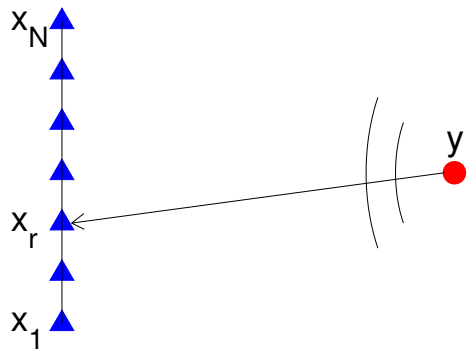
Source imaging

Source imaging: data acquisition



Passive array $\Longleftrightarrow$ The sensors are receivers.

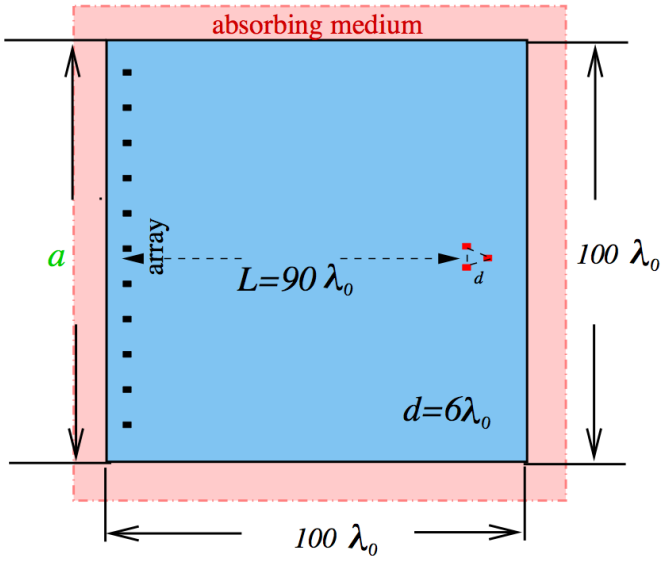
The source $\boldsymbol{y}$ emits a short pulse.

The sensors $(\boldsymbol{x}_r)_{r=1,\dots,N}$ record the waves.

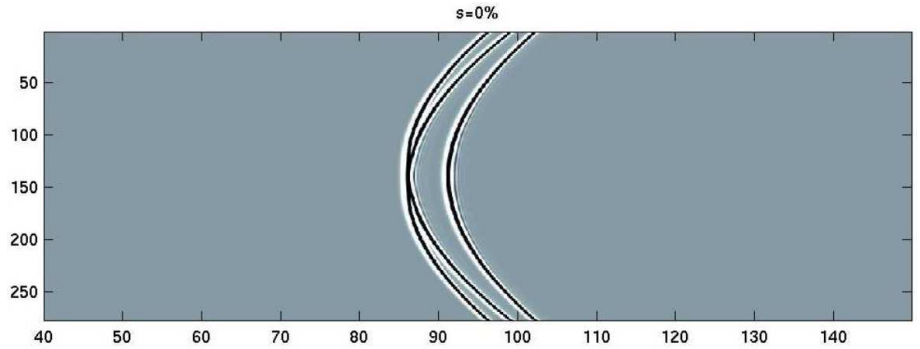
The data set is $\big(u(t,\boldsymbol{x}_r)\big)_{t\in\mathbb{R},r=1,\dots,N}$.

Goal: find the source position $\boldsymbol{y}$ (more generally, find the *source* region).

Source imaging: simulation

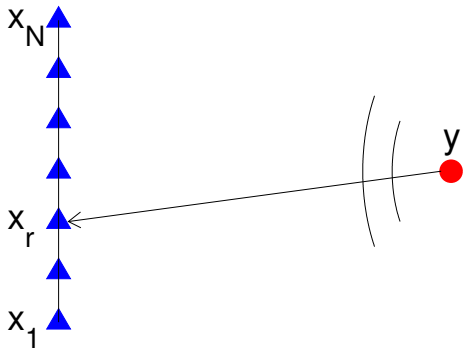


Configuration

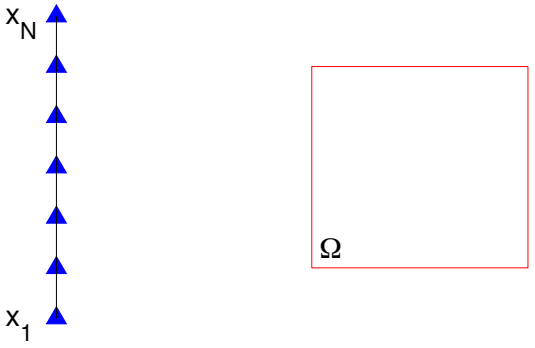


Data set $(u(t, \boldsymbol{x}_r))_{80 \leq t \leq 200, r=1,\ldots,270}$

Source imaging: imaging function



Data acquisition



Search region

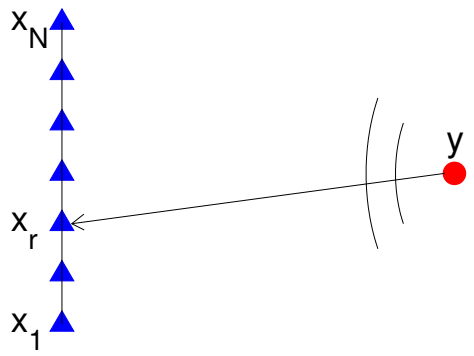
Goal: find the source point $\boldsymbol{y}$ (more generally, find the *source* region).

The data set is $\big(u(t,\boldsymbol{x}_r)\big)_{t\in\mathbb{R},r=1,\ldots,N}$.

$\hookrightarrow$ Given the data set, build an imaging function in the search region $\Omega\subset\mathbb{R}^3$:

$$\mathcal{I}:\left|\begin{array}{l}\Omega\to\mathbb{R}^+\\ \boldsymbol{y}^S\mapsto\mathcal{I}(\boldsymbol{y}^S)\end{array}\right.$$ which plots an image of the search region.

Source imaging - the linear forward operator



The source term is of the form $n(t, \boldsymbol{y}) = \rho_{\text{real}}(\boldsymbol{y})\delta(t)$.
Goal: find the source function ρ_{real}.
Here: the Green's function is known.

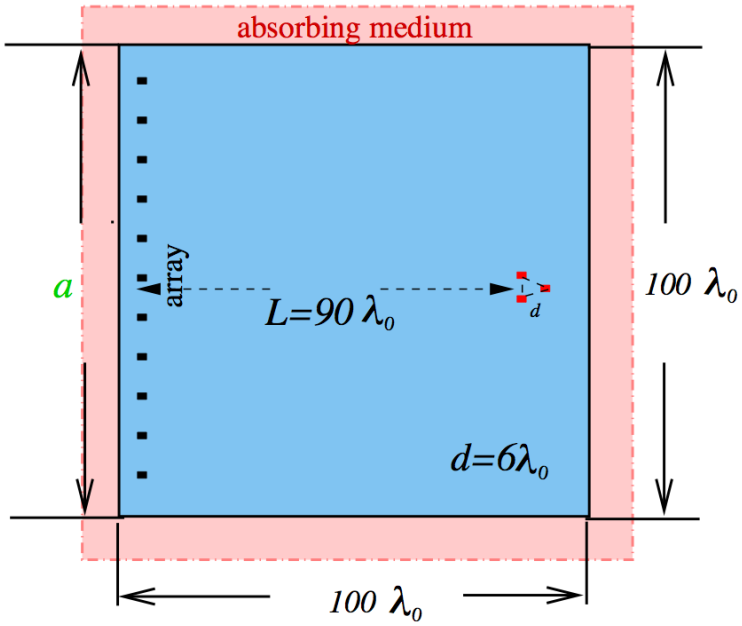
The data set is $\hat{\boldsymbol{u}} = \left(\hat{u}(\omega, \boldsymbol{x}_r)\right)_{\omega\in\mathcal{B}, r=1,\ldots,N}$ with $\hat{u}(\omega, \boldsymbol{x}_r) = \int_\Omega \hat{G}(\omega, \boldsymbol{x}_r, \boldsymbol{y})\rho_{\text{real}}(\boldsymbol{y})d\boldsymbol{y}$.
We define

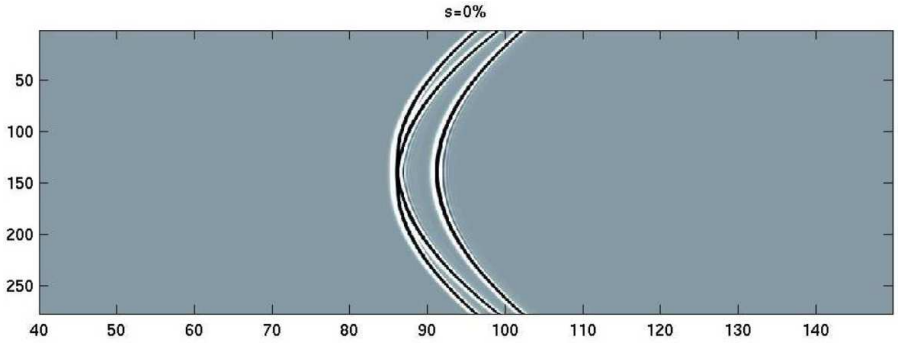
$$[\hat{\mathbf{A}}\rho](\omega, \boldsymbol{x}_r) = \int_\Omega \hat{G}(\omega, \boldsymbol{x}_r, \boldsymbol{y})\rho(\boldsymbol{y})d\boldsymbol{y}.$$

$\hat{\mathbf{A}}$ is the linear operator that maps the source function to the array data $\hat{\boldsymbol{u}}$:

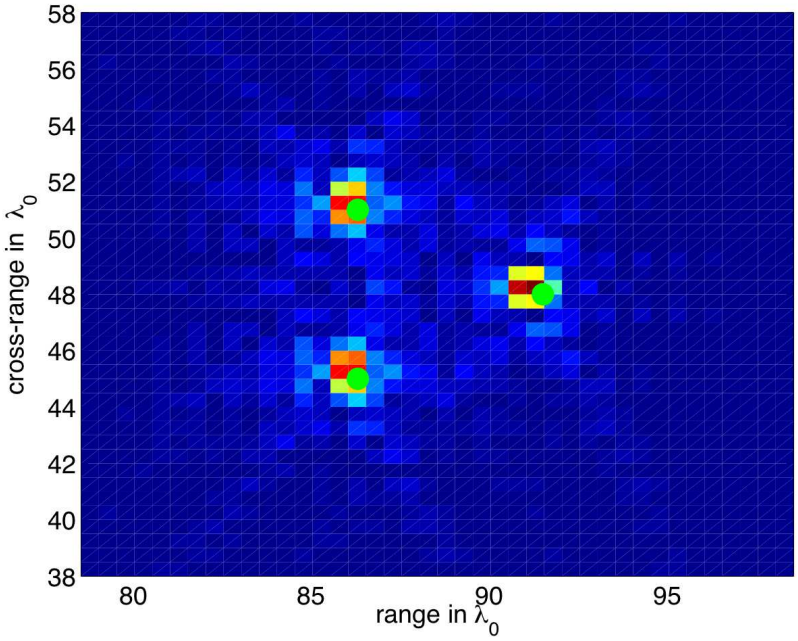
$$\hat{\boldsymbol{u}} = \hat{\mathbf{A}}\rho_{\text{real}}$$

Kirchhoff migration for source imaging: simulation





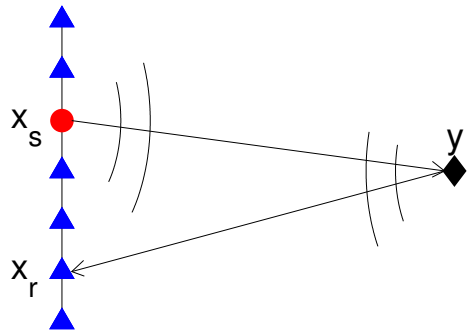
Data $\left(u(t, \boldsymbol{x}_r)\right)_{t\in[40,150], r=1,\ldots,264}$



KM imaging function $\mathcal{I}_{\mathrm{KM}}(\boldsymbol{y}^S)$

Reflector imaging

Reflector imaging: data acquisition



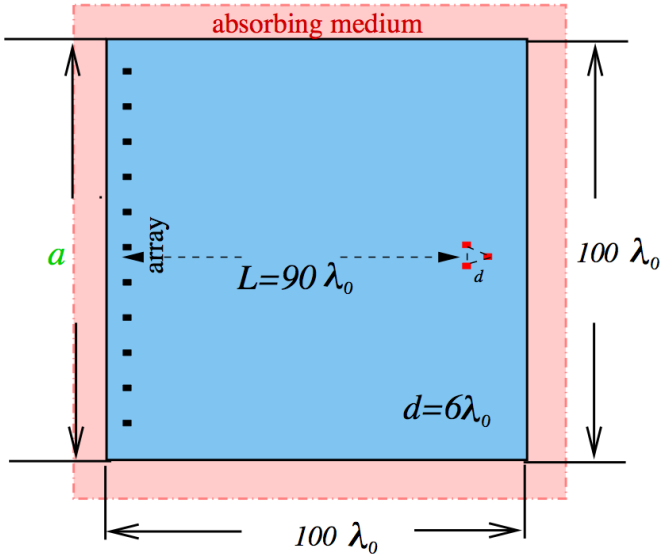
Active array $\Longleftrightarrow$ The sensors can be used as sources and/or as receivers.

For each $s = 1, \ldots, N$:
- The source $\boldsymbol{x}_s$ emits a short pulse.
- The sensors $(\boldsymbol{x}_r)_{r=1,\ldots,N}$ record the waves.

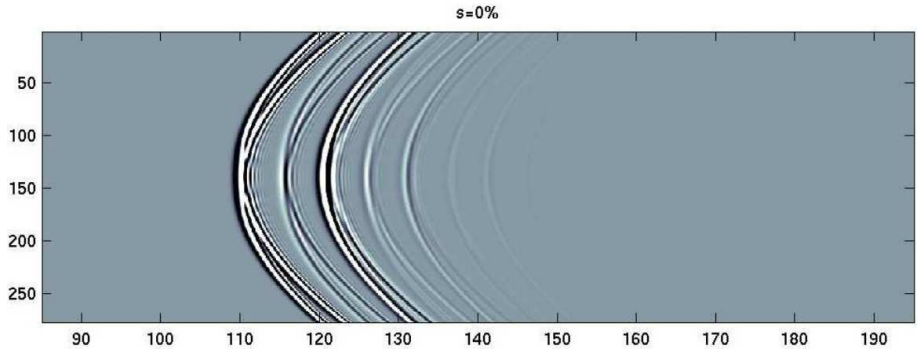
The data set is $(u(t, \boldsymbol{x}_r, \boldsymbol{x}_s))_{t\in\mathbb{R},r,s=1,\ldots,N}$ (the time-dependent response matrix).

Goal: find the reflector position $\boldsymbol{y}$ (more generally, find the reflectivity function of the *medium*).

Reflector imaging: simulation

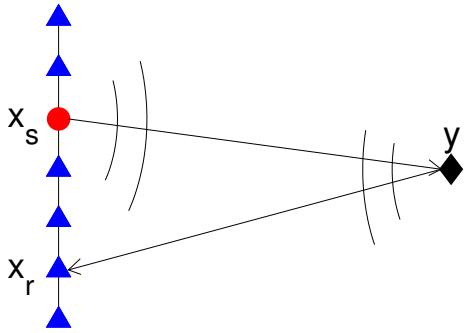


Configuration

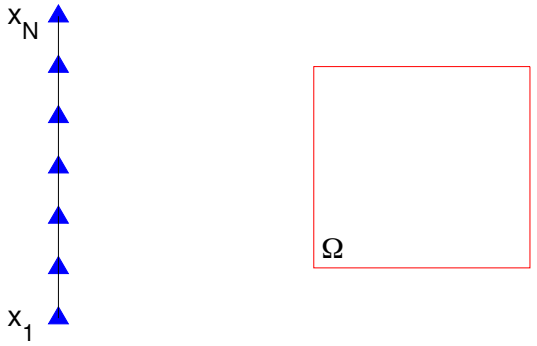


Data set $(u(t, \boldsymbol{x}_r, \boldsymbol{x}_{135}))_{80 \leq t \leq 200, r=1,\ldots,270}$
(traces recorded for a central illumination)

Reflector imaging: imaging function



Data acquisition



Search region

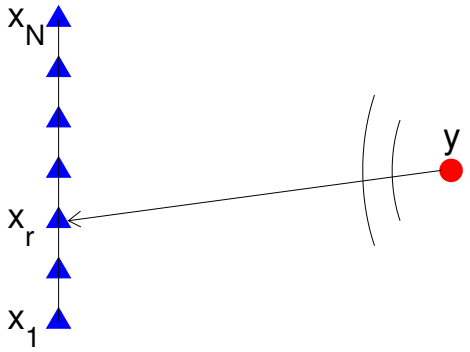
Goal: find the reflector position $\boldsymbol{y}$ (more generally, find the reflectivity function of the *medium*).

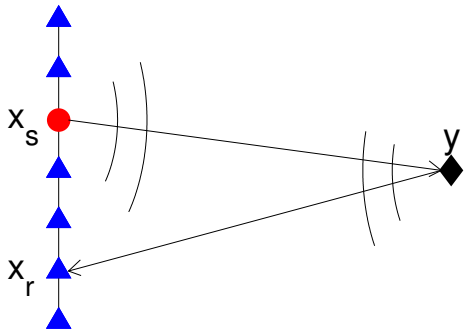
The data set is $(u(t, \boldsymbol{x}_r, \boldsymbol{x}_s))_{t\in\mathbb{R}, r,s=1,\dots,N}$.

$\hookrightarrow$ Given the data set, build an imaging function in the search region $\Omega \subset \mathbb{R}^3$:

$$\mathcal{I} : \left| \begin{array}{l} \Omega \to \mathbb{R}^+ \\ \boldsymbol{y}^S \mapsto \mathcal{I}(\boldsymbol{y}^S) \end{array} \right.$$ which plots an image of the search region.

Source and reflector imaging: comparison





Passive array of sensors

The sensors $(\boldsymbol{x}_r)_{r=1,\dots,N}$ record

$\boldsymbol{y}$ is a source

Data: $\big(u(t,\boldsymbol{x}_r)\big)_{t\in\mathbb{R},r=1,\dots,N}$

Active array of sensors/sources

The sensors $(\boldsymbol{x}_r)_{r=1,\dots,N}$ record

$\boldsymbol{y}$ is a reflector, $\boldsymbol{x}_s$ is a source

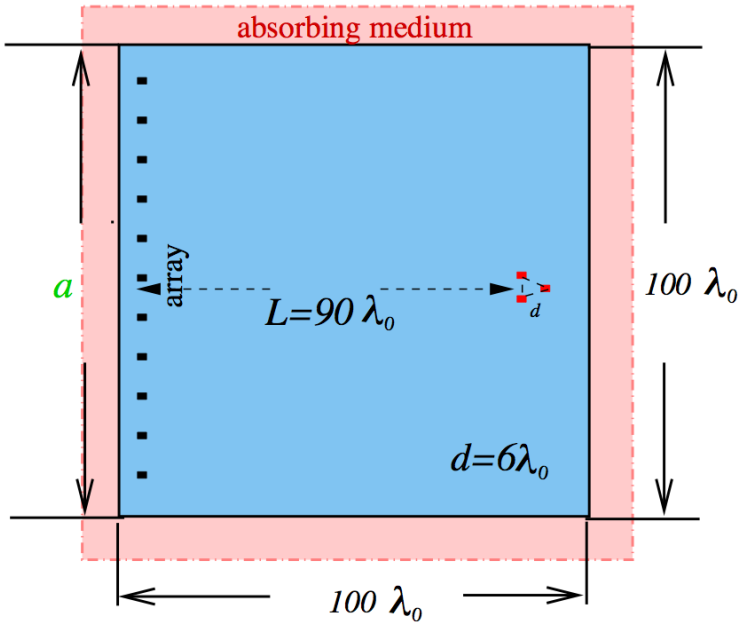
Data: $\big(u(t,\boldsymbol{x}_r,\boldsymbol{x}_s)\big)_{t\in\mathbb{R},r,s=1,\dots,N}$

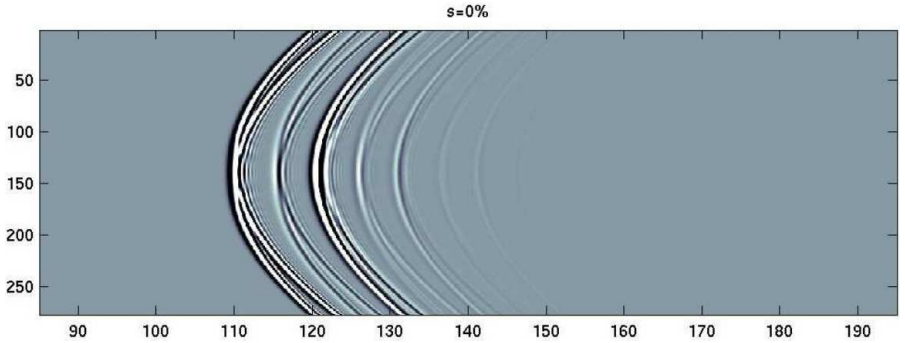
Goal: process the data to find $\boldsymbol{y}$.

Passive and active array imaging: important features

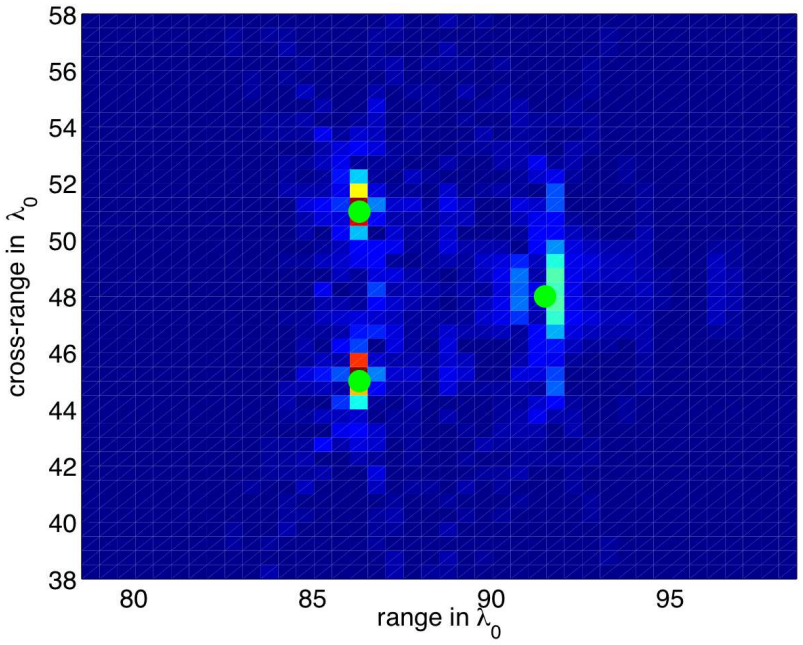
- Different data acquisition geometries: Dense arrays or distributed networks, synthetic aperture arrays (zero-offset), full aperture imaging (medical).
- Point targets or extended targets.
- Narrowband (microwaves), broadband (ultrasound), and noise probing signals.
- Coherent (radar, sonar, seismic ...) and incoherent (infrared, optical tomography, X-ray tomography) imaging.
- Amount of data.
- Real time processing.
- Signal-to-Noise ratio (SNR) issues. Stability with respect to additive noise (electronic noise) and to partial knowledge of the medium (clutter noise).

Kirchhoff migration for reflector imaging: simulation





Data $\left(u(t,\boldsymbol{x}_r,\boldsymbol{x}_{128})\right)_{t\in[80,200],r=1,\ldots,264}$

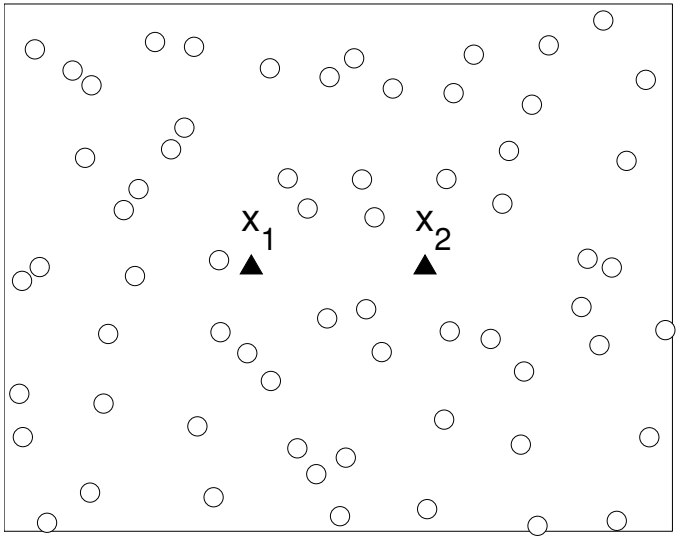


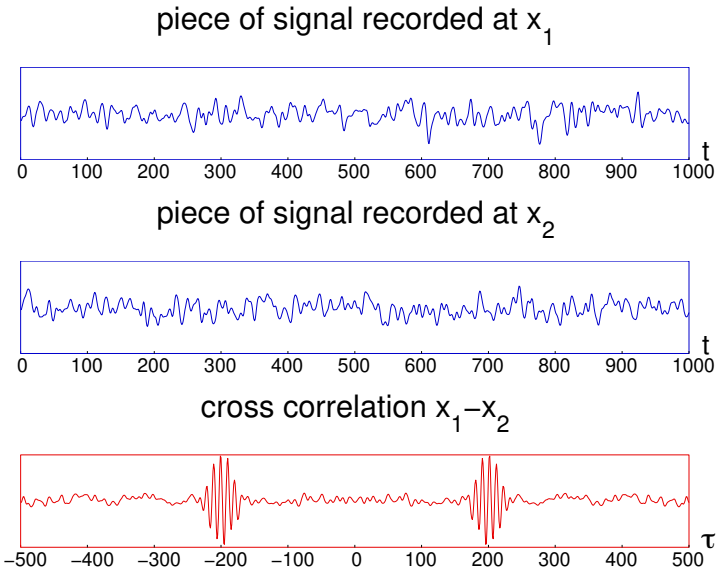
KM imaging function
$\mathcal{I}_{\mathrm{KM}}(\boldsymbol{y}^S)$

Passive imaging with noise sources

Green's function estimation by cross correlation of ambient noise signals

- Ambient noise sources ($\circ$) emit stationary random signals.
- The waves propagate in the (inhomogeneous) medium.
- The signals $u(t, \boldsymbol{x}_1)$ and $u(t, \boldsymbol{x}_2)$ are recorded at two sensors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.





- Compute the empirical cross correlation:

$$C_T(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{1}{T} \int_0^T u(t, \boldsymbol{x}_1) u(t + \tau, \boldsymbol{x}_2) dt$$

- $C_T(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2)$ is related to the Green's function from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$!

Wave equation

$$\frac{1}{c^2(\boldsymbol{x})}\frac{\partial^2 u}{\partial t^2} - \Delta_{\boldsymbol{x}} u = n(t, \boldsymbol{x})$$

- Three-dimensional inhomogeneous medium (background velocity $c(\boldsymbol{x})$).

- Sources $n(t, \boldsymbol{x})$: Gaussian process, with mean zero, with covariance

$$\langle n(t_1, \boldsymbol{y}_1) n(t_2, \boldsymbol{y}_2) \rangle = F(t_2 - t_1)\Gamma(\boldsymbol{y}_1, \boldsymbol{y}_2)$$

- Stationary in time: $\big(n(t, \boldsymbol{y})\big)_{t,\boldsymbol{y}}$ and $\big(n(t+h, \boldsymbol{y})\big)_{t,\boldsymbol{y}}$ have the same statistical distribution for any $h \Longrightarrow$ the time correlation function F depends only on $t_2 - t_1$.

- The spatial distribution of the sources is characterized by $\Gamma(\boldsymbol{y}_1, \boldsymbol{y}_2)$. To simplify:

$$\Gamma(\boldsymbol{y}_1, \boldsymbol{y}_2) = K(\boldsymbol{y}_1)\delta(\boldsymbol{y}_1 - \boldsymbol{y}_2)$$

The function K characterizes the spatial support of the sources.
[For more general Γ, use multiscale or microlocal analysis, see below.]

Remember: the Green's function

Solution u of the wave equation:

$$u(t, \boldsymbol{x}) = \int \int G(s, \boldsymbol{x}, \boldsymbol{y}) n(t - s, \boldsymbol{y}) ds d\boldsymbol{y}$$

Green's function:

$$\frac{1}{c^2(\boldsymbol{x})} \frac{\partial^2 G}{\partial t^2} - \Delta_{\boldsymbol{x}} G = \delta(t)\delta(\boldsymbol{x} - \boldsymbol{y})$$

starting from $G(0, \boldsymbol{x}, \boldsymbol{y}) = \partial_t G(0, \boldsymbol{x}, \boldsymbol{y}) = 0 \quad (G(t, \boldsymbol{x}, \boldsymbol{y}) = 0 \text{ for } t < 0)$.

Time-harmonic Green's function:

$$\hat{G}(\omega, \boldsymbol{x}, \boldsymbol{y}) = \int G(t, \boldsymbol{x}, \boldsymbol{y}) e^{i\omega t} dt$$

solution of

$$\frac{\omega^2}{c^2(\boldsymbol{x})} \hat{G} + \Delta_{\boldsymbol{x}} \hat{G} = -\delta(\boldsymbol{x} - \boldsymbol{y})$$

+ Sommerfeld radiation condition.

Example: If $c(\boldsymbol{x}) \equiv c_0$ (and in dimension 3)

$$G(t, \boldsymbol{x}, \boldsymbol{y}) = \frac{1}{4\pi|\boldsymbol{x} - \boldsymbol{y}|} \delta\Big(\frac{|\boldsymbol{x} - \boldsymbol{y}|}{c_0} - t\Big)$$

$$\hat{G}(\omega, \boldsymbol{x}, \boldsymbol{y}) = \frac{1}{4\pi|\boldsymbol{x} - \boldsymbol{y}|} e^{i\omega \frac{|\boldsymbol{x} - \boldsymbol{y}|}{c_0}}$$

Empirical cross correlation and statistical cross correlation

Empirical cross correlation:

$$C_T(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{1}{T} \int_0^T u(t, \boldsymbol{x}_1) u(t+\tau, \boldsymbol{x}_2) dt$$

with $u(t, \boldsymbol{x}) = \iint G(s, \boldsymbol{x}, \boldsymbol{y}) n(t-s, \boldsymbol{y}) ds d\boldsymbol{y}$ and $G =$ causal time-dependent Green's function.

1. The expectation of C_T (with respect to the distribution of the sources) is independent of the integration time T:

$$\langle C_T(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) \rangle = C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2)$$

where the statistical cross correlation $C^{(1)}$ is given by

$$C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{1}{2\pi} \int d\boldsymbol{y} \int d\omega \overline{\hat{G}}(\omega, \boldsymbol{x}_1, \boldsymbol{y}) \hat{G}(\omega, \boldsymbol{x}_2, \boldsymbol{y}) K(\boldsymbol{y}) \hat{F}(\omega) e^{-i\omega\tau}$$

Empirical cross correlation and statistical cross correlation

Empirical cross correlation:

$$C_T(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{1}{T} \int_0^T u(t, \boldsymbol{x}_1) u(t+\tau, \boldsymbol{x}_2) dt$$

with $u(t, \boldsymbol{x}) = \iint G(s, \boldsymbol{x}, \boldsymbol{y}) n(t-s, \boldsymbol{y}) ds d\boldsymbol{y}$ and $G =$ causal time-dependent Green's function.

1. The expectation of C_T (with respect to the distribution of the sources) is independent of the integration time T:

$$\langle C_T(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) \rangle = C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2)$$

where the statistical cross correlation $C^{(1)}$ is given by

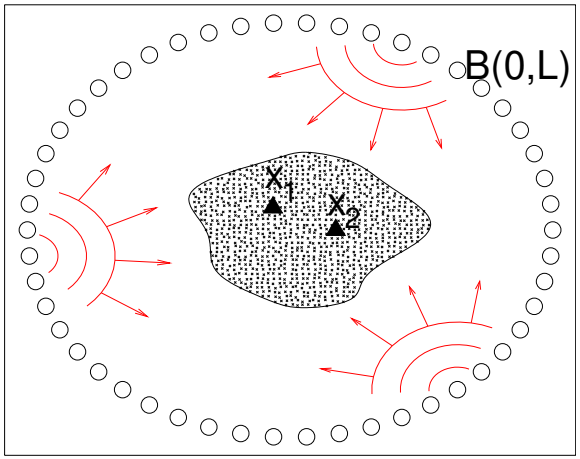
$$C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{1}{2\pi} \int d\boldsymbol{y} \int d\omega \overline{\hat{G}}(\omega, \boldsymbol{x}_1, \boldsymbol{y}) \hat{G}(\omega, \boldsymbol{x}_2, \boldsymbol{y}) K(\boldsymbol{y}) \hat{F}(\omega) e^{-i\omega\tau}$$

2. The empirical cross correlation is a self-averaging quantity:

$$C_T(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) \overset{T \to \infty}{\longrightarrow} C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2)$$

in probability.
More precisely, the fluctuations of C_T around its expectation $C^{(1)}$ are of order $T^{-1/2}$.

Emergence of the Green's function for an extended distribution of sources in an inhomogeneous open medium



Cross correlation with noise sources distributed on a closed surface $\partial B(\mathbf{0}, L)$:

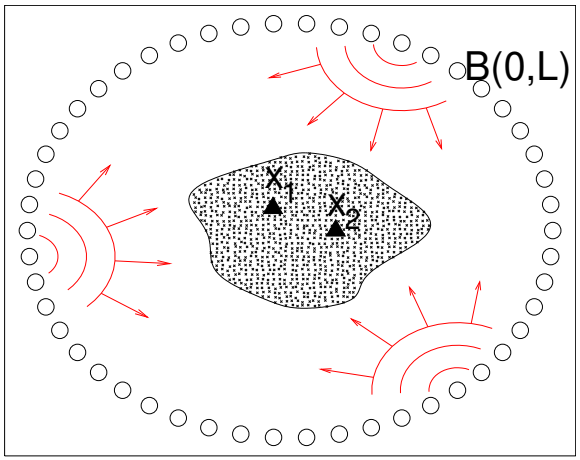
$$C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{1}{2\pi} \int d\omega \int_{\partial B(\mathbf{0},L)} d\sigma(\boldsymbol{y}) \overline{\hat{G}}(\omega, \boldsymbol{x}_1, \boldsymbol{y}) \hat{G}(\omega, \boldsymbol{x}_2, \boldsymbol{y}) \hat{F}(\omega) e^{-i\omega\tau}$$

By Helmholtz-Kirchhoff identity,

$$\hat{G}(\omega, \boldsymbol{x}_1, \boldsymbol{x}_2) - \overline{\hat{G}}(\omega, \boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{2i\omega}{c_0} \int_{\partial B(\mathbf{0},L)} d\sigma(\boldsymbol{y}) \overline{\hat{G}}(\omega, \boldsymbol{x}_1, \boldsymbol{y}) \hat{G}(\omega, \boldsymbol{x}_2, \boldsymbol{y})$$

we have

$$C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{c_0}{4\pi} \int \frac{\hat{F}(\omega)}{\omega} \mathrm{Im}\big(\hat{G}(\omega, \boldsymbol{x}_1, \boldsymbol{x}_2)\big) e^{-i\omega\tau} d\omega$$



$$\partial_\tau C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) = -\frac{ic_0}{4\pi} \int \hat{F}(\omega) \mathrm{Im}\big(\hat{G}(\omega, \boldsymbol{x}_1, \boldsymbol{x}_2)\big) e^{-i\omega\tau} d\omega$$

$$= -\frac{c_0}{2} \Big(F *_\tau G(\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) - F *_\tau G(-\tau, \boldsymbol{x}_1, \boldsymbol{x}_2) \Big)$$

• The cross correlation of noise signals recorded by two passive sensors is related to the Green's function between the sensors.

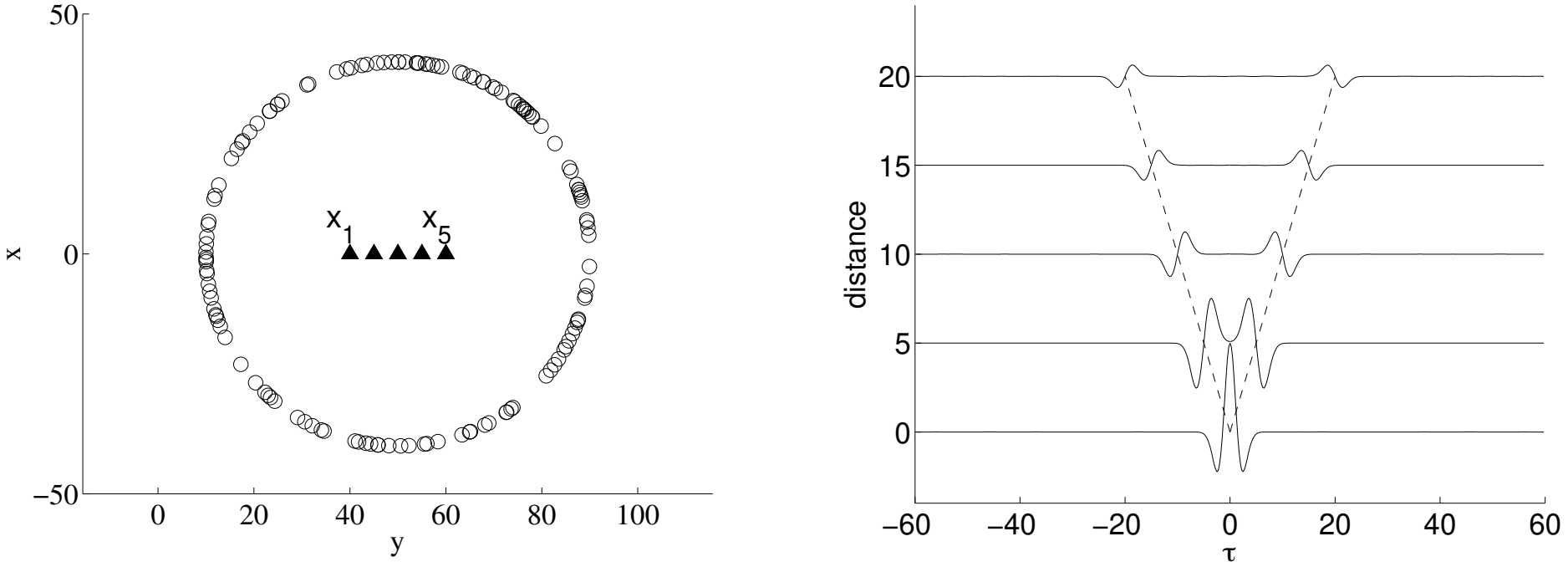
$\hookrightarrow$ this is the classical proof that passive sensors can be transformed into virtual sources (known in seismology).

• This proof requires the noise sources to surround the region of interest.

The previous proofs also establish exact relations in ideal conditions.

Other proofs can justify approximate relations (enough for travel time estimation) in realistic conditions.

Ideal illumination:

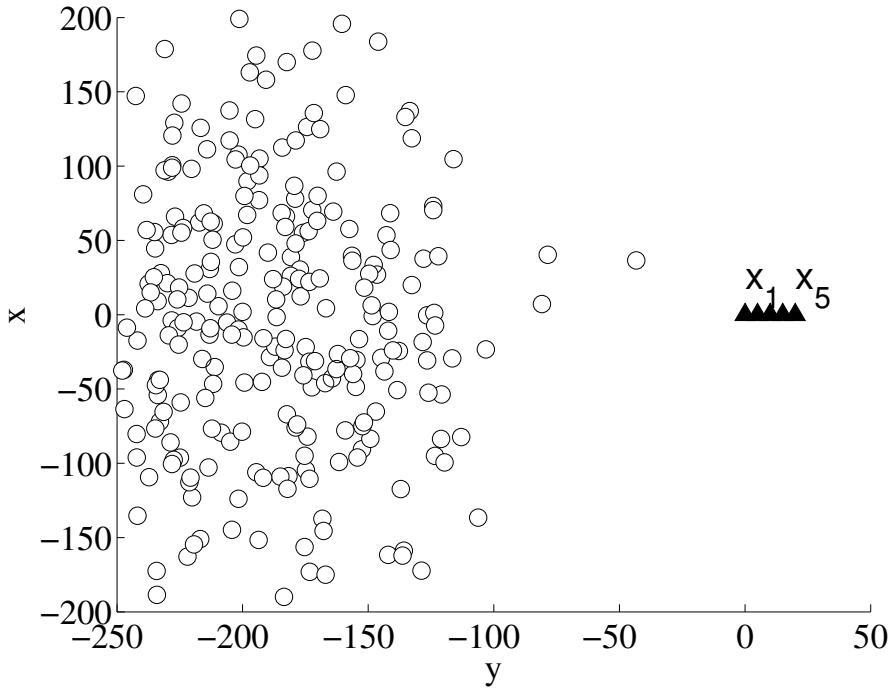


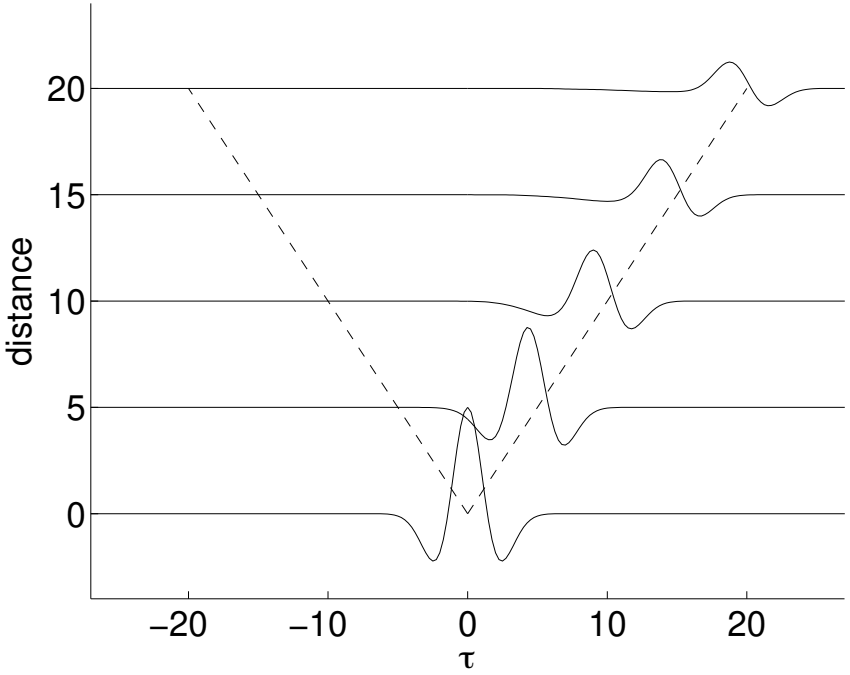
Left: The circles are the noise sources and the triangles are the sensors (the distance between two successive sensors is $\underline{5}$). Here $\hat{F}(\omega) = \omega^2 \hat{f}(\omega)$, $\hat{f}(\omega) = \exp(-\omega^2)$, $c_0 = 1$. Right: Cross correlation $\tau \to C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_j)$ between the pairs of sensors $(\boldsymbol{x}_1, \boldsymbol{x}_j)$, $j = 1, \ldots, 5$, versus the distance $|\boldsymbol{x}_j - \boldsymbol{x}_1|$. In theory:

$$C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_j) = \frac{c_0}{8\pi|\boldsymbol{x}_1 - \boldsymbol{x}_j|}\left[f'\left(\tau - \frac{|\boldsymbol{x}_j - \boldsymbol{x}_1|}{c_0}\right) - f'\left(\tau + \frac{|\boldsymbol{x}_j - \boldsymbol{x}_1|}{c_0}\right)\right], \quad j \geq 2$$

$$C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_1) = -\frac{1}{4\pi} f''(\tau)$$

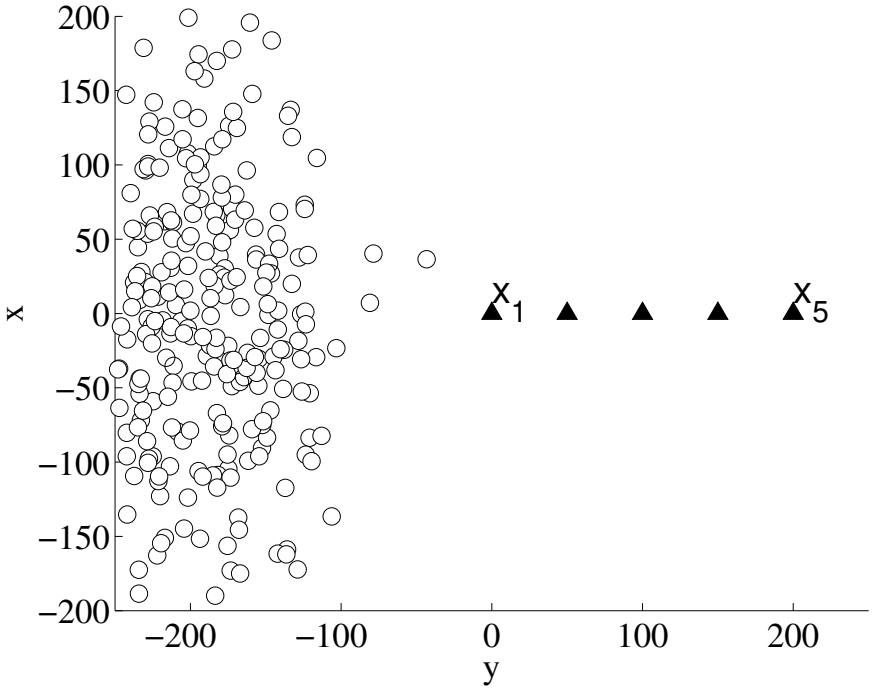
Realistic illumination:

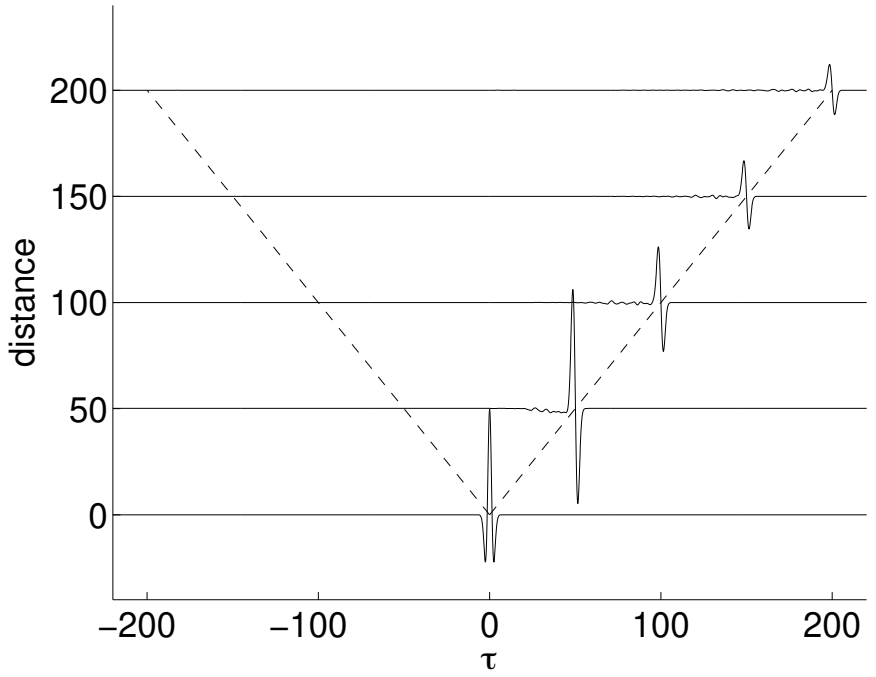




Left: The circles are the noise sources and the triangles are the sensors (the distance between two successive sensors is $\underline{5}$).
Right: Cross correlation $\tau \to C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_j)$ between the pairs of sensors $(\boldsymbol{x}_1, \boldsymbol{x}_j)$, $j = 1, \ldots, 5$, versus the distance $|\boldsymbol{x}_j - \boldsymbol{x}_1|$.

Realistic illumination, high-frequency regime:





Left: The circles are the noise sources and the triangles are the sensors (the distance between two successive sensors is 50).
Right: Cross correlation $\tau \to C^{(1)}(\tau, \boldsymbol{x}_1, \boldsymbol{x}_j)$ between the pairs of sensors $(\boldsymbol{x}_1, \boldsymbol{x}_j)$, $j = 1, \ldots, 5$, versus the distance $|\boldsymbol{x}_j - \boldsymbol{x}_1|$.

Introduction to probability theory

- Gaussian vectors and processes.
- Limit theorems.
- Simulation methods for random processes.

Random vector

- n-dimensional random vector $\boldsymbol{Z}$ = collection of n random variables $(Z_1, \ldots, Z_n)$.
- The distribution of a (continuous) random vector is characterized by the pdf:

$$\mathbb{P}(\boldsymbol{Z} \in A) = \int_A p_{\boldsymbol{Z}}(\boldsymbol{z}) d\boldsymbol{z}, \qquad \forall A \in \mathcal{B}(\mathbb{R}^n)$$

- The random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ is independent if

$$\mathbb{E}\big[\phi_1(Z_1) \cdots \phi_n(Z_n)\big] = \mathbb{E}\big[\phi_1(Z_1)\big] \cdots \mathbb{E}\big[\phi_n(Z_n)\big], \qquad \forall \phi_1, \ldots, \phi_n \in \mathcal{C}_b(\mathbb{R}, \mathbb{R})$$

for all test functions ϕ_j.

- The continuous vector $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ is independent if

$$p(\boldsymbol{z}) = \prod_{j=1}^{n} p_{Z_j}(z_j)$$

- Example: a normalized Gaussian random vector $\boldsymbol{Z}$ has the Gaussian pdf

$$p(\boldsymbol{z}) = \frac{1}{\sqrt{(2\pi)^d}} \exp(-\frac{|\boldsymbol{z}|^2}{2})$$

It is a vector of independent random normalized Gaussian variables.

- Let $\boldsymbol{Z} = (Z_j)_{j=1,\dots,n}$ be a random vector.
- The mean of $\boldsymbol{Z}$ is the vector $\boldsymbol{\mu} = (\mu_j)_{j=1,\dots,n}$:

$$\mu_j = \mathbb{E}[Z_j]$$

The covariance matrix of $\boldsymbol{Z}$ is $\mathbf{C} = (C_{jl})_{j,l=1,\dots,n}$:

$$C_{jl} = \mathbb{E}\big[(Z_j - \mathbb{E}[Z_j])(Z_l - \mathbb{E}[Z_l])\big]$$

- Let $\boldsymbol{\beta} = (\beta_j)_{j=1,\dots,n} \in \mathbb{R}^n$. The random variable $Z = \boldsymbol{\beta}^T \boldsymbol{Z} = \sum_{j=1}^n \beta_j Z_j$ has mean:

$$\mathbb{E}[Z] = \boldsymbol{\beta}^T \boldsymbol{\mu} \quad = \sum_{j=1}^n \beta_j \mathbb{E}[Z_j]$$

and variance:

$$\mathrm{Var}(Z) = \boldsymbol{\beta}^T \mathbf{C} \boldsymbol{\beta} \quad = \sum_{j,l=1}^n C_{jl} \beta_j \beta_l$$

(as a byproduct, this shows that $\mathbf{C}$ is a nonnegative matrix).

- If the variables are independent then the covariance matrix is diagonal. In particular:

$$\mathrm{Var}\Big(\sum_{j=1}^n Z_j\Big) = \sum_{j=1}^n \mathrm{Var}(Z_j)$$

The reciprocal is false (in general).

• A random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ is Gaussian iff any linear combination has normal distribution.

• A Gaussian random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ (write $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$) has characteristic function

$$\mathbb{E}[e^{i\boldsymbol{\lambda}^T \boldsymbol{Z}}] = \int_{\mathbb{R}^d} e^{i\boldsymbol{\lambda}^T \boldsymbol{z}} p(\boldsymbol{z}) d\boldsymbol{z} = \exp\Big(i\boldsymbol{\lambda}^T \boldsymbol{\mu} - \frac{\boldsymbol{\lambda}^T \mathbf{C} \boldsymbol{\lambda}}{2}\Big), \qquad \boldsymbol{\lambda} \in \mathbb{R}^n$$

• A Gaussian random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ with mean $\boldsymbol{\mu}$ and *invertible* covariance matrix $\mathbf{C}$ has the pdf

$$p(\boldsymbol{z}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathbf{C}}} \exp\Big(-\frac{(\boldsymbol{z}-\boldsymbol{\mu})^T \mathbf{C}^{-1} (\boldsymbol{z}-\boldsymbol{\mu})}{2}\Big)$$

• The expectations of high-order moments of a zero-mean Gaussian vector can be expressed as a sum of second-order moments.
For instance, if $\boldsymbol{Z} = (Z_1, Z_2, Z_3, Z_4)$ is a zero-mean Gaussian vector, then

$$\mathbb{E}\Big[\prod_{j=1}^{4} Z_j\Big] = \mathbb{E}\big[Z_1 Z_2\big]\mathbb{E}\big[Z_3 Z_4\big] + \mathbb{E}\big[Z_1 Z_3\big]\mathbb{E}\big[Z_2 Z_4\big] + \mathbb{E}\big[Z_1 Z_4\big]\mathbb{E}\big[Z_2 Z_3\big]$$

Gaussian conditioning

Let $\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}$ be a Gaussian random vector (with $\boldsymbol{X}_1$ of size n_1 and $\boldsymbol{X}_2$ of size n_2):

$$\mathcal{L}\left(\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \right) \quad = \quad \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix} \right)$$

with the means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ of sizes n_1 and n_2, the covariance matrices $\mathbf{R}_{11}$ of size $n_1 \times n_1$, $\mathbf{R}_{12}$ of size $n_1 \times n_2$, $\mathbf{R}_{21} = \mathbf{R}_{12}^T$ of size $n_2 \times n_1$, and $\mathbf{R}_{22}$ of size $n_2 \times n_2$. Then the distribution of $\boldsymbol{X}_1$ conditionally to $\boldsymbol{X}_2$ is Gaussian:

$$\mathcal{L}(\boldsymbol{X}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2) = \mathcal{N}(\boldsymbol{\mu}_1 + \mathbf{R}_{12}\mathbf{R}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \mathbf{R}_{11} - \mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21})$$

Proof: The posterior pdf of $\boldsymbol{X}_1$ given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$ is $p_{\text{post}}(\boldsymbol{x}_1) = \dfrac{p(\boldsymbol{x}_1, \boldsymbol{x}_2)}{\int p(\boldsymbol{x}_1', \boldsymbol{x}_2) d\boldsymbol{x}_1'}$.

Use

$$\begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{Q}^{-1} & -\mathbf{Q}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1} \\ -\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{Q}^{-1} & \mathbf{R}_{22}^{-1} + \mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{Q}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1} \end{pmatrix},$$

where $\mathbf{Q} = \mathbf{R}_{11} - \mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}$ is the Schur complement.

Example. Let us consider the Gaussian vector ($n_1 = n_2 = 1$)

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

with $\rho \in [-1, 1]$ correlation coefficient of X_1 and X_2.

The distribution of X_1 is

$$\mathcal{L}(X_1) = \mathcal{N}(0, 1)$$

If ones observes $X_2 = x_2$, then:

$$\mathcal{L}(X_1 | X_2 = x_2) = \mathcal{N}(\rho x_2, 1 - \rho^2)$$

- the mean of X_1 is attracted (if $\rho > 0$) or repeled (if $\rho < 0$) by the observation.
- the variance of X_1 is reduced (if $\rho \neq 0$).

Extreme cases: $\rho = 0$ (one learns nothing from the observation of X_2) and $\rho = 1$ (ones knows everything once X_2 is observed).

Limit theorems

Limit theorems for sums (1/2)

Let $(X_n)_{n\geq 0}$ be a sequence of independent and identically distributed (i.i.d.) random variables (square integrable). The empirical mean is the random variable:

$$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$$

Its mean is $\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mathbb{E}[X_1]$, and its variance is

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\sigma^2}{n}$$

with $\sigma^2 = \mathrm{Var}(X_1) = \cdots = \mathrm{Var}(X_n)$. This quantity goes to 0 as $n \to \infty$, which means that $\bar{X}_n$ concentrates on the deterministic value $\mathbb{E}[X_1]$:

$$\mathbb{P}\left(|\bar{X}_n - \mathbb{E}[X_1]| \geq \varepsilon\right) \leq \frac{\mathbb{E}[|\bar{X}_n - \mathbb{E}[X_1]|^2]}{\varepsilon^2} = \frac{\mathrm{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon}$$

- (Strong) law of large numbers.

Let $(X_n)_{n\geq 0}$ be a sequence of independent and identically distributed random variables (square integrable). The empirical mean $\bar{X}_n$ converges to $\mathbb{E}[X_1]$ with probability one:

$$\bar{X}_n \overset{n\to\infty}{\longrightarrow} \mathbb{E}[X_1] \quad a.s. \qquad \text{or } \mathbb{P}\left(\lim_{n\to\infty}\bar{X}_n = \mathbb{E}[X_1]\right) = 1$$

Limit theorems for sums (2/2)

- Central limit theorem

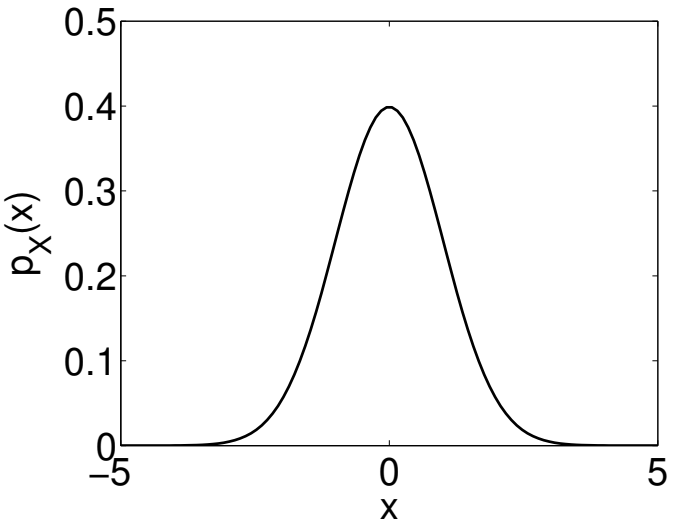
Let $(X_n)_{n\geq 0}$ be a sequence of independent and identically distributed random variables (square integrable), with the mean $\mu \in \mathbb{R}$ and the variance σ^2, $\sigma \in (0, \infty)$. Then

$$\sqrt{n}(\bar{X}_n - \mu) \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \sigma^2)$$

This means that, for any interval $I \subset \mathbb{R}$,

$$\mathbb{P}\left(\sqrt{n}(\bar{X}_n - \mu) \in I\right) \overset{n\to\infty}{\longrightarrow} \int_I p_{0,\sigma^2}(x)dx$$

where $p_{0,\sigma^2}(x) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/(2\sigma^2))$.



Limit theorems for maxima (1/2)

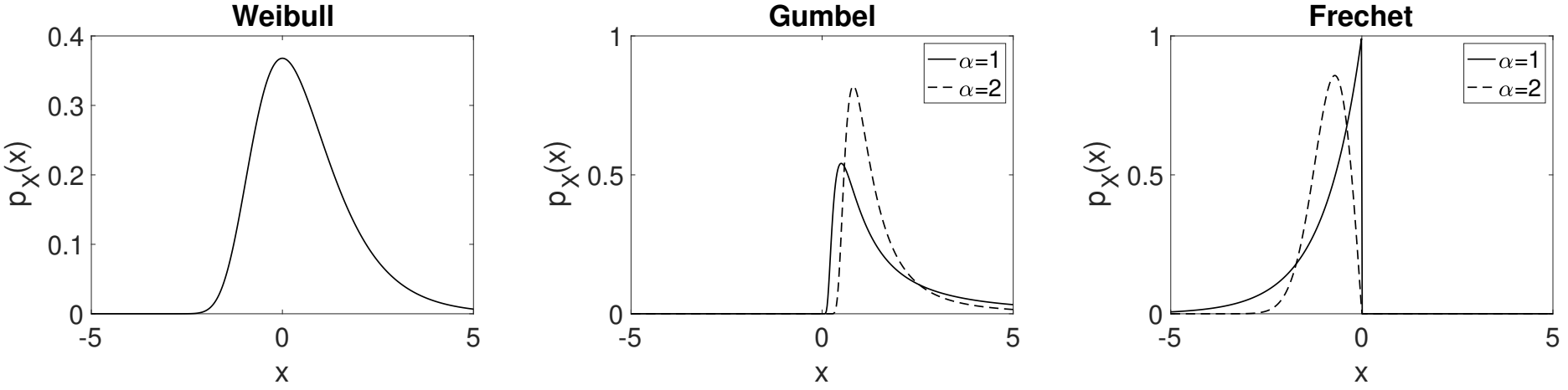
Let $M_n = \max(X_1, \dots, X_n)$ where the X_i are i.i.d. with pdf $p(x)$.

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) = \mathbb{P}(X \leq x)^n = \left(1 - \int_x^\infty p(s)ds\right)^n$$

Theorem (Fisher-Tippett-Gnedenko): Assume that there exist a_n and b_n such that $\mathbb{P}(M_n \leq a_n + b_n x)$ converges to a cdf $F(x)$. Then $F(x)$ belongs to one of the three following types (up to affine scaling):

- $F_1(x) = e^{-e^{-x}}$ with support $\mathbb{R}$ (Weibull).
- $F_{2,\alpha}(x) = e^{-1/x^\alpha}$ with support $\mathbb{R}^+$, $\alpha > 0$ (Gumbel).
- $F_{3,\alpha}(x) = e^{-(-x)^\alpha}$ with support $\mathbb{R}^-$, $\alpha > 0$ (Fréchet).

Conclusion: As $n \to \infty$, $M_n = a_n + b_n Y$ where Y follows one of the three "extreme" distributions.



Limit theorems for maxima (2/2)

Main idea to determine the asymptotic distribution of $M_n = \max(X_1, \dots, X_n)$:
1) Denote $x_{n,\alpha}$ such that $\int_{x_{n,\alpha}}^{\infty} p(s)ds = \frac{\alpha}{n}$. Then $\mathbb{P}(M_n \leq x_{n,\alpha}) = (1 - \frac{\alpha}{n}) \to e^{-\alpha}$.
2) Look for the expansion of $x_{n,\alpha}$ as $n \to \infty$.

Example 1: $X_i \sim \mathcal{U}(0,1)$, i.e. $p(x) = \mathbf{1}_{[0,1]}(x)$.
Then $a_n = 1$, $b_n = 1/n$, and $F = F_{3,1}$.

Example 2: $X_i \sim \mathcal{E}(1)$, i.e. $p(x) = e^{-x}\mathbf{1}_{[0,\infty)}(x)$.
Then $a_n = \ln(n)$, $b_n = 1$, and $F = F_1$.

Example 3: $X_i \sim \mathcal{N}(0,1)$, i.e. $p(x) = (2\pi)^{-1/2}e^{-x^2/2}$.
Then $a_n = \sqrt{2\ln(n)} - \ln\ln n/\sqrt{8\ln(n)} - \ln(4\pi)/\sqrt{8\ln(n)}$, $b_n = 1/\sqrt{2\ln(n)}$, and $F = F_1$.

Simulation of random variables and random processes

Simulation of random variables

How do we generate random numbers with a computer ? There is nothing random in a computer !

Strategy:
- find a pseudo random number generator that can generate a sequence of numbers that behave like independant copies of a random variable with uniform distribution over $(0, 1)$.
- use deterministic transforms to generate numbers with any prescribed distribution using only the uniform pseudo random number generator.

- Pseudo random number generator

A 32-bit multiplicative congruential generator:

$$x_{n+1} = a x_n \bmod b,$$

with $a = 7^5$, $b = 2^{31} - 1$, and some x_0.
This gives a sequence of integer numbers in $\{0, 1, ..., 2^{31} - 2\}$.
The sequence $u_n = x_n/(2^{31} - 1)$ gives a sequence of "quasi-real" number between 0 and 1.
Note: the sequence is periodic, with period $2^{31} - 2$.
This is the generator mcg16807 of matlab (used in early versions).
Today: matlab uses mt19937ar (the period is $2^{19936} - 1$).

Extension: Let F be a cdf. The generalized inverse of F is $F^{-1} : (0,1) \to \mathbb{R}$ defined by:

$$F^{-1}(u) := \inf A_u, \qquad \text{where } A_u := \{x \in \mathbb{R} \text{ such that } F(x) \geq u\}.$$

The generalized inverse always exists because, for any $u \in (0,1)$:
(i) $\lim_{x \to +\infty} F(x) = 1$, therefore $A_u \neq \emptyset$.
(ii) $\lim_{x \to -\infty} F(x) = 0$, therefore A_u is bounded from below.

Result: Let X be a random variable with the cdf $F(x)$. Let U be a random variable with the distribution $\mathcal{U}(0,1)$. Then X and $F^{-1}(U)$ have the same distribution.

- Inversion method.

A little bit of theory:

Result: Let X be a random variable with the cumulative distribution function (cdf) $F(x)$:

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} p(y)dy$$

Let U be a random variable with the distribution $\mathcal{U}(0,1)$. If F is one-to-one, then X and $F^{-1}(U)$ have the same distribution.

Proof: Set $Y = F^{-1}(U)$.

$$\mathbb{P}(Y \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)\,,$$

which shows that the cdf of Y is F.

Extension: Let F be a cdf. The generalized inverse of F is $F^{-1} : (0,1) \to \mathbb{R}$ defined by:

$$F^{-1}(u) := \inf A_u, \qquad \text{where } A_u := \{x \in \mathbb{R} \text{ such that } F(x) \geq u\}.$$

The generalized inverse always exists because, for any $u \in (0,1)$:
(i) $\lim_{x \to +\infty} F(x) = 1$, therefore $A_u \neq \emptyset$.
(ii) $\lim_{x \to -\infty} F(x) = 0$, therefore A_u is bounded from below.

Result: Let X be a random variable with the cdf $F(x)$. Let U be a random variable with the distribution $\mathcal{U}(0,1)$. Then X and $F^{-1}(U)$ have the same distribution.

• Example. Let us write a simulator of an exponential random variable:

$$p(x) = e^{-x} \mathbf{1}_{[0,\infty)}(x)$$

Its cdf is

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ \int_0^x e^{-y} dy = 1 - e^{-x} & \text{if } x \geq 0 \end{cases}$$

whose reciprocal is:

$$F^{-1}(u) = -\ln(1-u)\,.$$

Therefore if U is a uniform random variable on $[0,1]$, then the random variable $X := -\ln[(1-U)]$ obeys the exponential distribution.

Moreover, U and $1-U$ have the same distribution, therefore $-\ln(U)$ also obeys the exponential distribution.

- Simulation of Gaussian random variables.

The inversion method requires the knowledge of the reciprocal cdf F^{-1}.
We do not always have the explicit expression of this reciprocal function.
An important example is the Gaussian distribution such that
$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-s^2/2} ds$ for $x \in \mathbb{R}$.

Box-Muller algorithm:
Let U_1, U_2 two independent random variables with uniform distribution over $[0, 1]$. If

$$X = (-2 \ln U_1)^{1/2} \cos(2\pi U_2)$$

$$Y = (-2 \ln U_1)^{1/2} \sin(2\pi U_2)$$

then the random variables X and Y are independent and distributed according to $\mathcal{N}(0, 1)$.