

Hyperspectral unmixing (HSU)

- Linear model: $x_i = \sum_{k=1}^K a_k^i s_{ki} + n_i$
 - $x_i \in \mathbb{R}^m$: i th observation
 - $a_k^i \in \mathbb{R}^m$: k th endmember (signature)
 - $s_{ik} \in \mathbb{R}$: abundance of the k th endmember in the i th pixel (signature)
 - $n_i \in \mathbb{R}^m$: noise
 - $X = A^* S^* + N$
 - N : m observations and t samples columns ($m \times t$)
 - A^* : endmembers ($m \times n$)
 - S^* : abundances ($n \times t$) - the sources will be assumed to be sparse
 - N : noise and model imperfections ($m \times t$)
 - Goal: retrieve A^* and S^* from the sole knowledge of X (unmix some signals)
 - Assumption $m \ll t$
 - Two cases:
 - Exactly and over-determined: $r \leq m$
 - Under determined: $r > m$
 - Let A^*, S^* be such that $X = A^* S^* = A^* P P^{-1} S^* = AS$
 - Invertible matrix P
 - $A = A^* P$
 - $S = P^{-1} S^*$
 - Infinite number of possible solutions which do not correspond to the true generating A^*, S^*
 - HSU is an ill-posed problem: need to introduce additional information, priors, on the sought after factors A^*, S^*
 - Three main families of priors in HSU:
 - Assume the independence of S (ICA)
 - Assume the sparsity of S (SBSS)
 - Use non-negativity (NMF) of A and S
 - Deep learning methods
 - Each family has its strengths and weaknesses
- ### Independent Component Analysis
- Main principle
 - The sources $(s_k^*)_{k=1..n}$ are assumed to be mutually statistically independent but their mixtures are not: look for independent estimated sources $(\hat{s}_k)_{k=1..n}$
 - Roughly speaking, two types of methods to solve the problem
 - Minimize the mutual information
 - Maximize the non-Gaussianity
 - Various methods: infomax, JADE, FastICA, EFICA, ...
 - Pros (Darmois theorem):
 - All the sources s_k^* are statistically independent
 - There is at most one source following a Gaussian law
 - The mixing matrix A^* is square and full rank
 - Then, the estimated sources \hat{S} correspond to the ones S^* having generated the dataset, up to a (generally inconsequential) scaling and permutation indeterminacy
 - Cons: The Darmois theorem result only holds in the absence of noise, which is impractical
 - The statistical independence of the sources can be a too strong assumption in HSU:
 - If we choose the sources to be abundances: there is no independence since the concentrations have to sum to one
 - If we choose the sources to be the endmembers: unfortunately, the material spectra are often highly correlated

Sparse HSU / blind source separation

- Generally speaking, sparsity amounts to represent a signal with as few variables as possible
- Exact sparsity: a signal $s \in \mathbb{R}^n$ is said to be k -sparse if only $k \ll t$ of its element are non-zeros: $\|s\|_0 = k \ll t$, with the pseudo-norm $\|\cdot\|_0$ being the cardinal of the support of s
 - Most real life signals are not exactly sparse $\|s\|_0 \approx t$
- Approximate sparsity: only a small number k of the signal samples have a large amplitude: the signal can be well approximated by a k -sparse signal
 - This is, for instance, if the sorted magnitudes of the signal samples follow a power law
 - They admit approximately sparse representation in a transformed domain ϕ
 - In practice, it is often better to retain the spatial information (e.g. using a multi-scale transform - wavelet)
 - It is also possible to learn ϕ
- Mixing reduces the sparsity, to recover the abundances we must find the sparsest signals
 - Due to sparsity, the scatter plot of the abundances has a star shape
 - Multiplying the abundances by A^* changes the direction of the axes: recovering the abundances amounts to back-projet the observations on the canonical axes
- Rewrite sparse HSU as a Maximum A Posteriori (MAP) estimation $\arg\max_{A,S} P(A, S|X) = \arg\max_{A,S} P(X|A, S)P(A)P(S) = \arg\min_{A,S} -\log(P(X|A, S)) - \log(P(A)) - \log(P(S))$
 - Assuming a Gaussian noise: $P(X|A, S) \propto e^{-\frac{1}{2\sigma^2}\|X - AS\|_2^2}$
 - An exponential distribution for $S = e^{\beta|S|_{\mathcal{L}_1}}$
 - Uniform distribution for A

- We obtain: $\arg\min_{A,S} \frac{1}{2} \|X - AS\|_2^2 + \lambda \|S\|_1$
- Unfortunately, for any A, S and $\alpha > 1$: $\arg\min_{A,S} \frac{1}{2} \|X - AS\|_2^2 + \lambda \|S\|_1 > \arg\min_{A,S} \frac{1}{2} \|X - A\alpha\alpha^{-1}S\|_2^2 + \lambda \|\alpha^{-1}S\|_1$
 - Scaling indeterminacy
 - The solution $S' = \alpha^{-1}S$ and $A' = \alpha A$ has a lower cost than S, A
 - The larger the α , the better the results
 - Need to modify the cost function and limit the energy of the columns of A so that we do not obtain degenerated solutions:
 - Oblique constraint: we require each column of A to have unit energy: $\iota_{\{i \in [1,n] : \|a_i\|_2 \leq 1\}}(A)$
- $\arg\min_{A \in \mathbb{R}^{m \times n}, S \in \mathbb{R}^{n \times t}} \frac{1}{2} \|X - AS\|_2^2 + \lambda \|S\|_1 + \iota_{\{i \in [1,n] : \|a_i\|_2 \leq 1\}}(A)$ (data-fidelity + sparsity + oblique constraint)
 - λ regularization parameters
 - If needed, we can further use ϕ_S^* a sparsifying transform
 - Challenges: difficult optimization problem
 - Non-smooth (needs advanced optimization tools: proximal operators)
 - Non-convex (non-unique minima)
- Let us fix A : $\arg\min_{S \in \mathbb{R}^{n \times t}} \frac{1}{2} \|X - AS\|_2^2 + \lambda \|S\|_1$. In addition we will first work on a single column of X (and the corresponding column of S): $\arg\min_{S \in \mathbb{R}^n} \frac{1}{2} \|x - AS\|_2^2 + \lambda \|s\|_1$
 - non-smooth due to the l_1 norm
- Let us consider a problem of the form $\arg\min_{s \in \mathbb{R}^n} \frac{1}{2} h_s(s) + \mathcal{G}(s)$
 - With the assumptions:
 - The function $s \rightarrow h_s(s)$ is smooth and has a Lipschitz gradient
 - The $s \rightarrow \mathcal{G}(s)$ is a closed proper convex function
 - The function \mathcal{G} is said to be closed proper convex if its epigraph is a non-empty closed convex set $\text{epi}\mathcal{G} = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} | \mathcal{G}(x) \leq t\}$
- If we only have had the simpler (smooth) problem $\arg\min_{s \in \mathbb{R}^n} h_s(s)$
 - Gradient descent algorithm (with $\gamma < \frac{1}{L}$ the Lipschitz constant of ∇h_s) ($k = 0$)
 - while not converged do:
 - $S^{(k+1)} = S^{(k)} + \gamma \nabla h_s(S^{(k)})$
 - $k \leftarrow k + 1$
 - $\nabla h_s(S^{(k)}) = A^T(AS - X)$ and we can choose $\gamma \approx \frac{1}{\|A^T A\|_2}$
- Let us consider the minimization of a non-differentiable cost function: $\arg\min_{s \in \mathbb{R}^n} \mathcal{G}(S)$
 - A widely used tool is the proximal operator $\text{prox}_{\gamma \mathcal{G}}(s) = \arg\min_{y \in \mathbb{R}^n} \mathcal{G}(y) + \frac{1}{2} \|s - y\|_2^2$
 - Minimized function is strongly convex and not everywhere infinite so it has a unique minimizer
 - The regularization by the l_2 -norm makes the problem easier to solve
 - Compromises between minimizing \mathcal{G} and being close to y
 - Separability: if \mathcal{G} is separable across some variables s_1, \dots, s_k : $\mathcal{G}(s_1, \dots, s_k) = g_1(s_1) + \dots + g_k(s_k)$, then $\text{prox}_{\mathcal{G}}(s_1, \dots, s_k) = (\text{prox}_{g_1}(s_1), \dots, \text{prox}_{g_k}(s_k))$
 - Implies the problem can be minimized separately over the s_i variables
 - If $\mathcal{G}(\cdot)$ is the l_1 -norm $\mathcal{G}(s) = \lambda \|s\|_1$ for $s \in \mathbb{R}^n$, then $(\text{prox}_{\lambda \|\cdot\|_1}(s))_i = S_\lambda(s)_i = \begin{cases} s_i - \lambda & \text{if } s_i \geq \lambda \\ 0 & \text{if } |s_i| \leq \lambda \\ s_i + \lambda & \text{if } s_i \leq -\lambda \end{cases}$ (soft-thresholding operator)
 - If $\mathcal{G}(\cdot)$ is the indicator function of a closed non-empty convex set \mathcal{C} : $\mathcal{I}_{\mathcal{C}}(s) = \begin{cases} 0 & s \in \mathcal{C} \\ +\infty & s \notin \mathcal{C} \end{cases}$, then $\text{prox}_{\mathcal{I}_{\mathcal{C}}}(\cdot) = \Pi_{\mathcal{C}}(\cdot) = \arg\min_{x \in \mathcal{C}} \|x - s\|_2$ (euclidean projector onto \mathcal{C})
 - If $\mathcal{G}(\cdot)$ is the l_1 -norm, but the sparsity is enforced into a transformed domain, with ϕ an orthogonal transform, $\mathcal{G}(S) = \lambda \|S\phi\|_1$ for $s \in \mathbb{R}^n$, then $\text{prox}_{\lambda \|\cdot\|_1}(S) = S_\lambda(S\phi)\phi^T$
 - Fixed point property: the point S^* minimizes $\mathcal{G}(S) \Leftrightarrow \text{prox}_{\mathcal{G}(\cdot)}(S^*) = S^*$
 - Firm non-expansiveness: Let $x, y \in \mathbb{R}^n$, then $\|\text{prox}_{\mathcal{G}}(s_1) - \text{prox}_{\mathcal{G}}(s_2)\|_2^2 \leq (s_1 - s_2)^T (\text{prox}_{\mathcal{G}}(s_1) - \text{prox}_{\mathcal{G}}(s_2))$
 - Proximal point method ($k = 0$)
 - while not converged do:
 - $S^{(k+1)} = \text{prox}_{\mathcal{G}(\cdot)}(S^{(k)})$
 - $k \leftarrow k + 1$
 - Sum of previous cases:
 - Forward-backward splitting method (proximal gradient method) ($k = 0$)
 - while not converged do:
 - $S^{(k+1)} = \text{prox}_{\gamma \mathcal{G}(\cdot)}(S^{(k)} - \gamma \nabla h_s(S^{(k)}))$
 - $k \leftarrow k + 1$
 - Iterative Shrinkage Thresholding Algorithm (ISTA) ($k = 0$)

- $\gamma = \frac{0.9}{\|A^T A\|_2}$
- while not converged do:
 - $S^{(k+1)} = S_{\lambda \gamma}(S^{(k)} - \gamma A^T(AS - X))$
 - $k \leftarrow k + 1$
- Currently, no solution for solving general non-convex optimization problems, only heuristics
- Our problem is non-convex but convex in and separately: multi-convex
- Using the multi-convexity (and other hypotheses), some algorithms exist ensuring the convergence to a fixed point of the cost function (i.e. local min. / max. or saddle point)
- PALM ($k = 0$)
 - while not converged do:
 - $\gamma = \frac{0.9}{\|A^T A\|_2}$
 - $\eta = \frac{0.9}{\|S^{(k+1)}S^{(k+1)T}\|_2}$
 - $S^{(k+1)} = S_{\lambda \gamma}(S^{(k)} + \gamma \nabla A^{(k)}(X - A^{(k)}S^{(k)}))$
 - $S^{(k+1)} = \Pi_{\{i \in [1,n] : \|a_i\|_2 \leq 1\}}(A^{(k)} + \eta(X - A^{(k)}S^{(k+1)})S^{(k+1)T})$
 - $k \leftarrow k + 1$
- Nonnegative Matrix Factorization (+ deep learning methods)**
- There are applications in which it naturally makes sense, such as:
 - Multi/hyperspectral unmixing (abundances S^* are concentrations, endmember signatures A^* are spectra)
 - Text mining (corresponding to words counts)
 - Audio, with works on the modulus of the Fourier transform...
- In contrast to ICA, can cope with noise
- In contrast to sparsity, more theoretical results have been obtained
- $X = A^* S^* + N$, $A^* \geq 0$, $S^* \geq 0$ (elementwise)
 - Given a nonnegative matrix $X \in \mathbb{R}^{m \times t}$, a number of sources n (factorization rank), and a norm distance measure $D(\cdot, \cdot)$ between matrices, compute two nonnegative matrices $\hat{A} \in \mathbb{R}^{m \times n}$ and $\hat{S} \in \mathbb{R}^{n \times t}$ such that $\hat{A}, \hat{S} = \arg\min_{A \geq 0, S \geq 0} \frac{1}{2} \|X - AS\|_F^2$
- Key to understand the problem and design new algorithms
 - The sources are contained into the nonnegative orthant
 - The dataset is contained within a cone, from which the vertices are (under some conditions) the columns of A^*
- $\hat{A}, \hat{S} = \arg\min_{A \geq 0, S \geq 0} \frac{1}{2} \|X - AS\|_F^2 + \iota_{\geq 0}(A) + \iota_{\geq 0}(S)$ (can use PALM to solve)
- Plain NMF does not guarantee the unicity of the solution: is still an ill-posed problem (nonnegativity is not a strong enough prior)
- Near-separability
 - Near separable NMF assumes that $X = A^* S^*$ with S^* a separable matrix
 - Separable matrix: $S^* \in \mathbb{R}^{n \times t}$ separable if $\text{cone}(S^*) = \mathbb{R}_+^t$, with $\text{cone}(S) = \{y \in \mathbb{R}^n | y = Sx, x \geq 0\}$
 - Requires all unit columns to "hide" (up to a scaling) among the columns of S^*
- Separable NMF: Because of the scaling degree of freedom in the decomposition $X = A^* S^*$, separability matrix S^* amounts to assume that there exists an index set $\mathcal{K} \subset [1, t]$ such that $A^* = X(\cdot, \mathcal{K})$, this mixing becomes $X = X(\cdot, \mathcal{K})S^*$
 - The problem becomes identifiable
 - In practice, most datasets do not admit an exact separable NMF decomposition due to noise and model misfit: $X \approx X(\cdot, \mathcal{K})S^*$ or $X = X(\cdot, \mathcal{K})S^* + N$, with $\|n_j\|_2 \leq \epsilon, \forall j$
- Near-separable NMF: $\hat{A}, \hat{S} = \arg\min_{A \geq 0, S \geq 0} \frac{1}{2} \|X - X(\cdot, \mathcal{K})S\|_F^2$, such that $|\mathcal{K}| = n$
 - Three families:
 - Heuristic algorithms: no robustness guaranteed (PPI, N-FINDR)
 - Convex optimization algorithms: good robustness, but costly (MLP, self-dictionary)
 - Greedy algorithms: robust, fast (SPA, FAW, VCA, SNPA)
- SPNA ($R = X, k = 1, \mathcal{K} = \{\}$)
 - while $k \leq r$ do:
 - $p = \arg\max_j \|r_j\|_2$
 - $\mathcal{K} = \mathcal{K} \cup p$
 - $H' = \arg\min_{H \in \mathbb{R}^n} \|X - X(\cdot, \mathcal{K})H\|_F^2$
 - $R = \left(I - \frac{r_p r_p^T}{r_p^T r_p^T} \right) H$
 - $k \leftarrow k + 1$
- SPA ($R = X, k = 1, \mathcal{K} = \{\}$)
 - while $k \leq r$ do:
 - $p = \arg\max_j \|r_j\|_2$
 - $\mathcal{K} = \mathcal{K} \cup p$
 - $R = \left(I - \frac{r_p r_p^T}{r_p^T r_p^T} \right) R$
 - $k \leftarrow k + 1$
 - Pros
 - Easily geometrically interpretable
 - Robustness guaranteed in the presence of noise
 - Very fast (only n iterations)
 - Cons
 - Near separability can be a very strong assumption in HS

- The sufficiently scattered condition (SSC) is in general more general and also leads to identifiable results in the absence of noise
- Requires the column of H to span a "sufficiently" big area in the nonnegative orthant
- SSC 1: $\mathcal{C} = \{x \in \mathbb{R}_+^n | e^T x \geq \sqrt{n-1} \|x\|_2\} \subseteq \text{cone}(S^*)$ with $e^T = (1, \dots, 1)$
- SSC 2: There does not exist any orthogonal matrix Q ($Q^T Q = Q Q^T = I$) such that $\text{cone}(S^*) \subseteq \text{cone}(Q)$, except permutation matrices
- In practice, near-separable algorithm might be too simplistic

Minimum-volume NMF

- If we consider the noiseless mixture $X = A^* S^*$, then it makes sense to look among all the possible A such that $\text{conv}(X) \subset \text{conv}(A)$ for the ones such that $\text{conv}(A)$ is minimal
 - $\text{conv}(A)$ is defined as $\text{conv}(A) = \{x | x = Ay, y \in \mathbb{R}^n, y \geq 0, e^T y = 1\}$
 - Intuitively, minimum-volume NMF looks for a matrix \hat{A} as close as possible to the dataset X columns
 - If S^* is sufficiently scattered, then such a \hat{A} is unique
- To measure the volume of $\text{conv}(A)$: $\frac{1}{n} \sqrt{\det(A^T A)}$
 - Using the logarithm of $\det(A^T A)$ yields better practical results (less sensitive to very small and very large singular values of A)
 - Adding a small positive value δ prevents the logarithm to go to $-\infty$ in the rank deficient case: $\log\det(A^T A + \delta I)$
- There exists different min-vol NMF algorithm, in particular with respect to the normalization used:
 - $S^T e = e$
 - $He = e$
 - $A^T e = e$
- Minimization of the cost function: $\hat{A}, \hat{S} = \arg\min_{A \geq 0, S \geq 0} \frac{1}{2} \|X - AS\|_F + \lambda \log\det(A^T A + \delta I)$
 - The logarithmic function is concave
 - A way to apply to proximal algorithms is to find a convex majorizer of the cost function (Maximization Minimization algorithms)
 - Note however, that depending on the chosen normalization of the factors, the proximal operators might not be explicit
 - Pros
 - It is more general than near-separability
 - The sufficiently scattered condition is practically rather mild
 - It leads to better results in practice
 - Cons
 - Solving exactly min-vol NMF is NP-hard (alleviating the near-separability thus comes at a price)
 - An open question is thus when is it possible to solve it efficiently? And how to do it in practice?
 - The behavior of min-vol NMF is not well understood in the presence of noise
- Extension to deep learning
 - The most basic training function is $\arg\min_{\hat{A}, \hat{S}} \frac{1}{2} \|X - \hat{X}_f\|_F^2$
 - The use of a single linear layer makes the auto-encoder decoder architecture quite interpretable: the latent space is expected to correspond to S^* ($H \approx S^*$) since the output is $\hat{X} = WH$ and $X \approx \hat{X}$. The A^* matrix is approximated by the weights W . (need for regularization)
 - Nonnegativity of the H coefficients
 - Sum-to-one of the coefficients in each pixel of H
 - Both constraints are often implemented using a softmax nonlinearity at the end of the encoder if $h \in \mathbb{R}^n, S(h)_i = \frac{e^{h_i}}{\sum_{i=1}^n e^{h_i}}$
 - The problem is still ill-posed, requiring several optimization trick to avoid bad solution

Introduction to SAR imaging

Principle of radar acquisition

- Physic measurement
 - Passive sensors
 - Optic domain
 - Infra-red domain
 - Active sensors
 - RADAR (radio detection and ranging)
 - LIDAR
- Radar imaging
 - Emission of an electro-magnetic wave by an antenna
 - Recording of the signal backscattered by the ground by the antenna
 - Characteristics
 - Lateral viewing
 - Mono-static sensor
- SAR systems
 - Avantages
 - All time sensor (own source of illumination)

- All weather sensor (electro-magnetic wave penetrating the clouds)
- Phase information linked to the distance from sensor to target ($\phi = \frac{2\pi}{\lambda}r$)
- Complementary information on optic images
 - Sensitivity to different properties of the ground (roughness, soil moisture, dielectric properties...)
 - Sensitivity to geometric properties of the objects
 - Penetration capability of some surfaces
- Drawbacks
 - Speckle (strong radiometric noise)
 - Sensitivity to geometric distortions (lateral viewing and high incidence angle)

Examples of SAR images and backscattering mechanisms

- Parameters:
 - Depends on the radar sensor
 - Wavelength of the electro-magnetic wave
 - Incidence angle
 - Polarization of the wave
 - Depends on the surface
 - Roughness (compared to the wavelength)
 - Dielectric properties (moisture, ...)
 - Geometric shape
 - Penetration depth: depends on the wavelength and surface dryness (increases with λ and with dryness)
- Backscattering mechanisms and roughness:
 - Smooth surface: regular reflection
 - Flat and horizontal surface:
 - Reflection and refraction mechanisms
 - Case of metallic or water surfaces: full reflection
 - Dihedral configuration: multiple backscattered signals with the same distance to the sensor (influence of the orientation of the dihedral compared to the incidence direction)
 - Trihedral configuration: corner reflector (calibration purposes) backscattered signal ($\sigma_{\text{br}} = \frac{4\pi A^2}{3\lambda^2}$)
 - Rough surface: irregular reflection
 - Volume scattering: irregular reflection and absorption
 - Vegetation: highly complex interaction with branches trunk and ground
 - Backscattering coefficient \propto vegetation volume (biomass)
 - Wavelength penetration $\propto f^{-1}, \lambda$
- Object appearance on a SAR image:
 - Bright targets: trihedral/dihedral structures (man-made objects, urban areas, rocks, ...)
 - Surface area
 - Depends on the roughness (specular/lambertian backscattering)
 - Depends on the geometric configuration (object and incidence direction)
 - Dielectric properties (water content, ...)
 - If many elementary scatterers inside the resolution cell (rough surface): speckle

SAR image acquisition

- Radar imaging:
 - Active imaging by electro-magnetic pulse emission (5 - 10 GHz) which uses an antenna mounted on an airborne vehicle to emit radar pulses
 - Depending on the scene, the emitted pulse is backscattered specularly, or in a Lambertian way and part of the incident energy is backscattered towards the antenna
 - The time at which the signal is received t , is linked to the distance between the sensor and the target by $t = \frac{2R}{c}$
 - Sampling in time is sampling in distance
 - One pulse gives a line of the image: pixels are obtained by time sampling of the backscattered pulse
 - The incident direction is called the range direction
 - When the plane moves the antenna illuminates another area with another pulse, creating another line forming the 2D image
 - The sensor direction is called azimuth direction
 - The synthetic aperture is obtained by combining numerically the backscattered signals of a common area improving drastically the resolution
- Time, range, ground
 - Time sampling: Δt
 - Distance sampling: $\Delta r = \frac{c\Delta t}{2}$
 - Ground range sampling: $\Delta x = \frac{\Delta r}{\sin(\theta)}$ (depends on the local incidence angle in the antenna swath)
- Range resolution

- Pulse duration τ
- To be separated, two points M_1 and M_2 backscattering at times t_1 and t_2 should verify: $t_2 - t_1 \geq \tau \Leftrightarrow 2\frac{R_2}{c} - 2\frac{R_1}{c} \geq \tau$
 - $\Delta r \geq \frac{c\tau}{2}$
- Improving resolution by chirp emission:
 - $S_e(t) = e^{j2\pi f_0 t} e^{j\pi K t^2}$, $t \in \left[-\frac{\tau}{2}, \frac{\tau}{2}\right]$, compression factor K , signal duration τ
 - Linear frequency modulation (instantaneous frequency): $f_i = \frac{1}{2\pi} \frac{\partial \phi}{\partial t} = f_0 + Kt$
 - Bandwidth: $B_{\text{chirp}} = K\tau$
 - Matched filter: convolution by $s_e^*(-t)$ pf the received signal
 - The resulting signal is a cardinal sine after matched filtering $\approx \frac{\sin(\pi K \tau t)}{\pi K \tau t}$
 - Apparent duration: $\tau' = \frac{1}{K\tau} = \frac{1}{B_{\text{chirp}}}$
 - $\delta_r = \frac{c}{2B_{\text{chirp}}}$
- Azimuth resolution
 - Aperture of the antenna $\beta = \frac{\lambda}{L}$, length of the antenna in the azimuth direction L
 - Spread of the swath in the azimuth direction: $\delta_{\text{az}} = R \frac{\lambda}{L}$, distance of the sensor to the middle of the swath R
 - Visibility of a target
 - Start: $y_{\text{min}} = -\frac{R\lambda}{2L}$
 - End: $y_{\text{max}} = \frac{R\lambda}{2L}$
 - Closest point of approach (PCA): $y = 0$
 - Synthetic aperture
 - Azimuth resolution after aperture synthesis: $\delta_y = \frac{L}{2}$
 - $\beta = \frac{2\lambda}{L}$
 - $L_s = \frac{2R\lambda}{L}$

- Point Spread Function: $\text{PSF} = \frac{\sin(\pi B_{\text{chirp}} t)}{\pi B_{\text{chirp}} t} \frac{\sin(\pi B_y y)}{\pi B_y y} \propto \frac{\sin(\pi B_r r)}{\pi B_r r} \frac{\sin(\pi B_y y)}{\pi B_y y}$

Relief and phase information

- Distance sampling
 - The closer a point from the sensor, the sooner it will appear in the image
 - The order of appearance of the points in the image can be different from the order on the ground
 - The distance between points in the image is modified depending on the ground relief
 - Slope facing the sensor: foreshortening
 - Slope on the opposite side: dilation
 - Overlay: if the slope of the relief α is superior to the incidence angle θ , the points on the ground are inverted and there is a mixing of target responses
 - Shadow phenomenon: some parts of the ground are not illuminated (back-slope) when $\alpha' > \frac{\pi}{2} - \theta$

- Phase and geometry: $\phi(t) = 2\pi f_0 t + \phi_{\text{br}} = 4\pi \frac{r}{\lambda} + \phi_{\text{br}}$, intrinsic contribution of the pixel to the phase ϕ_{br} (phase shift due to cell scatterer organization)
- Interferometric phase: $\psi_{1,2} = 4\pi \frac{R_{\text{both}}}{R \sin(\theta) \lambda} h$
- Interferometric processing chain
 - Acquisition of two SAR images with a small difference of incidence angle
 - Fine registration of the two images
 - Computation of the phase difference
 - Phase unwrapping

Statistics of coherent imaging

Introduction

- Noise models in image processing
 - Most usual noise model: additive white Gaussian noise
 - Low light conditions: shot noise or Poisson noise (when the number of collected photons is small such as in fluorescence microscopy or astronomy); extension to Poisson-Gaussian noise when shot noise and thermal noise are mixed
 - Coherent imaging: speckle noise

Speckle and Goodman model

- Speckle phenomenon
 - Resolution cell much bigger than the wavelength: many elementary scatterers inside a resolution cell
 - Coherent sum of the waves:
 - Each scatterer backscatters with the e.m. wave
 - Vectorial addition in the complex plane of the backscattered waves
 - Interference phenomenon

- Statistical models
 - No acces to the organization of the elementary scatterers inside the resolution cell
 - Modelling with random variables at the pixel level
 - Why develop models for the backscattered electro-magnetic field?
 - Prediction of performances

- Adaptation of the processing to take into account the statistical models
- Considered random variables
 - Active sensor: emits a wave and measures its echoes
 - SAR: At each pixel: complex amplitude of the echo $z = Ae^{j\phi}$
 - Amplitude: $A = |z|$, Intensity: $I = A^2 = |z|^2$, Phase: $\phi = \arg z$
- Goodman model: coherent summation of N punctual echos: $z = \frac{1}{\sqrt{N}} \sum_{i=1}^N z_i$, $z_i \in \mathbb{C}$, $z_i = a_i e^{j\phi_i}$
 - Amplitude $|z_i|$ and phase $\arg z_i$ independent for each scatterer
 - Amplitude $|z_i|$ and phase $\arg z_i$ iid
 - Phase $\arg z_i$ uniformly distributed on $[-\pi, \pi]$ (non informative)
 - Implies that the surface is rough compared to the wavelength: $\delta > \frac{\lambda}{8 \cos(\theta)}$
 - $\mathbb{E}(\Re(z)) = \mathbb{E}(\Im(z)) = 0$
 - $\mathbb{E}(\Re(z)^2) = \mathbb{E}(\Im(z)^2) = \sigma^2$
 - Law of numbers: $p(z|s) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|z|^2}{2\sigma^2}\right)$
 - $p(\phi|R) = \frac{1}{2\pi}$
 - I exponentially distributed: $p(I|R) = \frac{1}{R} \exp\left(-\frac{I}{R}\right)$, $\mathbb{E}(I) = R$, $\text{Var}(I) = R^2$
 - $R = 2\sigma^2$ represents the reflectivity of the considered area
 - Noise is said multiplicative
 - Amplitude is distributed according to a Rayleigh distribution
 - $p(A|R) = \frac{2A}{R} \exp\left(-\frac{A^2}{R}\right)$, $\mathbb{E}(A) = \sqrt{\frac{\pi}{2}} R$
 - The mode does not correspond to the mean
 - The distribution has a heavy tail
 - Mean and standard deviation are proportional
- Measuring the heterogeneity of an area
 - Mean and standard deviation are proportional for a physically homogeneous area (both for intensity and amplitude data)
 - The standard deviation no longer measures the local heterogeneity
 - Standard deviation must be normalized by the mean: coefficient of variation
 - For intensity data on a homogeneous area (R constant): $\gamma_I = \frac{\sigma_I}{\mathbb{E}(I)} = 1$
 - For amplitude data on a homogeneous area (R constant): $\gamma_A = \frac{\sigma_A}{\mathbb{E}(A)} = \sqrt{\frac{\pi}{2}} - 1 \approx 0.523$

- Multi-look processing
 - Averaging L iid samples divides the variance by L
 - $X_{\text{ML}} = \frac{1}{L} \sum_{i=1}^L X_i$, $\mathbb{E}(X_{\text{ML}}) = \mathbb{E}(X_i)$, $\text{Var}(X_{\text{ML}}) = \frac{\text{Var}(X_i)}{L}$
 - $I = I_{\text{ML}} = \frac{1}{L} \sum_{i=1}^L I_i$, $p(I|R, L) = \frac{I^{L-1}}{\Gamma(L) R^L} \exp\left(-\frac{I}{R}\right)$
 - $A = A_{\text{ML}} = \sqrt{I_{\text{ML}}}$, $p(A|R, L) = \frac{2I^{L-1} A^{2(L-1)}}{\Gamma(L) R^L} \exp\left(-\frac{LA^2}{R}\right)$
 - $\gamma_{I_{\text{ML}}} = \frac{1}{\sqrt{L}}$, $\gamma_{A_{\text{ML}}} = \frac{0.523}{\sqrt{L}}$
 - Real data
 - In practice, samples are not iid: the number of looks will be less than announced and no more an integer value (ENL - Equivalent Number of Looks)
 - $\text{ENL} = \frac{1}{\gamma_{\text{ML}}}$
- Spatial multi-looking
 - Local averaging and subsampling
 - Drawback: loss of resolution
 - Advantage: the number of vertical and horizontal samples is chosen to obtain square pixels
 - Alternative: mean filter with moving window (no explicit loss in resolution)
- Temporal multi-looking (remote sensing context)
 - Use of the new acquisitions of the same area after a cycle of the satellite
 - Drawback: cost of acquisitions, temporal changes
 - Advantage: no loss in resolution
 - Very good solution for stable features in the image

- Multiplicative noise model
 - $I = RS$, normalized speckle S
 - $\mathbb{E}(S) = 1, \text{Var}(S) = \frac{1}{L}$, $p(S|L) = \frac{\mu^L}{\Gamma(L)} S^{L-1} \exp(-LS)$
 - $p(I = RS|L) = \int p(R) p(I|R, L) dR$
 - Coefficient of variation of the scene $\gamma_S^2 = \frac{I^2 - \mu_S^2}{I^2 - \mu_S^2}$, $\gamma_S^2 = \frac{1}{L}$
 - Lee filter: $\hat{R} = \bar{I} + k(I - \bar{I})$, $k = 1 - \frac{\gamma_S^2}{\gamma_I^2}$
 - $-\log p(I|L) = \frac{L}{R} + L \log(R) - (L-1) \log(I) + \text{cte}$
 - To get rid of the multiplicative noise (homomorphic approach): $\log(I) = \log(R) + \log(S)$
 - $\log(I)$ follows a Fisher-Tippett distribution, which can be approximated as additive white Gaussian noise: not very good for small L (asymmetry towards lower values), not centered (debiasing step is needed)
 - Isolated dark pixels appear

Extension to vectorial data

- Interferometry: 2 SAR images
 - Scattering vector $k = (z_1, z_2)^t$
 - Covariance matrix $\Sigma = \begin{pmatrix} R_1 & \sqrt{R_1 R_2} \rho_{1,2} \\ \sqrt{R_1 R_2} \rho_{1,2} & R_2 \end{pmatrix}$
 - Reflectivity of channel k : $R_k = \mathbb{E}(|z_k|^2)$
 - Coherence between the two acquisitions $\rho_{1,2} = \frac{\sqrt{\mathbb{E}(z_1 z_2^*) \mathbb{E}(\bar{z}_1 \bar{z}_2^*)}}{\sqrt{\mathbb{E}(z_1^2) \mathbb{E}(\bar{z}_2^2)}} = D_{1,2} e^{j\phi_{1,2}}$ and interferometric phase $\phi_{1,2} = \phi_1 - \phi_2$
- Polarimetry: 3 SAR images
 - $p(k|\Sigma) = \frac{1}{\pi^3 |\Sigma|} \exp(-k^t \Sigma^{-1} k)$
 - $C = \frac{1}{\pi} \sum_{k=1}^2 k_t k_t^t$, $I_k = C e_k$
 - $L \geq K$: $p(C|\Sigma, L) = \frac{L^{\frac{K(K+1)}{2}} \exp(-L C^t \Sigma^{-1} C)}{\Gamma_K(K) \Gamma_L^{\frac{K+1}{2}}}$
 - Empirical complex coherency: $d_{k,l} e^{j\phi_{k,l}} = \frac{C_{k,l}}{\sqrt{C_{k,k} C_{l,l}}}$

Applications industrielles

IDEMIA

End-to-End Trackers

- Transition from traditional tracker-by-detection methods to end-to-end tracking.
- Deep learning techniques for integrated detection and tracking.
- Overview of models like DETR, MOTR, CAROQ, and MOTIP.
- Performance evaluation using benchmarks like BDD100K, DanceTrack, and MOT17.
- Detection Transformer (DETR): The use of a transformer architecture with object queries and a bipartite matching process to enhance detection and tracking performance, representing a significant advancement in object tracking by handling complex motions and simplifying training processes.
- Object Detection for Video Analysis
 - Methods for object detection: Traditional, two-stage, and one-stage detectors.
 - Detailed discussion of YOLO and SSD architectures.
 - Applications in video analytics and biometric recognition.
 - Challenges related to fairness and deep fakes in biometric identification.
 - YOLO (You Only Look Once): Its real-time performance and fully convolutional structure that enables entire image analysis in a single forward pass, significantly improving speed and efficiency in object detection.

INCEPTO

AI Solutions Platform for Radiology

- Integration of AI with existing PACS and RIS infrastructures.
- AI applications for tasks like fracture detection, chest X-rays, breast cancer screening, and brain MRIs.
- Enhancements in diagnostic accuracy, productivity, and workflow efficiency.
- Emphasis on clean databases, robust data annotation, and clinical validation.
- ARVA (Automatic measure of Aortic Aneurysm): The AI algorithm's capability to automatically measure aortic diameters, incorporating advanced techniques like ResNet and guided filtering for handling complex clinical scenarios and providing high accuracy in comparison to expert measurements.

Annotation Strategy for Machine Learning in Medical Imaging

- Challenges and strategies in dataset creation and annotation.
- Types of annotations: simple class and advanced segmentation.
- Methods like active learning and self-supervised learning to optimize annotation.
- Visual interpretation techniques for machine learning models.
- Active Learning: Its ability to select the most informative data points for annotation, significantly reducing the annotation workload and ensuring that the most valuable data is used for training, thus enhancing model performance with fewer resources.

- DxO: Practical Aspects of Image Denoising
 - Noise types in image sensors: Photon shot noise, dark current noise, read noise.
 - Conversion from photons to gray levels and impact of noise.
 - Raw conversion process and denoising techniques.
 - Balance between noise reduction and detail preservation.
 - Wavelet-Based Denoising: This technique's multi-scale analysis and selective noise reduction capabilities, which maintain a critical balance between reducing noise and preserving important image details, making it suitable for high-precision applications like medical imaging.

THALES : Imagerie SAR dans un Contexte Aéroporté

- Principles of SAR imaging and its advantages (all-weather, high resolution).
- Antenna scanning modes: Spot, strip, and quicklook.
- Data processing and image formation techniques.
- Applications in terrain mapping, position estimation, and object recognition.
- Challenges related to hardware constraints, data acquisition costs, and image quality measurement.
- EPiK-SAR Demonstrator: A flexible prototype radar system for airborne data acquisition that supports a wide range of test scenarios at lower costs, equipped with advanced hardware for developing and validating SAR imaging techniques in real-world conditions.