{y},\boldsymbol{\vartheta})-\mu_i(\boldsymbol{\vartheta}^{\mathbf{M}'})\Psi_i^{\mathbf{T}}(\mathbf{y},\boldsymbol{\vartheta})\right)$$

$$\begin{aligned}0=\sum_{i=1}^{m}\frac{\partial\sigma_i(\boldsymbol{\vartheta}^{\mathbf{S}'})}{\partial\vartheta'_\gamma}\Bigg(&\frac{\Psi_i^{\mathbf{T}}(\mathbf{y},\boldsymbol{\vartheta})}{\sigma_i(\boldsymbol{\vartheta}^{\mathbf{S}'})}-\frac{1}{\sigma_i^3(\boldsymbol{\vartheta}^{\mathbf{S}'})}\big(\mu_i^2(\boldsymbol{\vartheta}^{\mathbf{M}'})\Psi_i^{\mathbf{T}}(\mathbf{y},\boldsymbol{\vartheta})\\ &+Y_i^{2\mathbf{T}}(\mathbf{y},\boldsymbol{\vartheta})-2\mu_i(\boldsymbol{\vartheta}^{\mathbf{M}'})Y_i^{1\mathbf{T}}(\mathbf{y},\boldsymbol{\vartheta})\big)\Bigg).\end{aligned}$$

B. Estimating the Parameters for the Initial State Distribution

Assume the initial probabilities parameterized by $\boldsymbol{\vartheta}^{\pi}$ except for one component, say, π_{i_0}, that is obtained by normalization to unity. Then, with ϑ'_δ a component of $\boldsymbol{\vartheta}^{\pi'}$, the re-estimation formulae are, with $1\le\delta\le\dim\boldsymbol{\vartheta}^{\pi}$

$$0=\sum_{\substack{i=1\\ i\ne i_0}}^{m}\frac{\partial\pi_i(\boldsymbol{\vartheta}^{\pi'})}{\partial\vartheta'_\delta}\left(\frac{\Psi_i(1,\mathbf{y},\boldsymbol{\vartheta})}{\pi_i(\boldsymbol{\vartheta}^{\pi'})}-\frac{\Psi_{i_0}(1,\mathbf{y},\boldsymbol{\vartheta})}{\pi_{i_0}(\boldsymbol{\vartheta}^{\pi'})}\right).$$

References

[1] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 257–285, 1989.
[2] S.-H. Chung, J. B. Moore, L. Xia, L. S. Premkumar, and P. W. Gage, "Characterization of single channel currents using digital signal processing techniques based on hidden Markov models," *Phil. Trans. R. Soc. Lond. B*, vol. 329, pp. 265–285, 1990.
[3] D. R. Fredkin and J. A. Rice, "Maximum likelihood estimation and identification directly from single-channel recordings," in *Proc. R. Soc. Lond. B.*, 1992, vol. 249, pp. 125–132.
[4] J. D. Becker, J. Honerkamp, J. Hirsch, E. Schlatter, and R. Greger, "Analyzing ion channels with hidden Markov models," *Pflügers Arch.*, vol. 426, pp. 328–332, 1994.
[5] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, vol. 41, pp. 164–171, 1970.
[6] B. Sakmann and E. Neher, Eds., *Single-Channel Recording*, 2nd ed. New York: Plenum, 1995.
[7] R. Horn and C. A. Vandenberg, "Statistical properties of single sodium channels," *J. Gen. Physiol.*, vol. 84, pp. 505–534, Oct. 1984.
[8] B. Hille, *Ionic Channels of Excitable Membranes*, 2nd ed. Sunderland, MA: Sinauer, 1992.
[9] D. Colquhoun and A. G. Hawkes, "On the stochastic properties of single ion channels," in *Proc. R. Soc. Lond. B*, 1981, vol. 211, pp. 205–235.
[10] D. R. Fredkin, M. Montal, and J. A. Rice, "Identification of aggregated Markovian models: Application to the nicotinic acetylcholine receptor," in *Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer*, L. M. Le Cam and R. A. Olshen, Eds., 1985, vol. I, pp. 269–289.
[11] S.-H. Chung and R. A. Kennedy, "Coupled Markov chain model: Characterization of membrane channel currents with multiple conductance sublevels as partially coupled elementary pores," *Math. Biosci.*, vol. 133, pp. 111–137, 1996.
[12] S. Draber, R. Schultze, and U.P. Hansen, "Cooperative behavior of K^+ channels in the tonoplast of *Chara corallina*," *Biophys. J.*, vol. 65, pp. 1553–1559, 1993.
[13] A. Albertsen and U.-P. Hansen, "Estimation of kinetic rate constants from multi-channel recordings by a direct fit of the time series," *Biophys. J.*, vol. 67, pp. 1393–1403, Oct. 1994.
[14] S. Klein, J. Timmer, and J. Honerkamp, "Analysis of multichannel patch clamp recordings by Hidden Markov models," *Biometr.*, vol. 53, pp. 870–884, Sept. 1997.
[15] G. W. Pulford, J. C. Gallant, R. A. Kennedy, and S. H. Chung, "Evaluation and estimation of various Markov models with applications to membrane channel kinetics," *Biomed. J.*, vol. 37, no. 1, pp. 39–63, 1995.
[16] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley, 1973, Wiley series in probability and mathematical statistics.
[17] A. P. Dempster, N. M. Laird, and D. R. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. B*, vol. 39, pp. 1–22, 1977.
[18] C. F. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, vol. 11, no. 1, pp. 95–103, 1983.
[19] I. Najfeld and T. F. Havel, "Derivatives of the matrix exponential and their computation," *Adv. Appl. Math.*, vol. 16, pp. 321–375, 1995.
[20] F. G. Ball and M. S. P. Sansom, "Ion-channel gating mechanisms: Model identification and parameter estimation from single channel recordings," in *Proc. R. Soc. Lond. B*, 1989, vol. 236, pp. 385–416.
[21] R. J. Elliott, L. Aggoun, and J. B. Moore, "Hidden Markov models: Estimation and control," in *Applications of Mathematics*. New York: Springer, 1995.
[22] J. D. Hamilton, *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press, 1994.