Inverse Problems

$$y = Kf + z$$

Goal: Reconstruct f from indirect (and noisy) measurements y.

y = measurements
K = linear operator from signal space to measurement space
f = (unknown) signal/image to acquire
z = noise (we'll just take $z \sim$ iid Gaussian w/ var. σ^2)

Of course, if K is an orthogonal transform, $K^{-1} = K^*$ is perfectly well-defined and conditioned, and this just becomes a denoising problem.

But in most interesting situations, K^{-1} is not well-defined at all.

Particular examples:

- K = Radon transform (tomography, medical imaging)
- K = blurring convolution operator (deconvolution, low-speed photography)

First, recall that if the matrix is invertible, and we recover by setting

$$\tilde{f} = K^{-1}y = K^{-1}(Kf + z) = f + K^{-1}z$$

The risk will be

$$E\|f-\tilde{f}\|_2^2 = E\|K^{-1}z\|_2^2$$

$$=$$

__Solving Inverse Problems using the SVD__

$$K = U\Lambda V^*$$

U = orthobasis for the measurement space
V = orthobasis for the signal space
Λ = singular values (square-root of the eigenvalues of K^*K)

Say $y = Kf = U\Lambda V^*$ $\Rightarrow$ $U^*y = \Lambda V^* f$

$$\Rightarrow \langle u_\gamma, y\rangle = \lambda_\gamma \langle v_\gamma, f\rangle$$

$$\Rightarrow \langle v_\gamma, f\rangle = \lambda_\gamma^{-1} \langle u_\gamma, y\rangle$$

Since

$$f = \sum_\gamma \langle v_\gamma, f \rangle v_\gamma$$

we have another reproducing formula

$$f = \sum_\gamma \lambda_\gamma^{-1} \langle u_\gamma, Kf \rangle v_\gamma$$

(assuming all the $\lambda_\gamma > 0$)

If we add noise

$$y = Kf + z$$

and reconstruct

$$\tilde{f} = \sum_\gamma \lambda_\gamma^{-1} \langle u_\gamma, y \rangle v_\gamma \quad \Rightarrow f - \tilde{f} = \sum_\gamma \lambda_\gamma^{-1} \langle u_\gamma, z \rangle v_\gamma$$

$$= \sum_\gamma \lambda_\gamma^{-1} \underbrace{z'[\gamma]}_{\text{also iid}} v_\gamma$$

we can see things go badly for the γ s.t. λ_γ is very small

$$E\|f - \tilde{f}\|_2^2 = E\|V\Lambda^{-1}U^* z\|_2^2 = \sigma^2 \cdot \mathrm{Trace}(\Lambda^{-2})$$

$$= \sigma^2 \sum_\gamma 1/\lambda_\gamma^2$$

To make things stable, we <u>dampen</u> the terms in this sum to keep things from "blowing up".

Damping the SVD

$$\tilde{f} = \sum_\gamma w_\gamma \lambda_\gamma^{-1} \langle u_\gamma, y \rangle v_\gamma$$

The idea is to choose the w_γ small when λ_γ is large to "zero out" the effects of the noise in subspaces we aren't observing so well.

The ideal weights are

$$w_\gamma = \frac{|\langle f, v_\gamma \rangle|^2}{|\langle f, v_\gamma \rangle|^2 + \sigma_\gamma^2}$$

where $\sigma_\gamma = \sigma / \lambda_\gamma$

(This is the Wiener filter, again)

But as before, we don't know the $|\langle f, v_\gamma \rangle|^2$ in general.

We see, though, that the situation is basically the same as in denoising, but now we have an additional weighting by λ_γ^{-1}.

Oracle projection

$$y = Kf + z$$

In the SVD domain, we have

$$\beta = \alpha + z' \quad \text{where} \quad z'(\gamma) \sim \text{Normal}\left(0, \sigma^2/\lambda_\gamma^2\right)$$

Suppose we had an oracle that told us which elements of α were above the noise level. Then a "smart" thing to do would be to just keep these, setting

$$W_\gamma = \begin{cases} 1 & |\alpha_\gamma| > \sigma/\lambda_\gamma \\ 0 & \text{otherwise} \end{cases}$$

note, we need <u>strong</u> evidence of signal when $\lambda_\gamma \approx 0$

The error is then

$$E\|f - \tilde{f}\|_2^2 = \sum_\gamma \min\left(|\alpha_\gamma|^2, \sigma^2/\lambda_\gamma^2\right)$$

$$= \underbrace{\sum_{\gamma \in \Gamma^c} |\alpha_\gamma|^2}_{\text{bias}^2} + \underbrace{\sum_{\gamma \in \Gamma} \sigma^2/\lambda_\gamma^2}_{\text{variance}}$$

where $\Gamma = \{\gamma : \lambda_\gamma |\alpha_\gamma| \geq \sigma\}$

$$E\|f-\tilde{f}\|_2^2 = \underbrace{\sum_{\gamma\in\Gamma^c} |\alpha_\gamma|^2}_{\substack{\text{approximation error} \\ \text{(signal we are leaving out)}}} + \underbrace{\sum_{\gamma\in\Gamma} \sigma^2/\lambda_\gamma^2}_{\substack{\text{noise error} \\ \text{(noise on coefficients we are keeping)}}}$$

So again, good recovery when

$$\sum_{\gamma\in\Gamma^c} |\alpha_\gamma|^2 \cong \text{nonlinear approximation error}$$

is small.

Q: But in what basis?

A:

For classical damping (via SVD) methods to work, the signal must be sparse/compressible (i.e. decay) in the same basis which diagonalizes K.

This often is not the case at all

Typical example:

- operator K is diagonal in Fourier domain (convolution, tomography)

- signal/image is piecewise smooth (cross section of earth, brain or photo w/ edges)

Our goal, then, will be to find a representation that

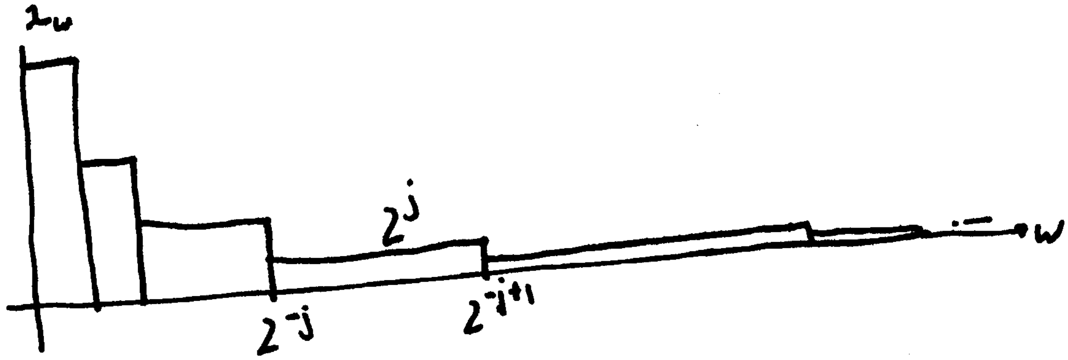
① is "almost" an SVD for operators K of interest

② is a good basis in which to approximate signals/images we wish to recover.

We will concentrate on a specific class of operators that are diagonalized in the Fourier domain, and whose spectrums obey a power law

We will see that wavelets almost diagonalize these types of operators.

Suppose we have an operator K which is diagonal in the Fourier domain whose spectrum looks like



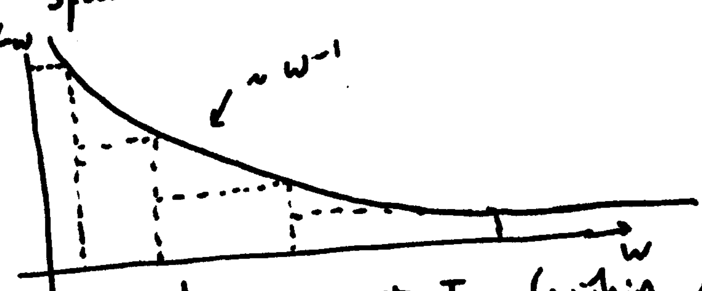
$j = 0, -1, -2, \ldots$

Since there are multiple eigenvectors of K^*K that have the same eigenvalue, the SVD is not unique.

What is another orthobasis which diagonalizes K ?

A:

Suppose the spectrum looks like



The spectrum is almost constant (within a factor of 2) over each subband

$\Rightarrow$ it will be almost diagonal in the wavelet domain

Wavelet - Vaguelette Decomposition (WVD)

We have an orthonormal wavelet decomposition $\{\psi_{j,k}\}$

Define a vaguelette basis $\{u_{j,k}\}$ which complements $\{\psi_{j,k}\}$ through

$$\langle f, \psi_{j,k} \rangle = 2^{-j/2} \langle u_{j,k}, Kf \rangle \quad \forall j,k$$

i.e.

$$\psi_{j,k} = 2^{-j/2} K^* u_{j,k}$$

and so $u_{j,k} = 2^{j/2} \underbrace{(K^*)^{+}}_{\text{pseudo-inverse of } K^*} \psi_{j,k}$

We have the reproducing formula

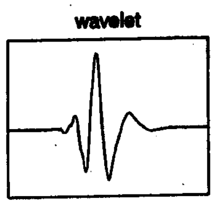
$$f = \sum_{j,k} \langle f, \psi_{j,k} \rangle \psi_{j,k} = \sum_{j,k} 2^{-j/2} \langle u_{j,k}, Kf \rangle \psi_{j,k}$$

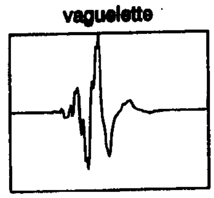
$$= \sum_{j,k} 2^{-j/2} \langle u_{j,k}, y \rangle \psi_{j,k}$$

where $y = Kf$
$\hookrightarrow$ observations

Take apart observations using vaguelettes, reweight by quasi-singular values $2^{-j/2}$, reconstruct using wavelets

Note: If $\lambda_w^2 \sim w^{-r}$, we use $2^{-jr/2}$ as the quasi-singular values above





The recovery procedure when y is contaminated with noise is

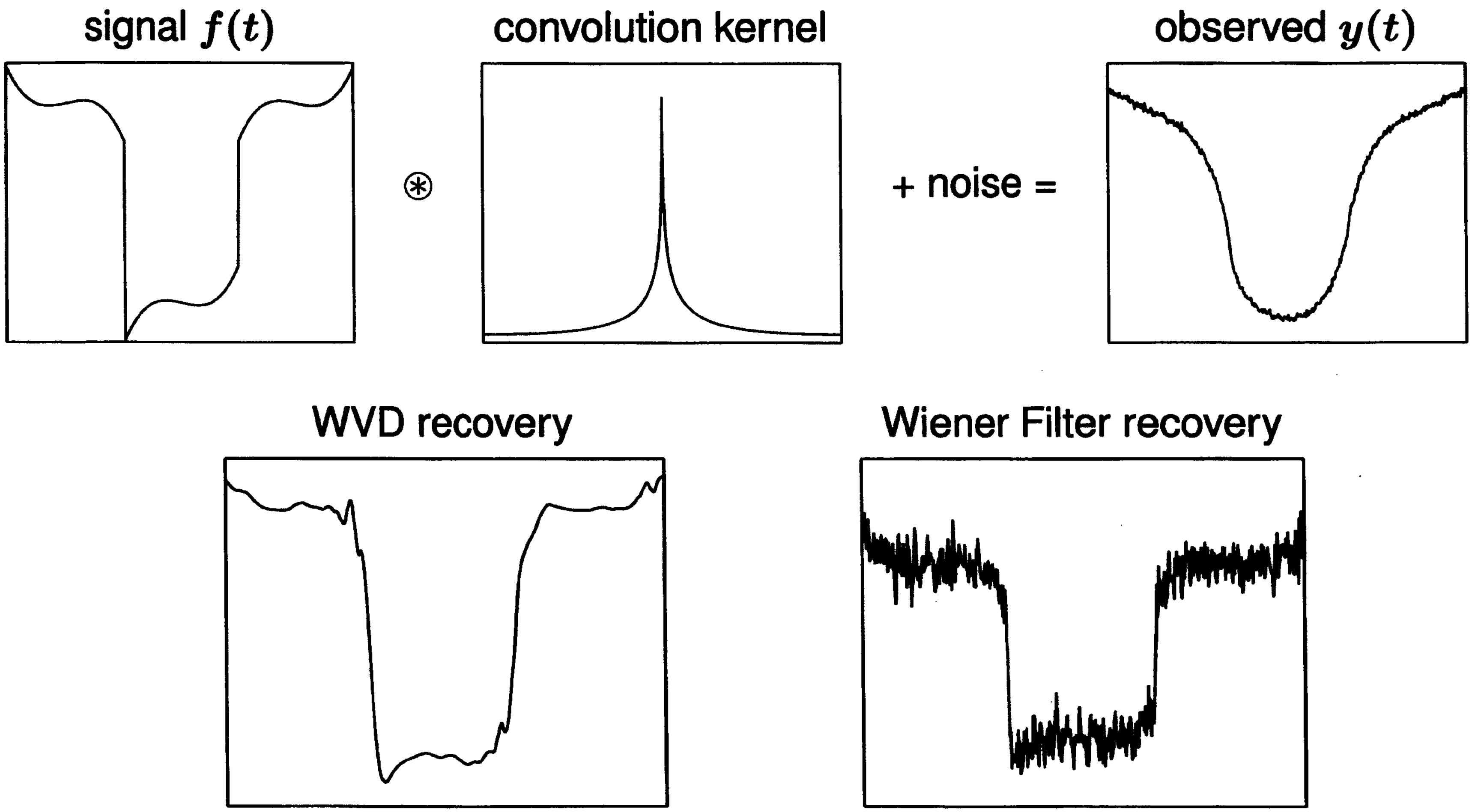
1. Decompose $y = Kf + z$ using vaguelettes
2. Threshold. The threshold is level dependent
   $T_j \sim 2^{-j/2}\sigma$ for $k_w^2 \sim w^{-1}$
3. Reconstruct using wavelets

The only real difference is that now we are thresholding noise which is <u>slightly correlated</u>.

Demonstrating the effectiveness of WVD boils down to showing that thresholding is also nearly minimax optimal for getting rid of slightly correlated noise.

Deconvolution Example

- Measure $y = Kf + \sigma z$, where K is $1/|\omega|$



WVD Optimality

Basically, if the wavelet coefficients of the signal we are interested in recovering decay fast enough, WVD thresholding is within a constant of minimax optimal.

Precisely, take as our class $\mathcal{F}_{\tau,p}$ the set of functions whose wavelet coefficients $\alpha_{j,k}$ satisfy

$$\left(\sum_{j \geq 0} 2^{jsp} \sum_{k} |\alpha_{j,k}|^p \right)^{1/p} \leq C$$

with $s = \tau + d(1/2 - 1/p)$.

d = dimension of domain ($=1$ for signals in time, $=2$ for images)

This is essentially a class of piecewise smooth functions (think of it as an extension of the l_p norm on the wavelet coefficients)

Say K has Fourier spectrum $\sim \omega^{-2r}$

The minimax risk obeys

$$r(\mathcal{F}_{s,r}) \sim \sigma^{2\tau}$$

for $t > \frac{2-p}{p} \cdot r + d(1/p - 1/2)$

this condition says that the operator spectrum cannot fall off too fast relative to the decay

where $\tau = \dfrac{t}{t + d/2 + r}$

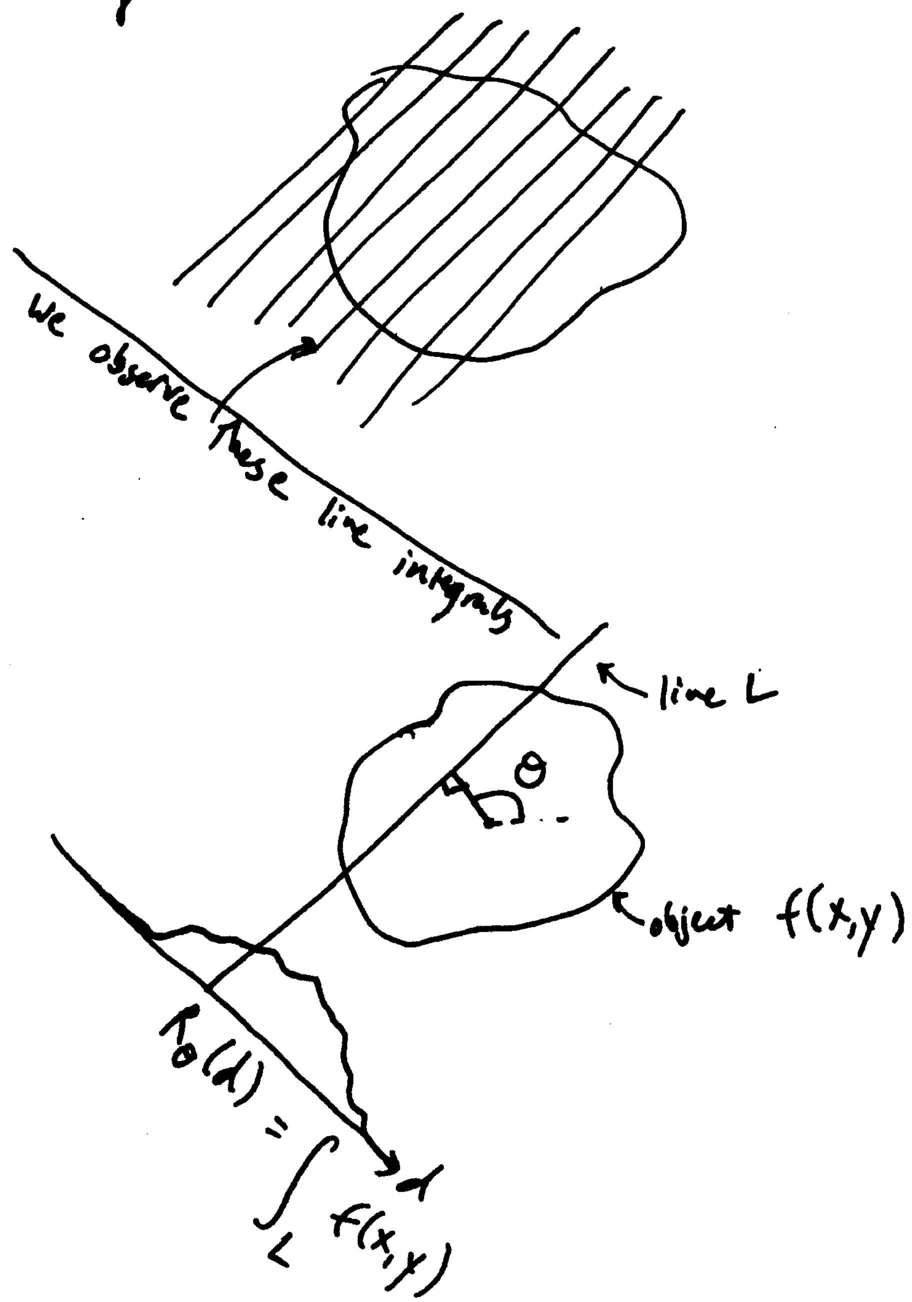
The WVD thresholding risk is also

$$r_{WVD}(\mathcal{F}_{s,r}) \sim \sigma^{2\tau}$$

The lesson is that WVD thresholding is a "good idea" for recovering piecewise smooth signals.

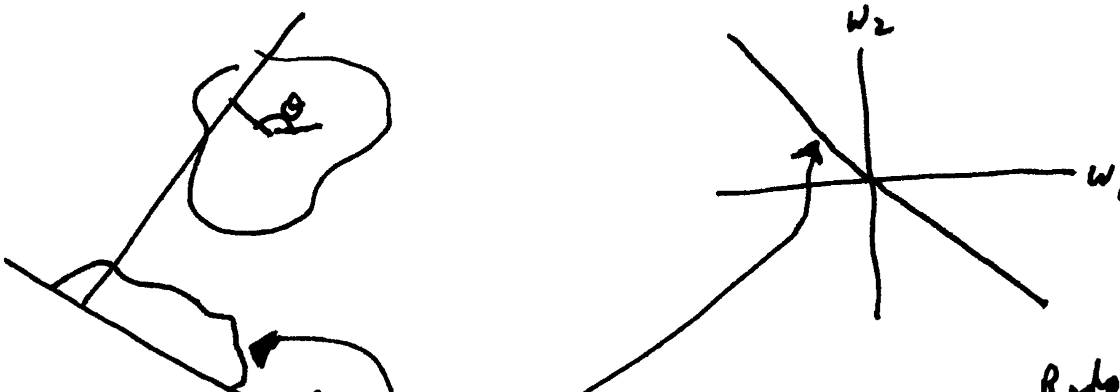
Interesting Operator: The Radon transform

In many scientific and engineering applications we are faced with the problem of recovering an image from 1D projections



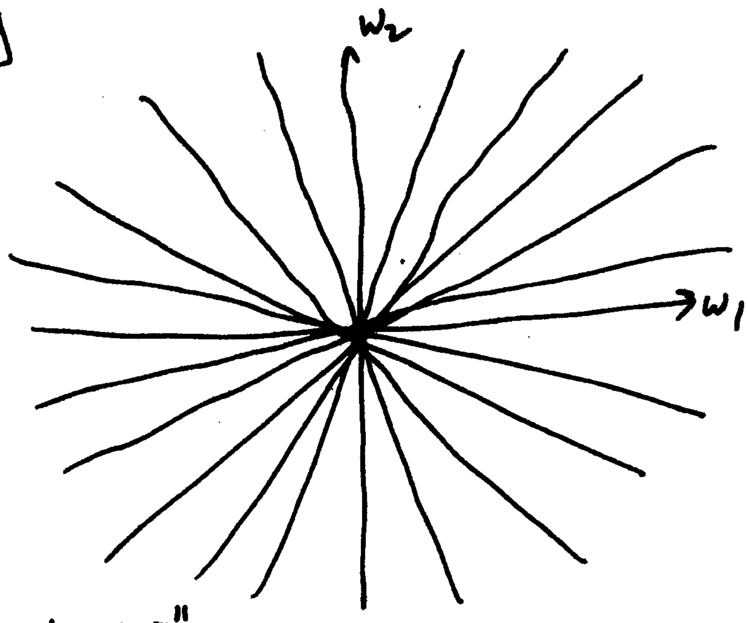
The "Radon slice" theorem says

$$\hat{\tilde{R}}_\theta = \hat{F}(w_1, w_2)\Big|_{w_1, w_2 \text{ on a radial line @ angle } \theta}$$



1D Fourier xform of this goes here in 2D Fourier plane $\Rightarrow$ Radon xform diagonalized in Fourier domain

Collecting a bunch of projections gives me this sampling



more at "center"

$\Rightarrow \Sigma_w^2 \sim w^{-1}$

$\Rightarrow$ Radon inversion w/ WVD is a "good idea"