Chapter 3

Deblurring Problem

3.1 Inverse problem

In this chapter, we consider an image deblurring problem as an illustrative example of an inverse problem. Although the forward operator is not a partial differential equation, the problem shares the essential features of typical inverse problems: ill-posedness, the need for regularization, and the challenge of parameter selection. The presentation follows [Ghattas, 2015] and [Vogel, 2002] closely.

Consider the following abstract problem:

$$d = Kp. \tag{3.1}$$

Here, K is an operator mapping p to d, and typically represents a physical process, often expressed as a differential equation. The variable d denotes the *data*, *measurements*, or *observables*. We classify the associated problems as follows:

- *Forward problem*: given K and p, find d.
- *Inverse problem, Type A*: given K and d, find p (e.g., seismic inversion, deblurring problem, and control problem).
- *Inverse problem, Type B*: given p and d, find K (e.g., imaging, system identification, and design problem).

In the inverse problem of Type A, p is the unknown to be reconstructed and is called the *parameter field*; it is typically a function or, after discretization, a finite-dimensional vector. If the inverse operator K^{-1} exists and is readily available, the problem reduces to $p = K^{-1}d$, which is simply a forward solve and is not the focus of these notes. The inverse problems of interest are *ill-posed*, meaning that at least one of the following conditions of well-posedness, due to Hadamard, fails:

- *Existence*: for all data d, there exists a parameter p satisfying $d = Kp$.
- *Uniqueness*: for all data d, the solution p is unique.

- *Stability* (continuous dependence on data): small perturbations in d result in small perturbations in p. Specifically, let p and q be parameters corresponding to data d and f, respectively, i.e., $d = Kp$ and $f = Kq$. Stability requires the existence of a constant $C > 0$ such that

$$\|p - q\|_* \leq C\|d - f\|_{**}. \tag{3.2}$$

3.2 Problem definition

As a model problem, we begin with image deblurring in one dimension. Let $p\colon [0,1] \to \mathbb{R}$ denote the unknown (true) image and $d\colon [0,1] \to \mathbb{R}$ the observed, blurred, and noisy data. The forward model is a Fredholm integral equation of the first kind,

$$\begin{aligned} d(x) &= (Kp)(x) + n(x) \\ &:= \int_0^1 k(x - x')p(x')\,dx' + n(x), \qquad x \in [0,1], \end{aligned} \tag{3.3}$$

where n denotes additive measurement noise and the convolution kernel k is the Gaussian

$$k(x) := C\exp\left(-\frac{x^2}{2\gamma^2}\right), \qquad C,\, \gamma > 0. \tag{3.4}$$

The parameter γ controls the strength of blur: a larger γ yields a wider kernel, so each output value averages p over a broader neighborhood. Equivalently, the action of K can be examined in the frequency domain. Since the Fourier transform of a Gaussian with standard deviation γ is a Gaussian with standard deviation $1/\gamma$, a larger γ narrows the transform of k in frequency, thereby attenuating high-frequency components of p more aggressively. The operator K thus acts as a low-pass filter, and γ governs the cutoff (Figure 3.1).

Fourier transform of a Gaussian $k(x)$ is another Gaussian:

$$\begin{aligned} &Fk\,(\omega) \\ &:= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} k\,(x)\,e^{i\omega x}dx \\ &= \gamma\exp\left(-\frac{\gamma^2\omega^2}{2}\right), \end{aligned}$$

where $k(x) = \exp\left[-x^2/(2\gamma^2)\right]$.

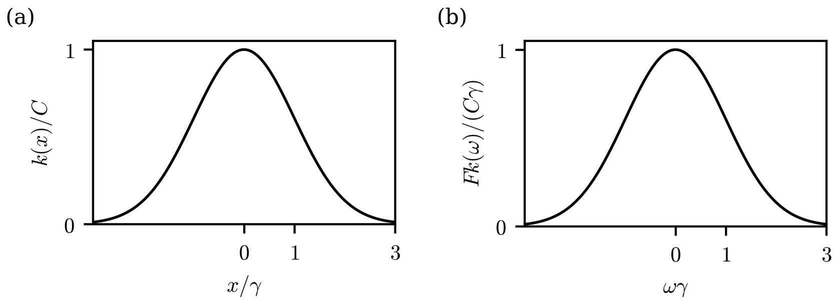

Figure 3.1: (a) Gaussian $k(x) = C\exp\left[-x^2/(2\gamma^2)\right]$ and (b) its Fourier transform $Fk\,(\omega) = C\gamma\exp\left[-\gamma^2\omega^2/2\right]$.

Next, we examine which of the three well-posedness conditions fails for the deblurring problem (3.3). The key structural property of the forward operator K is compactness.

Definition 3.2.1 (Compact operator) *Let X and Y be Banach spaces. A bounded linear operator $K\colon X \to Y$ is* compact *if every bounded sequence $\{p_n\} \subset X$ has a subsequence whose image under K converges in Y. Equivalently, K maps bounded sets in X to precompact sets in Y.*

Namely, the range space of K is smaller than the Y.

A standard class of compact operators is provided by integral operators with square-integrable kernels.

Definition 3.2.2 (Hilbert-Schmidt operator) *Let $k \in L^2((0,1) \times (0,1))$. The integral operator*

$$(Kp)(x) := \int_0^1 k(x,x')p(x')\,dx' \tag{3.5}$$

is called a Hilbert-Schmidt operator*. Every Hilbert-Schmidt operator is compact from $L^2(0,1)$ to $L^2(0,1)$.*

Since the Gaussian kernel (3.26) is smooth and bounded on $[0,1] \times [0,1]$, it belongs to $L^2((0,1) \times (0,1))$, and thus K in (3.3) is a compact operator on $L^2(0,1)$.

Lemma 3.2.1 (Riemann-Lebesgue) *Let $f \in L^1(\mathbb{R})$. Then its Fourier transform vanishes at infinity:*

$$\hat{f}(\omega) := \int_{-\infty}^{\infty} f(x)e^{-i\omega x}\,dx \to 0 \qquad \text{as } |\omega| \to \infty. \tag{3.6}$$

Now, we check the well-posedness of the given operator K:

- Existence: Since K is compact, its range $\mathrm{R}(K)$ is not closed in $L^2(0,1)$. Consequently, there exist data $d \in L^2(0,1)$ for which $Kp = d$ has no solution; existence fails in general.
- Uniqueness: The Gaussian kernel (3.26) is strictly positive, and K is injective on $L^2(0,1)$, i.e., $N(K) = \{0\}$. Hence, if a solution exists, it is unique; uniqueness holds.
- Stability: Consider a perturbed problem $K(p + \delta p) = d + \delta d$, where δp is the high-frequency perturbation

  $$\delta p(x) := \varepsilon \sin(\omega x), \qquad \varepsilon > 0, \quad \omega = 1 \cdot 2\pi, 2 \cdot 2\pi, 3 \cdot 2\pi, \ldots. \tag{3.7}$$

  The corresponding perturbation in the data is

  $$\delta d(x) = \varepsilon \int_0^1 k(x - x')\sin(\omega x')\,dx'. \tag{3.8}$$

  By the Riemann-Lebesgue lemma (Lemma 3.2.1), $\|\delta d\|_{L^2} \to 0$ as $\omega \to \infty$, while $\|\delta p\|_{L^2} = \varepsilon/\sqrt{2}$ remains fixed. Therefore,

  $$\frac{\|\delta p\|_{L^2}}{\|\delta d\|_{L^2}} \to \infty \qquad \text{as } \omega \to \infty, \tag{3.9}$$

  and no finite constant C can satisfy the stability condition. Arbitrarily small perturbations in the data thus correspond to arbitrarily large errors in the recovered parameter.

For numerical purposes, we discretize the problem $d = Kp+n$ to $\mathbf{d} = \mathbf{Kp}+\mathbf{n}$, where $\mathbf{d}, \mathbf{p}, \mathbf{n} \in \mathbb{R}^N$, $\mathbf{K} \in \mathbb{R}^{N\times N}$, and

$$K_{ij} = hC \exp\left(-\frac{(i-j)^2h^2}{2\gamma^2}\right), \qquad 1 \leq i, j \leq N. \tag{3.10}$$

Here, $h = 1/N$ is the mesh spacing, obtained by partitioning $[0, 1]$ into N uniform subintervals and applying the midpoint quadrature rule to the integral (3.3). Figure 3.2 illustrates the action of the Gaussian convolution operator K.

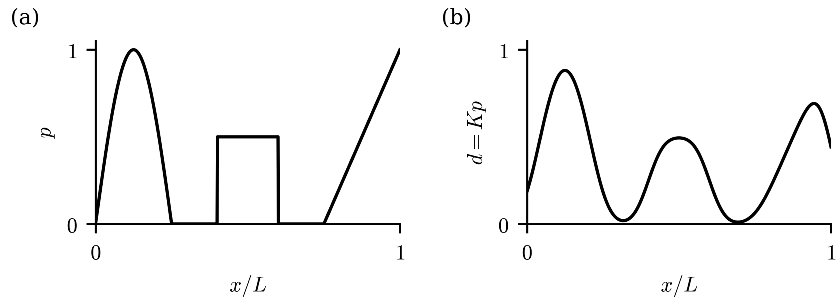

Figure 3.2: Gaussian blur with $C = 1/(\gamma\sqrt{2\pi})$, $\gamma = 0.04$. (a) Original image. (b) Blurred image.

Note that the discretized problem is stable; however, the stability constant grows as N increases. Equivalently, $\mathbf{K}$ becomes increasingly *ill-conditioned* as $N \to \infty$, so that a small noise vector $\mathbf{n}$ can result in large errors in the recovered $\mathbf{p}$. In this sense, the ill-posedness of the original infinite-dimensional problem is reflected in the ill-conditioning of its finite-dimensional discretization, with the *condition number* of $\mathbf{K}$ growing without bound as $N \to \infty$.

Definition 3.2.3 (Condition number) *Let $\mathbf{K} \in \mathbb{R}^{N\times N}$ be a nonsingular matrix. The* condition number *of $\mathbf{K}$ is defined as*

$$\kappa(\mathbf{K}) := \|\mathbf{K}\|\|\mathbf{K}^{-1}\|, \tag{3.11}$$

where $\|\cdot\|$ denotes a matrix norm. When the spectral norm $\|\mathbf{K}\|_2 := \sigma_{\max}(\mathbf{K})$ is used, the condition number takes the form

$$\kappa_2(\mathbf{K}) = \frac{\sigma_{\max}(\mathbf{K})}{\sigma_{\min}(\mathbf{K})}, \tag{3.12}$$

where $\sigma_{\max}$ and $\sigma_{\min}$ are the largest and smallest singular values of $\mathbf{K}$, respectively. A matrix with large $\kappa(\mathbf{K})$ is called ill-conditioned.

This instability has a natural interpretation: the Gaussian kernel acts as a low-pass filter, attenuating high-frequency components of p at a rate controlled by γ in (3.26). Inversion requires amplifying these attenuated components, which amplifies noise catastrophically.

Definition 3.2.4 (Singular Value Decomposition) *Let $\mathbb{F}$ denote either $\mathbb{R}$ or $\mathbb{C}$, and let $\mathbf{K} \in \mathbb{F}^{m\times n}$ with $p := \min\{m, n\}$. There exist unitary matrices*

$$\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \in \mathbb{F}^{m\times m}, \qquad \mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n] \in \mathbb{F}^{n\times n}, \tag{3.13}$$

and a rectangular diagonal matrix $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{m\times n}$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$, such that

$$\mathbf{K} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger = \sum_{i=1}^{p} \sigma_i\, \mathbf{u}_i \mathbf{v}_i^\dagger. \tag{3.14}$$

The σ_i are the singular values *of $\mathbf{K}$; the columns $\mathbf{u}_i$ and $\mathbf{v}_i$ are the* left *and* right singular vectors*, respectively. When $\mathbb{F} = \mathbb{R}$, unitarity reduces to orthogonality and $(\cdot)^\dagger$ reduces to $(\cdot)^T$.*

Unlike the eigendecomposition, which requires $\mathbf{K}$ to be square and may yield complex eigenvalues even for real matrices, the SVD exists for any $\mathbf{K} \in \mathbb{F}^{m\times n}$ and its singular values are always real and nonnegative. For a symmetric positive definite (SPD) matrix $\mathbf{K} \in \mathbb{R}^{n\times n}$, the eigenvalues are real and positive, the SVD coincides with the eigendecomposition, and $\sigma_i = \lambda_i$, $\mathbf{U} = \mathbf{V}$. The discrete convolution matrix (3.10) is SPD, so these simplifications apply throughout.

The SVD also reveals the four fundamental subspaces of $\mathbf{K}$. Let $r \leq p$ denote the rank of $\mathbf{K}$, and partition

$$\mathbf{U} = \begin{bmatrix}\mathbf{U}_r & \mathbf{U}_0\end{bmatrix}, \qquad \mathbf{V} = \begin{bmatrix}\mathbf{V}_r & \mathbf{V}_0\end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}, \tag{3.15}$$

where $\mathbf{U}_r := [\mathbf{u}_1, \ldots, \mathbf{u}_r]$, $\mathbf{U}_0 := [\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m]$, $\mathbf{V}_r := [\mathbf{v}_1, \ldots, \mathbf{v}_r]$, $\mathbf{V}_0 := [\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n]$, and $\boldsymbol{\Sigma}_r := \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$. Then $\mathbf{K} = \mathbf{U}_r\boldsymbol{\Sigma}_r\mathbf{V}_r^\dagger$, and the four fundamental subspaces are

$$\begin{aligned} \mathrm{R}(\mathbf{K}) &= \operatorname{col}(\mathbf{U}_r) \subset \mathbb{F}^m, & \mathrm{N}(\mathbf{K}) &= \operatorname{col}(\mathbf{V}_0) \subset \mathbb{F}^n,\\ \mathrm{R}(\mathbf{K}^\dagger) &= \operatorname{col}(\mathbf{V}_r) \subset \mathbb{F}^n, & \mathrm{N}(\mathbf{K}^\dagger) &= \operatorname{col}(\mathbf{U}_0) \subset \mathbb{F}^m. \end{aligned} \tag{3.16}$$

The columns of $\mathbf{U}$ form an orthonormal basis for $\mathbb{F}^m$ and the columns of $\mathbf{V}$ form an orthonormal basis for $\mathbb{F}^n$. The action of $\mathbf{K}$ is completely characterized: it maps $\mathbf{v}_i \mapsto \sigma_i\mathbf{u}_i$ for $i \leq r$ and annihilates $\mathbf{v}_i$ for $i > r$.

In the context of the deblurring problem, $\mathbf{K} \in \mathbb{R}^{N\times N}$ is SPD and hence invertible, so $r = N$, $\mathbf{U} = \mathbf{V}$, and the null space is trivial; however, the singular vectors $\mathbf{v}_i$ associated with small σ_i correspond to high-frequency modes that are nearly annihilated by $\mathbf{K}$, making their recovery from noisy data unreliable.

3.3 SVD-based approach

Since the discrete convolution matrix $\mathbf{K}$ is SPD, the SVD coincides with the eigendecomposition and $\mathbf{U} = \mathbf{V}$. The naive solution to the inverse problem is

$$\begin{aligned}\mathbf{p} &= \mathbf{K}^{-1}\mathbf{d} \\ &= \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{d} \\ &= \sum_{i=1}^{N} \sigma_i^{-1}\left(\mathbf{u}_i^T\mathbf{d}\right)\mathbf{u}_i \\ &= \mathbf{p}_{\text{true}} + \sum_{i=1}^{N} \sigma_i^{-1}\left(\mathbf{u}_i^T\mathbf{n}\right)\mathbf{u}_i. \end{aligned} \tag{3.17}$$

The last equation shows that the noise contribution $\mathbf{u}_i^T\mathbf{n}$ is amplified by σ_i^{-1}. Since the singular values decay rapidly (Figure 3.3), the high-frequency components of $\mathbf{n}$, which correspond to small σ_i, are amplified catastrophically, rendering the naive inversion unstable.

The remedy is to suppress the contributions from small singular values. The *truncated singular value decomposition* (TSVD) replaces the exact inverse by the filtered expansion

$$\mathbf{p} = \sum_{i=1}^{N} \Omega_{\text{TSVD}}\left(\sigma_i^2\right)\sigma_i^{-1}\left(\mathbf{u}_i^T\mathbf{d}\right)\mathbf{u}_i, \tag{3.18}$$

where the filter function Ω_{TSVD} is defined by

$$\Omega_{\text{TSVD}}\left(\sigma^2\right) := \begin{cases} 1, & \sigma^2 \geq \alpha, \\ 0, & \text{otherwise}, \end{cases} \tag{3.19}$$

and $\alpha > 0$ is the regularization parameter controlling the truncation threshold. Terms corresponding to $\sigma_i^2 < \alpha$ are discarded, eliminating the catastrophic amplification of high-frequency noise at the cost of losing the corresponding signal components.

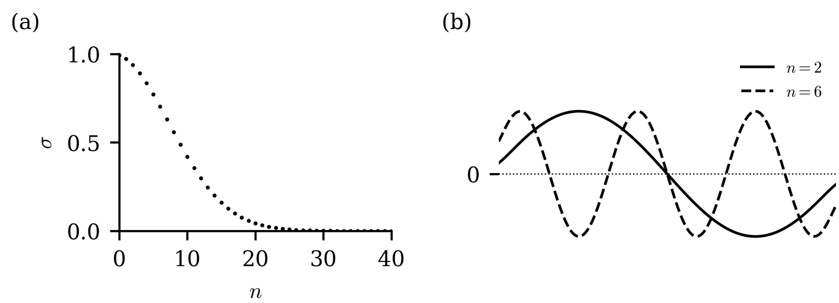

Figure 3.3: Singular value decomposition of $\mathbf{K}$. (a) Singular values. (b) Selected singular vectors.

3.4 Tikhonov regularization

Alternatively, the inverse problem can be formulated as an optimization problem. Given the noisy data $d^{\text{obs}} := Kp_{\text{true}} + n$, find p such that

$$\min_p \; J(p) := \frac{1}{2}\left\|Kp - d^{\text{obs}}\right\|^2 + \frac{\alpha}{2}\|p\|^2, \tag{3.20}$$

where $\alpha > 0$ is the regularization parameter and $\|\cdot\|$ denotes the $L^2(0,1)$ norm,

$$\|a\|^2 = (a,a) := \int_0^1 a(x)^2\,dx. \tag{3.21}$$

The first term penalizes the data misfit. The second term is called *Tikhonov regularization*, which penalizes the L^2 norm of p, with α controlling the trade-off between fidelity to the data and regularity of the reconstruction.

The minimizer is characterized by the stationary point condition $DJ(p)[\tilde{p}] = 0$ for all $\tilde{p} \in L^2(0,1)$. Since $\tilde{p}$ is arbitrary, the Euler equation reads

$$\left(K^\dagger K + \alpha\mathcal{I}\right)p = K^\dagger d^{\text{obs}}, \tag{3.22}$$

where $K^\dagger$ is the L^2-adjoint of K, defined by $(Kp,q) = (p,K^\dagger q)$ for all $p,q \in L^2(0,1)$, and $\mathcal{I}$ is the identity operator. Since K is compact and $\mathcal{I}$ is not, the operator $K^\dagger K + \alpha\mathcal{I}$ is boundedly invertible for any $\alpha > 0$; this is precisely the regularizing effect of the Tikhonov term. For the convolution operator (3.3) with the symmetric Gaussian kernel $k(x) = k(-x)$, K is self-adjoint, i.e., $K^\dagger = K$.

Discretizing the Euler equation and solving for $\mathbf{p}$ gives

$$\left(\mathbf{K}^T\mathbf{K} + \alpha\mathbf{I}\right)\mathbf{p} = \mathbf{K}^T\mathbf{d}^{\text{obs}}, \quad \text{i.e.,} \quad \mathbf{p} = \left(\mathbf{K}^T\mathbf{K} + \alpha\mathbf{I}\right)^{-1}\mathbf{K}^T\mathbf{d}^{\text{obs}}. \tag{3.23}$$

Substituting the SVD $\mathbf{K} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T$ (valid since $\mathbf{K}$ is SPD) and using $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, we obtain

$$\begin{aligned}\mathbf{p} &= \left(\mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}^T + \alpha\mathbf{U}\mathbf{U}^T\right)^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\mathbf{d}^{\text{obs}}\\ &= \mathbf{U}\left(\boldsymbol{\Sigma}^2 + \alpha\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}\,\mathbf{U}^T\mathbf{d}^{\text{obs}}\\ &= \sum_{i=1}^N \frac{\sigma_i}{\sigma_i^2 + \alpha}\left(\mathbf{u}_i^T\mathbf{d}^{\text{obs}}\right)\mathbf{u}_i. \end{aligned}\tag{3.24}$$

This is equivalent to the filtered expansion (3.18) with the *Tikhonov filter function*

$$\Omega_{\text{TIK}}\left(\sigma^2\right) := \frac{\sigma^2}{\sigma^2 + \alpha}, \tag{3.25}$$

which satisfies $\Omega_{\text{TIK}}(\sigma^2) \to 1$ as $\sigma^2 \gg \alpha$ and $\Omega_{\text{TIK}}(\sigma^2) \to 0$ as $\sigma^2 \ll \alpha$, and thus smoothly attenuates the contributions from small singular values. An important practical advantage of Tikhonov regularization over TSVD is that the solution $\mathbf{p}$ can be computed without explicit knowledge of the SVD of $\mathbf{K}$, since the discrete Euler equation is a linear system that can be solved by standard methods.

Exercise 3.4.1 *Show that the Gaussian blur*

$$Kp(x) = \int_0^1 k(x - x')p(x')dx', \quad \textit{where}$$
$$k(x) = C \exp\left(-\frac{x^2}{2\gamma^2}\right), \quad C, \gamma > 0 \tag{3.26}$$

is self-adjoint.

3.5 Regularization parameter choice

The choice of α is critical: too small a value yields an unstable reconstruction, while too large a value suppresses signal and degrades accuracy. We discuss *a posteriori* strategies for selecting α, both of which require solving the regularized problem over a sequence of parameter values.

Let $\mathbf{p}_\alpha$ denote the filtered solution, either by TSVD or Tikhonov regularization, with regularization parameter α, and let $\mathbf{K}_\alpha$ denote the corresponding filtered blur operator such that

$$\mathbf{K}_\alpha^{-1}\mathbf{d} := \mathbf{p}_\alpha = \sum_{i=1}^{N} \Omega_\alpha\big(\sigma_i^2\big)\, \sigma_i^{-1}\, \big(\mathbf{u}_i^T\mathbf{d}\big)\, \mathbf{u}_i. \tag{3.27}$$

The reconstruction error is defined as

$$\begin{aligned}\mathbf{e}_\alpha &:= \mathbf{p}_\alpha - \mathbf{p}_{\text{true}} \\ &= \mathbf{K}_\alpha^{-1}(\mathbf{K}\mathbf{p}_{\text{true}} + \mathbf{n}) - \mathbf{p}_{\text{true}},\end{aligned} \tag{3.28}$$

and is decomposed into a truncation error $\mathbf{e}_\alpha^{\text{trunc}}$ and a noise amplification error $\mathbf{e}_\alpha^{\text{noise}}$:

$$\mathbf{e}_\alpha = \mathbf{e}_\alpha^{\text{trunc}} + \mathbf{e}_\alpha^{\text{noise}}, \tag{3.29}$$

where

$$\begin{aligned}\mathbf{e}_\alpha^{\text{trunc}} &:= \big(\mathbf{K}_\alpha^{-1}\mathbf{K} - \mathbf{I}\big)\, \mathbf{p}_{\text{true}} \\ &= \sum_{i=1}^{N} \big(\Omega_\alpha\big(\sigma_i^2\big) - 1\big)\, \big(\mathbf{u}_i^T\mathbf{p}_{\text{true}}\big)\, \mathbf{u}_i,\end{aligned} \tag{3.30}$$

$$\begin{aligned}\mathbf{e}_\alpha^{\text{noise}} &:= \mathbf{K}_\alpha^{-1}\mathbf{n} \\ &= \sum_{i=1}^{N} \Omega_\alpha\big(\sigma_i^2\big)\, \sigma_i^{-1}\, \big(\mathbf{u}_i^T\mathbf{n}\big)\, \mathbf{u}_i.\end{aligned} \tag{3.31}$$

We show that both errors converge to zero as the noise level

$$\delta := \|\mathbf{n}\| \tag{3.32}$$

goes to zero, for an appropriate choice of α.

For both TSVD and Tikhonov regularization,

$$\Omega_\alpha\big(\sigma^2\big) \to 1 \quad \text{as} \quad \alpha \to 0, \tag{3.33}$$

which immediately implies

$$\|\mathbf{e}_\alpha^{\text{trunc}}\| \to 0 \quad \text{as} \quad \alpha \to 0. \tag{3.34}$$

For the noise amplification error, define the noise amplification factor $A_\alpha(\sigma) := \Omega_\alpha(\sigma^2)\,\sigma^{-1}$. We claim that

$$A_\alpha(\sigma) \leq \frac{1}{\sqrt{\alpha}} \tag{3.35}$$

for all $\sigma > 0$ and $\alpha > 0$.

- *TSVD.* If $\sigma^2 < \alpha$, then $\Omega_\alpha(\sigma^2) = 0$ and the bound holds trivially. If $\sigma^2 \geq \alpha$, then $\Omega_\alpha(\sigma^2) = 1$ and $\sigma \geq \sqrt{\alpha}$, so

$$A_\alpha(\sigma) = \sigma^{-1} \leq \frac{1}{\sqrt{\alpha}}. \tag{3.36}$$

- *Tikhonov.* We seek the supremum of $A_\alpha(\sigma) = \sigma/(\sigma^2 + \alpha)$ over $\sigma > 0$. The stationarity condition reads

$$\frac{dA_\alpha}{d\sigma} = \frac{(\sigma^2+\alpha) - 2\sigma^2}{(\sigma^2+\alpha)^2} = \frac{\alpha - \sigma^2}{(\sigma^2+\alpha)^2} = 0, \tag{3.37}$$

  yielding the unique critical point $\sigma^* = \sqrt{\alpha}$, which is a global maximum. Evaluating,

$$A_\alpha(\sigma^*) = \frac{\sqrt{\alpha}}{2\alpha} = \frac{1}{2\sqrt{\alpha}} \leq \frac{1}{\sqrt{\alpha}}. \tag{3.38}$$

Applying this bound and the orthonormality of $\mathbf{U}$, we obtain

$$\|\mathbf{e}_\alpha^{\text{noise}}\| \leq \frac{1}{\sqrt{\alpha}} \left\| \sum_{i=1}^{N} \left(\mathbf{u}_i^T \mathbf{n}\right) \mathbf{u}_i \right\| = \frac{\delta}{\sqrt{\alpha}}. \tag{3.39}$$

Thus, for $\alpha = \delta^p$ with $0 < p < 2$,

$$\|\mathbf{e}_\alpha^{\text{noise}}\| \to 0 \quad \text{as} \quad \delta \to 0. \tag{3.40}$$

Combining both estimates, the choice $\alpha = \delta^p$ with $0 < p < 2$ guarantees $\|\mathbf{e}_\alpha\| \to 0$ as $\delta \to 0$.

3.5.1 Discrepancy (Morozov) principle

The discrepancy principle selects the largest α such that the data misfit does not exceed the noise level $\delta := \|\mathbf{n}\|$. Defining $M(\alpha) := \|\mathbf{K}\mathbf{p}_\alpha - \mathbf{d}\|$, the criterion requires

$$M(\alpha) \leq \delta. \tag{3.41}$$

Thus, the above criterion avoids overfitting, i.e., fitting the noise rather than the signal.

We can show that for Tikhonov regularization such an α always exists, provided $\delta < \|\mathbf{d}\|$. Using the eigendecomposition $\mathbf{K} = \mathbf{U\Sigma U}^T$, one obtains

$$\mathbf{K}\mathbf{p}_\alpha - \mathbf{d} = \sum_{i=1}^{N} \left(\frac{\sigma_i^2}{\sigma_i^2 + \alpha} - 1 \right) \left(\mathbf{u}_i^T \mathbf{d} \right) \mathbf{u}_i, \tag{3.42}$$

and by orthonormality of $\mathbf{U}$,

$$M^2(\alpha) = \sum_{i=1}^{N} \left(\frac{\sigma_i^2}{\sigma_i^2 + \alpha} - 1 \right)^2 \left(\mathbf{u}_i^T \mathbf{d} \right)^2. \tag{3.43}$$

This expression shows that M is continuous and monotonically increasing in α, with $M(0) = 0$ and $M(\alpha) \to \|\mathbf{d}\|$ as $\alpha \to \infty$. Hence, provided $\delta\|\mathbf{d}\|$, the intermediate value theorem guarantees the existence of α such that $M(\alpha) = \delta$.

3.5.2 L-curve criterion

The L-curve criterion requires no knowledge of the noise level δ. For each α in a prescribed sequence, one solves for $\mathbf{p}_\alpha$ and plots $\|\mathbf{K}\mathbf{p}_\alpha - \mathbf{d}\|$ against $\|\mathbf{p}_\alpha\|$ on a log-log scale. The resulting curve is typically L-shaped, and the selected α corresponds to the point of maximum curvature. This choice balances data fidelity against solution stability: for α below the optimal value, the data misfit decreases only marginally while $\|\mathbf{p}_\alpha\|$ grows substantially; for α above it, $\|\mathbf{p}_\alpha\|$ decreases only marginally while the misfit increases substantially.

Figure 3.4 shows a deblurring example using Tikhonov regularization with varying α. Both the discrepancy principle and the L-curve criterion suggest $\alpha \approx 10^{-2}$, which yields a reasonable reconstruction (panel (g)). However, $\alpha = 10^{-12}$ produces the best visual result (panel (f)), indicating that the a posteriori parameter-choice methods are conservative for this problem. For $\alpha = 10^{-15}$, the regularization is too weak and the reconstruction is dominated by noise amplification (panel (e)). Conversely, $\alpha = 10^0$ over-smooths the reconstruction (panel (h)).

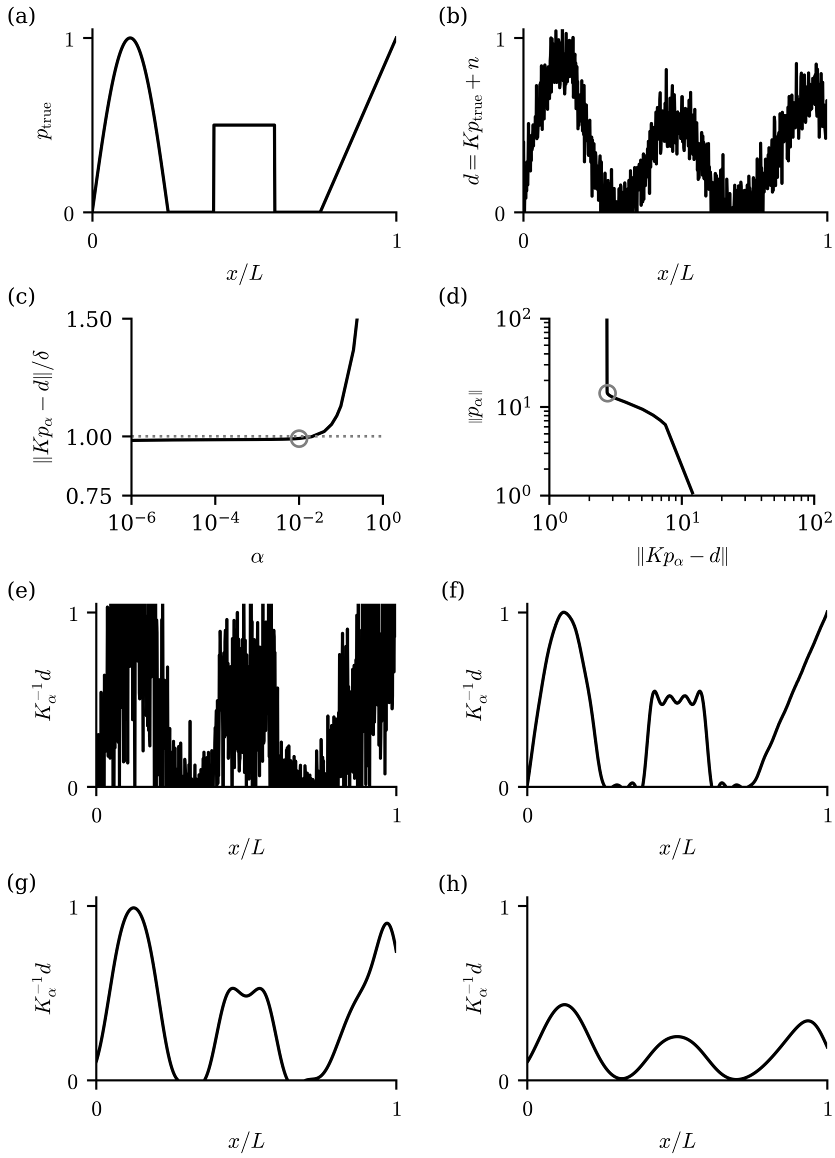

Figure 3.4: Deblurring example. (a) Original image. (b) Blurred image with noise. (c) Discrepancy plot. (d) L-curve plot. (e) Deblurred image with $\alpha = 1.0 \times 10^{-15}$. (f) Deblurred image with $\alpha = 1.0 \times 10^{-12}$. (g) Deblurred image with $\alpha = 1.0 \times 10^{-2}$. (h) Deblurred image with $\alpha = 1.0 \times 10^{0}$.

3.6 Image deblurring in two dimension

Consider a 2D image of standard definition, i.e., $N_x \times N_y = 720 \times 480$ pixels, giving $N = N_x N_y = 345{,}600$ unknowns. The 2D blurring operator $\mathbf{K}$ is an $N \times N$ matrix with $N^2 \approx 1.19 \times 10^{11}$ entries. At 8 bytes per entry (64-bit floating point), explicit storage requires

$$8 \times N^2 \approx 8 \times 1.19 \times 10^{11} \approx 9.5 \times 10^{11} \text{ bytes} \approx 953 \text{ GB}, \tag{3.44}$$

which is infeasible on standard hardware. Moreover, solving the Tikhonov normal equations

$$\left(\mathbf{K}^T\mathbf{K} + \alpha\mathbf{I}\right)\mathbf{p} = \mathbf{K}^T\mathbf{d} \tag{3.45}$$

via Gaussian elimination requires $\mathcal{O}(N^3)$ floating point operations. For $N =$ 345,600, this amounts to

$$N^3 \approx 4.13 \times 10^{16} \text{ flops}, \tag{3.46}$$

which is computationally intractable. Both difficulties are resolved by exploiting the separable structure of the 2D Gaussian kernel: the action of $\mathbf{K}$ on a vector can be computed as successive 1D convolutions in the x- and y-directions. This enables the use of iterative solvers such as the conjugate gradient method, which requires only matrix-vector operation.

The conjugate gradient method solves the symmetric positive definite system

$$\mathbf{A}\mathbf{p} = \mathbf{b}, \qquad \mathbf{A} := \mathbf{K}^T\mathbf{K} + \alpha\mathbf{I}, \qquad \mathbf{b} := \mathbf{K}^T\mathbf{d}, \tag{3.47}$$

iteratively without forming $\mathbf{A}$ explicitly. Starting from an initial guess $\mathbf{p}^{(0)}$, each iteration requires one matrix-vector product with $\mathbf{A}$. The algorithm is as follows [Quarteroni et al., 2006]:

Algorithm 1 Conjugate gradient method for $\mathbf{Ax} = \mathbf{b}$

Set initial value $\mathbf{x}^{(0)}$
$\mathbf{r}^{(0)} := \mathbf{b} - \mathbf{A}\mathbf{x}^{(0)}$ ▷ initial residual
$\mathbf{p}^{(0)} := \mathbf{r}^{(0)}$ ▷ initial search direction
for $k = 0, 1, 2, \ldots$ **do**
  $\alpha^{(k)} := \dfrac{\|\mathbf{r}^{(k)}\|^2}{\left(\mathbf{p}^{(k)}\right)^T \mathbf{A}\,\mathbf{p}^{(k)}}$ ▷ step length
  $\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)}$ ▷ update iterate
  $\mathbf{r}^{(k+1)} := \mathbf{r}^{(k)} - \alpha^{(k)}\mathbf{A}\mathbf{p}^{(k)}$ ▷ update residual
  if $\|\mathbf{r}^{(k+1)}\| \leq \varepsilon\,\|\mathbf{r}^{(0)}\|$ **then**
    break ▷ convergence criterion
  end if
  $\beta^{(k)} := -\dfrac{\|\mathbf{r}^{(k+1)}\|^2}{\|\mathbf{r}^{(k)}\|^2}$ ▷ direction update coefficient
  $\mathbf{p}^{(k+1)} := \mathbf{r}^{(k+1)} - \beta^{(k)}\mathbf{p}^{(k)}$ ▷ update search direction
end for
return $\mathbf{x}^{(k+1)}$

Figures 3.5 and 3.6 illustrate two-dimensional image deblurring. In each case, the original image is convolved with a Gaussian kernel and contaminated with additive noise, producing a severely degraded observation. Tikhonov regularization recovers the essential features of the original in both the binary and grayscale settings.

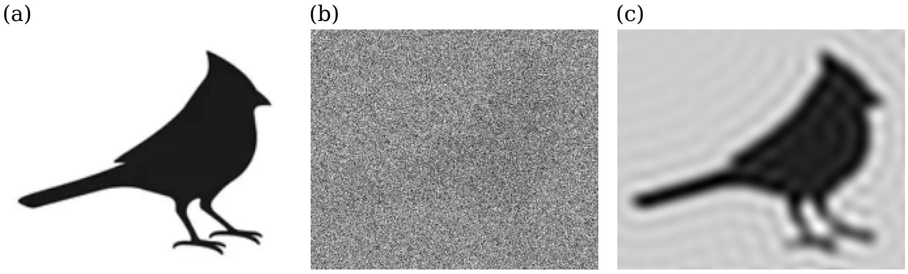

Figure 3.5: Two-dimensional image deblurring with a binary image. (a) Original image (Northern cardinal silhouette). (b) Blurred and noised image. (c) Reconstructed image via Tikhonov regularization.

Figure 3.6: Two-dimensional image deblurring with a grayscale image. (a) Original image (Dia Beacon, Beacon, NY, USA). (b) Blurred and noisy image. (c) Reconstructed image via Tikhonov regularization.