4

Statistical inversion theory

The majority of retrieval approaches currently used in atmospheric remote sensing belong to the category of statistical inversion methods (Rodgers, 2000). The goal of this chapter is to reveal the similarity between classical regularization and statistical inversion regarding

1. the regularized solution representation,
2. the error analysis,
3. the design of one- and multi-parameter regularization methods.

In statistical inversion theory all variables included in the model are absolutely continuous random variables and the degree of information concerning their realizations is coded in probability densities. The solution of the inverse problem is the a posteriori density, which makes possible to compute estimates of the unknown atmospheric profile.

In the framework of Tikhonov regularization we have considered the linear data model

$$y^{\delta} = Kx + \delta,$$  \hspace{1cm} (4.1)

where $y^{\delta}$ is the noisy data vector and $\delta$ is the noise vector. In statistical inversion theory all parameters are viewed as random variables, and since in statistics random variables are denoted by capital letters and their realizations by lowercase letters, the stochastic version of the data model (4.1) is

$$Y^{\delta} = KX + \Delta.$$  \hspace{1cm} (4.2)

The random vectors $Y^{\delta}$, $X$ and $\Delta$ represent the data, the state and the noise, respectively; their realizations are denoted by $Y^{\delta} = y^{\delta}$, $X = x$ and $\Delta = \delta$, respectively.

4.1 Bayes theorem and estimators

The data model (4.2) gives a relation between the three random vectors $Y^{\delta}$, $X$ and $\Delta$, and therefore, their probability densities depend on each other. The following probability densities are relevant for our analysis:
(1) the a priori density \( p_a(x) \), which encapsulates our presumable information about \( X \) before performing the measurement of \( Y^\delta \);

(2) the likelihood density \( p(y^\delta | x) \), which represents the conditional probability density of \( Y^\delta \) given the state \( X = x \);

(3) the a posteriori density \( p(x | y^\delta) \), which represents the conditional probability density of \( X \) given the data \( Y^\delta = y^\delta \).

The choice of the a priori density \( p_a(x) \) is perhaps the most important part of the inversion process. Different a priori models yield different objective functions, and in particular, the classical regularization terms correspond to Gaussian a priori models. Gaussian densities are widely used in statistical inversion theory because they are easy to compute and often lead to explicit estimators. Besides Gaussian densities other types of a priori models, as for instance the Cauchy density and the entropy density can be found in the literature (Kaipio and Somersalo, 2005).

The construction of the likelihood density \( p(y^\delta | x) \) depends on the noise assumption. The data model (4.2) operates with additive noise, but other explicit noise models including multiplicative noise models and models with an incompletely known forward model matrix can be considered. If the noise is additive and is independent of the atmospheric state, the probability density \( p_n(\delta) \) of \( \Delta \) remains unchanged when conditioned on \( X = x \). Thus, \( Y^\delta \) conditioned on \( X = x \) is distributed like \( \Delta \), and the likelihood density becomes

\[
p(y^\delta | x) = p_n(y^\delta - Kx).
\]

Assuming that after analyzing the measurement setting and accounting of the additional information available about all variables we have found the joint probability density \( p(x, y^\delta) \) of \( X \) and \( Y^\delta \), then the a priori density is given by

\[
p_a(x) = \int_{\mathbb{R}^m} p(x, y^\delta) \, dy^\delta,
\]

while the likelihood density and the a posteriori density can be expressed as

\[
p(y^\delta | x) = \frac{p(x, y^\delta)}{p_a(x)}, \quad (4.4)
\]

and

\[
p(x | y^\delta) = \frac{p(x, y^\delta)}{p(y^\delta)}, \quad (4.5)
\]

respectively.

The following result known as the Bayes theorem of inverse problems relates the a posteriori density to the likelihood density (cf. (4.4) and (4.5)):

\[
p(x | y^\delta) = \frac{p(y^\delta | x) \, p_a(x)}{p(y^\delta)}. \quad (4.6)
\]

In (4.6), the marginal density \( p(y^\delta) \) computed as

\[
p(y^\delta) = \int_{\mathbb{R}^n} p(x, y^\delta) \, dx = \int_{\mathbb{R}^n} p(y^\delta | x) \, p_a(x) \, dx,
\]