Humboldt-Universität zu Berlin
Institute for Theoretical Biology

Exercise Set 6
Summer 2006, May 29

# Computational Neuroscience IV: Analysis of Neural Systems

Dr. R. Kempter, Prof. Dr. A.V.M. Herz

## Estimation Theory

Assume that the parameters $\boldsymbol{\theta}$ and the observations $\mathbf{x}_T$ have the joint pdf $p_{\boldsymbol{\theta},\mathbf{x}}(\boldsymbol{\theta},\mathbf{x}_T)$. A theoretically significant, conceptually simple, general, and unbiased estimator of $\boldsymbol{\theta}$ is the **minimum mean-square error (MSE)** estimator $\hat{\boldsymbol{\theta}}_{MSE}$, which minimizes $E\{|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}|^2\}$). This MSE estimator is given by the conditional expectation

$$\hat{\boldsymbol{\theta}}_{MSE} = E_{\boldsymbol{\theta}|\mathbf{x}}\{\boldsymbol{\theta}|\mathbf{x}_T\}, \tag{1}$$

which is an expectation with respect to the so-called *posterior density* $p_{\boldsymbol{\theta}|\mathbf{x}}$. The posterior density can be derived from Bayes' formula,

$$p_{\boldsymbol{\theta}|\mathbf{x}}(\boldsymbol{\theta}|\mathbf{x}_T) = \frac{p_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}_T|\boldsymbol{\theta})\, p_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{p_{\mathbf{x}}(\mathbf{x}_T)}. \tag{2}$$

The computation of $\hat{\boldsymbol{\theta}}_{MSE}$ is difficult in practice because we may only know or assume the *prior distribution* $p_{\boldsymbol{\theta}}$ of the parameters $\boldsymbol{\theta}$ and the *conditional distribution* $p_{\mathbf{x}|\boldsymbol{\theta}}$ of the observations $\mathbf{x}_T$ given $\boldsymbol{\theta}$. The denominator is computed by integrating the numerator, $p_{\mathbf{x}}(\mathbf{x}_T) = \int \mathrm{d}\boldsymbol{\theta}'\, p_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}_T|\boldsymbol{\theta}')\, p_{\boldsymbol{\theta}}(\boldsymbol{\theta}')$. This integral and the one in (1) are usually difficult to evaluate.

To simplify the problem, one could instead estimate the parameter vector $\boldsymbol{\theta}$ that maximizes the posterior density $p_{\boldsymbol{\theta}|\mathbf{x}}$ in (2). Because $p_{\mathbf{x}}$ in (2) does not depend on $\boldsymbol{\theta}$, it is sufficient to maximize the numerator of (2), which is

$$p_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}_T|\boldsymbol{\theta})\, p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = p_{\boldsymbol{\theta},\mathbf{x}}(\boldsymbol{\theta},\mathbf{x}_T). \tag{3}$$

Maximizing $p_{\boldsymbol{\theta}|\mathbf{x}}$ we obtain the **maximum a posteriori (MAP)** estimator $\hat{\boldsymbol{\theta}}_{MAP}$ of $\boldsymbol{\theta}$. Furthermore, if the prior $p_{\boldsymbol{\theta}}$ is unknown, one can maximize $p_{\mathbf{x}|\boldsymbol{\theta}}$ alone, which leads to the **maximum likelihood (ML)** estimator $\hat{\boldsymbol{\theta}}_{ML}$ of $\boldsymbol{\theta}$.

**1.** Let the joint pdf of the parameter $\theta$ and the random variable $x$ be

$$p_{\theta,x}(\theta,x) = \begin{cases} 8\,\theta x & \text{for } 0 < \theta \le x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

a) Indicate the region where $p_{\theta,x}$ is nonzero in the $x$-$\theta$ plane.

b) Find the conditional density $p_{x|\theta}$ and show that it is normalized. Plot $p_{x|\theta}$ as a function of $x$ for different values of $\theta$. Hint: take care of the constraints on $x$ and $\theta$.

c) From $p_{x|\theta}$ derive the ML estimate $\hat{\theta}_{ML}$ of $\theta$ given $x$. Plot $p_{x|\theta}$ as a function of $\theta$ for different values of $x$ and argue why the 'naive' likelihood equation (cf. Exercises 4) leads to a wrong result here.

d) Compute the posterior density $p_{\theta|x}$ and derive the MAP estimate $\hat{\theta}_{MAP}$ of $\theta$ given $x$. Plot $p_{\theta|x}$ as a function of $\theta$ for different values of $x$.

e) Compute the optimal mean-square error estimate $\hat{\theta}_{MSE}$ of $\theta$ given $x$.

**Estimation of Noise-Free Independent Components**

In noisy ICA, where gaussian "sensor" noise $\mathbf{n}$ with covariance $\mathbf{\Sigma}$ is added to the observations $\mathbf{x}$,

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} \ ,$$

it is not enough to estimate the mixing matrix $\mathbf{A}$ because we get noisy estimates of the independent components $\mathbf{s}$. Therefore, we would like to obtain estimates of the original ICs that are somehow optimal, i.e., contain minimum noise.

We assume that we already have estimated $\mathbf{A}$. Given the data $\mathbf{x}_T$ where the subscript indicates that we have $T$ independent measurements $\mathbf{x}(t)$ $(t = 1, \ldots, T)$, we can use the MAP method to estimate the 'parameters' $\mathbf{s}$. The conditional density $p_{\mathbf{x},\mathbf{A}|\mathbf{s}}(\mathbf{x}_T, \mathbf{A}|\mathbf{s}_T) \propto \prod_{t=1}^T \exp[-||\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t)||^2_{\Sigma^{-1}}/2]$ of $\mathbf{x}_T$ and $\mathbf{A}$ given $\mathbf{s}_T$ is gaussian, where $||\mathbf{m}||^2_{\Sigma^{-1}}$ is defined as $\mathbf{m}^T\mathbf{\Sigma}^{-1}\mathbf{m}$. We also assume that we know the 'prior' distribution $p_\mathbf{s}(\mathbf{s}_T)$.

**2.** Show that the MAP log-likelihood is given by

$$\log L(\mathbf{s}) = -\sum_{t'=1}^{T} \left[ \frac{1}{2} ||\mathbf{x}(t') - \mathbf{A}\mathbf{s}(t')||^2_{\Sigma^{-1}} + \sum_{i=1}^{n} f_i(s_i(t')) \right] + C$$

where $C$ is an irrelevant constant. What is $f_i$?

**3.** To compute the MAP estimator $\hat{\mathbf{s}}(t)$, we take the gradient of the log-likelihood with respect to the elements of $\mathbf{s}(t)$ and equate this to 0. Show that this leads to an implicit condition on $\hat{\mathbf{s}}$ of the form

$$\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A}\hat{\mathbf{s}}(t) - \mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{x}(t) + \mathbf{f}'(\hat{\mathbf{s}}(t)) = 0 \qquad (4)$$

where the derivative, denoted by $\mathbf{f}'$, is applied separately to each component of the vector $\hat{\mathbf{s}}(t)$. This gives a nonlinear generalization of classic Wiener filtering. Hint: for a constant vector $\mathbf{w}$, a constant matrix $\mathbf{W}$, and some scalar function $g$, use

$$\frac{\partial[\mathbf{w}^T\,\mathbf{s}(t')]}{\partial\mathbf{s}(t)} = \mathbf{w}\,\delta_{t,t'} \ , \quad \frac{\partial[\mathbf{s}^T(t')\,\mathbf{W}\,\mathbf{s}(t')]}{\partial\mathbf{s}(t)} = \left[\mathbf{W}\,\mathbf{s}(t) + \mathbf{W}^T\,\mathbf{s}(t)\right]\,\delta_{t,t'} \ , \text{and} \quad \frac{\partial g}{\partial\mathbf{s}} = \left(\frac{\partial g}{\partial s_1}, \ldots, \frac{\partial g}{\partial s_n}\right)^T \ .$$

**4.** In order to interpret this result, consider Equation (4) in the 1-dimensional case where $\mathbf{A} = 1$, $\mathbf{\Sigma} = \sigma^2$, and $p_s(s') = \exp(-\sqrt{2}|s'|)/\sqrt{2}$ is Laplacian. Our goal is to find an estimate $\hat{s}$ of $s$ given $x$.

a) Plot $x$ as a function of $\hat{s}$ where we can write formally $x = g^{-1}(\hat{s})$ with some 'inverse' function $g^{-1}$.

b) Plot $\hat{s}$ as a function of $x$. Show that this 'shrinkage' function can be approximated by $\hat{s} = g(x)$ where $g(x) = \operatorname{sign}(x)\,\max(0, |x| - \sqrt{2}\sigma^2)$.

c) Plot the Laplacian pdf (supergaussian) and interpret the result in b) in the limits of small noise $(\sigma^2 \ll 1)$ and large noise $(\sigma^2 \gg 1)$

**5.** Repeat the calculations of Problem 4 for a uniform pdf, $p_s(s') = 1/2$ for $|s'| \leq 1$. Hint: use $\vartheta'(x) = \delta(x)$ where $\vartheta$ is the step function and $\delta$ is the Dirac delta function.

**6.** Which simplifying assumptions are necessary to derive from Equation (4) the 'linear-least square' estimator $\hat{\mathbf{s}}(t) = \mathbf{A}^{-1}\mathbf{x}(t)$ ? First state conditions for which the prior on the densities of $\hat{\mathbf{s}}$ can be neglected! Then write down an explicit equation for $\hat{\mathbf{s}}$. Finally, derive conditions on $\mathbf{\Sigma}$ and $\mathbf{A}$.

---

DR. R. KEMPTER, phone 2093-8925, room 2315, `r.kempter(AT)biologie.hu-berlin.de`
PROF. DR. A.V.M. HERZ, phone 2093-9112, room 2325, `a.herz(AT)biologie.hu-berlin.de`

Problems handed out on Monday, 29.05.2006. Solutions to be handed in by Monday, 12.06.2006, $12^{15}$ pm. Discussion and presentation of the problems on Friday, 16.06.2006, $8^{30} - 10^{00}$ am in front of room 2317 I-W (Zwischengeschoß).