Contents

Chapter 8

Parameter Estimation

The theory of parameter estimation is part of the statistical decision theory. The goal of parameter estimation is the evaluation of a parameter generated by a source of information in noisy conditions.

In our discussion the image $\mathbf{x} = [x_1, x_2, \ldots, x_N]$ is the source of information that is observed in the presence of noise $\mathbf{n} = [n_1, n_2, \ldots, n_N]$. $\mathbf{y} = [y_1, y_2, \ldots, y_N]$ are the *measured* pixel intensities.

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{8.1}$$

The image $\mathbf{x}$ is a random signal, a realization of a stochastic process. In some particular cases we will also consider the case of deterministic images. The noise n is a 2-dimensional random signal.

The problem statement is :

Given the observations $\mathbf{y}$ *and possibly some knowledge about* $\mathbf{x}$ *and* $\mathbf{n}$*, find a guess of* $\mathbf{x}$.

Example: Let x be the unknown gray-level of a uniform scene (or a part of it). A sensor provides us with N observations $y_1, y_2, \ldots, y_N$ of the image intensity. They can be considered as realizations of the N statistically independent and identically distributed random variables $Y_1, Y_2, \ldots, Y_N$. The probability density function is unknown. Assume that the measurement errors $y_i - x, \quad i = 1, \ldots, N$ satisfy the relation

$$\mathcal{E}[y_i - x] = 0 \qquad ; i = 1, \ldots, N \tag{8.2}$$

$$\Rightarrow \mathcal{E}[y_i] = x \tag{8.3}$$

and that they are not affected by any *systematic* errors. Further we assume that the higher order moments do not depend on x. Thus only the first order moment is under consideration, and the estimated value $\hat{x}$ is chosen as a linear form

$$\hat{x}(\mathbf{y}) = c_1 y_1 + c_2 y_2 + \ldots + c_N y_N \tag{8.4}$$

We find

$$\mathcal{E}[\hat{x}] = \sum_{i=1}^{N} c_i \, \mathcal{E}[y_i] = x \sum_{i=1}^{N} c_i \tag{8.5}$$

$$\sigma_{\hat{x}}^2 = \sum_{i=1}^{N} c_i^2 \, \sigma_{y_i}^2 = \sigma_x^2 \sum_{i=1}^{N} c_i^2 \tag{8.6}$$

Under the assumption of no systematic effects we set

$$\sum_{i=1}^{N} c_i = 1. \tag{8.7}$$

Thus the variance of our estimators will be minimized if the quantity $\sum_{i=1}^{N} c_i^2$ is minimal. Using the Lagrange multipliers we obtain

$$\Psi(c_1, \ldots, c_N) = \sum_{i=1}^{N} c_i^2 - 2\lambda \left(\sum_{i=1}^{N} c_i - 1 \right) \tag{8.8}$$

Setting the partial derivatives of Ψ to zero

$$\frac{\partial \Psi(c_1, \ldots, c_N)}{\partial(c_1, c_2 \cdots, c_N)} = 2 \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} - 2\lambda \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = 0 \tag{8.9}$$

we obtain the conditions for the minimum. With Eq. 8.7 follows

$$\lambda = \frac{1}{N} \, . \tag{8.10}$$

Thus

$$\hat{x} = \frac{1}{N} \sum_{i=1}^{N} y_i \, . \tag{8.11}$$

Remark: In the previous example very little assumptions about the stochastic process involved have been done. One can ask oneself: Can we derive better estimates?

8.1 Estimation of a random parameter

We consider the linearly ordered, observed pixel intensities of a random image $\mathbf{y} = [y_1, \ldots, y_N]$ that are characterized by the conditional p.d.f.

$$p(y_1, \ldots, y_N \mid x) = p(\mathbf{y} \mid x) \, . \tag{8.12}$$

Further we consider an estimate $\hat{x}$ of the unknown random gray-level x, and the estimation error

$$\epsilon_x = x - \hat{x}(\mathbf{y}) \tag{8.13}$$

Observation The error ϵ_x is only hypothetically defined, since the true value of x is unknown.

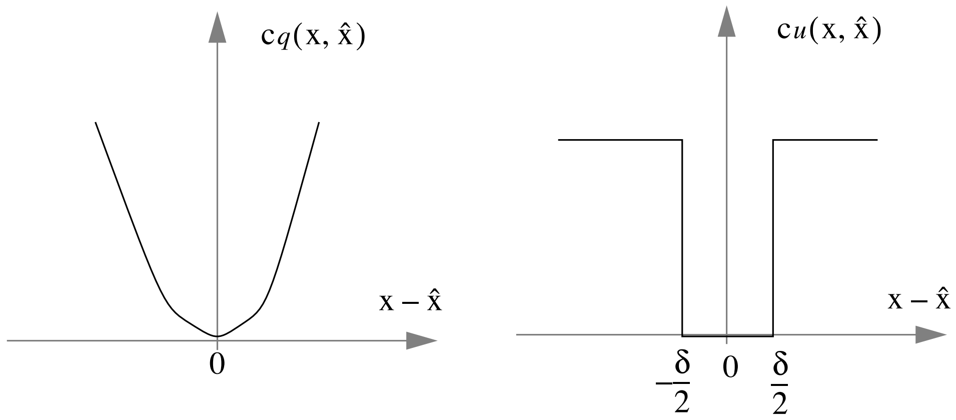

Figure 8.1: Left: Quadratic cost function which leads to the MMSE estimator. Right: Uniform cost function leading to the MAP estimator.

8.1.1 Bayes risk

The importance of the estimation error in a certain application, is measured using a *cost function* $c(\epsilon_x)$. Commonly used cost functions are the quadratic

$$c_q(\epsilon_x) = \epsilon_x^2 \tag{8.14}$$

and the uniform one (Fig. 8.1

$$c_u(\epsilon_x) = \begin{cases} 0 & : \quad |\epsilon_x| \leq \frac{\delta}{2} \\ 1 & : \quad |\epsilon_x| > \frac{\delta}{2} \end{cases} . \tag{8.15}$$

The expectation of the cost relative to the joint p.d.f. $p(x, \mathbf{y})$ is called *Bayes risk* R

$$R = \mathcal{E}[c(x - \hat{x})] = \int\int \mathrm{d}x\, \mathrm{d}\mathbf{y}\, c(x - \hat{x})\, p(x, \mathbf{y}).^1 \tag{8.16}$$

In the following the parameter estimation problem will be stated as a minimization of the Bayes risk.

8.1.2 Minimum mean square error estimators (MMSE)

Using the quadratic cost function the risk is

$$R_q = \int\int \mathrm{d}x\, \mathrm{d}\mathbf{y}\, (x - \hat{x})^2\, p(x, \mathbf{y}). \tag{8.17}$$

With the Bayes formula for conditioned p.d.f. we obtain

$$R_q = \int \mathrm{d}\mathbf{y}\, p(\mathbf{y}) \int \mathrm{d}x\, (x - \hat{x})^2\, p(x|\mathbf{y}). \tag{8.18}$$

[1] Here and in the following, the integrals are understood to be evaluated on the whole domain of definition for the variable, if not specified explicitly.

Since both integrals are non-negative, the minimum risk is obtained by minimizing

$$I(\hat{x}, \mathbf{y}) = \int \mathrm{d}x \, (x - \hat{x})^2 \, p(x|\mathbf{y}). \tag{8.19}$$

with respect to $\hat{x}$.

$$\begin{aligned} \frac{\partial}{\partial \hat{x}} I(\hat{x}, \mathbf{y}) &= 0 & (8.20) \\ 2\hat{x} \int p(x \,|\, \mathbf{y}) \, \mathrm{d}x - 2 \int x p(x \,|\, \mathbf{y}) \, \mathrm{d}x &= 0 \, . & (8.21) \end{aligned}$$

Knowing that

$$\int p(x \,|\, \mathbf{y}) \, \mathrm{d}x = 1 \tag{8.22}$$

we obtain the MMSE estimator $\hat{x}_{\mathrm{MMSE}}$

$$\hat{x}_{\mathrm{MMSE}} = \int x p(x \,|\, \mathbf{y}) \, \mathrm{d}x \tag{8.23}$$

The MMSE estimator is the conditional mean.

Observation The MMSE is a function of the observation: $\hat{x}_{MMSE} = \hat{x}_{MMSE}(\mathbf{y})$.

8.1.3 Maximum *a posteriori* estimator (MAP)

The MAP estimator assumes a uniform cost function. The risk is

$$\begin{aligned} R_u &= \int \mathrm{d}\mathbf{y} \, p(\mathbf{y}) \int \mathrm{d}x \, c_u(x - \hat{x}) \, p(x \,|\, \mathbf{y}) & (8.24) \\ &= \int \mathrm{d}\mathbf{y} \, p(\mathbf{y}) \left(1 - \int_{\hat{x} - \frac{\delta}{2}}^{\hat{x} + \frac{\delta}{2}} \mathrm{d}x \, p(x \,|\, \mathbf{y}) \right) & (8.25) \end{aligned}$$

The minimization of the risk R_u (in analogy with the previous example) requires the maximization of the integral

$$I(\hat{x}, \mathbf{y}) = \int_{\hat{x} - \frac{\delta}{2}}^{\hat{x} + \frac{\delta}{2}} \mathrm{d}x \, p(x|\mathbf{y}). \tag{8.26}$$

We observe that

$$\lim_{\delta \to 0} I(\hat{x}, \mathbf{y}) = \delta \, p(\hat{x}|\mathbf{y}). \tag{8.27}$$

The maximization of I is obtained by the maximization of the posterior density $p(x \,|\, \mathbf{y})$

$$\frac{\partial}{\partial \hat{x}} p(x \,|\, \mathbf{y})|_{x = \hat{x}_{\mathrm{MAP}}} = 0 \, . \tag{8.28}$$

The posterior can be evaluated using the Bayes formula

$$p(x \,|\, \mathbf{y}) = \frac{p(\mathbf{y} \,|\, x) \, p(x)}{p(\mathbf{y})} \, . \tag{8.29}$$

due to the fact that many p.d.f.'s used in practice have exponential forms the MAP estimator is evaluated under a logarithmic transform.

$$\frac{\partial}{\partial x}\log p(x\,|\,\mathbf{y}) = \frac{\partial}{\partial x}\log p(\mathbf{y}\,|\,x) + \frac{\partial}{\partial x}\log p(x) = 0 \tag{8.30}$$

$$\hat{x}_{\text{MAP}} = \arg\max_x \log p(x\,|\,\mathbf{y}) \tag{8.31}$$

We observe that both estimators –MMSE and MAP– use the posterior p.d.f. $p(x\,|\,\mathbf{y})$ (8.2). However the estimators extract different information and do not result in the same solution.

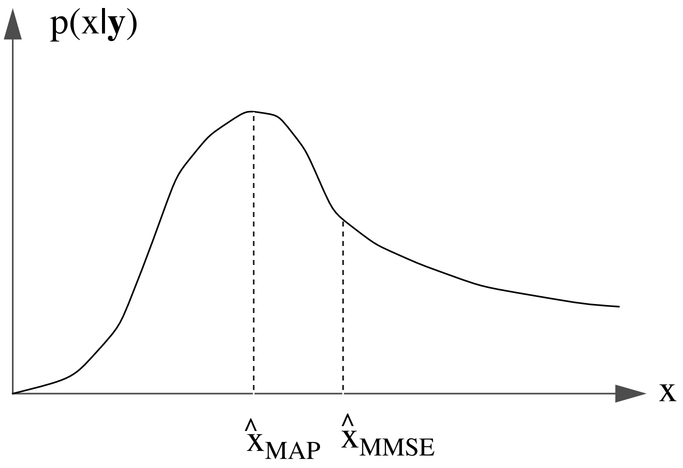

Figure 8.2: Both MMSE and MAP estimators are evaluated using the posterior p.d.f. $p(x|y)$. The MMSE is the center of mass, while the MAP is the mode of the p.d.f.

Remark In the particular case of symmetric posterior p.d.f. the MMSE and MAP estimators are equal.

Example We consider again the estimation of the intensity of a uniform (constant) image or image region. However this time we make the following assumptions:

- The image intensity x is a Gaussian r.v. of zero mean and variance σ_x^2.
- The noise n is also Gaussian of zero mean and variance σ_n^2.

We have

$$y_i = x + n_i, \qquad i = 1, 2, \ldots, N \tag{8.32}$$

$$p(\mathbf{y}\,|\,x) = \left(\frac{1}{\sqrt{2\pi\sigma_n^2}}\right)^N \prod_{i=1}^{N} \exp\left(-\frac{1}{2\sigma_n^2}(y_i - x)^2\right) \tag{8.33}$$

where we assumed independent pixels and

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right). \tag{8.34}$$

The MAP estimator is evaluated

$$\frac{\partial}{\partial x}\left(-\frac{1}{\sigma_n^2}\sum_{i=1}^{N}(y_i - x)^2\right) - \frac{\partial}{\partial x}\left(-\frac{1}{\sigma_x^2}x^2\right) = 0 \tag{8.35}$$

$$\frac{1}{\sigma_n^2}\sum_{i=1}^{N} y_i - \frac{N}{\sigma_n^2}\hat{x} - \frac{1}{\sigma_x^2}\hat{x} = 0 \tag{8.36}$$

Thus

$$\hat{x}_{\mathrm{MAP}} = \frac{\sigma_x^2}{\sigma_x^2 + \frac{1}{N}\sigma_n^2}\left(\frac{1}{N}\sum_{i=1}^{N} y_i\right) \tag{8.37}$$

If $N \to \infty$ or if the noise has small variance then

$$\tilde{x}_{\mathrm{MAP}} = \frac{1}{N}\sum_{i=1}^{N} y_i. \tag{8.38}$$

The sum is evaluated in a window of N pixels in the assumption of uniformity. Precise estimation in the vicinity of a border between two uniform areas requires a small analyzing window. That will result in an estimator strongly dependent on the prior knowledge of the noise and the *a priori* model $p(x)$.

For windows overlapping the borders of a uniform region the estimator will not be correct, the *a priori* information is not the same (8.3). The solution is the estimation with adaptive windows which follow the border. The algorithms are generally iterative and adapt the window according to a preceding estimation. One has to iterate until a certain measure of goodness reaches a heuristic threshold.

8.2 Estimation of a deterministic unknown parameter

If the parameter to be estimated is unknown but deterministic the previous shortcuts do not apply. The prior p.d.f. is a delta function

$$p(x) = \delta(x - x_0) \tag{8.39}$$

thus making the risk evaluation inconsistent. Instead we introduce the likelihood function $L(x)$

$$L(x) = p(\mathbf{y}|x) \tag{8.40}$$

$L(x)$ reaches its maximum when the noise is zero with high probability.

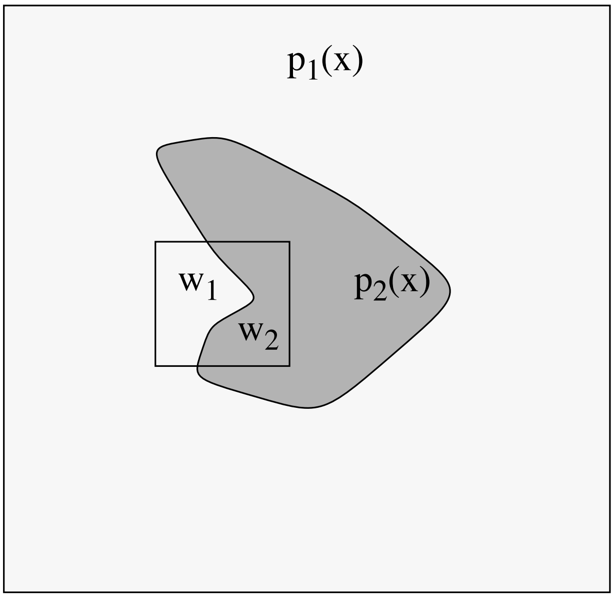

Figure 8.3: Example of image with two regions modelled with two priors, $p_1(x)$ and $p2(x)$. Precise parameter estimation at the region border requires computations in adaptive windows.

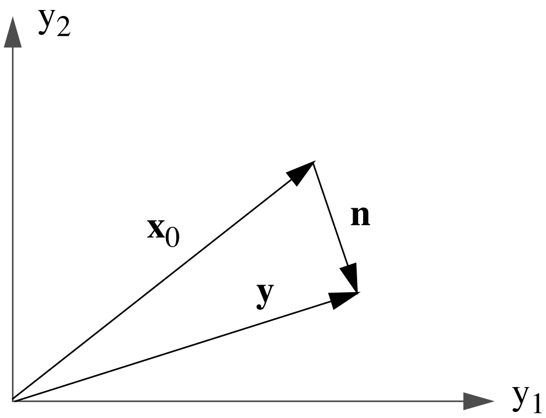

Figure 8.4: A deterministic parameter x_0 observed in noisy conditions. Where $\mathbf{n}$ is the noise and $\mathbf{y}$ the observation.

8.2.1 Maximum likelihood estimator (ML)

The maximum likelihood estimator is defined as

$$\hat{x}_{\mathrm{ML}} = \arg\max_{x} p(\mathbf{y} \,|\, x) \tag{8.41}$$

and it is evaluated in general using the log-likelihood

$$\frac{\partial}{\partial x} \log p(\mathbf{y} \,|\, x) = 0 \tag{8.42}$$

Remark We note that the ML estimator can be seen as a particular case of the MAP estimator, i.e. the p.d.f. p(x) is uniform. Thus if

$$p(x) = \begin{cases} \frac{1}{b-a} & : \ x \in [a,b] \\ 0 & : \ else \end{cases} \tag{8.43}$$

$$\Rightarrow \hat{x}_{\mathrm{MAP}} = \hat{x}_{\mathrm{ML}} \tag{8.44}$$

Also we observe that for symmetric $p(\mathbf{y} \,|\, x)$ and $p(x)$

$$\hat{x}_{\mathrm{ML}} = \hat{x}_{\mathrm{MMSE}} \tag{8.45}$$

the ML estimator is equal to the MMSE estimator.

We will call estimators that minimize the Bayes risk, like the MMSE and MAP estimators, *Bayes estimators*. Even if the ML is not based on a Bayes risk we will still consider it as a Bayes estimator.

8.3 Discussion and particular cases of the Bayes estimators

The Bayes estimators were introduced based on the risk minimization. Where the risk is defined *knowing*, or assuming a certain cost function. This can be interpreted as a model –the model for the penalty of the estimation error.

Further the Bayes estimation requires the knowledge of the joint p.d.f. $p(x, y)$ which is nothing but the complete statistical model of the stochastic processes involved: the random image and the noise.

The last assumed model was for the observation process

$$y_i = x + n_i \ , i = 1, 2, \ldots, N. \tag{8.46}$$

We already noted that the MAP estimator is –in particular cases– equal to the MMSE or the ML estimator with no reference to the initially defined cost function. This observation leads us to the following comparative analysis of the Bayes estimators.

8.3.1 Comparative analysis of the Bayes estimators

The MAP is the most complete approach.

$$\hat{x}_{\mathrm{MAP}} = \arg\max_{x} p(x \,|\, \mathbf{y}) \tag{8.47}$$

with

$$p(x \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid x)\, p(x)}{p(\mathbf{y})} , \tag{8.48}$$

where

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}} . \tag{8.49}$$

The likelihood represents the incertitude in the measurements that is introduced by the noise.

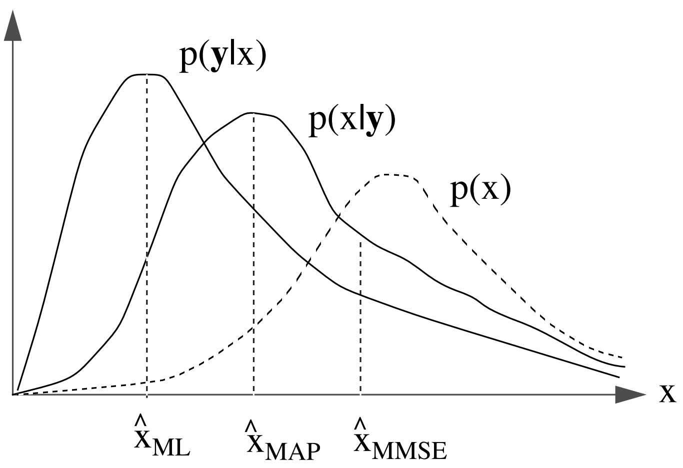

Figure 8.5: Comparison of the Bayes estimators. The MMSE is obtained as the center of mass of the posterior p.d.f., the MAP as the argument of the maximum of the posterior p.d.f. and the ML as argument of the maximum of the likelihood p.d.f.

From the modality of reasoning in the case of the ML estimator we observe that the likelihood

$$p(y_i \mid x) = p_n(y_i - x) \tag{8.50}$$

where $p_n(\cdot)$ is the p.d.f. of the noise process. The following diagram Fig. 8.5 shows qualitatively the MAP estimation. The ML is obtained from the maximum of $p(y \mid x)$. The MMSE is the expectation relative to the conditional p.d.f. $p(x \mid y)$.

If the prior is uniform (8.6)

$$\hat{x}_{\text{MAP}} = \hat{x}_{\text{ML}}. \tag{8.51}$$

If the posterior is symmetric (8.7)

$$\hat{x}_{\text{MAP}} = \hat{x}_{\text{MMSE}}. \tag{8.52}$$

If the likelihood and the prior are symmetric all these estimators are equal.

We conclude remarking that the MAP is the most complete model-based estimator. The used models in terms of prior knowledge encapsulated is the estimator can change the estimation values. Thus utilization of an inadequate model leads to erroneous results. The ML estimator is the least committed one. In absence of a reliable prior, it can be considered to give the best results.

In general, the Bayes estimators are non-linear.

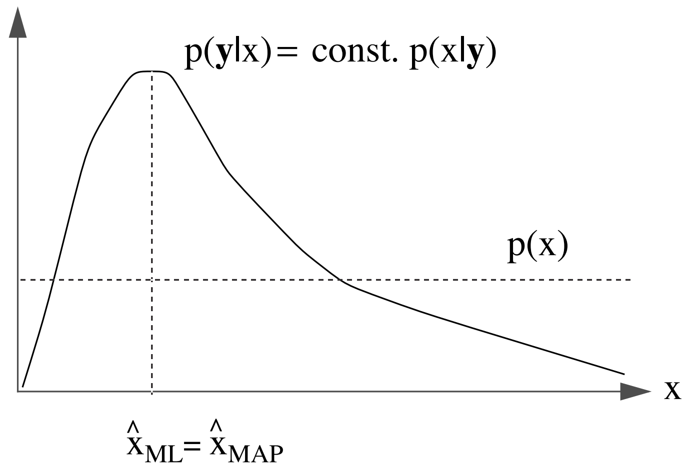

Figure 8.6: The particular case of uniform prior. The MAP and ML estimators are identical.

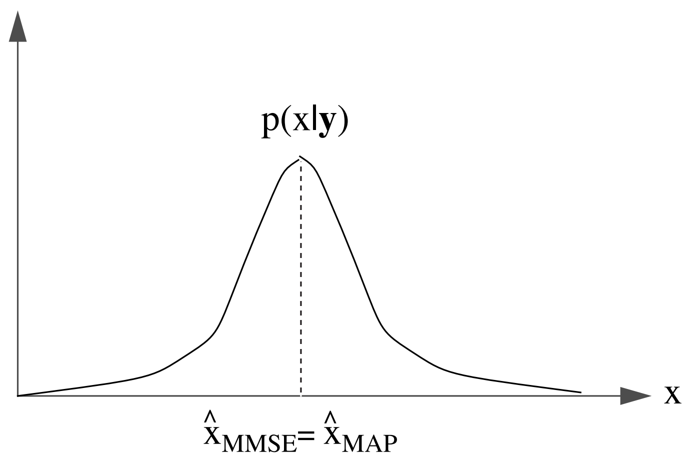

Figure 8.7: The particular case of symetric posterior. The MAP and MMSE estimators are identical.

8.3.2 Linear Gaussian models

We consider a more complete observation model.

$$\mathbf{y} = H\mathbf{x} + \mathbf{n} \tag{8.53}$$

where H is a linear operator in matrix form. We assume x and n to be statistically independent and the likelihood to be of Gaussian type

$$p(\mathbf{y} \,|\, \mathbf{x}) = (2\pi)^{-\frac{N}{2}} (\det C_n)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y} - H\mathbf{x})^T C_n^{-1} (\mathbf{y} - H\mathbf{x})\right) \tag{8.54}$$

where C_n is the covariance matrix of the noise. We let also the prior be Gaussian with zero mean

$$p(\mathbf{x}) = (2\pi)^{-\frac{N}{2}} (\det C_x)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{x}^T C_x^{-1} \mathbf{x}\right). \tag{8.55}$$

The MAP estimator is

$$\begin{aligned} \hat{\mathbf{x}}_{\text{MAP}} &= \arg\max_{\mathbf{x}} \log p(\mathbf{y} \,|\, \mathbf{x}) + \log p(\mathbf{x}) & (8.56)\\ &= \arg\min_{\mathbf{x}} \underbrace{\left((\mathbf{y} - H\mathbf{x})^T C_n^{-1} (\mathbf{y} - H\mathbf{x}) + \mathbf{x}^T C_x^{-1} \mathbf{x}\right)}_{\Psi(\mathbf{x})}. & (8.57) \end{aligned}$$

The estimator is obtained from the equation

$$\begin{aligned} \frac{\partial \Psi(\mathbf{x})}{\partial \mathbf{x}} &= 0 & (8.58)\\ -H^T C_n^{-1}(\mathbf{y} - H\mathbf{x}) + C_x^{-1}\mathbf{x} &= 0 & (8.59) \end{aligned}$$

$$\hat{\mathbf{x}}_{\text{MAP}} = \left(H^T C_n^{-1} H + C_x^{-1}\right)^{-1} H^T C_n^{-1} \mathbf{y} \tag{8.60}$$

For Gaussian prior and noise models with a linear observation process the MAP estimator is linear.

Exercise For scalar parameter x use Eq. 8.60 to derive Eg. 8.37

Remark Due to the symmetry of the Gaussian p.d.f. we have

$$\hat{\mathbf{x}}_{\text{MAP}} = \hat{\mathbf{x}}_{\text{MMSE}} \tag{8.61}$$

Remark If the prior is uniform we obtain the ML estimator

$$\hat{\mathbf{x}}_{\text{ML}} = (H^T C_n^{-1} H)^{-1} H^T C_n^{-1} \mathbf{y} \tag{8.62}$$

Remark If we are faced with an estimation problem where no prior knowledge is available, we make the simplest allowed assumptions:

$$C_x = 0 \qquad ; \qquad C_n = \mathrm{I}. \tag{8.63}$$

This results in the Least Squares Estimator (LSE)

$$\hat{\mathbf{x}}_{\text{LSE}} = (H^T H)^{-1} H^T \mathbf{y}. \tag{8.64}$$

However the LSE estimator still assumes uncorrelated, Gaussian noise.

8.4 The Wiener filter

We further deal with the parameter estimation in the special situation where only the covariance matrix of the stochastic process is known. We consider the observation model

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{8.65}$$

and we search a transform W which gives the estimate $\hat{x}$

$$\hat{\mathbf{x}} = W\,\mathbf{y} \tag{8.66}$$

such that the mean square error (MSE) is minimized

$$\epsilon^2 = \mathcal{E}[\,(\mathbf{x} - \hat{\mathbf{x}})^2\,] \to \min \tag{8.67}$$

For simplicity we assume

$$\mathcal{E}[\,\mathbf{x}\,] = \mathcal{E}[\,\mathbf{y}\,] = \mathcal{E}[\,\mathbf{n}\,] = \mathbf{0} \tag{8.68}$$

The random parameter to be estimated $\mathbf{x}$ is assumed uncorrelated with the noise process.

$$\mathcal{E}[\,\mathbf{x}\mathbf{n}^T\,] = 0 \tag{8.69}$$

The resulting expression of the mean square error is:

$$\begin{aligned}\epsilon^2 &= \mathcal{E}[(\mathbf{x} - W\mathbf{y})^T\,(\mathbf{x} - W\mathbf{y})] && (8.70)\\ &= \mathcal{E}[\,\mathbf{y}^T W^T W\,\mathbf{y}\,] - \mathcal{E}[\,\mathbf{x}^T\,W\,\mathbf{y}\,] - \mathcal{E}[\,\mathbf{y}^T\,W^T\,\mathbf{x}\,] + \mathcal{E}[\,\mathbf{x}^T\mathbf{x}\,]. && (8.71)\end{aligned}$$

Since $\mathbf{x}^T W\mathbf{y}$ is a scalar, we can write

$$\mathbf{x}^T W\mathbf{y} = \mathbf{y}^T W^T\mathbf{x} \tag{8.72}$$

so that Eq. (8.71) can be simplified to

$$\epsilon^2 = \mathcal{E}[\,\mathbf{y}^T W^T W\,\mathbf{y}\,] - 2\,\mathcal{E}[\,\mathbf{y}^T\,W^T\,\mathbf{x}\,] + \mathcal{E}[\,\mathbf{x}^T\mathbf{x}\,]. \tag{8.73}$$

We solve for the minimizing matrix W

$$\nabla_W \epsilon^2 = 0 \tag{8.74}$$

where ∇_W denotes the partial derivations relative to the matrix elements

$$\nabla_W g = \begin{bmatrix} \frac{\partial g}{\partial W_{11}} & \frac{\partial g}{\partial W_{12}} & \cdots \\ \frac{\partial g}{\partial W_{21}} & \frac{\partial g}{\partial W_{22}} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{8.75}$$

Using the following identities

$$\begin{aligned}\nabla_W \left\{ \mathbf{y}^T\,W^T W\,\mathbf{y} \right\} &= 2W\mathbf{y}\mathbf{y}^T && (8.76)\\ \nabla_W \left\{ \mathbf{y}^T\,W^T\,\mathbf{x} \right\} &= \mathbf{x}\,\mathbf{y}^T && (8.77)\\ \nabla_W \left\{ \mathbf{x}\,\mathbf{x}^T \right\} &= 0 && (8.78)\end{aligned}$$

the MMSE condition becomes

$$W\ \left(\mathcal{E}[\,\mathbf{x}\,\mathbf{x}^T\,] + \mathcal{E}[\,\mathbf{n}\,\mathbf{n}^T\,]\right) - \mathcal{E}[\,\mathbf{x}\,\mathbf{x}^T\,] = 0. \tag{8.79}$$

The expectation values are the covariance matrices of the corresponding variables and we write in a shorthand notation

$$W\,(C_x + C_n) - C_x = 0 \tag{8.80}$$

$$W = C_x\,(C_x + C_n)^{-1}. \tag{8.81}$$

Thus the MMSE estimator expressed in terms of the covariances is

$$\hat{\mathbf{x}} = C_x\,(C_x + C_n)^{-1}\,\mathbf{y}. \tag{8.82}$$

Observation In the stationary case, the Wiener filter can be computed in the frequency domain using the Wiener-Khintchine theorem (see Section 7.3.6.).

Exercise Derive the Wiener filter for the observation model $\mathbf{y} = H\mathbf{x} + \mathbf{n}$.

8.5 Outlook

Chapter 8 introduced the fundamental concepts of parameter estimation theory. The estimation of random parameters was defined as an optimization problem: the minimization of the Bayes risk. The risk was defined as the expected cost of the estimation error. The cost function is understood as a model for the penalty of the estimation error. Two basic estimators have been derived: the minimum mean square error (MMSE) assuming a quadratic cost function, and the maximum *a posteriori* (MAP) in the assumption of a uniform cost function.

The estimation of deterministic unknown parameters was treated as a separate case and solved introducing the maximum likelihood (ML) estimator.

It was shown that the MAP is the most complete estimator: it encapsulates the knowledge of the noise process, the description of the observation process and the *a priori* model of the desired parameter. It was also demonstrated that the MAP estimator includes as special cases the MMSE, and ML estimators.

The above mentioned estimators are in general nonlinear. As particular situations the MAP estimator for Gaussian assumptions was presented as an example of linear estimator, which further was used to define the least squares estimator (LSE). The last section was introduced the Wiener filter as a particular MMSE estimator based only on the knowledge of the covariances of the noise and of the desired process.