Wiener Deconvolution: Theoretical Basis

The *Wiener Deconvolution* is a technique used to obtain the *phase-velocity* dispersion curve and the *attenuation* coefficients, by a *two-stations method*, from two pre-processed traces (instrument corrected) registered in two different stations, located at the same great circle that the epicenter (Hwang and Mitchell, 1986). For it, the earth structure crossed by the waves between the two stations selected, can be considered as a filter that acts in the input signal (registered in the station more near to the epicenter) transforming this signal in the output signal (registered in the station more far to the epicenter). Thus, the problem can be written in convolution form as:

$$g(t) = \int_{-\infty}^{\infty} f(\tau)h(t-\tau)d\tau$$

where f(t) is the trace recorded at the station more near to the epicenter, g(t) is the trace recorded more far and h(t) is the time-domain response of the *inter-station media*.

Wiener Deconvolution: Theoretical Basis

In frequency domain, the above-mentioned convolution can be written as:

$$g(t) = f(t) * h(t) \quad \leftrightarrow \quad G(\omega) = F(\omega) \cdot H(\omega)$$

$$|G(\omega)| = |F(\omega)| \cdot |H(\omega)| \qquad \Phi_G(\omega) = \Phi_F(\omega) + \Phi_H(\omega)$$

amplitude spectrum phase spectrum

F(ω) = input-signal spectrum ,, H(ω) = media-response spectrum
G(w) = output-signal spectrum (Green Function)

Then, the *frequency-domain response* of the inter-station media H(ω) can be written as:

$$|H(\omega)| = \frac{|G(\omega)|}{|F(\omega)|} \qquad \Phi_H(\omega) = \Phi_G(\omega) - \Phi_F(\omega)$$

Wiener Deconvolution: Theoretical Basis

A problem arises from the *spectral holes* present in the input-signal spectrum F(ω). These points are small or zero amplitudes that can exit for some values of the frequency ω. At these points the *spectral ratio*, preformed to compute |H(ω)|, can be infinite or a big-meaningless number for these frequencies. A method to avoid this problem is the computation of the *Green Function* by means of a spectral ratio of the *cross-correlation* divided by the *auto-correlation*:

$$y(t) = g(t) \otimes f(t) \leftrightarrow Y(\omega) = G(\omega) \cdot F^*(\omega)$$
$$x(t) = f(t) \otimes f(t) \leftrightarrow X(\omega) = F(\omega) \cdot F^*(\omega)$$
$$\Longrightarrow \quad H(\omega) = \frac{G(\omega)F^*(\omega)}{F(\omega)F^*(\omega)} = \frac{Y(\omega)}{X(\omega)}$$

and the *amplitude* and *phase* spectra of H(ω) are computed by:

$$|H(\omega)| = \frac{|Y(\omega)|}{|X(\omega)|} = \frac{|G(\omega)||F(\omega)|}{|F(\omega)||F(\omega)|}$$

$$\Phi_H(\omega) = \Phi_Y(\omega) - \Phi_X(\omega) = \Phi_G(\omega) - \Phi_F(\omega)$$

$$\Phi_Y(\omega) = \Phi_G(\omega) - \Phi_F(\omega)$$
$$\Phi_X(\omega) = \Phi_F(\omega) - \Phi_F(\omega) = 0$$

Wiener Deconvolution: Theoretical Basis

The above-mentioned process is called *Wiener deconvolution*. The goal of this process is the determination of the *Green function* in frequency domain. This function H(ω) can be used to compute the inter-station *phase velocity* c and the *attenuation* coefficients γ, by means of:

$$c(f) = \frac{f\Delta}{ft_0 + (\Phi_H(f) \pm N)}$$

$$\gamma(f) = -\frac{\ln\left(|H(f)| \sqrt{\frac{\sin\Delta_2}{\sin\Delta_1}} \right)}{\Delta}$$

where f is the *frequency* in Hz (ω = 2πf), N is an integer number (Bath, 1974), t_0 is the *origin time* of the Green function in time domain, Δ is the *inter-station distance*, Δ_1 is the *epicentral distance* for the station *near* to the epicenter and Δ_2 is the *epicentral distance* for the station far to the epicenter. These distances are measured over the great circle defined by the stations and the epicenter.

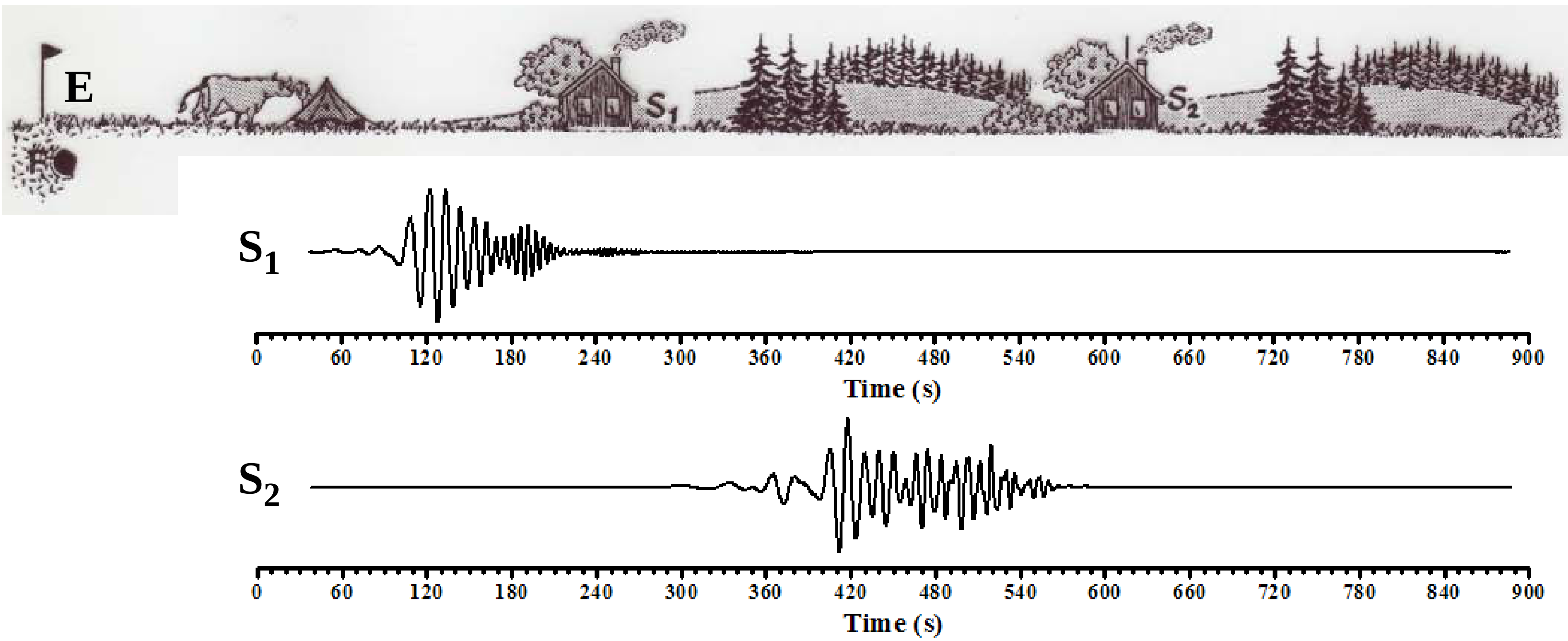
Wiener Deconvolution: Theoretical Basis

The inter-station *group velocity* U_g and the *quality factor* Q, can be obtained from c and γ by means of the well-known relationships (Ben-Menahem and Singh, 1981):

$$\frac{1}{U_g} = \frac{1}{c} + \frac{T}{c^2}\frac{dc}{dT} \qquad Q = \frac{\pi}{\gamma U_g T}$$

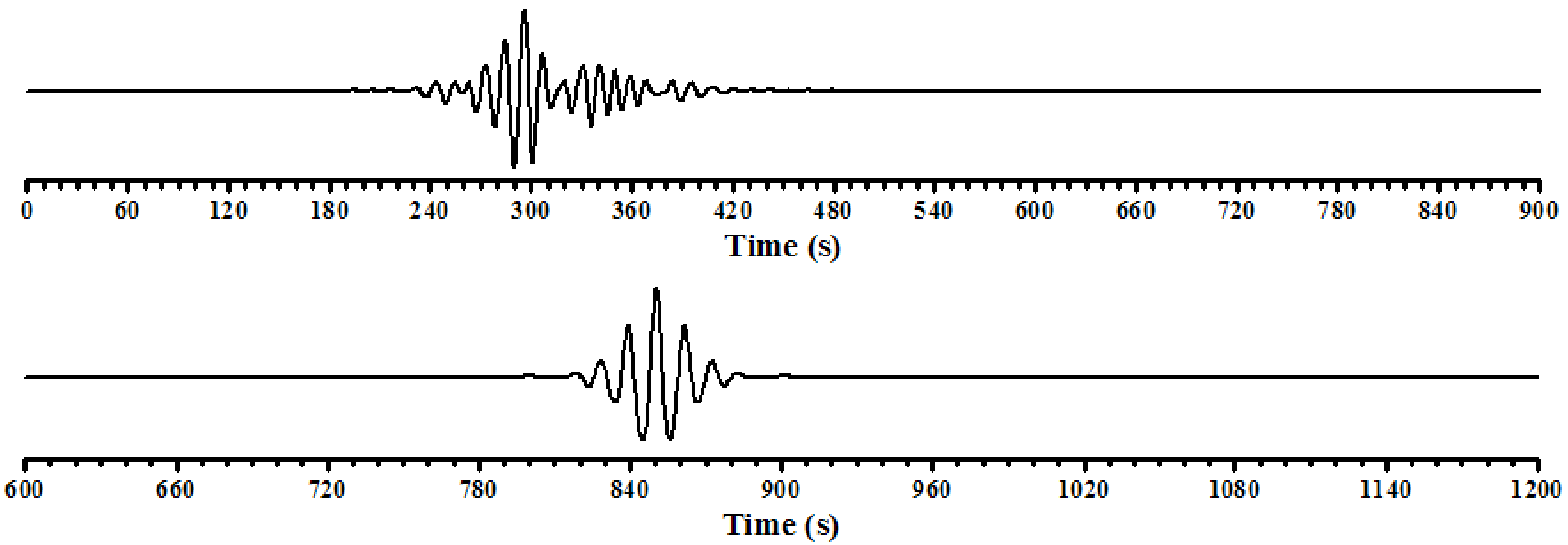
where T is the *period* (T = 1/f). The inter-station *group velocity* also can be computed applying the MFT to the Green function expressed on time domain (Hwang and Mitchell, 1986).

Wiener Deconvolution: An Example



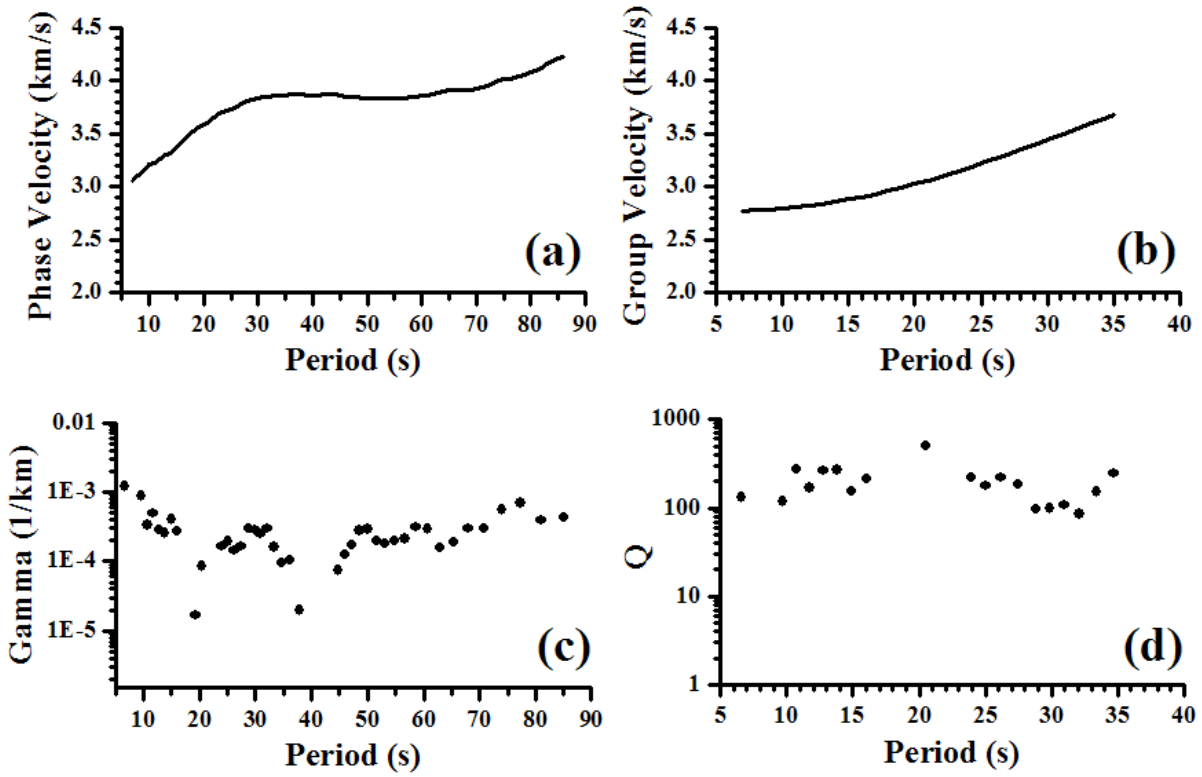
The Wiener filtering, as an example, has been applied to the traces shown above, which has been instrumentally corrected. These traces have been recorded at the stations S_1 more near to the epicenter E and S_2 more far.

Wiener Deconvolution: An Example



The cross-correlation of the traces S_1 and S_2 and the auto-correlation of the trace S_1 are performed. The auto-correlation can be *windowed* to remove noise and other undesirable perturbations, which can produce *spectral holes* in the auto-correlation spectrum and the Green function on frequency domain.

Wiener Deconvolution: An Example



The phase velocity (a), attenuation coefficients (c), group velocity (b) and quality factor (d); can be calculated through the Green function in frequency domain, as it has been explained in the previous slides.

Wiener Deconvolution: References

Bath M. (1974). *Spectral Analysis in Geophysics*. Elsevier, Amsterdam.

Ben-Menahem A. and Singh S. J. (1981). *Seismic Waves and Sources*. Springer-Verlag, Berlin.

Hwang H. J. and Mitchell B. J. (1986). *Interstation surface wave analysis by frequency-domain Wiener deconvolution and modal isolation*. Bull. of the Seism. Soc. of America, 76, 847-864.

Wiener Deconvolution: Web Page

http://airy.ual.es/www/Wiener.htm

Approximation Theory

Given a Hilbert space H with an inner product (,) we define

$$d(\varphi_1, \varphi_2) = ||\varphi_1 - \varphi_2|| = \sqrt{(\varphi_1 - \varphi_2, \varphi_1 - \varphi_2)}$$

Given a subspace $S \subset H$ with a basis $(\psi_1, \psi_2, , ..., \psi_n)$
Any $\widehat{\varphi} \in S$ can be written as $\widehat{\varphi} = \sum_{i=1}^{n} \alpha_i \psi_i$.

Problem: Given $\varphi \in H$ find $\widehat{\varphi} \in S$ so that $||\varphi - \widehat{\varphi}||$ is a minimum

Solution: Choose φ so that

$$(\varphi - \widehat{\varphi}, v) = 0 \quad \text{for every } v \in S$$

Wiener Filter

Notation:
$f(x,y)$ = "original" picture
$g(x,y)$= "actual" picture
$\widehat{f}(x,y)$ = estimate to $f(x,y)$

Assumption: the noise is additive. So

$$(*) \quad g(x,y) = \iint h(x,x';y,y')f(x',y')dx'dy' + n(x,y)$$

We **assume** $h(x,x';y,y')$ is known (?)
Assumption:

1. $E\left[n(x,y)\right] = 0$
2. $\left.\begin{array}{l} E\left[f(x,y)f(x',y')\right] = R_{ff}(x,y,x',y') \\ E\left[n(x,y)n(x',y')\right] = R_{nn}(x,y,x',y') \end{array}\right\}$ known

Problem:
Given $g(x,y)$ find $\widehat{f}(x,y)$ so that

$$\varepsilon^2 = E\left[\left(f(x,y) - \widehat{f}(x,y)\right)^2\right] \text{ is a minimum}$$

Solution:

To get a meaningful solution we need to assume that $\widehat{f}(x,y)$ is a linear function of $g(x,y)$.

$$\widehat{f}(x,y) = \iint m(x,x';y,y')g(x',y')dx'dy'$$

We need to calculate $m(x,x';y,y')$ **as a function of** g **and** h
We wish that

$$E\left[\left(\widehat{f}(x,y) - \iint m(x,x';y,y')g(x',y')dx'dy'\right)^2\right] \text{ be a minimum}$$

Using the orthogonality theorem we have

$$E\left[\left(\widehat{f}(x,y) - \iint m(x,x';y,y')g(x',y')dx'dy', g(\widetilde{x},\widetilde{y})\right)\right] = 0 \quad \text{all } \widetilde{x},\widetilde{y}$$

Interchanging the order of the integrals

$$E\left[\widehat{f}(x,y)g(\widetilde{x},\widetilde{y})\right] = \iint E\left[m(x,x';y,y')g(x',y')g(\widetilde{x},\widetilde{y})dx'dy'\right]$$

or using the cross-correlation

$$\begin{aligned} R_{fg}(x,y,\widetilde{x},\widetilde{y}) &= E\left[\widehat{f}(x,y)g(\widetilde{x},\widetilde{y})\right] \\ R_{gg}(x',y',\widetilde{x},\widetilde{y}) &= E\left[g(x',y')g(\widetilde{x},\widetilde{y})\right] \end{aligned}$$

$$R_{fg}(x,y,\widetilde{x},\widetilde{y}) = \iint_{\text{image}} R_{gg}(x',y',\widetilde{x},\widetilde{y})m(x,x';y,y')dx'dy'$$

This is still very difficult so we assume that all the statistics are homogeneous and invariant. So

$$\begin{aligned} R_{fg}(x,y,\widetilde{x},\widetilde{y}) &= R_{fg}(x-\widetilde{x},y-\widetilde{y}) \\ R_{gg}(x',y',\widetilde{x},\widetilde{y}) &= R_{gg}(x'-\widetilde{x},y-\widetilde{y}) \\ R_{nn}(x',y',\widetilde{x},\widetilde{y}) &= R_{nn}(x'-\widetilde{x},y-\widetilde{y}) \\ m(x,x';y,y') &= m(x-x'y-y') \end{aligned}$$

Then

$$\begin{aligned} R_{fg}(x-\widetilde{x},y-\widetilde{y}) &= \iint_{\text{image}} R_{gg}(x'-\widetilde{x},y'-\widetilde{y})m(x-x',y-y')dx'dy' \\ &= R_{gg}(x'-\widetilde{x},y'-\widetilde{y}) * m(x,y) \end{aligned}$$

Fourier transform yields

$$S_{fg}(u,v) = S_{gg}(u,v)H(u,v)$$

but from (*) we have

$$g(x,y) = \iint h(x,x';y,y')f(x',y')dx'dy' + n(x,y) = \iint h(x-x',y-y')f(x',y')dx'dy' + n(x,y)$$

and in Fourier space

$$G = HF + N$$

Wiener-Khinchine Theorem

Theorem 1 *The Fourier Transform of the spatial autocorrelation function is equal to the spectral density* $|\mathcal{F}(u,v)|^2$

Proof. The spatial autocorrelation function is defined by

$$R_{ff}(\widetilde{x},\widetilde{y}) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x+\widetilde{x}, y+\widetilde{y}) f(x,y) dx dy$$

Muliply both sides by the kernel of the Fourier transform and integrate. Then

$$\begin{aligned}
\widehat{R}_{ff}(\widetilde{x},\widetilde{y}) &= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} R_{ff}(\widetilde{x},\widetilde{y}) e^{-i(\widetilde{x}u+\widetilde{y}v)} d\widetilde{x} d\widetilde{y} \\
&= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x+\widetilde{x}, y+\widetilde{y}) f(x,y) e^{-i(\widetilde{x}u+\widetilde{y}v)} dx dy d\widetilde{x} d\widetilde{y}
\end{aligned}$$

Define

$$\begin{aligned}
s_1 &= x+\widetilde{x} \\
s_2 &= y+\widetilde{y}
\end{aligned}$$

Then

$$\begin{aligned}
\widehat{R}_{ff}(\widetilde{x},\widetilde{y}) &= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(s_1,s_2) f(x,y) e^{-i((s_1-x)u+(s_2-x)v)} dx dy ds_1 ds_2 \\
&= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(s_1,s_2) e^{-i(s_1u+s_2v)} ds_1 ds_2 \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x,y) e^{i(xu+yv)} dx dy \\
&= \mathcal{F}(u,v)\mathcal{F}^*(u,v) = |\mathcal{F}(u,v)|^2
\end{aligned}$$

■

Definition: A vector has a Gaussian (or normal) distribution if its joint probability density function has the form

$$p(x, \mu, C) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} e^{-\frac{1}{2}(x-\mu)^t C^{-1}(x-\mu)}$$

where C is a nxn symmetric positive definite matrix

$$E[X] = \mu$$
$$cov(X) = C$$

Definition: A vector has a Poisson distribution if its joint probability density function has the form

$$p(x, \lambda) = \prod_{i=1}^{n} \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!} \qquad x_i \text{ is a nonnegative integer}$$

where C is a nxn symmetric positive definite matrix

$$\begin{aligned} E[X] &= \lambda \\ cov(X) &= diag(\lambda_1, \lambda_2, ..., \lambda_n) \end{aligned}$$

Summary Wiener Filter

- The Wiener filter is the MSE-optimal stationary **linear** filter for images degraded by additive noise and blurring.
- Calculation of the Wiener filter requires the assumption that the signal and noise processes are second-order stationary (in the random process sense).
- Wiener filters are often applied in the frequency domain. Given a degraded image $x(n,m)$, one takes the Discrete Fourier Transform (DFT) to obtain $X(u,v)$. The original image spectrum is estimated by taking the product of $X(u,v)$ with the Wiener filter $G(u,v)$:

$$\hat{S}(u,v) = G(u,v)X(u,v)$$

The inverse DFT is then used to obtain the image estimate from its spectrum. The Wiener filter is defined in terms of these spectra:

$H(u,v)$	Fourier transform of the point-spread function (PSF)
$P_s(u,v)$	Power spectrum of the signal process, obtained by taking the Fourier transform of the signal autocorrelation
$P_n(u,v)$	Power spectrum of the noise process, obtained by taking the Fourier transform of the noise autocorrelation

The Wiener filter is:

$$G(u,v) = \frac{H^*(u,v)P_s(u,v)}{|H(u,v)|^2 P_s(u,v) + P_n(u,v)} \qquad (1.19)$$

Dividing through by P_s makes its behaviour easier to explain:

$$G(u,v) = \frac{H^*(u,v)\,P_s(u,v)}{|H(u,v)|^2 P_s(u,v) + P_n(u,v)}$$

Dividing through by P_s

$$G(u,v) = \frac{H^*(u,v)}{|H(u,v)|^2 + \frac{P_n(u,v)}{P_s(u,v)}}$$

The term P_n/P_s can be interpreted as the reciprocal of the signal-to-noise ratio. Where the signal is very strong relative to the noise, $P_n/P_s \approx 0$ and the Wiener filter becomes $H^{-1}(u,v)$ - the inverse filter for the PSF. Where the signal is very weak, $P_n/P_s \rightarrow \infty$ and $G(u,v) \rightarrow 0$.

- For the case of additive white noise and no blurring, the Wiener filter simplifies to:
- $$G(u,v) = \frac{P_s(u,v)}{P_s(u,v) + \sigma_n^2} \qquad (1.21)$$
- where σ_n^2 is the noise variance.

Wiener filters are unable to *reconstruct* frequency components which have been degraded by noise.

- They can only suppress them. Also, Wiener filters are unable to restore components for which $H(u,v)=0$. This means they are unable to undo blurring caused by bandlimiting of $H(u,v)$. Such bandlimiting which occurs in any real-world imaging system.

Steps

Obtaining P_s can be problematic.

- One can assume P_s has a parametric shape, for example exponential or Gaussian.
- Alternately, P_s can be estimated using images representative of the class of images being filtered.

- Wiener filters are comparatively slow to apply, since they require working in the frequency domain.
- To speed up filtering, one can take the inverse FFT of the Wiener filter $G(u,v)$ to obtain an impulse response $g(n,m)$.
- This impulse response can be truncated spatially to produce a convolution mask. The spatially truncated Wiener filter is inferior to the frequency domain version, but may be much faster.

Constrained Optimisation

- The algebraic framework can be used to develop a family of image restoration methods based on optimisation. Imposing appropriate constraints on the optimisation allows us to control the characteristics of the resulting estimate in order to enhance its quality. The degradation model is the same: an LSI degradation and additive noise:

$$\mathbf{f} = \mathbf{H_d}\mathbf{s} + \mathbf{n}$$

- Suppose we apply an LSI filter **H** to **f** and obtain an estimate **g:** $\mathbf{g} = \mathbf{H}\mathbf{f}$. Noting that the noise energy can be written as

$$\left|\mathbf{H_d}\mathbf{s} - \mathbf{f}\right|^2 = \left|-\mathbf{n}\right|^2 = \left|\mathbf{n}\right|^2$$

Filter Design Under Noise Constraints

- we would expect that a good estimate would satisfy the condition .

$$\left|\mathbf{H}_d\mathbf{g}-\mathbf{f}\right|^2=\left|\mathbf{n}\right|^2$$

A simple optimisation strategy would be to assume that the noise is low energy and simply choose the estimate that minimises $\left|\mathbf{H}_d\mathbf{g}-\mathbf{f}\right|^2$. Taking the derivative with respect to the estimate vector, we obtain:

$$\frac{\partial(\mathbf{H}_d\mathbf{g}-\mathbf{f})^T(\mathbf{H}_d\mathbf{g}-\mathbf{f})}{\partial\mathbf{g}}=2\mathbf{H}_d^T(\mathbf{H}_d\mathbf{g}-\mathbf{f})$$

Minimum Noise Assumption

Setting to zero and solving for **g**, we obtain what is known as the unconstrained optimisation estimate:

$$\mathbf{g} = \mathbf{H}_{\mathbf{d}}^{-1}\mathbf{f}\ .$$

This is the familiar inverse filter, the deterministic solution, appropriate for the zero noise case, but exacerbating any high frequency noise.

To avoid this result, the constrained optimisation approach introduces a new criterion, $J(\mathbf{g}) = \left|\mathbf{Q}\mathbf{g}\right|^2$ which is to be minimised subject to the noise energy constraint, $\left|\mathbf{H}_{\mathbf{d}}\mathbf{g} - \mathbf{f}\right|^{\mathbf{2}} = \left|\mathbf{n}\right|^{\mathbf{2}}$.

The criterion matrix Q **is an LSI system chosen to select for the undesirable components of the estimate.**

. If **Q** is high pass in nature, the estimate will be smoother than an unconstrained result.

Our problem can now be formalised as a classical constrained optimisation problem that can be solved using the method of Lagrange.

We want to minimise

$$J(\mathbf{g}) = \left|\mathbf{Qg}\right|^2$$

subject to the

$$\left|\mathbf{H}_d\mathbf{g} - \mathbf{f}\right|^2 = \left|\mathbf{n}\right|^2$$

- The method of Lagrange augments the criterion with a term incorporating the constraint multiplied by a constant:

$$J(\mathbf{g}) + \lambda\left[\left\|\mathbf{H}_{\mathbf{d}}\mathbf{g} - \mathbf{f}\right\|^2 - \left|\mathbf{n}\right|^2\right]$$

- Now, we take the derivative with respect to **g** and set to zero, which gives us an expression for the estimate in terms of the Lagrange multiplier constant.

$$\frac{\partial}{\partial \mathbf{g}}\left(\mathbf{g}^T\mathbf{Q}^T\mathbf{Q}\mathbf{g} + \lambda\left(\left(\mathbf{H}\mathbf{g} - \mathbf{f}\right)^T\left(\mathbf{H}\mathbf{g} - \mathbf{f}\right) - \mathbf{n}^T\mathbf{n}\right)\right) = 2\mathbf{Q}^T\mathbf{Q}\mathbf{g} + 2\lambda\mathbf{H}_\mathbf{d}^T\left(\mathbf{H}_\mathbf{d}\mathbf{g} - \mathbf{f}\right)$$

Setting to zero, we obtain:

$$\mathbf{g} = \left[\mathbf{H}_\mathbf{d}^T\mathbf{H}_\mathbf{d} + \gamma\mathbf{Q}^T\mathbf{Q}\right]^{-1}\mathbf{H}_\mathbf{d}^T\mathbf{f}$$

- In principle, we impose the constraint in order to find the multiplier $\lambda = 1/\gamma$ and thus obtain the estimate that satisfies the constraint.
- Unfortunately, there is no closed form expression for λ .
- While it is possible to employ an iterative technique, adjusting λ at each step to formally meet the constraint as closely as desired, in practice, λ is typically adjusted empirically along with the criterion matrix **Q**.

After diagonalization the frequency response of the filter is:

$$H(u) = \frac{H_d^*(u)}{\left|H_d(u)\right|^2 + \gamma \left|Q(u)\right|^2}$$

Now, consider the choice of the criterion matrix **Q**, and the Lagrange constant γ. Obviously if either or both are zero, the filter reduces to the unconstrained, inverse filter result. There are three common more interesting choices.

If we let $\mathbf{Q} = \mathbf{I}$, the criterion is just the estimate energy, $\left|\mathbf{g}\right|^2$, and the frequency response is:

$$H(u) = \frac{H_d^*(u)}{\left|H_d(u)\right|^2 + \gamma}.$$

- This simple filter is a pragmatic alternative to the Wiener filter when estimation of the signal and noise spectra are difficult.
- The constant is determined empirically to insure that at high frequencies, the degradation model has small frequency response,

$$H(u) \approx \frac{1}{\gamma} H_d^*(u)$$

rather than growing with frequency and amplifying high frequency noise like the inverse filter

A second choice is to set $\gamma |Q(u)|^2 = \frac{P_n(u)}{P_s(u)}$, which results in the Wiener deconvolution filter,

$$H(u) = \frac{H_d^*(u)}{|H_d(u)|^2 + \frac{P_n(u)}{P_s(u)}}$$

For our standard signal model, the signal spectrum decreases with frequency squared so the denominator of $H(u)$ grows as frequency squared, causing greater smoothing than the minimum energy estimate filter.

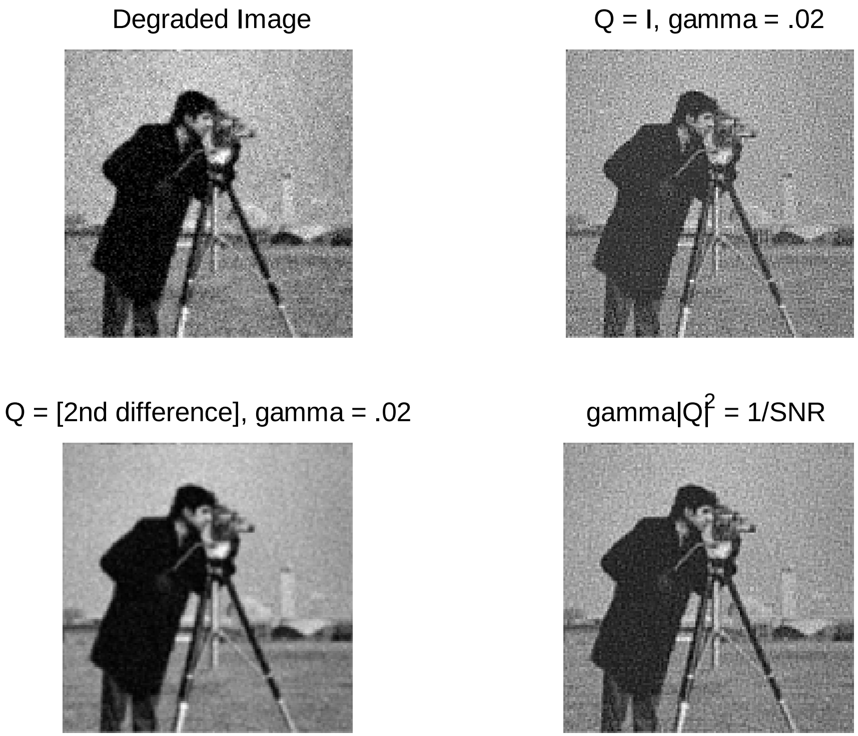
- A third choice of the criterion is the second difference operator, with impulse response

$$q(n) = \begin{bmatrix} -1 & 2 & -1 \end{bmatrix}$$

and the corresponding impulse response matrix **Q**.

- The continuous version of the second difference operator is the second derivative, with frequency domain behaviour proportional to u^2.
- The denominator $H(u)$ of now grows as and the estimate is even smoother than that of the Wiener deconvolution filter.

Figure 15.1 shows examples of constrained optimization results. The degradation model is Gaussian low pass filter blur and additive Gaussian white noise. Note that the $\mathbf{Q}=\mathbf{I}$ condition with $\gamma = 0.02$ does the least smoothing, the second difference criterion with $\gamma = 0.02$ does the most smoothing, and the Wiener deconvolution filter is intermediate between these two.



Adaptive Filters for Image Smoothing

Fundamental issues:

- Noise (typically high frequency) and signal (typically edges, also comprising high frequency components) overlap in frequency and cannot be separated by the simple frequency component weighting characteristic of LSI systems.
- Alternately, image signals are non-stationary and any global processing is likely to be sub-optimal in any local region.
- Early efforts to address this problem took an approach based on local statistics, designing operators that were optimised to the local image characteristics.

local linear minimum mean squared error (LLMMSE) filter

- introduced by Lee in 1980.
- In this approach, an optimal linear estimator for a signal in additive noise is formed as

$$\hat{s}(n) = \alpha f(n) + \beta$$

- $f(n) = s(n) + n(n)$ is the observation
- The noise and signal are assumed to be independent.

α

- Let the noise be a zero mean white Gaussian noise process with variance σ_n^2
- The parameters α and β are chosen to minimise the mean squared estimation error criterion

$$J(\alpha,\beta) = E\left[\left(\hat{s}(n) - s(n)\right)^2\right] = E\left[\left(\alpha f(n) + \beta - s(n)\right)^2\right]$$

- **Taking the derivative of the**

$$\frac{\partial J(\alpha,\beta)}{\partial \alpha} = 2E\left[\left(\alpha f(n) + \beta - s(n)\right) f(n)\right] = 0$$

$$\frac{\partial J(\alpha,\beta)}{\partial \beta} = 2E\left[\alpha f(n) + \beta - s(n)\right] = 0$$

Solving these equations for α and β, using the results $\overline{f(n)} = \overline{s(n)}$ and $\sigma_f^2 = \sigma_s^2 + \sigma_n^2$,

$$\alpha = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_n^2}$$

$$\beta = (1-\alpha)\overline{f(n)}.$$

The resulting LLMMSE estimate is then:

$$\hat{s}(n) = \frac{\sigma_s^2(n)}{\sigma_f^2(n)} f(n) + \frac{\sigma_n^2}{\sigma_f^2(n)} \overline{f(n)},$$

- **The estimate is a weighted sum of the observation and its local average, using local variance for the weighting.**
- **When the local signal variance is much greater than the constant noise variance, the estimate is just the observation—no smoothing occurs.**

- **When the local variance is entirely attributable to noise, the estimate is just the local average—maximum smoothing occurs**
- **The LLMMSE filter does little near edges or high contrast texture regions and smoothes as much as it can when the signal component is constant**

- Note that we have to estimate the noise variance somehow, as we did with the conventional Wiener filter.
- The choice of window size over which to estimate the local mean and variance is important.
- It needs to be at least 5×5 for reasonable variance estimates, but it should be small enough to insure local signal stationarity.
- Lee found that both 5x5 and 7x7 worked well.

- Results illustrating the effect of the LLMMSE filter for three different region sizes are shown in Figure 16.1. Note the residual noise within the local neighbourhood near strong edges.

Original plus noise

3x3

5x5

7x7

The LLMMSE filter can be interpreted as a local impulse response or mask whose weights are functions of the local statistics in the input. Specifically,

$$"h(n)" = \begin{cases} \alpha + (1-\alpha)/N & n = 0 \\ (1-\alpha)/N & |n| \le (N-1)/2 \\ 0 & \text{else} \end{cases}$$

where $\alpha = 1 - \sigma_n^2 \big/ \sigma_f^2(n)$ and N is the window size over which the local statistics are estimated. You can also determine a local frequency response whose bandwidth is a function of local signal-to-noise variance ratio.

Limitations of the LLMMSE filter are apparent.

- The need to restrict the window's size to achieve local stationarity restricts the amount of smoothing that can be achieved in constant signal regions.
- It may be desirable to allow the smoothing window to grow as large as the signal constant region allows.
- Some alternate noise suppression scheme to simple averaging should be employed, such as a local order-statistic filter (median filter and the like, which we will discuss further later).
- Alternately, it may be desirable to apply the LLMMSE filter iteratively, achieving repeated smoothing in constant signal regions. The window's square shape is also a limitation.

- Near an edge, the LLMMSE filter does no smoothing, allowing visible noise in close proximity to the edge.
- The window shape should also adapt, allowing the filter to smooth along but not across edges.
- The constant weights within the window, except for the centre point, limit the smoothing to the box filter, or simple local averaging variety.
- Perhaps some sort of locally variable weighting within the window would improve the performance near edges.

An extension of the local statistics filter that addresses these limitations is the local adaptive recursive structure where the output is formed with a local recursive difference equation whose coefficients depend on the local input statistics. A simple one dimensional version is:

$$g(n) = \alpha f(n) + \beta g(n-1)$$

If the parameter α is chosen the same as for the LLMMSE filter and $\beta = 1 - \alpha$, a similar edge-dependent smoothing behaviour occurs.

In constant signal regions, the next input is ignored in favour of the last output—smoothing occurs because noise is ignored

In strong signal regions the last output is ignored in favour of the next input and edges and textures are preserved. The local impulse response has an exponential form:

$$"h(n)" = \alpha \beta^{n} u(n)$$

- It is as if both the weights within the window and the window size are now dependent on the local signal-to-noise variance ratio (SNR).
- As the local SNR decreases, β increases, the weights become more uniform and the impulse response extends over a larger interval, and more smoothing occurs.

β can also be interpreted as the pole position in the Z plane for the local system transfer function ($H(z) = \sum_n h(n) z^{-n}$). The pole varies from 0, the all-pass case with no smoothing, at maximum SNR and approaches 1, the maximum smoothing case at minimum SNR.

Note that this is a causal, first order filter. If we construct a two dimensional version with different horizontal and vertical parameters, we can accomplish directionally dependent smoothing:

$$g(m,n) = \alpha f(m,n) + \beta_h g(m,n-1) + \beta_v g(m-1,n) .$$

We could even add in a diagonal term if desired. Figure 16.2 compares recursive and non-recursive versions of LLMMSE edge-dependent smoothing. In the next session we will explore a more general alternative to these simple local statistics filters.

Recursive Lee Filter

7x7 Nonrecursive Lee