



Wiener filtering

DS-GA 1013 / MATH-GA 2824 Mathematical Tools for Data Science

Carlos Fernandez-Granda

Prerequisites

Linear regression

Discrete Fourier transform

Linear translation-invariant models and convolution

Stationary signals and PCA

Signal estimation

Goal: Estimate N -dimensional signal \tilde{y} from N -dimensional measurements \tilde{x}

Optimal estimate? (in terms of MSE) $E(\tilde{y} | \tilde{x})$

Linear estimate?

Linear MMSE

$$E\left(\left\|\tilde{y} - B^T \tilde{x}\right\|_2^2\right) = \sum_{j=1}^N E\left[\left(\tilde{y}[j] - B_j^T \tilde{x}\right)^2\right]$$

Decouples into separate linear regression problems

$$\Sigma_{\tilde{x}}^{-1} (\Sigma_{\tilde{x}\tilde{y}})_j = \arg \min_{B_j} E\left[(\tilde{y}[j] - \tilde{x}^T B_j)^2\right]$$

$$\Sigma_{\tilde{x}}^{-1} \Sigma_{\tilde{x}\tilde{y}} = \arg \min_B E\left(\left\|\tilde{y} - B^T \tilde{x}\right\|_2^2\right)$$

$$\Sigma_{\tilde{x}\tilde{y}} := E\left(\tilde{x}\tilde{y}^T\right) = \begin{bmatrix} E(\tilde{x}[1]\tilde{y}[1]) & E(\tilde{x}[1]\tilde{y}[2]) & \cdots & E(\tilde{x}[1]\tilde{y}[N]) \\ E(\tilde{x}[2]\tilde{y}[1]) & E(\tilde{x}[2]\tilde{y}[2]) & \cdots & E(\tilde{x}[2]\tilde{y}[N]) \\ \vdots & \vdots & \ddots & \vdots \\ E(\tilde{x}[N]\tilde{y}[1]) & E(\tilde{x}[N]\tilde{y}[2]) & \cdots & E(\tilde{x}[N]\tilde{y}[N]) \end{bmatrix}$$

Problem

Covariance and cross-covariance matrices are $N \times N$

$$\Sigma_{\tilde{x}}^{-1} \Sigma_{\tilde{x}\tilde{y}} = \arg \min_B E \left(\left\| \tilde{y} - B^T \tilde{x} \right\|_2^2 \right)$$

For images, audio, and other signals, N is at least 10^4 (and often $\geq 10^6$)...

Fortunately, these signals often have translation-invariant statistics

Stationary signals

\tilde{x} is wide-sense or weak-sense stationary if

1. it has a constant mean

$$E(\tilde{x}[j]) = \mu, \quad 1 \leq j \leq N$$

2. there is an *autocovariance* function $a_{\tilde{x}}$ such that

$$\text{Cov}(\tilde{x}[j_1]\tilde{x}[j_2]) = \text{ac}_{\tilde{x}}(j_2 - j_1 \bmod N), \quad 0 \leq j_1, j_2 \leq N - 1$$

$$\begin{aligned}\Sigma_{\tilde{x}} &= \begin{bmatrix} a_{\tilde{x}} & a_{\tilde{x}}^{\downarrow 1} & a_{\tilde{x}}^{\downarrow 2} & \cdots a_{\tilde{x}}^{\downarrow N-1} \end{bmatrix} \\ &= \frac{1}{N} F^* \text{diag}(\hat{a}_{\tilde{x}}) F\end{aligned}$$

But what about $\Sigma_{\tilde{x}\tilde{y}}$?

Joint stationarity

\tilde{x} and \tilde{y} are jointly wide-sense or weak-sense stationary if

1. they are each wide-sense or weak-sense stationary
2. there is a function $cc_{\tilde{x}, \tilde{y}}$ such that

$$E(\tilde{x}[j_1]\tilde{y}[j_2]) = cc_{\tilde{x}\tilde{y}}(j_2 - j_1 \bmod N), \quad 0 \leq j_1, j_2 \leq N - 1$$

i.e. they have **translation-invariant** cross-covariance

Cross-covariance

$\text{cc}_{\tilde{x}\tilde{y}}$ is the cross-covariance of \tilde{x} and \tilde{y}

$$\begin{aligned}\Sigma_{\tilde{x}\tilde{y}} &:= E \left(\tilde{x} \tilde{y}^T \right) \\ &= \begin{bmatrix} c_{\tilde{x}\tilde{y}} & c_{\tilde{x}\tilde{y}}^{\downarrow 1} & c_{\tilde{x}\tilde{y}}^{\downarrow 2} & \cdots & c_{\tilde{x}\tilde{y}}^{\downarrow N-1} \end{bmatrix} \\ &= \begin{bmatrix} \text{cc}_{\tilde{x}\tilde{y}}(0) & \text{cc}_{\tilde{x}\tilde{y}}(N-1) & \cdots & \text{cc}_{\tilde{x}\tilde{y}}(1) \\ \text{cc}_{\tilde{x}\tilde{y}}(1) & \text{cc}_{\tilde{x}\tilde{y}}(0) & \cdots & \text{cc}_{\tilde{x}\tilde{y}}(2) \\ \text{cc}_{\tilde{x}\tilde{y}}(2) & \text{cc}_{\tilde{x}\tilde{y}}(1) & \cdots & \text{cc}_{\tilde{x}\tilde{y}}(3) \\ && \ddots & \\ \text{cc}_{\tilde{x}\tilde{y}}(N-1) & \text{cc}_{\tilde{x}\tilde{y}}(N-2) & \cdots & \text{cc}_{\tilde{x}\tilde{y}}(0) \end{bmatrix} \\ c_{\tilde{x}\tilde{y}} &:= \begin{bmatrix} \text{cc}_{\tilde{x}\tilde{y}}(0) \\ \text{cc}_{\tilde{x}\tilde{y}}(1) \\ \text{cc}_{\tilde{x}\tilde{y}}(2) \\ \cdots \\ \text{cc}_{\tilde{x}\tilde{y}}(N-1) \end{bmatrix}\end{aligned}$$

Eigendecomposition of circulant matrix

$$\begin{aligned} C &:= [h \ h^{\downarrow 1} \ h^{\downarrow 2} \ \dots \ h^{\downarrow N-1}] \\ &= \frac{1}{N} F^* \operatorname{diag}(\hat{h}) F \end{aligned}$$

where F is the DFT matrix and \hat{h} is the DFT of the first column

$$\begin{aligned} \Sigma_{\tilde{x}\tilde{y}} &= \left[c_{\tilde{x}\tilde{y}} \ c_{\tilde{x}\tilde{y}}^{\downarrow 1} \ c_{\tilde{x}\tilde{y}}^{\downarrow 2} \ \dots \ c_{\tilde{x}\tilde{y}}^{\downarrow N-1} \right] \\ &= \frac{1}{N} F^* \operatorname{diag}(\hat{c}_{\tilde{x}\tilde{y}}) F \end{aligned}$$

Linear MMSE estimate for stationary signals

$$\Sigma_{\tilde{x}} = \frac{1}{N} F^* \text{diag}(\hat{a}_{\tilde{x}}) F$$

$$\Sigma_{\tilde{x}\tilde{y}} = \frac{1}{N} F^* \text{diag}(\hat{c}_{\tilde{x}}) F$$

$$\begin{aligned}\Sigma_{\tilde{x}}^{-1} \Sigma_{\tilde{x}\tilde{y}} &= \left(\frac{1}{N} F^* \text{diag}(\hat{a}_{\tilde{x}}) F \right)^{-1} \frac{1}{N} F^* \text{diag}(\hat{c}_{\tilde{x}}) F \\ &= \frac{1}{N} F^* \text{diag}(\hat{a}_{\tilde{x}}^{-1}) F \frac{1}{N} F^* \text{diag}(\hat{c}_{\tilde{x}}) F \\ &= \frac{1}{N} F^* \text{diag}(\hat{a}_{\tilde{x}}^{-1} \hat{c}_{\tilde{x}}) F\end{aligned}$$

Linear MMSE estimate for stationary signals

The optimal linear estimate weights each Fourier coefficient separately

$$\begin{aligned}\tilde{y}_{\text{est}} &= \Sigma_{\tilde{x}\tilde{y}}^T \Sigma_{\tilde{x}}^{-1} \tilde{x} \\ &= \frac{1}{N} F^* \text{diag}(\hat{a}_{\tilde{x}}^{-1} \hat{c}_{\tilde{x}}) F \tilde{x}\end{aligned}$$

Weights depend on statistics of Fourier coefficients of measurements and signal

Covariance of Fourier coefficients of measurements

$$\begin{aligned}\Sigma_{\tilde{x}_F} &:= E(F\tilde{x}(F\tilde{x})^*) \\&= F E(\tilde{x}\tilde{x}^T) F^* \\&= F\Sigma_{\tilde{x}}F^* \\&= F \frac{1}{N} F^* \text{diag}(\hat{a}_{\tilde{x}}) F F^* \\&= N \text{diag}(\hat{a}_{\tilde{x}})\end{aligned}$$

$$\hat{a}_{\tilde{x}}[k] = \frac{\text{Var}(\tilde{x}_F[k])}{N}, \quad 0 \leq k \leq N-1$$

Cross-covariance of Fourier coefficients of signal and measurements

$$\begin{aligned}\Sigma_{\tilde{x}_F \tilde{y}_F} &:= E(F \tilde{x}(F \tilde{y})^*) \\&= F E\left(\tilde{x} \tilde{y}^T\right) F^* \\&= F \Sigma_{\tilde{x} \tilde{y}} F^* \\&= F \frac{1}{N} F^* \operatorname{diag}(\hat{c}_{\tilde{x}}) F F^* \\&= N \operatorname{diag}(\hat{c}_{\tilde{x}})\end{aligned}$$

$$\hat{c}_{\tilde{x} \tilde{y}}[k] = \frac{\operatorname{Cov}(\tilde{x}_F[k], \tilde{y}_F[k])}{N}, \quad 0 \leq k \leq N - 1$$

Linear MMSE estimate for stationary signals

$$\begin{aligned}\Sigma_{\tilde{x}}^{-1} \Sigma_{\tilde{x}\tilde{y}} &= \frac{1}{N} F^* \text{diag}(\hat{a}_{\tilde{x}}^{-1} \hat{c}_{\tilde{x}}) F \\ &= \frac{1}{N} F^* \text{diag}_{k=0}^{N-1} \left(\frac{\text{Cov}(\tilde{x}_F[k], \tilde{y}_F[k])}{\text{Var}(\tilde{x}_F[k])} \right) F\end{aligned}$$

Wiener filter

Let \tilde{x} and \tilde{y} be zero-mean and jointly stationary

The linear estimate of \tilde{y} given \tilde{x} that minimizes MSE as the convolution of \tilde{x} with the Wiener filter w , defined by

$$\hat{w}[k] := \frac{\text{Cov}(\tilde{x}_F[k], \tilde{y}_F[k])}{\text{Var}(\tilde{x}_F[k])}, \quad 0 \leq k \leq N - 1$$

where \tilde{x}_F and \tilde{y}_F denote the DFT coefficients of \tilde{x} and \tilde{y} , and

$$\text{Cov}(\tilde{x}_F[k], \tilde{y}_F[k]) := E\left(\tilde{x}_F[k]\overline{\tilde{y}_F[k]}\right)$$

$$\text{Var}(\tilde{x}_F[k]) := E\left(|\tilde{x}_F[k]|^2\right), \quad 0 \leq k \leq N - 1$$

Denoising

Measurements

$$\tilde{x} = \tilde{y} + \tilde{z},$$

where \tilde{z} is zero-mean Gaussian noise with variance σ^2 , independent of \tilde{y}

Noise

Linear transformation $A\tilde{z}$ of a Gaussian vector with mean μ and covariance matrix Σ is Gaussian with mean $A\mu$ and covariance matrix $A\Sigma A^*$

Fourier coefficients of noise are Gaussian with zero mean and covariance matrix $F\sigma^2 F^* = N\sigma^2 I$ (**iid Gaussian with variance $N\sigma^2$**)

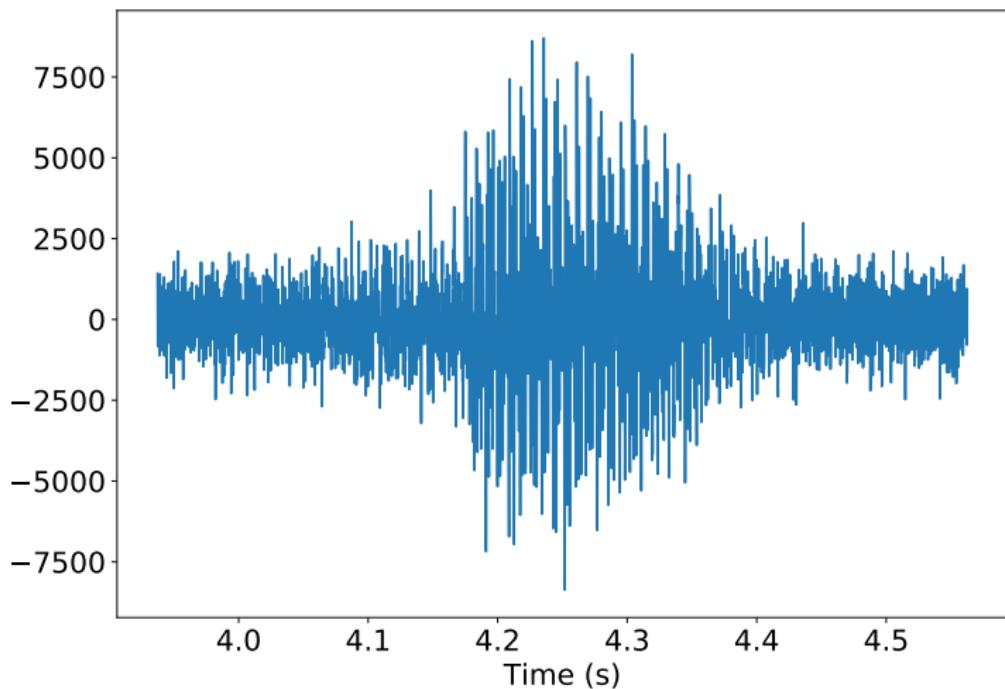
Wiener filter

$$\begin{aligned}\text{Cov}(\tilde{x}_F[k], \tilde{y}_F[k]) &= E\left(\tilde{x}_F[k]\overline{\tilde{y}_F[k]}\right) \\ &= E\left(\tilde{y}_F[k]\overline{\tilde{y}_F[k]}\right) + E\left(\tilde{z}_F[k]\overline{\tilde{y}_F[k]}\right) \\ &= \text{Var}(\tilde{y}_F[k])\end{aligned}$$

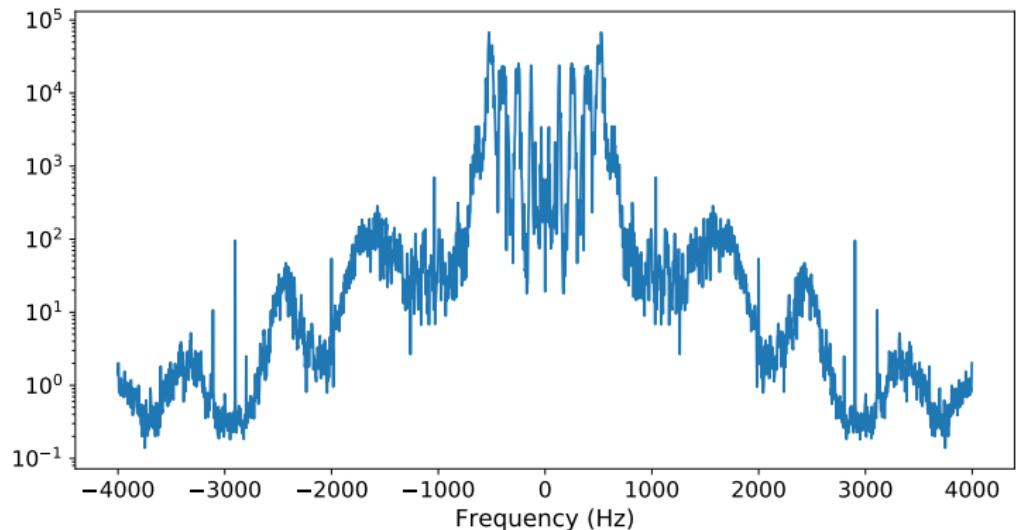
$$\begin{aligned}\text{Var}(\tilde{x}_F[k]) &= \text{Var}(\tilde{y}_F[k]) + \text{Var}(\tilde{z}_F[k]) \\ &= \text{Var}(\tilde{y}_F[k]) + N\sigma^2\end{aligned}$$

$$\hat{w}[k] = \frac{\text{Var}(\tilde{y}_F[k])}{\text{Var}(\tilde{y}_F[k]) + N\sigma^2}, \quad 0 \leq k \leq N-1$$

Audio data

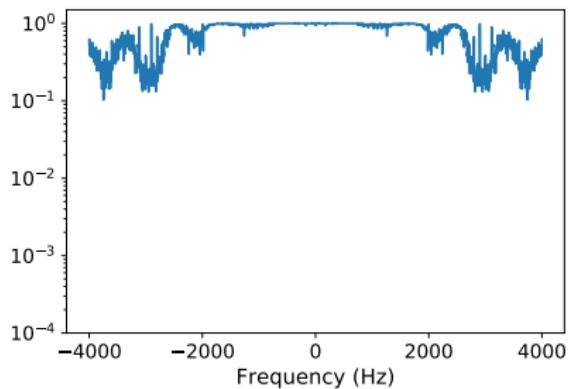


Audio data: Variance of Fourier coefficients

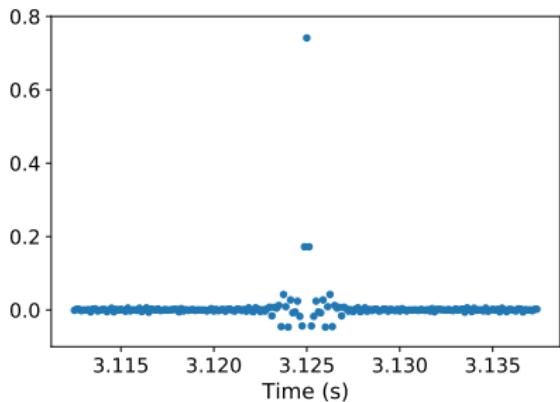


Wiener filter: $\sigma = 0.02$

Frequency

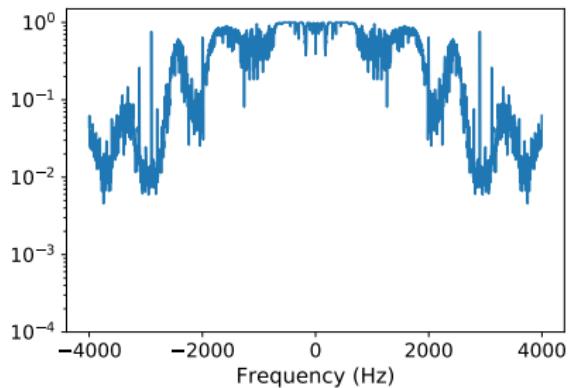


Time

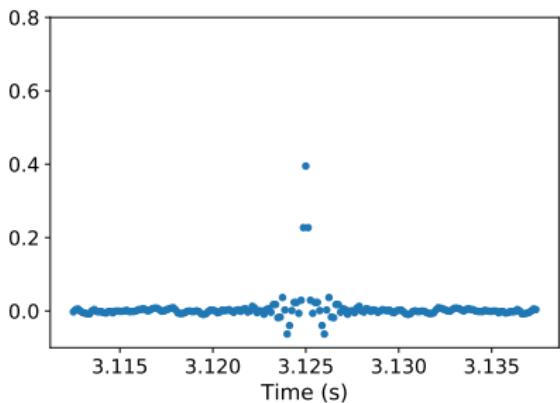


Wiener filter: $\sigma = 0.1$

Frequency

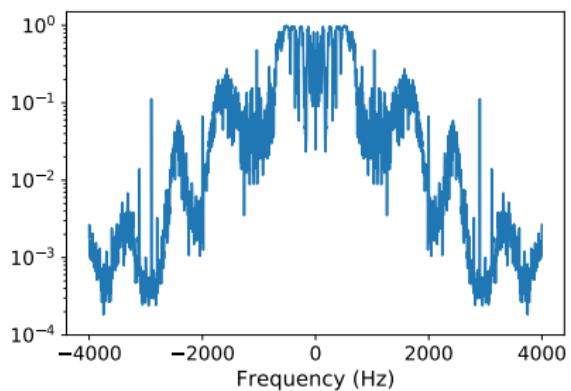


Time

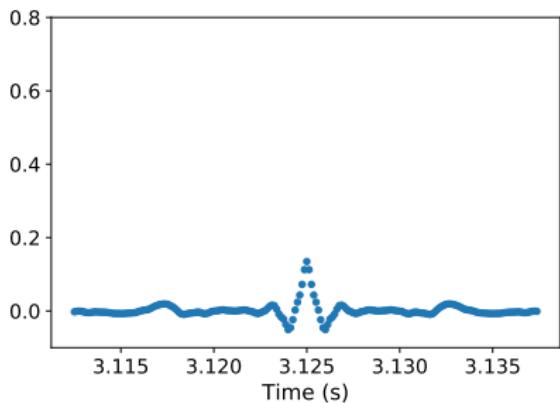


Wiener filter: $\sigma = 0.5$

Frequency

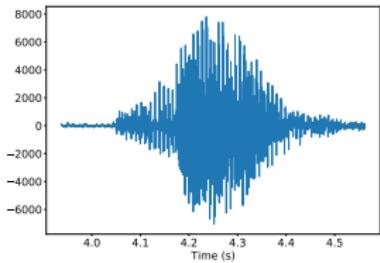


Time

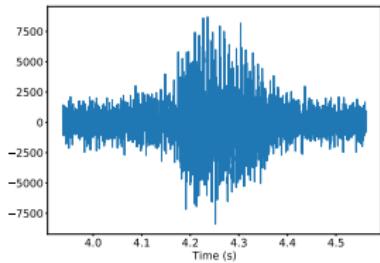


Example: $\sigma = 0.1$

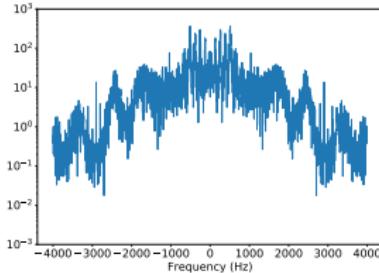
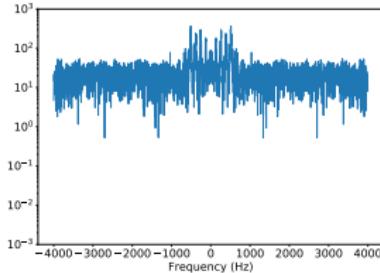
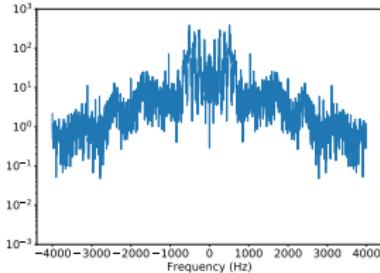
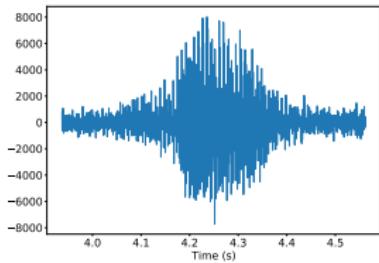
Clean



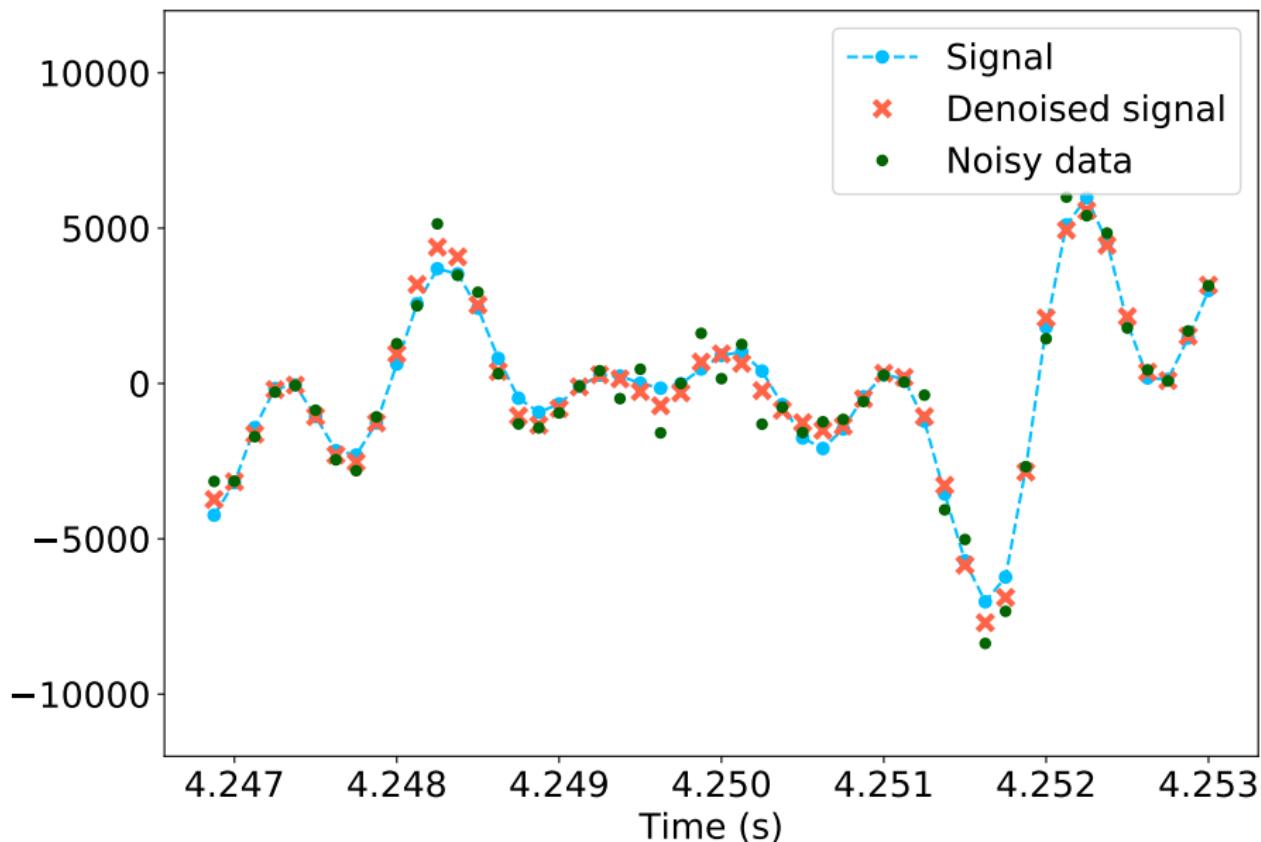
Noisy



Denoised

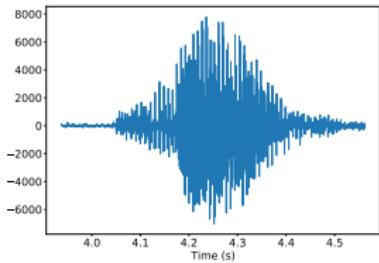


Example: $\sigma = 0.1$

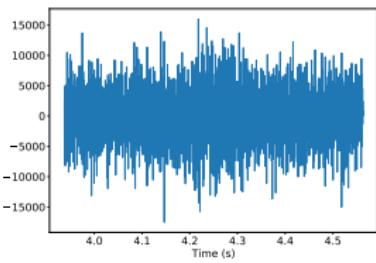


Example: $\sigma = 0.5$

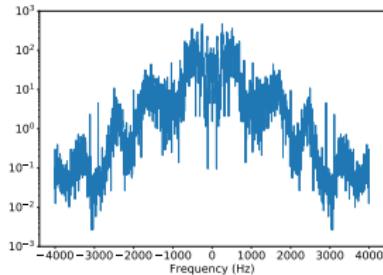
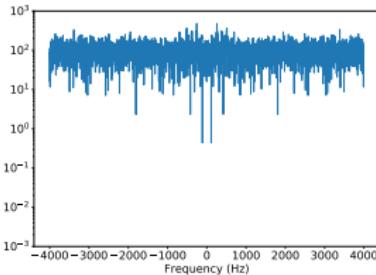
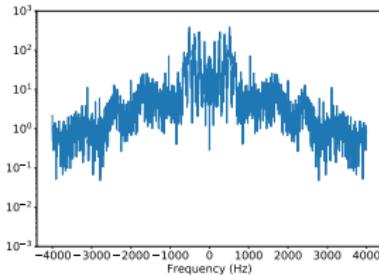
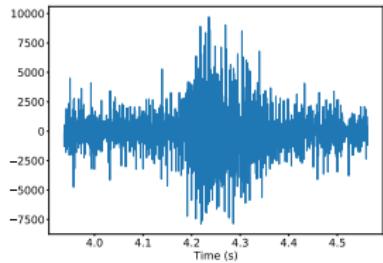
Clean



Noisy



Denoised



Example: $\sigma = 0.5$

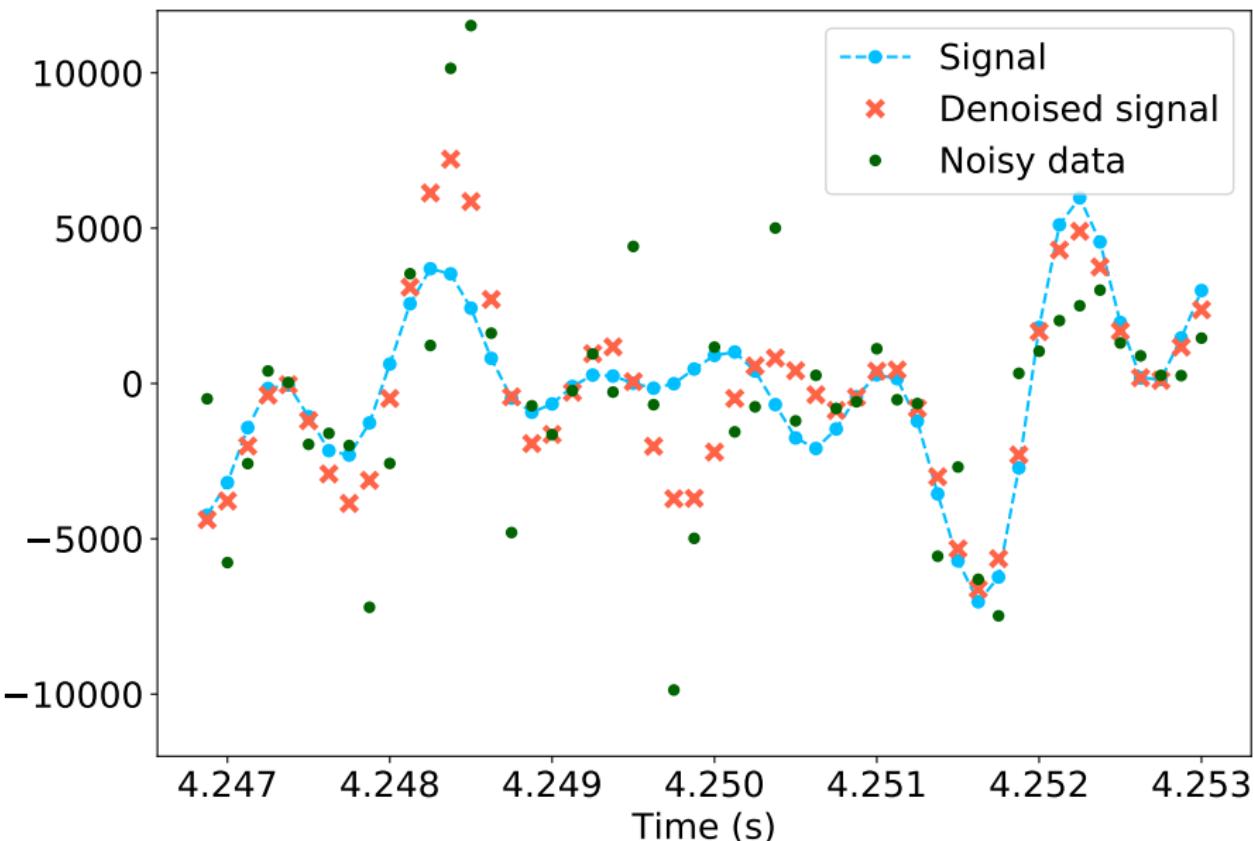


Image data

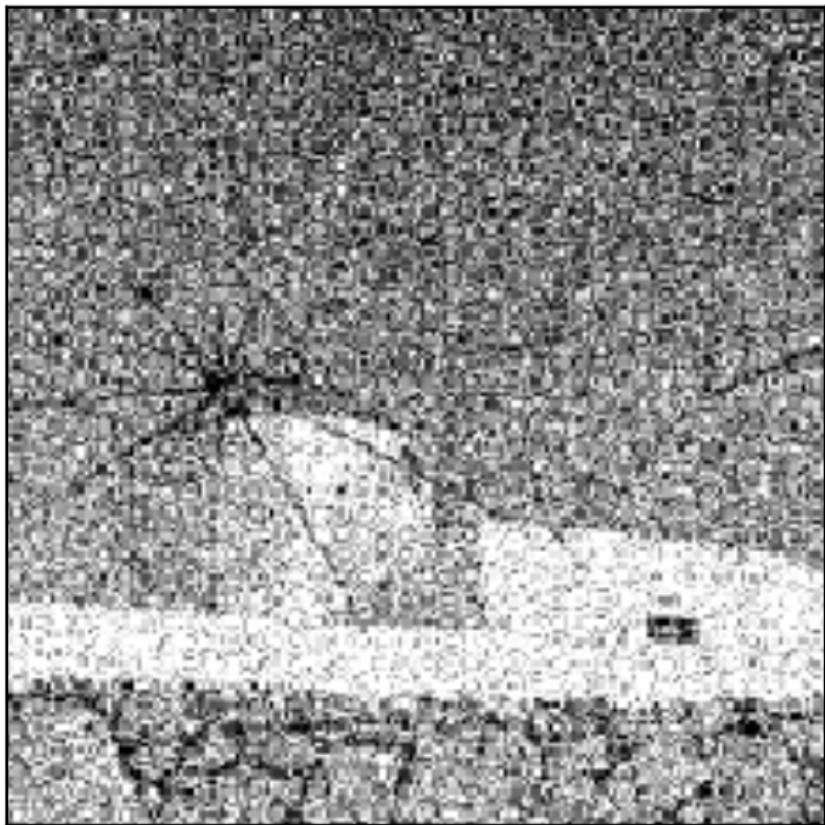
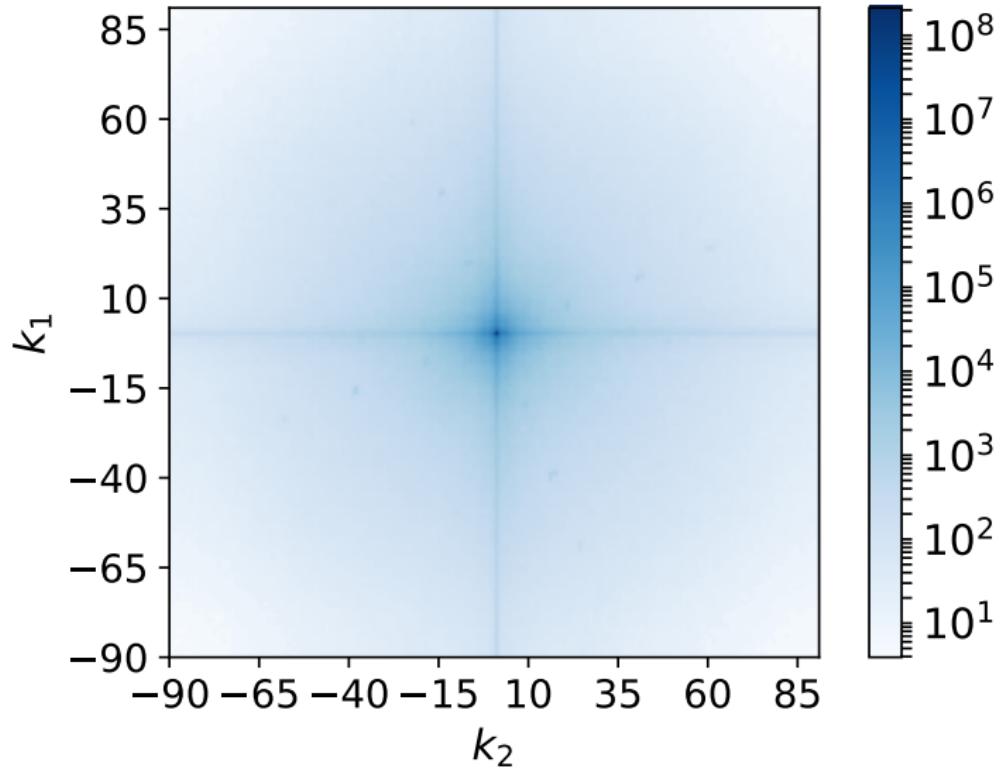
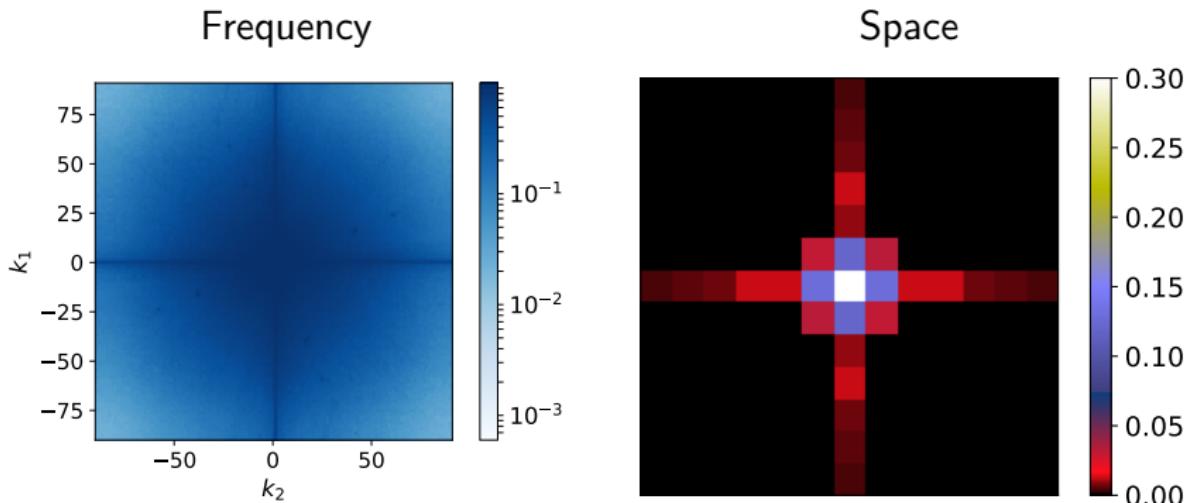


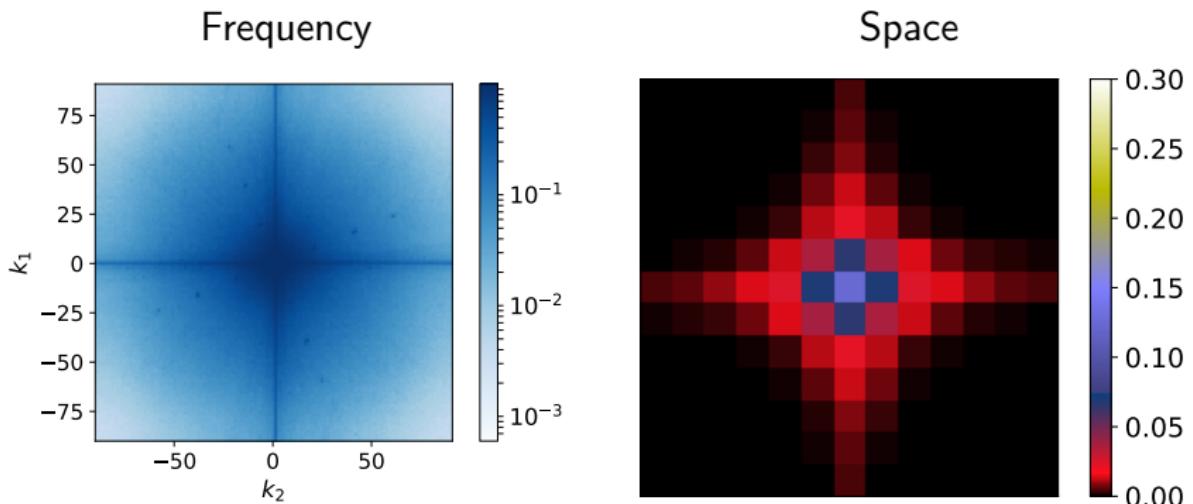
Image data: Variance of Fourier coefficients



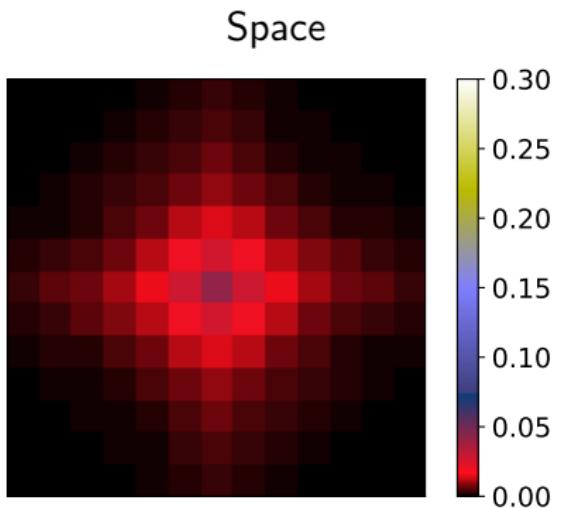
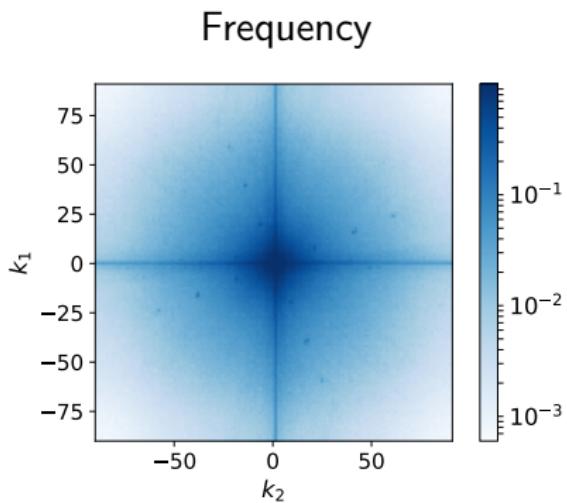
Wiener filter: $\sigma = 0.04$



Wiener filter: $\sigma = 0.1$



Wiener filter: $\sigma = 0.2$



Example: $\sigma = 0.04$

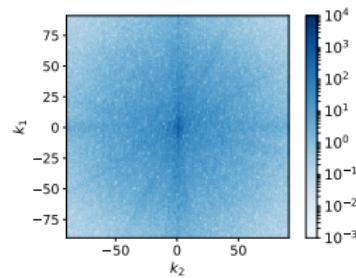
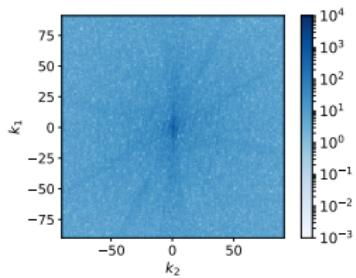
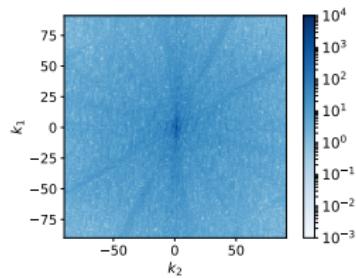
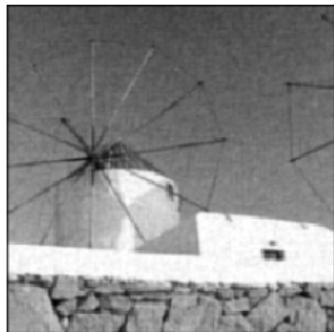
Clean



Noisy



Denoised

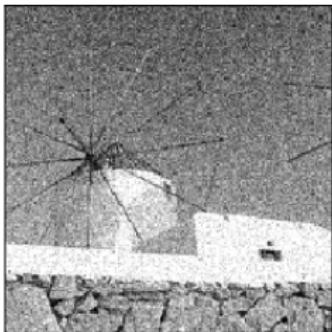


Example: $\sigma = 0.1$

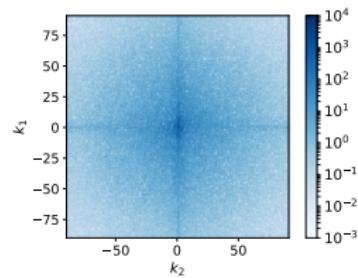
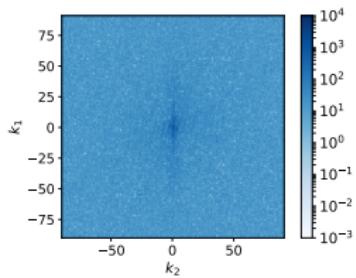
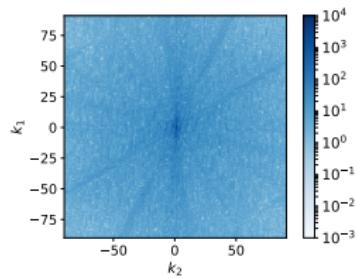
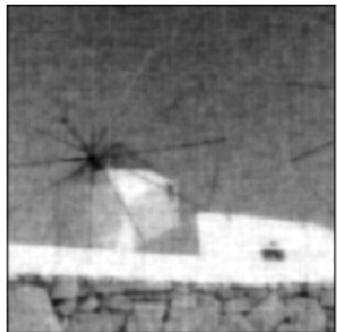
Clean



Noisy



Denoised

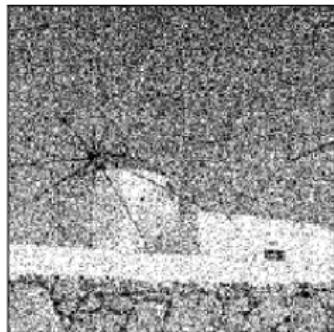


Example: $\sigma = 0.2$

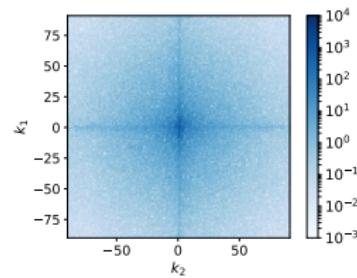
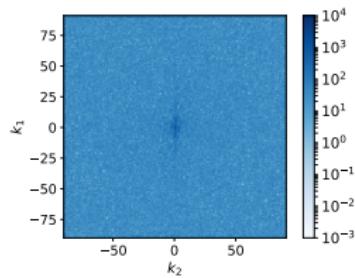
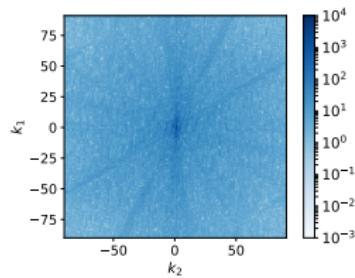
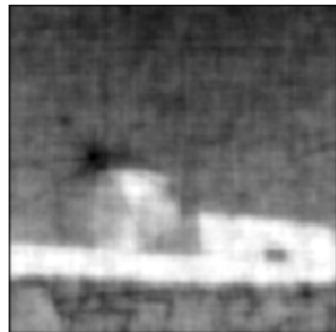
Clean



Noisy



Denoised



What have we learned?

Under stationarity assumptions, optimal linear estimation reduces to estimating each Fourier coefficient separately

This is equivalent to convolution with a fixed *Wiener* filter, which does not adapt to individual signals