Chapter 3

Ill-posedness in inverse problems

Inverse problems are characterized by ill-posedness—in other words extreme sensitivity to measurement noise and modeling errors. In this chapter we will look at ill-posedness of the infinite-dimensional inverse problem, which requires some knowledge of operator theory, and ill-posedness of the finite-dimensional inverse problem, which typically arises as a discretization of an infinite-dimensional inverse problem. For the necessary background on Banach spaces and compact linear operators, see Appendix B and references therein. We introduce the concept of regularization and the singular value decomposition.

3.1 Forward map and Hadamard's conditions

The examples in Sections 2.1–2.3 should convince the reader that there is something suspicious going on with the inversion of those three simple and indirect measurements. The reason for the observed instability is *ill-posedness*, which we discuss next.

The core of any inverse problem is the *forward map* $\mathcal{A} : \mathcal{D}(A) \to Y$, a mathematical model of the corresponding direct problem. Here X and Y are suitable Hilbert spaces called *model space* and *data space*, respectively, and the subset $\mathcal{D}(A) \subset X$ is the domain of definition of the bounded linear operator $\mathcal{A}$. The forward map is used as a mathematical model of the indirect measurement

$$m = \mathcal{A} f + \varepsilon. \tag{3.1}$$

Here $f \in \mathcal{D}(A) \subset X$ is the quantity of interest, $m \in Y$ is measurement data, and ε is noise satisfying $\|\varepsilon\|_Y \leq \delta$ with some known $\delta > 0$. (Here we denote the measurement by m instead of the vector notation $\mathbf{m}$ because we want to cover more general data spaces than just $Y = \mathbb{R}^k$.)

Given a particular inverse problem, it is not necessarily straightforward to choose the spaces X and Y and the forward map $\mathcal{A}$. Constructing them must be considered as a nontrivial mathematical modeling task. The following aspects need to be modeled: physical processes involved, technical properties of the measurement device, geometry of the measurement, and possible limitations in data sets.

According to Hadamard, a solution method is called *well-posed* if the following three conditions are satisfied:

H_1: **Existence.** There should be at least one solution.

H_2: **Uniqueness.** There should be at most one solution.

H_3: **Stability.** The solution must depend continuously on data.

We need to study the well-posedness of the naïve inversion $\mathcal{A}^{-1}m$ as a solution to the inverse problem "Given m, find f."

If the forward map is bijective from X to Y and allows a continuous inverse $\mathcal{A}^{-1}$, then naïve inversion satisfies all conditions H_1–H_3 and we are dealing with a well-posed inverse problem.

This book is about *ill-posed inverse problems*, defined as the complement of well-posed problems: at least one of the conditions H_1–H_3 must fail for naïve inversion $\mathcal{A}^{-1}m$. Condition H_1 is violated if the measured noisy data does not belong to the range of the forward map: $\mathcal{A}f + \varepsilon \notin \mathcal{A}(\mathcal{D}(\mathcal{A}))$. Condition H_2 fails if two quantities $f, g \in \mathcal{D}(\mathcal{A})$ give the same measurement: $\mathcal{A}f = \mathcal{A}g$, leading to nonuniqueness. The forward map does not always allow a continuous inverse, not even when restricted to the range $\mathcal{A}(\mathcal{D}(\mathcal{A}))$, so condition H_3 does not hold. See Figure 3.1 for a schematic illustration of the forward map and the related definitions.

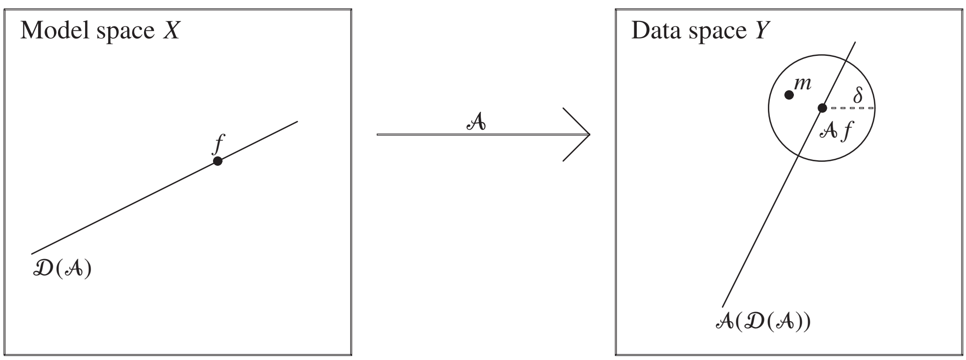

Figure 3.1. Schematic illustration of the linear forward map $\mathcal{A}$. The radius δ indicates the known maximum amplitude of measurement noise.

3.3 Ill-posedness in the continuous case

This section relies on some knowledge of operator theory, and the reader may need to first cover some or all of the material in the appendices before covering this section. Recall from Appendix B that $\mathcal{L}(X,Y)$ denotes the space of bounded linear mappings from a normed linear space X to a normed linear space Y.

In terms of the linear operator $\mathcal{A} : U \to V$ with U and V being subsets of normed spaces X and Y, the conditions of well-posedness imply

- $\mathcal{A}$ is surjective (onto),
- $\mathcal{A}$ is injective (one-to-one),
- $\mathcal{A}^{-1}$ is continuous.

These three properties are not independent. The following theorem is a consequence of the closed graph theorem, which implies that the range of a continuous operator from a Banach space to a Banach space is closed.

Theorem 3.1 (see [396, Theorem 7.72]). *Let X and Y be Banach spaces. If $\mathcal{A} \in \mathcal{L}(X,Y)$ is bijective, then $\mathcal{A}^{-1} \in \mathcal{L}(Y,X)$.*

By Theorem B.1 of Appendix B, boundedness implies continuity, so the problem $\mathcal{A}f = m$ is well-posed. From the following theorem, we see that an ill-posed problem is guaranteed to result when a compact linear operator acts on an infinite-dimensional Banach space.

Theorem 3.2. *Let X and Y be Banach spaces. Suppose $\mathcal{A} : U \subset X \to Y$ is a compact linear operator, and* $\dim U$ *is not finite. Then the problem $\mathcal{A}f = m$ is ill-posed.*

Proof. Suppose $\dim U = \infty$ and $\mathcal{A}^{-1} \in \mathcal{L}(Y,U)$, that is, the inverse of $\mathcal{A}$ exists and is continuous. Then $I_U = \mathcal{A}^{-1}\mathcal{A}$ is the composition of a compact and a continuous operator and thus compact. However, the identity map I_U on an infinite-dimensional Banach space is not compact. Thus, $\dim(U)$ is finite.

Two straightforward consequences of this theorem are given in the following two corollaries. Their proofs are Exercises 3.3.1 and 3.3.2.

Corollary 3.3.1. *If $\mathcal{A}$ is a compact, linear operator from H to H, where H is an infinite-dimensional Hilbert space, and $\mathcal{A}^{-1}$ exists, then $\mathcal{A}^{-1}$ is unbounded.*

Corollary 3.3.2. *If $\mathcal{A} : H \to H$ is a compact, linear operator and $\mathcal{A}^{-1}$ exists, then* $\dim(H)$ *is finite.*

The following theorem provides one way to prove that an operator is compact. Theorems 3.3 and 3.4 are very important because they explain why the ill-posed behavior of a large class of linear systems, $\mathcal{A} f = m$, cannot be rigorously detected by examining a single approximation matrix A_k to the operator $\mathcal{A}$, and why one needs to examine the sequence of approximations $\{A_k\}$ for large k. This leads to the test defined by condition (3.18).

Theorem 3.3. *Let $\mathcal{A} : H \to H$ be a linear operator and $K_n \in \mathcal{L}(H,H)$ a sequence of compact operators. If $K_n \to \mathcal{A}$ in the operator norm, then $\mathcal{A}$ is a compact operator.*

The proof can be found in [396]; it uses a diagonal sequence argument.

If K is a Hilbert–Schmidt kernel of an integral operator, the resulting integral equation, a Fredholm integral equation of the first kind when $\mathcal{A} f = m$, is an ill-posed problem, as we see from the following theorem.

Theorem 3.4. *Let*

$$(\mathcal{A} f)(x) = \int_\Omega K(x,y)\, f(y)\, dy$$

with kernel $K \in L^2(\Omega \times \Omega)$ (i.e., K is a Hilbert–Schmidt kernel from $\Omega \times \Omega \to \mathbb{R}$). Then $\mathcal{A} \in \mathcal{L}(L^2(\Omega), L^2(\Omega))$ is compact.

Proof. First note that if Φ_i is an orthonormal basis for $L^2(\Omega)$, then $\Phi_i(x)\,\Phi_j(y)$ is an orthonormal basis for $L^2(\Omega \times \Omega)$. Express K in this basis:

$$K(x,y) = \sum_{i,j=1}^{\infty} k_{i,j}\, \Phi_i(x)\, \Phi_j(y)$$

with

$$k_{i,j} = \int_\Omega \int_\Omega K(x,y)\, \Phi_i(x)\, \Phi_j(y)\, dx\, dy.$$

The convergence is in L^2, and we find

$$\|K\|_{L^2}^2 = \sum_{i,j=1}^{\infty} k_{i,j}^2.$$

Define

$$K_n(x,y) = \sum_{i,j=1}^{n} k_{i,j}\, \Phi_i(x)\, \Phi_j(y).$$

Then

$$(\mathcal{A}_n f)(x) = \int_\Omega K_n(x,y)\, f(y)\, dy = \sum_{i,j=1}^{n} k_{i,j} \int_\Omega \Phi_i(x)\, \Phi_j(y) f(y) dy.$$

Thus, $\mathcal{A}_n$ maps from $L^2(\Omega)$ to a finite-dimensional subspace of $L^2(\Omega)$, which we will denote by $\tilde{L}^2(\Omega)$. By the converse of Corollary 3.3.1, due to the fact that the Range$(\mathcal{A}_n)$ is finite-dimensional, $\mathcal{A}_n$ is compact. Now we see that

$$\begin{aligned}
\|(\mathcal{A}-\mathcal{A}_n)f\|^2_{\tilde{L}^2(\Omega)} &= \left\| \int_\Omega (K(x,y)-K_n(x,y))\, f(y)\, dy \right\|^2_{\tilde{L}^2(\Omega)} \\
&= \int_\Omega \left| \int_\Omega (K(x,y)-K_n(x,y))\, f(y)\, dy \right|^2 dx \\
&\le \int_\Omega \left(\int_\Omega |K(x,y)-K_n(x,y)|\, |f(y)|\, dy \right)^2 dx \\
&\quad \text{by Cauchy–Schwarz} \\
&\le \int_\Omega \left(\int_\Omega |K(x,y)-K_n(x,y)|^2\, dx \int_\Omega |f(y)|^2\, dy \right) dx \\
&= \left(\int_\Omega \int_\Omega |K(x,y)-K_n(x,y)|^2\, dx\, dy \right) \int_\Omega |f(y)|^2\, dy.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|\mathcal{A}-\mathcal{A}_n\|^2 &\le \int_\Omega \int_\Omega |K(x,y)-K_n(x,y)|^2\, dx\, dy \\
&= \sum_{i,j=n+1}^{\infty} |k_{i,j}|^2 \longrightarrow 0 \ \text{ as } \ n \to \infty.
\end{aligned}$$

Thus, $\mathcal{A}_n$ converges to $\mathcal{A}$ in the operator norm, and $\mathcal{A}_n$ is compact. We conclude that $\mathcal{A}$ is compact.

3.4 Regularized inversion

Having established that many inverse problems are ill-posed, we now introduce an indispensable technique known as *regularization* or *regularized inversion* to deal with the ill-posedness.

The basic inverse problem related to the indirect measurement (3.1) is this:

(IP_1) **Inverse problem:** *Let* $m = \mathcal{A} f + \varepsilon$ *as in* (3.1).
Given m *and* $\delta > 0$ *with* $\|m - \mathcal{A} f\|_Y \leq \delta$, *recover* f *approximately.*

In ill-posed inverse problems there does not exist any continuous function from Y to X that would map $\mathcal{A} f \in Y$ to $f \in X$. This can be viewed as extreme sensitivity to perturbations in $\mathcal{A} f$, which is inevitable because of measurement noise. Consequently, it is not straightforward to design a computational method that would map $m = \mathcal{A} f + \varepsilon$ to some point in X near f.

The naïve way of approaching the inverse problem IP_1 would be to approximate f by $\mathcal{A}^{-1} m$. Because m and $\mathcal{A} f$ are close to each other, the point $f = \mathcal{A}^{-1} \mathcal{A} f$ must be close to $\mathcal{A}^{-1} m$. This approach is fine for well-posed problems but does not work for ill-posed inverse problems since $\mathcal{A}$ may fail to be either injective or surjective, and even if $\mathcal{A}$ is invertible, its inverse may not be continuous. So what can be done?

Sometimes it helps to consider a restricted problem, such as the following:

(IP_2) **Restricted inverse problem:** *Let* $m = \mathcal{A} f + \varepsilon$ *as in* (3.1).
Given m *and* $\delta > 0$ *with* $\|m - \mathcal{A} f\|_Y \leq \delta$, *extract any information about* f.

For example, one might look for the locations of inclusions in known background material. In any case, the most important property of an inversion method is robustness against noise.

Let us define the notions of *regularization strategy* and *admissible choice of regularization parameter*. We need to assume that $\mathrm{Ker}(\mathcal{A}) = \{0\}$; however, this is not a serious lack of generality since we can always consider the restriction of $\mathcal{A}$ to $(\mathrm{Ker}(\mathcal{A}))^{\perp}$ by working in the linear space of equivalence classes $[f + \mathrm{Ker}(\mathcal{A})]$.

Definition 3.4.1. *Let X and Y be Hilbert spaces. Let $\mathcal{A} : X \to Y$ be an injective bounded linear operator. Consider the measurement $m = \mathcal{A}f + \varepsilon$. A family of linear maps $\mathcal{R}_\alpha : Y \to X$ parameterized by $0 < \alpha < \infty$ is called a* regularization strategy *if*

$$\lim_{\alpha \to 0} \mathcal{R}_\alpha \mathcal{A} f = f \tag{3.10}$$

for every $f \in X$.

Further, assume we are given a noise level $\delta > 0$ so that $\|m - \mathcal{A}f\|_Y \le \delta$. A choice of regularization parameter $\alpha = \alpha(\delta)$ as a function of δ is called admissible *if*

$$\begin{aligned} &\alpha(\delta) \to 0 \textit{ as } \delta \to 0, \textit{ and} \\ &\sup_m \left\{ \|\mathcal{R}_{\alpha(\delta)} m - f\| : \|\mathcal{A}f - m\| \le \delta \right\} \to 0 \quad \textit{as } \delta \to 0 \textit{ for every } f \in X. \end{aligned} \tag{3.11}$$

Figure 3.2 shows a schematic illustration of regularized inversion. See [132, 269] for more details about regularization.

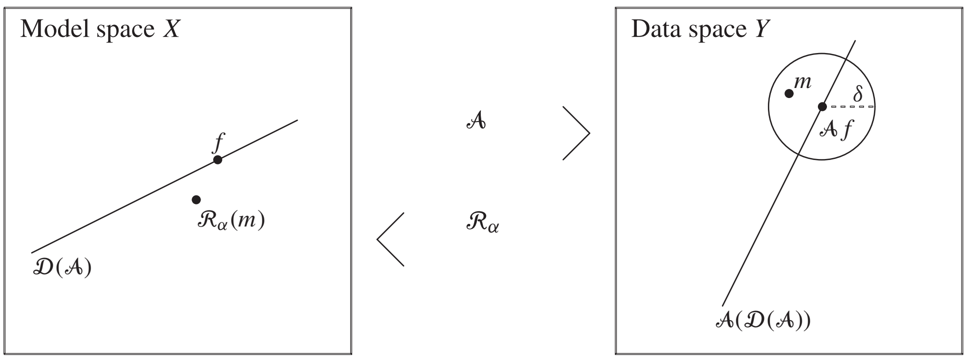

Figure 3.2. Schematic illustration of regularization for linear inverse problems. The linear forward map $\mathcal{A}$ does not necessarily have a continuous inverse. The regularized approximate inverses $\mathcal{R}_\alpha : Y \to X$ are continuous for any choice of regularization parameter $0 < \alpha < \infty$.

This book studies computational inversion methods that apply to practical inverse problems and allow precise mathematical analysis. Developing such methods typically involves considering the following series of questions:

Q_1: **Uniqueness.** Is the forward map $\mathcal{A}$ injective on $\mathcal{D}(\mathcal{A})$? In other words, are there two different objects $f \neq \widetilde{f}$ producing exactly the same infinite-precision measurement data: $\mathcal{A}f = \mathcal{A}\widetilde{f}$?

Q_2: **Reconstruction from ideal data.** Assume that $\mathcal{A}$ is injective on $\mathcal{D}(\mathcal{A})$. Given infinite-precision data $\mathcal{A}f$, how to calculate f?

Q_3: **Conditional stability.** Assume that $\mathcal{A}$ is injective on $\mathcal{D}(\mathcal{A})$. Is it possible to derive a formula of the type $\|f - \widetilde{f}\|_X \le g(\|\mathcal{A}f - \mathcal{A}\widetilde{f}\|_Y)$, where $g : \mathbb{R} \to \mathbb{R}$ is a continuous function satisfying $g(0) = 0$?

Q_4: **Characterization of the range.** How to decide whether a given $m \in Y$ belongs to $\mathcal{A}(\mathcal{D}(\mathcal{A}))$, in other words, whether $m = \mathcal{A}f$ for some $f \in X$?

Q_5: **Reconstruction from practical data.** Given the noisy data $m = \mathcal{A}f + \varepsilon$, how to approximate f in a noise-robust fashion?