UNIVERSITY OF MINNESOTA
Driven to Discover℠

EE 5561: Image Processing and Applications

Lecture 10

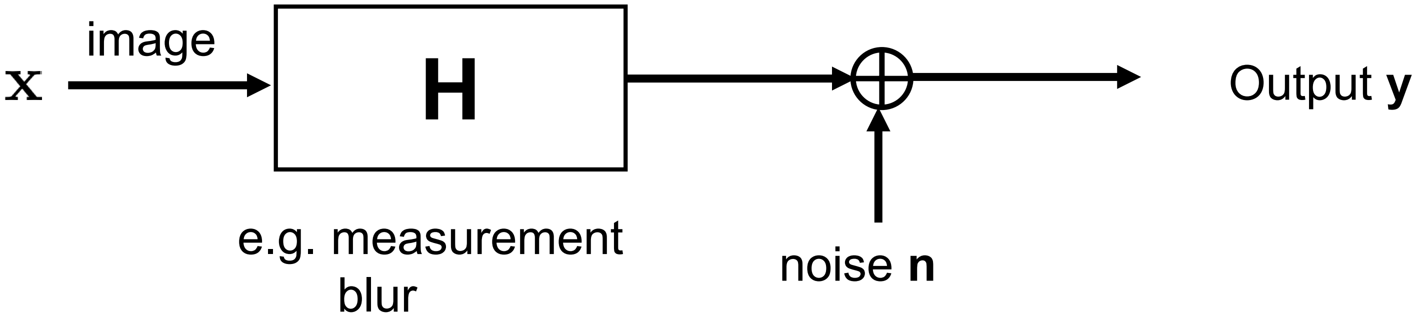
Mehmet Akçakaya

Recap of Last Lecture

- Start on image restoration
  - Denoising with Wiener filter
    - Statistical model for noise & image
    - Only need correlation matrices of these to generate a linear filter
  - Performance measures in restoration

- Today:
  - More general inverse methods

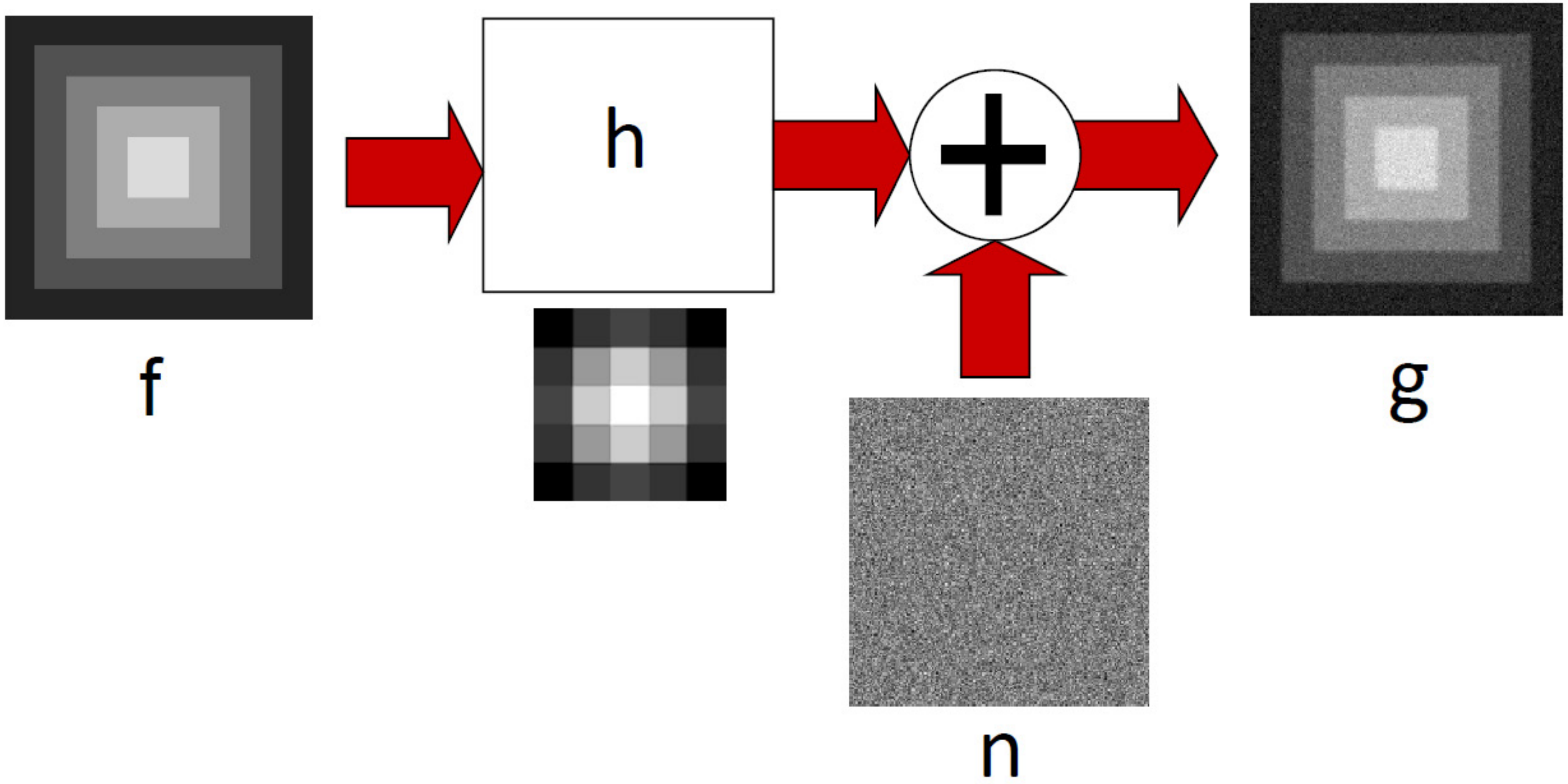
General Inverse Problem



$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

- We already saw Wiener filter, which has a statistical model for **x** and **n**
- What is a model for **n** in practice?
  - Most common (and physically relevant) is additive Gaussian $\mathbf{n} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K})$
  - Commonly we have white Gaussian noise, i.e. $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{K} = \sigma^2 \mathbf{I}$

Degraded System



Noise Examples

Additive Gaussian

Noise Examples

Poisson: Output pixel is from a Poisson distribution with parameter equal to input pixel value

Noise Examples

Salt & Pepper

Noise Examples

Speckle: Multiplicative (uniformly distributed) noise

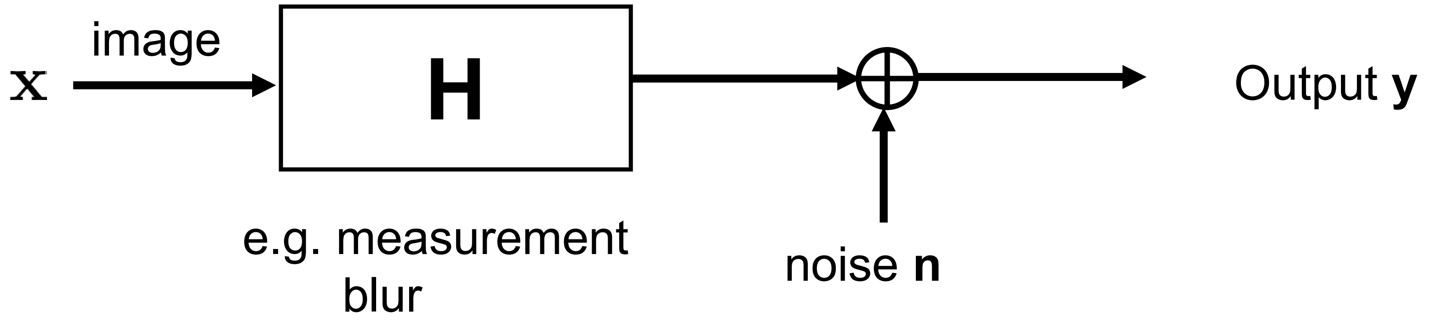
Blur Examples

Linear Motion Blur

Noise Examples

Atmospheric Turbulence: What you see when you look out the window behind an airplane engine

General Inverse Problem



$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

- What is the distribution of **y** then?
  - Is **x** a realization of a random vector?
  - If so, what's its distribution?
  - Or we can assume no statistical knowledge of **x**, then for i.i.d. Gaussian noise

$$p(\mathbf{y}|\mathbf{x}) \sim \mathcal{N}(\mathbf{H}\mathbf{x}, \sigma^2\mathbf{I})$$

  - How should we estimate **x** from **y**?

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-||\mathbf{y}-\mathbf{H}\mathbf{x}||_2^2/2\sigma^2}$$

General Inverse Problem

- One option is to do maximum likelihood estimation
- Find **x** that agrees "most" with **y**

based on the statistical model

$$
\begin{aligned}
\hat{\mathbf{x}}_{\mathrm{ML}} &= \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) \\
&= \arg\max_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x}) \\
&= \arg\max_{\mathbf{x}} \underbrace{-\frac{N}{2}\log(2\pi\sigma^2)}_{\text{no dependence on } \mathbf{x}} - \frac{1}{2\sigma^2}||\mathbf{y}-\mathbf{H}\mathbf{x}||_2^2 \\
&= \arg\max_{\mathbf{x}} -\frac{1}{2\sigma^2}||\mathbf{y}-\mathbf{H}\mathbf{x}||_2^2 = \arg\max_{\mathbf{x}} -||\mathbf{y}-\mathbf{H}\mathbf{x}||_2^2 \\
&= \arg\min_{\mathbf{x}} ||\mathbf{y}-\mathbf{H}\mathbf{x}||_2^2 = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}
\end{aligned}
$$

General Inverse Problem

- Maximum likelihood solution for white Gaussian noise is the least squares solution!

- This is a different viewpoint of getting to the same objective function/solution

- Possible issue: $\hat{\mathbf{x}}_{\mathrm{ML}}$ is noisy if **H** is ill-conditioned (usually the case)
- Condition number of **H** is the ratio of its largest and smallest singular values
- The smaller this number, the more "stable" the matrix is to small changes in its input argument

- Maximum-likelihood estimation is simple: Only needs information about **n**

General Inverse Problem

- What happens if the noise is not white, i.e. $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$?
  - Easy to show $$\log p(\mathbf{y}|\mathbf{x}) \sim -\frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x})^H \mathbf{K}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x})$$
  - Note **K**$^{-½}$ exists $$\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}^H \quad \Rightarrow \quad \mathbf{K}^{-\frac{1}{2}} = \mathbf{U}\mathbf{D}^{-\frac{1}{2}}\mathbf{U}^H$$
  - Define $$\mathbf{y}' = \mathbf{K}^{-\frac{1}{2}}\mathbf{y} \quad \text{and} \quad \mathbf{H}' = \mathbf{K}^{-\frac{1}{2}}\mathbf{H}$$
  - Then this is the same as the inverse problem

$$\mathbf{y}' = \mathbf{H}'\mathbf{x} + \mathbf{n}' \qquad \text{with} \qquad \mathbf{n}' \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Regularization

- When the objective function for the inverse problem is ill-posed, one can include additional information about the images to be recovered for finding a better solution

- What kind of additional information?
  - For instance, energy compaction of images in DCT/wavelet domains
  - Here we have additional knowledge about **x**

- This can also be viewed through a Bayesian viewpoint
  - First consider **x** as an instance of a random vector
  - Now if we have some knowledge about p(**x**), then we can change how we estimate **x**
  - Use maximum a-posteriori (MAP) estimation instead of maximum likelihood earlier

Regularization

- Bayesian viewpoint & MAP estimation

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y})$$

  - Contrast to $\hat{\mathbf{x}}_{\text{ML}} = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})$
  - Recall Bayes rule: $p(\mathbf{x}|\mathbf{y}) = \dfrac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}$
  - Thus

$$\begin{aligned}\hat{\mathbf{x}}_{\text{MAP}} &= \arg\max_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y}) = \arg\max_{\mathbf{x}} \log \big(p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\big) \\ &= \arg\max_{\mathbf{x}} \Big[\log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x})\Big]\end{aligned}$$

  - For $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ $\qquad \hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 - \underbrace{2\sigma^2}_{\alpha} \log p(\mathbf{x})$
  - In other words $\qquad \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \psi(\mathbf{x})$

Regularization

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \psi(\mathbf{x})$$

data fidelity/consistency:
make sure solution is consistent with data

regularization:
capture additional information about **x**

- For i.i.d. Gaussian noise, this is the general formulation for regularization
- This problem is called regularized/penalized least squares (very common in computational imaging)
- Bayesian formulation connects this with a prior distribution on random vectors (from which **x** is sampled), i.e. p(**x**)
- When the noise model changes the data fidelity term needs to be updated to match the appropriate log-likelihood, i.e. $-\log p(\mathbf{y}|\mathbf{x})$

Tikhonov Regularization

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \psi(\mathbf{x})$$

- Now we will start considering particular $\psi(\mathbf{x})$
- Going back to our original unregularized formulation, we have

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}$$

- i.e. We are inverting $(\mathbf{H}^H\mathbf{H})$ → Hermitian & square matrix → diagonalizable by unitary matrices, …
- Thus if the ratio of its largest and smallest eigenvalues are large, then it’s not a stable inversion (or extreme case if the smallest eigenvalues are ~ 0)

Tikhonov Regularization

- Tikhonov regularization (a.k.a. ridge regression) tackles this issue
- Essentially stabilizes the inversion process
- Consider

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \psi(\mathbf{x}) \qquad \text{with} \qquad \psi(\mathbf{x}) = \lambda||\mathbf{x}||_2^2$$

- This also has a closed form solution

$$\begin{aligned}
\hat{\mathbf{x}}_{\text{Tikh}} &= \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_2^2 \\
&= \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + ||\mathbf{0} - \sqrt{\lambda}\mathbf{I}\mathbf{x}||_2^2 \\
&= \arg\min_{\mathbf{x}} \left|\left| \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{H} \\ \sqrt{\lambda}\mathbf{I} \end{bmatrix} \mathbf{x} \right|\right|_2^2 \\
&= \left(\begin{bmatrix} \mathbf{H}^H & \sqrt{\lambda}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{H} \\ \sqrt{\lambda}\mathbf{I} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{H} \\ \sqrt{\lambda}\mathbf{I} \end{bmatrix}^H \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} = (\mathbf{H}^H\mathbf{H} + \lambda\mathbf{I})^{-1}\mathbf{H}^H\mathbf{y}
\end{aligned}$$

Tikhonov Regularization

$$\hat{\mathbf{x}}_{\text{Tikh}} = (\mathbf{H}^H\mathbf{H} + \lambda\mathbf{I})^{-1}\mathbf{H}^H\mathbf{y}$$

- In this new solution, we are inverting $(\mathbf{H}^H\mathbf{H} + \lambda\mathbf{I})$

- Increases all eigenvalues by λ → Ratio of largest to smallest is now smaller compared to no regularization

Tikhonov Regularization

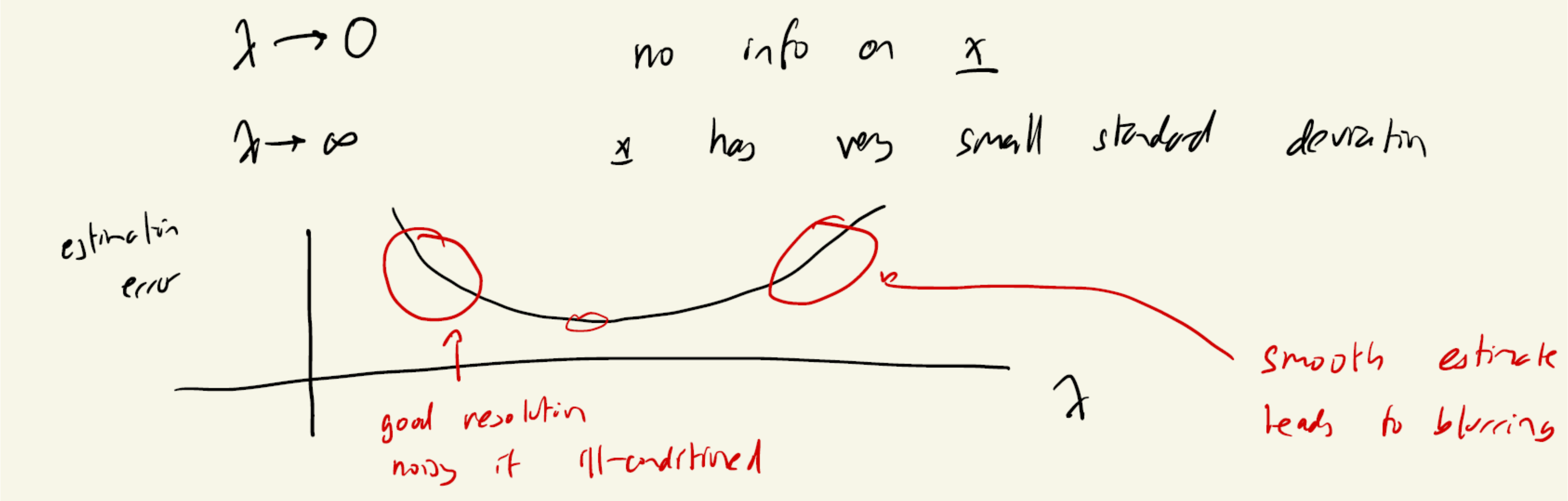
$$\hat{\mathbf{x}}_{\mathrm{Tikh}} = (\mathbf{H}^H\mathbf{H} + \lambda\mathbf{I})^{-1}\mathbf{H}^H\mathbf{y}$$

- What does $\psi(\mathbf{x}) = \lambda||\mathbf{x}||_2^2$ mean?
- All point with $||\mathbf{x}||_2^2 = C$ for some constant *C*, have the same $\psi(\mathbf{x})$
- i.e. All these points lie on an *n*-dimensional sphere
- A sphere is rotation-invariant → invariant under unitary transformations
- Thus, no difference calculating this in e.g. Fourier domain vs. image domain

- This regularizer has no information about **x** except some "energy regularization"

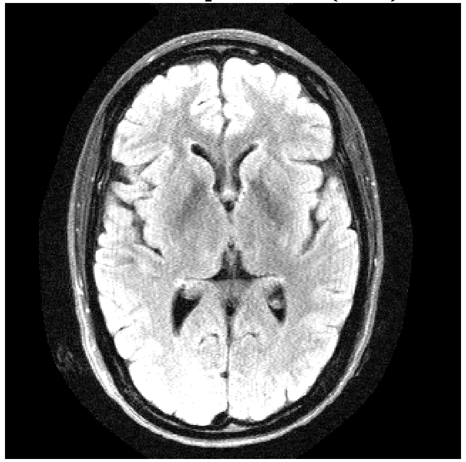
Tikhonov Regularization

- Practically
  - Helps numerically (as above) → noise reduction
  - Leads to blurring

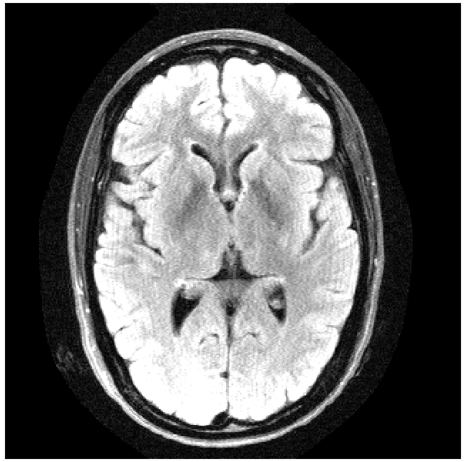


Example

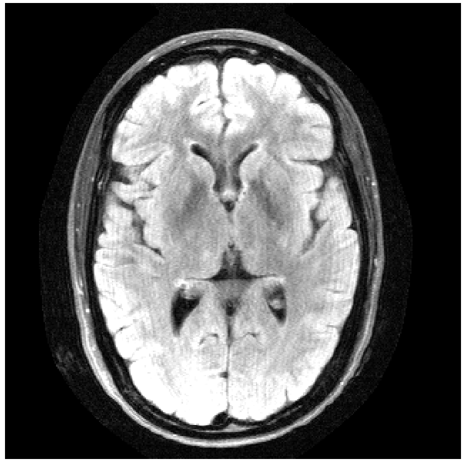
Unregularized Least Squares (LS)



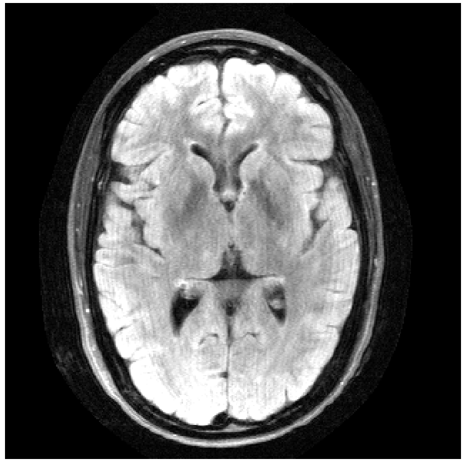
Tikhonov Regularized $\lambda = 0.01$



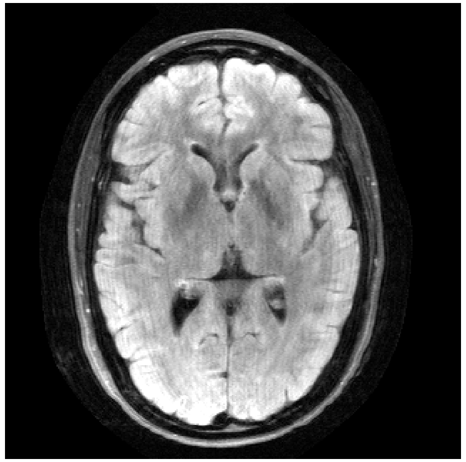
Tikhonov Regularized $\lambda = 0.05$



Tikhonov Regularized $\lambda = 0.1$



Tikhonov Regularized $\lambda = 0.2$



Tikhonov Regularization

- Variations:

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + ||\mathbf{\Gamma}\mathbf{x}||_2^2$$

- Here **Γ** captures additional information, e.g.
  - Discourage high spatial frequency oscillations, “roughness” penalty
  - Discourage disparities between neighboring pixels

- Solution

$$\begin{aligned}
\hat{\mathbf{x}}_{\text{G-Tikh}} &= \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + ||\mathbf{\Gamma}\mathbf{x}||_2^2 \\
&= \arg\min_{\mathbf{x}} \left|\left| \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{H} \\ \mathbf{\Gamma} \end{bmatrix} \mathbf{x} \right|\right|_2^2 \\
&= \left(\begin{bmatrix} \mathbf{H}^H & \mathbf{\Gamma}^H \end{bmatrix} \begin{bmatrix} \mathbf{H} \\ \mathbf{\Gamma} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{H} \\ \mathbf{\Gamma} \end{bmatrix}^H \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} = (\mathbf{H}^H\mathbf{H} + \mathbf{\Gamma}^H\mathbf{\Gamma})^{-1}\mathbf{H}^H\mathbf{y}
\end{aligned}$$

Tikhonov Regularization

- Example: Let's look at discouraging disparities between neighbor pixels
- In 1D: We want to minimize $\sum_{j=2}^{N}(x_j - x_{j-1})^2$
- i.e. The l_2 norm (squared) of the gradient
- Note we are using the l_2 norm to stay in the Tikhonov regularization framework

- Also note summation starts from 2 → We can add a circulant boundary condition to make our lives easier. Then define

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & \dots & & 0 & -1 \\ -1 & 1 & 0 & \dots & & 0 \\ \vdots & & \vdots & & & \\ 0 & \dots & 0 & -1 & & 1 \end{bmatrix}$$

N×*N* circulant

Easy to implement

Tikhonov Regularization

- Example: Let's look at discouraging disparities between neighbor pixels
- In 1D: We want to minimize $\sum_{j=2}^{N}(x_j - x_{j-1})^2$
- Then this regularization term is related to $||\mathbf{Cx}||_2^2$
- In the objective function, we will also add a weight term

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{Hx}||_2^2 + \lambda||\mathbf{Cx}||_2^2$$

or

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{Hx}||_2^2 + ||\mathbf{\Gamma x}||_2^2 \qquad \text{with} \qquad \mathbf{\Gamma} = \sqrt{\lambda}\mathbf{C}$$

- Solution: $\hat{\mathbf{x}}_{\text{G-Tikh}} = (\mathbf{H}^H\mathbf{H} + \lambda\mathbf{C}^H\mathbf{C})^{-1}\mathbf{H}^H\mathbf{y}$

- Same idea in 2D, with $\sum_{m,n}(x_{m,n} - x_{m,n-1})^2 + \sum_{m,n}(x_{m,n} - x_{m-1,n})^2$

Tikhonov regularization

- With proper design of Γ can be accurate & fast

JOURNAL OF MAGNETIC RESONANCE IMAGING 40:181–191 (2014)

Original Research

Fast Image Reconstruction With L2-Regularization

Berkin Bilgic, PhD,[1,2*] Itthi Chatnuntawech, SM,[1] Audrey P. Fan, SM,[1]
Kawin Setsompop, PhD,[2,3] Stephen F. Cauley, PhD,[2] Lawrence L. Wald, PhD,[2,3,4]
and Elfar Adalsteinsson, PhD[1,4]