Lecture 4: Optimization theory and random processes
Outline

1. Complex gradients
2. Optimization theory
3. Random variables and random processes
The local and global minima of an objective function $f(x)$, with real $x$, satisfy

$$\frac{\partial f(x)}{\partial x} = \nabla_x f(x) = 0 \quad \text{and} \quad \frac{\partial^2 f(x)}{\partial x^2} = \nabla_x^2 f(x) > 0$$

If $f(x)$ is convex, then the local minimum is the global minimum.

For $f(z)$ with complex $z$, we write $f(z)$ as $f(z, z^*)$ and treat $z = x + jy$ and $z^* = x - jy$ as independent variables and define the partial derivatives w.r.t. $z$ and $z^*$ as

$$\frac{\partial f}{\partial z} = \frac{1}{2} \left[ \frac{\partial f}{\partial x} - j \frac{\partial f}{\partial y} \right] \quad \text{and} \quad \frac{\partial f}{\partial z^*} = \frac{1}{2} \left[ \frac{\partial f}{\partial x} + j \frac{\partial f}{\partial y} \right]$$

For an objective function $f(z, z^*)$, the stationary points of $f(z, z^*)$ are found by setting the derivative of $f(z, z^*)$ w.r.t. $z$ or $z^*$ to zero.
For an objective function in two or more real variables, \( f(x_1, x_2, \ldots, x_N) = f(x) \), the first-order derivative (gradient) and the second-order derivative (Hessian) are given by

\[
[\nabla_x f(x)]_i = \frac{\partial f(x)}{\partial x_i} \quad \text{and} \quad [H(x)]_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}
\]

The local and global minima of an objective function \( f(x) \), with real \( x \), satisfy

\[
\nabla_x f(x) = 0 \quad \text{and} \quad H(x) > 0
\]

For an objective function \( f(z, z^*) \), the stationary points of \( f(z, z^*) \) are found by setting the derivative of \( f(z, z^*) \) w.r.t. \( z \) or \( z^* \) to zero.
Random variables

- A random variable $x$ is a function that assigns a number to each outcome of a random experiment.

- Probability distribution function

$$F_x(\alpha) = \Pr\{x \leq \alpha\}$$

- Probability distribution function

$$f_x(\alpha) = \frac{d}{d\alpha}F_x(\alpha)$$

- Mean or expected value

$$m_x = E\{x\} = \int_{-\infty}^{\infty} \alpha f_x(\alpha) d\alpha$$

- Variance

$$\sigma^2_x = \text{var}\{x\} = E\{(x - m_x)^2\} = \int_{-\infty}^{\infty} (\alpha - m_x)^2 f_x(\alpha) d\alpha$$
Random variables

- Joint probability distribution function

\[ F_{x,y}(\alpha, \beta) = \Pr\{x \leq \alpha, y \leq \beta\} \]

- Probability distribution function

\[ f_{x,y}(\alpha, \beta) = \frac{d^2}{d\alpha d\beta} F_{x,y}(\alpha, \beta) \]

- \( x \) and \( y \) are independent: \( f_{x,y}(\alpha, \beta) = f_x(\alpha)f_x(\beta) \)

- Correlation

\[ r_{xy} = E\{xy^*\} \]

- \( x \) and \( y \) are uncorrelated: \( E\{xy^*\} = E\{x\}E\{y^*\} \) or \( r_{xy} = m_x m_y^* \) or \( c_{xy} = 0 \).

- \( r_{xy} = 0 \) means \( x \) and \( y \) are statistically orthogonal.

- Covariance

\[ c_{xy} = \text{cov}\{x, y\} = E\{(x - m_x)(y - m_y)^*\} = r_{xy} - m_x m_y^* \]
A random process $x(n)$ is a sequence of random variables.

**Probability distribution function**

$$F_x(\alpha) = \Pr\{x \leq \alpha\}$$

**Mean and variance:**

$$m_x = E\{x\} \quad \text{and} \quad \sigma^2_x(n) = E\{|x(n) - m_x(n)|^2\}$$

**Autocorrelation and autocovariance**

$$r_x(k, l) = E\{x(k)x^*(l)\}$$

$$c_x(k, l) = E\{[x(k) - m_x(k)][x(l) - m_x(l)]^*\}$$
First-order stationarity if $f_{x(n)}(\alpha) = f_{x(n+k)}(\alpha)$. This implies $m_x(n) = m_x(0) = m_x$.

Second-order stationarity if, for any $k$, the process $x(n)$ and $x(n + k)$ have the same second-order density function:

$$f_{x(n_1),x(n_2)}(\alpha_1, \alpha_2) = f_{x(n_1+k),x(n_2+k)}(\alpha_1, \alpha_2).$$

This implies $r_x(k, l) = r_x(k - l, 0) = r_x(k - l)$. 
Wide-sense stationarity

- **Wide-sense stationary (WSS):**

  \[ m_x(n) = m_x; \quad r_x(k, l) = r_{xy}(k - l); \quad c_x(0) < \infty. \]

- **Properties of WSS processes:**
  - **Symmetry:** \( r_x(k) = r_x^*(-k) \)
  - **Mean-square value:** \( r_x(0) = E\{|x(n)|^2\} \geq 0. \)
  - **Maximum value:** \( r_x(0) \geq |r_x(k)| \)
  - **Mean-squared periodic:** \( r_x(k_0) = r_x(0) \)
Autocorrelation and autocovariance matrices

- Consider a WSS process $x(n)$ and collect $p + 1$ samples in
  
  \[ x = [x(0), x(1), \ldots, x(p)]^T \]

- Autocorrelation matrix:

  \[ R_x = E\{xx^H\} = \begin{bmatrix}
  r_x(0) & r_x^*(1) & r_x^*(2) & \cdots & r_x^*(p) \\
  r_x(1) & r_x(0) & r_x^*(1) & \cdots & r_x^*(p - 1) \\
  r_x(2) & r_x(1) & r_x(0) & \cdots & r_x^*(p - 2) \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  r_x(p) & r_x(p - 1) & r_x(p - 2) & \cdots & r_x(0)
  \end{bmatrix} \]

  - $R_x$ is Toeplitz, Hermitian, and nonnegative definite.

- Autocovariance matrix: $C_x = R_x - m_x m_x^H$, where $m_x = m_x \mathbf{1}$