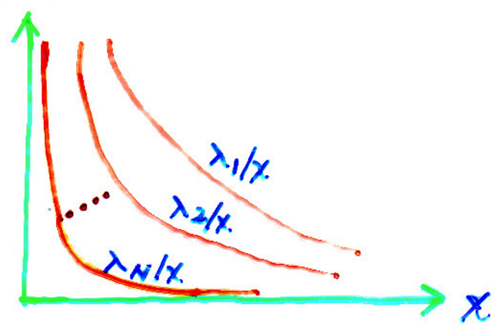
Chapter 3 Stochastic Processes in Optics

Consider a family of curves, *e.g.*, $\{ f(x;\lambda) = \lambda/x,\ \ with\ \ \lambda \in RV\}$. The individual members of the family are defined by their λ values. If λ is a RV, then $f(x;\lambda)$ is called a **stochastic process**. For example,



$f(x;\lambda)$ can be a smooth function of *x*; the randomness lies in which curve *f*(*x*) of the family λ was chosen. The occurrence probability of λ is depicted by its associated *pdf* $p_\Lambda(\lambda)$.

For clarity, in the following, we will depict two examples of stochastic processes:

- $\{ f(t;d) = \cos(\omega t - 2k\,d),\ \ with\ \ d \in RV\}$ are Radar signals, which are stochastic with the RV *d* (random distance to the unknown source). $f(t;d)$ is a smooth function of *t*.

- $\{ f(k;\vec{\phi}) = e^{jkz + jk\Delta(\beta)}\}$. $\phi = k\{\Delta(\beta_1), \Delta(\beta_2), ..., \Delta(\beta_N)\}$, which β_i relates to the inclination angle (or spatial frequency) of the optical wave, denotes a random phase across the pupil of an imaging system.

For an ensemble of functions { $f(x;\lambda)$ }, the average $< f(x;\lambda) >= \int f(x;\lambda)\, p_\Lambda(\lambda) d\lambda$ taking over the random values of λ, which is called an **ensemble average** of the stochastic process.

3.1 Power Spectrum

We can conveniently define a definite Fourier transform of a stochastic process $f(x;\vec{\lambda})$ as

$$F_L(\omega;\vec{\lambda}) \equiv \int_{-L}^{+L} f(x;\lambda) e^{-j\omega x} dx\,.$$

Thus, the power spectrum of $f(x;\vec{\lambda})$ can be revealed by

$$S_f(\omega) = \lim_{L\to\infty}\left[< \left| F_L(\omega;\vec{\lambda}) \right|^2 > \big/ (2L) \right].$$

The denominator $2L$ is needed in order to yield a finite value for $S_f(\omega)$. This formula describes the allocation of average power to the various frequencies comprising $f(x;\vec{\lambda})$.

3.2 Autocorrelation Function

We can define the autocorrelation function $R_f(x;x_0)$ of a stochastic process $f(x;\vec{\lambda})$ as

$$R_f(x;x_0) \equiv < f(x_0;\lambda) f^*(x+x_0;\lambda) >,$$

which can be viewed as an average over RV's $\vec{\lambda}$ of $f(x;\vec{\lambda})$ at same arbitrary point x_0 times $f^*(x+x_0;\vec{\lambda})$ at a point distance x away.

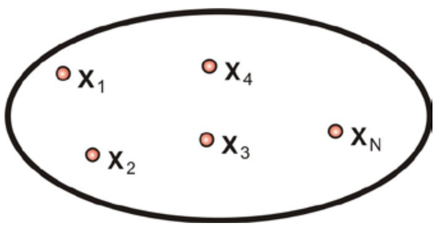
If $R_f(x;x_0)$ equals $R_f(x)$ regardless of position x_0, and if the mean value of $f(x;\vec{\lambda})$ is independent of x, *i.e.*,

$$< f(x;\lambda) f^*(x';\lambda) > \equiv R_f(x-x') \text{ and } < f(x;\lambda) > = \text{constant},$$

we call this stochastic process ***wide-sense stationary***. Furthermore, we can calculate correlation in fluctuation of $f(x;\vec{\lambda})$ from its means at the two points by

$$\rho(x;x_0) = <[f(x_0;\lambda) - \langle f(x_0;\lambda)\rangle][f^*(x+x_0;\lambda) - \langle f^*(x+x_0;\lambda)\rangle]>$$
$$= R_f(x) - |< f(x_0;\lambda) >|^2 .$$

Now consider *n* stochastic processes $f_1(x_1;\vec{\lambda})$, $f_2(x_2;\vec{\lambda})$, …, $f_n(x_n;\vec{\lambda})$ (note the processes are different f_i, and so are their observation points ***x****i*).



Note also that a stochastic process $f(x;\lambda)$ at fixed x_0 is essentially a random variable $f(x_0;\lambda)$. Therefore, we can define a joint probability density function as $p(f_1, f_2,...., f_n)$. Shift each of the processes by ***x***: $x_i' = x_i + x$.

If $p(f_1, f_2,...., f_n) = p(f_1', f_2',...., f_n')$, where $f_i'(x_i';\vec{\lambda}) = f_i(x_i + x;\vec{\lambda})$ and regardless of the size of *x*, we call $\{f_i, f_2,..., f_n\}$ to follow ***strict-sense stationarity***, which indicates that *the joint statistics* of the ***n*** stochastic processes *are independent of their absolute position* in the signal.

3.3 Fourier Transform Theorem

For a wide-sense stationary process, $R_f(x)$ and $S_f(\omega)$ form a Fourier transform pair,

$$
\begin{aligned}
S_f(\omega) &\equiv \lim_{L\to\infty} <| F_L(\omega; L) |^2> /(2L) \\
&= \lim_{L\to\infty} \frac{1}{2L}\int_{-L}^{+L} dx\ e^{-j\omega x}\int_{-L}^{+L} dx'\ e^{j\omega x'} < f(x;\vec{\lambda}) f^*(x';\vec{\lambda}) > \\
&= \lim_{L\to\infty} [\ \frac{1}{2L}\int_{-L}^{+L} dx\ e^{-j\omega x}\ e^{j\omega x}\ ]\cdot \int_{x-L}^{x+L} dy\ e^{-j\omega y} R_f(y) \qquad . \\
&= \int_{-\infty}^{+\infty} dy\, e^{-j\omega y}\ R_f(y)
\end{aligned}
$$

Here we have done an integration variable substitution $y = x - x' \quad \Rightarrow \quad dx' = -dy$.

Let us consider the power spectrum of a signal $f(x;\vec{\lambda})$. Assuming that f is uncorrelated at different points but correlates with itself at the same point, *i.e.*,

$$
< f(x;\vec{\lambda}) f^*(x';\vec{\lambda}) > = \begin{cases} 0; & x \neq x' \\ m_2(x); & x = x' \end{cases}.
$$

(i) Let $m_2(x) \equiv < | f(x;\vec{\lambda}) |^2 >$ depends only on x, then

$$
S_f(\omega) = \lim_{L\to\infty} \frac{1}{2L}\int_{-L}^{+L} m_2(x)\ dx = const\ .
$$

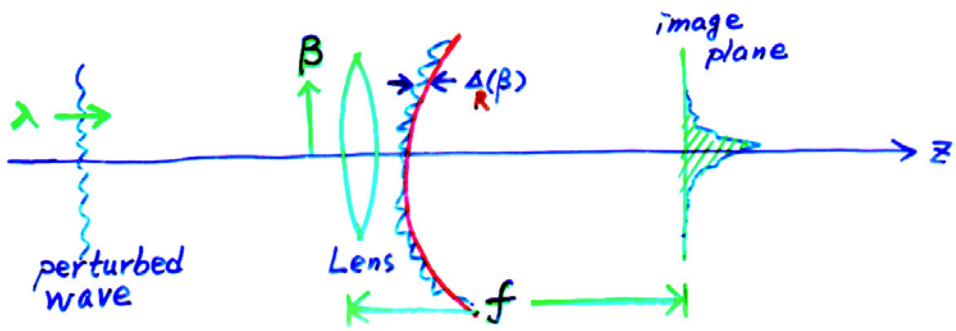
(ii) If $m_2(x) \equiv < | f(x;\vec{\lambda}) |^2 > = const = m_2$ independent of x, then $S_f(\omega) = m_2$.

Both cases indicate the uncorrelated stochastic process $f(x;\vec{\lambda})$ to have a white power spectrum.

An application example

Consider again an imaging system with a point spread function

$$
S(y) = \frac{ka}{\pi f}\mathbf{sinc}^2\left(\frac{ka}{f}y\right).
$$



At any instant of time, the net phase $\Delta(\beta)$ at the normalized lateral displacement point β on the pupil can be decomposed to be

$$\Delta(\beta) = \Delta_{o}(\beta) + \Delta_{R}(\beta),$$

where $\Delta_{o}(\beta)$ denotes a deterministic contribution from the imaging lens which is constant in time; $\Delta_{R}(\beta)$ is random in time due to the turbulence. Therefore, the transfer function in the image plane can be expressed as

$$T(\omega) = k\int_{\omega-\beta_{o}}^{\beta_{o}} e^{jk[\Delta(\beta)-\Delta(\beta-\omega)]}\, d\beta\,.$$

Due to a long-time exposure, the time averaged transfer function shall become

$$< T(\omega) >= k\int_{\omega-\beta_{o}}^{\beta_{o}} d\beta\; e^{jk[\Delta_{o}(\beta)-\Delta_{o}(\beta-\omega)]} \cdot < e^{jk[\Delta_{R}(\beta)-\Delta_{R}(\beta-\omega)]} > \;,$$

indicating that we need to know the statistics for $\Delta_{R}(\beta)$ at two points β and $\beta-\omega$.

- Let us assume $\Delta_{R}(\beta)$ obeys the central limit theorem, leading to that $\Delta_{R}(\beta)$ is normal at each β.
- $< \Delta_{R}(\beta) >= 0$, indicating the phase fluctuations can go negative as often as positive.

- $\Delta_R(\beta)$ is strict-sense stationary for two points β and $\beta-\omega$ in the pupil plane, *i.e.*,

$p[\Delta_R(\beta), \Delta_R(\beta-\omega)] = p[\Delta_R(\beta-\omega_{\mathbf{o}}), \Delta_R(\beta-\omega-\omega_{\mathbf{o}})]$ for arbitrary ω_o.

Thus, $p[\Delta_R(\beta), \Delta_R(\beta-\omega)] = p[\Delta_R, \Delta_R']$ is independent of β. The joint probability statistics for phase at two points of separation ω is independent of their absolute positions in the pupil plane.

Based on these results, we can further deduce

- $p(\Delta_R)$ = marginal probability of $\Delta_R = \int p[\Delta_R, \Delta_R']\, d\Delta_R'$ is independent of β.

- $\sigma^2 = \int \Delta_R{}^2 p(\Delta_R)\, d\Delta_R = \int \Delta_R'^2 p(\Delta_R')\, d\Delta_R' = \sigma'^2$. Every point in the pupil plane suffers the same turbulence effect. And

- $$\begin{aligned} \rho &= < \Delta_R(\beta)\Delta_R(\beta-\omega) > \\ &= \int \Delta_R(\beta)\Delta_R(\beta-\omega) p(\Delta_R(\beta), \Delta_R(\beta-\omega))\, d\Delta_R(\beta) d\Delta_R(\beta-\omega) \\ &= \rho(\omega) = depends\ on\ \omega\ only. \end{aligned}$$

Correlation of the phases at two points depends only on the distance between the two points.

Based on these findings, the average transfer function becomes

$< T(\omega) >= T_o(\omega)T_R(\omega) = T_o(\omega)\phi_{GB}(\omega)$, where

$T_o(\omega) = k\int_{\omega-\beta_o}^{\beta_o} e^{jk[\Delta_o(\beta)-\Delta_o(\beta-\omega)]}\, d\beta$ = from fixed, intrinsic lens property; and

$$\begin{aligned} \phi_{GB}(\omega) &= \int\int e^{jk[\Delta_R(\beta)-\Delta_R(\beta-\omega)]} p[\Delta_R(\beta), \Delta_R(\beta-\omega)] d\Delta_R(\beta)\, d\Delta_R(\beta-\omega) \\ &= e^{-\sigma^2\omega^2[1-\rho(\omega)]} \end{aligned}$$, from

turbulence effect, which depends on λ, σ, and ρ.

For a numerical estimation, let us assume $\sigma \simeq \lambda/4$, $k\sigma = \frac{2\pi}{\lambda}\sigma \simeq \pi/2$,

$$\therefore T_R \sim 0.08^{(1-\rho)} = \begin{cases} 0.08\,, & \rho = 0 \quad uncorrelated \\ 1\,, & \rho = 1 \quad perfect\ correlated \end{cases} \cdot$$

3.4 Transfer Theorem for Power Spectrum

Consider an image formation process $i(y) = \int_{-\infty}^{+\infty} dx\, s(y; x) o(x)$.

(i) If $o(x)$ is a stochastic process $o(x; \vec{\lambda})$, what is $S_i(\omega)$ of the image?

To answer the question, note first

$$\begin{aligned} I_L(\omega; \vec{\lambda}) &= \int_{-L}^{+L} dy\, i(y; \vec{\lambda}) e^{-j\omega y} \\ &= \int_{-\infty}^{+\infty} dx\, o(x; \vec{\lambda}) \int_{-L}^{+L} dy\, s(y-x) e^{-j\omega y} \\ &\overset{Let\ y-x=t}{=} \int_{-\infty}^{+\infty} dx\, o(x; \vec{\lambda}) e^{-j\omega x} \int_{-L-x}^{L-x} dt\, s(t) e^{-j\omega t} \end{aligned}$$

$$I_L(\omega; \vec{\lambda}) \overset{L\to\infty}{\to} \int_{-\infty}^{+\infty} dx\, o(x; \vec{\lambda}) e^{-j\omega x} \bullet \int_{-\infty}^{\infty} dt\, s(t) e^{-j\omega t} = O(\omega; \vec{\lambda}) \cdot T(\omega).$$

$$\begin{aligned} S_i(\omega) &= <|I_L(\omega; \vec{\lambda})|^2> = <|O(\omega; \vec{\lambda})|^2> \cdot |T(\omega)|^2 \\ &= S_o(\omega) \cdot MTF^2(\omega) \end{aligned}.$$

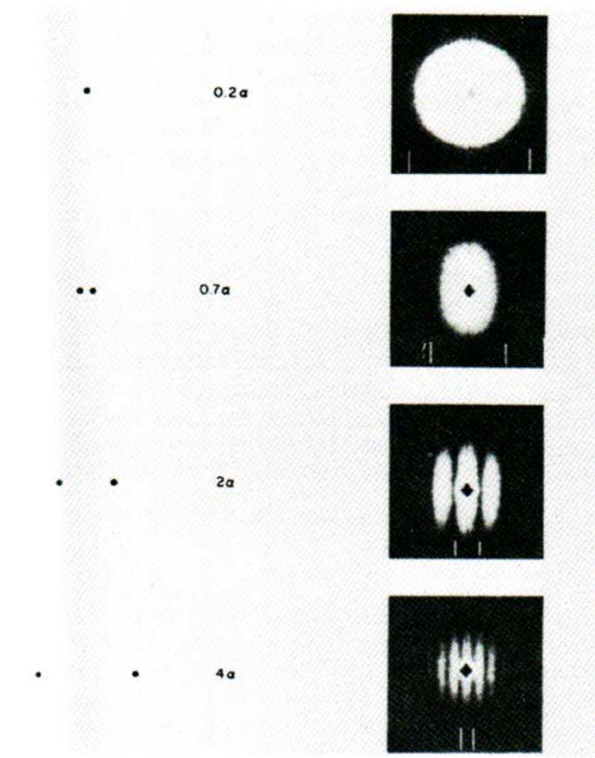
Here the modulus transfer function $MTF(\omega)$ of the imaging system is deduced from $|T(\omega)|$.

If $S_o(\omega)$ =constant=k (*i.e.*, white power spectrum), then

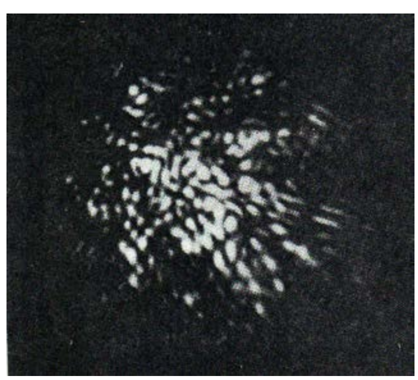
$$MTF(\omega) = [S_i(\omega)/k]^{1/2}.$$

This result provides **a simple way to determine** $MTF(\omega)$.

We consider an object formed by two point sources locating at $\pm a/2$ in the object plane: $o(x) = B\delta(x - a/2) + B\delta(x + a/2)$. The power spectrum of the object can be easily calculated to be: $|O(\omega)|^2 = 4B^2 \cos(a\omega/2)$, revealing a set of fringes with periods depending on *a*.



In the following figure, we show a typical speckle pattern from a short-exposure image of a single star


.

(ii) If $s(y-x;\vec{\lambda})$ is a stochastic process, for example, a telescope is repeatedly imaging an object $o(x)$ through short-term turbulence. Many images $\{i(x;\vec{\lambda})\}$ can be formed.

Note that

$$
\begin{aligned}
&< I(\omega;\vec{\lambda})I^*(\omega+\Delta\omega;\vec{\lambda}) >_{\vec{\lambda}} \\
&=< T(\omega;\vec{\lambda})T^*(\omega+\Delta\omega;\vec{\lambda}) >_{\vec{\lambda}} O(\omega)O^*(\omega+\Delta\omega)
\end{aligned}
.
$$

This equation suggests that the autocorrelation of the image spectrum contains phase information about the object and therefore provides a simple way to retrieve the phase of the object from the observed images.

3.5 Noise

Noise *n*(*x*) in a signal can be defined as any departure of the measured signal (*i.e.*, the data) *d*(*x*) from its ideal value (*i.e.*, the true signal) *s*(*x*):

$$n(x) \equiv d(x) - s(x) .$$

If both *s* and *n* are stochastic processes with each obeying arbitrary probability laws $p_S(s)$ and $p_N(n)$ at *x*, then

$$d(x;\vec{\lambda},\vec{\lambda}') \equiv s(x;\vec{\lambda}) + n(x;\vec{\lambda}') .$$

Here RVs $s(x)$ and $n(x')$ do not correlate at any *x* and *x*', *i.e.*, *s* and *n* are statistically independent RVs, then

$$< s(x;\vec{\lambda})n(x';\vec{\lambda}') >=< s(x;\vec{\lambda}) >< n(x';\vec{\lambda}') >= 0 \quad (\because < n(x';\vec{\lambda}') >= 0).$$

The noise $n(x;\vec{\lambda})$ is called ***additive noise***.

Defining a joint probability density function of input *s* and output *d* at an arbitrary but fixed *x*:

$$p_{DS}(d;s) = p(d\,|\,s)\,p_S(s).$$

If $p_S(s)$ is known, $p(d\,|\,s) = p(s+n\,|\,s) = p(n\,|\,s)$ from $d = s+n$ and a fixed *s*. Because *n* is additive $p(s\,|\,n) = p_S(s)$, we can further deduce

$$p(n\,|\,s) = \frac{p(s\,|\,n)\,p_N(n)}{p_S(s)} = \frac{p_S(s)\,p_N(n)}{p_S(s)} = p_N(n).$$

$$\therefore\ p(d\,|\,s) = p_N(n) = p_N(d-s).$$

Thus, $p_{DS}(d,s) = p(d\,|\,s)\,p_S(s) = p_N(d-s)\,p_S(s)$.

Note that

1. signal $s(x;\vec{\lambda})$ usually has a strong correlation over some finite range of *x*;
2. white noise $n(x;\vec{\lambda})$ is uncorrelated in *x*, leading to

   $<n(x;\vec{\lambda})n(x';\vec{\lambda})>= m_2(x)\delta(x-x')$, a random noise with a white power spectrum.

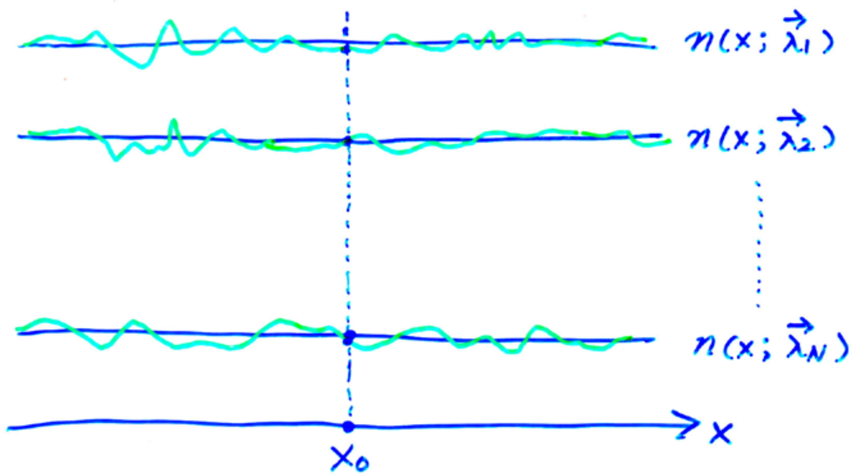
$$\begin{aligned}\sigma_n^2(x) &=< n^2(x;\vec{\lambda})> - < n(x;\vec{\lambda})>^2 \quad if\ < n(x;\vec{\lambda})>= 0 \\ &=< n^2(x;\vec{\lambda})> \equiv m_2(x)\end{aligned}.$$

Wide-sense stationary noise $\Rightarrow$ $<n(x;\vec{\lambda})n(x';\vec{\lambda})>= \sigma_n^2\delta(x-x')$.

3.6 Ergodic Property

Let $\vec{\lambda}_k$ indicates the k-th particular set of $\vec{\lambda}$ values, $n(x;\vec{\lambda}_k)$ k=1, 2, 3 …, N.

At any fixed x=x_o, there is a probability law $p_{x_o}(n)$ for random values of n.



A histogram built up of noise occurrences across curve $n(x;\vec{\lambda}_k)$ (along x) should be identical to the probability law $p_{x_o}(n)$, with x_o fixed, as formed from curve to curve (*i.e.*, $p_{x_o}(n)$ is independent of x_o). This is the principle of **Ergodic Theorem**:

$$< n(x;\vec{\lambda}_k) >_{\vec{\lambda}} = \int_{-\infty}^{+\infty} n(x;\vec{\lambda}_k)\, dx \,.$$

3.7 Optimum Restoring Filter

If $i(y;\vec{\lambda},\vec{\lambda}') = \int_{-\infty}^{+\infty} dx\, s(y-x) o(x;\vec{\lambda}) + n(y;\vec{\lambda}')$ = blur due to deterministic spread function + additive noise.

By taking Fourier transform, we obtain

$$\overset{FT}{\rightarrow} \quad I(\omega;\vec{\lambda},\vec{\lambda}') = T(\omega)O(\omega;\vec{\lambda}) + N(\omega;\vec{\lambda}').$$

We can construct a restoring filter function $Y(\omega)$ to retrieve $\hat{O}(\omega) = Y(\omega)I(\omega)$ from an image.

An appropriate restoring filter $Y(\omega)$ can be found by defining

$$e = <\int_{-\infty}^{+\infty} [o(x;\lambda) - \hat{o}(x;\lambda)]^2 dx>_\lambda = <\int_{-\infty}^{+\infty} |O(\omega;\lambda) - \hat{O}(\omega;\lambda)|^2 d\omega>_\lambda.$$

By substituting $\hat{O}(\omega) = Y(\omega)[T(\omega)O(\omega) + N(\omega)]$ into the above equation

$$\begin{aligned} e &= <\int_{-\infty}^{+\infty} |O(\omega;\lambda) - \hat{O}(\omega;\lambda)|^2 d\omega>_\lambda \\ &= \int_{-\infty}^{+\infty} [<|O|^2> + |Y|^2 (|T|^2 <|O|^2> + T^* <O^* N> + T <ON^*> + <|N|^2>) \\ &\quad - (Y^*T^* + YT) <|O|^2> - Y^* <ON^*> - Y <O^* N>] \, d\omega \end{aligned}.$$

By a minimum mean-square error (mmse) criterion: $\delta e = 0$, we can obtain $Y(\omega)$.

Assuming $n(x;\vec{\lambda})$ is additive to $o(x;\vec{\lambda},\vec{\lambda}')$, *i.e.*, $<o^*(x;\vec{\lambda}) n(x';\vec{\lambda}')> = 0$.

By taking Fourier transform,

$$\overset{FT}{=} \quad <O^* N> = \int_{-\infty}^{+\infty} dx \, e^{-j\omega x} \int_{-\infty}^{\infty} dx' \, e^{-j\omega x'} <o(x;\vec{\lambda}) n(x';\vec{\lambda}')> = 0,$$

$$\therefore \quad e = \int_{-\infty}^{+\infty} d\omega \, \{S_o(\omega) + YY^*[|T|^2 S_o(\omega) + S_n(\omega)] - (Y^*T^* + YT) S_o(\omega)\} = \text{minimum}.$$

$$\delta e = 0 \quad \Rightarrow \quad \frac{d}{d\omega}\left(\frac{\partial L}{\partial \dot{Y}^*}\right) = \frac{\partial L}{\partial Y^*} = 0, \text{ where } \dot{Y} = \frac{dY}{d\omega}.$$

Thus, $Y(\omega) = \dfrac{T^*(\omega) S_o(\omega)}{|T(\omega)|^2 S_o(\omega) + S_n(\omega)}$ *and* $e_{\min} = \displaystyle\int_{-\infty}^{\infty} \frac{S_n(\omega) S_o(\omega)}{|T(\omega)|^2 S_o(\omega) + S_n(\omega)} d\omega$.

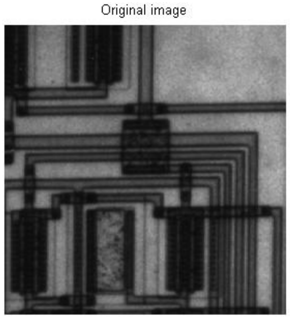
$Y(\omega)$ obtained by this criterion is called **Wiener-Helstrom** filter.

(i) If $S_n \gg S_o \Rightarrow \quad Y(\omega) \simeq 0$. $Y(\omega)$ blanks out data frequencies which are too noisy.

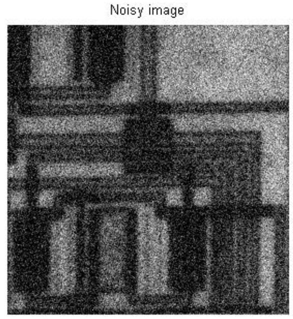
(ii) If $S_n \ll S_o \Rightarrow \quad Y(\omega \simeq \frac{1}{T(\omega)}$, corresponding to an inverse filter.

Therefore, optimal Wiener-Helstrom filter does its job by rejecting high-noise frequency components and accepting low-noise components.

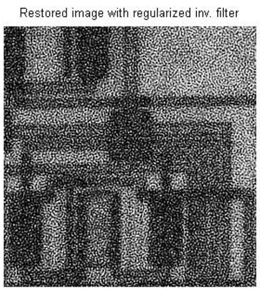
In the following we present the effect of an optimal Wiener-Helstrom filter. The input image was first blurred with a Gaussian filter using sigma=1. The bluured image was then degraded by an additive white noise (20% to that of maximum signal). The blurred and noisy image has pSNR=65.1 relative to the original image. The restored image with regulated direct inversion was shown in the third picture, which yields pSNR=56.3. Much better image with pSNR=77.1 can be restored using an optimal Wiener-Helstrom filter as long as the noise statistics was known. We can construct a Wiener-Helstrom filter without knowing the noise statistics to restore an image with pSNR=73.7, performance matching to that of the optimal Wiener-Helstrom filter.

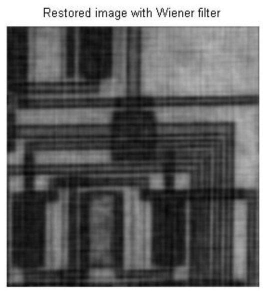


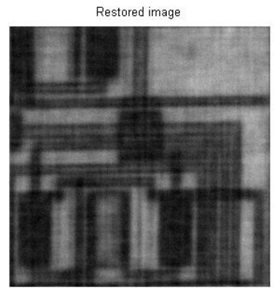
Input Image



Blurred and Degraded by Noise (pSNR=65.1)







Restored by Regulated directed inversion (pSNR=56.3)

Wiener-Helstrom optimal filter (pSNR=77.1)

Estimated Wiener-Helstrom filter (pSNR=73.7)