MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

Working Paper 100   1974 (revised 1981)

The Application of Linear Systems Analysis
to Image Processing.
Some Notes.

Berthold K.P. Horn
Robert W. Sjoberg

ABSTRACT

The Fourier transform is a convenient tool for analyzing the performance of an image-forming system, but must be treated with caution. One of its major uses is turning convolutions into products. It is also used to transform a problem that is more naturally thought of in terms of frequency than time or space. We define the point-spread function and modulation transfer function in a two-dimensional linear system as analogues of the one-dimensional impulse response and its Fourier transform, the frequency response, respectively. For many imaging devices, the point-spread function is rotationally symmeteric. Useful transforms are developed for the special cases of a "pill box,", a gaussian blob, and an inverse scatter function.

Fourier methods are appropriate in the analysis of a defocused imaging system. We define a focus function as a weighted sum of high frequency terms in the spectrum of the system. This function will be a maximum when the image is in focus, and we can hill-climb on it to determine the best focus. We compare this function against two others, the sum of squares of intensities, and the sum of squares of first differences, and show it to be superior.

Another use of the Fourier transform is in optimal filtering, that is, of filtering to separate additive noise from a desired signal. We discuss the theory for the two-dimensional case, which is actually easier than for a single dimension since causality is not an issue. We show how to construct a linear, shift-invariant filter for imaging systems given only the input power spectrum and cross-power spectrum of input versus desired output.

Finally, we present two ways to calculate the line-spread function given the point-spread function.

FOURIER TRANSFORMS AND IMAGE FORMING DEVICES

The Fourier transform should be treated with caution when applied to problems in image analysis. It should not be considered as a wonderful black-box like tool that has the potential to solve many problems without involving us in the usual painful process of actually analyzing these problems. Instead it is a convenient tool for analyzing image forming system performance, much as it allows us to evaluate the performance of audio-reproduction equipment. One of the main uses of the Fourier transform is in turning convolutions into products, the latter usually being much simpler to understand. Certain problems also can be thought about more clearly in terms of the frequency domain, rather than the spatial domain.

It is easy to extend the concept of *convolution* to two dimensions:

$$\text{one-d: } h = f \otimes g \quad h(\tau) = \int f(t) \, g(\tau - t) \, dt$$

$$\text{two-d: } h = f \otimes g \quad h(\zeta, \eta) = \int \int f(x,y) \, g(\zeta - x, \eta - y) \, dx \, dy$$

Convolutions most often occur in the analysis of linear systems. If the *impulse response* of a mechanism is $g(t)$, we can calculate its response to an arbitrary stimulus $f(t)$ simply by convolving these two functions. The impulse response is obtained by applying an impulse to the input of the system and observing the behavior of its output over time. Note that causality requires that this response be one-sided, that is $g(t)$ better be zero for t less than zero.
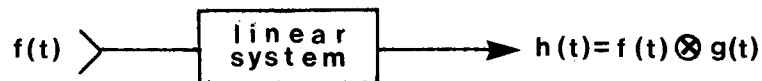
A similar situation obtains in the case of image forming systems. Instead of an impulse (in time) we here use a point source as the test input. The pattern recorded in the image plane is the output. This is called the *point-spread function* and depends on two (spatial) dimensions. It differs from the impulse response in that it depends on two spatial dimensions instead of one time dimension and in that there is no restriction on one-sidedness.

Time-invariance in a linear system means it produces the same impulse response at any time. This must be the case for convolution to apply to the problem of calculating the response to arbitrary inputs. In imaging systems we often find that the point-spread function is not quite constant as we image point sources in different areas of the image plane. The performance tends to deteriorate as we move away from the center, that is the point-spread function spreads out. This implies that the convolution method has to be applied with a little caution and is exact only in small regions in which the point-spread function does not vary.

It is easy also to extend the concept of *Fourier transform* to two dimensions:

$$\text{one-d: } g(w) = (2\pi)^{-1/2} \int f(t) \, e^{-iwt} \, dt$$

$$f(t) = (2\pi)^{-1/2} \int g(w) \, e^{iwt} \, dw$$

$$\text{two-d:} \quad g(u,v) = (2\pi)^{-1} \int \int f(w,y) \, e^{-i(ux+vy)} \, dx \, dy$$

$$f(x,y) = (2\pi)^{-1} \int \int g(u,v) \, e^{i(ux+vy)} \, du \, dv$$

We will ignore questions of existence and convergence and so on. Note that f and g may be complex-valued. Also note that if we transform functions of two variables we must also have two variables in the frequency domain. This makes sense if we realize that the transformed function contains exactly the same amount of information and that the original function can be reconstructed from it by applying the inverse transform.

We have already seen how linear systems can be characterized by their impulse response or point-spread function. One may also be interested in their *frequency response*. The frequency response is simply the Fourier transform of the impulse response. The frequency response then can be used equally well to characterize the system. Similarly we may define the *modulation transfer function* (or MTF) as the Fourier transform of the point-spread function and it too can be used to characterize imaging systems.

The frequency response tells us how much attenuation or amplification and how much phase shift a sine wave of any given frequency will suffer as it passes through the system. Similarly the modulation transfer function tells us how much amplification or attenuation a scene consisting of sinusoidally varying intensity will suffer in the imaging process. The sinusoidal variation here is along a particular spatial direction. Looking at such a scene we would see a series of parallel stripes at right angles to this direction, the separation between successive maxima being inversely proportional to the frequency. Unfortunately such scenes occur more rarely in images than sine waves happen in electrical linear systems. This makes it a bit harder to see what is going on.

Now we come to one of the main reasons one cares to think about the apparently unnecessary complication of Fourier transforms at all. And that is that Fourier transformation turns convolutions into *products*. In our simple linear system we have that:

$$\text{one-d:} \quad h = f \otimes g \quad \text{and} \quad \mathcal{F}(h) = \mathcal{F}(f) * \mathcal{F}(g), \quad H(\omega) = F(\omega)*G(\omega)$$

So if we feed a function with frequency spectrum F into a linear system with frequency response G, we get a function at the output which has frequency spectrum $H = F*G$. This can be inverse transformed to get $h(t)$ if we want. Similarly for our imaging system we have:

$$\text{two-d:} \quad h = f \otimes g \quad \text{and} \quad \mathcal{F}(h) = \mathcal{F}(f) * \mathcal{F}(g), \quad H(u,v) = F(u,v)*G(u,v)$$

The only difference is that the convolutions and Fourier transforms here are two-dimensional.

With imaging devices one often finds that the point-spread function is rotationally symmetric, at least for points close to the center of the image. It is thus helpful to know how the two-dimensional Fourier transform simplifies for such functions. The introduction of polar coordinates r, θ for the source domain (x,y) and of ρ,α for the frequency domain (u,v), leads to the following formula:

$$g(\rho) = \int_0^\infty f(r) \, r J_0(r\rho) \, dr$$

$$f(r) = \int_0^\infty g(\rho)\rho \, J_0(r\rho) \, d\rho$$

J_0 here is the Bessel function of order zero. Note the symmetry in the two transformations and that both f and g now are real-valued.

Some examples might be helpful:

A)   A "pill-box" of unit volume. That is a function which is constant within a disc centered on the origin and zero outside.

$$\frac{1}{\pi R^2} \text{ if } r < R \text{ otherwise } 0 \quad \text{transforms into } \frac{1}{\pi} \frac{J_1(R\rho)}{(R\rho)}$$

J_1 here is the Bessel function of order zero and $J_1(x)/x$ plays the same role here as $\sin(x)/x$ does for one-dimensional problems. It behaves vaguely similar, starting positive, becoming zero for x about 3.8 and then continuing with diminishing oscillations.

B)   A gaussian blob of unit volume.

$$\frac{1}{2\pi\sigma^2} \exp(-\tfrac{1}{2} r^2/\sigma^2) \quad \text{transforms into } \frac{1}{2\pi} \exp(-\tfrac{1}{2}\sigma^2\rho^2)$$

So it transforms into itself in some sense. This is another interesting property of this function. It is the only function, by the way, that is rotationally symmetric and can be decomposed into the product of two functions of one variable.
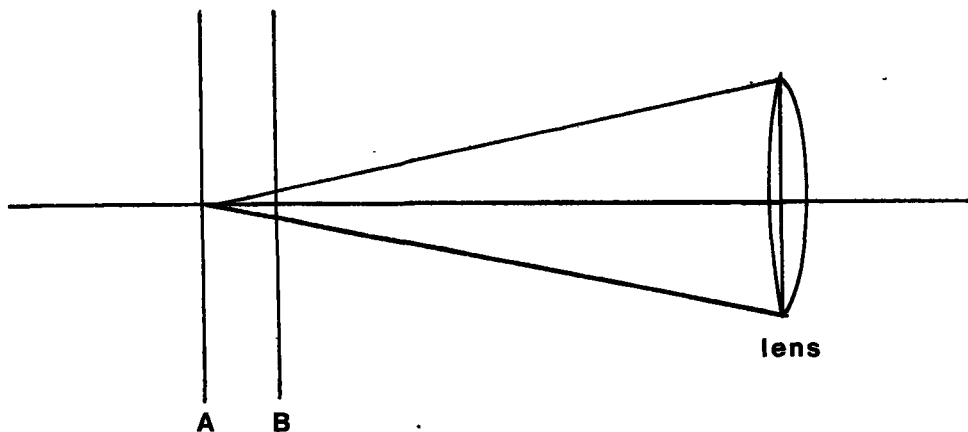
$$\exp(-\tfrac{1}{2}r^2/\sigma^2) = \exp(-\tfrac{1}{2}x^2/\sigma^2)\exp(-\tfrac{1}{2}y^2/\sigma^2)$$

C)   A scatter function of unit volume.

$$\frac{\sigma}{2\pi} \frac{e^{-\sigma r}}{r} \text{ transforms into } \frac{1}{2\pi} \frac{\sigma}{\sqrt{\sigma^2}}$$

Note that all of the above transformations are also valid if we exchange r and ρ. So we have a total of six useful transforms.

APPLICATION OF THESE TECHNIQUES TO ANALYSIS OF A DEFOCUSED IMAGE

lens

A   B

To characterize the defocused imaging system we will determine its point-spread function. Consider a point source imaged perfectly on the image plane A. Clearly if we now move the image plane to B we will cut the cone of light emerging from the lens and focused on A in a small disc. Inside this disc the intensity will be uniform, outside it will be zero. Hence the point-spread function of defocused imaging system is a pill-box.

The image in plane B then is the perfect image on plane A convolved with this pill-box. Since a convolution is not intuitively easy to deal with, we will transform it into a product. That is, the Fourier transform of the image at B will be the Fourier transform of the image at A multiplied by the Fourier transform of the point-spread function, our pill-box. We already know that this comes out to $(1/\pi)\, J_1(R\rho)/(R\rho)$, where R is the radius of the pill-box. $R = d\,(a/f)$, where d is the separation of plane A and plane B, a is the aperture of the lens, while f is its focal length. So R varies directly with the amount of image plane displacement. So as defocusing increases, the first zero of $J_1(R\rho)$ will approach closer and closer to the origin. That is, more and more of the high frequency components will be attenuated. So the effect of defocusing is mainly one of progressively reducing the high frequency components of the image.

One can then define a *focus function*, the sum of higher frequency components in the spectrum. This function will be at a maximum when the image is in best focus, and one can hill-climb on it to determine this position. In practice a weighted sum of high frequency terms is taken to account for the fact that beyond a certain frequency no detail will be seen even in the ideal image due to defects of the imaging device or because they are absent in the scene. One also wants to give increasing importance to higher frequency terms aside from this limitation.

This focus function can be compared against two others that are used at times:

a) SUM OF SQUARES OF INTENSITIES. It can be shown that the integral of the square of the original function is equal to the integral of the square of the transformed function. That is, total power can be calculated by summing up in the original domain or the transform domain. We can thus conclude that the sum of squares of intensities will equal the sum of squares of the transformed values. This then clearly will have a peak at best focus, but a much wider peak than our focus function defined above, since it includes the relatively constant low frequency terms as well.

b) SUM OF SQUARES OF FIRST DIFFERENCES. This is like summing the square of the derivative. Once again we can instead sum the square of the transform of the derivative and get the same result. Now the transform of the derivative of a function is the transform of the function multiplied by the frequency ($i\omega$ in the one-D case, $-\rho^2$ in the two-D case of applying the Laplacian). So we are getting an extra emphasis on the higher frequencies. This function is much better than a) above, but not yet as good as our focus function, defined earlier.

If this focusing method is to be implemented on a computer controlled system, we will have to deal with intensity functions defined only at discrete points and will thus have to extend our Fourier transform from the continuous case so far to the *discrete Fourier transform*. This is fairly simple and if we are dealing with a function defined at N equally spaced points we get:

$$\text{one-d:} \quad g(u\omega) = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} f(xT)\, w^{-xu} \quad u = 0,1\ldots N-1$$

$$f(xT) = \frac{1}{\sqrt{N}} \sum_{u=0}^{N-1} g(u\omega)\, w^{xu} \quad x = 0,1\ldots N-1$$

$$\text{where } w = e^{-2\omega j/N} \quad \omega T = 2\pi/N \quad j = \sqrt{-1}$$

T is the separation between points in the original domain, while ω is the separation of points in the transform domain. It appears at first sight that the calculation of a transform involves N^2 operations since we have to evaluate the sum over N terms a total of N times to get the N points in the transform. Fortunately it is possible to combine intermediate results in such a way that a total of $N \log_2 N$ operations are required. This is called the fast Fourier transform (or FFT). What is more, the above can be readily extended to two dimensions:

$$\text{two-d:} \quad g(uw_u, vw_v) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(xT_x, yT_y)\, w^{-(xu+vy)} \quad u,v = 0,1\ldots N-1$$

$$f(xT_x, yT_y) = \frac{1}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} g(u\omega_u, v\omega_v)\, w^{+(xu+vy)} \quad x,y = 0,1\ldots N-1$$

Fortunately the transform decomposes into succesive application of one-dimensional transforms along each row in turn followed by one-dimensional transforms along each column of values. This implies that the fast Fourier transform trick can still be used for two-dimensional data.
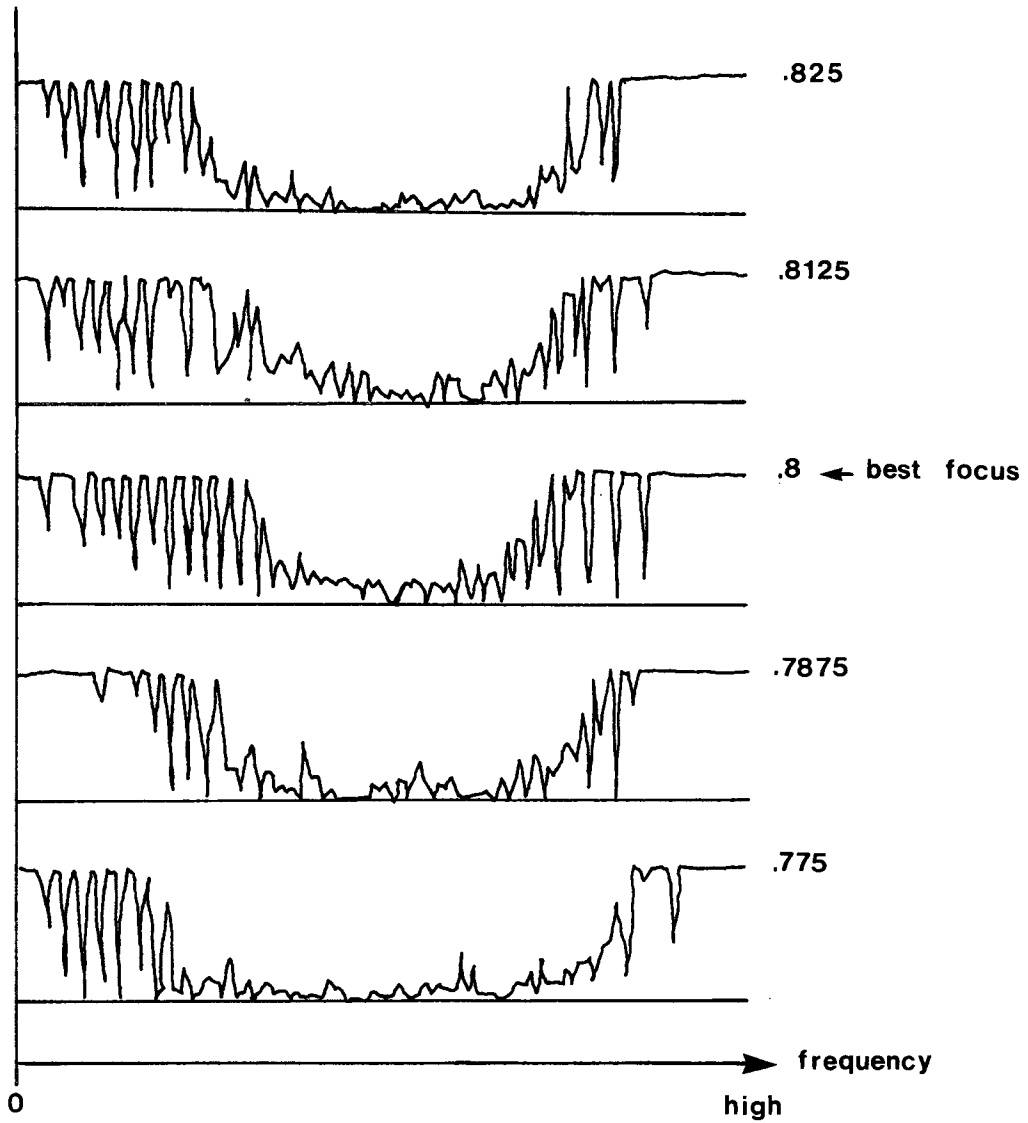
FOCUSING

Focusing is one aspect of the operation of optical systems which has so far eluded automation. Here we have a method which can be used to automatically focus projectors, cameras and image sensors. The same method of course can be used for determining range. This is one of the few image processing methods that depends in any significant way on Fourier transforms. The focusing method processes image intensities at a number of points at a very low level without trying to do anything akin to feature extraction. It thus requires an external agent such as the person pointing the camera or the program using the image sensor to decide which image area is to be used for focusing.

The central idea is that the effect of defocusing can best be described in the transform domain of the image as a reduction of the higher frequencies. To focus then one has to try and maximize the highest frequencies that are likely to be in the image at optimal focus. These will depend on both the scene being viewed and the limitations of the optical and image sensing system. In the absence of noise and other image sensing defects this would be very simple. In practice one has to investigate various trade-offs between noisiness and width of the function to be maximized if one wants to achieve near-optimal accuracy.
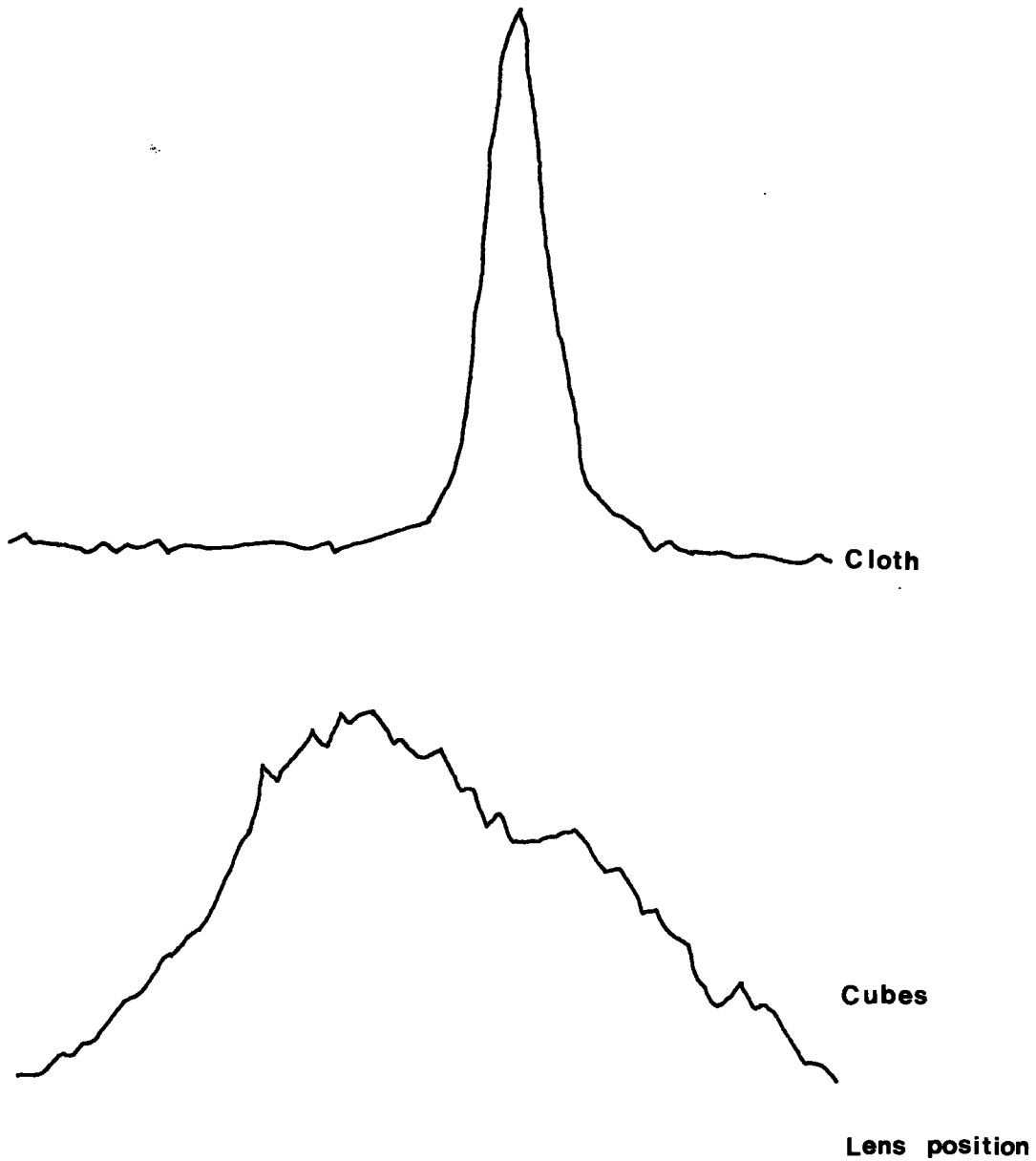
THE EFFECT OF DEFOCUSING

Consider first the image of a point source in an optical system, ignoring any effects of distortion and diffraction. It will be a point if the image plane is in the correct place. If however we displace the image plane it will cut the cone of light emerging from the lens and instead form a disc. The intensity inside this disc will be uniform and decreasing as we increase the size of the disc by further defocusing. The size of the disc will depend on the amount of image displacement and the aperture of the lens.

Limited power spectrum versus lens position. This is the minimum of some value and the sum of squares of real and imaginary parts of the Fourier transform of the series of intensities measured in the image.

Focusing function for two different scenes. This is the normalized sum of high frequency components of the power spectrum. Note that the maximum for the cloth scene is much tighter. This is in part because the lines in the cloth produce more high frequency components than the edges on the cubes and because the cubes are not all at the same distance.
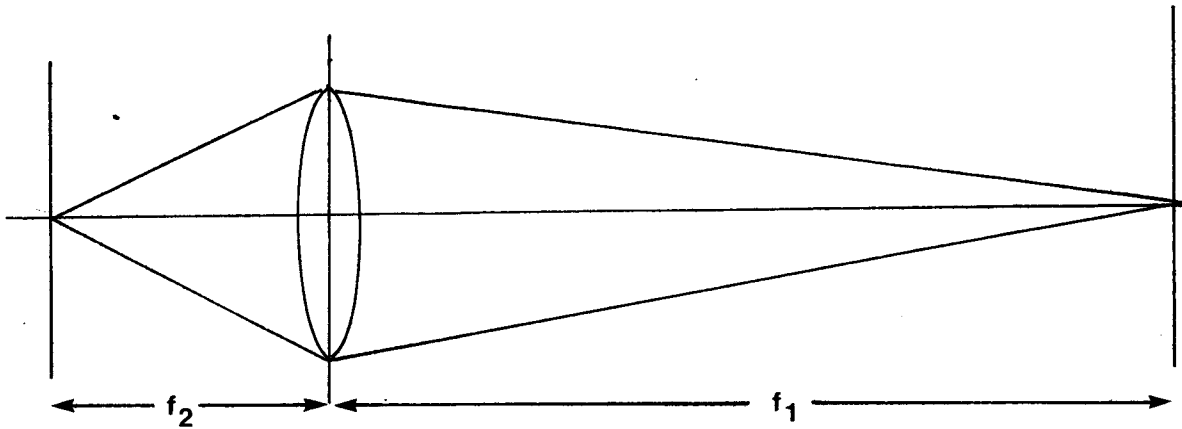

Cloth

Cubes

Lens position

Several tries at getting the focusing function for a particular scene. The focusing function is often noisier than this and it is non-trivial to hill-climb on it.

1 mm ⟷   Lens   position
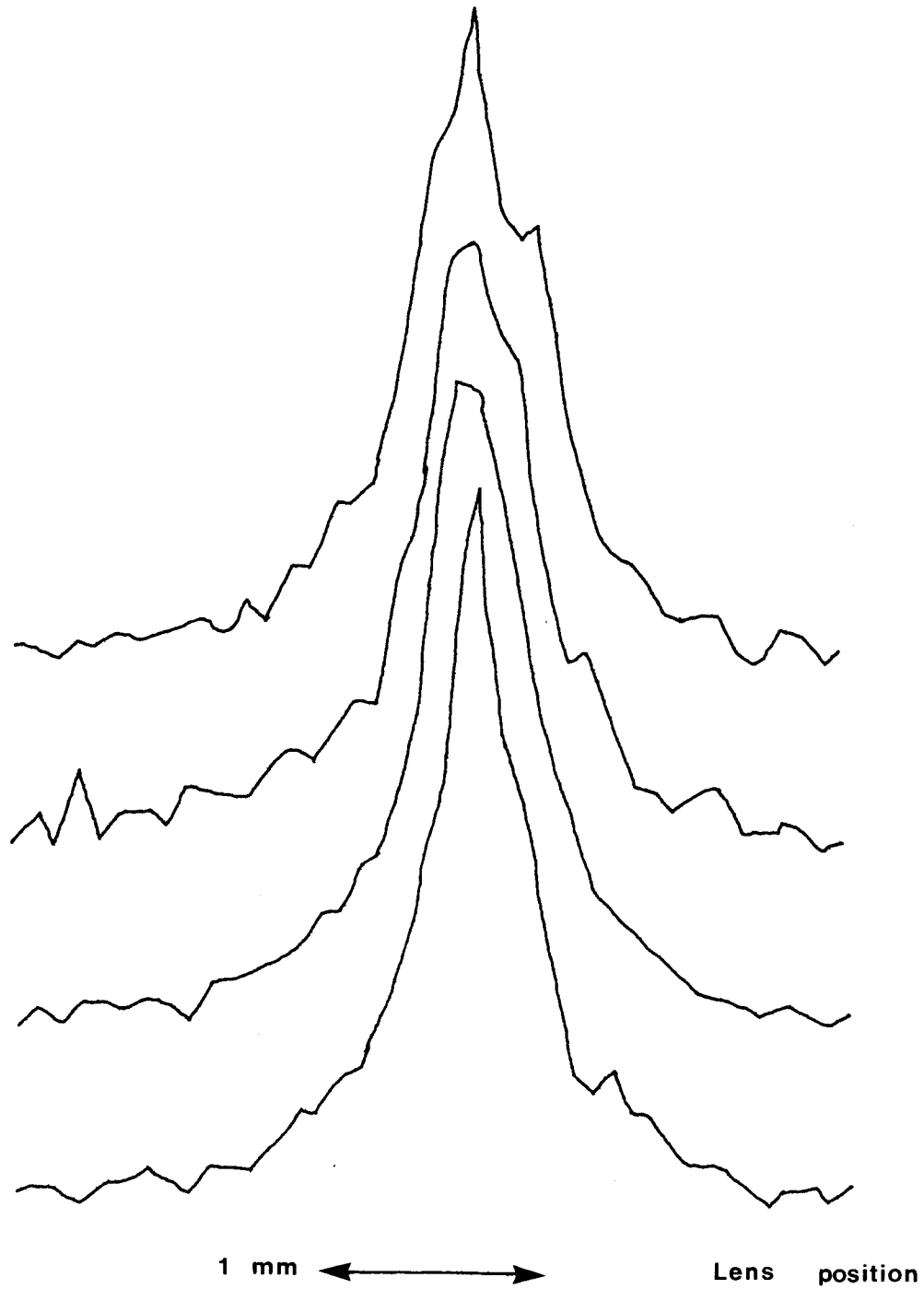
Effect of spherical abberation on ideal defocused point-spread function.

lens

actual   ideal
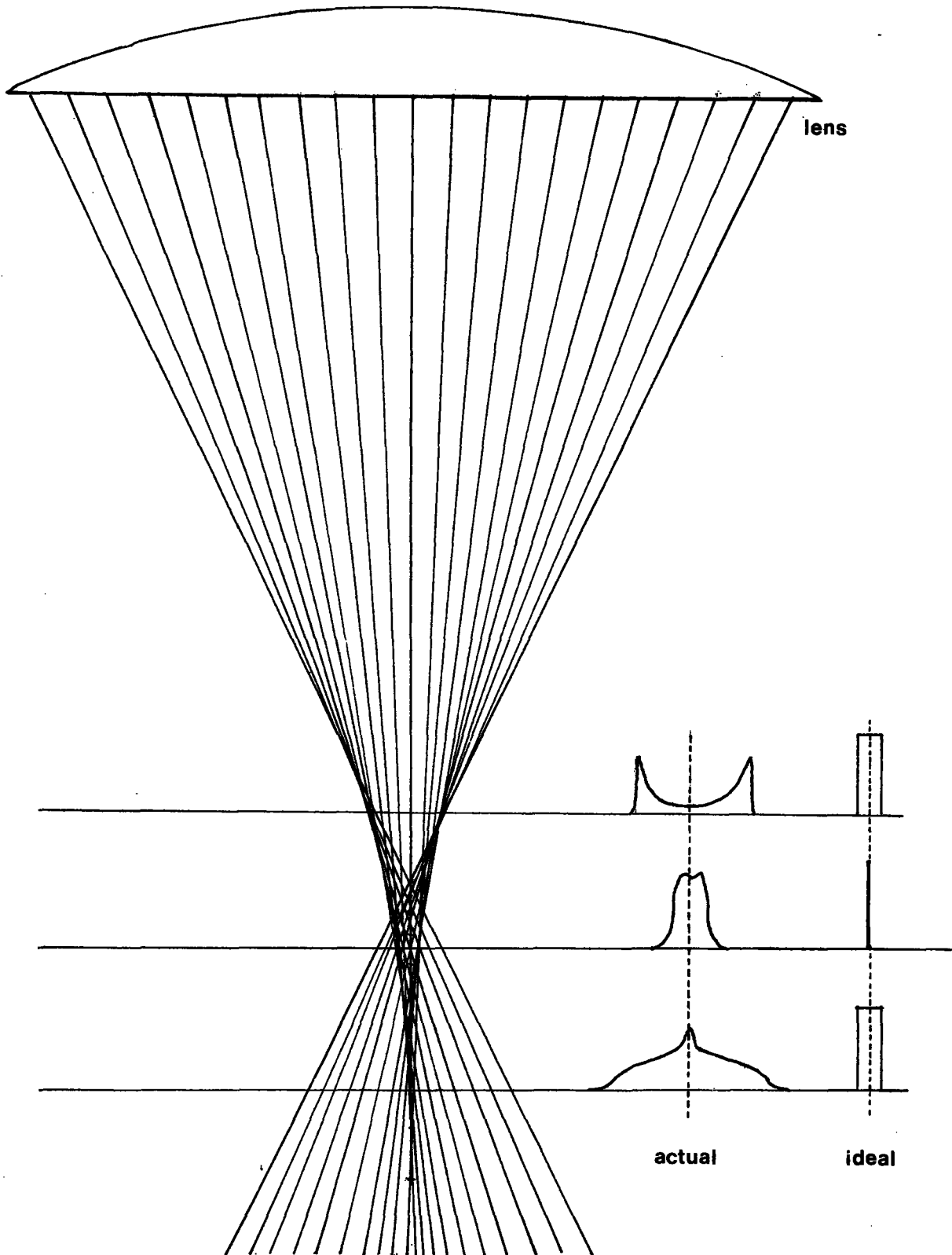
The lens formula tells us that $1/f_1 + 1/f_2 = 1/f$. It is useful to introduce the magnification parameter $M = f_2/f_1$. We then have $f_1 = (1 + 1/M)f$ and $f_2 = (1 + M)f$. If we let the distance from the object image be $d = f_1 + f_2$ we have:

$$\frac{df_1}{dM} = -\frac{1}{M^2}f$$

$$\frac{df_2}{dM} = f$$

$$\frac{dd}{dM} = (1 - \frac{1}{M^2})f$$

$$\frac{df_2}{df_1} = -M^2$$

$$\frac{dd}{df_1} = 1 - M^2$$

$$\frac{dd}{df_2} = 1 - \frac{1}{M^2}$$

OPTIMAL FILTERING

Consider the problem of designing a linear system which will recover a desired signal, given an input signal corrupted by additive noise. Obviously we need some criterion for deciding how well a given filter will work. The one most tractable is a least squares norm. That is we try and minimize the average square of the difference between the desired signal and the output of the filter when given the corrupted input. We also have to specify the class of possible filters. We have already said that we will consider only linear translation-invariant filters. We further need to define how we will specify the three signals: input, additive noise, and desired output. It turns out that all we need are the auto- and cross-correlations of these signals. The problem of finding the optimal filter given these constraints was first formulated and solved by Norbert Wiener. It turns out to be quite hard because of the fact that in the systems he was interested in, the response is one-sided. That is he had to restrict the possible solutions to those obeying causality. This condition is hard to express in the Fourier domain where the rest of the solution can be dealt with quite simply. Wiener solved this by a method of factorization of the spectrum.
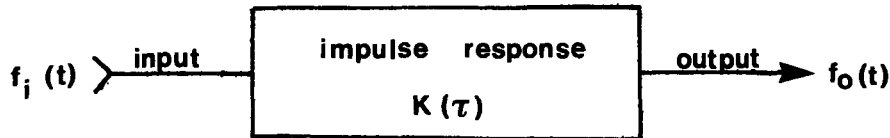
$$f_i(t) = f_m(t) + f_n(t)$$

$$f_o(t) \int_0^\infty f_i(t\text{-}\tau) K(\tau) d\tau$$

The average square error is then:

$$E = \lim_{T \to \infty} \frac{1}{2\tau} \int_T^{-T} \left[f_d(t) - f_o(t) \right]^2 dt$$

This must be minimized by choosing $K(\tau)$ properly.

The average square error can be expanded to give:

$$E = \lim_{T \to \infty} \int_{-T}^{T} \left[f_d^2(t) - 2f_d(t)f_o(t) + f_o^2(t) \right] dt$$

$$E = \lim_{T \to \infty} \int_{-T}^{T} \left[f^2{}_d(t) - 2f_d(t) \int_0^\infty f_i(t\text{-}\tau) K(\tau) d\tau \right.$$
$$\left. + \int_0^\infty f_i(t-\tau)K(\tau)d\tau \int_0^\infty f_i(t-\sigma) K(\sigma)d\sigma \right] dt$$

$$E = \varphi_{dd}(0) - 2 \int_0^\infty \varphi_{id}(\tau)K(\tau)d\tau + \int_0^\infty \int_0^\infty \varphi_{ii}(\tau\text{-}\sigma)K(\sigma)K(\tau)d\sigma d\tau$$

$$\varphi_{ii}(\tau) = \int f_i(\tau\text{-}t) f_i(t)dt$$

$$\varphi_{id}(\tau) = \int f_i(\tau\text{-}t) f_d(t)dt$$

Now to minimize the average squared error E we need something more than simple calculus since we are not varying merely a parameter, but are looking for a function. Optimization with respect to functions can be

handled by the calculus of variations.

The calculus of variation is invoked to solve this minimization problem. At a minimum we would expect that adding a small function to $K(\tau)$ would not change the resultant error very much. To be precise we have

$$\frac{d}{d\varepsilon} \text{norm}\,(K(\tau) + \varepsilon\,\delta K(\tau)) = 0 \quad \text{for } \varepsilon = 0 \text{ for all } \delta K(\tau)$$

where norm(...) is the result of applying the above average square operation. (As in the calculus case we get a local extremum; proof of global minimum is harder.) Substituting $K(\tau) + \varepsilon\delta K(\tau)$ in the above expression for E instead of $K(\tau)$, then differentiating with respect to ε and then setting ε equal to zero:

$$0 = \int_0^\infty \left[\varphi_{id}(\tau) - \int_0^\infty \varphi_{ii}(\tau - \sigma)\,K(\sigma)d\sigma \right] \delta K(\tau)\,d\tau$$

Since this must be true for all $\delta K(\tau)$ we can conclude that (for $\tau > 0$):

$$\varphi_{id}(\tau) = \int_0^\infty \varphi_{ii}(\tau - \sigma)\,K(\sigma)\,d\sigma$$

Now, given all of the various auto- and cross-correlations indicated how do we find $K(\tau)$? One idea might be to note that the integral is actually just a convolution:

$$\varphi_{id} = \varphi_{ii} \otimes K$$

We could then take Fourier transforms and convert this into a product. Its then easy to solve for the required filter function K:

$$\Phi_{id} = \Phi_{ii} * \mathcal{F}(K)$$

Here we are dealing with power spectrums and cross-power spectrums, which can be used equally well to describe the signals. The solution seems at hand. But alas, we have a bug! We forgot that the equation only applies for $\tau > 0$. And this is where all the hair comes in, that is, spectrum factorization.

Example: We want a predictive filter, given a noise-free input. Let the prediction interval be α, then:

$$f_d(t) = f_i(t+\alpha), \quad f_n(t) = 0$$

So $$\varphi_{id}(\tau) = \varphi_{ii}(\tau + \alpha)$$

If we were not constrained by causality, we could simply choose $K(\tau)$ to be a unit pulse at $t = -\alpha$! And the above convolutional equation would be satisfied and we would have zero error.

OPTIMAL FILTERING IN THE CASE OF IMAGING SYSTEMS

These systems are described by their point-spread functions, much as a linear system in time is described by its impulse response. The major differences are:

1)   We have two dimensions for the point-spread function, instead of the one in the case of the impulse response.

2)   The point-spread function is a function of position, not time.

3)   The response is not restricted to any part of the plane for the point-spread function, while the impulse response of a realizable linear device is nonzero only for $t > 0$.

The analysis proceeds exactly as above except that f_i, f_o, f_m, f_n, K are all functions of two variables instead of one. Similarly the correlations φ_{ii} and φ_{id} are also functions of two variables, as will be the power spectra Φ_{ii} and Φ_{id}.

But the most important difference is that we can use the Fourier transform of the indicated convolution to get at the filter response, since we are no longer restricted to solutions that have to be zero in certain areas.

$$\varphi_{id}(\alpha,\beta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi_{ii}(\alpha - \gamma, \beta - \delta)\, K(\gamma,\delta)\, d\gamma d\delta$$
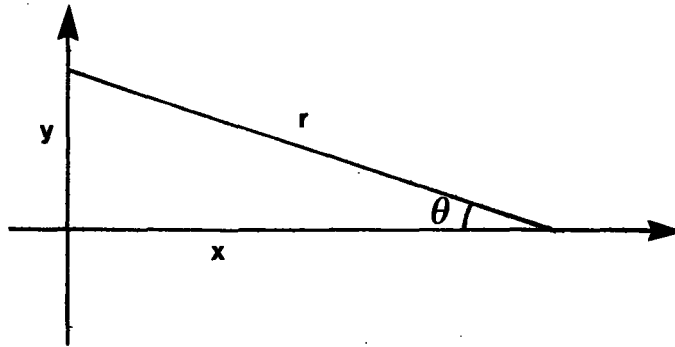
This has to be satisfied for all α, β

$$\Phi_{id}(u,v) = \Phi_{ii}(u,v) * K(u,v)$$

So finally we get: $K(u,v) = \Phi_{id}(u,v)/\Phi_{ii}(u,v)$

So given the power spectrum of the input, and the cross-power spectrum of input versus desired output, we can easily calculate the filter required for optimal filtering.

CALCULATING LINE-SPREAD FROM POINT-SPREAD

Let the point-spread function be a(r) and the line-spread function b(x).

$x = r\cos\theta,\ y = r\sin\theta,\ y = x\tan\theta$

$r = x\sec\theta,\ dy = x\sec^2\theta\, d\theta$

$$b(x) = \int_{-\infty}^{\infty} a(r)\, dy = \int_{-\pi/2}^{\pi/2} a(x\sec\theta)\, x\sec^2\theta\, d\theta$$

$$b(x) = 2\int_{0}^{\pi/2} a(x\sec\theta)\, x\sec^2\theta\, d\theta$$

$$\theta = \cos^{-1}(x/r), \quad d\theta = (x/r^2) / \sqrt{1 - (x/r)^2} \, dr$$

$$b(x) = 2 \int_x^\infty a(r) \, x \, (r/x)^2 \, (x/r^2) / \sqrt{1 - (x/r)^2} \, dr$$

$$b(x) = 2 \int_x^\infty a(r) \, r / \sqrt{r^2 - x^2} \, dr$$

ALTERNATIVE DERIVATION OF LINE-SPREAD FROM POINT-SPREAD FORMULA

The line-spread b(x) is the convolution of the point-spread a(r) and the impulse-line $\delta(x)$ in two dimensions.

$$b(x,y) = a(x,y) \otimes \delta(x)$$

$$B(u,v) = A(u,v) \, \delta(v)$$

where A and B are the two-dimensional Fourier transforms of a and b. Since a is rotationally symmetric we can use the special formula for the forward transform. Since $A(u,v) \, \delta(v)$ will be zero except on the u-axis we can use a one-dimensional inverse transform to obtain the essentially one-dimensional b(x,y).

$$g(\rho) = \int_0^\infty a(r) \, r \, J_0(r\rho) \, dr$$

$$b(x) = 2 \int_0^\infty \cos(x\rho) \, g(\rho) \, d\rho$$

$$b(x) = 2 \int_0^\infty a(r) \, r \int_0^\infty J_0(r\rho) \cos(x\rho) \, d\rho \, dr$$

The inner integral is zero for $r < x$ and $1/\sqrt{r^2 - x^2}$ for $r > x$

$$b(x) = 2 \int_x^\infty a(r) \, r / \sqrt{r^2 - x^2} \, dr$$

It usually makes more sensee though to actually perform the indicated transformations.