2

Image and Noise Models

Maximum a posteriori (MAP) estimation is a statistical method for denoising of data, which takes into account statistical prior information on the clean data and on the noise process. The maximum a posteriori estimate is the most likely data under the assumption of priors for the data and the noise.

Typically, noise is assumed to be Gaussian, Laplacian, or Poisson distributed. Prior distributions of images are derived from histograms of training data. Under such assumptions, MAP estimation reduces to a discrete variational regularization problem.

In this chapter, we first review basic statistical concepts. Applying these concepts to discrete, digital images, we discuss several noise models and derive priors for image data from histograms of "comparable" image data. Finally, we show how this information can be used for MAP estimation.

2.1 Basic Concepts of Statistics

A *random experiment* is a "process, whose outcome is not known in advance with certainty" (see [129, p. 5]). The set of possible outcomes is referred to as the *sampling space* of the process. A *probability distribution* or *probability measure* P on a sampling space Ω is a measure that satisfies $P(\Omega) = 1$.

Let Ω be a sampling space with probability distribution P. A measurable function $\Delta : \Omega \to \mathbb{R}$ is called *random variable*. By $\mathrm{Ran}(\Delta) := \{\Delta(\omega) : \omega \in \Omega\}$ we denote the *range* of Δ. The random variable Δ induces a measure P_Δ on $\mathbb{R}$ by

$$P_\Delta(A) := P(\Delta^{-1}A)\,, \qquad A \subset \mathbb{R} \text{ measurable}\,.$$

An element $x \in \mathrm{Ran}(\Delta)$ is called *realization* of Δ, and a P_Δ-measurable subset of $\mathbb{R}$ is called an *event*. For simplicity we write $P_\Delta(x) := P_\Delta(\{x\})$.

If $\mathrm{Ran}(\Delta)$ is discrete, then P_Δ is called *discrete* probability distribution. In this case, the probability distribution is uniquely determined by the values $P_\Delta(x)$, $x \in \mathrm{Ran}(\Delta)$.

If there exists a non-negative Borel function $p_\Delta : \mathbb{R} \to \mathbb{R}_{\geq 0}$ such that

$$P_\Delta(A) = \int_A p_\Delta\,, \qquad A \subset \mathbb{R} \text{ measurable}\,,$$

then P_Δ is called a *(absolutely) continuous* probability distribution. In this case, the function p_Δ is called the *probability density* of Δ.

Assume that Ω is a sampling space with probability distribution P. An n-dimensional *random vector* $\boldsymbol{\Delta} = (\Delta_1, \dots, \Delta_n)$ is a measurable function $\boldsymbol{\Delta} : \Omega \to \mathbb{R}^n$. The *joint probability* $P_{\boldsymbol{\Delta}}$ of $\boldsymbol{\Delta}$ is the measure on $\mathbb{R}^n$ defined by

$$P_{\boldsymbol{\Delta}}(A) := P\bigl(\boldsymbol{\Delta}^{-1}(A)\bigr)\,, \qquad A \subset \mathbb{R}^n \text{ measurable}\,.$$

The probability density of a random vector $\boldsymbol{\Delta}$ is defined analogously to the probability density of a random variable.

If $\boldsymbol{\Delta}$ is an n-dimensional random vector on Ω, then its components Δ_i, $1 \leq i \leq n$, are themselves random variables on Ω. We say that the random vector $\boldsymbol{\Delta}$ consists of *independent* random variables Δ_i, if

$$P_{\boldsymbol{\Delta}}(A_1 \times \cdots \times A_n) = P_{\Delta_1}(A_1) \cdots P_{\Delta_n}(A_n)\,, \qquad A_1, \dots, A_n \subset \mathbb{R} \text{ measurable}\,,$$

where P_{Δ_i} are the probability distributions of Δ_i, $1 \leq i \leq n$. If additionally $P_{\Delta_i} = P_{\Delta_j}$ for all $1 \leq i, j \leq n$, then $\boldsymbol{\Delta}$ consists of *independent and identically distributed*, in short i.i.d., random variables.

The probability density of a random vector of independent continuous random variables can be determined by the following result:

Theorem 2.1. *Let $\boldsymbol{\Delta}$ be a random vector consisting of independent random variables Δ_i, $1 \leq i \leq n$, with continuous probability distributions P_{Δ_i} and corresponding densities p_{Δ_i}. Then $P_{\boldsymbol{\Delta}}$ is continuous, and its probability density $p_{\boldsymbol{\Delta}}$ is given by*

$$p_{\boldsymbol{\Delta}} = \prod_{i=1}^{n} p_{\Delta_i}\,.$$

Proof. See, e.g., [321, Thm. I.3.2.]. □

Definition 2.2. *Assume that $\boldsymbol{\Delta}$ is an n-dimensional random vector with probability distribution $P_{\boldsymbol{\Delta}}$, and that $f : \mathbb{R}^n \to \mathbb{R}^m$, $1 \leq m \leq n$, is continuous. The* push forward *$f^{\#}\boldsymbol{\Delta}$ of $\boldsymbol{\Delta}$ is the m-dimensional random vector defined by the probability distribution*

$$P_{f^{\#}\boldsymbol{\Delta}}(A) := P_{\boldsymbol{\Delta}}(f^{-1}A)\,, \qquad A \subset \mathbb{R}^m \text{ measurable}\,.$$

For a Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}^m$, $1 \leq m \leq n$, the Jacobian is defined as

$$J_f := \sqrt{\det(\nabla f\, \nabla f^T)}\,. \tag{2.1}$$

If f is Lipschitz and has a non-vanishing Jacobian almost everywhere and $\mathbf{\Delta}$ has a continuous probability distribution, then also $P_{f\#\mathbf{\Delta}}$ is a continuous probability distribution. In this case, its density can be determined by means of the following lemma.

Lemma 2.3. *Let $\mathbf{\Delta}$ be an n-dimensional continuous random vector with probability distribution $P_{\mathbf{\Delta}}$ and density $p_{\mathbf{\Delta}}$. Assume that $f : \mathbb{R}^n \to \mathbb{R}^m$, $1 \le m \le n$, is locally Lipschitz such that its Jacobian satisfies $J_f \neq 0$ almost everywhere in $\mathbb{R}^n$. Then*

$$p_{f\#\mathbf{\Delta}}(\mathbf{y}) = \int_{f^{-1}(\mathbf{y})} \frac{p_{\mathbf{\Delta}}(\mathbf{x})}{J_f(\mathbf{x})} \, \mathrm{d}\mathcal{H}^{n-m}\,, \qquad \mathbf{y} \in \mathbb{R}^m\,,$$

where $\mathcal{H}^{n-m}$ denotes the $(n-m)$-dimensional Hausdorff measure (see (9.1)).

Proof. By definition, we have for every measurable set $A \subset \mathbb{R}^m$ that

$$\int_A p_{f\#\mathbf{\Delta}}(\mathbf{y}) = P_{f\#\mathbf{\Delta}}(A) = P_{\mathbf{\Delta}}\bigl(f^{-1}(A)\bigr) = \int_{f^{-1}(A)} p_{\mathbf{\Delta}}(\mathbf{x})\,. \tag{2.2}$$

Using the coarea formula (see [159, Thm. 3.2.12], where as function g there we use $g = (p_{\mathbf{\Delta}}/J_f)\chi_{f^{-1}(A)})$, we find that

$$\int_{f^{-1}(A)} p_{\mathbf{\Delta}}(\mathbf{x}) = \int_A \int_{f^{-1}(\mathbf{y})} \frac{p_{\mathbf{\Delta}}(\mathbf{x})}{J_f(\mathbf{x})} \, \mathrm{d}\mathcal{H}^{n-m}\,. \tag{2.3}$$

Combining (2.2) and (2.3), it follows that

$$\int_A p_{f\#\mathbf{\Delta}}(\mathbf{y}) = \int_A \int_{f^{-1}(\mathbf{y})} \frac{p_{\mathbf{\Delta}}(\mathbf{x})}{J_f(\mathbf{x})} \, \mathrm{d}\mathcal{H}^{n-m}\,, \qquad A \subset \mathbb{R}^m \text{ measurable}\,.$$

This shows the assertion. □

Definition 2.4 (Mean and variance). *Let Δ be a random variable with probability distribution P_Δ. We define the* mean *(or* expectation*)* $\mathrm{E}(\Delta)$ *and the* variance $\mathrm{Var}(\Delta)$ *by*

$$\mathrm{E}(\Delta) := \int_{\mathbb{R}} x \, \mathrm{d}P_\Delta\,, \qquad \mathrm{Var}(\Delta) := \int_{\mathbb{R}} \bigl(x - E(\Delta)\bigr)^2 \, \mathrm{d}P_\Delta\,,$$

provided the integrals exist. If the distribution P_Δ is continuous with density p_Δ, then we have

$$\mathrm{E}(\Delta) := \int_{\mathbb{R}} p_\Delta(x)\, x\,, \qquad \mathrm{Var}(\Delta) := \int_{\mathbb{R}} p_\Delta(x)\bigl(x - E(\Delta)\bigr)^2\,.$$

We call $\sqrt{\mathrm{Var}(\Delta)}$ the standard deviation *of Δ.*

Remark 2.5. Repeating a random experiment, we obtain a finite number of realizations (a *sample*) of a random variable. Based on this sample, we can define a discrete probability distribution on $\mathbb{R}$:

Let $\delta_1, \ldots, \delta_n$ denote n realizations of a random variable Δ. Then the vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$ defines a probability distribution on $\mathbb{R}$ by

$$P_{\boldsymbol{\delta}}(x) := \frac{1}{n} \left| \left\{ i \in \{1, \ldots, n\} : \delta_i = x \right\} \right| . \tag{2.4}$$

We refer to $P_{\boldsymbol{\delta}}(x)$ as the *empirical* probability distribution of $\boldsymbol{\delta}$. We denote

$$\mathrm{E}(\boldsymbol{\delta}) := \frac{1}{n} \sum_{i=1}^{n} \delta_i$$

the *sample mean* and

$$\mathrm{Var}(\boldsymbol{\delta}) := \frac{1}{n} \sum_{i=1}^{n} \bigl(\delta_i - E(\delta)\bigr)^2$$

the *sample variance* of $\boldsymbol{\delta}$, i.e., $\mathrm{E}(\boldsymbol{\delta})$ and $\mathrm{Var}(\boldsymbol{\delta})$ are the mean and variance of the probability density $P_{\boldsymbol{\delta}}(x)$ defined in (2.4), respectively.

In particular, $\mathrm{E}(\boldsymbol{\delta})$ and $\mathrm{Var}(\boldsymbol{\delta})$ are the mean and variance, respectively, of the empirical probability distribution of $\boldsymbol{\delta}$. ♢

Remark 2.6. Let Δ be a random variable. Assume that $\mathrm{Var}(\Delta)$ and $\mathrm{E}(\Delta)$ exist. Then

$$\mathrm{Var}(\Delta) = \mathrm{E}(\Delta^2) - \mathrm{E}(\Delta)^2 ,$$

where Δ^2 is the push-forward of Δ by the function $f(x) = x^2$ (see, e.g., [129, Thm. 4.3.3]). ♢

Example 2.7. We recall some important distributions on $\mathbb{R}$ and $\mathbb{R}^n$, which are required below for the definitions of image noise models. Details and motivations for these distributions can be found in [129].

1. The *Poisson distribution* is a discrete distribution P with range $\mathrm{Ran}(P) = \mathbb{N} \cup \{0\}$. It is given by

$$P(k) = \frac{\lambda^k}{k!} \exp(-\lambda) , \qquad k \in \mathbb{N} \cup \{0\} , \tag{2.5}$$

   where the parameter $\lambda \geq 0$ is at the same time the mean and the variance of P.
2. Let $I \subset \mathbb{R}$ be measurable with $0 < \mathcal{L}^1(I) < \infty$. The *uniform distribution* on I is given by the probability density

$$p(x) = \begin{cases} \mathcal{L}^1(I)^{-1} , & \text{if } x \in I , \\ 0 , & \text{if } x \notin I . \end{cases}$$

3. The *Laplacian distribution* on $\mathbb{R}$ with mean $\bar{x} \in \mathbb{R}$ and $\sigma_1 > 0$ is given by the probability density

$$p(x) = \frac{1}{2\sigma_1} \exp\left(-\frac{|x - \bar{x}|}{\sigma_1}\right), \qquad x \in \mathbb{R}. \tag{2.6}$$

4. The *Gaussian distribution* on $\mathbb{R}$, also called *normal distribution*, with mean $\bar{x}$ and standard deviation $\sigma_2 > 0$ is given by the probability density

$$p(x) = \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{|x - \bar{x}|^2}{2\sigma_2^2}\right). \tag{2.7}$$

5. If $\boldsymbol{\Delta}$ is a random vector consisting of i.i.d. random variables, then the probability density of $\boldsymbol{\Delta}$ is given as the product of the probability densities of Δ_i (cf. Theorem 2.1). For example, for i.i.d. Gaussian random variables we have

$$p(\mathbf{x}) = \frac{1}{\left(\sigma_2\sqrt{2\pi}\right)^n} \exp\left(-\frac{|\mathbf{x} - \bar{\mathbf{x}}|^2}{2\sigma_2^2}\right),$$

where $\bar{\mathbf{x}} = (\bar{x}, \dots, \bar{x})^T \in \mathbb{R}^n$ (compare with the more general definition of *multivariate* (or *vectorial*) Gaussian distribution in the literature, see for example [321, Sect. VIII.4]). Note that here $|\mathbf{x} - \bar{\mathbf{x}}|$ denotes the Euclidean norm on $\mathbb{R}^n$. $\diamondsuit$

2.2 Digitized (Discrete) Images

In this section, we give the basic model of discrete and continuous images as used in the sequel.

Let $h > 0$ and $n_x, n_y \in \mathbb{N}$. *Discrete images* of size $n_x \times n_y$ are given as matrices $\mathbf{u} = (u_{ij})_{(i,j)\in\mathcal{I}_1}$, where

$$\boxed{u_{ij} \in \mathbb{R}, \qquad (i,j) \in \mathcal{I}_1 := \{1, \dots, n_x\} \times \{1, \dots, n_y\},}$$

describe the intensity values of a digital image at the nodal points

$$\boxed{x_{ij} = (ih, jh), \qquad (i,j) \in \mathcal{I}_1,}$$

of a regular rectangular *pixel grid* $\mathbf{x} = (x_{ij})$. The parameter h controls the resolution of the image, that is, the horizontal and vertical distance of the *pixels* x_{ij} (see Fig. 2.1). Note that in the literature, sometimes pixels are defined as rectangles with midpoints x_{ij}.

In contrast with digital photography, where intensities are assumed to be integers in a certain range (for instance, between 0 and 255), we allow for arbitrary real values in the consideration below.

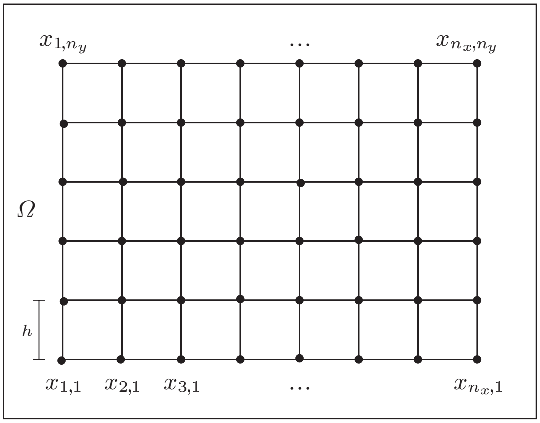

Fig. 2.1. Pixel grid with nodes $x_{ij} = (ih, jh)$.

A *continuous image* is given by its *intensity function* $u : \Omega \to \mathbb{R}$, where

$$\Omega := \big(0, (n_x + 1)h\big) \times \big(0, (n_y + 1)h\big) .$$

Note that Ω is chosen in such a way that the pixel grid $\mathbf{x}$ is contained in the *interior* of Ω.

To every pair (i, j) in the set

$$\mathcal{I}_2 := \{1, \ldots, n_x - 1\} \times \{1, \ldots, n_y - 1\}$$

we assign the discrete gradient v_{ij} of u at x_{ij} setting

$$\boxed{v_{ij} := \frac{1}{h} \begin{pmatrix} u_{i+1,j} - u_{ij} \\ u_{i,j+1} - u_{ij} \end{pmatrix} .} \tag{2.8}$$

The resulting mapping $\mathbf{v} : \mathcal{I}_2 \to \mathbb{R}^2$ is called the *discrete gradients matrix*. Note that this matrix is not an ordinary matrix of scalars, but its entries are actually vectors. Moreover, we denote the matrix of norms of the discrete gradients $|v_{ij}|$ by $|\mathbf{v}|$.

We distinguish discrete gradients $\mathbf{v}$ of a *discrete image* from one-sided discrete gradients $\nabla_h u$ of a *continuous image* u, which are defined by

$$\boxed{\nabla_h u(x_{ij}) := \frac{1}{h} \begin{pmatrix} u(x_{i+1,j}) - u(x_{ij}) \\ u(x_{i,j+1}) - u(x_{ij}) \end{pmatrix} , \qquad (i, j) \in \mathcal{I}_2 .}$$

In the special case that the discrete image $\mathbf{u}$ is given as pointwise discretization of a continuous image u, that is, $u_{ij} = u(x_{ij})$, we obtain the equality of gradients $v_{ij} = \nabla_h u(x_{ij})$. It is, however, convenient in certain applications to also allow more general discretizations with respect to which the equality does not necessarily hold.

2.3 Noise Models

In this section, we discuss noise models corresponding to different distortions in image recording. We concentrate first on intensity errors, which are realizations of independent random variables, acting on each pixel location separately, and then on sampling errors, where the observed error depends on surrounding pixels as well.

Intensity Errors

The simplest model for intensity errors is *additive noise*. Let $\mathbf{u}$ be a discrete image and $\boldsymbol{\delta} = (\delta_{ij})_{ij}$ be an $n_x \times n_y$ matrix of realizations of i.i.d. random variables. If the recorded data are

$$\boxed{\mathbf{u}^\delta = \mathbf{u} + \boldsymbol{\delta}\,,} \tag{2.9}$$

then we speak of *additive intensity errors* in the image data. If each random variable is Gaussian distributed, we speak of *Gaussian intensity errors.* Other commonly used noise models assume a Laplacian, uniform, or Poisson distribution (with constant parameter) of the random variables. Variational approaches for removing additive Gaussian intensity errors are discussed in the subsequent sections.

A model of *multiplicative noise* is given by

$$\boxed{\mathbf{u}^\delta = \mathbf{u} \cdot \boldsymbol{\delta}\,,}$$

where, again, $\boldsymbol{\delta} = (\delta_{ij})_{ij}$ is a matrix of realizations of (non-negative) i.i.d. random variables, and the multiplication is understood pointwise, that is, $u_{ij}^\delta = u_{ij}\delta_{ij}$. We then speak of *multiplicative intensity errors.* A variational denoising approach taking into account such a noise model has been studied in [337, 346, 347]. Aubert & Aujol [25] have considered multiplicative *Gamma noise* and developed an adequate variational denoising approach.

Poisson noise and *Salt-and-Pepper noise* are prominent noise models with a functional dependency of the noise $\boldsymbol{\delta}$ on $\mathbf{u}$, which is neither multiplicative nor additive, that is,

$$\boxed{\mathbf{u}^\delta = \boldsymbol{\delta}(\mathbf{u})\,.}$$

Photon counting errors produced by CCD sensors are typically modeled by Poisson noise [40, 223, 359]. Let us consider a camera with a two-dimensional array of CCD sensors, each sensor (i, j) corresponding to a position x_{ij} of the sensor. During exposure, each sensor counts the number of incoming photons at x_{ij}. Because this number is non-negative, the vector $\mathbf{u}$ has non-negative entries.

The number of photons $\delta_{ij}(\mathbf{u})$ detected by the CCD sensor can be modeled as a realization of a Poisson distributed random variable with mean u_{ij}. Then the probability for measuring the value $k \in \mathbb{N} \cup \{0\}$ at the pixel position x_{ij} is given by the probability distribution $P_{\Delta_{ij}} =: P_{ij}$ defined by (cf. (2.5))

$$P_{ij}(k) = \frac{u_{ij}^k}{k!} \exp(-u_{ij}) , \qquad k \in \mathbb{N} \cup \{0\} .$$

In the case of Salt-and-Pepper noise, it is assumed that uniform bounds $c_{\min} \le u_{ij} \le c_{\max}$ of the data $\mathbf{u}$ are given. On each pixel x_{ij}, the noise process either sets the intensity u_{ij} to $c_{\min}$ or $c_{\max}$, or leaves the intensity unchanged. This can be modeled by considering $\delta_{ij}(\mathbf{u})$ a realization of the random variable P_{ij} with range $\{c_{\min}, u_{ij}, c_{\max}\}$ given by

$$P_{ij}(c_{\min}) = \lambda_1 , \qquad P_{ij}(u_{ij}) = \lambda_2 , \qquad P_{ij}(c_{\max}) = \lambda_3 ,$$

where $\lambda_i \ge 0$ satisfy $\lambda_1 + \lambda_2 + \lambda_3 = 1$. One application is the modeling of corrupt sensors that are either in an "always on" or "always off" state. In this case, $c_{\min} = 0$ represents black (off) pixels and $c_{\max} = 1$ white (on) pixels. For more details, we refer to [184, p. 316] or [92].

Sampling Errors

We consider the noise model

$$\boxed{\mathbf{u}^\delta = \mathbf{u} + \boldsymbol{\delta}\,|\mathbf{v}|\ ,} \tag{2.10}$$

where $|\mathbf{v}|$ is the matrix of the norms of the discrete gradients defined in (2.8) and $\boldsymbol{\delta}$ is an $(n_x - 1) \times (n_y - 1)$ matrix of realizations of i.i.d. Gaussian random variables Δ_{ij}. We assume that each Δ_{ij} has zero mean and standard deviation $\sigma_{\Delta_{ij}} := \sigma_\Delta > 0$. As in the case of multiplicative intensity errors, all operations in (2.10) are understood pointwise. For the sake of simplicity of presentation, we do not notationally distinguish between the $n_x \times n_y$ matrices $\mathbf{u}$ and $\mathbf{u}^{\boldsymbol{\delta}}$ on the one hand and the sub-matrices consisting of the first $(n_x - 1)$ columns and first $(n_y - 1)$ rows on the other hand.

The relevance of this noise model becomes evident from the following considerations: Let us assume that u_{ij}, $(i,j) \in \mathcal{I}_2$, are obtained by sampling a function $u \in C_0^2(\Omega)$ at sampling points $x_{ij} \in \Omega$, $(i,j) \in \mathcal{I}_2$. The following results state that the error model defined in (2.10) approximates an error model, where each sampling point is randomly shifted in direction of $\nabla u(x_{ij})$.

Theorem 2.8. *Let $h > 0$ fixed. Assume that $u \in C_0^2(\mathbb{R}^2)$ satisfies*

$$u_{ij} = u(x_{ij}) , \qquad (i,j) \in \mathcal{I}_2 .$$

Moreover, let

$$\boxed{x_{ij}^\delta := x_{ij} + \delta_{ij} n_{ij} , \qquad n_{ij} := \begin{cases} \dfrac{\nabla u(x_{ij})}{|\nabla u(x_{ij})|}, & \textit{if } \nabla u(x_{ij}) \neq 0 , \\ 0, & \textit{else} , \end{cases}}$$

that is, n_{ij} is orthogonal to the level line $\partial\,\mathrm{level}_{u_{ij}}(u)$ at x_{ij}. Then there exists a constant C only depending on u, such that

$$\frac{1}{|\mathcal{I}_2|}\sum_{(i,j)\in\mathcal{I}_2}\left|u(x_{ij}^\delta)-u_{ij}^\delta\right| \le \frac{C}{|\mathcal{I}_2|}\Big(h\sum_{(i,j)\in\mathcal{I}_2}|\delta_{ij}| + \sum_{(i,j)\in\mathcal{I}_2}\delta_{ij}^2\Big)\,. \tag{2.11}$$

Proof. Because $u(x_{ij}) = u_{ij}$, it follows that also $\nabla_h u(x_{ij}) = v_{ij}$. Because by assumption $u \in C_0^2(\mathbb{R}^2)$, Taylor's theorem shows that there exists $C_1 > 0$ only depending on $\left\|\nabla^2 u\right\|_\infty$, such that

$$\left|u(x_{ij}+\delta_{ij}n_{ij}) - u(x_{ij}) - \delta_{ij}\nabla u(x_{ij})\cdot n_{ij}\right| \le C_1\delta_{ij}^2\,, \qquad (i,j)\in\mathcal{I}_2\,. \tag{2.12}$$

Using (2.12) shows that

$$\begin{aligned}\left|u(x_{ij}^\delta)-u_{ij}^\delta\right| &= \left|u(x_{ij}+\delta_{ij}n_{ij}) - u(x_{ij}) - \delta_{ij}\left|\nabla_h u(x_{ij})\right|\right| \\ &\le \left|\delta_{ij}\nabla u(x_{ij})\cdot n_{ij} - \delta_{ij}\left|\nabla_h u(x_{ij})\right|\right| + C_1\delta_{ij}^2\,.\end{aligned} \tag{2.13}$$

Because $\nabla u(x_{ij})\cdot n_{ij} = |\nabla u(x_{ij})|$, it follows from (2.13) that

$$\begin{aligned}\left|u(x_{ij}^\delta)-u_{ij}^\delta\right| &\le |\delta_{ij}|\left||\nabla u(x_{ij})| - |\nabla_h u(x_{ij})|\right| + C_1\delta_{ij}^2 \\ &\le |\delta_{ij}|\left|\nabla u(x_{ij}) - \nabla_h u(x_{ij})\right| + C_1\delta_{ij}^2\,.\end{aligned} \tag{2.14}$$

Moreover, there exists $C_2 > 0$, again only depending on $\left\|\nabla^2 u\right\|_\infty$, such that

$$\left|\nabla u(x_{ij}) - \nabla_h u(x_{ij})\right| \le C_2 h\,, \qquad (i,j)\in\mathcal{I}_2\,. \tag{2.15}$$

Inserting (2.15) in (2.14), we derive

$$\frac{1}{|\mathcal{I}_2|}\sum_{(i,j)\in\mathcal{I}_2}\left|u(x_{ij}^\delta)-u_{ij}^\delta\right| \le \frac{C_2\,h}{|\mathcal{I}_2|}\sum_{(i,j)\in\mathcal{I}_2}|\delta_{ij}| + \frac{C_1}{|\mathcal{I}_2|}\sum_{(i,j)\in\mathcal{I}_2}\delta_{ij}^2\,,$$

which proves the assertion. □

Remark 2.9. We now study the influence of the mesh size h on the above defined sampling errors. To that end, we indicate the parameter h by a superscript in all occurring variables and sets.

Recall that the sample means of $\boldsymbol{\delta}^h$ and $\left|\boldsymbol{\delta}^h\right|$ and the sample variance of $\boldsymbol{\delta}^h$ are defined as (see Definition 2.4 and Remark 2.6)

$$\mathrm{E}(\boldsymbol{\delta}^h) = \frac{1}{\left|\mathcal{I}_2^h\right|}\sum_{(i,j)\in\mathcal{I}_2^h}\delta_{ij}^h\,, \qquad \mathrm{E}(|\boldsymbol{\delta}^h|) = \frac{1}{\left|\mathcal{I}_2^h\right|}\sum_{(i,j)\in\mathcal{I}_2^h}\left|\delta_{ij}^h\right|\,,$$

$$\mathrm{Var}(\boldsymbol{\delta}^h) = \frac{1}{\left|\mathcal{I}_2^h\right|}\sum_{(i,j)\in\mathcal{I}_2^h}\left(\delta_{ij}^h - \mathrm{E}(\boldsymbol{\delta}^h)\right)^2 = \frac{1}{\left|\mathcal{I}_2^h\right|}\sum_{(i,j)\in\mathcal{I}_2^h}(\delta_{ij}^h)^2 - \mathrm{E}(\boldsymbol{\delta}^h)^2\,.$$

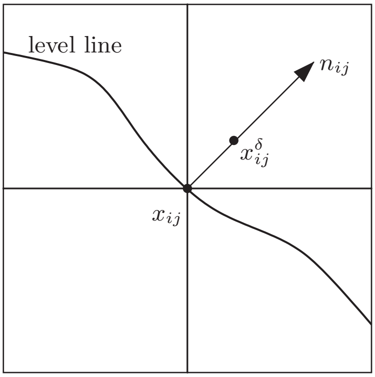

Fig. 2.2. Distortion of a sampling point in 2D. The shift is assumed to be orthogonal to the level line.

Inserting these definitions in the right-hand side of (2.11) yields

$$\frac{1}{\left|\mathcal{I}_2^h\right|} \sum_{(i,j)\in\mathcal{I}_2^h} \left|u(x_{ij}^{h,\delta}) - u_{ij}^{h,\delta}\right| \le C\big(h\ \mathrm{E}(|\boldsymbol{\delta}^h|) + \mathrm{E}(\boldsymbol{\delta}^h)^2 + \mathrm{Var}(\boldsymbol{\delta}^h)\big)\ .$$

For $h > 0$, denote by $P_{\boldsymbol{\Delta}^h}$ the distribution of the random vector $\boldsymbol{\Delta}^h$. The law of large numbers (see, e.g., [160, VII.7, Thm. 1]) implies that $\mathrm{E}(\boldsymbol{\delta}^h) \to 0$ in probability, that is,

$$\lim_{h\to 0} P_{\boldsymbol{\Delta}^h}\big(\big\{\left|\mathrm{E}(\boldsymbol{\delta}^h)\right| > \varepsilon\big\}\big) = 0\,, \qquad \varepsilon > 0\ .$$

Similarly, the law of large numbers implies that $\mathrm{E}(|\boldsymbol{\delta}^h|)$ converges in probability to a finite number, which implies that $h\,\mathrm{E}(|\boldsymbol{\delta}^h|) \to 0$. As a consequence, it follows from Theorem 2.8 that

$$\limsup_{h\to 0} \frac{1}{\left|\mathcal{I}_2^h\right|} \sum_{(i,j)\in\mathcal{I}_2^h} \left|u(x_{ij}^{h,\delta}) - u_{ij}^{h,\delta}\right| \le C\,\mathrm{Var}(\boldsymbol{\delta}^h) \quad \text{in probability}\,, \tag{2.16}$$

that is,

$$\lim_{h\to 0} P_{\boldsymbol{\Delta}^h}\Big(\Big\{\frac{1}{\left|\mathcal{I}_2^h\right|} \sum_{(i,j)\in\mathcal{I}_2^h} \left|u(x_{ij}^{h,\delta}) - u_{ij}^{h,\delta}\right| > C\,\mathrm{Var}(\boldsymbol{\delta}^h) + \varepsilon\Big\}\Big) = 0\,, \qquad \varepsilon > 0\ .$$

Using (2.16), it follows that the error model (2.10) for small variances approximately describes displacement errors of the sampling points in direction orthogonal to the level lines (compare Fig. 2.2). ◊

2.4 Priors for Images

In the following, we show how images themselves can be modeled as realizations of a random vector, the distribution of which is called *prior distribution* or *prior* (see [129, 231]). The method of MAP estimation, to be introduced in

Section 2.5, then provides a statistical motivation for variational methods for denoising. We attempt to use as simple as possible priors, and assume that either the intensities of the image or the discrete gradients are i.i.d. Below we show with three test examples that this assumption, though extremely simplifying, still provides enough information to be used in MAP estimation for efficient denoising.

In this book, we consider three digitized test images shown in Figs. 2.3, 2.5, and 2.7:

- a digital photo, which we refer to as the *mountain* image,
- a synthetic image, which we refer to as the *cards* image, and
- ultrasound data.

As additional test data, we use noisy variants of the mountain and cards images. We have artificially distorted the images by adding either *Gaussian intensity errors* or by simulating *sampling errors*.

The test data with Gaussian intensity errors are plotted in Figs. 2.9 and 2.11. The test data with sampling errors are shown in Figs. 2.10 and 2.12.

Histograms of the Intensities

Histograms are important for motivating variational regularization techniques. The *histogram* of an image is determined by partitioning $\mathbb{R}$ into congruent half-open sub-intervals of length $\Delta I > 0$,

$$I_k := \big[k\,\Delta I, (k+1)\Delta I\big)\,, \qquad k \in \mathbb{Z}\,,$$

and counting the occurrences of $\mathbf{u}$ in the sub-intervals, that is,

$$c_k := \left|\{(i,j) \in \mathcal{I}_1 : u_{ij} \in I_k\}\right| \,.$$

The histogram is represented as a probability density p on $\mathbb{R}$ that is constant on each interval I_k and there attains the value

$$\boxed{p|_{I_k} := \frac{c_k}{\Delta I\ |\mathcal{I}_1|}\,, \qquad k \in \mathbb{Z}\,.}$$

Comparing the histograms of the intensities of the test images with the corresponding histograms of the distorted images reveals that, by adding Gaussian noise to an image, the histogram of the intensities becomes smoother (compare the histograms of Figs. 2.4 and 2.6).

The ultrasound image in Fig. 2.7 contains speckle noise. Because no noise-free version is available, we compare the original data with a filtered version of the image (see Fig. 2.8). For filtering, the total variation regularization method discussed in Chapter 4 is used. Again, the histogram of the noisy data is smoother than that of the filtered image.

Fig. 2.3. Mountain image.

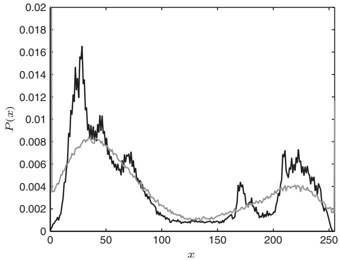

Fig. 2.4. Histogram of mountain image (*black line*) and histogram of the image distorted with Gaussian noise (*gray line*).

Fig. 2.5. Cards image.

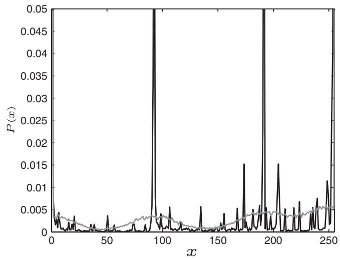

Fig. 2.6. Histogram of cards image (*black line*) and histogram of the image distorted with Gaussian noise (*gray line*).

Fig. 2.7. Ultrasound data.

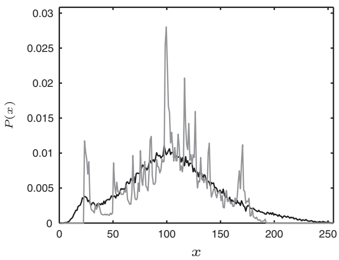

Fig. 2.8. Histogram of original ultrasound data (*black line*) and of filtered data (*gray line*).

Fig. 2.9. Mountain image distorted by additive Gaussian noise.

Fig. 2.10. Mountain image distorted by sampling point errors.

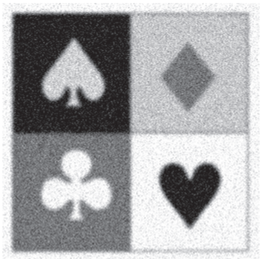

Fig. 2.11. Cards image distorted by additive Gaussian noise.

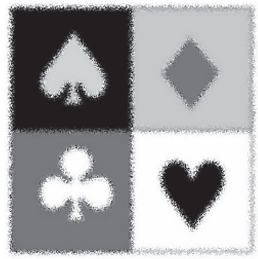

Fig. 2.12. Cards image distorted by sampling point errors.

The above examples show that the intensity histograms of images strongly depend on the image content. Therefore it is difficult to provide an *a priori* probability density $p(\mathbf{u})$ that approximates the histograms of a variety of different images.

Histograms of the Discrete Gradients

In image processing, commonly the histograms of *norms of the discrete gradients* of intensities are preferred to intensity histograms. Figures 2.14, 2.16, and 2.18 show the histograms of $|\mathbf{v}|$ for our test images. It can be recognized that the histograms are pronounced at around 0 and look very similar to the probability distributions considered above. In Figs. 2.13, 2.15, and 2.17, the histograms for the distorted and the original test images are compared to highlight the differences.

For both the card and the mountain image without distortions, the histograms of the discrete gradients are concentrated around zero, indicating that the images have dominant flat regions. For the data distorted with Gaussian noise, the histogram is significantly flatter. Distortions of sampling points strongly change the histogram in the mountain image but not in the cards image. This is due to the fact that the cards image consists of piecewise constant parts, in which sampling errors have no effect.

Distribution of $|\mathbf{v}|$ and fitted Gaussian and Laplacian distribution

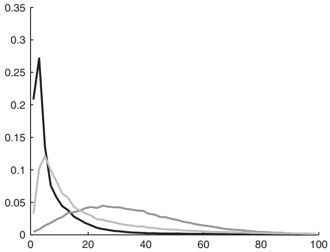

Fig. 2.13. Empirical distribution of the discrete gradient: mountain image (*black line*), distorted by Gaussian noise (*dark gray line*) and distorted by sampling errors (*light gray line*).

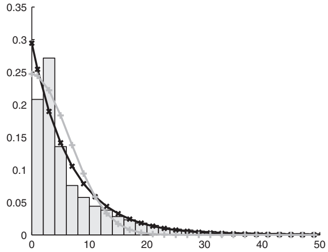

Fig. 2.14. Histogram of $|\mathbf{v}|$ (*bar plot*) for the mountain image and fitted Laplacian (*black line*) and Gaussian (*gray line*) distribution.

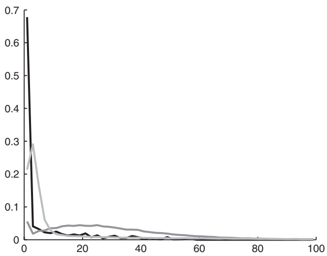

Fig. 2.15. Empirical density of $|\mathbf{v}|$ for the cards image (*black line*), distorted by Gaussian noise (*dark gray line*) and distorted by sampling errors (*light gray line*).

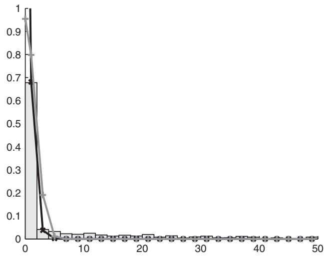

Fig. 2.16. Histogram of $|\mathbf{v}|$ (*bar plot*) for the cards image and fitted Laplacian (*black line*) and Gaussian (*gray line*) distribution.

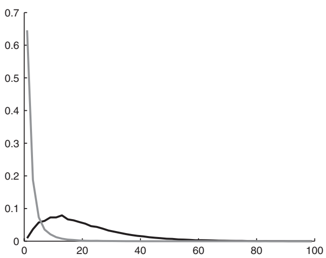

Fig. 2.17. Histogram of $|\mathbf{v}|$ for the ultrasound (*black line*) and filtered ultrasound data (*gray line*).

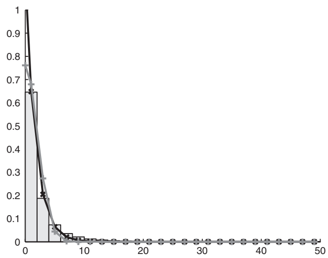

Fig. 2.18. Histogram of $|\mathbf{v}|$ for the filtered ultrasound data and fitted Laplacian (*black line*) and Gaussian (*gray line*) distribution.

Table 2.1. Optimal approximation (w.r.t. l^2-error) by Gaussian and Laplacian probability densities to the histograms of the absolute value of discrete gradients of the images.

Test image	l^2-error – Gauss	l^2-error – Laplace
Mountain	3.13×10^{-3}	2.61×10^{-3}
Cards	10.25×10^{-3}	1.14×10^{-3}

In order to derive image priors, we compare the histograms of $|\mathbf{v}|$ with an appropriate subset of well-established continuous probability density functions supported in $[0, \infty)$. For a continuous density function $\tilde{p}$ we use the approximation

$$\tilde{P}|_{I_k} := \frac{1}{|I_k|} \int_{I_k} \tilde{p}(s) \approx \tilde{p}(k)\,, \qquad k \in \mathbb{Z}\,,$$

and minimize the l^2-error between the histogram and the vector $\big(\tilde{p}(k)\big)$.

In the following, we denote by $\mathbf{U}$ a random vector and by $p_{\mathbf{U}}$ the probability density of $\mathbf{U}$. The image $\mathbf{u}$ is considered as a realization of $\mathbf{U}$.

We now assume that the probability density $p_{\mathbf{U}}(\mathbf{u})$ only depends on the matrix $|\mathbf{v}|$ of the norms of the discrete gradients $\mathbf{v}$ of $\mathbf{u}$. Additionally, we assume that the norms of the discrete gradients are i.i.d. In this case, the probability density of $\mathbf{U}$ is the product of the densities of $|v_{ij}|$.

A typical assumption on the absolute values of the discrete gradients is that they are Gaussian distributed, in which case the prior is

$$\boxed{p_{\mathbf{U}}(\mathbf{u}) := C \exp\left(-\frac{1}{2\sigma_2^2} \sum_{(i,j)\in\mathcal{I}_2} |v_{ij}|^2\right)\,,} \tag{2.17}$$

or that they are Laplacian distributed (see [39]), in which case the prior is

$$\boxed{p_{\mathbf{U}}(\mathbf{u}) := C \exp\left(-\frac{1}{\sigma_1} \sum_{(i,j)\in\mathcal{I}_2} |v_{ij}|\right)\,.}$$

We refer to these priors as the *Gaussian prior* and the *Laplacian prior*, respectively.

Example 2.10. We determine the best approximation of discrete gradients of the cards and mountain histogram, respectively, within the set of Laplacian and Gaussian densities. To that end, we have to determine the parameters $\sigma_q > 0$, $q \in \{1, 2\}$, in such a way that the density p as introduced in (2.6) and

(2.7), respectively, optimally fits the histogram. In Figs. 2.14 and 2.16, we have plotted the optimal Laplacian density ($q = 1$) and the optimal Gaussian density ($q = 2$). Table 2.1 shows that the histogram can be better approximated within the set of Laplacian distributions than within the set of Gaussian distributions. $\diamondsuit$

In the case of the mountain image, one can see that the histogram of the discrete gradients attains its maximum away from zero (see Fig. 2.14). The reason is that natural images often include regions containing texture, where small oscillations cause a non-vanishing discrete gradient. The Gaussian and Laplacian prior, however, both attain their maximum at zero. In order to mirror this situation, we introduce a new density, in the following referred to as *log-prior* (see Fig. 2.19),

$$\boxed{p_{\mathbf{U}}(\mathbf{u}) := C \exp\left(\sum_{(i,j)\in\mathcal{I}_2} -\frac{|v_{ij}|^q}{q\,\sigma_3^q} + \log|v_{ij}|\right),}$$

where $C > 0$ is a normalizing constant, and $q = 1$ or $q = 2$.

We motivate the log-prior as follows: Let $v \in \mathbb{R}^2$ be a realization of a two-dimensional random vector V, which is Gaussian or Laplacian distributed, that is, it has a probability density of the form

$$p_V(v) = C \exp\left(-\frac{|v|^q}{q\,\sigma_3^q}\right),$$

where $\sigma_3 > 0$, $q \in \{1, 2\}$, and $C := \left(\int_{\mathbb{R}^2} \exp\left(-|\tilde{v}|^q / q\sigma_3^q\right)\right)^{-1}$. We are interested in the distribution of $|V|$, and therefore we consider its probability density $p_{|V|}$. Using Lemma 2.3 with $f(v) = |v|$, which implies that the Jacobian of f (see (2.1)) satisfies $J_f = 1$ almost everywhere, we find that

$$p_{|V|}(s) = C \int_{|\tilde{v}|=s} \exp\left(-\frac{|\tilde{v}|^q}{q\sigma_3^q}\right) \mathrm{d}\mathcal{H}^1, \qquad s \geq 0, \tag{2.18}$$

where $\mathcal{H}^1$ is the one-dimensional Hausdorff measure (see (9.1)). Because the integrand in (2.18) is constant on $\{|\tilde{v}| = s\}$, it follows from the fact that $\mathcal{H}^1(\{|\tilde{v}| = s\}) = 2\pi s$ that

$$p_{|V|}(s) = 2\pi s\, C \exp\left(-\frac{s^q}{q\sigma_3^q}\right), \qquad s \geq 0, \tag{2.19}$$

the maximum of which is attained for $s = \sigma_3$. Figure 2.19 shows the graphs of the probability density (2.19) for $q = 1$ and $q = 2$. For $q = 2$, the function $p_{|V|}$ is the density of the *Rayleigh distribution* (see, for example, [388]).

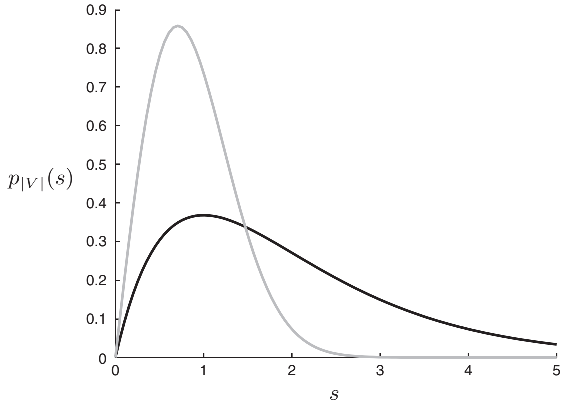

Fig. 2.19. Probability density $p_{|V|}(s) = Cs \exp(-s^q/q\sigma_3^q)$ with $\sigma_3 = 1$, $C = \left(\int_0^\infty s \exp(-s^q/q\sigma_3^q)\right)^{-1}$ for $q = 1$ (*black line*) and $q = 2$ (*gray line*).

2.5 Maximum A Posteriori Estimation

We consider the following situation:

Let $\tilde{\mathbf{U}} = (\mathbf{U}, \mathbf{U}^\delta)$ be an $(n+m)$-dimensional random vector. The probability distribution of $\tilde{\mathbf{U}}$ is just the joint probability distribution of $\mathbf{U}$ and $\mathbf{U}^\delta$, denoted by $P_{\mathbf{U},\mathbf{U}^\delta}$.

Moreover, let $\mathbf{u}^\delta$ be a realization of the m-dimensional random vector $\mathbf{U}^\delta$. We want to find a realization $\mathbf{u}^0$ of $\mathbf{U}$ that makes the pair $(\mathbf{u}, \mathbf{u}^\delta)$ most likely. Typically, $\mathbf{u}^\delta$ is interpreted as noisy data, which are formed from the clean data by means of a known noise process.

If $\mathbf{U}$ is a discrete random vector, the task of reconstructing $\mathbf{u}^0$ is comparatively easy. The most likely realization $\mathbf{u}^0$ is the one that, for fixed $\mathbf{u}^\delta$, maximizes the joint probability $P_{\mathbf{U},\mathbf{U}^\delta}(\cdot, \mathbf{u}^\delta)$. In order to make the definition suited for generalization to the non-discrete case, we define maximum a posteriori estimation for discrete random vectors by means of conditional probabilities:

Definition 2.11. *Let* $\mathbf{U}$ *and* $\mathbf{U}^\delta$ *be discrete random vectors. The* conditional probability *of* $\mathbf{u}$ *for given realization* $\mathbf{u}^\delta$ *of* $\mathbf{U}^\delta$ *is defined by*

$$\boxed{P_{\mathbf{U}|\mathbf{U}^\delta}(\mathbf{u}|\mathbf{u}^\delta) := \begin{cases} \dfrac{P_{\mathbf{U},\mathbf{U}^\delta}(\mathbf{u}, \mathbf{u}^\delta)}{P_{\mathbf{U}^\delta}(\mathbf{u}^\delta)}, & \text{if } P_{\mathbf{U}^\delta}(\mathbf{u}^\delta) > 0\,, \\ 0\,, & \text{if } P_{\mathbf{U}^\delta}(\mathbf{u}^\delta) = 0\,. \end{cases}} \tag{2.20}$$

The mapping

$$\boxed{\mathbf{u}^\delta \mapsto \mathbf{u}^0 := \arg\max_{\mathbf{u}} P_{\mathbf{U}|\mathbf{U}^\delta}(\mathbf{u}|\mathbf{u}^\delta)}$$

is called maximum a posteriori estimator*, in short MAP estimator, and the function* $\mathbf{u}^0$ *is called* MAP estimate *(see [383, 391]).*

Example 2.12. We apply MAP estimation to a simple example: Let U and Δ be two independent discrete random variables with values in $I_1 := \{1,2,3\}$ and $I_2 := \mathbb{Z}$, respectively. We assume that the corresponding probability distributions are defined by

$$P_U(u) = \frac{1}{3} \qquad \text{and} \qquad P_\Delta(\delta) = \begin{cases} 0.4 & \text{if } \delta = 0\,, \\ 0.24 & \text{if } |\delta| = 1\,, \\ 0.055 & \text{if } |\delta| = 2\,, \\ 0.005 & \text{if } |\delta| = 3\,, \\ 0 & \text{else}\,. \end{cases}$$

Let $U^\delta = U + \Delta$. Then

$$\begin{aligned} P_{U^\delta}(u^\delta) &= \sum_{u\in I_1} P_{U,\Delta}(u, u^\delta - u) = \sum_{u\in I_1} P_U(u)\, P_\Delta(u^\delta - u) \\ &= \frac{1}{3}\sum_{u\in I_1} P_\Delta(u^\delta - u) = \begin{cases} 0.002 & \text{if } u^\delta = -2 \text{ or } 6\,, \\ 0.02 & \text{if } u^\delta = -1 \text{ or } 5\,, \\ 0.1 & \text{if } u^\delta = 0 \text{ or } 4\,, \\ 0.232 & \text{if } u^\delta = 1 \text{ or } 3\,, \\ 0.293 & \text{if } u^\delta = 2\,, \\ 0 & \text{else}\,. \end{cases} \end{aligned}$$

For $u^\delta \in \{-2,\dots,6\}$, the probabilities $P_{U,U^\delta}(u,u^\delta)$ and $P_{U|U^\delta}(u|u^\delta)$ can be read from the following tables, for $u^\delta \notin \{-2,\dots 6\}$ we have $P_{U,U^\delta}(u,u^\delta) = P_{U|U^\delta}(u|u^\delta) = 0$ for every u.

u^δ

P_{U,U^δ}	-2	-1	0	1	2	3	4	5	6
u=1	0.002	0.018	0.080	0.133	0.080	0.018	0.002	0.000	0.000
u=2	0.000	0.002	0.018	0.080	0.133	0.080	0.018	0.002	0.000
u=3	0.000	0.000	0.002	0.018	0.080	0.133	0.080	0.018	0.002

$P_{U\|U^\delta}$	-2	-1	0	1	2	3	4	5	6
u=1	1.0	0.917	0.800	0.576	0.273	0.079	0.017	0.000	0.0
u=2	0.0	0.083	0.183	0.345	0.455	0.345	0.183	0.083	0.0
u=3	0.0	0.000	0.017	0.079	0.273	0.576	0.800	0.917	1.0

(Note that these values have been rounded.)

For given u^δ, we can determine from $P_{U|U^\delta}$ the most probable value $u \in \{1,2,3\}$. For example, the probability $P_{U|U^\delta}$ for the value of $U^\delta = 0$ attains the maximum at $U = 1$. $\diamondsuit$

In the following, we study the problem of MAP estimation for absolutely continuous distributions. The argumentation follows [321, pp. 98–99]. We assume that the random vectors $\mathbf{U}$, $\mathbf{U}^\delta$, and $\tilde{\mathbf{U}} = (\mathbf{U}, \mathbf{U}^\delta)$ have absolutely continuous probability distributions $P_{\mathbf{U}}$, $P_{\mathbf{U}^\delta}$, and $P_{\mathbf{U},\mathbf{U}^\delta}$ with according densities $p_{\mathbf{U}}$, $p_{\mathbf{U}^\delta}$, and $p_{\mathbf{U},\mathbf{U}^\delta}$.

Analogously to (2.20), we define the conditional probability of a measurable set $A \subset \mathbb{R}^n$ for given measurable $B \subset \mathbb{R}^m$ by

$$P_{\mathbf{U}|\mathbf{U}^\delta}(A|B) := \begin{cases} \dfrac{P_{\mathbf{U},\mathbf{U}^\delta}(A,B)}{P_{\mathbf{U}^\delta}(B)}, & \text{if } P_{\mathbf{U}^\delta}(B) > 0\,, \\ 0, & \text{if } P_{\mathbf{U}^\delta}(B) = 0\,. \end{cases}$$

Now let $\mathbf{u}^\delta$ be a realization of $\mathbf{U}^\delta$. We define the *conditional density* $p_{\mathbf{U}|\mathbf{U}^\delta}$ of $\mathbf{u} \in \mathbb{R}^n$ given $\mathbf{u}^\delta$ by

$$p_{\mathbf{U}|\mathbf{U}^\delta}(\mathbf{u}|\mathbf{u}^\delta) := \begin{cases} \dfrac{p_{\mathbf{U},\mathbf{U}^\delta}(\mathbf{u},\mathbf{u}^\delta)}{p_{\mathbf{U}^\delta}(\mathbf{u}^\delta)}, & \text{if } p_{\mathbf{U}^\delta}(\mathbf{u}^\delta) > 0\,, \\ 0, & \text{if } p_{\mathbf{U}^\delta}(\mathbf{u}^\delta) = 0\,. \end{cases} \tag{2.21}$$

The next result reveals the connection between conditional density and conditional probability:

Theorem 2.13. *Let* $\mathbf{u}$ *and* $\mathbf{u}^\delta$ *be realizations of the random vectors* $\mathbf{U}$ *and* $\mathbf{U}^\delta$*, respectively. Assume that the densities* $p_{\mathbf{U}^\delta}$ *and* $p_{\mathbf{U},\mathbf{U}^\delta}$ *are continuous, and* $p_{\mathbf{U}^\delta}(\mathbf{u}^\delta) > 0$.

For $\rho > 0$*, let* $\mathcal{U}_\rho(\mathbf{u})$ *and* $\mathcal{U}_\rho(\mathbf{u}^\delta)$ *denote the open cubes with side length* 2ρ *around* $\mathbf{u}$ *and* $\mathbf{u}^\delta$*,*

$$\begin{aligned} \mathcal{U}_\rho(\mathbf{u}) &:= (u_1 - \rho, u_1 + \rho) \times \cdots \times (u_n - \rho, u_n + \rho)\,, \\ \mathcal{U}_\rho(\mathbf{u}^\delta) &:= (u_1^\delta - \rho, u_1^\delta + \rho) \times \cdots \times (u_m^\delta - \rho, u_m^\delta + \rho)\,. \end{aligned}$$

Then

$$p_{\mathbf{U}|\mathbf{U}^\delta}(\mathbf{u}|\mathbf{u}^\delta) = \lim_{\rho \to 0} 2^{-n} \rho^{-n}\, P_{\mathbf{U}|\mathbf{U}^\delta}\big(\mathcal{U}_\rho(\mathbf{u}) | \mathcal{U}_\rho(\mathbf{u}^\delta)\big)\,.$$

Proof. Because the probability densities $p_{\mathbf{U}^\delta}$ and $p_{\mathbf{U},\mathbf{U}^\delta}$ are continuous, it follows from the mean value theorem for integration that

$$p_{\mathbf{U}^\delta}(\mathbf{u}^\delta) = \lim_{\rho \to 0} \frac{1}{2^m \rho^m} \int_{\mathcal{U}_\rho(\mathbf{u}^\delta)} p_{\mathbf{U}^\delta}(\mathbf{u}^\delta) = \lim_{\rho \to 0} \frac{P_{\mathbf{U}^\delta}\big(\mathcal{U}_\rho(\mathbf{u}^\delta)\big)}{2^m \rho^m}\,, \tag{2.22}$$

$$p_{\mathbf{U},\mathbf{U}^\delta}(\mathbf{u},\mathbf{u}^\delta) = \lim_{\rho \to 0} \frac{P_{\mathbf{U},\mathbf{U}^\delta}\big(\mathcal{U}_\rho(\mathbf{u}) \times \mathcal{U}_\rho(\mathbf{u}^\delta)\big)}{2^{n+m} \rho^{n+m}}\,. \tag{2.23}$$

Thus the assertion follows from the definitions of conditional probability in (2.20) and conditional density in (2.21). □

Note that (2.22) and (2.23) are simple versions of the Lebesgue–Besicovitch differentiation theorem (see, e.g., [157, Sect. 1.7] for a formulation with balls instead of cubes), which also applies to discontinuous densities, in which case (2.22) and (2.23) only hold almost everywhere.

As a consequence of Theorem 2.13, maximization of $p_{\mathbf{U}|\mathbf{U}^\delta}(\cdot|\mathbf{u}^\delta)$ can be considered as continuous analogue to discrete MAP estimation.

In many applications, the vector $\mathbf{u}^\delta$ is considered a noisy perturbation of some unknown data $\mathbf{u}$. The noise process that generates $\mathbf{u}^\delta$ is described by the conditional density $p_{\mathbf{U}^\delta|\mathbf{U}}(\mathbf{u}^\delta|\mathbf{u})$ of $\mathbf{u}^\delta$ given $\mathbf{u}$. Thus we have to find a way that links the two conditional densities $p_{\mathbf{U}|\mathbf{U}^\delta}(\mathbf{u}|\mathbf{u}^\delta)$ and $p_{\mathbf{U}^\delta|\mathbf{U}}(\mathbf{u}^\delta|\mathbf{u})$. This is achieved by means of the *formula of Bayes* (see, for instance, [129]),

$$\boxed{p_{\mathbf{U}|\mathbf{U}^\delta}(\mathbf{u}|\mathbf{u}^\delta) = \begin{cases} \dfrac{p_{\mathbf{U}^\delta|\mathbf{U}}(\mathbf{u}^\delta|\mathbf{u})\, p_{\mathbf{U}}(\mathbf{u})}{p_{\mathbf{U}^\delta}(\mathbf{u}^\delta)}, & \text{if } p_{\mathbf{U}^\delta}(\mathbf{u}^\delta) > 0\,, \\ 0\,, & \text{if } p_{\mathbf{U}^\delta}(\mathbf{u}^\delta) = 0\,. \end{cases}}$$

Therefore, we call *continuous MAP estimation* the problem of maximizing the functional

$$\mathcal{T}^{\mathrm{MAP}}(\mathbf{u}) = \frac{p_{\mathbf{U}^\delta|\mathbf{U}}(\mathbf{u}^\delta|\mathbf{u})\, p_{\mathbf{U}}(\mathbf{u})}{p_{\mathbf{U}^\delta}(\mathbf{u}^\delta)}\,. \tag{2.24}$$

Note that in (2.24), the constant factor $p_{\mathbf{U}^\delta}(\mathbf{u}^\delta)$ can be omitted without affecting the maximization problem. A maximizer of (2.24) is called *MAP estimate.*

To simplify the maximization, the logarithmic MAP estimator

$$\mathcal{T}^{\mathrm{logMAP}}(\mathbf{u}) := -\log p_{\mathbf{U}^\delta|\mathbf{U}}(\mathbf{u}^\delta|\mathbf{u}) - \log p_{\mathbf{U}}(\mathbf{u}) \tag{2.25}$$

is often used in applications. Because the logarithm is a strictly increasing function, the transformation does not change the extrema. The problem of minimization of $\mathcal{T}^{\mathrm{logMAP}}$ is referred to as log MAP estimation.

2.6 MAP Estimation for Noisy Images

We now show how the method of MAP estimation can be applied to image denoising and analysis.

We always assume that we are given a noisy image $\mathbf{u}^\delta$ that is a distortion of the clean image by one of the noise processes introduced in Section 2.3. Moreover, we denote by $\mathbf{U}$ a random variable associated with one of the image priors introduced in Section 2.4. In addition, $\mathbf{u}$ denotes a realization of $\mathbf{U}$.

Intensity Errors

We first assume additive Gaussian intensity errors on the image. In this case, the data $\mathbf{u}^\delta$ are given as (see (2.9))

$$\mathbf{u}^\delta = \mathbf{u} + \boldsymbol{\delta}\,,$$

where $\boldsymbol{\delta}$ is a realization of the random vector $\boldsymbol{\Delta} = (\Delta_{ij})$, $(i,j) \in \mathcal{I}_2$, where Δ_{ij} are i.i.d. Gaussian random variables with zero mean and variance σ^2. For fixed $\mathbf{u}$, the random vector $\mathbf{U}^\delta$ is given by

$$\mathbf{U}^\delta = \mathbf{u} + \boldsymbol{\Delta}\,.$$

We immediately see that U_{ij}^δ for given $\mathbf{u}$ are independently Gaussian distributed with mean u_{ij} and variance σ^2. Thus the conditional probability density $p(\mathbf{u}^\delta|\mathbf{u}) := p_{\mathbf{U}^\delta|\mathbf{U}}(\mathbf{u}^\delta|\mathbf{u})$ is given by

$$p(\mathbf{u}^\delta|\mathbf{u}) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{|\mathcal{I}_2|} \prod_{(i,j)\in\mathcal{I}_2} \exp\left(-\frac{(u_{ij}^\delta - u_{ij})^2}{2\sigma^2}\right)\,. \tag{2.26}$$

For simplicity of presentation, we now omit the subscripts of the probability densities $p_{\mathbf{U}^\delta|\mathbf{U}}(\mathbf{u}^\delta|\mathbf{u})$ and $p_{\mathbf{U}}(\mathbf{u})$, which can always be identified from the context. From (2.26), it follows that

$$-\log p(\mathbf{u}^\delta|\mathbf{u}) = |\mathcal{I}_2| \log\bigl(\sigma\sqrt{2\pi}\bigr) + \sum_{(i,j)\in\mathcal{I}_2} \frac{(u_{ij}^\delta - u_{ij})^2}{2\sigma^2}\,.$$

The goal of maximum a posteriori estimators (see also (2.24)) is to determine $\mathbf{u}$ by maximizing the product of the conditional probability density $p(\mathbf{u}^\delta|\mathbf{u})$ and the probability density of $\mathbf{u}$, which is given by its image prior $p(\mathbf{u})$. Maximizing the conditional probability density is equivalent to minimization of the negative logarithm of the conditional probability density.

Assuming a Gaussian prior (2.17), the second term in (2.25) reads as

$$-\log p(\mathbf{u}) = \sum_{(i,j)\in\mathcal{I}_2} \frac{1}{2\sigma_2^2}\,|v_{ij}|^2 + C\,.$$

Thus, the *log MAP estimator* for denoising images with intensity errors and Gaussian prior consists in minimization of the functional

$$\boxed{\operatorname*{arg\,min}_{\mathbf{u}\in\mathbb{R}^{\mathcal{I}_1}} \sum_{(i,j)\in\mathcal{I}_2} \Bigl((u_{ij} - u_{ij}^\delta)^2 + \alpha\,|v_{ij}|^2\Bigr)\,,}$$

where $\alpha := \sigma^2/\sigma_2^2 > 0$.

Sampling Errors

As above, we now determine MAP estimators for the model of sampling errors, where the noise model is given by (2.10).

Again we assume that $\boldsymbol{\delta}$ is a realization of a random vector $\boldsymbol{\Delta} = (\Delta_{ij})$, $(i,j) \in \mathcal{I}_2$, consisting of i.i.d. Gaussian random variables Δ_{ij} all having zero mean and variance σ^2. Let $\mathbf{u}$ be fixed (and therefore also $\mathbf{v}$), then it follows that the random variables

$$U_{ij}^\delta = u_{ij} + |v_{ij}|\, \Delta_{ij}\,, \qquad (i,j) \in \mathcal{I}_2\,, \tag{2.27}$$

are independent.

Assuming that $|v_{ij}| > 0$ for all $(i,j) \in \mathcal{I}_2$, it follows from (2.27) by using Lemma 2.3 with $f(x) = u_{ij} + |v_{ij}|\, x$ (and therefore $J_f = |v_{ij}|$) that

$$\begin{aligned} p(u_{ij}^\delta|\mathbf{u}) &= \int_{\left\{\delta_{ij} = \frac{u_{ij}^\delta - u_{ij}}{|v_{ij}|}\right\}} \frac{1}{|v_{ij}|}\, p_{\Delta_{ij}}(\delta_{ij})\, \mathrm{d}\mathcal{H}^0 \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right) \frac{1}{|v_{ij}|} \exp\left(-\frac{(u_{ij} - u_{ij}^\delta)^2}{2\sigma^2\, |v_{ij}|^2}\right)\,. \end{aligned} \tag{2.28}$$

Because U_{ij}^δ, $(i,j) \in \mathcal{I}_2$, are independent, we have that

$$p(\mathbf{u}^\delta|\mathbf{u}) = \prod_{(i,j)\in\mathcal{I}_2} p(u_{ij}^\delta|\mathbf{u})\,. \tag{2.29}$$

Inserting (2.28) into (2.29), it follows that

$$p(\mathbf{u}^\delta|\mathbf{u}) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{|\mathcal{I}_2|} \prod_{(i,j)\in\mathcal{I}_2} \frac{1}{|v_{ij}|} \exp\left(-\frac{(u_{ij} - u_{ij}^\delta)^2}{2\sigma^2\, |v_{ij}|^2}\right)\,. \tag{2.30}$$

As an example, the log MAP estimator, defined in (2.25), according to the conditional probability density (2.30) and the log-prior (2.19) is given by

$$\operatorname*{arg\,min}_{\mathbf{u}\in\mathbb{R}^{\mathcal{I}_1}} \sum_{(i,j)\in\mathcal{I}_2} \left(\frac{1}{2} \frac{(u_{ij} - u_{ij}^\delta)^2}{|v_{ij}|^2} + \frac{\alpha}{q} |v_{ij}|^q\right)\,, \qquad q = 1, 2\,. \tag{2.31}$$

Here $\alpha := \sigma^2/\sigma_3^q > 0$.

It is convenient for this book to study (2.31) in a more general setting. We consider

$$\operatorname*{arg\,min}_{\mathbf{u}\in\mathbb{R}^{\mathcal{I}_1}} \sum_{(i,j)\in\mathcal{I}_2} \left(\frac{1}{p} \frac{(u_{ij} - u_{ij}^\delta)^p}{|v_{ij}|^q} + \frac{\alpha}{r} |v_{ij}|^r\right) \tag{2.32}$$

with $p > 1$ and $r \geq 1$, and $q \geq 0$. In Chapters 4 and 5, we investigate continuous formulations

$$\boxed{\operatorname*{arg\,min}_{u\in X} \left(\frac{1}{p} \int_\Omega \frac{(u - u^\delta)^p}{|\nabla u|^q} + \frac{\alpha}{r} \int_\Omega |\nabla u|^r\right)} \tag{2.33}$$

of the discrete variational problem defined in (2.32), where X is an appropriate space of functions $u : \Omega \to \mathbb{R}$.

Further Reading

Background on statistical modeling of MAP estimators can be found for instance in [129, 321].

The standard reference for statistical approaches in inverse problems is [231]. Computational methods for statistical inverse problems are discussed in [378].

The relation between variational methods and MAP estimation is discussed in [39, 79, 186, 201, 202, 296]. An early reference on the topic on MAP estimators in imaging is [179].

http://www.springer.com/978-0-387-30931-6

Variational Methods in Imaging
Scherzer, O.; Grasmair, M.; Grossauer, H.; Haltmeier, M.; Lenzen, F.
2009, XIV, 320 p., Hardcover
ISBN: 978-0-387-30931-6