Lecture 13

Theory of random processes

Part I: Basic concepts

Harrison H. Barrett
University of Arizona

PLAN

Introduce random processes in three parts:

I. Basic concepts

II. Poisson and Gaussian random processes

III. Complex random processes and random waves

Then might go on to talk about speckle, SCIDAR, etc.

Outline for Part I

- Some key points from Lecture 8, Basic Probability Theory

- Linear transformations of random vectors
  - Transformation of the mean and covariance
  - Karhunen-Loève transformations

- Introduction to random processes
  - Averages; moments and variance
  - Single-point and multiple-point PDFs
  - Characteristic functionals
  - Linear transformations
  - Correlation analysis

- Stationarity and quasi-stationarity
  - Karhunen-Loève in the stationary case, power spectral density
  - Local autocovariance and stochastic Wigner distribution function

References

Barrett and Myers, Foundations of Image Science

Chap. 8, Stochastic Properties of Objects and Images

Chap. 11, Poisson Statistics and Photon Counting

Some key points from Lecture 8

An M-dimensional random vector g is a set of M scalar random variables, $\{g_m, m = 1, ..., M\}$. (Think of measurements from some detector array.)

If each g_m can take a continuous range of values, the vector g is described by a probability density function (PDF) $\mathrm{pr}(\mathrm{g}) = \mathrm{pr}(\{g_m\})$, and the mean of g is given by

$$\langle\, \mathrm{g} \,\rangle = \overline{\mathrm{g}} = \{\overline{g}_m\} = \int_{-\infty}^{\infty} dg_1 \int_{-\infty}^{\infty} dg_2 \cdots \int_{-\infty}^{\infty} dg_M \, \mathrm{g} \; \mathrm{pr}(\mathrm{g}) = \int_{\infty} d^M g \; \mathrm{g} \; \mathrm{pr}(\mathrm{g}) \,.$$

If each g_m can take on only integer values, the vector g is described by a probability (not PDF) $\mathrm{Pr}(\mathrm{g}) = \mathrm{Pr}(\{g_m\})$, and the mean of g is given by

$$\langle\, \mathrm{g} \,\rangle = \overline{\mathrm{g}} = \{\overline{g}_m\} = \sum_{g_1=0}^{\infty} \sum_{g_2=0}^{\infty} \cdots \sum_{g_M=0}^{\infty} \mathrm{g} \; \mathrm{Pr}(\mathrm{g}) \,.$$

(Note that the sum here is over each of the g_m, not over m.)

Covariance matrix

The *covariance matrix* is a generalization of the variance to random vectors.

For an MD random vector $\mathbf{g}$, the covariance matrix $\mathbf{K}$ is an $M \times M$ matrix with elements given by

$$K_{ij} = \left\langle (g_i - \overline{g}_i)(g_j - \overline{g}_j)^* \right\rangle \,, \tag{8.16}$$

where the asterisk indicates complex conjugate, allowing for the possibility that components of $\mathbf{g}$ might be complex. It follows from this definition that $\mathbf{K}$ is Hermitian, *i.e.*, $K_{ij} = K^*_{ji}$.

In outer product form,

$$\mathbf{K} = \left\langle \, [\mathbf{g} - \overline{\mathbf{g}}] \, [\mathbf{g} - \overline{\mathbf{g}}]^\dagger \, \right\rangle \,.$$

Any random variable covaries with itself. The diagonal elements of the covariance matrix are the variances of the components:

$$K_{jj} = \mathrm{Var}\{g_j\} \,. \tag{8.18}$$

Characteristic functions (from Lecture 8)

For any random variable or vector, the *characteristic function* is the expectation of the appropriate Fourier kernel:

$$\psi(\xi) \equiv \left\langle\, e^{-2\pi i \xi x} \,\right\rangle .$$

If x is scalar and real-valued, then

$$\psi(\xi) = \int_{-\infty}^{\infty} dx \; \mathrm{pr}(x)\, e^{-2\pi i \xi x} , \tag{C.53}$$

and the PDF and characteristic function form a Fourier transform pair:

$$\mathrm{pr}(x) = \int_{-\infty}^{\infty} d\xi \; \psi(\xi)\, e^{2\pi i \xi x} . \tag{C.54}$$

Caution: Do not confuse ξ with a spatial frequency.

Moments of the random variable x can be derived through differentiation of $\psi(\xi)$:

$$\left\langle\, x^k \,\right\rangle = (-2\pi i)^{-k} \left. \frac{\partial^k}{\partial \xi^k} \psi(\xi) \right|_{\xi=0} . \tag{C.55}$$

Characteristic function of a real random vector

For a real $M \times 1$ random vector g (column vector), the characteristic function is defined as

$$\psi_{\mathrm{g}}(\boldsymbol{\xi}) = \left\langle \exp(-2\pi i \boldsymbol{\xi}^t \mathrm{g}) \right\rangle \,, \tag{8.26}$$

where $\boldsymbol{\xi}^t$ is a real $1 \times M$ vector.

For the case of a continuous-valued random vector, $\psi_{\mathrm{g}}(\boldsymbol{\xi})$ can be written as

$$\psi_{\mathrm{g}}(\boldsymbol{\xi}) = \int_{\infty} d^M g \; \mathrm{pr}(\mathrm{g}) \exp(-2\pi i \boldsymbol{\xi}^t \mathrm{g}) \,. \tag{8.27}$$

This integral is the MD Fourier transform of the PDF, so

$$\mathrm{pr}(\mathrm{g}) = \int_{\infty} d^M \xi \; \psi_{\mathrm{g}}(\boldsymbol{\xi}) \exp(2\pi i \boldsymbol{\xi}^t \mathrm{g}) \,. \tag{8.28}$$

Linear transformations of random vectors

If the random vector $\mathbf{g}$ is generated as the output of a linear filter acting on the random vector $\mathbf{f}$, we can characterize the linear transformation by an $M \times N$ matrix $\mathbf{H}$. Then we can write the $M \times 1$ output vector $\mathbf{g}$ in terms of the $N \times 1$ input vector $\mathbf{f}$ as

$$\mathbf{g} = \mathbf{H}\mathbf{f}. \tag{8.40}$$

From the linearity of the expectation operator, we have immediately for the mean of $\mathbf{g}$,

$$\overline{g} = \langle \mathbf{g} \rangle = \langle \mathbf{H}\mathbf{f} \rangle = \mathbf{H} \langle \mathbf{f} \rangle = \mathbf{H}\overline{\mathbf{f}}. \tag{8.49}$$

The covariance matrix of $\mathbf{g}$ is found as

$$\mathbf{K_g} = \left\langle \Delta\mathbf{g}\Delta\mathbf{g}^\dagger \right\rangle = \left\langle (\mathbf{H}\mathbf{f} - \mathbf{H}\overline{\mathbf{f}})(\mathbf{H}\mathbf{f} - \mathbf{H}\overline{\mathbf{f}})^\dagger \right\rangle = \mathbf{H} \left\langle \Delta\mathbf{f}\,\Delta\mathbf{f}^\dagger \right\rangle \mathbf{H}^\dagger = \mathbf{H}\mathbf{K_f}\mathbf{H}^\dagger, \tag{8.50}$$

where $\Delta\mathbf{f} \equiv \mathbf{f} - \overline{\mathbf{f}}$.

Linear transformation of the characteristic function

Transformation of the PDF is tricky, but the characteristic function is easy:

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \left\langle \exp(-2\pi i \boldsymbol{\xi}^t \mathbf{H}\mathbf{f}) \right\rangle = \left\langle \exp\left[-2\pi i (\mathbf{H}^t\boldsymbol{\xi})^t \mathbf{f}\right] \right\rangle , \tag{8.42}$$

where the last step has used the definition of the adjoint, (1.39). Thus

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \psi_{\mathbf{f}}(\mathbf{H}^t\boldsymbol{\xi}) , \tag{8.43}$$

so knowledge of $\psi_{\mathbf{f}}$ and $\mathbf{H}$ immediately gives $\psi_{\mathbf{g}}$.

The PDF on $\mathbf{g}$ can in principle be found by taking an inverse MD Fourier transform of (8.43). Formally, we can write

$$\mathsf{pr}(\mathbf{g}) = \int_\infty d^M\xi \; \psi_{\mathbf{f}}(\mathbf{H}^t\boldsymbol{\xi}) \exp(2\pi i \boldsymbol{\xi}^t \mathbf{g}) , \tag{8.44}$$

but in practice the integral might not be easy. The problem is that we are integrating a function of an ND vector over an MD space.

Eigenanalysis of the covariance matrix

A covariance matrix is Hermitian, so the eigenvalues are real, and the eigenvectors can be chosen to form a complete, orthonormal set.

Let $\mathbf{K_g}$ be the $M \times M$ covariance matrix for a random vector $\mathbf{g}$. The eigenvalue equation for this matrix is

$$\mathbf{K_g}\boldsymbol{\phi}_m = \mu_m \boldsymbol{\phi}_m \,, \qquad m = 1, ..., M \,, \tag{8.51}$$

where $\boldsymbol{\phi}_m$ is an $M \times 1$ eigenvector and μ_m is the corresponding eigenvalue.

Order eigenvalues as usual:

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_R > 0 \,, \tag{8.52}$$

where R is the rank.

Orthonormality and completeness:

$$\boldsymbol{\phi}_m^\dagger \boldsymbol{\phi}_n = \delta_{mn} \,, \qquad \sum_{m=1}^{M} \boldsymbol{\phi}_m \boldsymbol{\phi}_m^\dagger = \mathbf{I} \,.$$

Karhunen-Loève transformation

Since the eigenvectors of a Hermitian operator form a complete, orthonormal set in the relevant space, any $M \times 1$ vector $\mathbf{g}$ can be expressed as

$$\mathbf{g} = \sum_{m=1}^{M} \beta_m \boldsymbol{\phi}_m \,, \tag{8.58}$$

where the coefficients are given by

$$\beta_m = \boldsymbol{\phi}_m^\dagger \mathbf{g} \,. \tag{8.59}$$

If $\mathbf{g}$ is a random vector and the vectors $\{\boldsymbol{\phi}_m\}$ are eigenvectors of its covariance matrix, then the coefficients $\{\beta_m\}$ are uncorrelated random variables:

$$\langle \Delta\beta_n \Delta\beta_m^* \rangle = \left\langle \left[\boldsymbol{\phi}_n^\dagger \Delta\mathbf{g}\right] \left[\boldsymbol{\phi}_m^\dagger \Delta\mathbf{g}\right]^* \right\rangle = \left\langle \boldsymbol{\phi}_n^\dagger \Delta\mathbf{g} \Delta\mathbf{g}^\dagger \boldsymbol{\phi}_m \right\rangle$$

$$= \boldsymbol{\phi}_n^\dagger \left\langle \Delta\mathbf{g} \Delta\mathbf{g}^\dagger \right\rangle \boldsymbol{\phi}_m = \boldsymbol{\phi}_n^\dagger \mathbf{K}_\mathbf{g} \boldsymbol{\phi}_m = \mu_m \boldsymbol{\phi}_n^\dagger \boldsymbol{\phi}_m = \mu_m \, \delta_{nm} \,, \tag{8.61}$$

Uses of KL transformation

- Efficient representation of random vectors (minimum EMSE)
- Diagonal representation of covariance matrix
- Spectral representation of covariance and its inverse
- Prewhitening operator (useful in signal detection)

Intro to random processes

Simple definition: A random process is a random variable that varies with space and/or time. More formally, a spatial random process is a function of two variables, $\mathbf{r}$ and ζ. Depending on the context, $f(\mathbf{r}, \zeta)$ can refer to:

1. The family of spatial functions, referred to as the ensemble; in this case, $\mathbf{r}$ and ζ are variables (e.g., all possible images)

2. A single realization or sample of the spatial functions; in this case, $\mathbf{r}$ is variable and ζ is fixed (one particular image)

3. The random variable at a single point; in this case, $\mathbf{r}$ is fixed and ζ is variable (value at one point in all possible images)

4. A single number; in this case, $\mathbf{r}$ is fixed and ζ is fixed (value at one point in one image)

Notational quirks: Usually drop the ζ designator and hope the differences above will be clear by context

Classifications of random processes

- Spatial vs. temporal (or spatiotemporal)
- Scalar vs. vector valued
- Continuous vs. discrete valued
- Complex or real
- Kind of function:

  Square-integrable (finite "energy")

  Finite power (energy per unit time)

  Generalized (e.g., delta functions)

Averages

Consider a scalar-valued continuous random process. For fixed $\mathbf{r}$, $f(\mathbf{r})$ is simply a random variable (interpretation 3), and its expectation is defined just as for any other random variable;

$$\mathrm{E}\{f(\mathbf{r})\} = \langle f(\mathbf{r}) \rangle = \bar{f}(\mathbf{r}) = \int_{-\infty}^{\infty} df(\mathbf{r}) \; f(\mathbf{r}) \, \mathrm{pr}[f(\mathbf{r})] \, . \tag{8.71}$$

Computation of this expectation requires only the univariate PDF $\mathrm{pr}[f(\mathbf{r})]$.

N.B. The integral is over $f(\mathbf{r})$, not $\mathbf{r}$; result can still be a function of $\mathbf{r}$.

Moments and variance

Moments of $f(\mathbf{r})$ are defined easily. For example, the j^{th} moment is given by

$$\left\langle [f(\mathbf{r})]^j \right\rangle = \int_{-\infty}^{\infty} df(\mathbf{r})\ [f(\mathbf{r})]^j \,\mathrm{pr}[f(\mathbf{r})]\,. \tag{8.72}$$

Again, the integral is over $f(\mathbf{r})$, not over $\mathbf{r}$; The resultant, $\left\langle [f(\mathbf{r})]^j \right\rangle$, can still be a function of $\mathbf{r}$.

Having defined moments, we can also define the variance of a random process. In the general complex case, the variance is given by

$$\mathrm{Var}\{f(\mathbf{r})\} = \mathrm{E}\left\{|f(\mathbf{r})| - |\mathrm{E}\{f(\mathbf{r})\}|^2\right\} = \mathrm{E}\left\{|f(\mathbf{r})|^2\right\} - |\mathrm{E}\{f(\mathbf{r})\}|^2$$

$$= \int_{-\infty}^{\infty} df(\mathbf{r})\ |f(\mathbf{r})|^2 \,\mathrm{pr}[f(\mathbf{r})] - \left| \int_{-\infty}^{\infty} df(\mathbf{r})\ f(\mathbf{r}) \,\mathrm{pr}[f(\mathbf{r})] \right|^2 . \tag{8.73}$$

Still need only the univariate density $\mathrm{pr}[f(\mathbf{r})]$

Multiple-point expectations

$$\langle\, f(\mathbf{r}_1) f(\mathbf{r}_2)\,\rangle = \int_{-\infty}^{\infty} df(\mathbf{r}_1) \int_{-\infty}^{\infty} df(\mathbf{r}_2)\;\; f(\mathbf{r}_1)\, f(\mathbf{r}_2)\, \mathrm{pr}[f(\mathbf{r}_1), f(\mathbf{r}_2)]\,. \tag{8.74}$$

Here, $f(\mathbf{r}_1)$ and $f(\mathbf{r}_2)$ must be regarded as two *distinct* random variables and $\mathrm{pr}[f(\mathbf{r}_1), f(\mathbf{r}_2)]$ is their joint density. Only in very special circumstances will it be possible to write $\mathrm{pr}[f(\mathbf{r}_1), f(\mathbf{r}_2)]$ as $\mathrm{pr}[f(\mathbf{r}_1)]\,\mathrm{pr}[f(\mathbf{r}_2)]$.

A general two-point moment is defined by

$$\langle\, [f(\mathbf{r}_1)]^m\, [f(\mathbf{r}_2)]^n\,\rangle =$$

$$\int_{-\infty}^{\infty} df(\mathbf{r}_1) \int_{-\infty}^{\infty} df(\mathbf{r}_2)\; [f(\mathbf{r}_1)]^m\, [f(\mathbf{r}_2)]^n\, \mathrm{pr}[f(\mathbf{r}_1), f(\mathbf{r}_2)]\,. \tag{8.75}$$

Any moment involving the K points $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_K$ can be computed if $\mathrm{pr}[f(\mathbf{r}_1), f(\mathbf{r}_2), ..., f(\mathbf{r}_K)]$ is known. If this K-fold joint density is known for all values of each of the $\mathbf{r}_k$, the process is said to be *fully characterized* to order K.

PDFs and characteristic functionals

A spatial random process $f(\mathbf{r})$ is an infinite-dimensional vector in a Hilbert space – *if* every sample function is square-integrable.

Thus, at best, an infinite-dimensional PDF would be needed for a complete characterization: very tricky mathematically.

BUT, an infinite-dimensional characteristic *functional* can always be defined – and often calculated explicitly!

ALL statistical properties of a random process are contained in its characteristic functional.

Characteristic functionals – definition

Recall the definition of the characteristic *function* for a MD real random vector:

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \left\langle \exp(-2\pi i \boldsymbol{\xi}^t \mathbf{g}) \right\rangle , \tag{8.26}$$

Here, $\boldsymbol{\xi}$ is a real $M \times 1$ vector, and $\boldsymbol{\xi}^t \mathbf{g}$ denotes a scalar product.

In the case of a random process $f(\mathbf{r})$, each sample function corresponds to a vector $\mathbf{f}$ in an infinite-dimensional Hilbert space, so the frequency vector $\boldsymbol{\xi}$ in (8.26) must be replaced by an infinite-dimensional vector $\mathbf{s}$ in the same Hilbert space as $\mathbf{f}$. That means that $\mathbf{s}$ describes a function $s(\mathbf{r})$, so the characteristic function becomes a characteristic *functional* $\Psi_{\mathbf{f}}\{s(\mathbf{r})\}$ or $\Psi_{\mathbf{f}}(\mathbf{s})$ for short. It is defined by

$$\Psi_{\mathbf{f}}(\mathbf{s}) = \langle \exp[-2\pi i (\mathbf{s}, \mathbf{f})] \rangle , \tag{8.94}$$

where $(\mathbf{s}, \mathbf{f})$ is the usual $\mathbb{L}_2$ scalar product.

Mathematical detail: generalized random processes

Suppose $f(\mathbf{r})$ is a generalized function, say a delta function at a random location:

$$f(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}_0)\,, \qquad \mathbf{r}_0 \text{ random}$$

(Note that $\mathbf{r_0}$ plays the role of ζ here.)

The sample functions are not in $\mathbb{L}_2$, but we can still define the scalar product we need in the characteristic functional:

$$(\mathbf{s}, \mathbf{f}) \equiv s(\mathbf{r}_0)\,.$$

Thus, if we know the statistics of $\mathbf{r}_0$, we can compute the characteristic functional of this generalized random process (and will do so in the next lecture).

Linear transformations of random processes

Consider the familiar form of a general linear mapping, $\mathrm{g} = \boldsymbol{\mathcal{H}}\mathrm{f}$. If $\boldsymbol{\mathcal{H}}$ is a CD mapping, then

$$g_m = \int d^q r \, h_m(\mathbf{r}) \, f(\mathbf{r}) \, .$$

If f denotes a random process, this mapping defines a random vector. Specifically, if one sample function of $f(\mathbf{r})$ is denoted $f(\mathbf{r}, \zeta)$, then $\mathrm{g} = \boldsymbol{\mathcal{H}}\mathrm{f}$ is to be interpreted as

$$g_m(\zeta) = \int d^q r \, h_m(\mathbf{r}) \, f(\mathbf{r}, \zeta) \, .$$

A similar interpretation applies to integral transforms. For example, the Fourier transform of a random process is a new random process obtained by taking the Fourier transform of each sample function.

Linear transformation of the characteristic functional

Back to the definition, with a different notation for the scalar product:

$$\Psi_{\mathbf{f}}(\mathbf{s}) = \left\langle \exp[-2\pi i\, \mathbf{s}^\dagger \mathbf{f}] \right\rangle , \tag{8.94}$$

For now, assume $\mathbf{s}$ and $\mathbf{f}$ are real.

Consider the general CD mapping $\mathbf{g} = \boldsymbol{\mathcal{H}}\mathbf{f}$. The characteristic *function* of the random vector $\mathbf{g}$ is given by

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) \equiv \left\langle \exp[-2\pi i\, \boldsymbol{\xi}^\dagger \mathbf{g}] \right\rangle = \left\langle \exp[-2\pi i\, \boldsymbol{\xi}^\dagger (\boldsymbol{\mathcal{H}}\mathbf{f})] \right\rangle = \left\langle \exp[-2\pi i\, (\boldsymbol{\mathcal{H}}^\dagger \boldsymbol{\xi})^\dagger \mathbf{f}] \right\rangle ,$$

where the last step follows from the definition of the adjoint. Thus

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \Psi_{\mathbf{f}}(\boldsymbol{\mathcal{H}}^\dagger \boldsymbol{\xi}) . \tag{8.96}$$

Linear transformation of the characteristic functional (cont.)

From last slide, if $\mathbf{g} = \boldsymbol{\mathcal{H}}\mathbf{f}$, where $\boldsymbol{\mathcal{H}}$ is any linear transformation, then

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \Psi_{\mathbf{f}}(\boldsymbol{\mathcal{H}}^{\dagger}\boldsymbol{\xi}\,)\,. \tag{8.96}$$

If $\boldsymbol{\mathcal{H}}$ is specifically a CD mapping with kernel $h_m(\mathbf{r})$, then

$$[\boldsymbol{\mathcal{H}}^{\dagger}\boldsymbol{\xi}\,](\mathbf{r}) = \sum_{m=1}^{M} \xi_m\, h_m(\mathbf{r})\,.$$

Very powerful result: If we know the characteristic *functional* for $\mathbf{f}$, we immediately have the characteristic *function* for $\boldsymbol{\mathcal{H}}\mathbf{f}$, and therefore we know *all statistical properties* of the random vector (e.g., image) $\mathbf{g}$.

Correlation analysis

The autocorrelation function $R(\mathbf{r}_1, \mathbf{r}_2)$ of a random process $f(\mathbf{r})$ is defined by

$$R(\mathbf{r}_1, \mathbf{r}_2) = \langle f(\mathbf{r}_1)\, f^*(\mathbf{r}_2) \rangle \ , \tag{8.97}$$

which is the two-point expectation defined in (8.74), with the minor modification of the complex conjugate on the second factor [irrelevant if $f(\mathbf{r})$ is real].

The autocovariance function $K(\mathbf{r}_1, \mathbf{r}_2)$ is defined by

$$\begin{aligned} K(\mathbf{r}_1, \mathbf{r}_2) &= \big\langle [f(\mathbf{r}_1) - \langle f(\mathbf{r}_1) \rangle] \left[f^*(\mathbf{r}_2) - \langle f^*(\mathbf{r}_2) \rangle \right] \big\rangle \\ &= R(\mathbf{r}_1, \mathbf{r}_2) - \bar{f}(\mathbf{r}_1)\, \bar{f}^*(\mathbf{r}_2) \, . \end{aligned} \tag{8.98}$$

The autocovariance function is thus the two-point moment that is the generalization of the variance; it reduces to the variance when $\mathbf{r}_2 = \mathbf{r}_1 = \mathbf{r}$, *i.e.*,

$$K(\mathbf{r}, \mathbf{r}) = R(\mathbf{r}, \mathbf{r}) - |\bar{f}(\mathbf{r})|^2 = \mathrm{Var}\{ f(\mathbf{r}) \} \, . \tag{8.99}$$

Temporal stationarity

A temporal random process $f(t)$ is said to be stationary in the strict sense if, for any K, its K-point PDF $\mathrm{pr}[f(t_1), \cdots, f(t_K)]$ is such that

$$\mathrm{pr}[f(t_1), \cdots, f(t_K)] = \mathrm{pr}[f(t_1+\tau), \cdots, f(t_K+\tau)] \tag{8.107}$$

for any τ. In particular, this requires that the single-point density function be independent of time,

$$\mathrm{pr}[f(t)] = \mathrm{pr}[f(t+\tau)]\,, \tag{8.108}$$

and therefore the mean of the random process is also independent of time,

$$\langle f(t)\rangle = \langle f(t+\tau)\rangle\ . \tag{8.109}$$

Similarly, the two-point density function must be independent of time,

$$\mathrm{pr}[f(t_1), f(t_2)] = \mathrm{pr}[f(t_1+\tau), f(t_2+\tau)]\,, \tag{8.110}$$

and so the autocorrelation function $R(t_1, t_2)$, is also independent of the absolute time and can be written as

$$R(t_1,t_2) = \langle f(t_1)\, f^*(t_2)\rangle = \langle f(t_1+\tau)\, f^*(t_2+\tau)\rangle = R(t_1{-}t_2) = R(\Delta t)\,,$$

where $\Delta t \equiv t_1 - t_2$.

Wide-sense temporal stationarity

As just defined, stationarity in the strict sense requires that all K-point PDFs be independent of the time origin.

Stationarity in the wide (or loose) sense requires only that the mean and autocorrelation have no preferred origin:

$$\langle f(t) \rangle = \text{constant} \,;$$

$$R(t_1, t_2) = R(t_1 - t_2) \,.$$

A stationary random process is said to be *ergodic in the mean* if a time average gives the (constant) ensemble average.

Spatial stationarity?

Strict-sense and wide-sense stationarity can be defined for spatial random processes just as for temporal ones, but the conditions are much harder to satisfy.

- Electrical noise from a resistor is temporally stationary if T = const.
- Poisson noise in an image is stationary only for a uniform grey scene

My advice: Never assume spatial stationarity in imaging problems

Possible exception: Well, you might get away with assuming that atmospheric turbulence is stationary.

Quasi-stationarity

To reiterate the advice, you should always write a spatial autocorrelation function as $R(\mathbf{r}_1, \mathbf{r}_2)$ and resist the temptation to assume that $R(\mathbf{r}_1, \mathbf{r}_2) = R(\mathbf{r}_1 - \mathbf{r}_2)$.

That means that the autocorrelation of a 2D continuous image is a function of 4 variables, not 2.

Can we *ever* simplify and use 2D functions?

Consider a ground glass illuminated with a Gaussian laser beam. If the beam width is large compared to the grinding grit, then it is valid to assume that the field emerging from the ground glass is *quasi-stationary*:

$$R(\mathbf{r}_1, \mathbf{r}_2) = a(\Delta\mathbf{r})\, b(\mathbf{r}_0)\,, \tag{8.119}$$

where

$$\mathbf{r}_0 = \tfrac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2)\,, \qquad \Delta\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2\,. \tag{8.120}$$

Now have a product of two 2D functions rather than one 4D one.

"But, but, but ...", you sputter, "if I don't assume stationarity, how can I use Fourier methods? What will I do without my power spectrum?"

"Have no fear!", say I.

Spectral analysis

For a stationary temporal random process, we can follow Khinchin and define the *power spectrum* or *power spectral density* as

$$S(\nu) = \mathcal{F}\{R(\Delta t)\} = \int_{-\infty}^{\infty} d\Delta t \ \langle f(t+\Delta t)\, f^*(t)\rangle \exp(-2\pi i \nu \Delta t)\,. \tag{8.133}$$

This equation is often called the Wiener-Khinchin theorem, but I like to think of it as a definition. (Besides, Wiener didn't consider ensemble averages at all.)

What to do if the process is not stationary? Use the same equation anyway! (You just can't assume that the lefthand side is a function only of ν.)

And if you do that, you can use spatial as well as temporal random processes.

Stochastic Wigner distribution function

Rewrite the spatial autocorrelation function in symmetrized form:

$$R(\mathbf{r}_1, \mathbf{r}_2) \rightarrow R(\mathbf{r}_0, \Delta\mathbf{r})$$

where

$$\mathbf{r}_0 = \tfrac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2)\,, \qquad \Delta\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2\,. \tag{8.120}$$

There is no loss of generality if we don't assume the function factors; $R(\mathbf{r}_0, \Delta\mathbf{r})$ is still a general function of four variables if $\mathbf{r}$ is 2D.

Now take a FT with respect to $\Delta\mathbf{r}$ only, and define the *stochastic Wigner distribution function* as

$$W_f(\mathbf{r}_0, \boldsymbol{\rho}) = \int_\infty d^q\Delta r \;\left\langle f(\mathbf{r}_0 + \tfrac{1}{2}\Delta\mathbf{r})\, f^*(\mathbf{r}_0 - \tfrac{1}{2}\Delta\mathbf{r})\right\rangle \exp(-2\pi i\boldsymbol{\rho}\cdot\Delta\mathbf{r})\,. \tag{8.140}$$

If (somehow), $f(\mathbf{r})$ really was stationary, $W_f(\mathbf{r}_0, \boldsymbol{\rho})$ would be independent of $\mathbf{r}_0$ and equal to the spatial power spectrum.

Stochastic WDF for the quasi-stationary case

We defined quasi-stationarity by

$$R(\mathbf{r}_1, \mathbf{r}_2) = a(\Delta\mathbf{r})\, b(\mathbf{r}_0)\,. \tag{8.119}$$

For this model, the stochastic WDF is given by

$$W_f(\mathbf{r}_0, \boldsymbol{\rho}) = b(\mathbf{r}_0) \int_\infty d^q\Delta r\ a(\Delta\mathbf{r}) \exp(-2\pi i \boldsymbol{\rho}\cdot\Delta\mathbf{r}) = b(\mathbf{r}_0)\, A(\boldsymbol{\rho})\,. \tag{8.142}$$

Thus the WDF is just the Fourier transform of the short-range part of the autocorrelation function, modulated by the shift-variant strength of the slowly varying component at $\mathbf{r}_0$.

Spectral analysis from the Karhunen-Loève perspective

For discrete random vectors, KL transformation is equivalent to diagonalizing the covariance matrix.

The components of the random vector are uncorrelated in the KL basis.

For random processes, we want to diagonalize the autocorrelation operator, i.e., the integral operator whose kernel is the autocorrelation function.

For stationary random processes, the autocorrelation operator is a convolution, and we know that convolutions are diagonalized by Fourier transformation.

Hence, KL $\Leftrightarrow$ Fourier for stationary processes, and the components of the random process are uncorrelated in the Fourier basis:

$$\left\langle F(\nu)\, F^*(\nu') \right\rangle = S(\nu)\, \delta(\nu - \nu')\,.$$

KL for nonstationary random processes

In general, Karhunen-Loève $\Leftrightarrow$ eigenalysis of the autocorrelation operator.

The autocorreleation operator acting on an arbitrary function $u(\mathbf{r})$ is defined by

$$[\boldsymbol{\mathcal{R}}\mathbf{u}](\mathbf{r}) = \int_{\infty} d^q r' \, R(\mathbf{r}, \mathbf{r}') \, u(\mathbf{r}') \, . \tag{8.162}$$

For most spatial random processes (but *not* for stationary ones), this operator is compact, so it has a discrete eigenvalue spectrum. The eigenvalue problem is

$$\boldsymbol{\mathcal{R}}\phi_n(\mathbf{r}) = \mu_n \phi_n(\mathbf{r}) \, . \tag{8.167}$$

Any sample function of the random process can be expanded as a *sum* of these eigenfunctions, and the coefficients will be uncorrelated.

Preview of coming attractions

All of this machinery will be put to work in the coming lectures. I will:

- Show why Gaussian random processes are so ubiquitous
- Demonstrate the wonderful properties of Gaussian random processes
- Surprise and amaze you with the peculiarities of complex Gaussians
- Derive all statistical properties of nonstationary Poisson processes
- Show you that you have been living in the KL domain all your life
- Show how transformations of char. fcnals. can be put to work.