Adaptive Image Filtering

(In Handbook of Medical Imaging, editor Isaac Bankman, Academic Press, 2000)

Carl-Fredrik Westin[1] Hans Knutsson[2] Ron Kikinis[1]

[1]Harvard Medical School, Brigham & Women's Hospital, Boston MA, USA

[2]Linköping University, Computer Vision Laboratory, Linköping, Sweden

westin@bwh.harvard.edu

1. Introduction

Adaptive filters are commonly used in image processing to enhance or restore data by removing noise without significantly blurring the structures in the image. The adaptive filtering literature is vast and cannot adequately be summarized in a short chapter. However, a large part of the literature concerns one-dimensional (1D) signals [1]. Such methods are not directly applicable to image processing and there are no straightforward ways to extend 1D techniques to higher dimensions primarily because there is no unique ordering of data points in dimensions higher than one. Since higher-dimensional medical image data are not uncommon (2D images, 3D volumes, 4D time-volumes), we have chosen to focus this chapter on adaptive filtering techniques that can be generalized to multidimensional signals.

This chapter assumes that the reader is familiar with the fundamentals of 1D signal processing [2]. Section 2. addresses spatial frequency and filtering. 2D spatial signals and their Fourier transforms are shown to illuminate the similarities to signals in 1D. Unsharp masking is described as an example of simple image enhancement by spatial filtering. Section 3. covers random fields and is intended as a primer for the Wiener filter, which is introduced in Section 3.2. The Wiener formulation gives a lowpass filter with a frequency characteristic adapted to the noise level in the image. The higher the noise level, the more smoothing of the data. In Section 4., adaptive Wiener formulations are presented. By introducing local adaptation of the filters, a solution more suitable for nonstationary signals such as images can be obtained. For example, by using a "visibility function," which is based on local edge strength, the filters can be made locally adaptive to structures in the image so that areas with edges are less blurred. Section 5. is about adaptive anisotropic filtering. By allowing filters to change from circularly/spherically symmetric (isotropic) to shapes that are closely related to the image structure, more noise can be suppressed without severe blurring of lines and edges. A computationally efficient way of implementing shift-variant anisotropic filters based on a non-linear combination of shift-invariant filter responses is described.

2. Multidimensional Spatial Frequencies and Filtering

At a conceptual level, there is a great deal of similarity between 1D signal processing and signal processing in higher dimensions. For example, the intuition and knowledge gained from

working with the 1D Fourier transform extends fairly straightforwardly to higher dimensions. For overviews of signal processing techniques in 2D see Lim [3], or Granlund and Knutsson for higher dimensional signal processing [4].

2.1. Spatial Frequency

The only difference between the classical notions of time frequencies and spatial frequencies is the function variable used: instead of time, the variable is spatial position in the latter case. A multidimensional sinusoidal function can be written as

$$f = \cos(\mathbf{x}^T \hat{\mathbf{e}}), \tag{1}$$

where $\mathbf{x}$ is the spatial position vector and $\hat{\mathbf{e}}$ is a normalized vector defining the orientation of the wave ($\mathbf{x}$, $\hat{\mathbf{e}} \in \mathcal{R}^n$). The signal is constant on all hyperplanes normal to $\hat{\mathbf{e}}$. For any 1D function, g, the multidimensional function

$$f = g(\mathbf{x}^T \hat{\mathbf{e}}) \tag{2}$$

will have a Fourier transform in which all energy is concentrated on a line through the origin with direction $\mathbf{e}$. For example, for a plane in 3D given by

$$f = \delta(\mathbf{x}^T \hat{\mathbf{e}}), \tag{3}$$

where δ denotes the Dirac function, the Fourier transform will be concentrated on a line with orientation normal to the plane, and the function along this line will be constant.

Examples of 2D sinusoidal functions and their Fourier transforms are shown in Figure 1. The top figures show a transform pair of a sinusoid of fairly low frequency. The Fourier transform, which contains two Dirac impulse functions, is shown to the left and may be expressed as

$$F_1 = \delta(\mathbf{u} - \boldsymbol{\omega}_1) + \delta(\mathbf{u} + \boldsymbol{\omega}_1), \tag{4}$$

where $\mathbf{u}$ denotes the 2D frequency variable, $\boldsymbol{\omega}_1$ the spatial frequency of the signal. The bottom figures show the transform pair of a signal consisting of the same signal in F_1 plus another sinusoidal signal with frequency $\boldsymbol{\omega}_2$.

$$F_2 = F_1 + \delta(\mathbf{u} - \boldsymbol{\omega}_2) + \delta(\mathbf{u} + \boldsymbol{\omega}_2), \tag{5}$$

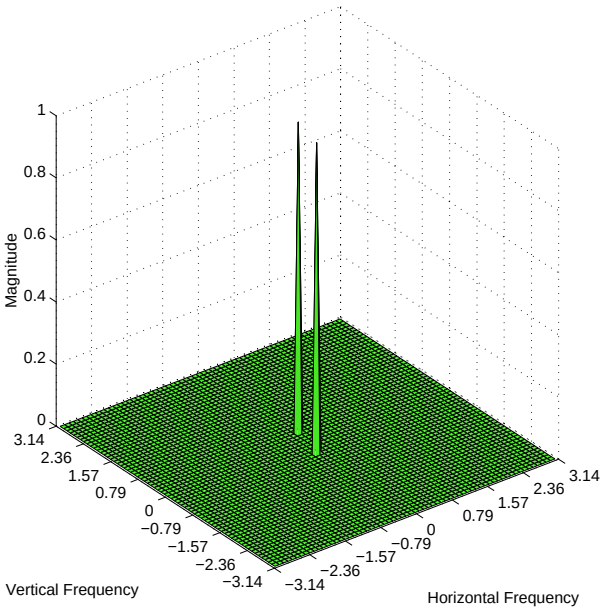
2.2. Filtering

Linear filtering of a signal can be seen as a controlled scaling of the signal components in the frequency domain. Reducing the components in the center of the frequency domain (low frequencies), gives the high-frequency components an increased relative importance, and thus highpass filtering is performed. Filters can be made very selective. Any of the Fourier coefficients can be changed independently of the others. For example, let H be a constant function minus a pair of Dirac functions symmetrically centered in the Fourier domain with a distance $|\boldsymbol{\omega}_1|$ from the center,

$$H = 1 - \delta(\mathbf{u} - \boldsymbol{\omega}_1) + \delta(\mathbf{u} + \boldsymbol{\omega}_1), \tag{6}$$

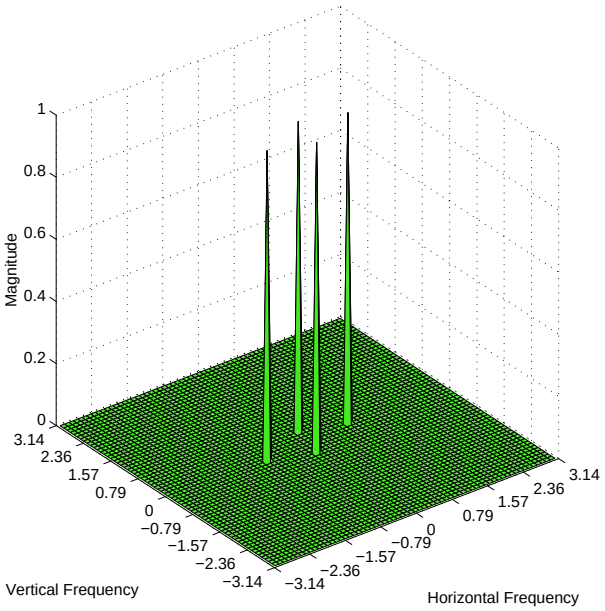
Frequency domain

Spatial domain

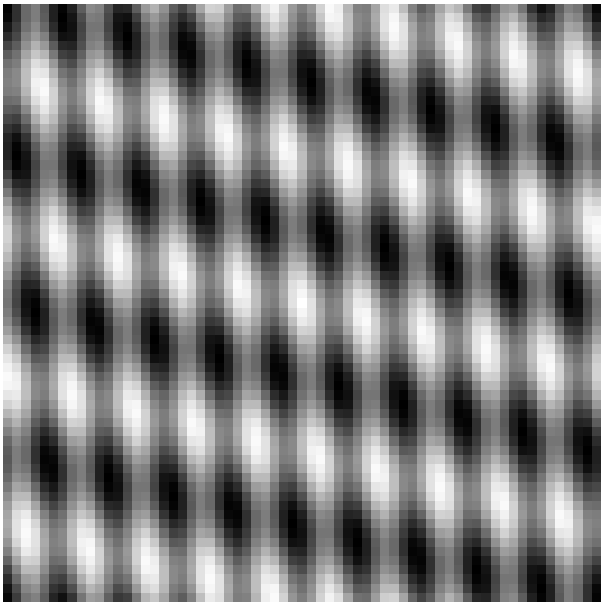


F_1

f_1



$F_1 + F_2$



$f_1 + f_2$

Figure 1: Top: Sinusoidal signal with low spatial frequency. Bottom: Sum of the top signal and a sinusoidal with a higher spatial frequency.

This filter, known as a notch filter, will leave all frequency components untouched, except the component that corresponds to the sinusoid in Fig. 1, which will be completely removed. A weighting of the Dirac functions will control how much of the component is removed. For example, the filter

$$H = 1 - 0.9\delta(\mathbf{u} - \boldsymbol{\omega}_1) + \delta(\mathbf{u} + \boldsymbol{\omega}_1), \tag{7}$$

will reduce the signal component to 10% of its original value. The result of the application of this filter to the signal $F_1 + F_2$ (Fig. 1, bottom) is shown in Fig. 2. The lower-frequency component is almost invisible.

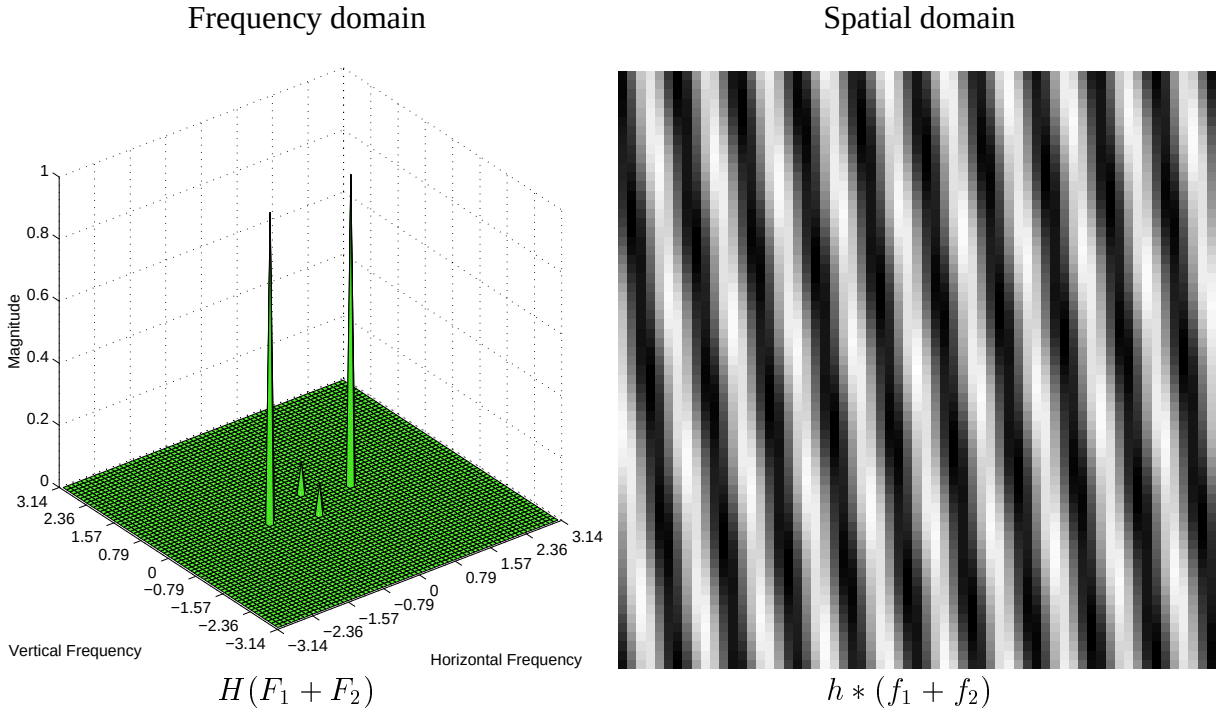


Figure 2: Notch filtering of the of the signal $f_1 + f_2$, a sum of the sinusoids. The application of the filter h in Eq. 7 reduces the low-frequency component to one-tenth of its original value.

Filters for practical applications have to be more general than "remove sinusoidal component $\cos(\boldsymbol{\omega}^T\mathbf{x})$." In image enhancement, filters are designed to remove noise that is spread out all over the frequency domain. It is a difficult task to design filters that remove as much noise as possible without removing important parts of the signal.

2.3. Unsharp Masking

Unsharp masking, an old technique known to photographers, is used to change the relative highpass content in an image by subtracting a blurred (lowpass filtered) version of the image [5]. This can be done optically by first developing an unsharp picture on a negative film and then using this film as a mask in a second development step. Mathematically, unsharp masking can be expressed as

$$\hat{f} = \alpha f - \beta f_{lp} \tag{8}$$

where α and β are positive constants, $\alpha \geq \beta$. When processing digital image data, it is desirable to keep the local mean of the image unchanged. If the coefficients in the lowpass filter f_{lp} are normalized, i.e. their sum equals one, the following formulation of unsharp masking ensures unchanged local mean in the image:

$$\hat{f} = \frac{1}{\alpha - \beta}(\alpha f - \beta f_{lp}). \tag{9}$$

By expanding the expression in the parentheses $(\alpha f - \beta f_{lp}) = \alpha f_{lp} + \alpha(f - f_{lp}) - \beta f_{lp}$, we can write Eq. (9) as

$$\hat{f} = f_{lp} + \frac{\alpha}{\alpha - \beta}(f - f_{lp}), \tag{10}$$

which provides a more intuitive way of writing unsharp masking. Further rewriting yields

$$\begin{aligned} \hat{f} &= f_{lp} + \gamma(f - f_{lp}) && (11) \\ &= f_{lp} + \gamma f_{hp}, && (12) \end{aligned}$$

where γ can be seen as a gain factor of the high frequencies. For $\gamma = 1$, the filter is the identity map, $f_{lp} + f_{hp} = f_{lp} + (f - f_{lp}) = f$, and the output image equals the input image. For $\gamma > 1$ the relative highpass content in the image is increased, resulting in higher contrast; for $\gamma < 1$ the relative highpass content is decreased and the image gets blurred. This process can be visualized by looking at the corresponding filters involved. In the Fourier domain, the lowpass image f_{lp} can be written as the product of a lowpass filter H_{lp} and the Fourier transform of the original image,

$$F_{lp} = H_{lp}F. \tag{13}$$

Figure 3 (top left) shows H_{lp} and the highpass filter that can be constructed thereof, $1 - H_{lp}$ (top right). Figure 3 (bottom left) shows the identity filter from adding the two top filters, and (bottom right) a filter where the highpass component has been amplified by a factor of 2. Figure 4 shows a slice of a CT data set through a skull (left), and the result of unsharp masking with $\gamma = 2$, i.e., a doubling of the highpass part of the image (right). The lowpass filter used (f_{lp} in Eq. (11)) was a Gaussian filter with a standard deviation of 4, implemented on a 21×21 grid.

A natural question arises: how should the parameters for the lowpass filter be chosen? It would be advantageous if the filter could adapt to the image automatically either through an a priori model of the data or by some estimation process. There are two main categories of such adaptive filters: filters for image enhancement and filters for image restoration. The two categories mainly differ in the view of the data that is to be filtered. The method of unsharp masking belongs to the first category, image enhancement. The image is made crisper by increasing the image contrast. The input image was not considered to be degraded in any way, and the purpose of the algorithm was just to improve the appearance of the image. In image restoration, as the name implies, the image data are modeled as being degraded by a (normally unknown) process, and the task at hand is to "undo" this degradation and restore the image. The models of image degradation commonly involve random noise processes. Before we introduce the well-known Wiener filter, a short description of stochastic processes in multiple dimensions is given. In multiple dimensions stochastic processes are customary referred to as random fields.

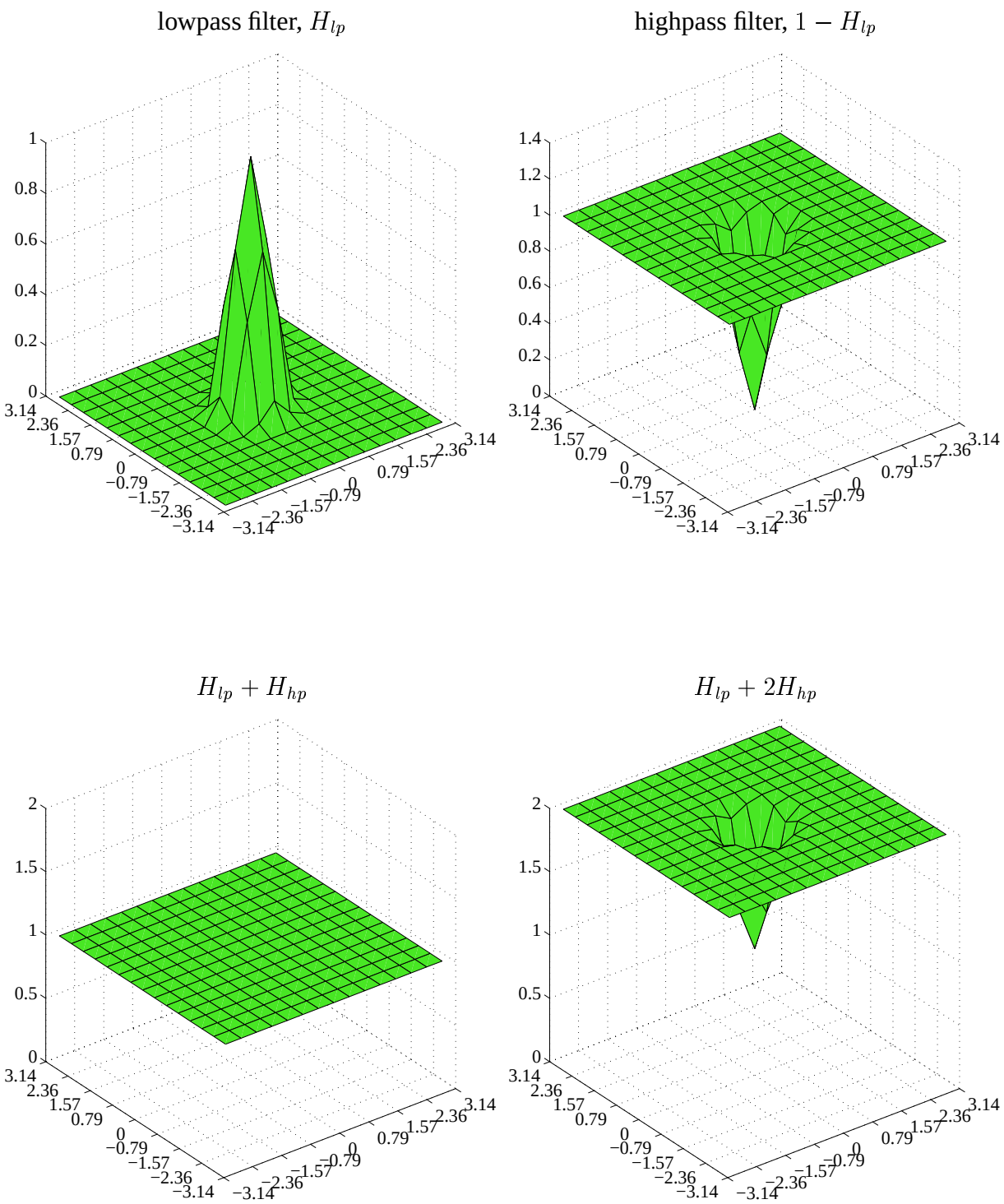


Figure 3: Visualization of the filters involved in unsharp masking.

Original CT data

Filtered CT data

Figure 4: Filtering CT data using unsharp masking. The highpass information of the original image (left) is twice as high in the result image (right). Note how details have been amplified. This technique works well due to the lack of noise in the image.

3. Random Fields and Wiener Filtering

A central problem in the application of random fields is the estimation of various statistical parameters from real data. The Wiener filter that will be discussed later requires knowledge of the spectral content of the image signal and the background noise. In practice these are, in general, not known and have to be estimated from the image.

3.1. Autocorrelation and Power Spectrum

A collection of an infinite number of random variables defined on an n-dimensional space ($\mathbf{x} \in \mathbb{R}^n$) is called a random field. The autocorrelation function of a random field $f(\mathbf{x})$ is defined as the expected value of the product of two samples of the random field,

$$R_{ff}(\mathbf{x}, \mathbf{x}') = E\{f(\mathbf{x})f(\mathbf{x}')\}, \tag{14}$$

where E denotes the statistical expectation operator. A random field is said to be stationary if the expectation value and the autocorrelation function are shift-invariant, i.e., the expectation value is independent of the spatial position vector $\mathbf{x}$, and the autocorrelation is a function only of $\boldsymbol{\tau} = \mathbf{x} - \mathbf{x}'$:

$$R_{ff}(\boldsymbol{\tau}) = E\{f(\mathbf{x} + \boldsymbol{\tau})f(\mathbf{x})\}. \tag{15}$$

The power spectrum of a stationary random process is defined by the Fourier transform of the autocorrelation function:

$$S_{ff} = \mathcal{F}(R_{ff}). \tag{16}$$

Since the autocorrelation function is always symmetric, the power spectrum is always a real function.

A random process $n(\mathbf{x})$ is called a white noise process if

$$R_{nn}(\boldsymbol{\tau}) = \sigma_n^2 \delta(\boldsymbol{\tau}). \tag{17}$$

From Eq. (16), the power spectrum of a white process is a constant:

$$S_{nn} = \sigma_n^2. \tag{18}$$

The Wiener-Khinchin theorem [6] states that

$$E\{f^2(\mathbf{x})\} = R(0) = \frac{1}{(2\pi)^n} \int S(\mathbf{u}) d\mathbf{u}, \tag{19}$$

where n is the dimension of the signal (the theorem follows directly from Eq. 16, by taking the inverse Fourier transform of $S(\mathbf{u})$ for $\boldsymbol{\tau} = 0$). This means that the integral of the power spectrum of any process is positive. It can also be shown that the power spectrum is always nonnegative [6].

The cross-correlation function of two random processes is defined as

$$R_{fg}(\mathbf{x}, \mathbf{x}') = E\{f(\mathbf{x})g(\mathbf{x}')\}. \tag{20}$$

3.2. The Wiener filter

The restoration problem is essentially one of optimal filtering with respect to some error criterion. The mean-squared error (MSE) criterion has formed the basis for most published work in this area [7, 8, 9].

In the case of linear stationary estimation we wish to estimate the process f with a linear combination of values of the data g. This can be expressed by a convolution operation

$$\hat{f} = h * g, \tag{21}$$

where h is a linear filter. A very general statement of the estimation problem is: given a set of data g, find the estimate $\hat{f}$ of an image f that minimizes some distance $\|f - \hat{f}\|$.

By using the mean-squared error, it is possible to derive the estimate with the *principle of orthogonality* [6]:

$$E\{(f - \hat{f})g\} = 0. \tag{22}$$

Inserting the convolution expression in Eq. (21) thus gives

$$E\{(f - h * g)g\} = 0. \tag{23}$$

When the filter operates in its optimal condition, the estimation error $(f - \hat{f})$ is orthogonal to the data g. In terms of the correlation functions R_{fg} and R_{gg}, Eq. (23) can be expressed as [2]

$$R_{fg} = h * R_{gg}. \tag{24}$$

By Fourier transforming both sides of the equation, we get $S_{fg} = H S_{gg}$, resulting in the familiar *Wiener filter*

$$H = \frac{S_{fg}}{S_{gg}}. \tag{25}$$

When the data are the sum of the image and stationary white noise of zero mean and variance σ_n^2,

$$g = f + n \tag{26}$$
$$R_{nn} = \sigma_n^2 \delta(\boldsymbol{\tau}), \tag{27}$$

then

$$R_{fg} = R_{ff} \tag{28}$$
$$R_{gg} = R_{ff} + R_{nn}, \tag{29}$$

and the stationary Wiener filter is given by

$$H = \frac{S_{ff}}{S_{ff} + \sigma_n^2}. \tag{30}$$

This means that for a Gaussian signal, the Wiener filter is the optimal linear estimator (among all estimators). The Wiener filter optimizes the trade-off between smoothing the signal discontinuities and removal of the noise.

4. Adaptive Wiener filters

The standard formulation of the Wiener filter has met limited success in image processing because of its lowpass characteristics, which give rise to unacceptable blurring of lines and edges. If the signal is a realization of a non-Gaussian process such as in natural images, the Wiener filter is outperformed by nonlinear estimators. One reason why the Wiener filter blurs the image significantly is that a fixed filter is used throughout the entire image: the filter is space invariant.

A number of attempts to overcome this problem have adopted a nonstationary approach where the characteristics of the signal and the noise are allowed to change spatially [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23].

4.1. Local adaptation

The Wiener filter in Eq. (30) defines a shift-invariant filter, and thus the same filter is used throughout the image. One way to make the filter spatially variant is by using a local spatially varying model of the noise parameter σ_n. Such a filter can be written as

$$H = \frac{S_{ff}}{S_{ff} + \sigma_n^2(\mathbf{x})}. \tag{31}$$

This filter formulation is computationally quite expensive since the filter changes from pixel to pixel in the image.

Lee derived an efficient implementation of a noise-adaptive Wiener filter by modeling the signal locally as a stationary process [14, 3] that results in the filter

$$\hat{f}(\mathbf{x}) = m_f(\mathbf{x}) + \frac{\sigma_f^2(\mathbf{x})}{\sigma_f^2(\mathbf{x}) + \sigma_n^2}(g(\mathbf{x}) - m_f(\mathbf{x})) \tag{32}$$

where m_f is the local mean of the signal g, and σ_f^2 is the local signal variance. A local mean operator can be implemented by a normalized lowpass filter where the support region of the filter defines the locality of the mean operator. It is interesting to note that this formulation gives an expression very similar to the expression of unsharp masking in Eq. (11), although the latter is not locally adaptive. Figure 5 shows the result of Lee's filter on MR data through the pelvis.

4.2. Nonlinear Extension by a Visibility Function

The Wiener filter that results from a minimization based on the MSE criterion can only relate to second-order statistics of the input data and no higher [1]. The use of a Wiener filter or a linear adaptive filter to extract signals of interest will therefore yield suboptimal solutions. By introducing nonlinearities in the structure, some limitations can be taken care of.

Abramatic and Silverman [12] modified the stationary white noise Wiener solution (Eq. (30)) by introducing a visibility function α, $0 \leq \alpha \leq 1$, which depends on the magnitude of the image gradient vector, where α is 0 for "large" gradients and 1 in areas of no gradient. They showed that the generalized Backus-Gilbert criterion yields the solution

$$H_\alpha = \frac{S_{ff}}{S_{ff} + \alpha\sigma^2} \tag{33}$$

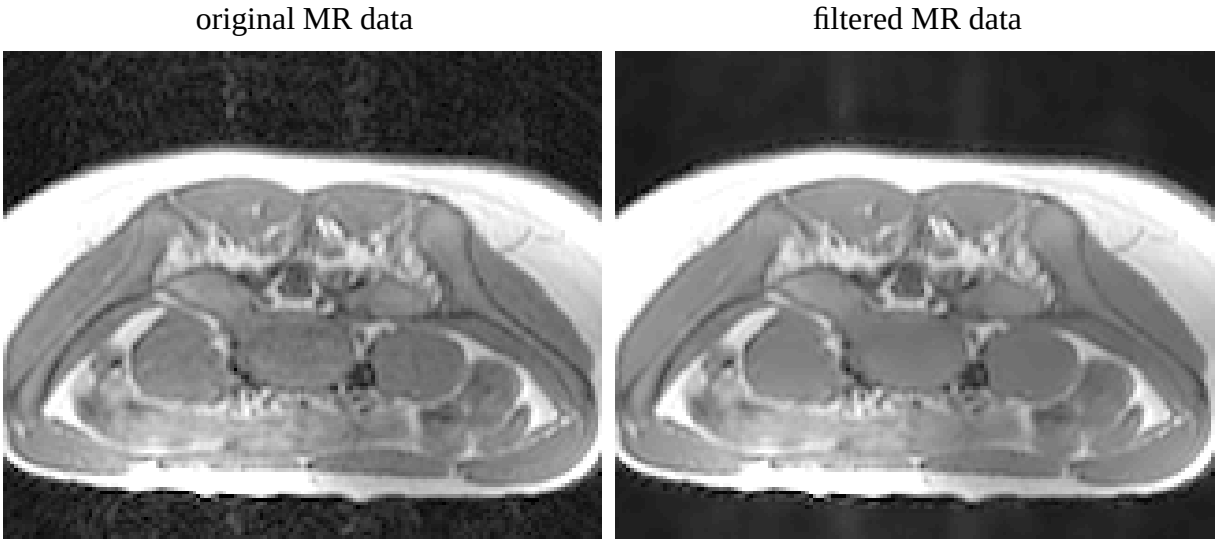


Figure 5: Filtering MR data through the pelvis using Lee's method. Note how the background noise the original image (left) has effectively been suppressed in result image (right). Note how the motion related artifacts are reduced, but blurring is introduced.

for an optimal filter. The Backus-Gilbert method [24, 25] is a regularization method that differs from others in that it seeks to maximize *stability* of the solution rather than, in the first instance, its *smoothness*. The Backus-Gilbert method seeks to make the mapping from $\hat{f}$ and f as close to the identity as possible in the limit of error-free data. Although the Backus-Gilbert philosophy is rather different from standard linear regularization methods, in practice the differences between the methods are small. A stable solution is almost inevitably smooth.

Equation (33) shows a very explicit trade-off between resolution and stability. $\alpha = 0$ gives a filter that is the identity mapping (maximum resolution) and $\alpha = 1$ gives the smoother Wiener solution. By choosing α to be spatially variant, i.e., a function of the position in the image, $\alpha = \alpha(\mathbf{x})$, a simple adaptive filter is obtained. For large gradients, the alpha function cancels the noise term and the function H becomes the identity map. This approach has the undesired feature that the filter changes from point to point in a way that is generally computationally burdensome. Abramatic and Silverman [12] proposed an "signal equivalent" approach giving a filter that is a linear combination of the stationary Wiener filter and the identity map:

$$H_\alpha = \alpha \frac{S_{ff}}{S_{ff} + \sigma^2} + (1 - \alpha). \tag{34}$$

Note that H_α equals the Wiener solution (Eq. (30)) for $\alpha = 1$, and for $\alpha = 0$ the filter becomes the identity map. It is interesting to note that Eq. (34) can be rewritten as

$$H_\alpha = \frac{S_{ff}}{S_{ff} + \sigma^2} + (1 - \alpha) \frac{\sigma^2}{S_{ff} + \sigma^2}. \tag{35}$$

Equation (35) shows that Abramatic and Silverman's model can be seen as a linear combination of a stationary lowpass component and a nonstationary highpass component [16]. Inserting H for the stationary Wiener solution we can write the filter in Eq. (35) as

$$H_\alpha = H + (1 - \alpha)(1 - H). \tag{36}$$

5. Anisotropic Adaptive Filtering

5.1. Anisotropic Adaptive Filtering in Two Dimensions

On the basis of the characteristics of the human visual system, Knutsson et al. [16] argued that local anisotropy is an important property in images and introduced an anisotropic component in Abramatic and Silverman's model (Eq. (36))

$$H_{\alpha,\gamma} = H + (1-\alpha)(\gamma + (1-\gamma)\cos^2(\varphi - \theta))(1-H), \tag{37}$$

where the parameter γ controls the level of anisotropy, φ defines the angular direction of the filter coordinates, and θ is the orientation of the local image structure. The specific choice of weighting function $\cos^2(\varphi - \theta)$ was imposed by its ideal interpolation properties; the directed anisotropy function could be implemented as a steerable filter from three fixed filters [16]

$$\cos^2(\varphi),\ \cos^2(\varphi - \pi/3), \text{ and } \cos^2(\varphi - 2\pi/3)$$

(these filters span the same space as the three filters 1, $\cos 2(\varphi)$, $\sin 2(\varphi)$, which also can be used). Freeman and Adelson later applied this concept to several problems in computer vision [26].

Knutsson et al. estimated the local orientation and the degree of anisotropy with three oriented Hilbert transform pairs, so-called quadrature filters, with the same angular profiles as the three basis functions describing the steerable weighting function. Figure 6 shows one of these Hilbert transform pairs. In areas of the image lacking a dominant orientation, γ is set to 1, and Eq. (37) reverts to the isotropic Abramatic and Silverman solution. The more dominant the local orientation, the smaller the γ value and the more anisotropic the filter.

5.2. Multidimensional Anisotropic Adaptive Filtering

The $\cos^2(\cdot)$ term in Eq. (37) can be viewed as a squared inner product of a vector defining the main directionality of the signal (θ) and the frequency coordinate vectors. This main direction is denoted $\hat{\mathbf{e}}_1$, and an orthogonal direction $\hat{\mathbf{e}}_2$, Eq. (37) can be rewritten as

$$H_\gamma = H + (1-H)[\gamma_1(\hat{\mathbf{e}}_1^T\hat{\mathbf{u}})^2 + \gamma_2(\hat{\mathbf{e}}_2^T\hat{\mathbf{u}})^2], \tag{38}$$

where $\mathbf{u}$ is the 2D frequency variable, and the parameters γ_1 and γ_2 defines the anisotropy of the filter. For $\gamma_1 = \gamma_2 = 1-\alpha$, the filter becomes isotropic ($\gamma = 1$ in Eq. (37)), and for $\gamma_1 = 1-\alpha$ and $\gamma_2 = 0$ the filter becomes maximally anisotropic ($\gamma = 0$ in Eq. (37)) and mainly favors signals oriented as $\hat{\mathbf{e}}_1$.

The inner product notation in Eq. (38) allows for a direct extension to multiple dimensions:

$$H_\gamma = H + (1-H)\sum_{k=1}^{N}\gamma_k(\hat{\mathbf{e}}_k^T\hat{\mathbf{u}})^2, \tag{39}$$

where N is the dimension of the signal (N=2 for images and N=3 for volumes). However, two important elements remain to be defined: how to define the orthogonal directions $\hat{\mathbf{e}}_k$ and the coefficients γ_k so that the equation describes a useful adaptive filter.

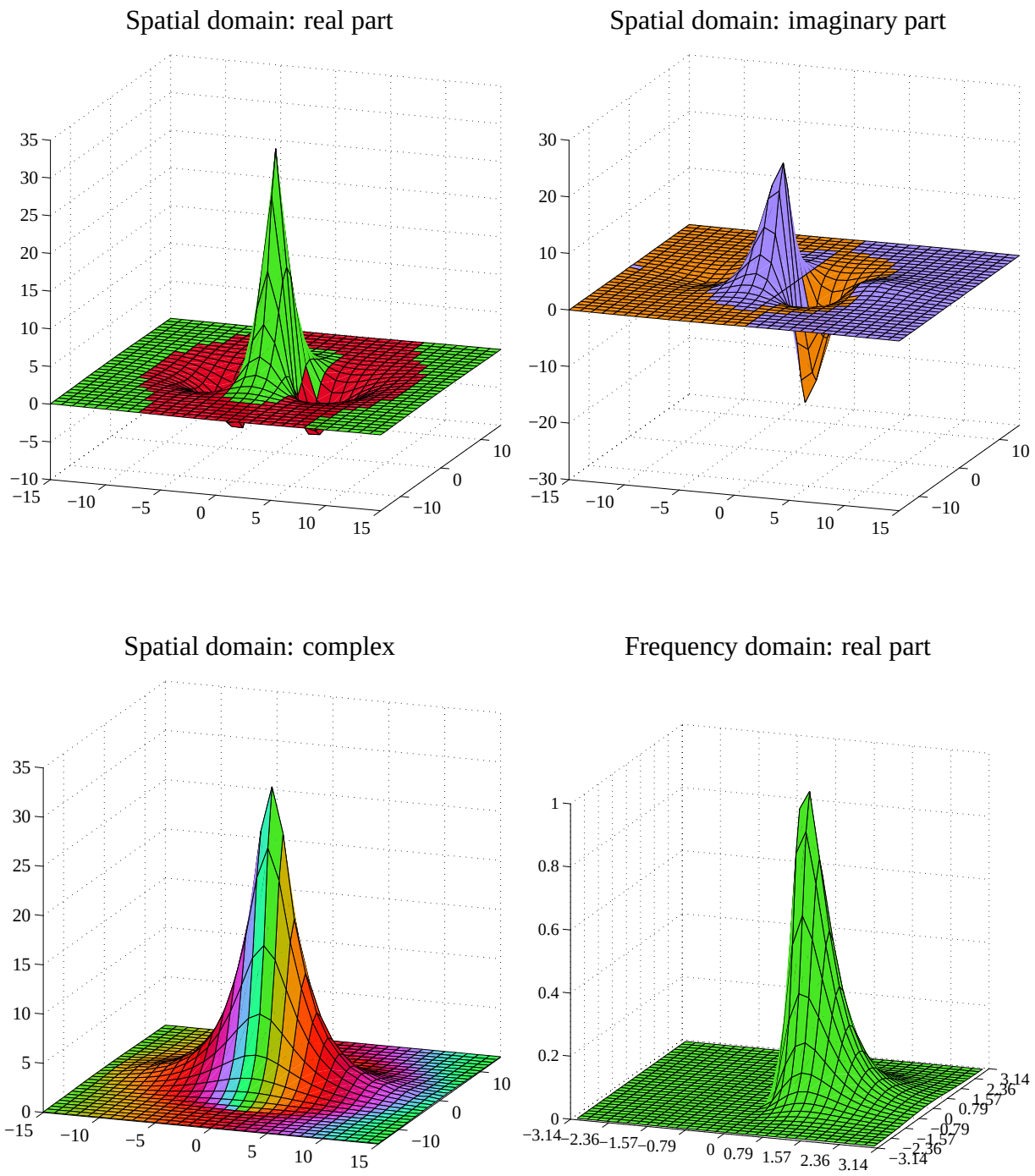


Figure 6: Visualization of a quadrature filter (Hilbert transform pair) used in the estimation of local anisotropy. Top: The plots show the filter in the spatial domain: the real part (left) and the imaginary part (right). It can be appreciated that the real part can be viewed as a line filter and the imaginary part an edge filter. The color coding is green, positive real; red, negative real; blue, positive imaginary; and orange, negative imaginary. Bottom: The left plot shows the magnitude of the filter with the phase of the filter color coded. The right plot shows the quadrature filter in the Fourier domain. Here the filter is real and zero on one half of the Fourier domain.

By expressing the squared inner product $(\hat{\mathbf{e}}_k^T \hat{\mathbf{u}})^2$ as the inner product of two outer products $< \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^T, \hat{\mathbf{u}} \hat{\mathbf{u}}^T >$

$$H_\gamma = H + (1-H) \sum_{k=1}^{N} < \gamma_k \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^T, \hat{\mathbf{u}} \hat{\mathbf{u}}^T > \quad (40)$$

$$= H + (1-H) < \mathbf{C}, \mathbf{U} >, \quad (41)$$

we can define a term $\mathbf{C}$ as a "control" tensor,

$$\mathbf{C} = \sum_{k=1}^{N} \gamma_k \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^T. \quad (42)$$

The tensor $\mathbf{C}$ controls the filter adaptation by weighting the components in the outer product description of the Fourier domain $\mathbf{U}$ according to its "shape". For a 3D signal this shape can be thought of as an ellipsoid with principal axes $\hat{\mathbf{e}}_k$. Similar to the 2D case in Eq. (37), where θ describes the main orientation of the local spectrum, the tensor $\mathbf{C}$ should describe the major axes of the local spectrum in the Fourier domain. The consequence of the inner product in Eq. (41) is a reduction of the high-frequency components in directions where the local spectrum is weak. In those directions the adaptive filter will mainly act as the stationary Wiener component $\mathbf{H}$, which has lowpass characteristics.

5.3. Adaptation Process

Before we describe the adaptation process, we will introduce a set of fixed filters

$$H_k(\mathbf{u}) = (1-H)(\hat{\mathbf{n}}_k^T \hat{\mathbf{u}})^2, \quad (43)$$

where $\hat{\mathbf{n}}_k$ define the directions of the filters and $\hat{\mathbf{u}} = \frac{\mathbf{u}}{|\mathbf{u}|}$.

It turns out that the adaptive filter in Eq. (39) can be written as linear combination of these fixed filters and the stationary Wiener solution where the weights are defined by inner products of a control tensor $\mathbf{C}$ and the filter associated dual tensors $\mathbf{M}_k$:

$$H_c = H + \sum_{k=1}^{N(N+1)/2} < \mathbf{M}_k, \mathbf{C} > H_k. \quad (44)$$

where the tensors $\mathbf{M}_k$ define the so-called dual tensor basis to the one defined by the outer product of the filter directions. The dual tensors $\mathbf{M}_k$ are defined by

$$< \mathbf{M}_k, \mathbf{N}_l >= \delta_{kl},$$

where $\mathbf{N}_k = \hat{\mathbf{n}}_k \hat{\mathbf{n}}_k^T$ are the outer products of the filter directions. The number of fixed filters needed, $N(N+1)/2$, is the number of independent coefficients in the tensor $\mathbf{C}$, which is described by a symmetric $N \times N$ matrix for a signal of dimension N. This gives that 3 filters are required in 2D, and 6 filters in 3D.

The rest of this section is somewhat technical and may be omitted on a first reading.

Equation (44) can be validated be inserting the expressions for the control tensor $\mathbf{C}$ and the fixed filter functions in Eq. (44),

$$H_c \;=\; H + (1-H) \sum_{k=1}^{N(N+1)/2} < \mathbf{M}_k, \sum_{i=1}^{N} \gamma_i \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T > (\hat{\mathbf{n}}_k^T \hat{\mathbf{u}})^2. \tag{45}$$

By expressing the squared inner product $(\hat{\mathbf{e}}_k^T \hat{\mathbf{u}})^2$ as the inner product of two outer products $< \hat{\mathbf{n}}_k \hat{\mathbf{n}}_k^T, \hat{\mathbf{u}} \hat{\mathbf{u}}^T >$, switching the order of the terms in the first inner product

$$H_c \;=\; H + (1-H) \sum_{k=1}^{N(N+1)/2} < \sum_{i=1}^{N} \gamma_i \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T, \mathbf{M}_k >< \hat{\mathbf{n}}_k \hat{\mathbf{n}}_k^T, \hat{\mathbf{u}} \hat{\mathbf{u}}^T >, \tag{46}$$

and reordering the summation

$$\begin{aligned} H_c \;&=\; H + (1-H) < \sum_{i=1}^{N} \gamma_i \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T, \underbrace{\sum_{k=1}^{N(N+1)/2} \mathbf{M}_k}_{=1} >< \mathbf{N}_k, \hat{\mathbf{u}} \hat{\mathbf{u}}^T > & (47) \\ &=\; H + (1-H) < \sum_{i=1}^{N} \gamma_i \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T, \hat{\mathbf{u}} \hat{\mathbf{u}}^T > & (48) \\ &=\; H + (1-H) \sum_{i=1}^{N} \gamma_i (\hat{\mathbf{e}}_i^T \hat{\mathbf{u}}_i)^2, & (49) \end{aligned}$$

which is equal to the adaptive filter we wanted to construct (Eq. (39)). The under-braced terms sums to 1 because of the dual basis relation between the two bases $\mathbf{M}_k$ and $\mathbf{N}_k$.

5.4. Estimation of Multidimensional Local Anisotropy Bias

Knutsson [27] has described how to combine quadrature filter responses into a description of local image structure using tensors. His use of tensors was primarily driven by the urge to find a continuous representation of local orientation. The underlying issue here is that orientation is a feature that maps back to itself modulo π under rotation, and direction is a feature that maps back to itself modulo 2π. That is why vector representations work well in the latter case (e.g., representing velocity) but not in the former case.

Knutsson used spherically separable quadrature filters [28],

$$Q(\mathbf{u}) \;=\; R(\rho) D_k(\hat{\mathbf{u}}), \tag{50}$$

where $\mathbf{u}$ is the vector valued frequency variable, $\rho = |\mathbf{u}|$, $\hat{\mathbf{u}} = \frac{\mathbf{u}}{|\mathbf{u}|}$, and $R(\rho)$ and $D_k(\hat{\mathbf{u}})$ are the radial and the directional functions, respectively,

$$\begin{cases} D_k(\hat{\mathbf{u}}) \;=\; (\hat{\mathbf{u}}^T \hat{\mathbf{n}}_k)^2 & \text{if } \hat{\mathbf{u}}^T \hat{\mathbf{n}}_k \;>\; 0 \\ D_k(\hat{\mathbf{u}}) \;=\; 0 & \text{otherwise}, \end{cases} \tag{51}$$

where $\hat{\mathbf{n}}_k$ is the filter direction, i.e., $D(\hat{\mathbf{u}})$ varies as $\cos^2(\varphi)$, where φ is the angle between $\mathbf{u}$ and the filter direction, and

$$R(\rho) \;=\; \mathrm{e}^{-\frac{4}{B^2 \ln 2} \ln^2(\rho/\rho_0)} \tag{52}$$

is the radial frequency function. Such functions are Gaussian functions on a logarithmic scale and are therefore termed lognormal functions. B is the relative bandwidth in octaves and ρ_0 is the center frequency of the filter. $R(\rho)$ defines the frequency characteristics of the quadrature filters.

A tensor $\mathbf{T}$ describing local structure is obtained by a linear summation of quadrature filter magnitude responses, $|q_k|$, weighted by predefined tensors $\mathbf{M}_k$, associated with each filter (defined as the dual tensors in Eq. 44):

$$\mathbf{T} = \sum_{k=1}^{N(N+1)/2} \mathbf{M}_k |q_k|. \tag{53}$$

where $|q_k|$ is the output magnitude from the quadrature filter k.

The tensor in Eq. (53) is real and symmetric and can thus can be written as a weighted sum of outer products of its orthogonal eigenvectors:

$$\mathbf{T} = \sum_{k=1}^{N} \lambda_k \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^T, \tag{54}$$

where the vectors $\hat{\mathbf{e}}_k$ describe the principal axes of the local signal spectrum (the locality is defined by the radial frequency functions of the quadrature filters).

The distribution of the eigenvalues, $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_N$, describes the anisotropy of the local spectrum. If the tensor is close to rank 1, i.e. there is only one large eigenvalue ($\lambda_1 >> \lambda_k, k \in \{2, ..., N\}$), the spectrum is concentrated to the line in the Fourier domain defined by $\hat{\mathbf{e}}_1$. Further, if the tensor is close to rank 2 the spectrum is concentrated to the plane in the Fourier domain spanned by the two eigenvectors having non-zero eigenvalues ($\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$).

5.5. Tensor mapping

The control tensor $\mathbf{C}$ used in the adaptive scheme is a based on a normalization of the tensor descibed in the previous section (Eq. 54):

$$\mathbf{C} = \sum_{k=1}^{N} \gamma_k \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^T \tag{55}$$

$$= \frac{\lambda_1}{\lambda_1^2 + \alpha^2} \mathbf{T}. \tag{56}$$

where α is a term defining the trade-off between resolution and stability similar to Eq. (33). However, here the resolution trade-off is adaptive in the same way as a the stationary Wiener filter. $\alpha = 0$ gives maximum resolution, and the larger the α, the more of the smoother Wiener solution. Maximum resolution is given when λ_1 is large compared to α, and the smaller the λ_1 the smoother the solution. If the "resolution parameter" $\alpha = 0$, the control tensor will be

$$\mathbf{C} = \frac{1}{\lambda_1} \mathbf{T} \tag{57}$$

$$= \hat{\mathbf{e}}_1 \hat{\mathbf{e}}_1^T + \frac{\lambda_2}{\lambda_1} \hat{\mathbf{e}}_2 \hat{\mathbf{e}}_2^T + \ldots + \frac{\lambda_N}{\lambda_1} \hat{\mathbf{e}}_N \hat{\mathbf{e}}_N^T. \tag{58}$$

With this normalization, the largest eigenvalue of $\mathbf{C}$ is $\gamma_1 = 1$, and the resulting adaptive filter H_c becomes an allpass filter along signal direction $\hat{\mathbf{e}}_1$,

$$H_c = H + (1 - H)\gamma_1 = 1. \tag{59}$$

If it is desired to increase the image contrast, this can be done by increasing the high frequency content in dim parts of the image data adaptively. Assuming a normalization of the tensors $\mathbf{T}$ so that the largest eigenvalue is 1 globally, $\max(\lambda_1) = 1$, this can be achieved by increasing the exponent of the λ_1 in the denominator.

$$\mathbf{C} = \frac{\lambda_1}{\lambda_1^{2.5} + \alpha^2}\mathbf{T}. \tag{60}$$

This will increase the relative weights for low and medium-low signal components compared to the largest ($\lambda_1 = 1$). More elaborate functions for remapping of the eigenvalues can be found in [4, 29, 30].

5.6. Examples of Anisotropic Filtering in 2D and 3D

The filters used in the examples below have in the 2D examples size 15×15, and $15 \times 15 \times 15$ in the 3D. The center frequencies of the quadrature filters differ in the examples, but the relative bandwidth is the same, B=2. We have for simplicity approximated the stationary Wiener filter with the following lowpass filter:

$$\begin{cases} H(\rho) = \cos^2(\frac{\pi\rho}{2\rho_{lp}}) & \text{if } 0 \leq \rho < \rho_{lp} \\ H(\rho) = 0 & \text{otherwise} \end{cases} \tag{61}$$

Here ρ is the radial frequency variable, and ρ_{lp} is the cutoff frequency. The more noise in the image, the lower the cutoff frequency ρ_{lp} should be used.

Figure 7 shows the result of 2D adaptive filtering of MR data from breast imaging. The original image is shown to the left. To the right, the result from anisotropic adaptive filtering is shown. The quadrature filters used for estimating the local structure had center frequency $\omega_0 = \pi/3$ and the lowpass filter H had a cutoff frequency $\rho_{lp} = \pi/4$. The control tensor $\mathbf{C}$ was defined by Eq. (56) with $\alpha = 1\%$ of the largest λ_1 globally.

Figure 8 shows the result of 3D adaptive filtering of MR data through the skull. The original image is shown to the left. The result from adaptive filtering is shown to the right. The quadrature filters used had center frequency $\omega_0 = \pi/2$ and the lowpass filter H had a cutoff frequency $\rho_{lp} = \pi/3$. The control tensor $\mathbf{C}$ was defined by Eq. (60) with $\alpha = 1\%$ of the largest λ_1 globally. Note that details with low contrast in the original image have higher contrast in the enhanced image.

Figure 9 shows the result of 3D (2D+time) adaptive filtering of ultrasound data of a beating heart. The top row shows images from the original time sequence. The bottom row shows the result after 3D filtering. The quadrature filters used had center frequency $\omega_0 = \pi/6$ and the cutoff frequency of the lowpass filter H was $\rho_{lp} = \pi/4$. The control tensor $\mathbf{C}$ is defined by Eq. (60) with $\alpha = 5\%$ of the largest λ_1 globally.

More details on implementing filters for anisotropic adaptive filtering can be found in [4] where also the estimation of local structure using quadrature filters is described in detail. An important issue that we have not discussed in this chapter is that in medical imaging we often

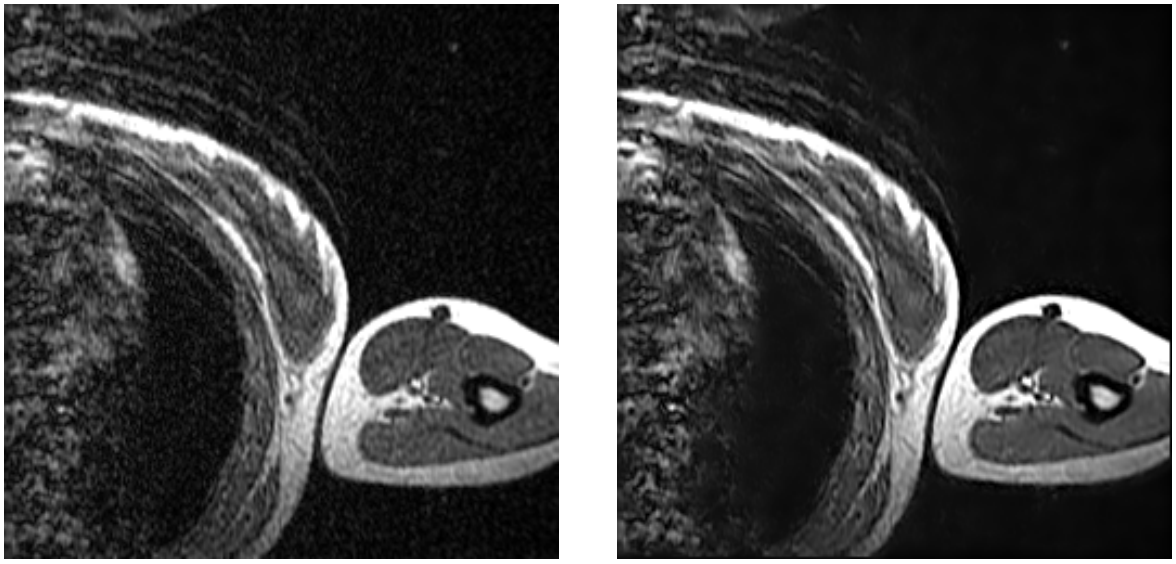


Figure 7: 2D adaptive filtering of data from MR breast imaging. Left: Original image. Right: Adaptively filtered image. The adaptive filtering reduces the unstructured component of the motion related artifacts.

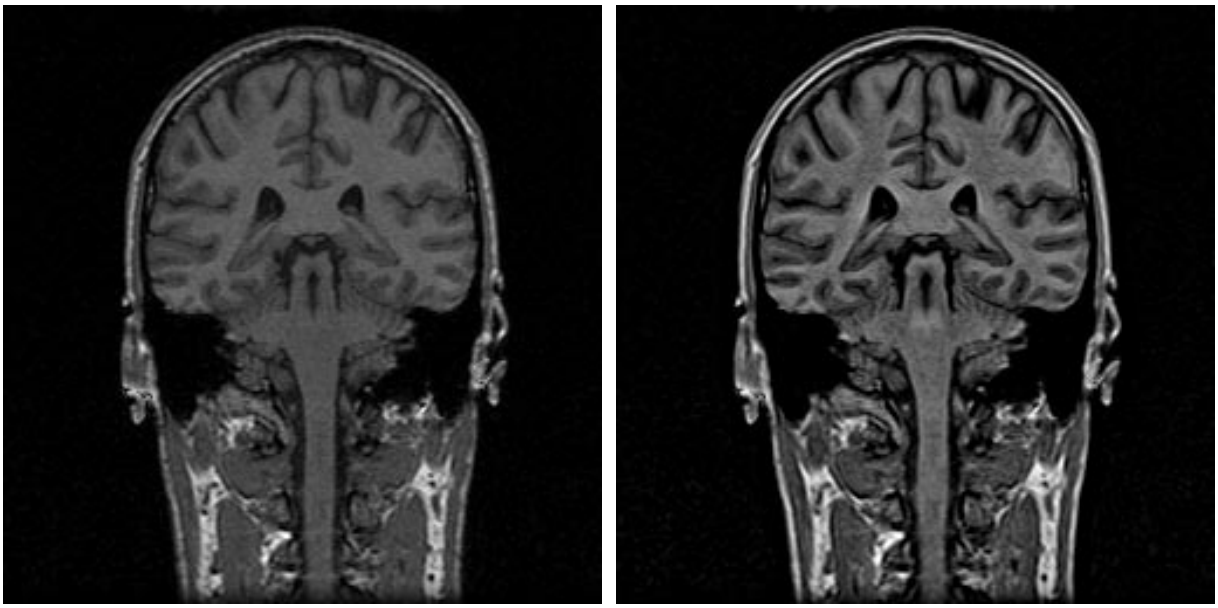


Figure 8: 3D adaptive filtering of a coronal MR data set of the head with dynamic compression of the signal. Left: Original image. Right: Adaptively filtered image. Note the improved contrast between brain and cerebrospinal fluid (CSF).

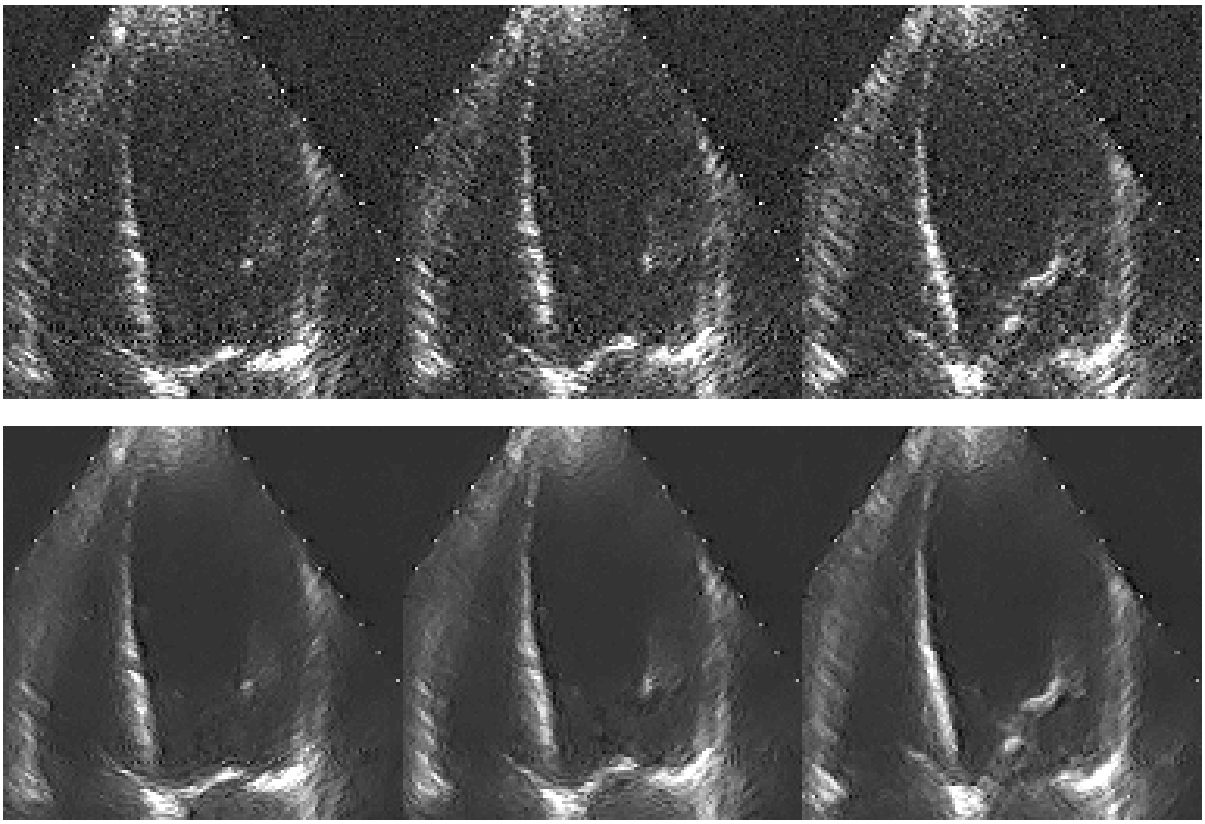


Figure 9: Spatio-temporal adaptive filtering of ultrasound data of the heart. Top: The images for the original image sequence. Bottom: Result after 3D adaptive filtering. Note the reduction of the specular noise when comparing between filtered and unfiltered image sets.

face data with a center-to-center spacing between slices that is larger than the in-plane pixel size. Westin et al. [30] introduced an affine model of the frequency characteristic of the filters to compensate for the data sampling anisotropy. In addition to changing the frequency distribution of the filters, they also show how this affine model can provide subvoxel shifted filters that can be used for interpolation of medical data.

6. Acknowledgments

This work was supported by CIMIT, and NIH grants P41-RR13218 and R01-RR11747. The authors gratefully acknowledges Dr. Raj Rangayyan, Dr. Klas Nordberg, Lars Wigström, and Dr. Juan Ruiz-Alzola for valuble comments on drafts of this chapter.

References

[1] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, 3rd edition edition, 1996. ISBN 0-13-322760-X.

[2] A. Papoulis. *Signal Analysis*. McGraw-Hill, 1977. ISBN 007484600.

[3] J. S. Lim. *Two-Dimensional Signal and Image Processing*. Prentice Hall, 1980. ISBN 0139353224.

[4] G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Kluwer Academic Publishers, 1995. ISBN 0-7923-9530-1.

[5] W. F. Schriber. Wirephoto quality improvement by unsharp masking. *J. Pattern Recognition*, 2:117–121, 1970.

[6] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1965. ISBN 0-07-048447-5.

[7] C. W. Helstrom. Image restoration by the methods of least squares. *J. Opt. Soc. Amer.*, 57(3):297–303, 1967.

[8] W. K. Pratt. *Digital Image Processing*. New York: Wiley, 1978.

[9] B. R. Hunt. The application of constrained least squares estimation to image restoration by digital computer. *IEEE Trans. Comput.*, C-22:805–812, 1973.

[10] G. L. Anderson and A. N. Netravali. Image restoration based on a subjective critera. *IEEE Trans. Sys. Man. Cybern.*, SMC-6:845–853, 1976.

[11] V. K. Ingle and J. W. Woods. Multiple model recursive estimation of images. In *Proc. IEEE Conf. Acoust., Speech, Signal Processing*, pages 642–645, 1979.

[12] J. F. Abramatic and L. M. Silverman. Nonlinear restoration of noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(2):141–149, March 1982.

[13] M. Nagao and T. Matsuyama. Edge preserving smoothing. *Computer Graphics Image Proc.*, 9:394–407, 1979.

[14] J. S. Lee. Digital image enhancement and noise filtering by local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2:165–168, 1980.

[15] J. S. Lim. Image restoration by short space spectral subtraction. *IEEE Trans. Acoust., Speech Sig. Proc.*, 28:191–197, 1980.

[16] H. Knutsson, R. Wilson, and G. H. Granlund. Anisotropic non-stationary image estimation and its applications — part I: Restoration of noisy images. *IEEE Transactions on Communications*, COM–31(3):388–397, March 1983.

[17] D. T. Kuan, A. A. Sawchuck, T. C. Strand, and P. Chavel. Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:165–177, 1985.

[18] A. C. Bovik, T. S. Huang, and D. C. Munson Jr. Edge-sensitive image restoration using order-constraint least squares method. *IEEE Trans. Acoust. Speech Sig. Proc.*, ASSP-33:1253–1263, 1985.

[19] K. Conradsen and G. Nilsson. Data dependent filters for edge enhancement of landsat images. *Computer Vision, Graphics, and Image Processing*, 38:101–121, 1987.

[20] H. Soltanianzadeh, J. P. Windham, and A. E. Yagle. A multidimensional nonlinear edge-preserving filter for magnetic resonance image restoration. *IEEE Transactions on Medical Imaging*, 4(2):147–161, 1995.

[21] X. You and G. Crebbin. A robust adaptive estimator for filtering noise in images. *IEEE Transactions on Image Processing*, 4(5):693–699, 1995.

[22] J. A. S. Centeno and V. Haertel. An adaptive image enhancement algorithm. *Pattern Recognition*, 30(7):1183–1189, 1997.

[23] B. Fischi and E. Schwartz. Adaptive nonlocal filtering:a fast alternative to anisotropic diffusion for image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(1):42–48, 1999.

[24] G. E Backus and F. Gilbert. The resolving power pf gross earth data. *Geophysical Journal of the Royal Astronomical Society*, 16:169–205, 1968.

[25] Ling Guang and Rabab K. Ward. Restoration of randomly blurred images by the Wiener filter. *IEEE Trans Acoustics, Speech, and Signal Processing*, 37(4):589–592, 1989.

[26] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(9):891–906, September 1991.

[27] H. Knutsson. Representing local structure using tensors. In *The 6th Scandinavian Conference on Image Analysis*, pages 244–251, Oulu, Finland, June 1989.

[28] H. Knutsson and G. H. Granlund. Fourier domain design of line and edge detectors. In *Proceedings of the 5th International Conference on Pattern Recognition*, Miami, Florida, December 1980.

[29] C.-F. Westin, S. Warfield, A. Bhalerao, L. Mui, J. Richolt, and R. Kikinis. Tensor Controlled Local Structure Enhancement of CT Images for Bone Segmentation. In *MICCAI'98, First Int Conf on Medical Image Computing and Computer-Assisted Intervention, 1998*, Lecture Notes in Computer Science 1496, pages 1205–1212. Springer Verlag, 1998.

[30] C.-F. Westin, J. Richolt, V. Moharir, and R. Kikinis. Affine adaptive filtering of CT data. *Medical Image Analysis*, 4(2):1–21, 2000.