Lecture 12: MAP Image Restoration; Computing the MAP Estimate

Chapter 5.1–5.2

Charles A. Bouman

Purdue University

Monday, September 21, 2026

Outline

1. **The image restoration problem**
   A blurred and noisy measurement of an unknown image.
2. **MAP estimation for inverse problems**
   The forward model and the prior model.
3. **Why the ML estimate fails**
   Noise amplification, the condition number, and the case $A = I$.
4. **The GMRF prior and the Gaussian posterior**
   The posterior mean, the MAP estimate, and the MMSE estimate.
5. **Computing the MAP estimate**
   The exact solution is too large, so we minimize the MAP cost.

Reading: Chapter 5.1–5.2, *Foundations of Computational Imaging*.

Today's Big Ideas

- MAP combines a forward model (data) with a prior model (image).
- ML has no prior, so it amplifies noise when it undoes blur.
- A GMRF prior adds a quadratic penalty, which keeps it stable.
- Gaussian posterior: MAP $=$ MMSE, from one linear system.
- The exact inverse is too large, so we minimize $f(x)$ numerically.

The Gaussian prior smooths across the real edges in an image. Chapter 6 replaces it with a prior that keeps them.

The image restoration problem

Let $X \in \mathbb{R}^N$ be an N-pixel image that we would like to measure, and let $Y \in \mathbb{R}^N$ be the noisy measurements given by

$$Y = AX + W \ ,$$

where A is a nonsingular $N \times N$ matrix and W is a vector of i.i.d. Gaussian noise with distribution $N(0, \sigma^2 I)$.



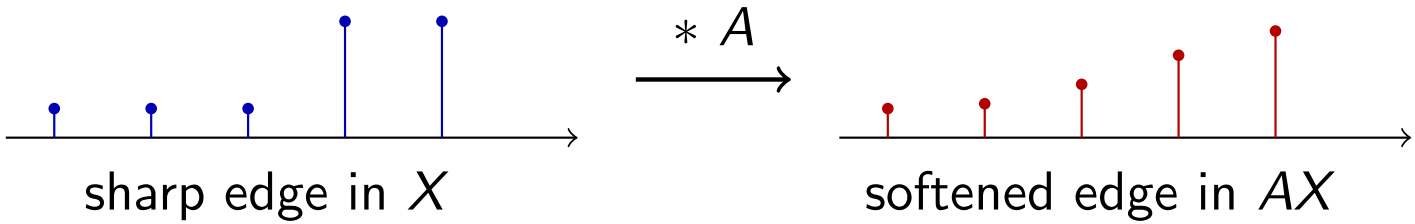
sharp image X | blurred AX | measured Y

This type of problem occurs in imaging applications where Y consists of noisy and blurred scanner measurements.

Blur as a linear space-invariant filter

The matrix A may represent the blurring operation of a linear space-invariant filter. In this case, the elements of Y and X are related by

$$Y_s = \sum_{r \in S} A_{s-r} X_r + W_s \ .$$



- In raster order, A is Toeplitz block Toeplitz.
- With generalization, $Y = AX + W$ covers tomography and astronomy.

The MAP estimate

Using Bayes' rule, the MAP estimate becomes

$$\begin{aligned}\hat{x}_{MAP} &= \arg\max_{x\in\Omega} \left\{\log p_{x|y}(x|y)\right\} \\ &= \arg\max_{x\in\Omega} \left\{\log p_{y|x}(y|x) + \log p_x(x) - \log p_y(y)\right\} \\ &= \arg\min_{x\in\Omega} \left\{-\log p_{y|x}(y|x) - \log p_x(x)\right\} \ .\end{aligned}$$

$$\underbrace{-\log p_{y|x}(y|x)}_{\text{forward model}} \quad + \quad \underbrace{-\log p_x(x)}_{\text{prior model}}$$

- The **forward model** $p_{y|x}(y|x)$: how data Y depends on X.
- The **prior model** $p_x(x)$: what we know before measuring.

The maximum likelihood estimate inverts A

Using the model $Y = AX + W$, the conditional distribution of Y given X is

$$p_{y|x}(y|x) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}||y - Ax||^2\right\} ,$$

so the ML estimate of Chapter 2 is

$$\hat{x}_{ML} = \arg\max_{x\in\mathbb{R}^N} \log p_{y|x}(y|x) = \arg\min_{x\in\mathbb{R}^N} ||y - Ax||^2 = A^{-1}y .$$

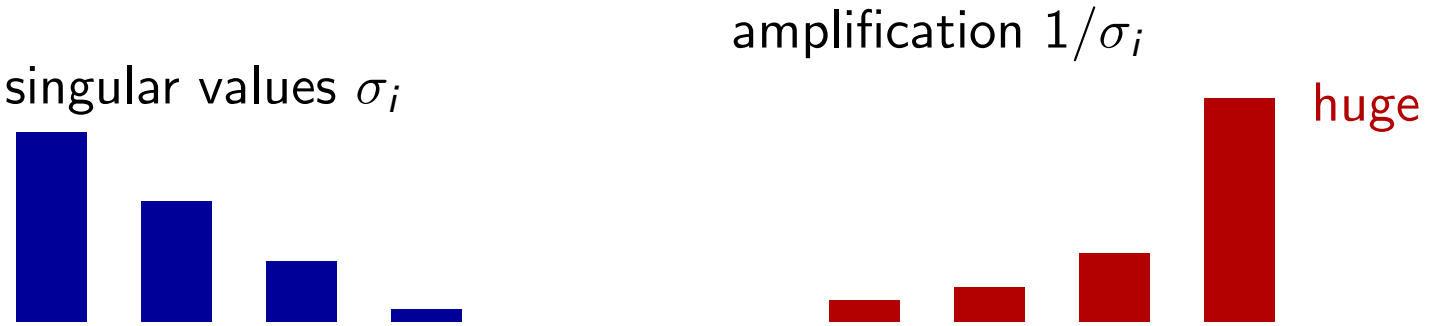
- If A blurred the image, apply A^{-1}: simple, and it uses no prior.

However, this ML solution has a number of problems.

Why inverting A amplifies the noise

The **singular value decomposition (SVD)** of A is $A = U\Sigma V^t$, where U and V are orthonormal $N \times N$ matrices and $\Sigma = \text{diag}\{\sigma_1, \cdots, \sigma_N\}$ holds the singular values in monotone decreasing order. Then

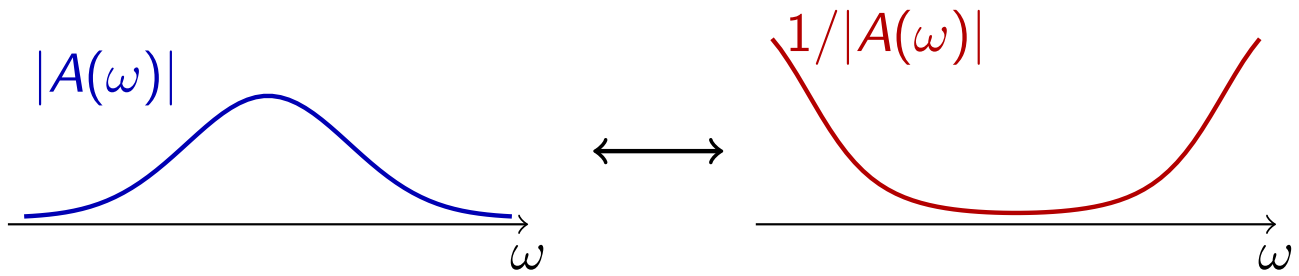
$$\begin{aligned} \hat{X}_{ML} &= A^{-1}Y = X + A^{-1}W = X + V\Sigma^{-1}U^tW \\ &= X + V\Sigma^{-1}\tilde{W}, \qquad \tilde{W} \sim N(0, \sigma^2 I) . \end{aligned}$$



If the small singular values are far below the noise standard deviation ($\sigma_N \ll \sigma$), the inversion **amplifies the noise** in W and renders the ML estimate unusable.

Inverting a blurring filter amplifies suppressed frequencies

A blurring filter often has a transfer function that is nearly zero at certain frequencies. Inverting A then amplifies exactly those frequencies.



This amplification increases both the signal and the noise in Y, so it produces a very noisy result.

- The **condition number** of A is $\kappa(A) = |\sigma_1/\sigma_N|$. The ML estimate tends to be unstable whenever $\kappa(A)$ is very large.
- Then the **dynamic range** is large, causing numerical instabilities.

MAP restoration is data fit plus a quadratic image penalty

From Bayes' rule, evaluating $\log p_{x|y}(x|y)$ as a function of x and ignoring any dependence on y,

$$\log p_{x|y}(x|y) = \underbrace{-\frac{1}{2\sigma^2}||y - Ax||^2}_{\text{fit the data}} \quad \underbrace{-\frac{1}{2}x^t Bx}_{\text{prefer plausible images}} + c(y) \ , \qquad (1)$$

where $c(y)$ is some as yet unknown function of y.

- Terms with only y do not change the optimizer; $p_y(y)$ is not needed.
- The prior term regularizes the inversion, our first such problem.

The posterior distribution

Equation (1) is a quadratic function of x, so completing the square puts it in the standard form

$$\log p_{x|y}(x|y) = -\frac{1}{2}(x-\mu(y))^t R^{-1}(x-\mu(y)) + c'(y) \ ,$$

where $\mu(y)$ and R are given by

$$\begin{aligned} \mu(y) &= \left(A^t A + \sigma^2 B\right)^{-1} A^t y \ , \\ R &= \left(\frac{1}{\sigma^2} A^t A + B\right)^{-1} \ . \end{aligned}$$

- $\mu(y)$ is a linear function of y.
- R is the conditional variance of x given y

The MMSE and MAP estimates are equal

The MMSE estimator is the conditional mean, and the MAP estimate is the conditional mode. The Gaussian posterior is symmetric and unimodal, so its mean and its mode are the same point:

$$\hat{x}_{MMSE} = \hat{x}_{MAP} = \mu(y) = \left(A^t A + \sigma^2 B\right)^{-1} A^t y \ .$$

- Always true when all random quantities are **jointly Gaussian**.
- They differ for non-Gaussian priors, which model image edges better.

Computing the MAP estimate: direct inversion is impractical

The MAP (and MMSE) estimate for $Y = AX + W$ with a GMRF prior is $\hat{x} = H^{-1}A^t y$, where $H \stackrel{\triangle}{=} A^t A + \sigma^2 B$. It seems our problem is solved, but the matrix H is enormous:

$$\begin{aligned} &1 \text{ megapixel image} \\ &N = 10^6 \text{ unknowns} \\ &H \text{ is } 10^6 \times 10^6 \qquad \text{~~form and invert } H\text{~~} \\ &10^{12} \text{ numbers} \\ &\approx 8 \text{ terabytes} \end{aligned}$$

So we go back to first principles. Substituting $p_{y|x}(y|x)$ and $p_x(x)$ into the MAP equation yields

$$\hat{x}_{MAP} = \arg\min_{x \in \mathbb{R}^N} f(x) \ , \qquad f(x) \stackrel{\triangle}{=} \frac{1}{2\sigma^2}||y - Ax||^2 + \frac{1}{2}x^t Bx \ .$$

We do not need the inverse. We need the minimizer of $f(x)$.

The limitation of the Gaussian prior

The GMRF prior has an important practical shortcoming.

Its most important shortcoming is that the GMRF **does not properly model the discontinuous edges** that occur in real image data, so the associated restorations tend to be **overly smooth**.

- Chapter 6 addresses this with non-Gaussian MRFs.

Summary

- The MAP estimate splits an inverse problem into a forward model and a prior model, $\hat{x}_{MAP} = \arg\min_x \left\{ -\log p_{y|x}(y|x) - \log p_x(x) \right\}$, and the example today is $Y = AX + W$ with $W \sim N(0, \sigma^2 I)$.
- The ML estimate $\hat{x}_{ML} = A^{-1}y$ amplifies the noise, because the noise term in $\hat{X}_{ML} = X + V\Sigma^{-1}\tilde{W}$ is large whenever $\sigma_N \ll \sigma$, and at $A = I$ it returns the data unchanged.
- The GMRF prior $p_x(x) \propto \exp\{-\frac{1}{2}x^t Bx\}$ adds the penalty $\frac{1}{2}x^t Bx$ to the data fit term $\frac{1}{2\sigma^2}||y - Ax||^2$.
- The posterior is Gaussian and its precisions add, $R^{-1} = \frac{1}{\sigma^2}A^t A + B$, so $\hat{x}_{MAP} = \hat{x}_{MMSE} = \mu(y) = \left(A^t A + \sigma^2 B\right)^{-1} A^t y$.
- The closed form is impractical at image scale, since H is $10^6 \times 10^6$ for a one megapixel image, so we minimize the MAP cost $f(x)$ by numerical optimization (next lecture).

Reading: Chapter 5.1–5.2. **Next lecture:** Chapter 5.3, gradient descent optimization.