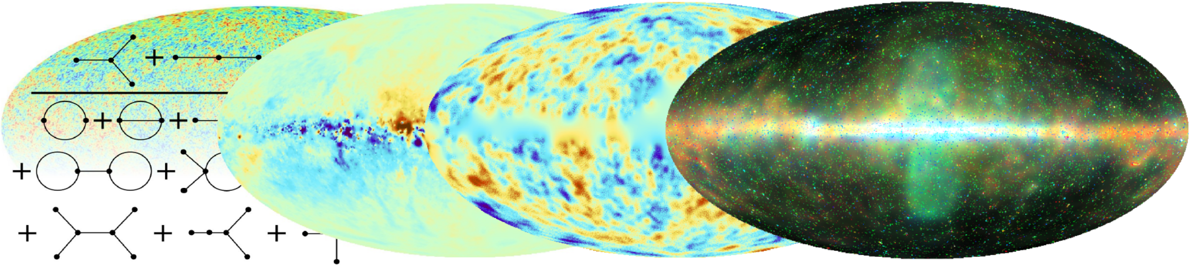


Information Field Theory

PD Dr. Torsten Enßlin

MPI for Astrophysics
LMU Munich

June 22, 2020

6. Information Hamiltonian

Recap:

Bayes theorem:

$$\mathcal{P}(s|d,\, I) \quad = \quad \frac{\mathcal{P}(d|s,\, I)\mathcal{P}(s|I)}{\mathcal{P}(d|I)} = \frac{\mathcal{P}(d,\, s|I)}{\int ds\, \mathcal{P}(d,\, s|I)} = \frac{e^{-\mathcal{H}(d,\, s|I)}}{\mathcal{Z}(d)}$$

information Hamiltonian : *surprise or information*

$$\mathcal{H}(d,\, s|I) \quad := \quad -\ln \mathcal{P}(d,\, s|I)$$

partition function: *evidence*

$$\mathcal{Z}(d|I) \quad := \quad \mathcal{P}(d|I) = \int ds\, \mathcal{P}(d,\, s|I) = \int ds\, e^{-\mathcal{H}(d,\, s|I)}$$

6.1 Linear Measurement of a Gaussian Signal with Gaussian Noise

Background information I:

$$
\begin{aligned}
d &= R\,s + n, \text{ either } d_i = \sum_j R_{ij} s_j + n_i \text{ or } d_i = \int dx\, R_{ix} s(x) + n_i,\ i.g.\ R \neq R^\dagger \\
\mathcal{P}(s|I) &= \mathcal{G}(s,\, S) = \frac{1}{\sqrt{|2\pi S|}} \exp\left(-\frac{1}{2} s^\dagger S^{-1} s\right),\ S = \langle s\, s^\dagger \rangle_{(s)} = \left(\langle s_x\, \overline{s_y} \rangle_{(s)}\right)_{xy} = S^\dagger > 0 \\
\mathcal{P}(n|s,\, I) &= \mathcal{G}(n,\, N),\ N = \langle n\, n^\dagger \rangle_{(n)} = \left(\langle n_i\, \overline{n_j} \rangle_{(n)}\right)_{ij} = N^\dagger > 0,\ N^{-1} = \left(N^{-1}\right)^\dagger > 0
\end{aligned}
$$

Calculation of the information Hamiltonian $\mathcal{H}(d, s|I)$:

$$
\begin{aligned}
\mathcal{H}(d, s|I) &= -\ln \mathcal{P}(d, s|I) \\
&= -\ln \left[\mathcal{P}(d|s,\, I)\, \mathcal{P}(s|I)\right] \\
&= -\ln \mathcal{P}(d|s,\, I) - \ln \mathcal{P}(s|I) \\
&= \mathcal{H}(d|s,\, I) + \mathcal{H}(s|I)
\end{aligned}
$$

6.1 Linear Measurement of a Gaussian Signal with Gaussian Noise

$$\mathcal{P}(s|I) = \frac{1}{\sqrt{|2\pi S|}} \exp\left(-\frac{1}{2}s^\dagger S^{-1} s\right)$$

$$\mathcal{H}(s|I) = -\ln \mathcal{P}(s|I) = \frac{1}{2}s^\dagger S^{-1} s + \frac{1}{2}\ln|2\pi S|$$

$$\mathcal{P}(d|s,\, I) = \int dn\, \mathcal{P}(d,\, n|s,\, I) = \int dn\, \underbrace{\mathcal{P}(d|s,\, n,\, I)}_{=\delta(d-(Rs+n))}\, \underbrace{\mathcal{P}(n|s,\, I)}_{=\mathcal{G}(n,N)} = \mathcal{G}(d-Rs, N)$$

$$\begin{aligned}
\mathcal{H}(d|s,\, I) &= -\ln \mathcal{P}(d|s,\, I) \\
&= \frac{1}{2}(d-Rs)^\dagger N^{-1}(d-Rs) + \frac{1}{2}\ln|2\pi N| \\
&= \frac{1}{2}[d^\dagger N^{-1} d - s^\dagger \underbrace{R^\dagger N^{-1} d}_{=:j} - \underbrace{d^\dagger N^{-1} R}_{=j^\dagger} s + s^\dagger R^\dagger N^{-1} R s + \ln|2\pi N|] \\
&= \frac{1}{2}[s^\dagger R^\dagger N^{-1} R s - s^\dagger j - j^\dagger s + d^\dagger N^{-1} d + \ln|2\pi N|]
\end{aligned}$$

6.1 Linear Measurement of a Gaussian Signal with Gaussian Noise

Summary:

$$
\begin{aligned}
d &= R\,s + n \\
\mathcal{P}(n,s) &= \mathcal{G}(n,N)\,\mathcal{G}(s,S) \\
\mathcal{H}(d,\,s) &= \mathcal{H}(d|s) + \mathcal{H}(s) \\
&= \frac{1}{2}[s^\dagger D^{-1} s - s^\dagger j - j^\dagger s] + \mathcal{H}_0 \\
D^{-1} &= S^{-1} + R^\dagger N^{-1} R = \text{information propagator}^{-1} \\
j &= R^\dagger N^{-1} d = \text{information source} \\
\mathcal{H}_0 &= d^\dagger N^{-1} d + \ln|2\pi N| + \ln|2\pi S| = \text{constants in } s \\
\left(D^{-1}\right)^\dagger &= \left(S^{-1}\right)^\dagger + \left(R^\dagger N^{-1} R\right)^\dagger = S^{-1} + R^\dagger\left(N^{-1}\right)^\dagger\left(R^\dagger\right)^\dagger = S^{-1} + R^\dagger N^{-1} R = D^{-1} \\
S, N &> 0 \Rightarrow S^{-1}, N^{-1} > 0 \Leftrightarrow \forall\, s, n \neq 0 :\; s^\dagger S^{-1} s,\; n^\dagger N^{-1} n > 0 \\
D^{-1} &= \underbrace{\underbrace{S^{-1}}_{>0} + \underbrace{R^\dagger N^{-1} R}_{\geq 0}}_{>0} \Rightarrow D \text{ exists}, D = D^\dagger > 0
\end{aligned}
$$

Quadratic Completion

"$\widehat{=}$": equal up to irrelevant constants, logarithmic brother of proportionality sign "$\propto$"

$$\begin{aligned}
\mathcal{H}(d,s|I) &\widehat{=} \frac{1}{2}[s^\dagger D^{-1} s - j^\dagger s - s^\dagger j] \\
&= \frac{1}{2}\left[s^\dagger D^{-1} s - j^\dagger D\, D^{-1}\, s - s^\dagger D^{-1} \underbrace{D j}_{=:m}\right] \\
&\widehat{=} \frac{1}{2}\left[s^\dagger D^{-1} s - \underbrace{(D j)^\dagger}_{=m^\dagger} D^{-1}\, s - s^\dagger D^{-1} m + m^\dagger D^{-1} m\right] \\
&= \frac{1}{2}(s-m)^\dagger\, D^{-1} (s-m)
\end{aligned}$$

Gaussian Posterior

$$
\begin{aligned}
\mathcal{H}(d, s|I) &= \mathcal{H}(d|s, I) + \mathcal{H}(s|I) = \frac{1}{2}(s-m)^\dagger D^{-1}(s-m) + \mathcal{H}_0' \\
\Rightarrow \mathcal{Z}(d) &= \int ds\, e^{-\mathcal{H}(d,s)} = \int ds\, e^{-\frac{1}{2}(s-m)^\dagger D^{-1}(s-m) - \mathcal{H}_0'} \\
&= e^{-\mathcal{H}_0'} \int ds'\, e^{-\frac{1}{2}s'^\dagger D^{-1} s'} = e^{-\mathcal{H}_0'}\sqrt{|2\pi D|} \\
\Rightarrow \mathcal{P}(s|d, I) &= \frac{\mathcal{P}(d, s|I)}{\mathcal{P}(d|I)} = \frac{e^{-\mathcal{H}(d,s|I)}}{\mathcal{Z}(d)} \\
&= \frac{e^{-\frac{1}{2}(s-m)^\dagger D^{-1}(s-m) - \mathcal{H}_0'}}{\sqrt{|2\pi D|}e^{-\mathcal{H}_0'}} = \frac{e^{-\frac{1}{2}(s-m)^\dagger D^{-1}(s-m)}}{\sqrt{|2\pi D|}} \\
&= \mathcal{G}(s-m, D)
\end{aligned}
$$

Gaussian Posterior

$$\mathcal{P}(s|d, I) = \mathcal{G}(s - m, D)$$

Mean:

$$m = \langle s \rangle_{(s|d,I)} = D\,j = \underbrace{\left(S^{-1} + R^\dagger N^{-1} R\right)^{-1} R^\dagger N^{-1}}_{=:F_\mathrm{W}=\text{Wiener filter}} d$$

Covariance:

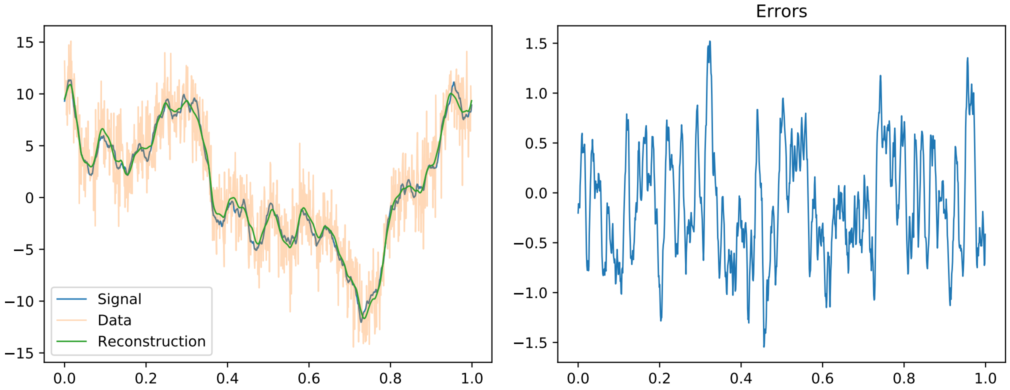
$$D = \langle (s-m)(s-m)^\dagger \rangle_{(s|d,I)} = \left(S^{-1} + R^\dagger N^{-1} R\right)^{-1} \leq S$$

$$D^{-1} = S^{-1} + R^\dagger N^{-1} R \geq S^{-1} \quad \square$$

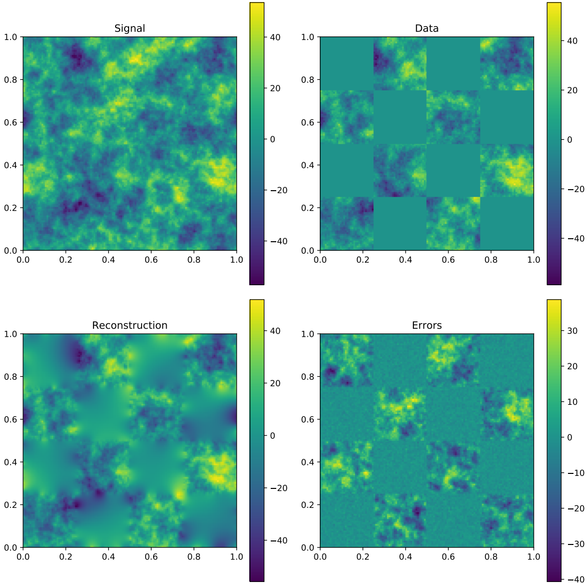
Information source: $j = R^\dagger N^{-1} d$

1σ-range: $s_x = m_x \pm \sqrt{D_{xx}}$

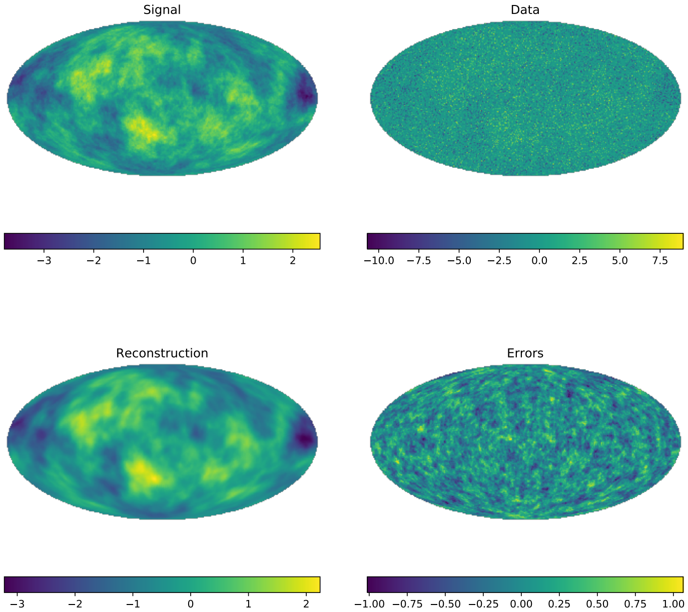
Wiener Filter in Action



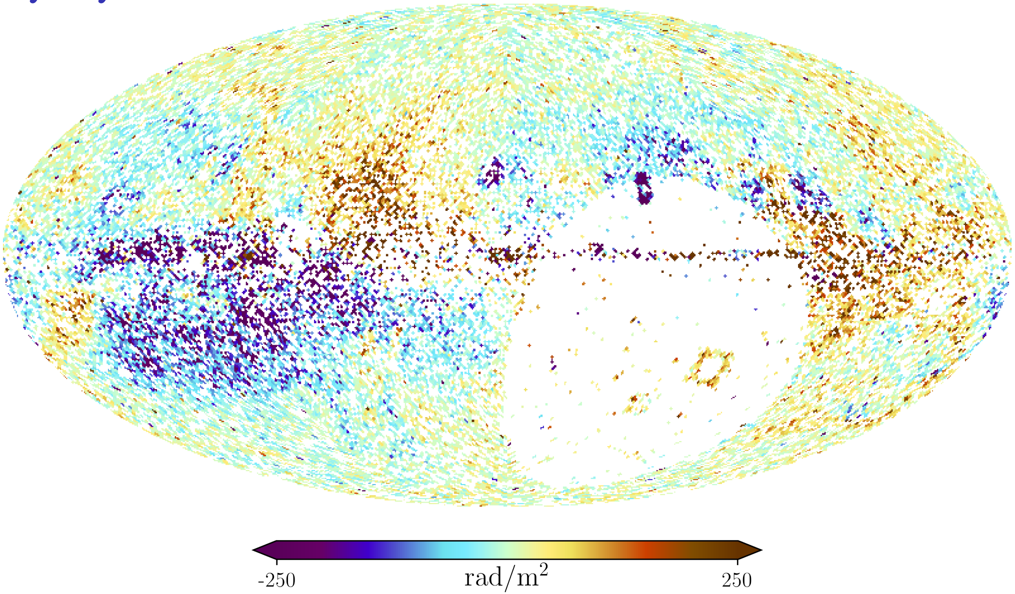
Wiener Filter in Action



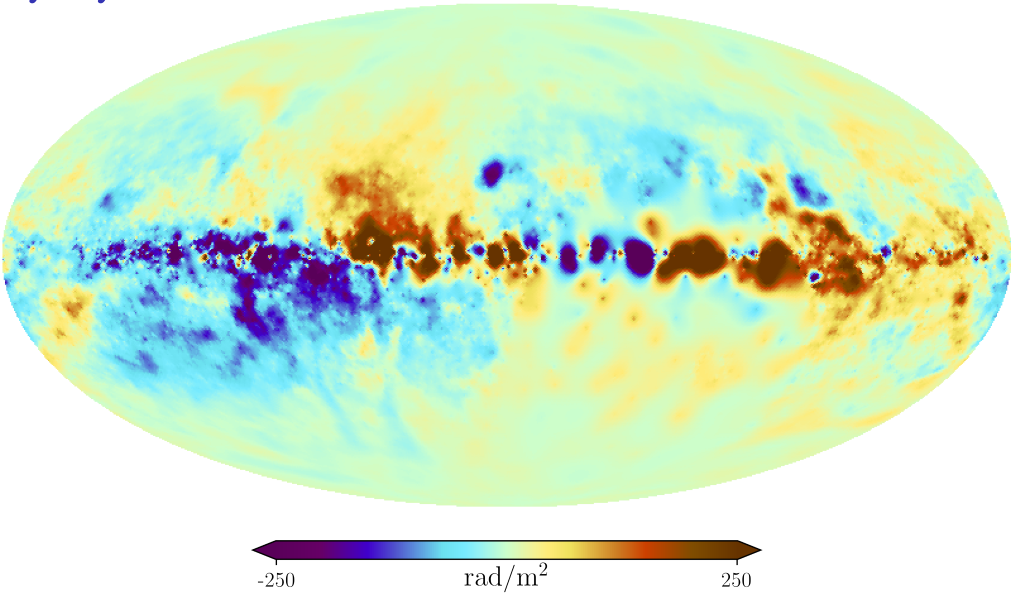
Wiener Filter



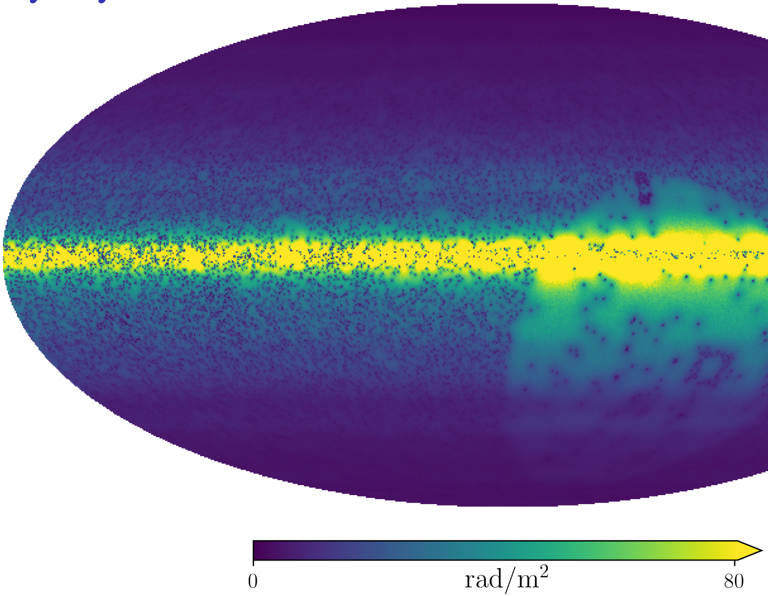
Faraday Sky



Faraday Sky



Faraday Sky



NIFTy

6.0

Search docs

- IFT – Information Field Theory
- Discretisation and Volume in NIFTy
- Gallery
- Installation
- Code Overview
- NIFTy-related publications
- Package Documentation

NIFTy – Numerical Information Field Theory

NIFTy [1], [2], "**N**umerical **I**nformation **F**ield **T**heory", is a versatile library designed to enable the development of signal inference algorithms that are independent of the underlying grids (spatial, spectral, temporal, …) and their resolutions. Its object-oriented framework is written in Python, although it accesses libraries written in C++ and C for efficiency.

NIFTy offers a toolkit that abstracts discretized representations of continuous spaces, fields in these spaces, and operators acting on these fields into classes. This allows for an abstract formulation and programming of inference algorithms, including those derived within information field theory. NIFTy's interface is designed to resemble IFT formulae in the sense that the user implements algorithms in NIFTy independent of the topology of the underlying spaces and the discretization scheme. Thus, the user can develop algorithms on subsets of problems and on space where the detailed performance of the algorithm can be properly evaluated and then easily generalize them to other, more complex spaces and the full problem, respectively.

The set of spaces on which NIFTy operates comprises point sets, *n*-dimensional regular grids, spherical spaces, their harmonic counterparts, and product spaces constructed as combinations of those. NIFTy takes care of numerical subtleties like the normalization of operations on fields and the numerical representation of model components, allowing the user to focus on formulating the

http://ift.pages.mpcdf.de/nifty/

7. Linear Filter Theory

7.1 Optimal Linear Filter: reconstruct linear signal estimate via $m = F_\mathrm{L}\, d$

Background information I:

- s: unknown signal
- d: measured data
- $\langle s s^\dagger \rangle_{(d,s)}$, $\langle d s^\dagger \rangle_{(d,s)}$, $\langle d d^\dagger \rangle_{(d,s)}$: assumed to be known covariances

Expected root mean square (RMS) error E:

$$
\begin{aligned}
E^2 &= \langle (s-m)^\dagger (s-m) \rangle_{(d,s)} \\
&= \sum_i \langle |s_i - m_i|^2 \rangle_{(d,s)} \\
&= \langle s^\dagger s \rangle_{(d,s)} - \langle s^\dagger m \rangle_{(d,s)} - \langle m^\dagger s \rangle_{(d,s)} + \langle m^\dagger m \rangle_{(d,s)}
\end{aligned}
$$

RMS Error

$$
\begin{aligned}
E^2 &= \langle s^\dagger s\rangle_{(d,\,s)} - \langle s^\dagger m\rangle_{(d,\,s)} - \langle m^\dagger s\rangle_{(d,\,s)} + \langle m^\dagger m\rangle_{(d,\,s)} \\
m &= F_{\mathrm{L}} d
\end{aligned}
$$

$$
\begin{aligned}
\langle s^\dagger s\rangle_{(d,\,s)} &= \mathrm{Tr}\langle s^\dagger s\rangle_{(d,\,s)} = \mathrm{Tr}\langle s s^\dagger\rangle_{(d,\,s)} = \mathrm{Tr}\ S \\
\langle s^\dagger m\rangle_{(d,\,s)} &= \mathrm{Tr}\langle m s^\dagger\rangle_{(d,\,s)} = \mathrm{Tr}\left(F_{\mathrm{L}}\langle d s^\dagger\rangle_{(d,\,s)}\right) \\
\langle m^\dagger s\rangle_{(d,\,s)} &= \mathrm{Tr}\langle s m^\dagger\rangle_{(d,\,s)} = \mathrm{Tr}\left(\langle s d^\dagger\rangle_{(d,\,s)} F_{\mathrm{L}}^\dagger\right) \\
\langle m^\dagger m\rangle_{(d,\,s)} &= \mathrm{Tr}\langle m m^\dagger\rangle_{(d,\,s)} = \mathrm{Tr}\left(F_{\mathrm{L}}\langle d d^\dagger\rangle_{(d,s)} F_{\mathrm{L}}^\dagger\right)
\end{aligned}
$$

$$
E^2 = \mathrm{Tr}\left[\langle s s^\dagger\rangle - F_{\mathrm{L}}\langle d s^\dagger\rangle - \langle s d^\dagger\rangle F_{\mathrm{L}}^\dagger + F_{\mathrm{L}}\langle d d^\dagger\rangle F_{\mathrm{L}}^\dagger\right]
$$

Minimization of RMS Error

$$E^2 = \mathrm{Tr}\left[\langle ss^\dagger\rangle - F_\mathrm{L}\langle ds^\dagger\rangle - \langle sd^\dagger\rangle F_\mathrm{L}^\dagger + F_\mathrm{L}\langle dd^\dagger\rangle F_\mathrm{L}^\dagger\right]$$

$$\frac{\partial E^2}{\partial F_\mathrm{L}^\dagger} = \left(0 - 0 - \langle sd^\dagger\rangle_{(d,s)} + F_\mathrm{L}\langle dd^\dagger\rangle_{(d,\,s)}\right)^\dagger \stackrel{!}{=} 0$$

$$F_\mathrm{L}\langle dd^\dagger\rangle_{(d,\,s)} \stackrel{!}{=} \langle sd^\dagger\rangle_{(d,s)}$$

$$F_\mathrm{L} = \underbrace{\langle sd^\dagger\rangle_{(d,s)}}_{\text{cross-correlation}} \underbrace{\langle dd^\dagger\rangle^{-1}_{(d,s)}}_{\text{auto-correlation}^{-1}} = \text{optimal linear filter}$$

Claim: Wiener filter is optimal linear filter, $F_\mathrm{L} = F_\mathrm{W}$

Wiener filter: $F_\mathrm{W} = \left(S^{-1} + R^\dagger N^{-1} R\right)^{-1} R^\dagger N^{-1}$

- linear measurement: $d = R\,s + n$
- Gaussian signal and Gaussian noise: $P(s,\,n) = \mathcal{G}(s,\,S)\,\mathcal{G}(n,\,N)$

Proof A: Wiener Filter is Optimal Linear Filter

Covariances:

$$
\begin{aligned}
\langle ss^\dagger \rangle_{(d,\,s)} &= \langle ss^\dagger \rangle_{(n,s)} =: S \\
\langle ds^\dagger \rangle_{(d,\,s)} &= \langle (Rs+n)s^\dagger \rangle_{(n,s)} = R \underbrace{\langle ss^\dagger \rangle_{(n,s)}}_{=S} + \underbrace{\langle ns^\dagger \rangle_{(n,s)}}_{=0} =: R\,S \\
\langle sd^\dagger \rangle_{(d,\,s)} &= S\,R^\dagger \\
\langle dd^\dagger \rangle_{(d,\,s)} &= \langle (Rs+n)(Rs+n)^\dagger \rangle_{(s)} \\
&= R\langle ss^\dagger \rangle_{(n,s)} R^\dagger + R\underbrace{\langle sn^\dagger \rangle_{(n,s)}}_{=0} + \underbrace{\langle ns^\dagger \rangle_{(n,s)}}_{=0} R^\dagger + \langle nn^\dagger \rangle_{(n,s)} \\
&=: RSR^\dagger + N
\end{aligned}
$$

Proof A: Wiener Filter is Optimal Linear Filter

Optimal linear filter: $F_{\mathrm{L}} = \langle sd^\dagger \rangle_{(d,s)} \langle dd^\dagger \rangle_{(d,s)}^{-1} = SR^\dagger (RSR^\dagger + N)^{-1}$

Wiener filter: $F_W = (S^{-1} + R^\dagger N^{-1} R)^{-1} R^\dagger N^{-1}$

$$
\begin{aligned}
F_{\mathrm{L}} &\stackrel{?}{=} F_{\mathrm{W}} \\
SR^\dagger (RSR^\dagger + N)^{-1} &\stackrel{?}{=} (S^{-1} + R^\dagger N^{-1} R)^{-1} R^\dagger N^{-1} && \mid \cdot (RSR^\dagger + N) \text{ right} \\
SR^\dagger &\stackrel{?}{=} (S^{-1} + R^\dagger N^{-1} R)^{-1} R^\dagger N^{-1} (RSR^\dagger + N) && \mid \cdot (S^{-1} + R^\dagger N^{-1} R) \text{left} \\
(S^{-1} + R^\dagger N^{-1} R) SR^\dagger &\stackrel{?}{=} R^\dagger N^{-1} (RSR^\dagger + N) \\
R^\dagger + R^\dagger N^{-1} RSR^\dagger &\stackrel{?}{=} R^\dagger N^{-1} RSR^\dagger + R^\dagger \\
R^\dagger + R^\dagger N^{-1} RSR^\dagger &= R^\dagger N^{-1} RSR^\dagger + R^\dagger \quad \square
\end{aligned}
$$

Proof B: Wiener Filter is Optimal Linear Filter

- $\mathbb{D}$: data space
- $\mathbb{S}$: signal space
- $\mathbb{X}$: some space
- linear operators $A : \mathbb{D} \to \mathbb{S}$, $B : \mathbb{S} \to \mathbb{D}$, and $X, Y : \mathbb{X} \to \mathbb{X}$

$$
\begin{aligned}
A(BA + \mathbb{1}_{\mathbb{D}}) &= (AB + \mathbb{1}_{\mathbb{S}})A \\
\Rightarrow (AB + \mathbb{1}_{\mathbb{S}})^{-1}A &= A(BA + \mathbb{1}_{\mathbb{D}})^{-1} \\
\mathbb{1}_{\mathbb{X}} + X^{-1}\,Y &= X^{-1}\,(X + Y) \\
\Rightarrow (X + Y)^{-1} &= (\mathbb{1}_{\mathbb{X}} + X^{-1}\,Y)^{-1}X^{-1}
\end{aligned}
$$

Proof B: Wiener Filter is Optimal Linear Filter

$$
\begin{aligned}
(AB + \mathbb{1}_{\mathbb{S}})^{-1} A &= A(BA + \mathbb{1}_{\mathbb{D}})^{-1} \\
(X + Y)^{-1} &= (\mathbb{1}_{\mathbb{X}} + X^{-1} Y)^{-1} X^{-1}
\end{aligned}
$$

- Optimal linear filter: $F_{\mathrm{L}} = \langle s d^\dagger \rangle_{(d,s)} \langle d d^\dagger \rangle_{(d,s)}^{-1} = S R^\dagger (R S R^\dagger + N)^{-1}$
- Wiener filter: $F_W = (S^{-1} + R^\dagger N^{-1} R)^{-1} R^\dagger N^{-1}$

With $A = R^\dagger$, $X = N$ it follows:

$$
\begin{aligned}
F_L &= S R^\dagger (R S R^\dagger + N)^{-1} \\
&= S R^\dagger (N^{-1} R S R^\dagger + \mathbb{1}_{\mathbb{D}})^{-1} N^{-1} \\
&= S (R^\dagger N^{-1} R S + \mathbb{1}_{\mathbb{S}}) R^\dagger N^{-1} \\
&= (R^\dagger N^{-1} R + S^{-1})^{-1} R^\dagger N^{-1} \\
&= F_W
\end{aligned}
$$

Summary: Wiener Filter as Optimal Linear Filter

- linear measurement: $d = Rs + n$
- Gaussian signal and Gaussian noise: $P(s,\, n) = \mathcal{G}(s,\, S)\, \mathcal{G}(n,\, N)$

$$\Rightarrow F_{\mathrm{L}} = F_{\mathrm{W}}$$

Wiener filter in signal space:

$$F_{\mathrm{W}} = \underbrace{(S^{-1} + R^{\dagger} N^{-1} R)^{-1}}_{=D} R^{\dagger} N^{-1}$$

Wiener filter in data space:

$$F_{\mathrm{L}} = \underbrace{SR^{\dagger}}_{=\langle s d^{\dagger} \rangle_{(d,\, s)}} \underbrace{(RSR^{\dagger} + N)^{-1}}_{=\langle d d^{\dagger} \rangle^{-1}_{(d,s)}}$$

Linear Response and Noise

- signal covariance:

$$\langle s s^\dagger \rangle_{(d,s)} =: S$$

- signal response: $R\,S = \langle d s^\dagger \rangle_{(d,s)}$

$$R = \langle d s^\dagger \rangle_{(d,s)} S^{-1} = \langle d s^\dagger \rangle_{(d,s)} \langle s s^\dagger \rangle_{(d,s)}^{-1}$$

- noise covariance: $R\,S\,R^\dagger + N = \langle d d^\dagger \rangle_{(d,s)}$

$$\begin{aligned} N &= \langle d d^\dagger \rangle_{(d,s)} - R\,S\,R^\dagger \\ &= \langle d d^\dagger \rangle_{(d,s)} - \langle d s^\dagger \rangle_{(d,s)} \langle s s^\dagger \rangle_{(d,s)}^{-1} \langle s s^\dagger \rangle_{(d,s)} \langle s s^\dagger \rangle_{(d,s)}^{-1} \langle s d^\dagger \rangle_{(d,s)} \\ &= \langle d d^\dagger \rangle_{(d,s)} - \langle d s^\dagger \rangle_{(d,s)} \langle s s^\dagger \rangle_{(d,s)}^{-1} \langle s d^\dagger \rangle_{(d,s)} \end{aligned}$$

- linear noise:

$$n = d - R\,s$$

End