Lecture 14

Theory of random processes
Part II: Real Gaussian random processes

Harrison H. Barrett
University of Arizona

OUTLINE

- Review of Lecture 13

- Real Gaussian random processes

  Review of Gaussian random variables and vectors

  Gaussian random processes: Definition and basic properties

  Characteristic functions and functionals

  Linear transformations

  From Gaussian process to Gaussian vector or variable

- Gaussian noise in imaging

  Origins

  Statistics of $\mathbf{g} = \boldsymbol{\mathcal{H}}\mathbf{f} + \mathbf{n}$

Review: Definition and characterization of random processes

Random processes are random functions of some independent variable, usually spatial position and/or time.

The statistics of random processes can be specified by:

- First- and second-order statistics
  Mean: $\overline{f}(\mathbf{r}) \equiv \langle\, f(\mathbf{r}) \,\rangle$
  Variance: $\sigma^2(\mathbf{r}) \equiv \langle\, [\Delta f(\mathbf{r})]^2 \,\rangle$, where $\Delta f(\mathbf{r}) \equiv f(\mathbf{r}) - \overline{f}(\mathbf{r})$
  Autocorrelation: $\langle\, f(\mathbf{r})\, f(\mathbf{r}') \,\rangle$
  Autocovariance: $\langle\, \Delta f(\mathbf{r})\, \Delta f(\mathbf{r}') \,\rangle$

- Single-point PDF: pr$[f(\mathbf{r})]$

- Multi-point PDF: pr$[f(\mathbf{r}_1), f(\mathbf{r}_2), ..., f(\mathbf{r}_K)]$

- Characteristic functionals

Characteristic functionals – definition

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

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

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

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

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

Linear transformations of random processes

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

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

If $\mathbf{f}$ 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 $\mathbf{g} = \boldsymbol{\mathcal{H}}\mathbf{f}$ 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:

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

For now, assume $\mathbf{s}$ and $\mathbf{f}$ are real, so $\mathbf{s}^{\dagger}\mathbf{f} = \int_{\infty} d^q r\, s(\mathbf{r})\, f(\mathbf{r})$.

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

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

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

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

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

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

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

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

Relation to variance:

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

Eigenanalysis of the covariance matrix

A covariance matrix is Hermitian, so:

- Its eigenvalues are real
- Its eigenvectors can be chosen to form a complete, orthonormal set.
- It can be diagonalized by a unitary transformation, called
  the Karhunen-Loève transformation

Key advantage::

- The expansion coefficients are uncorrelated in the KL representation

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

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

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

Strict-sense stationarity implies wide sense, but not vice versa (*except* for Gaussian random processes)

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.

Temporal processes are stationary if the basic processes that generate them are independent of time

- Example: electrical noise from a resistor at const. temp.

But images are inherently functions of position, so any image-dependent randomness must be non-stationary.

- Example 1: Poisson noise (variance $=$ mean $\neq$ constant)

- Example 2: Object randomness

KL and power spectrum

For WSS random processes, the autocorrelation operator is a convolution, so its eigenfunctions are complex exponentials (Fourier kernels)

For WSS: KL $\Leftrightarrow$ Fourier

Any sample function of a stationary temporal random process can be expanded in terms of its Fourier transform,

$$f(t) = \int_{-\infty}^{\infty} d\nu \; F(\nu) \, \mathrm{exp}(2\pi i \nu t) \, ,$$

and the "expansion coefficients" (Fourier transform values) are delta-correlated:

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

where

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

The function $S(\nu)$ is called the *power spectrum* or *power spectral density*, and (8.133) is called the *Wiener-Khinchin theorem*.

Dealing with nonstationarity: Stochastic Wigner distribution function

For a WSS spatial random process, the *power spectrum* or *power spectral density* can be defined in a symmetrized form as

$$S(\boldsymbol{\rho}) = \boldsymbol{\mathcal{F}}\{R(\mathbf{\Delta r})\} = \int_{\infty} d^q\Delta r \ \left\langle f(\mathbf{r}_0 + \tfrac{1}{2}\mathbf{\Delta r})\, \mathrm{f}^*(\mathbf{r}_0 - \tfrac{1}{2}\mathbf{\Delta r})\right\rangle \exp(-2\pi i\boldsymbol{\rho}\cdot\mathbf{\Delta r})\,. \tag{8.133}$$

The *stochastic Wigner distribution function* is defined as

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

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.

Univariate and multivariate normal or Gaussian PDFs

From Lecture 8, the standard form of the normal PDF is

$$\text{pr}(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left[-\frac{(x-\overline{x})^2}{2\sigma^2}\right] . \tag{C.108}$$

This is a two-parameter PDF, with $\overline{x}$ being the mean and σ^2 being the variance, as the notation implies.

The corresponding multivariate form is

$$\text{pr}(\mathbf{g}) = \left[(2\pi)^M \det(\mathbf{K})\right]^{-1/2} \exp\left[-\tfrac{1}{2}(\mathbf{g}-\overline{\mathbf{g}})^t \mathbf{K}^{-1}(\mathbf{g}-\overline{\mathbf{g}})\right] , \tag{8.185}$$

where $\overline{\mathbf{g}}$ is the mean vector and $\mathbf{K}$ is the covariance matrix of $\mathbf{g}$.

Comments on the multivariate normal law

$$\mathrm{pr}(\mathbf{g}) = \left[(2\pi)^M \det(\mathbf{K})\right]^{-1/2} \exp\left[-\tfrac{1}{2}(\mathbf{g}-\overline{\mathbf{g}})^t\,\mathbf{K}^{-1}(\mathbf{g}-\overline{\mathbf{g}})\right]\,, \qquad (8.185)$$

• PDF is fully specified by mean vector and covariance matrix. Therefore all statistical properties are determined by $\overline{\mathbf{g}}$ and $\mathbf{K}$.

• Exponent is a scalar quadratic form:

$$(\mathbf{g}-\overline{\mathbf{g}})^t\,\mathbf{K}^{-1}(\mathbf{g}-\overline{\mathbf{g}}) = \sum_{m=1}^{M}\sum_{m'=1}^{M}(g_m-\overline{g}_m)\,[\mathbf{K}^{-1}]_{mm'}\,(g_{m'}-\overline{g}_{m'})$$

• Normalization requires determinant (easy with KL representation).

• Central limit theorem: The PDF of a sum of N independent $M\times 1$ random vectors of (almost) arbitrary PDF tends to a multivariate normal as $N\to\infty$.

Characteristic function for multivariate Gaussian

If

$$\mathrm{pr}(\mathbf{g}) = \left[(2\pi)^M \det(\mathbf{K})\right]^{-1/2} \exp\left[-\tfrac{1}{2}(\mathbf{g}-\overline{\mathbf{g}})^t\, \mathbf{K}^{-1}(\mathbf{g}-\overline{\mathbf{g}})\right] , \qquad (8.185)$$

then

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \boldsymbol{\mathcal{F}}_M\left\{\mathrm{pr}(\mathbf{g})\right\} = \exp\left(-2\pi i \boldsymbol{\xi}^t \overline{\mathbf{g}}\right) \exp\left(-2\pi^2 \boldsymbol{\xi}^t \mathbf{K} \boldsymbol{\xi}\right) . \qquad (8.196)$$

Comments:

- For $\overline{g} = \mathbf{0}$, $\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \exp(-2\pi^2 \boldsymbol{\xi}^t \mathbf{K} \boldsymbol{\xi})$ (FT of Gaussian is Gaussian)
- Get $\mathbf{K}$ instead of $\mathbf{K}^{-1}$ in characteristic function
- Factor $\exp(-2\pi i \boldsymbol{\xi}^t \overline{\mathbf{g}})$ comes from shift theorem
- No normalizing factor or determinant needed, $\psi_{\mathbf{g}}(\mathbf{0}) = 1$
- Dimension M does not appear in characteristic function

Moments

We can use the characteristic function given in (8.196) to determine the moments of a multivariate normal random vector. By differentiating, we find $\langle \mathbf{g} \rangle = \overline{\mathbf{g}}$ and $\left\langle \mathbf{g}\, \mathbf{g}^t \right\rangle = \mathbf{K} + \overline{\mathbf{g}}\overline{\mathbf{g}}^t$. If $\overline{\mathbf{g}} = 0$, then $\left\langle \mathbf{g}\, \mathbf{g}^t \right\rangle = \mathbf{K}$.

All odd moments are zero for $\overline{\mathbf{g}} = 0$, and all even moments are expressible in terms of $\mathbf{K}$.

We frequently need fourth moments of the form $\left\langle g_i g_j g_k g_l \right\rangle$ where the g_i, etc., are components of a zero-mean Gaussian random vector, the

$$\left\langle g_i g_j g_k g_l \right\rangle = \left(\frac{\partial^4 \psi_{\mathbf{g}}(\boldsymbol{\xi})}{\partial \xi_l \partial \xi_k \partial \xi_j \partial \xi_i} \right)_{\boldsymbol{\xi}=\mathbf{0}} = K_{ij} K_{kl} + K_{jk} K_{il} + K_{ik} K_{jl}\,. \tag{8.197}$$

This result is referred to as the *Gaussian moment theorem*

For the case where $i = j = k = l$, we find $\left\langle g_i^4 \right\rangle = 3\sigma_i^4$, which is a familiar result for univariate normals.

Gaussian random processes

Conventional definition:

A random process $f(\mathbf{r})$ is normal (Gaussian) if all M-point PDFs, pr$[f(\mathbf{r}_1), f(\mathbf{r}_2), ..., f(\mathbf{r}_M)]$ for all M, are normal.

Unconventional definition used in Barrett & Myers:

A real-valued random process is normal if its characteristic functional is given by

$$\Psi_{\mathbf{f}}(\mathbf{s}) = \exp\left(-2\pi i \mathbf{s}^\dagger \bar{\mathbf{f}}\right) \exp\left(-2\pi^2 \mathbf{s}^\dagger \mathcal{K}_{\mathbf{f}}\, \mathbf{s}\right), \tag{8.216}$$

where $\mathcal{K}_{\mathbf{f}}$ is the autocovariance operator, *i.e.*, the integral operator with kernel $K_{\mathbf{f}}(\mathbf{r}, \mathbf{r}')$.

Compare:

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \exp\left(-2\pi i \boldsymbol{\xi}^t \overline{\mathbf{g}}\right) \exp\left(-2\pi^2 \boldsymbol{\xi}^t \mathbf{K} \boldsymbol{\xi}\right) . \tag{8.196}$$

We will now show that the second definition implies the first.

Linear functionals of Gaussian random processes

Basic definition (from last slide):

$$\Psi_{\mathbf{f}}(\mathbf{s}) = \exp\left(-2\pi i \mathbf{s}^{\dagger}\overline{\mathbf{f}}\right) \exp\left(-2\pi^2 \mathbf{s}^{\dagger} \boldsymbol{\mathcal{K}}_{\mathbf{f}}\, \mathbf{s}\right), \tag{8.216}$$

Recall the transformation rule for a general linear mapping, $\mathbf{g} = \boldsymbol{\mathcal{H}}\mathbf{f}$:

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

For a Gaussian process,

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \exp\left(-2\pi i \boldsymbol{\xi}^{\dagger}\boldsymbol{\mathcal{H}}\overline{\mathbf{f}}\right) \exp\left(-2\pi^2 \boldsymbol{\xi}^{\dagger}\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{K}}_{\mathbf{f}}\boldsymbol{\mathcal{H}}^{\dagger}\boldsymbol{\xi}\right). \tag{8.218}$$

Compare

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \exp\left(-2\pi i \boldsymbol{\xi}^{t}\overline{\mathbf{g}}\right) \exp\left(-2\pi^2 \boldsymbol{\xi}^{t}\mathbf{K}\boldsymbol{\xi}\right). \tag{8.196}$$

Thus $\mathbf{g}$ is an MD random vector with mean $\boldsymbol{\mathcal{H}}\overline{\mathbf{f}}$ and covariance $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{K}}_{\mathbf{f}}\boldsymbol{\mathcal{H}}^{\dagger}$.

Conclusion: All linear functionals of a Gaussian random process are Gaussian. Hence our definition of Gaussian random process implies the conventional one.

Multipoint densities and autocovariance functions

A general linear CD mapping is defined by

$$g_m = \int_\infty d^q r \; h_m(\mathbf{r}) \, f(\mathbf{r}) \, , \qquad m = 1, ..., M \, .$$

To sample (evaluate) the random process at a set of points, just let

$$h_m(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}_m) \, . \tag{8.219}$$

Thus $g_m = f(\mathbf{r}_m)$, and it follows from the last slide that pr$[f(\mathbf{r}_1), f(\mathbf{r}_2), ..., f(\mathbf{r}_M)]$ is an MD normal density if $f(\mathbf{r})$ is a normal random process. Explicitly,

$$\text{pr}[f(\mathbf{r}_1), f(\mathbf{r}_2), ..., f(\mathbf{r}_M)] = \text{pr}(\mathbf{f}_M)$$

$$= (2\pi)^{-\frac{1}{2}M} |\det \mathbf{K}_M|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2} (\mathbf{f}_M - \overline{\mathbf{f}}_M)^t \, \mathbf{K}_M^{-1} (\mathbf{f}_M - \overline{\mathbf{f}}_M) \right] \, , \tag{8.220}$$

where $\overline{\mathbf{f}}_M$ is the $M \times 1$ mean vector, with components $\langle f(\mathbf{r}_m) \rangle$, and $\mathbf{K}_M$ is the $M \times M$ covariance matrix. with components given by

$$[\mathbf{K}_M]_{mn} = \langle [f(\mathbf{r}_m) - \langle f(\mathbf{r}_m) \rangle] \, [f(\mathbf{r}_n) - \langle f(\mathbf{r}_n) \rangle] \rangle \, . \tag{8.221}$$

Multipoint densities and autocovariance functions – cont.

Comparison with (8.98) shows that

$$[\mathbf{K}_M]_{mn} = K_{\mathbf{f}}(\mathbf{r}_m, \mathbf{r}_n)\,. \tag{8.222}$$

Thus the covariance *matrix* in an M-point PDF for a normal random process is fully determined by the autocovariance *function* of the process.

Gaussian noise in imaging

Many noise sources are well modeled as Gaussian by the central-limit theorem:

- Thermal (Johnson) noise
- Flicker ($1/f$) noise

Sometimes the noise is actually Poisson but well approximated as Gaussian:

- Shot noise (if number of electrons is large)
- Photon noise (if number of photons is large)

Overall noise in a CD imaging system

Basic doubly stochastic CD model:

$$\mathbf{g} = \boldsymbol{\mathcal{H}}\mathbf{f} + \mathbf{n}\,,$$

where $\mathbf{n}$ is a zero-mean random *vector* and $\mathbf{f}$ is a general random process, If $\mathbf{n}$ is statistically independent of $\mathbf{f}$, the characteristic *function* for $\mathbf{g}$ is

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

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

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

Important special case: components of $\mathbf{n}$ are i.i.d. normal, so $\mathbf{K_n} = \sigma^2\mathbf{I}$. Then, from (8.196),

$$\psi_{\mathbf{n}}(\boldsymbol{\xi}) = \exp\left(-2\pi^2\boldsymbol{\xi}^t\mathbf{K}\boldsymbol{\xi}\right) = \exp\left(-2\pi^2\sigma^2|\boldsymbol{\xi}|^2\right)\,.$$

Overall noise in a CD imaging system – cont.

Collecting results from last slide, we get

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \Psi_{\mathbf{f}}(\boldsymbol{\mathcal{H}}^{\dagger}\boldsymbol{\xi}\,)\ \exp\left(-2\pi^2\sigma^2|\boldsymbol{\xi}|^2\right)\ .$$

PDF on $\mathbf{g}$ is obtained by inverse MD Fourier transform; result is convolution of PDFs for $\boldsymbol{\mathcal{H}}\mathbf{f}$ and $\mathbf{n}$.

We will see later how to use this result for specific object models.

$\boxed{\text{Still to come}}$

- Complex Gaussian random processes
- Poisson random processes
- Object models
- Applications in astronomy