Mathematics for Imaging and Signal Processing

Andrea Aspri - Artificial Intelligence

Reference: *Introduction to Inverse Problems in Imaging (2nd Ed.)*
by Bertero, Boccacci, and De Mol

Contents

1 Lecture 1: Mathematical Foundations and Physics of Imaging

1.1 The Physics of Image Formation: Blurring and Noise

Before deriving algorithms, we must understand the physical processes that degrade an image. QUESTION: What is an image?
ANSWER: An image is a signal carrying information about a physical object, but it is always a *degraded representation*.
We distinguish between two fundamental sources of degradation:

1. **Image Formation Process (Blurring):**
   - **Souce:** This degradation is due to the optical instrument itself (e.g., diffraction in microscopes, atmospheric turbulence in telescopes, relative motion).
   - **Nature:** It is a **deterministic process**. We usually have a sufficiently accurate mathematical model to describe it (e.g., convolution).
   - **Effect:** We will see shortly that it acts as a band-limiting filter, effectively suppressing high-frequency components.
2. **Image Recording Process (Noise):**
   - **Source:** This degradation occurs at the detection stage (e.g., photon counting errors, thermal noise in electronics).
   - **Nature:** It is a **stochastic process**. We do not know the specific noise realization affecting a specific image; we only know its *statistical properties* (e.g., mean, variance, probability distribution).
   - **Effect:** It introduces random fluctuations that corrupt the data, particularly affecting high-frequency components where the signal is often weak, thereby reducing the Signal-to-Noise Ratio (SNR). This will be clarified shortly.

In this course, we focus mainly on **BLURRING** and its inverse problem, called **DECONVOLUTION**.

From Object to Noisy Image

We can formalize this two-step degradation as follows:

1. **The Object (f):** We start with an unknown spatial radiance distribution (the physical reality), which we denote as f. This is what we want to recover.

2. **The Noise-Free Image ($g^{(0)}$):** We apply an imaging system (such as a microscope, telescope, or scanner). If the recording were perfect, we would obtain the *noise-free image*, denoted as $g^{(0)}$.

$$f \xrightarrow{\text{Imaging System}} g^{(0)}$$

3. **The Noisy Image (g):** Finally, a recording system (e.g., a CCD sensor) measures $g^{(0)}$, introducing random errors. The result is the *noisy image* g.

$$g^{(0)} \xrightarrow{\text{Recording System}} g$$

Remark: So, when we talk about the recording system, we often refer to a CCD, which stands for Charge-Coupled Device.

Think of it as the 'digital retina' of our imaging system. Physically, it is a sensor consisting of a grid of independent photo-sensitive elements. You can imagine each element as a small 'bucket' that collects photons during the exposure time.

Why is this important for our mathematical models? First, the CCD is the hardware responsible for discretizing the continuous light signal into a digital matrix of numbers. Second, and most importantly for us, it introduces a specific type of noise. While the arrival of photons follows a Poisson distribution, the electronic process of measuring the charge accumulated in each bucket adds a further error, known as Read-Out Noise (RON). This RON is typically modeled as additive Gaussian noise.

Therefore, in high-precision applications like astronomy, the data is actually corrupted by a mixed Poisson-Gaussian noise, which justifies the use of advanced statistical methods.

1.2 1.2 Mathematical Formulation of the Forward Problem

The Continuous Model

- We model the object f as a function of two variables $x = \{x_1, x_2\}$.

- **Functional Space:** We assume $f \in L^2(\mathbb{R}^2)$, the space of square-integrable functions. This is crucial because it is a **Hilbert space**, allowing us to use tools like

Schematic of Image Formation

Object $f \longrightarrow$ [Imaging System (K)] $\longrightarrow$ Noise-free Image $g^{(0)}$
$\downarrow$
[Recording System (Noise w)]
$\downarrow$
Noisy Image g

Figure 1: The two stages of image degradation: deterministic blurring followed by stochastic noise addition.

the inner product $\langle f, g \rangle$ (to measure similarity) and the norm $||f||_2$ (to measure energy/error).

- The **imaging process** (i.e., the transformation performed by the physical system on the object) is modeled by a linear operator A.

The Point Spread Function (PSF)

The key principle of *optical systems* is how they image a single point.

- A point source at location x is mathematically modeled as a Dirac delta distribution centered at x: $f(x') = \delta(x' - x)$.
- The optical system does not produce a perfect point but a small spot (a "little cloud") with a peak at x.
- This response is called the **Point Spread Function (PSF)** or Impulse Response, denoted by $K(x, x')$.
- This means that PSF describes how the system "spreads" the energy of a point over a region (blurring).

Assumption: Space Invariance (Isoplanatism)

We assume the imaging system is invariant with respect to translations. This means the image of a point source depends only on the relative distance between the source and the sensor point, not on the absolute position of the source.

$$K(x, x_0) = K(x - x_0, 0) := K(x - x_0) \tag{1}$$

We call $K(x)$ the space-invariant PSF. We typically assume it is normalized: $\int_{\mathbb{R}^2} K(x)dx = 1$ (conservation of energy).

Coming back to an image

- **Linearity:** The image of a sum of sources is the sum of their individual images.
- **Space Invariance (Isoplanatism):** The response of the system to a point source does not depend on the position of the source.

Under these assumptions, how do we mathematically model the degradation (blur)? For a linear and shift-invariant (isoplanatic) system, the relationship between the true object f and the noise-free image $g^{(0)}$ is a **convolution integral**:

$$g^{(0)}(x) =: Af(x) = (K * f)(x) = \int_{\mathbb{R}^2} K(x - x')f(x')\, dx' \tag{2}$$

The measured noisy image $g(x)$ is then:

$$g(x) = g^{(0)}(x) + w(x) = Af(x) + w(x) = (K * f)(x) + w(x) \tag{3}$$

where $w(x)$ represents the additive noise term

- $w(x)$ is independent of the noise-free image (often assumed additive Gaussian).
- **Noise Level (ϵ):** In mathematical proofs (e.g., regularization convergence), we assume a bound on the noise energy (norm):

$$\|w\|_{L^2} \leq \epsilon \tag{4}$$

The value of ϵ alone is poorly informative if not compared to the energy of the signal itself.
An example to understand: To understand why the bound on the noise energy, $\|w\|_{L^2} \leq \epsilon$, is poorly informative on its own, let us use a practical analogy.

Let's imagine that the **Noise Level (ϵ)** represents the constant hum of an air conditioner in a room. We fix this noise energy at a specific value, for example, $\epsilon = 10$ units.

Here is why knowing only ϵ is insufficient:

Scenario A: The Whisper (Weak Signal)

Imagine you are trying to listen to someone whispering in that room.

- **Signal Energy ($\|u\|$):** 12 units.

– **Noise Energy (ϵ):** 10 units (the air conditioner).

Outcome: The noise is almost as loud as the signal. You can barely understand the words. The error is massive relative to the data.

$$\text{Ratio} \approx \frac{10}{12} \implies \text{High impact on quality.}$$

Scenario B: The Rock Concert (Strong Signal)

Now, imagine a rock band starts playing in the same room.

– **Signal Energy ($\|u\|$):** 10,000 units.

– **Noise Energy (ϵ):** 10 units (the air conditioner is still running at the same volume).

Outcome: You do not even notice the air conditioner. It is completely irrelevant. Even though ϵ is exactly the same as in Scenario A, here the impact is negligible.

$$\text{Ratio} = \frac{10}{10,000} \implies \text{Zero impact on quality.}$$

Conclusion

This demonstrates why looking at the bound ϵ alone is misleading. An error of 10 is catastrophic for a whisper, but irrelevant for a concert. **SNR (Signal-to-Noise Ratio)** solves this by normalizing the error with respect to the signal strength:

$$\text{Relative Error} = \frac{\epsilon}{\|u\|_{L^2}} \tag{5}$$

SNR solves this problem by normalizing the error with respect to the signal.

1.3 Definition: Signal-to-Noise Ratio (SNR)

The Discrete Model

In computer vision and AI, we deal with digital images. Digital sensors (CCD/CMOS) cannot record continuous functions; they spatially integrate the signal over a grid of pixels. The continuous function g is sampled on a grid of $N \times M$ pixels. In the sequel, we assume that $N = M$ for simplicity.

- **Vectorization:** For algebraic manipulation, we often "vectorize" this matrix into a vector $f \in \mathbb{R}^2$ (where $D = N \times M$) using lexicographic ordering. By lexicographic ordering, we stack the pixel values into vectors $f, g, w \in \mathbb{R}^2$ (where $D = N \times M$).

- **Sampling Model:** The value of the m-th pixel, g_m, is not the value of the function at a point, but a weighted average:

$$g_m = \int_{\text{pixel area}} f(x)\, dx \approx \int p_m(x) f(x)\, dx$$

  where $p_m(x)$ is the *pixel response function* (often modeled as a boxcar function, 1 inside the pixel and 0 outside).
  Analogy to understand: Imagine you want to measure rainfall (representing the continuous image $f(x)$). You cannot measure it at a single infinitesimal point. You need to place buckets (representing the pixels). The water collected in a bucket (g_m) is the sum (the integral) of all the raindrops that fell within the opening of that bucket ($p_m(x)$).

- **Matrix Form:** The convolution operator becomes a matrix A.

- **The Linear System:**

$$\mathbf{g} = A\mathbf{f} + \mathbf{w} \tag{6}$$

The **Signal-to-Noise Ratio (SNR)** is a fundamental metric used to quantify how much the useful signal (the noise-free image) prevails over the noise corrupting it.

1. General Formula

SNR is defined as the ratio between the mean value of the signal (image) and the standard deviation of the noise. Considering an image discretized on a grid of $N \times N$ pixels, where $g_{m,n}$ represents the pixel value, the formula is:

$$\text{SNR} = \frac{\text{Mean Image Value}}{\text{Noise Standard Deviation}} = \frac{\frac{1}{N^2}\sum_{m,n=1}^{N} g_{m,n}}{\sigma} \tag{7}$$

Where:

- **Numerator:** It is the mean of the pixel intensity values. In the case of zero-mean noise, the mean of the noisy image coincides with that of the noise-free image.

- σ **(Sigma):** It is the standard deviation (i.e. a statistical measure of the noise variability) of the noise (the square root of the variance).

$$\sigma = \sqrt{\mathrm{Var}(w)} = \sqrt{\mathbb{E}[(w - \mathbb{E}[w])^2]} \tag{8}$$

It represents the average "spread" or magnitude of the noise fluctuations away from the true signal.

Remark: In a real-world scenario, w is unknown. However, in a **simulation**, we have a significant advantage: we know the ground truth $g^{(0)}$. This allows us to invert the equation and **mathematically isolate** the specific noise realization for every single pixel:

$$w_{m,n} = g_{m,n} - g^{(0)}_{m,n} \tag{9}$$

At this point, the noise is no longer an unknown random variable mixed with the signal; it becomes a **dataset of observable values**.

Assume the noise has zero mean (standard for Gaussian noise). Here lies the crucial mathematical step. We cannot compute the theoretical expectation $\mathbb{E}$ because we cannot generate infinite versions of the image to average over. However, we have a large sample size: the image itself contains N^2 pixels.

Thanks to the **Law of Large Numbers**, we can approximate the theoretical expected value with the **sample mean** (the arithmetic average) calculated over all the pixels in the image. We replace the expectation operator $\mathbb{E}[\cdot]$ with the spatial average, i.e. the **Mean Squared Error (MSE)**:

$$\mathbb{E}[w^2] \approx \frac{1}{N^2} \sum_{m,n} (w_{m,n})^2 = \frac{1}{N^2} \sum_{m,n=1}^{N} |g_{m,n} - g^{(0)}_{m,n}|^2 =: MSE \tag{10}$$

Therefore

$$\sigma =: RMSE = \sqrt{\frac{1}{N^2} \sum_{m,n} (g_{m,n} - g^{(0)}_{m,n})^2} \tag{11}$$

Remark: In real-world applications (e.g., astronomy, microscopy), we do not know $g^{(0)}$. We estimate σ by selecting a background region Ω_{bg} of the image that is known to be featureless (constant). The standard deviation of the pixel values in this region provides an estimate of σ:

$$\sigma \approx \sqrt{\frac{1}{P-1} \sum_{(m,n)\in\Omega_{bg}} (g_{m,n} - \bar{g}_{bg})^2} \tag{12}$$

where P is the number of pixels in the background region and $\bar{g}_{bg}$ is their mean.

2. Decibel Scale (dB)

Since images can have a very wide dynamic range, SNR is almost always expressed in a logarithmic scale using decibels (dB). The conversion is given by:

$$(\text{SNR})_{\text{dB}} = 20 \log_{10}(\text{SNR}) \tag{13}$$

Note: Sometimes you can find a factor 10 instead of 20. Here, the multiplicative factor is **20** because we are working with signal amplitudes (pixel intensities), not powers.

- **Noise Visibility:** Empirically, the human eye struggles to distinguish noise when the SNR is above 40 dB, while below this threshold, noise becomes progressively more visible.
- **Local vs. Global SNR:** SNR can vary across different regions of the image (e.g., it is higher in bright areas and lower in dark ones). Although the formula above calculates a global value, it is possible to calculate local SNRs on specific regions of interest.

1.4 Definition: Peak Signal-to-Noise Ratio (PSNR)

While SNR relates the noise to the *average* power of the signal, the **Peak Signal-to-Noise Ratio (PSNR)** relates the error to the maximum possible value (dynamic range) of the signal. It is the standard metric used to evaluate the quality of a reconstruction in simulations where the ground truth $g^{(0)}$ is known.

1. General Formula

The PSNR is defined as:

$$\text{PSNR} = \frac{\text{Peak Value}}{\text{RMSE}} = \frac{\max_{m,n} g^{(0)}_{m,n}}{\sqrt{\text{MSE}}} \tag{14}$$

Where:

- **Peak Value:** It is the maximum fluctuation of the signal. For digital images, this is determined by the bit depth (e.g., 255 for 8-bit images) or is 1.0 for normalized floating-point images.
- **RMSE:** The Root Mean Squared Error, representing the average magnitude of the error (noise) per pixel.

3. Comparison with SNR

- **Independence from Image Content:** SNR depends on the mean value of the image; a dark image will have a low SNR even with low noise. PSNR uses the *peak* value (a constant of the system), making it a more absolute measure of image fidelity relative to the sensor's dynamic range.
- **Interpretation:** Higher PSNR indicates better quality. In image restoration:
  - **20-25 dB:** Acceptable quality for wireless transmission.
  - **30-40 dB:** Good to excellent quality (typical of high-quality compression or good denoising).

1.5 The Fourier Perspective

Using the Fourier Transform (FT), the convolution becomes multiplication (i.e. the convolution product is diagonalized by the Fourier transform).

- **Convolution Theorem:** Convolution in the spatial domain corresponds to pointwise multiplication in the frequency domain.

$$\widehat{(K * f)}(\omega) = \hat{K}(\omega) \cdot \hat{f}(\omega) \tag{15}$$

This turns the integral equation into a simpler algebraic equation and to analyze how the system processes spatial frequencies.

$$\hat{g}(\omega) = \hat{K}(\omega)\hat{f}(\omega) + \hat{w}(\omega) \tag{16}$$

where:

- $\hat{h}(\omega) = \int_{\mathbb{R}^2} h(x) e^{-i\omega \cdot x} dx$ is the Fourier Transform.
- $\hat{K}(\omega)$ is the so-called **Transfer Function (TF)**.

Recap: Some Properties of the Fourier Transform

1. **Riemann-Lebesgue Theorem:** If f is absolutely integrable (L^1), then $\hat{f}(\omega)$ is bounded, and tends to 0 as $|\omega| \to \infty$.

2. **Integrability Issue:** The FT of an integrable function is *not necessarily* integrable. This means that while the spatial function is well-behaved (finite area), its spectrum might decay so slowly that it is not absolutely integrable.

   Examples:

   - **1D Case (Box Function):** Let $\chi_D(x)$ be the characteristic function of the interval $D = [-\pi, \pi]$. Its Fourier Transform is the *sinc* function:

     $$\hat{\chi}_D(\omega) = 2\pi \frac{\sin(\pi\omega)}{\pi\omega} = 2\pi \,\mathrm{sinc}(\omega) \tag{17}$$

     The sinc function is not absolutely integrable (its integral diverges).

   - **2D Case (Circular Aperture):** Let $\chi_D(x)$ be the characteristic function of the disk $D = B_\pi(0)$ (disk of radius π). Its Fourier Transform involves the Bessel function of the first kind of order 1, J_1:

     $$\hat{\chi}_D(\omega) = 2\pi^2 \frac{J_1(\pi|\omega|)}{|\omega|} \tag{18}$$

     Similarly, this function is not absolutely integrable due to the slow decay of the Bessel function ($\sim |\omega|^{-3/2}$).

3. L^2 **Isometry:** If f is square-integrable (L^2), then $\hat{f}$ is also in L^2.

4. **Parseval's Identity:**

   $$\int |f(x)|^2 dx = \frac{1}{(2\pi)^d} \int |\hat{f}(\omega)|^2 d\omega, \qquad d = \text{space dimension} \tag{19}$$

   Energy is conserved. This allows us to define Least Squares (we will see it later) in the frequency domain.

Remark:

- **Physical Implication of Riemann-Lebesgue:** For integrable functions (like physical PSFs), $\hat{K}(\omega) \to 0$ as $|\omega| \to \infty$.
  - This implies that Optical systems act as **Low-Pass Filters**. They transmit low frequencies (structure) well but attenuate or block high frequencies (corresponding to fine details/edges).

1.6 Real-world Point Spread Function (PSF)

In real-world applications, the specific shape of the kernel $K(x)$ is dictated by the physical process causing the degradation. We analyze the three most fundamental models found in photography, microscopy, and astronomy.

1. Photography: Linear Motion Blur

This blur occurs when there is relative motion between the camera and the object during the exposure interval T.

- **Physics:** Assuming a constant velocity v, the object "sweeps" a path of length $L = vT$ on the sensor.
- **Spatial Model:** The PSF is uniform along the direction of motion (a 1D boxcar function embedded in 2D). If the motion is along the x_1 axis:

$$K(x_1, x_2) = \begin{cases} \frac{1}{L} & \text{if } 0 \leq x_1 \leq L \text{ and } x_2 = 0 \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

- **Frequency Analysis:** The Fourier Transform of a boxcar is a **Sinc function**.

$$\hat{K}(\omega) = e^{-i\frac{L}{2}\omega_1}\text{sinc}\left(\frac{L\omega_1}{2\pi}\right) = e^{-i\frac{L}{2}\omega_1}\frac{\sin(L\omega_1/2)}{L\omega_1/2} \tag{21}$$

  Consequence: The Sinc function oscillates and has **zeros** whenever $L\omega_1/2 = n\pi$ (for $n \in \mathbb{Z} \setminus \{0\}$).

  - At these specific frequencies (parallel lines in the Fourier plane), $\hat{K}(\omega) = 0$.
  - **Information Loss:** The imaging system completely destroys information at these frequencies. $\hat{g}(\omega) = 0 \cdot \hat{f}(\omega) + \text{noise}$. We cannot recover $\hat{f}$ at these points (division by zero).

Figure 2: Original image (left). A linear motion blur is applied (with added Gaussian noise), producing the degraded observation (center). Right: the phase field in the Fourier domain, showing the characteristic frequency-domain signature of motion blur.

2. Microscopy: Out-of-Focus Blur (Defocus)

This occurs when the object plane is not conjugate to the sensor plane according to the lens equation $\frac{1}{d_0} + \frac{1}{d_i} = \frac{1}{d_f}$. Remember that

- d_0 is the distance of the object from the sensor;
- d_1 is the distance between the image plane and lens;
- d_f focal length of the lens.
- **Physics:** According to Geometrical Optics, a point source is imaged as a uniform disc called the *Circle of Confusion* (COC) with radius R.
- **Spatial Model:**

$$K(x) = \begin{cases} \frac{1}{\pi R^2} & \text{if } |x| \leq R \\ 0 & \text{if } |x| > R \end{cases} \tag{22}$$

- **Frequency Analysis:** The Fourier Transform of a disc is described by the Bessel function of the first kind of order 1, denoted as J_1:

$$\hat{K}(\omega) = 2\frac{J_1(R|\omega|)}{R|\omega|} \tag{23}$$

  Consequence: Like the Sinc function, the Bessel function oscillates and has infinite **zeros**. These zeros form concentric circles in the frequency domain at radii $\omega_n \approx \frac{n\pi}{R}$.

  – Information is lost on these rings.

  – The decay of the envelope is $\sim |\omega|^{-3/2}$, meaning high frequencies are severely attenuated even where $\hat{K} \neq 0$.

Focus: Why Defocus Creates "Holes" in Information

To understand why out-of-focus blur creates irreversible data loss, we must analyze the phenomenon from both physical and mathematical perspectives.

1. The Physical View: The Circle of Confusion

In the spatial domain, defocus is described by geometrical optics. When a point source is not in focus, light rays do not converge to a single point on the sensor but expand to form a uniform luminous disc. This is technically called the **Circle of Confusion (COC)**.

Mathematically, the Point Spread Function $K(x)$ becomes the indicator function of this disc, normalized to conserve energy:

$$K(x) = \frac{1}{\pi R^2}\chi_{\text{coc}}(x) \tag{24}$$

where R is the radius of the disc. Spatial information is "lost" because fine details are "smeared" over a circular area, overlapping and blurring into one another.

2. The Mathematical View: Zeros of the Transform

The rigorous explanation for information loss is found in the frequency domain. The Fourier Transform of a uniform disc is described by the Bessel function of the first kind of order 1, denoted as J_1:

$$\hat{K}(\omega) = 2\frac{J_1(R|\omega|)}{R|\omega|} \tag{25}$$

The problem arises from the specific properties of this function:

- **Oscillation and Zeros:** The function J_1, much like the *sinc* function in 1D, oscillates and takes the value **zero** at specific points. In the 2D frequency plane, these zeros form a series of concentric circles (dark rings in the spectrum) at radii $\omega_n \approx \pi x_n/R$.
- **Signal Cancellation:** Recall the fundamental relation in the frequency domain:

  $$\hat{g}(\omega) = \hat{K}(\omega) \cdot \hat{f}(\omega)$$

  At the specific frequencies where the Transfer Function $\hat{K}(\omega)$ is zero, the object's contribution $\hat{f}(\omega)$ is multiplied by zero.
- **Consequence:** At those specific spatial frequencies, information about the original object is completely erased. Mathematically, we cannot recover $\hat{f}(\omega)$ because it would require dividing by zero.

Furthermore, even in the regions between the zeros, the function decays rapidly (proportional to $|\omega|^{-3/2}$). This means high-frequency details (sharp edges) are attenuated so strongly that they quickly fall below the noise floor, making recovery practically impossible even where $\hat{K}(\omega) \neq 0$.

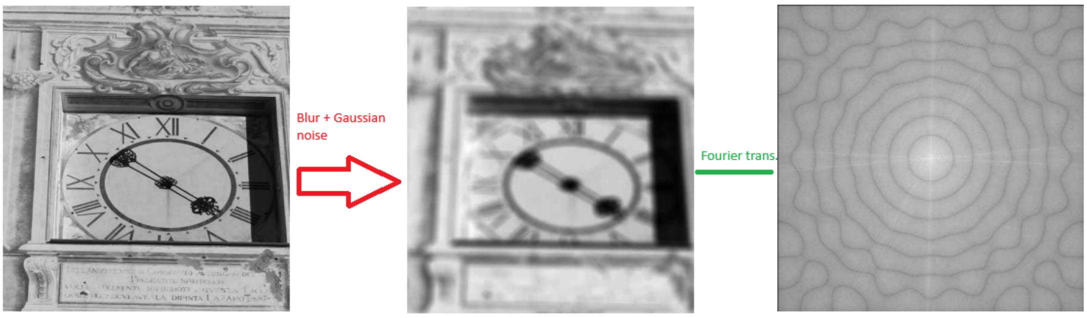

Figure 3: Original image (left). An out-of-focus (defocus) blur is applied, yielding the blurred observation (center). Right: the phase field in the Fourier domain, illustrating how defocus alters the frequency content of the image.

3. Diffraction-Limited Blur (astronomy)

Even with a perfectly focused, stationary, and aberration-free lens, the resolution is limited by the wave nature of light. This is the fundamental limit of microscopes and telescopes. To understand the origin of the Airy Pattern, we must look at the physics of diffraction.

- **The Physics of the Aperture (The Pupil Function):** Light rays carrying information about fine details (high spatial frequencies) travel at steeper angles than rays carrying coarse information (low frequencies).

  The lens has a finite physical aperture (the **exit pupil**).

  - Low-angle rays (low frequencies) pass through the aperture.
  - Steep-angle rays (high frequencies) miss the aperture and are blocked.

  Because spatial position on the lens pupil maps directly to spatial frequency ω (via the relation $\omega \propto x_{\text{pupil}}/\lambda$), we model the physical aperture as a set P in the frequency domain. The system acts as a "brick-wall" filter: it allows light to pass only if its corresponding spatial frequency lies within the set P.

  We define the **Pupil Function** $\chi_P(\omega)$ as the characteristic function of this set:

$$\chi_P(\omega) = \begin{cases} 1 & \text{if } \omega \in P \quad \text{(light passes)} \\ 0 & \text{if } \omega \notin P \quad \text{(light is blocked)} \end{cases} \tag{26}$$

- **From Coherent Amplitude to Incoherent Intensity:**

  1. **Coherent Response (Amplitude):** If we consider the complex amplitude of the electric field (coherent light), the impulse response $h(x)$ is the Inverse Fourier Transform of the pupil function:

$$h(x) = \mathcal{F}^{-1}[\chi_P](\omega) \tag{27}$$

     For a circular aperture of radius Ω (in frequency units), the inverse transform of the disk function χ_P is analytically given by the Bessel function term:

$$K(x) = h(x) \propto \frac{J_1(\Omega|x|)}{\Omega|x|} \tag{28}$$

2. **Incoherent Response (Intensity):** Most imaging systems (cameras, fluorescence microscopes) measure light *intensity*, not complex amplitude. This is the case of **Spatially Incoherent Illumination**. The PSF for intensity is the squared modulus of the amplitude response:

$$K_{inc}(x) = |h(x)|^2 \tag{29}$$

- **The Spatial Model (Airy Pattern):** Combining the steps above, for a circular aperture, we obtain the famous **Airy Pattern**:

$$K_{inc}(x) = \left(\frac{2J_1(\Omega|x|)}{\Omega|x|} \right)^2 \tag{30}$$

  Visually, this is a bright central spot (the Airy disk) surrounded by faint concentric diffraction rings.

- **Frequency Analysis (Band-Limited):** In the frequency domain, the Transfer Function (OTF) of the incoherent system is the autocorrelation of the pupil function χ_P.
  - Since χ_P has finite support (the aperture), its autocorrelation also has finite support.
  - **Cut-off Frequency:** There exists a strict frequency Ω_c (twice the aperture radius) beyond which the OTF is exactly zero.

$$\hat{K}(\omega) = 0 \quad \text{for all } |\omega| > \Omega_c \tag{31}$$

- **Consequence:** The system acts as a perfect **Low-Pass Filter**. Information about object details smaller than the diffraction limit (related to the Rayleigh distance $1.22\pi/\Omega$) is completely cut off. Unlike motion or defocus, where information is hidden in the zeros, here high-frequency information is non-existent in the data.

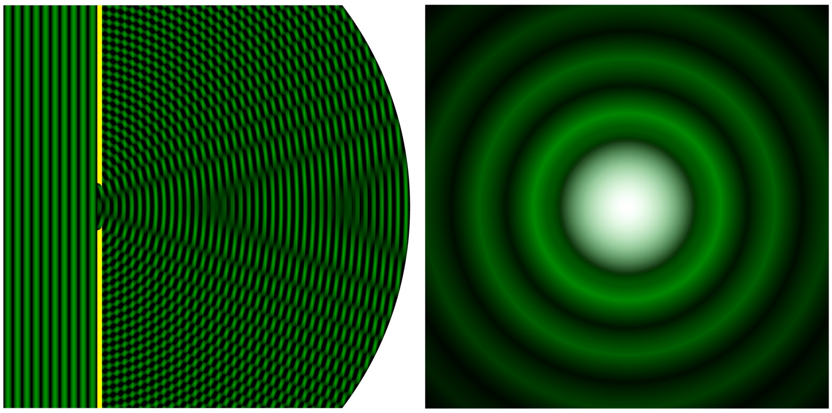

Figure 4: **The Airy Pattern (Diffraction-Limited PSF).** Visual representation of the Point Spread Function for a circular aperture. The image shows a bright central spot, known as the *Airy Disk*, which contains approximately 84% of the total light energy. It is surrounded by fainter concentric diffraction rings. Crucially, the dark rings (where intensity drops to zero) correspond to the zeros of the Bessel function J_1 in the frequency domain; at these specific spatial frequencies, information about the object is completely lost.

1.7 Ill-Posedness of Deconvolution and the Failure of Naive Inversion

From $\hat{g} = \hat{K}\hat{f} + \hat{w}$, If we try to invert the model of blurring in the Fourier domain, we need to solve the following inverse problem:

$$\hat{f}_{rec}(\omega) = \frac{\hat{g}(\omega)}{\hat{K}(\omega)} = \hat{f}(\omega) + \frac{\hat{w}(\omega)}{\hat{K}(\omega)} \tag{32}$$

According to Hadamard, a problem is **Well-Posed** if:

1. A solution exists.
2. The solution is unique.
3. The solution depends continuously on the data (stability).

We will now show that the deconvolution (deblurring) problem is ill-posed.

1.7.1 Uniqueness and "Invisible Objects"

Is the solution unique? Suppose we have two solutions f_1, f_2 such that $Af_1 = g$ and $Af_2 = g$. Then $A(f_1 - f_2) = 0$. Uniqueness holds if the null-space $\mathcal{N}(A) = \{0\}$. In the Fourier domain: $\hat{K}(\omega)\hat{f}(\omega) = 0$.

- If the support of $\hat{K}$ is $\mathbb{R}^2$ (e.g. Motion blur, Defocus), then $\hat{f} = 0 \implies f = 0$. Uniqueness holds.
- If the system is **Band-Limited** (e.g., Diffraction-limited optical systems), then $\hat{K}(\omega) = 0$ for $\omega \notin B$ (the band). Any function f whose spectrum lies outside B is an **"invisible object"**. Uniqueness fails.

1.7.2 Existence and Stability (The Real Problem)

Even if uniqueness holds, existence and stability are problematic. Using Inverse Filtering:

$$\hat{f}_{rec}(\omega) = \frac{\hat{g}(\omega)}{\hat{K}(\omega)} = \hat{f}(\omega) + \frac{\hat{w}(\omega)}{\hat{K}(\omega)} \tag{33}$$

- **Zeros of $\hat{K}$:** If $\hat{K}(\omega^*) = 0$ (e.g., motion blur zeros), we divide by zero. Since noise $\hat{w}$ is random, likely $\hat{w}(\omega^*) \neq 0$. The solution could not exist in L^2.
- **Decay at Infinity:** Even if $\hat{K} \neq 0$, Riemann-Lebesgue says $\hat{K}(\omega) \to 0$ as $|\omega| \to \infty$. The noise $\hat{w}$ is usually white (constant power). Thus, the ratio $\frac{\hat{w}}{\hat{K}}$ diverges. Small noise in high frequencies is amplified to infinity.

The above considerations show that the deconvolution (deblurring) problem is, in general, **ill-posed**.

1.7.3 Discretization: From Ill-Posed to Ill-Conditioned

Let us consider for a moment the discretized version of the continuous problem studied above

$$\mathbf{g} = A\mathbf{f}, \qquad \text{and} \qquad \mathbf{g}^\delta = \mathbf{g} + \mathbf{w} \tag{34}$$

Here $\mathbf{w}$ is the (unknown) noise vector, and the superscript δ indicates a contaminated measurement by noise.

Well-posedness in finite dimensions (when A is invertible). In the discretized version, $A \in \mathbb{R}^{N\times N}$ is the matrix representation of the continuous convolution operator restricted to a *finite set of discrete frequencies* (those available on the chosen sampling grid). Because discretization involves only finitely many frequencies, it may happen that the transfer function of the blur, when sampled on this discrete frequency grid, is *nonzero at all sampled frequencies* (and therefore some of the pathologies observed in the continuous setting—for instance the black lines or circles for linear motion blur and out-of-focus blur respectively—need not necessarily occur at the discrete level). Moreover, small modeling and numerical approximation effects (e.g., discretization, rounding) can prevent exact zeros from appearing at the discrete level. Therefore in some situations, the corresponding matrix A is not singular, that is invertible. If A is invertible, then for *every* right-hand side $\mathbf{g}^\delta \in \mathbb{R}^N$ there **exists a unique solution** (so the problem is well-posed)

$$\mathbf{f}^\delta = A^{-1}\mathbf{g}^\delta,$$

and the map $\mathbf{g}^\delta \mapsto \mathbf{f}^\delta$ is continuous. In particular, defining

$$\delta\mathbf{g} := \mathbf{g}^\delta - \mathbf{g} = \mathbf{w}, \qquad \delta\mathbf{f} := \mathbf{f}^\delta - \mathbf{f}.$$

for any perturbation $\delta\mathbf{g}$ one has

$$\delta\mathbf{f} := \mathbf{f}^\delta - \mathbf{f} = A^{-1}\mathbf{g}^\delta - A^{-1}\mathbf{g} = A^{-1}(\mathbf{g}^\delta - \mathbf{g}) = A^{-1}\delta\mathbf{g}. \tag{35}$$

Warning. *Well-posedness does not guarantee a meaningful reconstruction.* Even if the discrete system admits a unique solution (i.e., A is invertible), the matrix A is often **severely ill-conditioned**. In this regime, small perturbations in the data (most notably measurement noise) are strongly amplified by the inverse, so the computed solution may be dominated by artifacts and high-frequency oscillations rather than by the true underlying image. This is the key connection between deblurring and ill-conditioning: invertibility ensures existence and uniqueness, whereas conditioning determines whether the inversion is numerically stable and physically reliable. Before going into the details of the ill-conditioning, let us recall some important notions on singular values:

Mathematical Digression: Singular Values

Since the concept of ill-conditioning revolves entirely around singular values, let us briefly recall what they are. The **Singular Value Decomposition (SVD)** is a fundamental tool in linear algebra that generalizes diagonalization to any matrix (including rectangular ones).

1. Algebraic Definition

Let $A \in \mathbb{R}^{N\times M}$. Then A admits an SVD of the form

$$A = U\Sigma V^T, \tag{36}$$

Singular Values

where

- $U \in \mathbb{R}^{N\times N}$ and $V \in \mathbb{R}^{M\times M}$ are **orthogonal** matrices (their columns are the left and right singular vectors, respectively);
- $\Sigma \in \mathbb{R}^{N\times M}$ is a (rectangular) **diagonal** matrix, whose diagonal entries are the singular values.

Denoting by $R = \text{rank}(A) \le \min\{N, M\}$ the rank of A, the nonzero singular values satisfy

$$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_R > 0,$$

while $\sigma_{R+1} = \cdots = 0$ if $R < \min\{N, M\}$. Using the singular vectors $\{\mathbf{u}_i\}_{i=1}^N$ (columns of U) and $\{\mathbf{v}_i\}_{i=1}^M$ (columns of V), one can write the expansion

$$A = \sum_{i=1}^{R} \sigma_i \, \mathbf{u}_i \mathbf{v}_i^T. \tag{37}$$

2. Geometric Meaning (Gain)

In an imaging system, σ_i represents the **gain** (or transmission factor) of the system along the direction $\mathbf{v}_i$:

- If $\sigma_i \approx 1$, the corresponding component is transmitted almost unchanged.
- If $\sigma_i \ll 1$ (typically for high-frequency components), that component is strongly attenuated.

3. The Matrix 2-Norm and Singular Values

The **matrix 2-norm** (or **spectral norm**) of A is defined as the largest amplification factor that A can produce on a vector:

$$\|A\|_2 := \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}. \tag{38}$$

A key fact is that this norm is completely characterized by the singular values:

$$\|A\|_2 = \sigma_1.$$

Singular Values

Moreover, if A is square and invertible (so that $N = M$ and $R = N$), then

$$\|A^{-1}\|_2 = \frac{1}{\sigma_N},$$

i.e., the inverse has large norm precisely when the smallest singular value is small. In this course, when we discuss the decay of the Transfer Function $|\hat{K}(\omega)|$, we are essentially discussing the decay of the singular values.

Ill-conditioning is *precisely* the fact that $\|A^{-1}\|_2$ can be very large. Using the singular values of A, we have

$$\|A^{-1}\|_2 = \frac{1}{\sigma_N},$$

so (35) becomes

$$\|\mathbf{f}^\delta - \mathbf{f}\| \leq \frac{1}{\sigma_N}\|\mathbf{w}\|.$$

Hence, even if the noise $\|\mathbf{w}\|$ is small, the error can be large whenever σ_N is very small. A common relative version uses the **Condition Number**

$$\kappa(A) := \|A\|_2 \, \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_N},$$

and gives (for small perturbations) the rule of thumb

$$\frac{\|\mathbf{f}^\delta - \mathbf{f}\|}{\|\mathbf{f}\|} \lesssim \kappa(A) \, \frac{\|\mathbf{g}^\delta - \mathbf{g}\|}{\|\mathbf{g}\|} = \kappa(A) \, \frac{\|\mathbf{w}\|}{\|\mathbf{g}\|}.$$

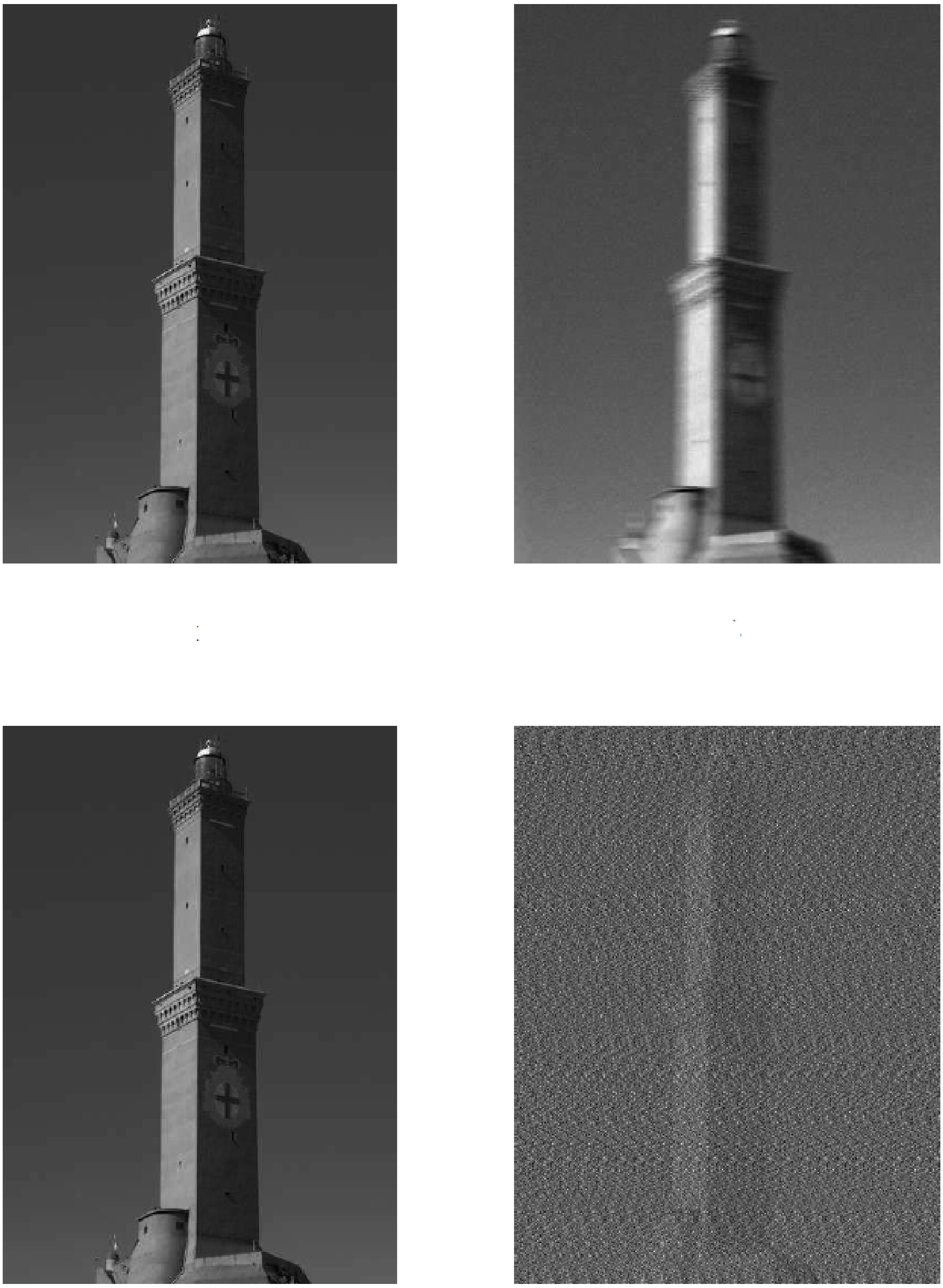

Figure 5: Effect of noise in discrete deblurring. **Top-left:** the original image $\mathbf{f}^{(0)}$. **Top-right:** the blurred image $\mathbf{g}^{(0)} = A\mathbf{f}^{(0)}$. **Bottom-left:** reconstruction obtained by inverting the blur when no noise is added. **Bottom-right:** reconstruction obtained by inverting the blur from noisy data $\mathbf{f}^{(0)} + \mathbf{w}$.

Figure 5 highlights the main practical issue of discrete deblurring. In the noise-free case (bottom-left), the inversion can recover a visually meaningful approximation of

$\mathbf{f}^{(0)}$. However, as soon as noise is present in the data (bottom-right), the reconstruction quality deteriorates dramatically: since the problem is *ill-conditioned*, the inverse operator amplifies the noise, producing a poor and highly artifact-contaminated image.
Take-home message: the discrete problem can be *well-posed* (because A^{-1} exists in finite dimensions), but still *ill-conditioned* because A^{-1} may have a huge norm, which greatly amplifies the noise.

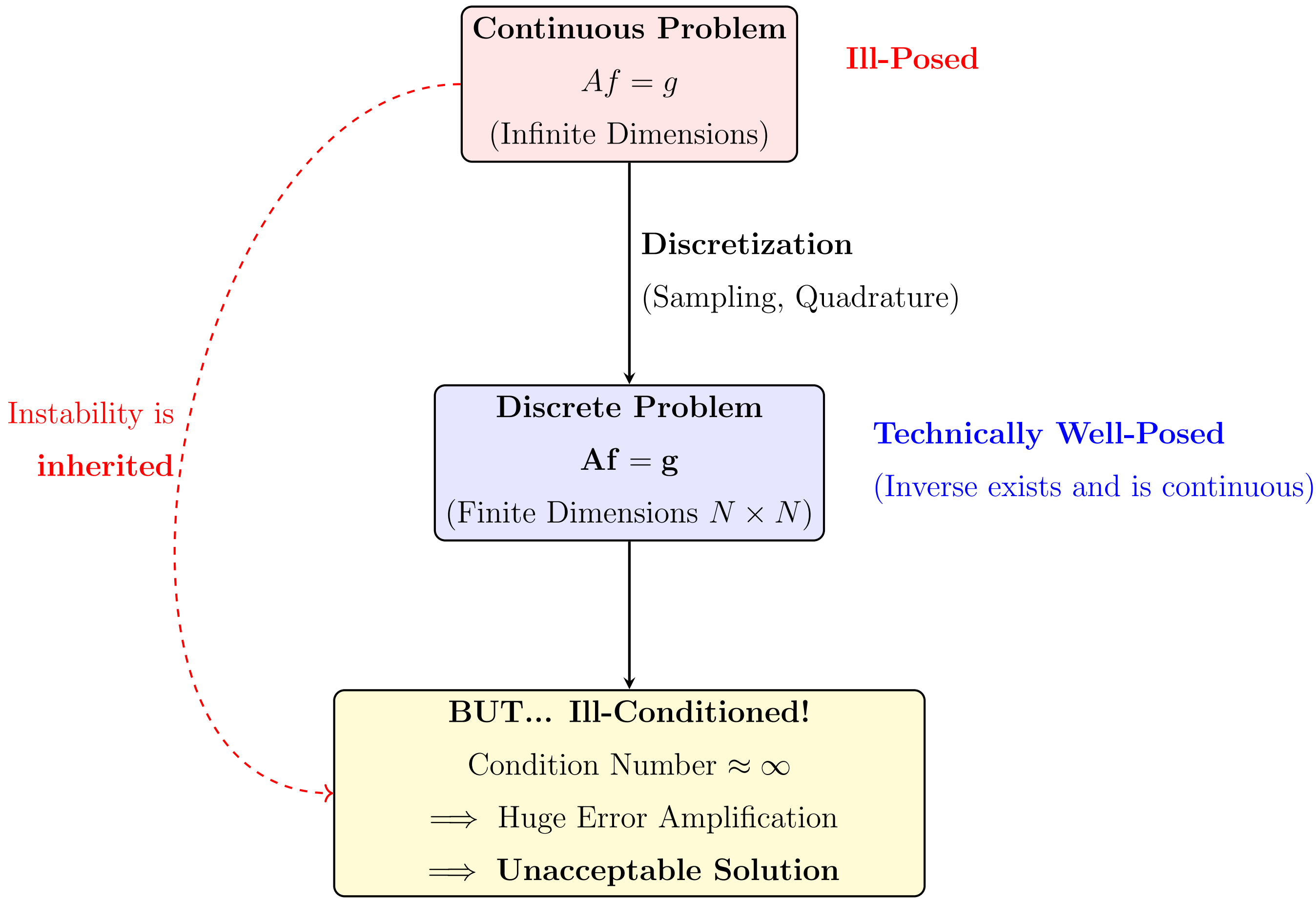

Figure 6: The transition from the continuous to the discrete setting fixes the mathematical definition of well-posedness (existence/uniqueness) but retains the fundamental physical instability in the form of a huge Condition Number $\kappa(A)$, rendering the naive solution useless in the presence of noise.

This demonstrates that we cannot simply "undo" the blur without handling the stochastic noise component. This necessitates **Regularization**.

2 Lecture 2: Quadratic Tikhonov Regularization and Filtering

Objective: Overcome the ill-posedness by introducing physical constraints. We replace the original problem with a well-posed approximation.
Notation: To simplify the calculations, in this section we incorporate the noise directly into the data and adopt the following notation:

$$g = Af, \qquad \text{and} \qquad g^\delta = g + w \tag{39}$$

Here w is the (unknown) noise vector, and the superscript δ indicates a contaminated measurement by noise.

2.1 Tikhonov regularization

Tikhonov proposed a **Regularization Method** to solve ill-posed problems: Consider families of *approximate solutions* depending on a *positive parameter* called *regularization parameter*.

2.1.1 Background: The Least Squares Problem

Before introducing regularization, recall that when the problem $Af = g$ is ill-posed and in presence of noisy data g^δ, an exact solution might not exist (the data g might imply features that cannot be generated by the physics A). In this case, the standard approach is to look for a **Least Squares Solution**, defined as any object f that minimizes *the discrepancy between the model and the data*:

$$\text{minimize } \epsilon^2(f) = ||Af - g^\delta||^2 \tag{40}$$

However, for ill-posed problems, the unconstrained least squares solution is typically unstable: it tries to fit the noise, resulting in a solution with huge, oscillating values (infinite energy).
A naïve explanation: Imagine the unconstrained Least Squares method as an **obsessive perfectionist**. Its only goal is to find an object that, when blurred, matches the noisy data *exactly*, down to the last pixel.

Here is why this approach leads to disaster in ill-posed problems:

1. **The Physics (The Dampener):** The imaging system (the blur) acts like a heavy dampener. It crushes fine details and sharp edges, turning them into smooth, weak signals. It effectively "whispers" the high-frequency information.

2. **The Data (The Noise):** Your recorded image inevitably contains noise. To the algorithm, this noise looks indistinguishable from the "fine details" that survived the blur.

3. **The Mistake (The Amplification):** The Least Squares method looks at a tiny speck of noise and infers: *"Wait, the blur is supposed to destroy fine details. If I can still see this tiny speck, the original object must have had a* ***massive*** *spike here for it to survive the blur!"*

4. **The Result (Oscillations):** To explain that tiny bit of noise, the algorithm creates a gigantic, artificial spike in the reconstruction. Since noise is everywhere, the algorithm does this all over the image. It effectively creates a violently oscillating, high-energy mess just to "explain" the random static in the data.

Conclusion: The algorithm mistakes the noise for a very faint signal and amplifies it by a factor of millions trying to restore it. The result is an image dominated by mathematical artifacts (infinite energy) rather than physical reality.

2.2 Least-Squares Solutions with Prescribed Energy

To address these issues, we adopt a constrained least-squares approach. **Idea:** Find a least squares solution with prescribed energy. Find the minimum (or the minima) of:

$$\epsilon^2(f; g^\delta) = \left\| Af - g^\delta \right\|^2 \tag{41}$$

in the set of all objects f such that:

$$E^2(f) := \|f\|^2 \leq E^2 \tag{42}$$

where E is a constant.

$\epsilon^2(f; g^\delta)$ and $E^2(f)$ are examples of **functionals**, that is a mapping from a functional space to $\mathbb{R}$:

$$E^2 : L^2(\mathbb{R}^2) \to \mathbb{R}_{\geq 0} := \{x \in \mathbb{R} : x \geq 0\}$$

$$f \to E^2(f) = \|f\|^2$$

Terminology:

- $\epsilon^2(f; g^\delta)$ is the **COST FUNCTION** (or Loss Function).
- Minimizers are called "ARGUMENTS OF THE MINIMUM", shortly "argmin".

Remark: Ideally, we defined the constraint as an inequality:

$$||f||^2 \leq E^2 \tag{43}$$

However, in the formulation of the method, we often consider the equality constraint $||f||^2 = E^2$. Why is this justified?
We must consider the position of the **unconstrained global minimizer** (the naive least-squares solution $f^\dagger$) with respect to the feasible set (the ball of radius E).

1. **Case 1 (Inactive Constraint):** If $||f^\dagger|| \leq E$, then the unconstrained solution is already feasible. The solution to the constrained problem would simply be $f^\dagger$ itself.
2. **Case 2 (Active Constraint):** If $||f^\dagger|| > E$, the unconstrained minimum lies outside the feasible set. Since the functional $\epsilon^2(f) = ||Af - g^\delta||^2$ is convex, strictly decreasing directions point towards $f^\dagger$. Therefore, the minimum over the feasible set must lie on the **boundary**.

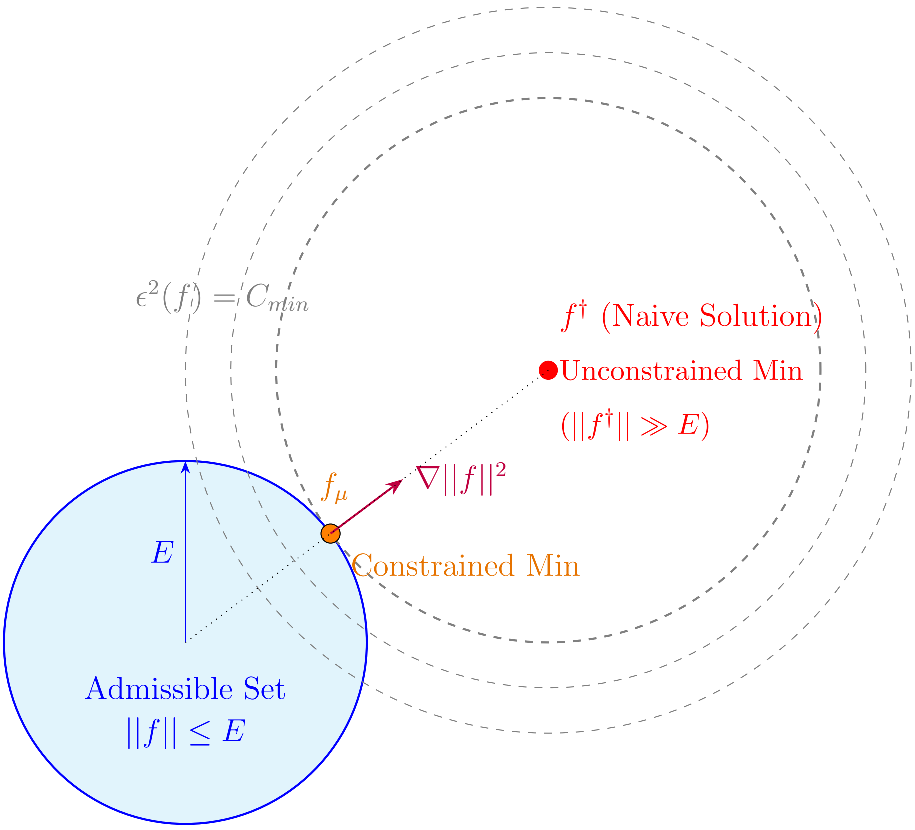

Figure 7: **Geometry of the Active Constraint.** The unconstrained minimizer $f^\dagger$ (red) lies far outside the feasible set (blue circle) because noise amplification gives it huge energy. The constrained minimizer f_μ (orange) is found on the **boundary** of the feasible set, exactly where a level curve of the cost function is tangent to the constraint circle. This justifies replacing the inequality $||f|| \leq E$ with the equality $||f|| = E$.

As shown previously, the naive solution $f^\dagger$ is dominated by noise amplification, leading to huge (or infinite) energy:

$$||f^\dagger|| \gg E \tag{44}$$

Since the unconstrained minimizer is far outside the feasible ball (which represents physical, finite energy), the constrained solution is forced to lie on the boundary (see Figure 7). Thus, the inequality constraint is essentially equivalent to the equality constraint:

$$||f||^2 = E^2 \tag{45}$$

Least-squares solutions could not satisfy this constraint due to the presence of noise in the data g^δ. In fact, as said above, the naive solution often has infinite energy due to noise amplification at high frequencies.

3 Lagrange Multipliers Method

This constrained minimization problem is equivalent to the unconstrained minimization of the following functional (using Lagrange multipliers):

$$\Phi_\mu(f) = \left\|Af - g^\delta\right\|^2 + \mu \left\|f\right\|^2, \qquad f \in L^2(\mathbb{R}^2) \tag{46}$$

where μ is the **Regularization Parameter**. This functional $\Phi_\mu(f)$ is called the **Tikhonov functional**. We look for the minimizer f_μ:

$$f_\mu \;=\; \underset{f\in L^2(\mathbb{R}^2)}{\arg\min} \left\{\Phi_\mu(f)\right\}.$$

Finding f_μ

We work in the frequency domain, always because this simplifies the convolution operator into a multiplication. Using Parseval's identity, we can rewrite the functional norms as integrals in the frequency domain:

$$\Psi_\mu(f) := (2\pi)^2\widehat{\Phi_\mu(f)} = \int_{\mathbb{R}^2} |\widehat{Af}(\omega) - \hat{g}^\delta(\omega)|^2 d\omega + \mu \int_{\mathbb{R}^2} |\hat{f}(\omega)|^2 d\omega \tag{47}$$

$$= \int_{\mathbb{R}^2} |\hat{K}(\omega)\hat{f}(\omega) - \hat{g}^\delta(\omega)|^2 d\omega + \mu \int_{\mathbb{R}^2} |\hat{f}(\omega)|^2 d\omega \tag{48}$$

Case 1: Compact Support Assumption

Let us assume, for simplicity, that the Transfer Function $\hat{K}(\omega)$ has **compact support**

$$B := \operatorname{supp}(\hat{K}) = \overline{\{\omega \in \mathbb{R}^2 : \hat{K}(\omega) \neq 0\}}.$$

Outside this region ($\omega \notin B$), we have $\hat{K}(\omega) = 0$.

We can split the integral of the functional into two parts: over the support B and over its complement B^c:

$$\Psi_\mu(f) = \int_B \left(|\hat{K}(\omega)\hat{f}(\omega) - \hat{g}^\delta(\omega)|^2 + \mu|\hat{f}(\omega)|^2\right) d\omega \tag{49}$$

$$+ \int_{B^c} \left(|0 \cdot \hat{f}(\omega) - \hat{g}^\delta(\omega)|^2 + \mu|\hat{f}(\omega)|^2\right) d\omega \tag{50}$$

Minimizing $\Psi_\mu(f)$ is equivalent to minimizing the integrand at each frequency ω pointwise (since the integral is a sum of non-negative terms).

1. For frequencies outside the support ($\omega \notin B$): The term to minimize is:

$$J(\hat{f}) = |\hat{g}^\delta(\omega)|^2 + \mu|\hat{f}(\omega)|^2$$

To minimize this with respect to $\hat{f}(\omega)$, since $\mu > 0$ and $|\hat{g}^\delta|^2$ is constant with respect to f, we must set the variable term to zero:

$$\mu|\hat{f}(\omega)|^2 = 0 \implies \hat{f}_\mu(\omega) = 0$$

This is consistent: if the system creates no signal at these frequencies, any energy in the reconstruction is pure hallucination/noise, so regularization sets it to zero.

2. For frequencies inside the support ($\omega \in B$): We minimize the integrand with respect to the complex variable $\hat{f}(\omega)$. Recall the basic identity for complex numbers: $z\,\overline{z} = |z|^2$. Let us denote the integrand by $\mathcal{L}(\hat{f})$:

$$\mathcal{L}(\hat{f}) = (\hat{K}\hat{f} - \hat{g}^\delta)\,\overline{(\hat{K}\hat{f} - \hat{g}^\delta)} + \mu\,\hat{f}\,\overline{\hat{f}}$$

$$= |\hat{K}\hat{f} - \hat{g}^\delta|^2 + \mu\,|\hat{f}|^2$$

hence,

$$\begin{aligned}
\mathcal{L}(\hat{f}) &= (\hat{K}\hat{f} - \hat{g}^\delta)\overline{(\hat{K}\hat{f} - \hat{g}^\delta)} + \mu\hat{f}\overline{\hat{f}} \\
&= (\hat{K}\hat{f} - \hat{g}^\delta)(\overline{\hat{K}\hat{f}} - \overline{\hat{g}^\delta}) + \mu\hat{f}\overline{\hat{f}} \\
&= |\hat{K}|^2|\hat{f}|^2 - \hat{K}\hat{f}\overline{\hat{g}^\delta} - \overline{\hat{K}\hat{f}}\hat{g}^\delta + |\hat{g}^\delta|^2 + \mu|\hat{f}|^2
\end{aligned}$$

Now, we collect the terms containing $\hat{f}$:

$$\mathcal{L}(\hat{f}) = (|\hat{K}|^2 + \mu)|\hat{f}|^2 - (\hat{K}\overline{\hat{g}^\delta})\hat{f} - (\overline{\hat{K}}\hat{g}^\delta)\overline{\hat{f}} + |\hat{g}^\delta|^2$$

Our goal is to rewrite this expression as a perfect square of the form $|\hat{f} - \hat{f}_\mu|^2 + \text{Residual}$ (where $\hat{f}_\mu$ is to be determined and "Residual" doesn't depend on $\hat{f}$). To do this, we add and subtract the term necessary to complete the square. The term to add and subtract is:

$$\frac{|\hat{K}\overline{\hat{g}^\delta}|^2}{|\hat{K}|^2 + \mu} = \frac{|\hat{K}|^2|\hat{g}^\delta|^2}{|\hat{K}|^2 + \mu}$$

Substituting this back, we can group the terms as follows:

$$\mathcal{L}(\hat{f}) = (|\hat{K}|^2+\mu)\left(|\hat{f}|^2 - \frac{\hat{K}\overline{\hat{g}^\delta}\hat{f}}{|\hat{K}|^2+\mu} - \frac{\overline{\hat{K}}\hat{g}^\delta\overline{\hat{f}}}{|\hat{K}|^2+\mu} + \frac{|\hat{K}|^2|\hat{g}^\delta|^2}{(|\hat{K}|^2+\mu)^2}\right) \\ + |\hat{g}^\delta|^2 - \frac{|\hat{K}|^2|\hat{g}^\delta|^2}{|\hat{K}|^2+\mu}$$

Let us focus on the cross-terms inside the parenthesis. Notice that the second term is the complex conjugate of the first:

$$\overline{\left(\frac{\hat{K}\overline{\hat{g}^\delta}\hat{f}}{|\hat{K}|^2+\mu}\right)} = \frac{\overline{\hat{K}}\hat{g}^\delta\overline{\hat{f}}}{|\hat{K}|^2+\mu}$$

Note that the term in the large parenthesis is equal to

$$\begin{aligned}|\hat{f}|^2 - \frac{\hat{K}\overline{\hat{g}^\delta}\hat{f}}{|\hat{K}|^2+\mu} - \frac{\overline{\hat{K}}\hat{g}^\delta\overline{\hat{f}}}{|\hat{K}|^2+\mu} + \frac{|\hat{K}|^2|\hat{g}^\delta|^2}{(|\hat{K}|^2+\mu)^2} &= \left(\hat{f} - \frac{\overline{\hat{K}}\hat{g}^\delta}{|\hat{K}|^2+\mu}\right)\left(\overline{\hat{f}} - \overline{\frac{\overline{\hat{K}}\hat{g}^\delta}{|\hat{K}|^2+\mu}}\right) \\ &= \left(\hat{f} - \frac{\overline{\hat{K}}\hat{g}^\delta}{|\hat{K}|^2+\mu}\right)\overline{\left(\hat{f} - \frac{\overline{\hat{K}}\hat{g}^\delta}{|\hat{K}|^2+\mu}\right)}\end{aligned}$$

Recognizing this structure allows us to write the term inside the large parenthesis exactly as a modulus squared:

$$\mathcal{L}(\hat{f}) = (|\hat{K}|^2+\mu)\left|\hat{f} - \frac{\overline{\hat{K}}\hat{g}^\delta}{|\hat{K}|^2+\mu}\right|^2 + \frac{\mu|\hat{g}^\delta|^2}{|\hat{K}|^2+\mu}$$

Conclusion: The expression is the sum of two terms:

1. A non-negative term that depends on $\hat{f}$ (the perfect square).

2. A constant term (Residual) that does not depend on $\hat{f}$.

Since $(|\hat{K}|^2+\mu) > 0$, the global minimum is achieved when the squared term is zero:

$$\hat{f} - \frac{\overline{\hat{K}}\hat{g}^\delta}{|\hat{K}|^2+\mu} = 0$$

Recalling that $\overline{\hat{K}} = \hat{K}^*$, we obtain the solution:

$$\hat{f}_\mu(\omega) = \frac{\hat{K}^*(\omega)}{|\hat{K}(\omega)|^2+\mu}\hat{g}^\delta(\omega) \tag{51}$$

Mathematical Insight: Why is $\hat{K}^* = \overline{\hat{K}}$?

To answer to the box question, we need to determine the expression of the adjoint operator A^* (of the convolution operator) in the frequency domain. Let A be the convolution operator $Af = K * f$. In the Fourier domain, this becomes a multiplication: $\widehat{Af} = \hat{K}\hat{f}$.

By definition of the adjoint operator A^* in the physical space $L^2(\mathbb{R}^2)$:

$$\langle Af, g \rangle_{L^2} = \langle f, A^* g \rangle_{L^2} \tag{52}$$

Since the Fourier Transform is an isometry (preserving inner products), we have:

$$\langle Af, g \rangle_{L^2} = \langle \widehat{Af}, \hat{g} \rangle_{L^2} = \langle \hat{K}\hat{f}, \hat{g} \rangle_{L^2} \tag{53}$$

Now we work entirely in the frequency domain. Writing out the integral definition:

$$\langle \hat{K}\hat{f}, \hat{g} \rangle = \int_{\mathbb{R}^2} (\hat{K}(\omega)\hat{f}(\omega))\overline{\hat{g}(\omega)}\, d\omega \tag{54}$$

Since scalar multiplication is commutative, we rearrange terms to isolate $\hat{f}$:

$$= \int_{\mathbb{R}^2} \hat{f}(\omega) \left(\hat{K}(\omega)\overline{\hat{g}(\omega)} \right) \, d\omega \tag{55}$$

We rewrite the term in parentheses as the complex conjugate of a product to match the inner product structure $\langle u, v \rangle = \int u\bar{v}$:

$$= \int_{\mathbb{R}^2} \hat{f}(\omega) \overline{\left(\overline{\hat{K}(\omega)}\hat{g}(\omega) \right)} \, d\omega \tag{56}$$

This is exactly the inner product in the frequency domain between $\hat{f}$ and a new function $\overline{\hat{K}}\hat{g}$:

$$= \langle \hat{f}, \overline{\hat{K}}\hat{g} \rangle_{L^2} \tag{57}$$

3. Identification of the Adjoint

By comparing the start and end of the chain (and using Plancherel again to go back to f):

$$\langle f, A^* g \rangle_{L^2} = \langle \hat{f}, \widehat{A^* g} \rangle_{L^2} \tag{58}$$

Matching the terms acting on g, we identify the adjoint operator in the frequency domain:

$$\widehat{A^* g}(\omega) = \overline{\hat{K}(\omega)}\hat{g}(\omega) \tag{59}$$

Thus, the transfer function of the adjoint operator is the complex conjugate of the original transfer function:

$$\widehat{A^*} \equiv \hat{K}^*(\omega) = \overline{\hat{K}(\omega)} \tag{60}$$

Remark: Back to Physical Space

What does A^* look like in the physical space? By applying the Inverse Fourier Transform to $\overline{\hat{K}(\omega)}$, we find the impulse response of the adjoint. Since $\overline{\hat{K}(\omega)} = \hat{K}(-\omega)$ (assuming $K(x)$ is real-valued), the inverse transform gives:

$$K^*(x) = K(-x) \tag{61}$$

Therefore, while A is a convolution with $K(x)$, the adjoint A^* is a convolution with the time-reversed kernel $K(-x)$ (which corresponds to the cross-correlation operation).

Generalization

We can now merge the two cases. Since for $\omega \notin B$ we have $\hat{K}(\omega) = 0$, the formula derived for the support:

$$\frac{0}{0+\mu}\hat{g}^\delta(\omega) = 0$$

correctly yields the zero solution found in step 1.

Final step

Therefore, the **Tikhonov Filter** formula is valid for all $\omega \in \mathbb{R}^2$:

$$\hat{f}_\mu(\omega) = \frac{\hat{K}^*(\omega)}{|\hat{K}(\omega)|^2 + \mu}\hat{g}^\delta(\omega) \tag{62}$$

Therefore, by applying the Inverse Fourier Transform, we derive the explicit expression of the solution in the physical space:

$$f_\mu(x) = (R_\mu * g^\delta)(x) = \int_{\mathbb{R}^2} R_\mu(x-y)g^\delta(y)\,dy \tag{63}$$

where the function $R_\mu(x)$ is the **Regularized Reconstruction Kernel**. It is defined as the Inverse Fourier Transform of the spectral filter derived above:

$$R_\mu(x) = \mathcal{F}^{-1}\left[\frac{\hat{K}^*(\omega)}{|\hat{K}(\omega)|^2 + \mu}\right](x) \tag{64}$$

Physical Interpretation: The regularized solution f_μ is obtained by convolving the noisy data g^δ with a specific kernel R_μ. This kernel acts simultaneously as:

- An **Inverse Filter** (restoring the blur) where the signal is strong.
- A **Smoothing Filter** (suppressing noise) where the signal is weak.

Properties of the Tikhonov Solution f_μ.

We analyze the properties of the solution f_μ derived above:

$$f_\mu(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^2} \frac{\hat{K}^*(\omega)}{|\hat{K}(\omega)|^2 + \mu}\hat{g}^\delta(\omega)e^{ix\cdot\omega}d\omega \tag{65}$$

The solution satisfies two fundamental properties:

1. $f_\mu \in L^2(\mathbb{R}^2)$ for any $g \in L^2$ and $\mu > 0$.
2. f_μ depends continuously on the data g (Stability).

Proofs: Conditions (1) and (2) can be proved using the following elementary inequality. Here, with g we denote a generic function in L^2. Consider the function $h(\eta) = \frac{\eta^2}{\eta^2+\mu}$, we have

$$\frac{\eta^2}{(\eta^2+\mu)^2} \leq \frac{1}{4\mu} \tag{66}$$

Now, let us compute the energy (squared L^2 norm) of the solution f_μ. Using Parseval's equality:

$$E^2(f_\mu) = ||f_\mu||^2 = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^2} |\hat{f}_\mu(\omega)|^2 d\omega \tag{67}$$

Substituting the expression of $\hat{f}_\mu(\omega)$:

$$E^2(f_\mu) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^2} \left|\frac{\hat{K}^*(\omega)\hat{g}(\omega)}{|\hat{K}(\omega)|^2 + \mu}\right|^2 d\omega \tag{68}$$

Since $|\hat{K}^*| = |\hat{K}|$, the integrand becomes:

$$\frac{|\hat{K}(\omega)|^2}{(|\hat{K}(\omega)|^2 + \mu)^2}|\hat{g}(\omega)|^2 \tag{69}$$

We can apply the elementary inequality (66) by setting $\eta = |\hat{K}(\omega)|$:

$$\frac{|\hat{K}(\omega)|^2}{(|\hat{K}(\omega)|^2 + \mu)^2} \leq \frac{1}{4\mu} \tag{70}$$

Therefore, the integral is bounded by:

$$E^2(f_\mu) \leq \frac{1}{(2\pi)^d}\int_{\mathbb{R}^2} \frac{1}{4\mu}|\hat{g}(\omega)|^2 d\omega = \frac{1}{4\mu}\left(\frac{1}{(2\pi)^d}\int_{\mathbb{R}^2}|\hat{g}(\omega)|^2 d\omega\right) \tag{71}$$

Using Parseval again on g:

$$||f_\mu||^2 \leq \frac{1}{4\mu}||g||^2 \tag{72}$$

Taking the square root:

$$||f_\mu|| \leq \frac{1}{2\sqrt{\mu}}||g|| \tag{73}$$

Conclusion: Since $\mu > 0$ is a fixed constant, the operator mapping $g \to f_\mu$ is bounded with norm $\leq \frac{1}{2\sqrt{\mu}}$. This proves that the solution has finite energy (Property 1) and depends continuously on the data (Property 2). A small perturbation in g results in a bounded perturbation in f_μ.

Existence and Uniqueness of μ

For the Lagrange multipliers method to be valid, once we have shown that the minimum f_μ exists for all $\mu > 0$, we need to prove that there exists **only one** μ such that:

$$E(f_\mu) = ||f_\mu|| = E \tag{74}$$

This is simple to show by analyzing the behavior of the energy function $E(f_\mu)$ with respect to μ. From the explicit formula of the solution, we can observe the following properties:

- **Monotonicity:** The energy $E(f_\mu) = \|f_\mu\|$ is a **non-increasing** function of μ (and it is **strictly decreasing** under mild non-degeneracy assumptions).
  - As μ increases (stronger regularization), high-frequency components are damped

more, hence the energy cannot increase.

Proof: Why does the Energy Decrease?

Using the Fourier-domain formula

$$\hat{f}_\mu(\omega) = \frac{\overline{\hat{K}(\omega)}\,\hat{g}(\omega)}{|\hat{K}(\omega)|^2 + \mu},$$

Parseval's identity gives

$$E^2(\mu) = \|f_\mu\|^2 = \int_{\mathbb{R}^2} |\hat{f}_\mu(\omega)|^2\,d\omega = \int_{\mathbb{R}^2} \frac{|\hat{K}(\omega)|^2\,|\hat{g}(\omega)|^2}{\left(|\hat{K}(\omega)|^2 + \mu\right)^2}\,d\omega.$$

If $0 < \mu_1 < \mu_2$, then for every ω

$$\frac{1}{\left(|\hat{K}(\omega)|^2 + \mu_2\right)^2} \le \frac{1}{\left(|\hat{K}(\omega)|^2 + \mu_1\right)^2},$$

hence $|\hat{f}_{\mu_2}(\omega)|^2 \le |\hat{f}_{\mu_1}(\omega)|^2$ pointwise. Integrating yields

$$E^2(\mu_2) \le E^2(\mu_1), \qquad \text{and therefore} \qquad E(\mu_2) \le E(\mu_1).$$

Moreover, the inequality is *strict* (so $E(\mu)$ strictly decreases) whenever the set $\{\omega : \ |\hat{K}(\omega)|\,|\hat{g}(\omega)| > 0\}$ has positive measure.

- **Limit $\mu \to 0$:**
$$\lim_{\mu\to 0} E(f_\mu) = E(f^\dagger) \tag{75}$$
where $f^\dagger$ is the naive (unconstrained) solution. As discussed, for ill-posed problems with noise, this energy is huge (virtually infinite):
$$E(f^\dagger) > E \tag{76}$$

- **Limit $\mu \to \infty$:**
$$\lim_{\mu\to\infty} E(f_\mu) = 0 \tag{77}$$
(Since the penalty term $\mu||f||^2$ dominates, forcing the solution to zero).

Conclusion: Since the function $E(f_\mu)$ is continuous and decreases monotonically from a value greater than E (at $\mu = 0$) to 0 (at $\mu \to \infty$), by the Intermediate Value

Theorem, there must exist a **unique** value μ^* such that:

$$E(f_{\mu^*}) = E \tag{78}$$

This guarantees that the Lagrange Multipliers method provides a unique regularized solution satisfying the prescribed energy constraint.

Case 2: Generalize to an arbitrary kernel K

The calculations in the previous section relied on the simplifying assumption that K has compact support. We now show that the final result remains valid in full generality, for an arbitrary kernel K.

Let us manipulate the filter equation (62) by multiplying both sides for the denominator:

$$(|\hat{K}(\omega)|^2 + \mu)\hat{f}_\mu(\omega) = \hat{K}^*(\omega)\hat{g}^\delta(\omega) \tag{79}$$

We can split the term on the left-hand side and identify the operators acting on the signal:

$$\underbrace{\hat{K}^*(\omega)\hat{K}(\omega)\hat{f}_\mu(\omega)}_{\text{Term 1}} + \underbrace{\mu\hat{f}_\mu(\omega)}_{\text{Term 2}} = \underbrace{\hat{K}^*(\omega)\hat{g}^\delta(\omega)}_{\text{Term 3}} \tag{80}$$

Now, we map these spectral terms back to the physical space using the properties of the Fourier Transform and the Adjoint Operator:

- **Term 3:** $\hat{K}^*(\omega)\hat{g}^\delta(\omega)$ is the Fourier transform of applying the adjoint to the data:

$$\mathcal{F}^{-1}[\hat{K}^*\hat{g}^\delta] = A^*g^\delta$$

- **Term 1:** $\hat{K}^*(\omega)\hat{K}(\omega)\hat{f}_\mu(\omega)$ corresponds to applying the operator A, followed by its adjoint A^*:

$$\mathcal{F}^{-1}[\hat{K}^*(\hat{K}\hat{f}_\mu)] = A^*(Af_\mu) = (A^*A)f_\mu$$

- **Term 2:** Multiplication by a constant μ corresponds to the identity operator scaled by μ:

$$\mathcal{F}^{-1}[\mu\hat{f}_\mu] = \mu I f_\mu$$

Combining terms, the frequency equation is equivalent to the following operator equation in L^2:

$$(A^*A + \mu I)f_\mu = A^*g^\delta \tag{81}$$

Which yields the formal solution:

$$f_\mu = (A^*A + \mu I)^{-1} A^* g^\delta \tag{82}$$

Generalization: The Normal Equations. The equation derived above is not specific to convolution operators or band-limited kernels. It is a general result known as the **Regularized Normal Equation**.

It arises naturally when minimizing the Tikhonov functional for *any* linear operator A (even those that are not diagonalized by the Fourier Transform, such as space-variant blur).

General Variational Derivation: Let us minimize the general Tikhonov functional:

$$\Phi_\mu(f) = ||Af - g^\delta||^2 + \mu||f||^2$$

To find the minimum, we compute the gradient $\nabla\Phi_\mu(f)$ and set it to zero. Recall the gradients of the quadratic terms:

$$\nabla\left(\|Af - g^\delta\|^2\right) = 2A^*(Af - g^\delta), \tag{83}$$

$$\nabla\left(\|f\|^2\right) = 2f. \tag{84}$$

Therefore:

$$\nabla\Phi_\mu(f) = 2A^*(Af - g^\delta) + 2\mu f = 0$$
$$A^*Af - A^*g^\delta + \mu f = 0$$
$$(A^*A + \mu I)f = A^*g^\delta$$

Since $(A^*A + \mu I)$ is positive definite for any $\mu > 0$, the inverse exists and the solution is unique. This confirms that the formula (82) we "discovered" from the filter is the universal solution to the quadratic regularization problem.

Mathematical Insight: Deriving the Gradients above

To get formulas (83) and (84), we need the definition of the Fréchet or Gâteaux derivative in Hilbert spaces. Let us denote by F a generic functional.

The gradient $\nabla F(f)$ is defined as the unique vector satisfying the linear approximation:

$$F(f + \epsilon h) \approx F(f) + \epsilon \langle \nabla F(f), h \rangle \tag{85}$$

for a small perturbation ϵ in any direction h. Let us apply this to our two terms.

1. Gradient of the Norm Squared $P(f) = ||f||^2$

We expand the squared norm using the inner product:

$$\begin{aligned} P(f + \epsilon h) &= \langle f + \epsilon h, f + \epsilon h \rangle & (86) \\ &= \langle f, f \rangle + \epsilon \langle f, h \rangle + \epsilon \langle h, f \rangle + \epsilon^2 \langle h, h \rangle & (87) \end{aligned}$$

Assuming a real Hilbert space (where $\langle f, h \rangle = \langle h, f \rangle$):

$$P(f + \epsilon h) = ||f||^2 + 2\epsilon \langle f, h \rangle + O(\epsilon^2) \tag{88}$$

Comparing this with the definition, the term multiplying the direction h in the inner product is $2f$. Thus:

$$\nabla(||f||^2) = 2f \tag{89}$$

2. Gradient of the Data Fidelity $J(f) = ||Af - g||^2$

We apply the same perturbation:

$$\begin{aligned} J(f + \epsilon h) &= ||A(f + \epsilon h) - g||^2 & (90) \\ &= ||(Af - g) + \epsilon A h||^2 & (91) \end{aligned}$$

Expanding the norm:

$$\begin{aligned} &= \langle (Af - g) + \epsilon A h, (Af - g) + \epsilon A h \rangle & (92) \\ &= ||Af - g||^2 + 2\epsilon \langle Af - g, Ah \rangle + \epsilon^2 ||Ah||^2 & (93) \end{aligned}$$

The linear part (the derivative) is $2\langle Af - g, Ah \rangle$. However, to identify the gradient, the scalar product must be in the form $\langle \textbf{Gradient}, h \rangle$. We must move the operator A from the right side to the left side using the definition of the adjoint operator (A^*):

$$\langle Af - g, Ah \rangle = \langle A^*(Af - g), h \rangle \tag{94}$$

Now we can identify the gradient as the vector multiplying h:

$$\nabla(||Af - g||^2) = 2A^*(Af - g) \tag{95}$$

4 Error Analysis and the Bias-Variance Trade-off

To understand the role of the regularization parameter μ, we use the **Regularized Reconstruction Operator** R_μ defined in (64). This is the linear operator derived in equation (82) that maps the data g to the solution f_μ:

$$f_\mu = R_\mu g^\delta \quad \text{with} \quad R_\mu = (A^*A + \mu I)^{-1}A^* \tag{96}$$

Let us analyze the error between the reconstructed solution f_μ and the true object $f^{(0)}$. Assume the data is generated by the model with additive noise:

$$g = Af^{(0)} + w$$

Substituting this into the reconstruction formula:

$$\begin{aligned} f_\mu &= R_\mu(Af^{(0)} + w) && (97) \\ &= R_\mu Af^{(0)} + R_\mu w && (98) \end{aligned}$$

Now we subtract the true object $f^{(0)}$ from both sides to find the reconstruction error:

$$\begin{aligned} f_\mu - f^{(0)} &= (R_\mu Af^{(0)} - f^{(0)}) + R_\mu w && (99) \\ &= \underbrace{(R_\mu A - I)f^{(0)}}_{\text{Bias (Approximation)}} + \underbrace{R_\mu w}_{\text{Noise}} && (100) \end{aligned}$$

Note that if we used the exact inverse A^{-1} (assuming it exists), then $R_\mu A = A^{-1}A = I$, and the first term would vanish. However, for ill-posed problems, we cannot use the exact inverse.

The total error is the sum of the squared norms of these two competing terms:

$$\text{Total Error}^2(\mu) = \underbrace{\left\|(R_\mu A - I)f^{(0)}\right\|^2}_{\text{Approximation Error}} + \underbrace{\|R_\mu w\|^2}_{\text{Noise Propagation Error}} \tag{101}$$

Interpretation of the Trade-off

- **Approximation Error (Bias):** Represents the loss of resolution due to regularization.
  - As $\mu \to \infty$, $R_\mu \to 0$, so the error tends to $||f^{(0)}||^2$ (maximum error, complete information loss).
  - As $\mu \to 0$, $R_\mu A \to I$, so the approximation error tends to 0.
- **Noise Error (Variance):** Represents the noise amplified by the inversion.
  - As $\mu \to \infty$, the filter suppresses everything, so noise error $\to 0$.
  - As $\mu \to 0$, the operator approaches the unstable inverse (the naive least squares solution), so noise error $\to \infty$.

The optimal μ is found at the intersection of these two trends, as shown in the plot below.

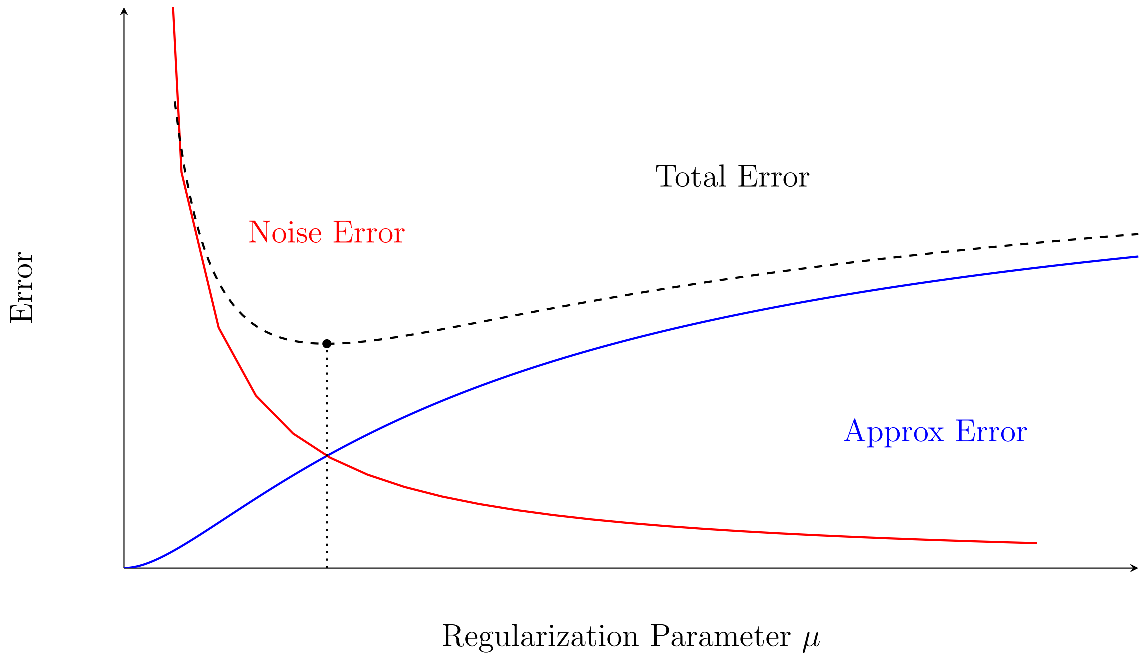

Figure 8: The Bias-Variance Trade-off. The total error (dashed) is minimized at an intermediate value μ_{opt}.

4.1 Alternative Strategy: Approximate Solutions with Minimal Energy

We can approach the regularization problem from a **dual perspective**. Sometimes we have information regarding the noise level ϵ, i.e., $||w|| \approx \epsilon$. In this case, instead of minimizing the error with a constraint on energy, we can look for the solution with **minimal energy** that is compatible with the data.

Problem Statement: Find the approximate solution f that minimizes the energy:

$$E^2(f) = ||f||^2 \tag{102}$$

subject to the discrepancy constraint:

$$\epsilon^2(f; g^\delta) = ||Af - g^\delta||^2 \leq \epsilon^2 \tag{103}$$

Equivalence to Tikhonov: This problem is, in some sense, the **dual** of the previous one, because the roles of the energy $E^2(f)$ and the misfit functional $||Af - g^\delta||^2$ are exchanged.

Using Lagrange multipliers, this is equivalent to minimizing:

$$\mathcal{L}(f) = ||f||^2 + \lambda(||Af - g^\delta||^2 - \epsilon^2) \tag{104}$$

Dividing by λ (and setting $\mu = 1/\lambda$), we obtain the exact same Tikhonov functional $\Phi_\mu(f)$ analyzed before. Therefore, the minimizer is the **same function** $f_\mu(x)$ derived in the previous sections.

5 Spectral Analysis (Filtering)

Regularization using Differential Operators (1D Examples)

So far, we have considered the constraint on the total energy of the object ($||f||^2 \leq E^2$), which corresponds to the penalty term $\mu||f||^2$ (Tikhonov regularization of order 0). However, often we want to force the solution to be **smooth** rather than just small. To do this, we can penalize the derivatives of the function.

Let us consider the 1D case ($x \in \mathbb{R}$) and generalize the functional as:

$$\Phi_\mu(f) = ||Af - g^\delta||^2 + \mu||Lf||^2 \tag{105}$$

where L is a differential operator.

In the Fourier domain, differentiation corresponds to multiplication by $i\omega$. Therefore, by Parseval's relation:

$$(2\pi)^2||Lf||^2 = \int_{-\infty}^{+\infty} |\widehat{Lf}(\omega)|^2 d\omega = \int_{-\infty}^{+\infty} |P(\omega)|^2|\hat{f}(\omega)|^2 d\omega \tag{106}$$

where $P(\omega)$ is the polynomial associated with the operator L.

Hence, the general form of the Tikhonov filter becomes:

$$\hat{f}_\mu(\omega) = \frac{\hat{K}^*(\omega)}{|\hat{K}(\omega)|^2 + \mu|P(\omega)|^2}\hat{g}^\delta(\omega) \tag{107}$$

Derivation of the General Tikhonov Filter

We perform the minimization of the functional with the differential operator L explicitly. In the frequency domain, the term to minimize at each frequency ω is:

$$\mathcal{L}(\hat{f}) = |\hat{K}(\omega)\hat{f}(\omega) - \hat{g}^\delta(\omega)|^2 + \mu|P(\omega)\hat{f}(\omega)|^2$$

We expand the squares using the identity $|z|^2 = z\bar{z}$.

$$\mathcal{L}(\hat{f}) = (\hat{K}\hat{f} - \hat{g}^\delta)\,\overline{(\hat{K}\hat{f} - \hat{g}^\delta)} + \mu\,(P\hat{f})\,\overline{(P\hat{f})}$$

Expanding terms (omitting ω for brevity):

$$\begin{aligned}\mathcal{L}(\hat{f}) &= (\hat{K}\hat{f} - \hat{g}^\delta)(\overline{\hat{K}\hat{f}} - \overline{\hat{g}^\delta}) + \mu P\hat{f}\overline{P}\overline{\hat{f}} \\ &= |\hat{K}|^2|\hat{f}|^2 - \hat{K}\hat{f}\overline{\hat{g}^\delta} - \overline{\hat{K}\hat{f}}\hat{g}^\delta + |\hat{g}^\delta|^2 + \mu|P|^2|\hat{f}|^2\end{aligned}$$

Collecting terms containing $|\hat{f}|^2$:

$$\mathcal{L}(\hat{f}) = (|\hat{K}|^2 + \mu|P|^2)|\hat{f}|^2 - (\hat{K}\overline{\hat{g}^\delta})\hat{f} - (\overline{\hat{K}}\hat{g}^\delta)\overline{\hat{f}} + |\hat{g}^\delta|^2$$

We complete the square by adding and subtracting:

$$\frac{|\hat{K}\overline{\hat{g}^\delta}|^2}{|\hat{K}|^2+\mu|P|^2} = \frac{|\hat{K}|^2|\hat{g}^\delta|^2}{|\hat{K}|^2+\mu|P|^2}$$

Substituting and grouping:

$$\mathcal{L}(\hat{f}) = (|\hat{K}|^2+\mu|P|^2)\left(|\hat{f}|^2 - \frac{\hat{K}\overline{\hat{g}^\delta}\hat{f}}{|\hat{K}|^2+\mu|P|^2} - \frac{\overline{\hat{K}}\hat{g}^\delta\overline{\hat{f}}}{|\hat{K}|^2+\mu|P|^2} + \frac{|\hat{K}|^2|\hat{g}^\delta|^2}{(|\hat{K}|^2+\mu|P|^2)^2}\right) + \left(|\hat{g}^\delta|^2 - \frac{|\hat{K}|^2|\hat{g}^\delta|^2}{|\hat{K}|^2+\mu|P|^2}\right)$$

Recognizing the perfect square:

$$\mathcal{L}(\hat{f}) = (|\hat{K}|^2+\mu|P|^2)\left|\hat{f} - \frac{\overline{\hat{K}}\hat{g}^\delta}{|\hat{K}|^2+\mu|P|^2}\right|^2 + \frac{\mu|P|^2}{|\hat{K}|^2+\mu|P|^2}$$

Since $(|\hat{K}|^2+\mu|P|^2) > 0$ and $\frac{\mu|P|^2}{|\hat{K}|^2+\mu|P|^2} > 0$, the minimum is at:

$$\hat{f} - \frac{\overline{\hat{K}}\hat{g}^\delta}{|\hat{K}|^2+\mu|P|^2} = 0 \implies \hat{f}_\mu(\omega) = \frac{\hat{K}^*(\omega)}{|\hat{K}(\omega)|^2+\mu|P(\omega)|^2}\hat{g}^\delta(\omega)$$

Let us analyze two specific examples.

Example 1: Derivative of first order

We want to *penalize rapid variations* in the solution. We choose the penalty term based on the first derivative:

$$||Lf||^2 = ||f'||^2 = \int |f'(x)|^2 dx \tag{108}$$

In the Fourier domain:

- $f'(x) \xrightarrow{\mathcal{F}} i\omega\hat{f}(\omega)$
- $||f'||^2 = \int |i\omega\hat{f}(\omega)|^2 d\omega = \int \omega^2|\hat{f}(\omega)|^2 d\omega$

Substituting this into the minimization problem, the regularization term in frequency is $\mu\omega^2$. The resulting filter is:

$$\hat{f}_\mu(\omega) = \frac{\hat{K}^*(\omega)}{|\hat{K}(\omega)|^2+\mu\omega^2}\hat{g}^\delta(\omega) \tag{109}$$

Interpretation: The term $\mu\omega^2$ acts as a frequency-dependent regularization parameter.

- At **low frequencies** ($\omega \approx 0$), the penalty $\mu\omega^2$ is negligible. The filter preserves the signal.
- At **high frequencies** ($\omega \to \infty$), the term $\mu\omega^2$ becomes very large, suppressing the solution much more aggressively than standard Tikhonov (where μ is constant).

This enforces **smoothness** (Sobolev regularization H^1). So, in the images, edges and sharp interfaces appear smoothed out.

5.1 Example 2: Regularization with Second Derivative

We want to penalize the "roughness" or curvature of the solution. We use the second derivative:

$$||Lf||^2 = ||f''||^2 = \int |f''(x)|^2 dx \tag{110}$$

In the Fourier domain:

- $f''(x) \xrightarrow{\mathcal{F}} (i\omega)^2 \hat{f}(\omega) = -\omega^2 \hat{f}(\omega)$
- $||f''||^2 = \int |-\omega^2 \hat{f}(\omega)|^2 d\omega = \int \omega^4 |\hat{f}(\omega)|^2 d\omega$

The resulting filter is:

$$\hat{f}_\mu(\omega) = \frac{\hat{K}^*(\omega)}{|\hat{K}(\omega)|^2 + \mu\omega^4} \hat{g}^\delta(\omega) \tag{111}$$

Interpretation: The penalty term scales with ω^4. This provides an extremely strong suppression of high-frequency noise, forcing the reconstructed object to be very smooth. This corresponds to regularization in the Sobolev space H^2.

Conclusion: As the order of the derivative increases, the window function decays faster to zero as $|\omega| \to \infty$, acting as a stronger Low-Pass filter.

5.1.1 Physical Interpretation in Image Deblurring

To understand the filter behavior, we rewrite the Tikhonov solution as a modification of the Naive Inverse Filter:

$$\hat{f}_\mu(\omega) = \underbrace{\left(\frac{|\hat{K}(\omega)|^2}{|\hat{K}(\omega)|^2 + \mu} \right)}_{\text{Window Function } \hat{W}_\mu(\omega)} \cdot \underbrace{\left(\frac{\hat{g}^\delta(\omega)}{\hat{K}(\omega)} \right)}_{\text{Naive Inverse}} \tag{112}$$

In image deblurring, the blur kernel $\hat{K}(\omega)$ is typically a **Low-Pass Filter** (e.g., Gaussian blur or Motion blur). This means it preserves low frequencies but attenuates high frequencies. The noise, however, is distributed across the entire spectrum.

The term $\hat{W}_\mu(\omega)$ acts as a "soft switch" that depends on the Signal-to-Noise Ratio (SNR) at each frequency:

- **Low Frequencies (Coarse Structures):**
  These frequencies correspond to the general shapes and homogeneous regions of the image. Here, the blur kernel transmits most of the energy ($|\hat{K}(\omega)|^2 \gg \mu$).

  $$\hat{W}_\mu(\omega) \approx 1$$

  Result: The filter acts as an **Inverse Filter**. It successfully deconvolves the image, correcting the blur on large structures.

- **High Frequencies (Fine Details vs. Noise):**
  These frequencies correspond to sharp edges, fine textures, and noise. Here, the blur kernel has destroyed the signal ($|\hat{K}(\omega)|^2 \ll \mu$), so the measured data is dominated by noise.

  $$\hat{W}_\mu(\omega) \approx 0$$

  Result: The filter acts as a **Regularizer**. It suppresses these frequencies.

  - *Consequence:* We avoid amplifying the noise (which would look like "salt and pepper" grain), but we inevitably lose the finest details of the image (the "smoothness" constraint).

Visual Trade-off: The parameter μ determines the cut-off frequency.

- If μ is too small: We recover high frequencies, but we amplify noise artifacts (image looks grainy).

- If μ is too large: We suppress noise, but we cut off too many high frequencies (image looks overly smooth/blurred).

6 Window Functions (generalization)

Every function $\hat{W}_\mu$ satisfying specific conditions defines a regularization procedure. In particular, from (112), to have a regularization we need the following

Properties of Window Functions:

1. $|\hat{W}_\mu(\omega)| \leq 1, \quad \forall \mu > 0$
2. $\lim_{\mu \to 0} \hat{W}_\mu(\omega) = 1, \quad \forall \omega$ s.t. $\hat{K}(\omega) \neq 0$
3. $\hat{W}_\mu(\omega)/\hat{K}(\omega)$ is bounded $\forall \mu > 0$ (Stability).

6.1 Examples of Window Functions

In the 1D case, let us assume we have a cutoff frequency B.

- **Rectangular Window (Truncated SVD/Spectral Cutoff):**

$$\hat{W}_\Omega(\omega) = \begin{cases} 1 & \text{if } |\omega| < \Omega \\ 0 & \text{if } |\omega| \geq \Omega \end{cases} \tag{113}$$

- **Triangular Window:**

$$\hat{W}_\Omega(\omega) = \begin{cases} 1 - \frac{|\omega|}{\Omega} & \text{if } |\omega| < \Omega \\ 0 & \text{if } |\omega| \geq \Omega \end{cases} \tag{114}$$

- **Generalized Hamming Window:**

$$\hat{W}_\Omega(\omega) = \begin{cases} \alpha + (1-\alpha)\cos\left(\frac{\pi\omega}{\Omega}\right) & \text{if } |\omega| < \Omega \\ 0 & \text{if } |\omega| \geq \Omega \end{cases} \tag{115}$$

  (Common values: $\alpha = 0.5$ or $\alpha = 0.54$).

7 Beyond Tikhonov: Edge-Preserving Regularization

While Tikhonov regularization is robust and computationally efficient (thanks to the FFT), it has a major drawback in image processing: it tends to **over-smooth** the solution.

7.1 The Smoothing Effect of the L^2 Norm

Recall the derivative-based Tikhonov penalty:

$$\Phi_{Tik}(f) = ||Af - g^\delta||^2 + \mu \int |\nabla f(x)|^2 dx \tag{116}$$

The term $\int |\nabla f|^2$ penalizes large gradients quadratically. This forces the derivative to be small everywhere, effectively preventing the reconstruction of sharp edges (discontinuities). The result is often a blurry image, even if the noise is removed.

7.2 Total Variation (TV) Regularization

To preserve edges, Rudin, Osher, and Fatemi (ROF, 1992) proposed replacing the squared L^2 norm of the gradient with the L^1 norm. This is known as **Total Variation (TV)** regularization:

$$\Phi_{TV}(f) = ||Af - g^\delta||^2 + \mu \underbrace{\int_\Omega |\nabla f(x)|\, dx}_{TV(f)} \tag{117}$$

Key Differences:

- **Edge Preservation:** The L^1 norm is "lenient" on large values. It allows the gradient ∇f to be very large on a set of measure zero (i.e., at the edges), provided the image is flat elsewhere. This favors *piecewise constant* solutions.
- **Non-Linearity:** unlike the squared norm, the L^1 norm is **not differentiable** at 0 and leads to a **non-linear** Euler-Lagrange equation.

Computational Warning

Unlike Tikhonov regularization, the Total Variation problem **cannot be solved in closed form** using the Fourier Transform. Minimizing the TV functional requires iterative numerical algorithms (such as Primal-Dual methods or Split-Bregman), which are computationally more expensive than the simple Tikhonov filter. Something, anyway, that we do not address in these notes.