Bayesian inference for Gaussian models: Inverse problems and evolution equations

Yan, D.

Citation

Yan, D. (2020, March 3). *Bayesian inference for Gaussian models: Inverse problems and evolution equations*. Retrieved from https://hdl.handle.net/1887/86070

Author: Yan, D.
Title: Bayesian inference for Gaussian models: Inverse problems and evolution equations
Issue Date: 2020-03-03

Chapter 5

Linear Inverse Problems

5.1 Introduction

In a statistical inverse problem one observes a noisy version of a transformed signal $\mathcal{A}f$ and wishes to recover the unknown parameter f. In Part II, we consider linear inverse problems with different observation schemes. The continuously observed model is the white noise model in the given form,

$$Y^{(n)} = \mathcal{A}f + \frac{1}{\sqrt{n}}\xi, \tag{5.1}$$

where $\mathcal{A} : H \to G$ is a known bounded linear operator between separable Hilbert spaces H and G, and ξ is a stochastic 'noise' process, which is multiplied by the scalar 'noise level' $n^{-1/2}$. The white noise model represents a limiting case (in an appropriate sense) of the inverse regression model

$$Y_i = (\mathcal{A}f)(x_i) + z_i, \quad i = 1, \cdots, n, \tag{5.2}$$

where z_i are independent standard normal random variables. Insights gained in inverse problems in the white noise model shed light on the behaviour of statistical procedures in the inverse regression model, which is the one encountered in actual practice, as the signal f can be typically observed only on a discrete grid of points. Both models belong to the form of Gaussian linear models introduced in Section 4.2,

$$Y^{(n)} = \mathcal{A}^{(n)}f + \xi^{(n)},$$

where $\mathcal{A}^{(n)} = \mathcal{A}$ in the white noise case and $\mathcal{A}^{(n)} = \mathcal{E}^{(n)}\mathcal{A}$ with evaluation operator $\mathcal{E}^{(n)}$ in the regression case (see Section 4.2.1 and Section 4.2.2). In this chapter, we explore the structure of operator $\mathcal{A}$, and in the subsequent chapters in Part II we study the relevant statistical models.

The problem is to infer f from the observation $Y^{(n)}$. To this purpose we assume that the *forward operator* $\mathcal{A}$ is injective, but we shall be interested in the case that the inverse $\mathcal{A}^{-1}$, defined on the range of $\mathcal{A}$ is not continuous (or equivalently the range of $\mathcal{A}$ is not closed in G). The problem of recovering f from $Y^{(n)}$ is then *ill-posed*, and *regularization* methods are necessary in order to 'invert' the operator $\mathcal{A}$. These consist of constructing an approximation to $\mathcal{A}^{-1}$, with natural properties such as boundedness and whose domain includes the data $Y^{(n)}$, and applying this to $Y^{(n)}$. By the discontinuity of the inverse $\mathcal{A}^{-1}$, the noise present in the observation is necessarily multiplied, and regularization is focused on balancing the error in the approximation to $\mathcal{A}^{-1}$ to the size of the magnified noise, in order to obtain a solution that is as close as possible to the true signal f. In this article we study this through the convergence rates of the regularized solutions to a true parameter f, as $n \to \infty$, i.e. as the noise level tends to zero. In particular, we consider contraction rates of posterior distributions resulting from a Bayesian approach to the problem.

It is also possible to consider the model (5.1) with a noise variable ξ that takes its values inside the Hilbert space G. In Section 6.7 we briefly note some results on this 'coloured noise' model, but our main focus is the white noise case.

The chapter is organized as follows. In Section 5.2 we introduce in greater detail our setup along with the assumptions that will be used in this part. We also present some examples for illustration. Then, an estimator, particularly suitable for the framework of inverse problems introduced in the former section, is constructed in Section 5.3. We conclude this short chapter with notes and comments in Section 5.4.

5.2 Inverse Nature

The forward operator $\mathcal{A}$ in the model (5.1) and (5.2) is a bounded linear operator $\mathcal{A} : H \to G$ between the separable Hilbert spaces H and G, and is assumed to be smoothing. The following assumption makes this precise. This assumption is satisfied in many examples and is common in the literature (for instance [19, 38, 71]).

Recall the concept of smoothness scales from Chapter 2. In Definition 2.1 the space H is embedded as $H = H_0$ in the smoothness scale $(H_s)_{s\in\mathbb{R}}$ and hence has norm $\|\cdot\|_0$.

Assumption 5.1 (Smoothing property of $\mathcal{A}$). For some $\gamma > 0$ the operator $\mathcal{A} : H_{-\gamma} \to G$ is injective and bounded and, for every $f \in H_0$,

$$\|\mathcal{A}f\| \simeq \|f\|_{-\gamma}. \tag{5.3}$$

Example 5.2 (SVD). If the operator $\mathcal{A} : H \to G$ is compact, then the positive self-adjoint operator $\mathcal{A}^*\mathcal{A} : H \to H$ possesses a countable orthonormal basis of eigenfunctions ϕ_i, which can be arranged so that the corresponding sequence of eigenvalues λ_i decreases to zero. If $\mathcal{A}$ is injective, then all eigenvalues, whose roots are known as the *singular values* of $\mathcal{A}$, are strictly positive. Suppose that there exists $\gamma > 0$ such that

$$\lambda_i \simeq i^{-2\gamma}. \tag{5.4}$$

If we construct the smoothness classes $(H_s)_{s\in\mathbb{R}}$ from the basis $(\phi_i)_{i\in\mathbb{N}}$ and the numbers $b_i = i$ as in Example 2.6, then (5.3) is satisfied.

Indeed, we can write $\mathcal{A}$ in polar decomposition as $\mathcal{A}f = U(\mathcal{A}^*\mathcal{A})^{1/2}f$, for a partial isometry $U : \mathrm{Ran}(\mathcal{A}) \to G$, and then have $\mathcal{A}f = U\sum_i f_i\sqrt{\lambda_i}\phi_i$, so that $\|\mathcal{A}f\| = \|\sum_i f_i i^{-\gamma}\phi_i\|_0 \simeq \|f\|_{-\gamma}$.

Thus constructions using the singular value decomposition of $\mathcal{A}$ can always be accommodated in the more general setup described in the preceding.

For more interesting illustrations of the preceding setup, consider linear differential equations of the form

$$Du(x) = f(x), \qquad x \in \mathcal{D} \subset \mathbb{R}^d,$$

where D is a differential operator. Under appropriate boundary conditions, the solution u can often be expressed in terms of the Green's function associated with D, through a kernel operator

$$u(x) = \int_{\mathcal{D}} k(x,t) f(t)\, dt =: \mathcal{A}f(x). \tag{5.5}$$

The operator $\mathcal{A}$ typically lifts a function $f \in L^2$ to a Sobolev space of functions, as in Example 2.5. The ill-posedness surfaces when one observes the state u with noise (which deteriorates the smoothness), and tries to recover the source function f. For illustration we include two concrete examples from the literature.

Example 5.3 (Poisson equation)**.** The following example can be found in Sections 10.4 and 11.2 in [44]. Let $(H_s)_{s\in\mathbb{R}}$ be the periodic Sobolev spaces of (generalized) functions satisfying the boundary condition $f(0) = f(1) = 0$. Consider the following boundary problem,

$$u'' := \frac{d^2u}{dx^2} = -f, \qquad f \in H_0,$$

with the Dirichlet boundary condition: $u(0) = u(1) = 0$. The unique solution $u \in H_2$ is given by

$$u(x) = \mathcal{A}f(x) = \int_0^1 k(x,t) f(t)\, dt,$$

where

$$k(x,t) = \begin{cases} (1-x)t, & \text{if } x \geq t, \\ (1-t)x, & \text{otherwise.} \end{cases}$$

The operator $\mathcal{A} : H_0 \to H_0$ is Hilbert-Schmidt and hence compact, and therefore has no bounded inverse. On the other hand the inverse exists as bounded operator $\mathcal{A}^{-1} : H_2 \to H_0$, and is given by $\mathcal{A}^{-1}f = -f''$.

When $\mathscr{D} \subset \mathbb{R}$, the Sobolev norm is equivalent to $\|f\|_2 = \|f\|_0 + \|f''\|_0$ (Page 217 in [10]). Since $(\mathcal{A}f)'' = f$, we have $\|f\|_0 \leq \|\mathcal{A}f\|_2 = \|\mathcal{A}f\|_0 + \|(\mathcal{A}f)''\|_0 \leq (\|\mathcal{A}\| + 1)\|f\|_0$, i.e $\|\mathcal{A}f\|_2 \simeq_{\mathcal{A}} \|f\|_0$.

Since the kernel is symmetric, $\mathcal{A}$ is self-adjoint. Besides, $\mathcal{A}$ is an isomorphism between H_2 and H_0 as shown above. Hence

$$\|\mathcal{A}f\|_0 = \sup_{\|h\|_0 \leq 1} |\langle h, \mathcal{A}f\rangle_0| = \sup_{\|h\|_0 \leq 1} |\langle \mathcal{A}h, f\rangle_0| \simeq_{\mathcal{A}} \sup_{\|h\|_2 \leq 1} |\langle h, f\rangle_0| = \|f\|_{-2},$$

by norm duality argument, for all $f \in H_0$. This shows that (5.3) holds with $\gamma = 2$.

Example 5.4 (Symm's equation [57])**.** Consider the Laplace equation $\Delta u = 0$ in a bounded set $\Omega \subset \mathbb{R}^2$ with boundary condition $u = g$ on the boundary $\partial\Omega$. The singular layer potential, a boundary integral

$$u(x) = -\frac{1}{\pi}\int_{\partial\Omega} h(y)\ln|x-y|\,ds(y), \quad x \in \Omega,$$

solves the boundary value problem if and only if the density h, belonging to the space $C(\partial\Omega)$ of continuous functions on $\partial\Omega$, solves *Symm's equation*

$$-\frac{1}{\pi}\int_{\partial\Omega} h(y)\ln|x-y|\,ds(y) = g(x), \quad x \in \partial\Omega. \tag{5.6}$$

Assume the boundary $\partial\Omega$ has a parametrization of the form $\{\rho(s), s \in [0, 2\pi]\}$, for some 2π-periodic analytic function $\rho : [0, 2\pi] \to \mathbb{R}^2$ such that $|\dot{\rho}(s)| > 0$ for all s. Then Symm's equation takes the following form,

$$\mathcal{A}f(z) := -\frac{1}{\pi}\int_0^{2\pi} \log|\rho(z)-\rho(s)|\,f(s)\,ds = g(\rho(z)), \quad z \in [0, 2\pi],$$

where $f(s) = h(\rho(s))|\dot{\rho}(s)|$. As shown in Theorem 3.18 from [57], the operator $\mathcal{A}$ satisfies (5.3), with $\gamma = 1$ and $(H_s)_{s\in\mathbb{R}}$ being periodic Sobolev spaces on $[0, 2\pi]$.

The following example is an inverse problem in Hilbert scales (see Section 2.2).

Example 5.5 (Abel operator)**.** For a given kernel function $K : (0,1) \times (0,1) \to \mathbb{R}$ and $\alpha \in (0, 1]$, consider the operator $A : L^2(0,1) \to L^2(0,1)$ given by

$$Af(x) = \frac{1}{\Gamma(\alpha)}\int_0^x (x-s)^{\alpha-1}K(x,s)f(s)\,ds.$$

For $K = 1$ this gives the classical *Abel operator*. Under mild smoothness conditions on K, it is shown in [39], Theorem 1, that A is smoothing (i.e. (5.3) holds) of order $\gamma = 1$ for the Sobolev scale generated by the root negative Laplacian under the Cauchy boundary condition, described in Example 2.10.

In the Bayesian setup we model a function through a prior. When a true function is known to satisfy certain boundary conditions, as in many problems involving differential forward operators, we can incorporate these in the by choosing an appropriate generating operator. For an operator $\mathcal{A}$ defined in terms of the Laplacian and the same boundary conditions the smoothing condition (5.3) will be satisfied. The following is a another example of a pair of Λ and $\mathcal{A}$.

Example 5.6 (Volterra operator)**.** Consider the operator $A : L^2((0,1)^2) \to L^2((0,1)^2)$ on functions $f : (0,1)^2 \to \mathbb{R}$ on the unit square satisfying the differential equation

$$D_{x,y}\mathcal{A}f = f, \qquad D_{x,y} = \frac{\partial^2}{\partial x \partial y}.$$

We can render the solution of the equation unique by imposing boundary conditions. Two solutions are given by

$$\begin{aligned}\mathcal{A}f(x,y) &= \int_0^x\int_0^y f(s,t)\,ds\,dt,\\ \mathcal{A}_0f(x,y) &= \mathcal{A}f(x,y) - \int_0^1 \mathcal{A}f(x,t)\,dt - \int_0^1 \mathcal{A}f(s,y)\,ds + \int_0^1\int_0^1 \mathcal{A}f(s,t)\,ds\,dt.\end{aligned}$$

The first satisfies the boundary conditions $\mathcal{A}f(x,0) = \mathcal{A}f(0,y) = 0$, while the second is obtained from the first by subtracting its projection on the set of all functions of the form $(x,y)\mapsto g_1(x)+g_2(y)$, which forms the kernel of the differential operator. Other boundary conditions will still give different versions of the operator.

We claim that $\mathcal{A}_0$ is smoothing of order $\gamma = 1$ for the Hilbert scale generated by the root Λ of $D^2_{x,y}$ with Dirichlet boundary condition, while $\mathcal{A}$ is smoothing relative to the scale of L combined with Cauchy boundary condition.

The scale under the Dirichlet boundary condition is generated by the orthogonal system of eigenfunctions $e_{k,l} : (x,y)\mapsto \sin(k\pi x)\sin(l\pi y)$, for $(k,l)\in\mathbb{N}^2$, the tensor product of the basis of the one-dimensional Dirichlet-Laplacian as in Example 2.10, with corresponding eigenvalues are $k^2l^2\pi^4$. By explicit calculation

$$\begin{aligned}\mathcal{A}e_{k,l}(x,y) &= \frac{1}{kl\pi^2}[\cos(k\pi x)\cos(l\pi y) - \cos(k\pi x) - \cos(l\pi y) + 1],\\ \mathcal{A}_0e_{k,l}(x,y) &= \frac{1}{kl\pi^2}\cos(k\pi x)\cos(l\pi y).\end{aligned}$$

The functions $(x,y)\mapsto\cos(k\pi x)\cos(l\pi y)$, for $(k,l)\in(\mathbb{N}\cup\{0\})^2$ form an orthogonal basis of $L^2((0,1)^2)$. We conclude that for $f=\sum_{k,l}f_{k,l}e_{k,l}$,

$$\begin{aligned}\|\mathcal{A}f\|^2 &\simeq \sum_{k,l}\frac{f_{k,l}^2}{k^2l^2} + \sum_k\Bigl(\sum_l\frac{f_{k,l}}{kl}\Bigr)^2 + \sum_l\Bigl(\sum_k\frac{f_{k,l}}{kl}\Bigr)^2 + \Bigl(\sum_{k,l}\frac{f_{k,l}}{kl}\Bigr)^2,\\ \|\mathcal{A}_0f\|^2 &\simeq \sum_{k,l}\frac{f_{k,l}^2}{k^2l^2} \simeq \|f\|_{-1}^2,\end{aligned}$$

where $\|\cdot\|_{-1}$ refers to the scale of Λ with Dirichlet boundary condition. The first equation shows that the operator $\mathcal{A}$ is not smoothing in this scale, but in general satisfies $\|\mathcal{A}f\|\gtrsim\|f\|_{-1}$.

On the other hand, the Cauchy boundary condition generates the system of eigenfunctions $(x,y)\mapsto\cos((k-1/2)\pi x)\cos((l-1/2)\pi y)$, for $(k,l)\in\mathbb{N}^2$. These can be seen to be also the eigenfunctions of $\mathcal{A}^*\mathcal{A}$, and hence the smoothing property of A fits the SVD framework, as in Example 5.2.

The two versions $\mathcal{A}$ and $\mathcal{A}_0$ possess the same inverse operator, namely the differential operator $D_{x,y}$ used for their definitions. This suggests that from the point of view of reconstructing f in the inverse problem it should not matter whether one is provided with a noisy version of either $\mathcal{A}f$ or $\mathcal{A}_0f$ as input data, seemingly contradicting the fact that the operators are smoothing in different scales. This paradox may be resolved by considering $\mathcal{A}$ or $\mathcal{A}_0$ as maps into the quotient space $L^2((0,1)^2)/N(D_{x,y})$, where N denotes the kernel of the operator. The map $f \mapsto [\mathcal{A}f] = [\mathcal{A}_0 f]$ into the class of $\mathcal{A}f$ in this quotient space is injective and can be shown to be appropriately smoothing (see (5.10)-(5.11)), and consequently both scales can be used with both operators (cf. Remark 5.7).

Remark 5.7. For all our purposes the smoothing condition (5.3) can be relaxed to (5.10)-(5.11). This relaxation covers the situation where there exists an operator A_0 that satisfies (5.3) and is a 'version' of A in that the two operators possess a common inverse, such as when $\mathcal{A}$ and $\mathcal{A}_0$ are defined to solve a differential equation with different boundary conditions. Lemma 5.9 shows that the relaxed version of the smoothing condition is then satisfied by the map $f \mapsto [\mathcal{A}f]$ of f in the class of $\mathcal{A}f$ in the quotient space $G/R(\mathcal{A}-\mathcal{A}_0)$.

5.3 Galerkin Projection

In this section we collect some (well known) results on the Galerkin method.

Consider a scale of smoothness classes $(H_s)_{s\in\mathbb{R}}$ as in Definition 2.1. Let $\mathcal{A}: H \to G$ be an injective bounded operator between separable Hilbert spaces, and let V_j be a finite-dimensional subspace of H. The Galerkin solution $f^{(j)} \in V_j$ to the image $\mathcal{A}f$ of an element f is defined as the element in V_j such that $\mathcal{A}f^{(j)}$ is equal to the orthogonal projection of $\mathcal{A}f$ onto the image space $W_j = \mathcal{A}V_j$. Thus, if $Q_j : G \to W_j$ denotes the orthogonal projection onto W_j, then the Galerkin solution can be written as

$$f^{(j)} = R_j \mathcal{A} f, \quad \text{for} \qquad R_j = \mathcal{A}^{-1} Q_j,$$

where the inverse $\mathcal{A}^{-1}$ is well defined on the linear subspace W_j.

If the operators $R_j\mathcal{A}$ are uniformly bounded with respect to j, then the convergence rate $\|f^{(j)} - f\|_0$ of the Galerkin solution to f is known to be of the same order as the distance $\|P_j f - f\|_0$ of f to its projection on V_j. (See Section 3.2 and Theorem 3.7 in [57], or the proof below.) In particular, if $f \in H_s$ and V_j satisfies (2.2), then the convergence rate is given by $\delta(j,s)$.

In order to control the stochastic noise term ξ in the observation schemes (5.1) and (5.2), it is necessary also to control the norms of the operators R_j. The following lemma summarizes the properties of the Galerkin projection needed in the proof of our main result.

Lemma 5.8. *If V_j is a finite-dimensional space as in Assumption 2.3 such that (2.2) and (2.3) hold, and $\mathcal{A}: H_0 \to G$ is a bounded linear operator satisfying $\|\mathcal{A}f\|_0 \simeq \|f\|_{-\gamma}$ for every $f \in H_0$, then the norms of the operators $R_j : G \to H_0$ and $R_j\mathcal{A} : H_0 \to H_0$ satisfy*

$$\|R_j\| \lesssim_{\mathcal{A}} \frac{1}{\delta(j,\gamma)}, \tag{5.7}$$

$$\|R_j\mathcal{A}\| \lesssim_{\mathcal{A}} 1. \tag{5.8}$$

Furthermore, for $f \in H_s$ the Galerkin solution $f^{(j)} \in V_j$ to $\mathcal{A}f$ satisfies

$$\|f^{(j)} - f\|_0 \lesssim_{\mathcal{A}} \delta(j,s)\, \|f\|_s. \tag{5.9}$$

Proof. For $g \in G$ we have $R_j g \in V_j$ and hence by (2.5),

$$\|R_j g\|_0 \lesssim \frac{1}{\delta(j,\gamma)}\|R_j g\|_{-\gamma} \simeq \frac{1}{\delta(j,\gamma)}\|\mathcal{A}R_j g\|_0 = \frac{1}{\delta(j,\gamma)}\|Q_j g\|_0,$$

since $\mathcal{A}R_j g = Q_j g$. Because $\|Q_j g\|_0 \leq \|g\|_0$, we conclude that $\|R_j\| \lesssim 1/\delta(j,\gamma)$.

By definition $f^{(j)} = R_j\mathcal{A}f$, and $R_j\mathcal{A}$ acts as the identity on V_j. Therefore $f^{(j)} - P_j f = R_j\mathcal{A}(f - P_j f)$, and hence

$$\|f^{(j)} - P_j f\|_0 \leq \|R_j\|\,\|\mathcal{A}(f - P_j f)\|_0 \simeq \|R_j\|\,\|f - P_j f\|_{-\gamma} \leq \|R_j\|\,\delta(j,\gamma)\|f\|_0,$$

by (2.4). By the preceding paragraph $\|R_j\|\,\delta(j,\gamma) \lesssim 1$, so that the right side is bounded above by $\|f\|_0$. By the triangle inequality

$$\|R_j\mathcal{A}f\|_0 = \|f^{(j)}\|_0 \leq \|f^{(j)} - P_j f\|_0 + \|P_j f - f\|_0 + \|f\|_0 \lesssim \|f\|_0,$$

in view of the preceding display and the fact that $\|P_j f - f\|_0 \leq \|f\|_0$. This shows that $\|R_j\mathcal{A}\| \lesssim 1$.

Finally, since $f^{(j)} - f = (R_j\mathcal{A} - I)(f - P_j f)$, we have that

$$\|f^{(j)} - f\|_0 = \|(R_j\mathcal{A} - I)(f - P_j f)\|_0 \leq (\|R_j\mathcal{A}\| + 1)\|f - P_j f\|_0.$$

Inequality (5.9) follows by the boundedness of $\|R_j\mathcal{A}\|$ and (2.2). ☐

As is clear from the proof, the smoothing assumption $\|\mathcal{A}f\| \simeq \|f\|_{-\gamma}$ can be relaxed to the pair of inequalities

$$\|\mathcal{A}f\| \lesssim \|f\|_{-\gamma}, \qquad f \perp V_j, \tag{5.10}$$

$$\|\mathcal{A}f\| \gtrsim \|f\|_{-\gamma}, \qquad f \in R(R_j). \tag{5.11}$$

This helps to cover cases in which the smoothing condition is satisfied for a modification of the operator $\mathcal{A}$, but not $\mathcal{A}$ itself, for example a modification taking different boundary conditions of a differential operator into account.

We introduce a *modified Galerkin solution* to $\mathcal{A}f$ to cover such a case. Let $\mathcal{A}_0, \mathcal{A} : H \to G$ be injective bounded operators between separable Hilbert spaces that possess a common inverse in the sense of existence of a linear map $B : D(B) \subset G \to H$ with domain $D(B)$ containing the linear span of the ranges of $\mathcal{A}_0$ and $\mathcal{A}$ such that that $B\mathcal{A}_0 = I = B\mathcal{A}$. For simplicity of notation, write $B = \mathcal{A}^- = \mathcal{A}_0^-$. Intuitively, for the inverse problem, taking $\mathcal{A}_0 f$ or $\mathcal{A}f$ as input data should be equivalent. However, it may be that $\mathcal{A}_0$ is smoothing in a given scale $(H_s)_{s\in\mathbb{R}}$, whereas $\mathcal{A}$ is not. In that case we reconstruct as follows. Assume that $\Phi = \mathcal{A} - \mathcal{A}_0$ has closed range, and let $P_\Phi : G \to G$ be the orthogonal projection onto this range. Now let $Q_j : G \to G$ be the orthogonal projection onto the finite-dimensional space $(I - P_\Phi)AV_j$, and set

$$f^{(j)} = R_j\mathcal{A}f, \quad \text{for} \qquad R_j = \mathcal{A}^- Q_j(I - P_\Phi). \tag{5.12}$$

Thus after removing the “irrelevant part” of $\mathcal{A}f$ that does not influence the inversion, we project onto the finite-dimensional space $(I - P_\Phi)\mathcal{A}V_j$ of similarly cleaned functions $\mathcal{A}f$ with $f \in V_j$, and finally invert.

Lemma 5.9. *If* V_j *is a finite-dimensional space as in Assumption 2.3 such that (2.2) and (2.3) hold, and* $\mathcal{A}_0, \mathcal{A} : H_0 \to G$ *are bounded linear operators with common inverse satisfying* $\|\mathcal{A}_0 f\| \simeq \|f\|_{-\gamma}$ *for every* $f \in H_0$*, then the operators* $R_j : G \to H_0$ *and* $R_j\mathcal{A} : H_0 \to H_0$ *and* $f^{(j)} = R_j\mathcal{A}f$ *as in* (5.12) *satisfy (5.7), (5.8), and (5.9).*

Proof. The operator $[\mathcal{A}] : H \to G/\Phi(H)$ mapping $f \in H$ into the class of $\mathcal{A}f$ in the quotient space $G/\Phi(H)$ is one-to-one, since $[\mathcal{A}f] = 0$ implies $\mathcal{A}f \in R(\Phi)$ and hence $f = B\mathcal{A}f = 0$, since $B\Phi = 0$. Identifying $[g] \in \tilde{G} := G/\Phi(H)$ with the function $(I - P_\Phi)g$ with norm $\|[g]\|_{\tilde{G}} = \|(I - P_\Phi)g\|_G$, we see that $R_j\mathcal{A}f$ as in (5.12) is actually the Galerkin solution to $[\mathcal{A}]f$. It suffices to show that $[\mathcal{A}] : H \to \tilde{G}$ is smoothing in the sense of (5.10). Now $\|[\mathcal{A}f]\|_{\tilde{G}} = \|(I - P_\Phi)\mathcal{A}_0 f\|_G \leq \|\mathcal{A}_0 f\|_G \simeq \|f\|_{-\gamma}$, for every $f \in H$. Furthermore, for every f such that $\mathcal{A}_0 f \perp R(\Phi)$, the inequality is an equality. This is true for $f = R_j g$, since $\mathcal{A}_0 R_j g = Q_j(I - P_\Phi)g \in (I - P_\Phi)\mathcal{A}V_j$. □

5.4 Notes

Statistical Inverse Problems

The study of statistical (nonparametric) linear inverse problems was initiated by Wahba in 1970s in [101]. The 1990s paper [24] used wavelet shrinkage methods, while around 2000, the authors of [17] investigated (5.1) in the linear partial differential equations setting, while a systematic study of Gaussian sequence models was presented in [16]. A review of work until 2008 is given in [15]. The connection of regularization methods to the Bayesian approach was recognized early on. However, the study of the recovery properties of posterior distributions was started only in [59, 60]. A review of the Bayesian approach to inverse problems, with many examples, is given in [88].

Much of the existing work on statistical inverse problems is based on the singular value decomposition (SVD) of the operator $\mathcal{A}$; see, e.g., [15]. When $\mathcal{A}$ is compact, the operator $\mathcal{A}^*\mathcal{A}$, where $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$, can be diagonalized with respect to an orthonormal *eigenbasis*, with eigenvalues tending to zero. The observation $Y^{(n)}$ can then be reduced to noisy observations on the Fourier coefficients of $\mathcal{A}f$ in the eigenbasis, which are multiples of the Fourier coefficients of f, and the problem is to recover the latter. In the frequentist setup thresholding or other regularization methods can be applied to reduce the weight of estimates on coefficients corresponding to smaller eigenvalues, in which the noise will overpower the signal. In the Bayesian setup one may design a prior by letting the Fourier coefficients be (independent) random variables, with smaller variances for smaller eigenvalues. These singular value methods have several disadvantages, as pointed out in [19, 24]. First, the eigenbasis functions might not be easy to compute. Second, and more importantly, these functions are directly linked to the operator $\mathcal{A}$, and need not be related to the function space (smoothness class) that is thought to contain the true signal f. Consequently, the parameter of interest f may not have a simple, parsimonious representation in the eigenbasis expansion, see [24]. Furthermore, it is logical to consider the series expansion of the signal f in other bases than the eigenbasis, for instance, in the situation that one can only measure noisy coefficients of the signal f in a given basis expansion, due to a particular experimental setup. See [37, 70] for further discussion.

Deterministic Inverse Problems
There is a rich literature on inverse problems. The case that the noise ξ is a bounded *deterministic* perturbation, has been particularly well studied, and various general procedures and methods to estimate the convergence rates of regularized solutions have been proposed. See the monographs [29, 57]. The case of stochastic noise is less studied, but is receiving increasing attention.

Inverse Problems in Scales and Bayesian Approach
A canonical example are Sobolev spaces, with the operator $\mathcal{A}$ being an integral operator. This Sobolev space setup with wavelet basis was investigated in [19, 24]. In deterministic inverse problems, a more general setup, considering $\mathcal{A}$ that acts along nested Hilbert spaces, *Hilbert scales*, was initiated by Natterer in [71] and further developed in, amongst others, [46, 69, 70]. In the Bayesian context Hilbert scales were used in [30], under the assumption that the noise ξ is a proper Gaussian element in G, and in [1], but under rather intricate assumptions. The second question, to allow priors that are not conjugate, can also be answered under the condition of $\mathcal{A}$. In the linear inverse problem Gaussian priors are easy, as they lead to Gaussian posterior distributions, which can be studied by direct means. Most of the results on Bayesian inverse problems fall in this framework [1, 30, 59, 60], exceptions being [80] and [58].

Generalized Random Elements
An alternative method to give a rigorous interpretation to white noise ξ, is to embed G into a bigger space in which ξ can be realized as a Borel measurable map, or to think of ξ as a cylindrical process. See e.g., [87]. For G a set of functions on an interval, one can also realize ξ as a stochastic integral relative to Brownian motion, which takes its values in the 'abstract Wiener space' attached to G.