# Spatialized Bayesian inference and uncertainty quantification on constrained co-domains with Gaussian Processes.

## Application to the simplex.

**Lucas Drumetz**
lucas.drumetz@imt-atlantique.fr

IMT Atlantique, Dept. Mathematical and Electrical Engineering, Lab-STICC, Brest, France

MASCOT-NUM, ENSAI, Rennes, 1-3 April 2026





# GENESIS AI Chair

GENerative modeling with kErnelS for Inverse problemS



UQ in Bayesian inverse problems in imaging:

- Sample spatialized posterior distributions with generative models as priors
- Learn prior distributions from data
- Handle constrained valued-images

Example applications:

- Interpolation/forecasting of oceanic variables at the global scale
- Source separation in optical remote sensing
- Geometric ML

$\rightarrow$ This presentation: handling UQ for images with constrained values, using Gaussian Processes.

*Aitchison Geometry on the Simplex for Uncertainty Quantification in Bayesian Hyperspectral Image Unmixing*, H. Blondel, L. Drumetz, T. Chonavel., Submitted to IEEE ICIP 2026.

# Outline

# Bayesian inverse problems

Recover parameter $\mathbf{x} \in \mathbb{R}^N$ given observations $\mathbf{y} \in \mathbb{R}^n$ (typically $n < N$), a forward model and a prior distribution on $\mathbf{x}$.

$$p_{X|Y}(\mathbf{x}|\mathbf{y}) \propto p_{Y|X}(\mathbf{y}|\mathbf{x}) p_X(\mathbf{x})$$

Classically, $p_{X|Y}(\mathbf{x}|\mathbf{y})$ has a closed form but it is intractable, but requires:

- Sampling algorithms
- Variational inference

to compute expectations using $M$ i.i.d. samples from the posterior:

$$\mathbb{E}[f(X)|Y = y] = \int_{\mathbb{R}^n} f(\mathbf{x}) p_{X|Y}(\mathbf{x}|\mathbf{y}) d\mathbf{x} \approx \frac{1}{M} \sum_{i=1}^{M} f(\mathbf{x}_i)$$

$\rightarrow$ obtain posterior averages, covariances... For example, taking $f = \mathrm{Id}$, we get the posterior mean:

$$\mathbb{E}[X|Y = y] = \int_{\mathbb{R}^n} \mathbf{x} p_{X|Y}(\mathbf{x}|\mathbf{y}) d\mathbf{x} \approx \frac{1}{M} \sum_{i=1}^{M} \mathbf{x}_i$$

which is the minimum (Euclidean) mean square error estimator.

# Objectives/requirements

In this talk, $\mathbf{x}$ and $\mathbf{y}$ are spatialized and are observed at locations $\mathbf{u} \in \mathbb{R}^d$.

$$\mathbf{x} \in \mathcal{C}^N, \quad \mathbf{y} \in \mathcal{D}^n,$$

for some $\mathcal{C} \subset \mathbb{R}^P, \mathcal{D} \subseteq \mathbb{R}^L$

- We want to be able to include any location $\mathbf{u} \in \mathbb{R}^q$ inside $\mathbf{x}$ (In imaging, typically locations are on a grid)
- $\mathbf{y}$ can be observations of the same quantity at some locations, or live in another space entirely.
- We want to handle constrained $\mathbf{x}$, i.e. images with constrained values.
- Need for a parametric prior distribution of $\mathbf{x}$ to handle spatial correlations

# Constrained inference

Assumption: $\mathcal{C} \subset \mathbb{R}^P$ is diffeomorphic to a Euclidean space $\mathbb{R}^D$, i.e. there exists a diffeomorphism $\phi : \mathcal{C} \to \mathbb{R}^D$.

$$\mathbf{z} = \phi(\mathbf{x})$$

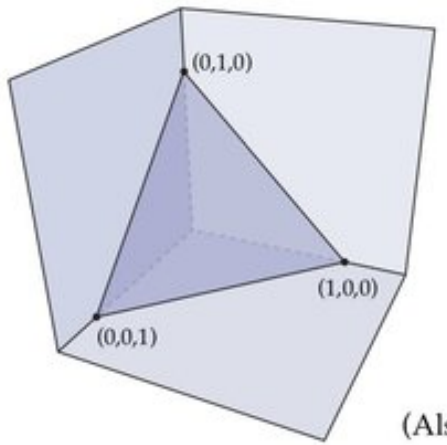
This encompasses many sets of practical interest:

- The *strictly* positive orthant $\{\mathbf{x} \in \mathbb{R}^P, x_i > 0\}$
- **The *interior* of the probability simplex with $P$ vertices**, $\Delta_{P-1} = \{\mathbf{x} \in \mathbb{R}^P, x_i > 0, \sum_{i=1}^P x_i = 1\}$ and variants
- The *interior* of the unit disk: $\{\mathbf{x} \in \mathbb{R}^P, ||\mathbf{x}|| < 1\}$
- The set of symmetric positive definite matrices: $\{\mathbf{X} \in \mathbb{R}^{P \times P}, \forall \mathbf{v} \in \mathbb{R}^P, \mathbf{v}^T \mathbf{X} \mathbf{v} > 0\}$
- and others...

# Running Example: the probability simplex

**Definition.** The *standard n-simplex* is the collection of points

$$\sigma := \left\{ (x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \,\middle|\, \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0\ \forall i \right\}.$$




(Also known as the "probability simplex.")

2-simplex and its natural embedding in $\mathbb{R}^3$. From [Zheng et al, AAAI, 2026]

Applications:

- compositional data analysis
- modeling categorical probability distributions (e.g. classification)
- spectral unmixing

# Related work

- Manifold-valued GP for SPD matrices/punctured sphere [1]
- Generative models on the simplex (flow matching/GP) with the same idea [2](very recent!)

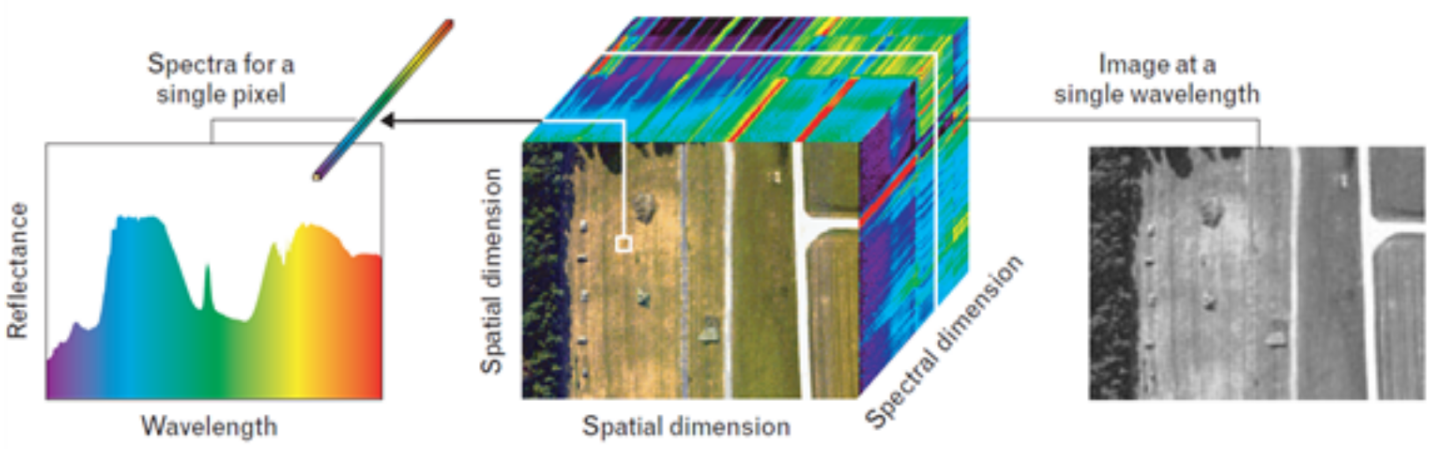
Our work:

- Generic construction of GP with values in constrained sets
- Application to the simplex and source separation
- Geometric implications
- Geometry-compliant UQ metrics and visualization for $P = 3$.

[1] A. Mallasto and A. Feragen. Wrapped gaussian process regression on riemannian manifolds. In Proceed- ings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5580–5588, 2018.

[2] Williams, B., Tetali, H. V., Klami, A., & Hartmann, M. (2026). Simplex-to-Euclidean Bijection for Conjugate and Calibrated Multiclass Gaussian Process. arXiv preprint arXiv:2603.16621.

# Hyperspectral imaging

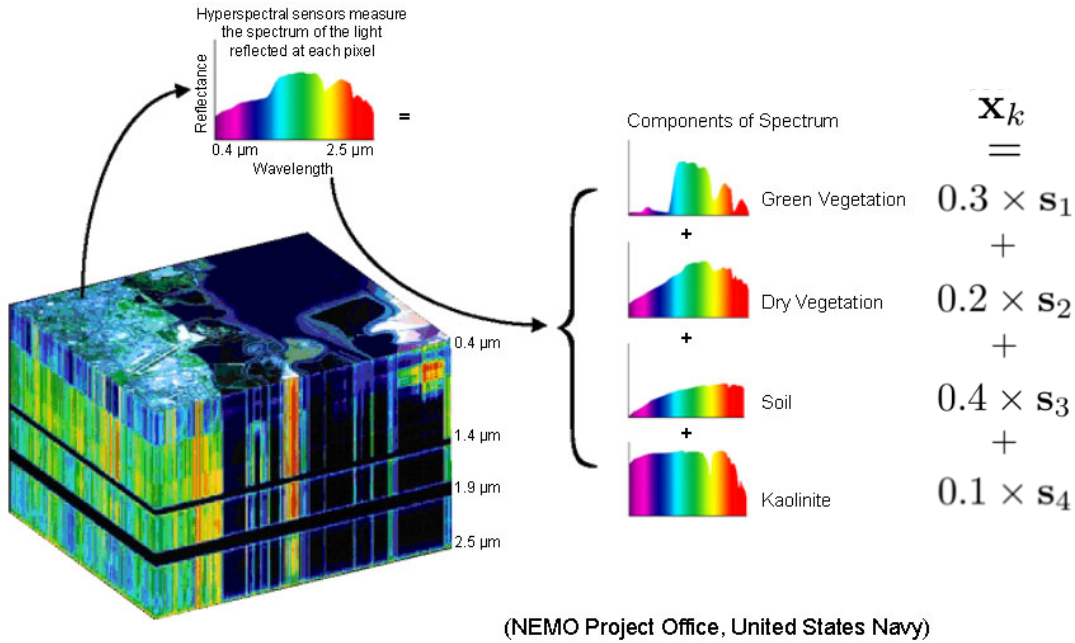



Hyperspectral imaging concept.

- Each pixel: reflectance spectrum for many contiguous and narrow wavelengths (visible and near-IR).
- Also a collection of gray-level reflectance images for each wavelength.
- The spectrum characterizes a material, but because of the limited spatial resolution, mixed pixels are present.

Applications in: **Remote sensing** – environment monitoring, geology, precision agriculture, planetary science, defense ...
but also food processing, chemometrics, document analysis...

# Spectral Unmixing




Linear Spectral Unmixing

A blind source separation problem[3] whose goals are to:

- extract the signatures of the *materials* (***endmembers***) in the image
- estimate their *relative proportions* (***fractional abundances***) in each pixel.

[3] Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., & Chanussot, J. (2012). Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. IEEE journal of selected topics in applied earth observations and remote sensing, 5(2), 354-379.

# Spectral Unmixing

Each pixel $\mathbf{y}_n \in \mathbb{R}^L$ of the image is decomposed as:

$$\mathbf{y}_n = \mathbf{S}\mathbf{x}_n + \mathbf{e}_n$$

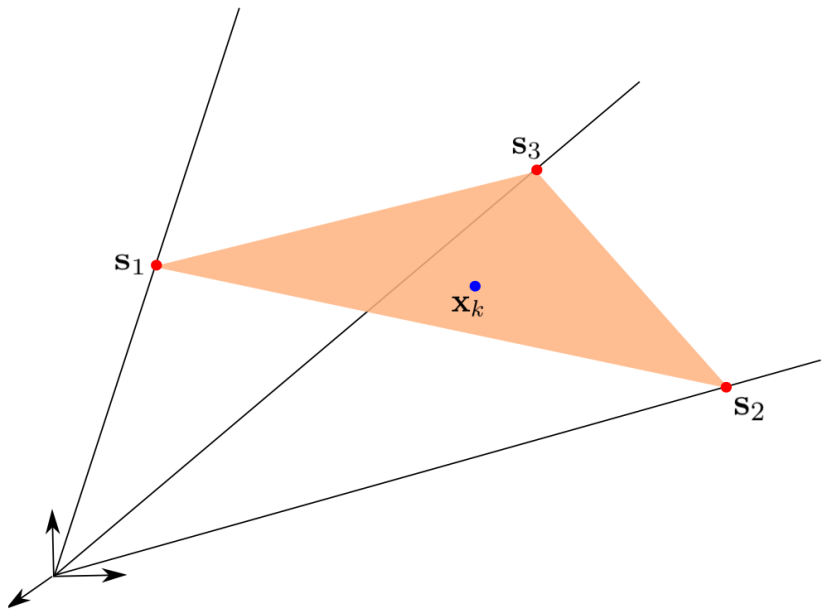
with $\mathbf{x}_n \in \Delta_{P-1}$ for all $n$. [Bioucas-Dias et al. 2012]

$$\mathbf{Y} = \mathbf{S}\mathbf{X} + \mathbf{E}$$

with $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N] \in \mathbb{R}^{L \times N}$, $\mathbf{E} \in \mathbb{R}^{L \times N}$ i.i.d Gaussian noise, $\mathbf{S} \in \mathbb{R}^{L \times P}$ and $\mathbf{X} \in \Delta_{P-1}^N \subset \mathbb{R}^{P \times N}$.

$L$: number of wavelengths
$N$: number of pixels
$P$: number of endmembers

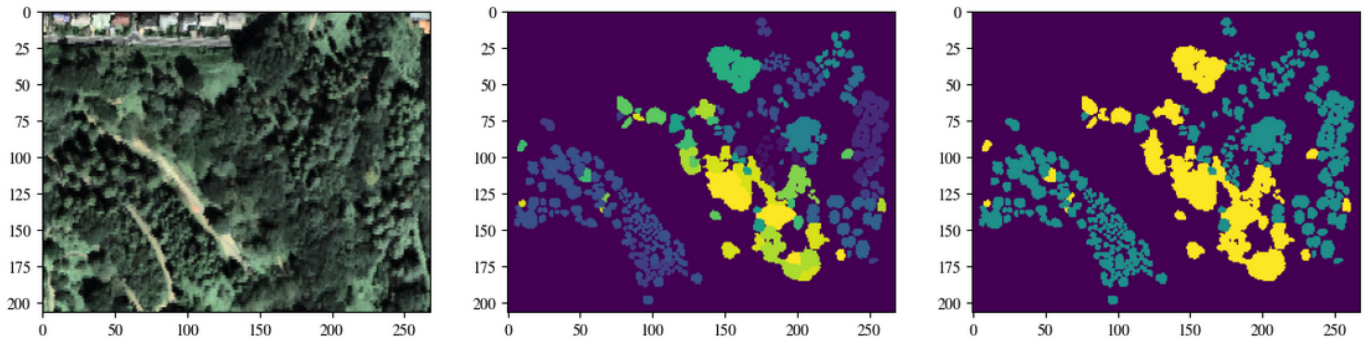
$\mathbf{Y}$ : data matrix
$\mathbf{S}$ : endmember matrix
$\mathbf{X}$ : abundance matrix



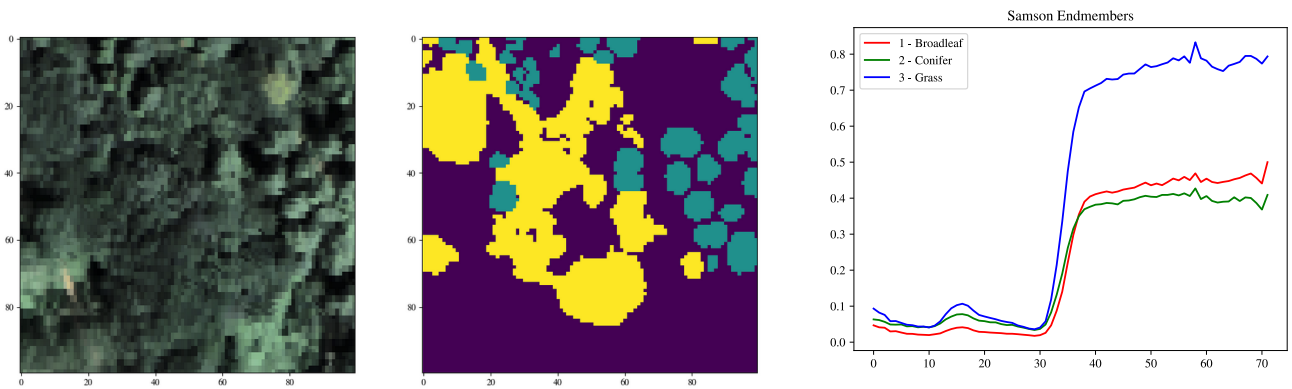

Geometric interpretation of the noiseless Linear SU problem, in the case of 3 endmembers in a 3D ambient space.

# Data and problem

Tama Forest Garden, near Tokyo, Japan. 72 spectral bands (wavelengths) in the visible and near infrared. 1m spatial resolution. Partial (classification) GT.[4].



RGB image (left), GT for several tree species (middle), organized into broadleaf (yellow) and conifer trees (green)




100×100 crop. We see that the sensor introduces spatial artifacts.

Task: separate conifer, broadleaf trees, grass (1 signature given per material).

[4] T. Matsuki, N. Yokoya, and A. Iwasaki, "Hyperspectral tree species classification of japanese complex mixed forest with the aid of lidar data," IEEE JSTARS, vol. 8, no. 5, pp. 2177–2187, May 2015.

# Spectral Unmixing processing chain

- Blind setup: $P$, $\mathbf{S}, \mathbf{X}$ unknown $\rightarrow$ elegant but unrealistic
- "Supervised setup": Here, $\mathbf{S}$ is known or partially known

Need for ecology, environmental monitoring, planetology... But very ill posed problem:

- Variability of $\mathbf{S}$
- Proximity/correlation between materials in practical scenarios.
- Ground truth is scarce, expensive and unreliable for remote sensing images.

Classically: Constrained Least Square estimation (MAP) of $\mathbf{x}_n$, in each pixel:

$$\hat{\mathbf{X}} = \underset{\mathbf{x} \in \Delta_{P-1}}{N\arg\min} \quad ||\mathbf{Y} - \mathbf{S}\mathbf{X}||_F^2$$

Need for tools and diagnostics for feasibility, assessing the difficulty of the problem for practitioners.

# Bayesian modeling for spectral unmixing

$$p_{X|Y}(\mathbf{x}|\mathbf{y}) \propto p_{Y|X}(\mathbf{y}|\mathbf{x})p_X(\mathbf{x})$$

with $p_{Y|X}(\mathbf{y}|\mathbf{x}) \sim \mathcal{N}(\mathbf{y}; \mathbf{S}\mathbf{x}, \sigma^2\mathbf{I}_L)$ and $p_X(\mathbf{x}) \sim \mathrm{Dir}(\boldsymbol{\alpha})$

But the posterior samples on $\mathbf{x}$ are

- Typically underexploited: No UQ diagnostics, just MMSE Estimators[5].
- Results in cropped posterior distributions on the simplex: e.g. Truncated Gaussian for $\boldsymbol{\alpha} = \mathbf{1}$ (Uniform prior)
- Dirichlet allows mass on the boundary and is tied to the Euclidean Geometry of the simplex.
- all variables are usually assumed i.i.d. spatially: not robust to noise.

[5] Eches, O., Dobigeon, N., Mailhes, C., & Tourneret, J. Y. (2010). Bayesian estimation of linear mixtures using the normal compositional model. Application to hyperspectral imagery. IEEE Transactions on Image Processing, 19(6), 1403-1413.

# Outline

# Compositional Data Analysis: Aitchison geometry

Well known in geostatistics[6] but not at all in the signal processing community.

Define the centered log-ratio transform $\mathrm{clr} : \mathrm{int}\ \Delta_{P-1} \to \mathbb{R}^{P-1}$ as

$$\mathbf{w} = \psi(\mathbf{x}) \triangleq \mathrm{clr}(\mathbf{x}) := \left[\log x_p - \frac{1}{P}\sum_{k=1}^{P} \log x_k\right]_{p=1,\ldots,P}$$

The image of the simplex is the hyperplane $\mathbf{1}^T\mathbf{u} = 0$.

Choosing an orthonormal basis $\mathbf{H} \in \mathbb{R}^{P\times P-1}$, we get a true bijection, called "isometric log ratio", ilr:

$$\mathbf{z} = \phi(\mathbf{x}) = \mathrm{ilr}(\mathbf{x}) \triangleq \mathbf{H}^T\mathbf{w}$$

And the inverse is given by

$$\mathrm{ilr}^{-1}(\mathbf{z}) = \mathrm{softmax}(\mathbf{Hz}) = \left[\frac{\exp w_p}{\sum_{k=1}^{P}\exp(w_k)}\right]_{p=1,\ldots,P}.$$

[6] Aitchison, J. (1982). The statistical analysis of compositional data. Journal of the Royal Statistical Society: Series B (Methodological), 44(2), 139-160.

# Pushforward Gaussian prior

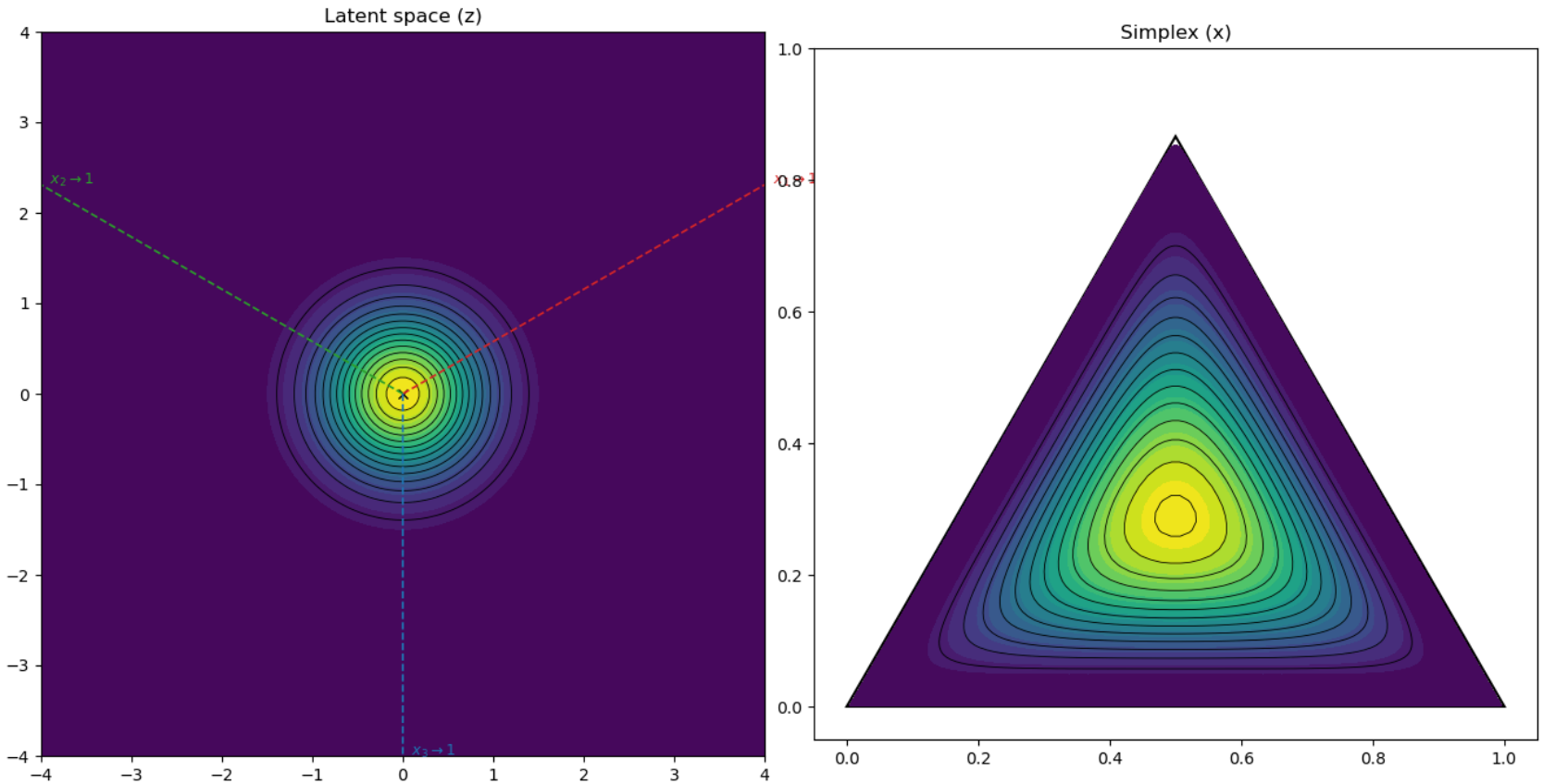
Now we have a diffeomorphism $\phi : \mathcal{C} \to \mathbb{R}^d$, we can put a Gaussian prior on $\mathbf{z}$ and push it to $\mathcal{C}$ using $\phi^{-1}$.

$$\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Simplex example: defines a family of distributions in $\Delta_{P-1}$ such that:

- Zero mass is put on the boundary $\to$ same for the posterior
- Can be made permutation invariant (uninformative) if $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{P-1}$
- May be multimodal depending on the latent prior.

# Examples of ilr-Gaussian distributions ($P = 3$)

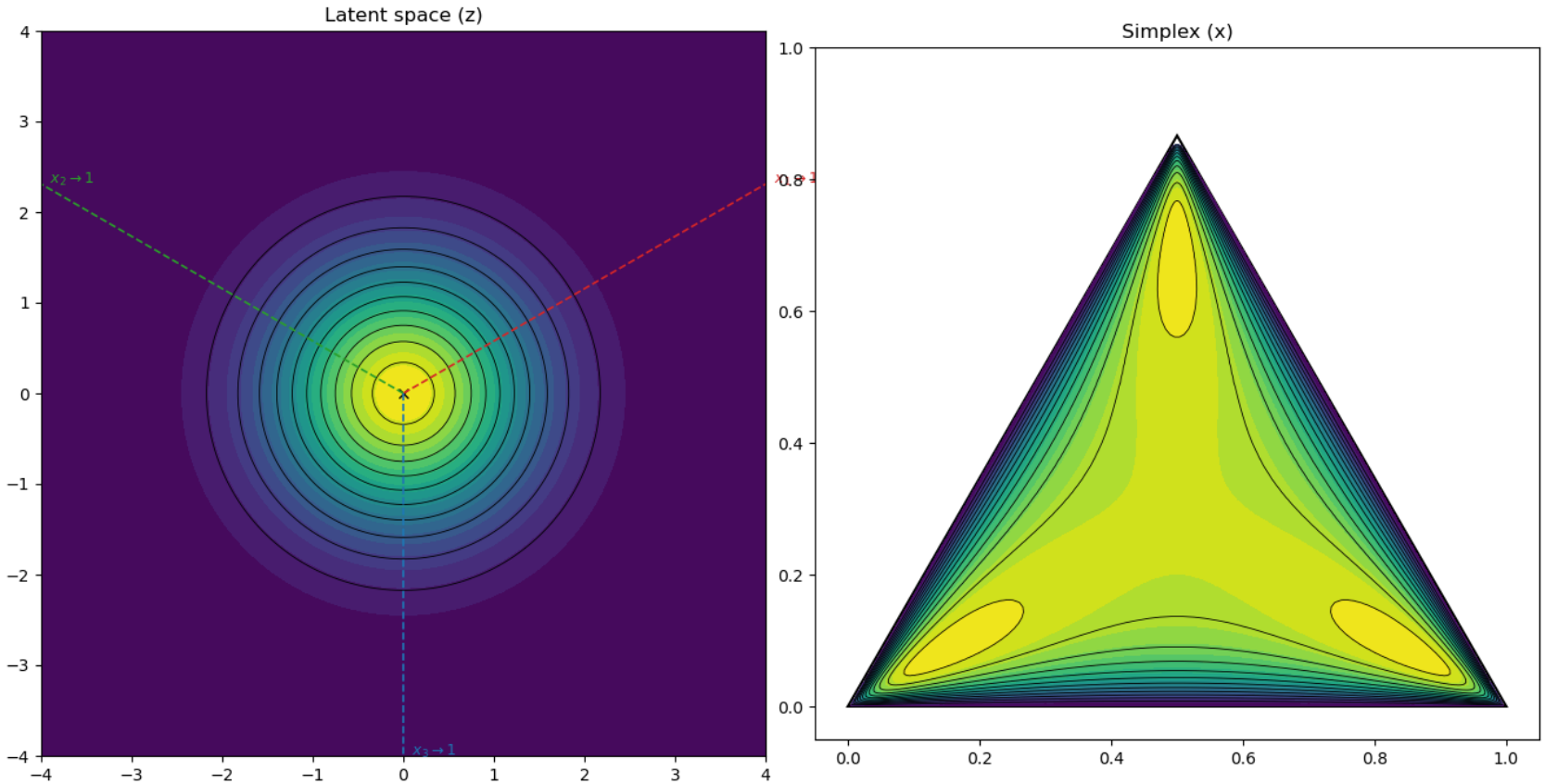



$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{P-1})$$

$$\sigma^2 = 0.6$$

Zero mean and isotropic latent $\rightarrow$ permutation invariant in the simplex

# Examples of ilr-Gaussian distributions ($P = 3$)

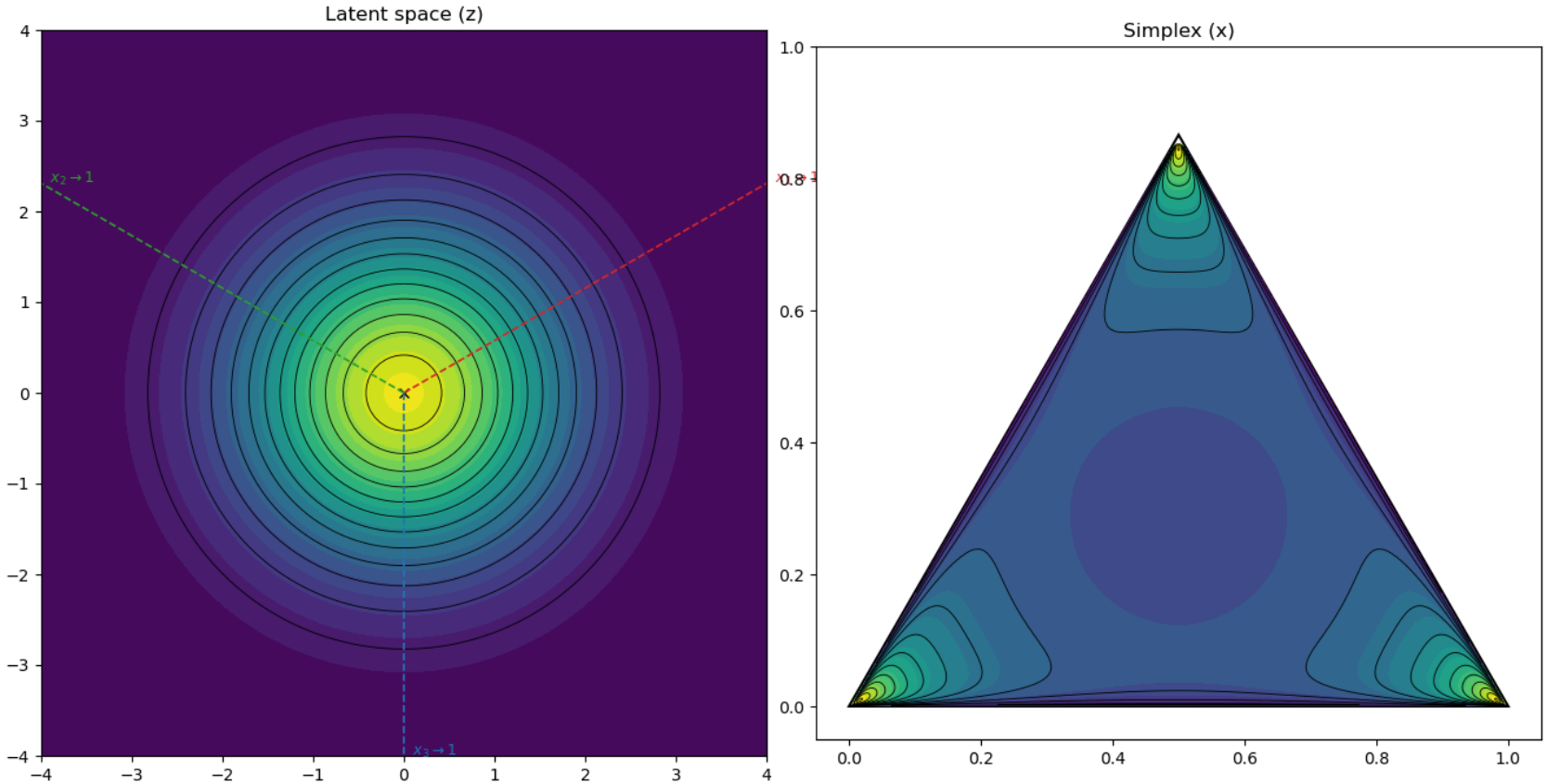



$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{P-1})$$

$$\sigma^2 = 1$$

Multimodal for large enough $\sigma$ (still isotropic)

# Examples of ilr-Gaussian distributions ($P = 3$)

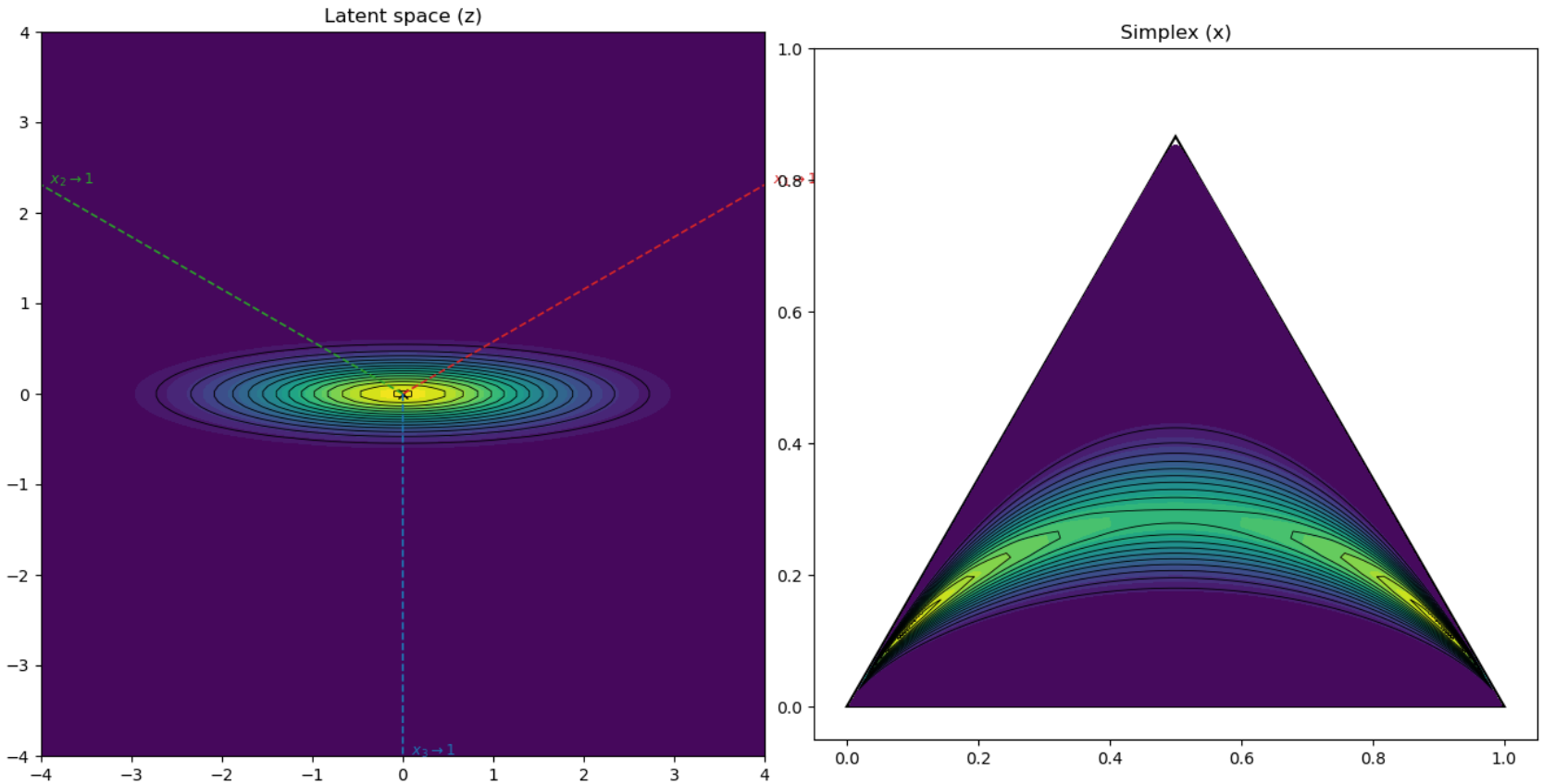



$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{P-1})$$

$$\sigma^2 = 1.25$$

For large values of $\sigma$, distributions favor edges with equal probability.

# Examples of ilr-Gaussian distributions ($P = 3$)

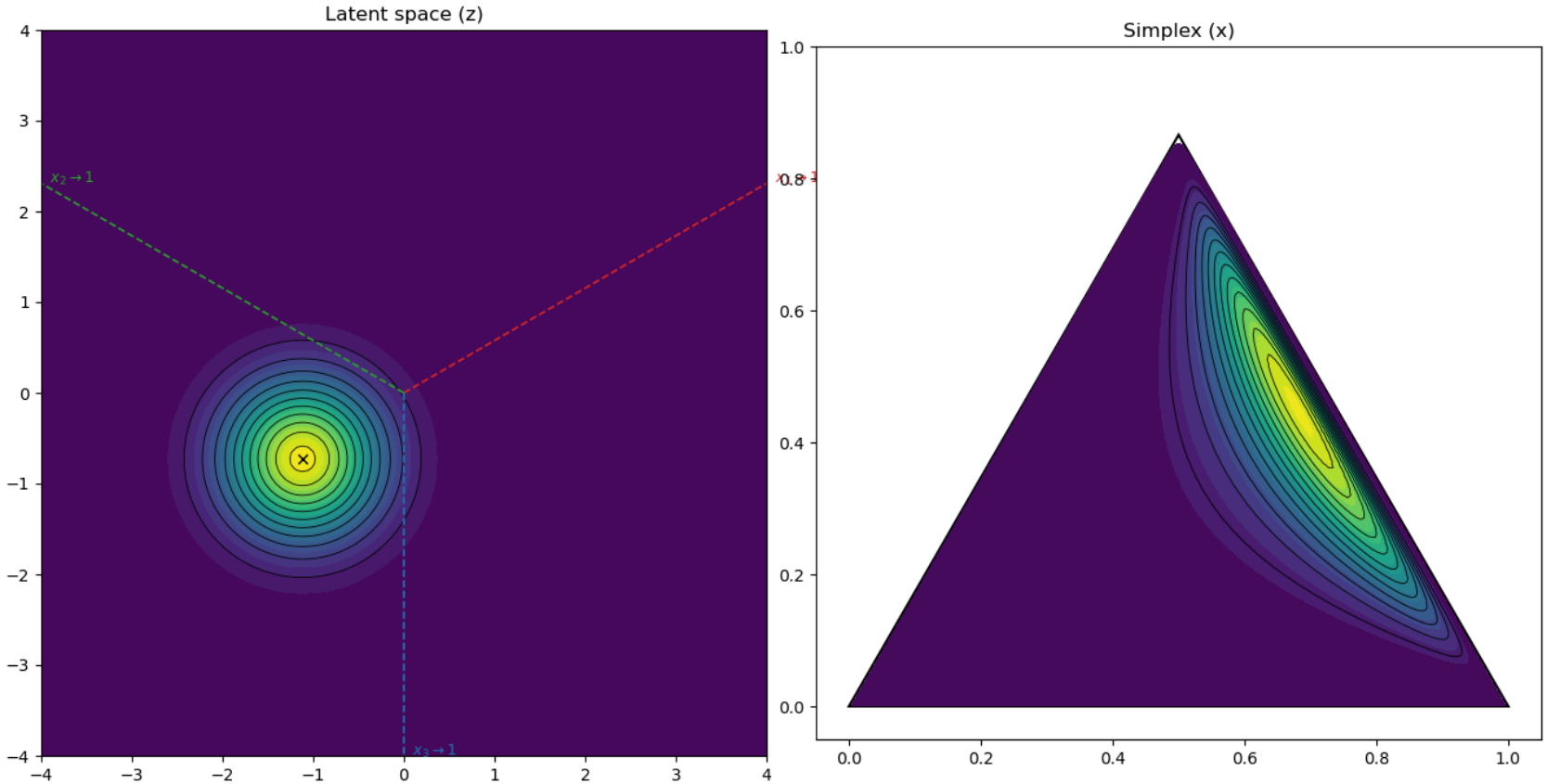



$$\mathbf{z} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}\sigma_1^2, 0\\ 0, \sigma_2^2\end{bmatrix} \mathbf{I}_{P-1}\right)$$

$$\sigma_1^2 = 1, \sigma_2^2 = 0.2$$

For non isotropic covariances, behavior is more complex.

# Examples of ilr-Gaussian distributions ($P = 3$)




$$\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_{P-1})$$

$$\sigma^2 = 0.6$$

The mean also affects how mass distributes in the simplex.

# Gaussian Process regression for constrained fields

Working in **z**-space allows Gaussian priors on matrix $\mathbf{Z} \in \mathbb{R}^{(P-1)\times N}$ (N latent values) at target locations $\mathbf{u}_n \in \mathbb{R}^q$, $n = 1, \ldots N$

Control spatial correlations via separable covariances in space and simplex dimensions:

$$\mathbf{X} = \psi(\mathbf{X}) \sim \mathcal{MN}(\mathbf{0}_{(P-1)\times N}, \sigma_a^2 \mathbf{I}_{P-1}, \mathbf{K_U}). \tag{1}$$

$\mathbf{K}_U$ encodes spatial correlations (i.e. a discretized Gaussian process): $\mathbf{K}_{U\,i,j} = k(\mathbf{u}_i, \mathbf{u}_j)$ for a given kernel function.

- Latent components are uncorrelated between one another:

$$\mathrm{Cov}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{0}_N \quad \text{if} \quad i \neq j$$

- For any given component, spatial correlation is governed by the kernel:

$$\mathrm{Cov}(\mathbf{x}_i, \mathbf{x}_i) = \sigma_a^2 \mathbf{K}_U$$

Equivalently

$$\mathbf{x} = \mathrm{vec}(\mathbf{X}) \sim (\mathbf{0}_{(P-1)N}, \sigma_a^2 \mathbf{I}_{(P-1)} \otimes \mathbf{K}_U)$$

# Latent GP prior for spatial interpolation on $\mathcal{C}$

Assuming observations $\mathbf{y} = \mathrm{vec}(\mathbf{Y}) \in \mathbb{R}^{PN}$ are also $\mathcal{C}$-valued, the likelihood can be written in latent space. If in addition we assume that the noise is additive and Gaussian there:

$$p_{\phi(Y)|Z}(\phi(\mathbf{y})|\mathbf{z}) = \mathcal{N}(\phi(\mathbf{y}); \boldsymbol{\mu}, \sigma^2 \mathbf{I}_{PN})$$

Then the prior is conjugate in $\mathbf{z}$-space and the posterior has a closed form in latent space:

$$p_{Z|\phi(Y)}(\mathbf{z}|\phi(\mathbf{y})) \propto p_{Y|Z}(\phi(\mathbf{y})|\mathbf{z}) p_Z(\mathbf{z}) \sim \mathcal{N}(\mathbf{z}|\boldsymbol{\mu}_{Z|Y}, \boldsymbol{\Sigma}_{Z|Y})$$

is Gaussian and the moments are obtained in closed-form (GP regression in latent space).

This covers denoising or interpolation/extrapolation of signals with values on $\mathcal{C}$, assuming additive noise in latent space.

# Latent GP prior for Inverse problems

If the likelihood is specified in **x**-space, the situation is more complex.

We can rewrite Bayes' theorem for the unconstrained variable **z**:

$$p_{Z|Y}(\mathbf{z}|\mathbf{y}) \propto p_{Y|Z}(\mathbf{y}|\phi^{-1}(\mathbf{z}))p_Z(\mathbf{z})$$

In our unmixing case on the simplex, for the whole matrix:

$$-\log p_{X|Y}(\mathbf{X}|\mathbf{Y}) = \frac{1}{2\sigma^2}||\mathbf{Y} - \mathbf{S}\mathbf{X}||_F^2 + \frac{1}{2\sigma_a^2}||\phi(\mathbf{X})\mathbf{K}_U^{-1/2}||_F^2 + \text{cst}$$

However, better to work in **Z** space, since **Z** is unconstrained:

$$-\log p_{Z|Y}(\mathbf{Z}|\mathbf{Y}) = \frac{1}{2\sigma^2}||\mathbf{Y} - \mathbf{S}\phi^{-1}(\mathbf{Z})||_F^2 + \frac{1}{2\sigma_a^2}||\mathbf{Z}\mathbf{K}_U^{-1/2}||_F^2 + \text{cst}$$

Here the likelihood and the prior live in different spaces: no closed form inference. Sampling is needed.

# Inference

We need constrained sampling algorithms to obtain samples of $\mathbf{x}$ and compute the posterior mean (MMSE estimator):

$$\mathbb{E}[X|Y=y] = \int_{\mathbb{R}^n} \mathbf{x} p_{X|Y}(\mathbf{x}|\mathbf{y}) d\mathbf{x} \approx \frac{1}{M} \sum_{i=1}^{M} \mathbf{x}_i$$

Inference can also been done in $\mathbf{z}$-space, using the same samples:

$$\mathbb{E}[Z|Y=y] = \int_{\mathbb{R}^d} \mathbf{z} p_{Z|Y}(\mathbf{z}|\mathbf{y}) d\mathbf{z} \approx \frac{1}{M} \sum_{i=1}^{M} \mathbf{z}_i$$

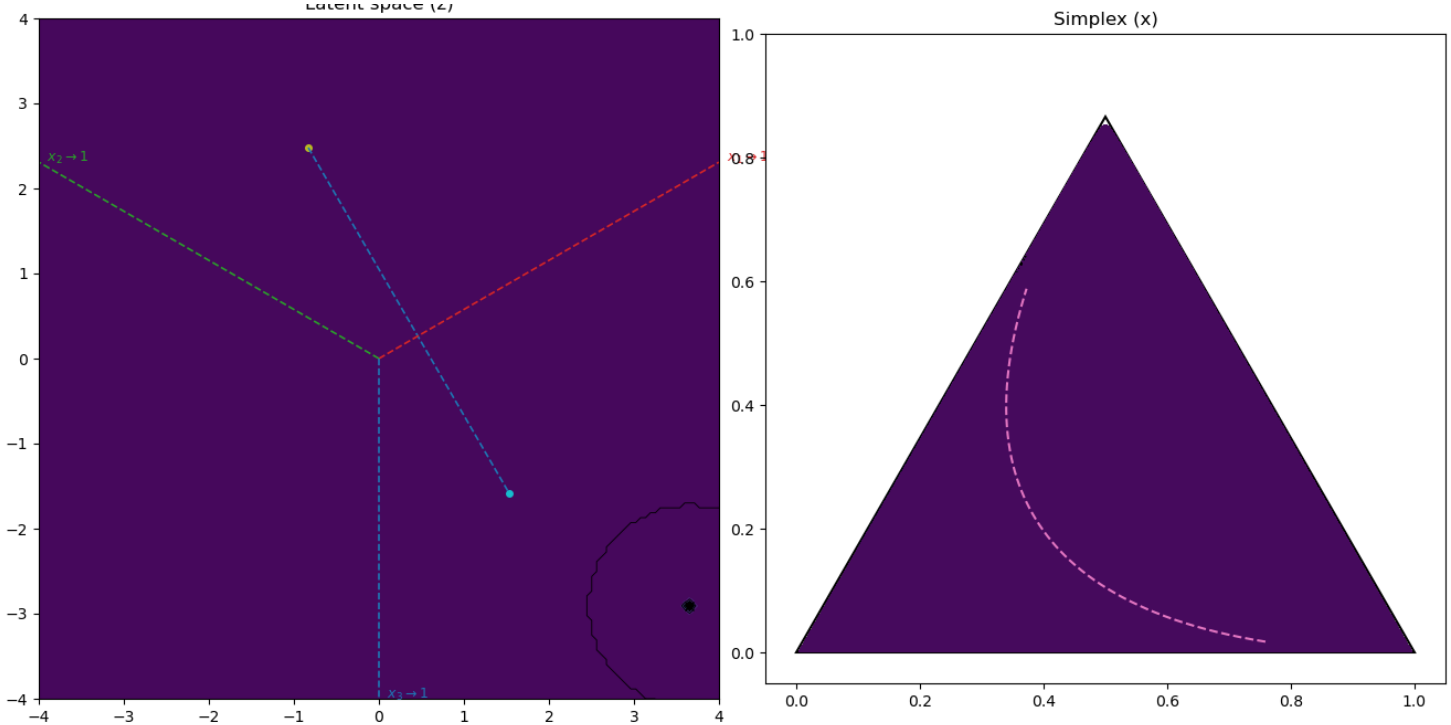
$\phi$ makes $\mathcal{C}$ a (globally flat) Riemannian manifold structure with a pullback metric $g = \phi^* \delta$, where $\delta$ is the Euclidean metric on $\mathbb{R}^d$. In particular this defines a geodesic distance:

$$d_\phi(\mathbf{x}_1, \mathbf{x}_2) = ||\phi(\mathbf{x}_1) - \phi(\mathbf{x}_2)||_2^2 = ||\mathbf{z}_1 - \mathbf{z}_2||_2^2$$

for which $\phi^{-1}(\mathbb{E}[Z|Y=y])$ is the geodesic MMSE estimator of $\mathbf{x}$.

Not clear which is better (if $\mathcal{C}$ is convex, at least both remain in $\mathcal{C}$).

# In the simplex: Aitchison distance vs Euclidean distance




Aitchison geodesic in the simplex. The geodesic is a straight line in latent space.

- Geodesics are straight lines in latent space
- The ilr-Gaussian distribution has Gaussian Riemannian density $d\nu_{\mathcal{C}}(\mathbf{x}) = \mathcal{N}(\phi(\mathbf{x}); \boldsymbol{\mu}, \boldsymbol{\Sigma}) d\mathrm{vol}_g(\mathbf{x})$, but the corresponding density in the Euclidean geometry is corrected by the Jacobian of $\phi$:

$$d\nu_{\mathcal{C}}(\mathbf{x}) = \mathcal{N}(\phi(\mathbf{x}); \boldsymbol{\mu}, \boldsymbol{\Sigma}) |\det(D\phi(\mathbf{x}))| d\mathbf{x}$$

# Constrained sampling

Also related to geometry.
Unconstrained $\mathbf{x}$: One can use the Unadjusted Langevin Algorithm:

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \beta \nabla U(\mathbf{x}_t) + \sqrt{2\beta}\boldsymbol{\xi}_t$$

where $\boldsymbol{\xi}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and $U(\mathbf{x}) = -\log(p_{X|Y}(\mathbf{x}|\mathbf{y}))$

For $\mathbf{x}$ in $\mathcal{C}$, the posterior lives inside $\Delta_{P-1}$. Several options for sampling:

- Mirror Langevin[7] (simplex native): simple exponential updates on $U(\mathbf{x})$
- Riemannian Langevin (directly uses the pullback geometry[8]):

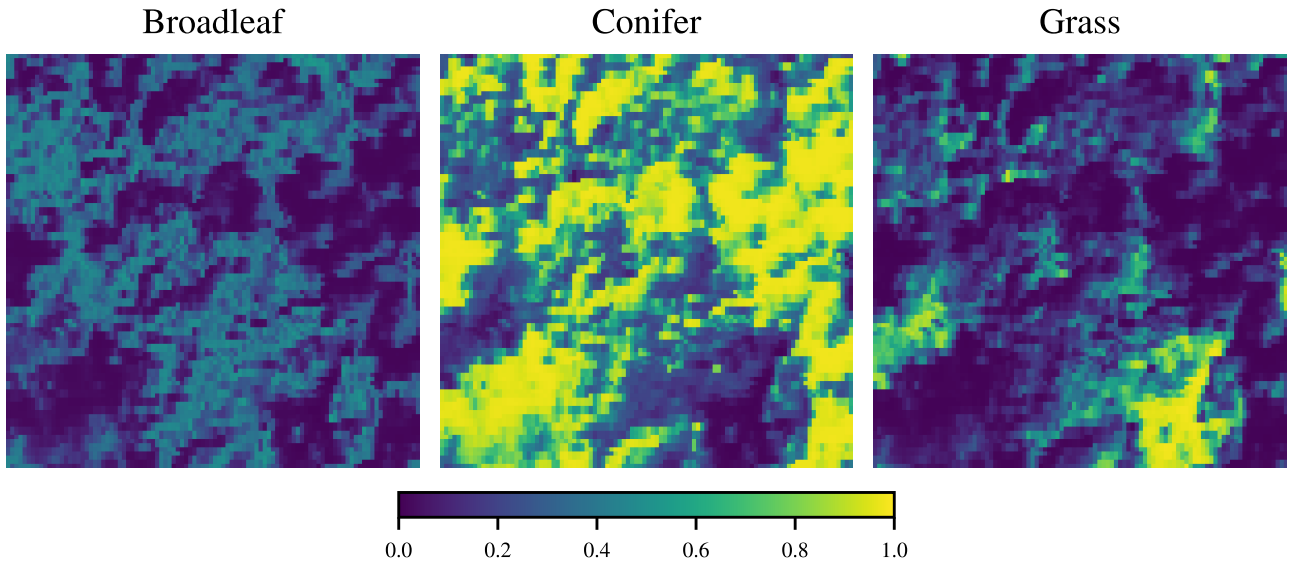
$$\mathbf{z}_{t+1} = \mathbf{z}_t - \beta \nabla V(\mathbf{z}_t) + \sqrt{2\beta}\boldsymbol{\xi}_t$$

with $V(\mathbf{z}) = -\log(p_{Z|Y}(\mathbf{z}|\mathbf{y}))$

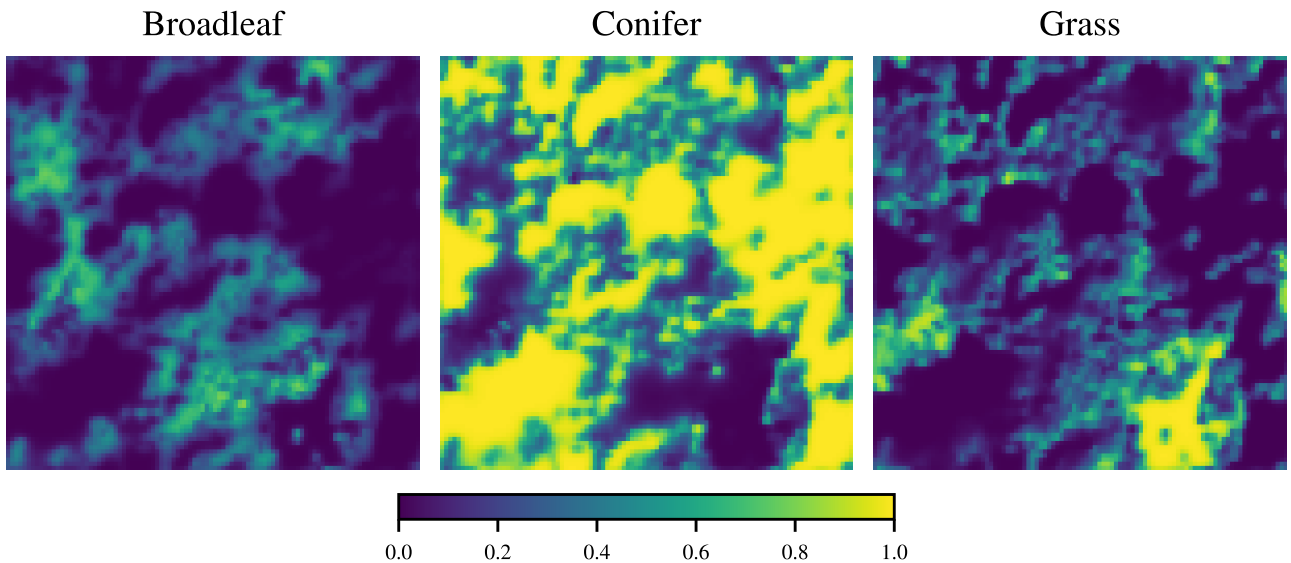
[7] Hsieh, Y. P., Kavis, A., Rolland, P., & Cevher, V. (2018). Mirrored langevin dynamics. Advances in Neural Information Processing Systems, 31.

[8] Wang, X., Lei, Q., & Panageas, I. (2020). Fast convergence of langevin dynamics on manifold: Geodesics meet log-sobolev. Advances in Neural Information Processing Systems, 33, 18894-18904.

# First order information




Geodesic mean for each component, dirac kernel




Geodesic mean for each component, with an exponential kernel

# Second order information: constrained UQ

UQ in constrained domains: covariance of posterior samples (for 1 pixel first)?

Same thing: covariances make sense both in **x** and **z** space:

$$\boldsymbol{\Sigma}_X = \mathbb{E}[(X - \boldsymbol{\mu}_X)(X - \boldsymbol{\mu}_X)^T | Y], \quad \boldsymbol{\mu}_X = \mathbb{E}[X|Y]$$

But:

- If $\mathcal{C}$ is not fully dimensional in $\mathbb{R}^d$, $\boldsymbol{\Sigma}_X$ will be singular.
- Gaussian approximations put mass outside $\mathcal{C}$ and credible regions can be meaningless.

Defining covariance in **Z**-space fixes both issues:

$$\boldsymbol{\Sigma}_Z = \mathbb{E}[(Z - \boldsymbol{\mu}_Z)(Z - \boldsymbol{\mu}_Z)^T | Y], \quad \boldsymbol{\mu}_Z = \mathbb{E}[Z|Y]$$

so that Gaussian approximations in **z**-space make perfect sense.

# Credible regions in the simplex

For (unconstrained) Gaussian distributions in $\mathbb{R}^d$ $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, credible ellipsoid $E_\alpha$ containing a proportion $\alpha \in [0, 1]$ of the total mass:

$$E_\alpha = \{\mathbf{x} \in \mathbb{R}^n, (\mathbf{x} - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) < F^{-1}_{\chi^2_d}(\alpha)\}$$

- Gaussian approximation of $p_{X|Y}(\mathbf{x}|\mathbf{y})$ in $\mathbf{x}$-space as a measure of uncertainty in the simplex can be misleading.
- Better strategy: Gaussian approximation of $p_{Z|Y}(\mathbf{z}|\mathbf{y})$ in $\mathbf{z}$-space, and pushforward $\alpha$-credible ellispoids to $\mathcal{C}$ using $\phi^{-1}$: defines credible regions in $\mathcal{C}$ with same total mass $\alpha$
- For multimodal densities in $\mathcal{C}$, better to consider the Highest Density Region (HDR): the region with given proportion $\alpha$ of the total mass and with minimum (Euclidean) volume; equivalently the set:

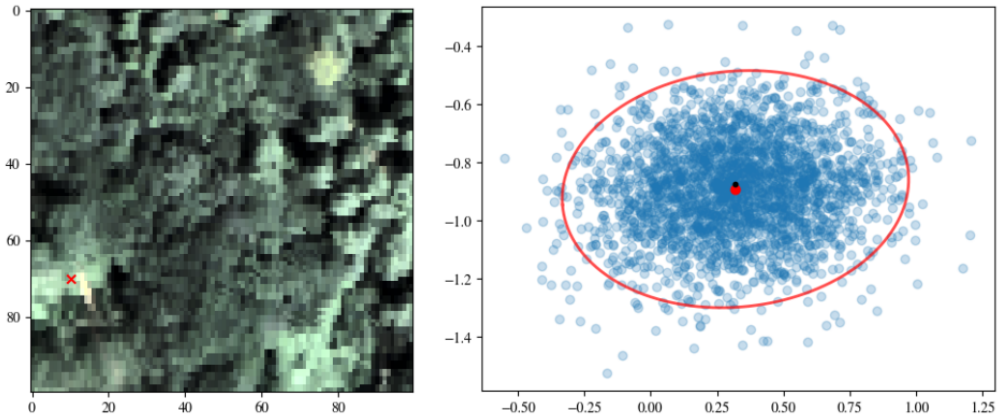
$$H_t = \{\mathbf{x} \in \mathcal{C}, p_{X|Y}(\mathbf{x}|\mathbf{y}) \geq t)\}$$

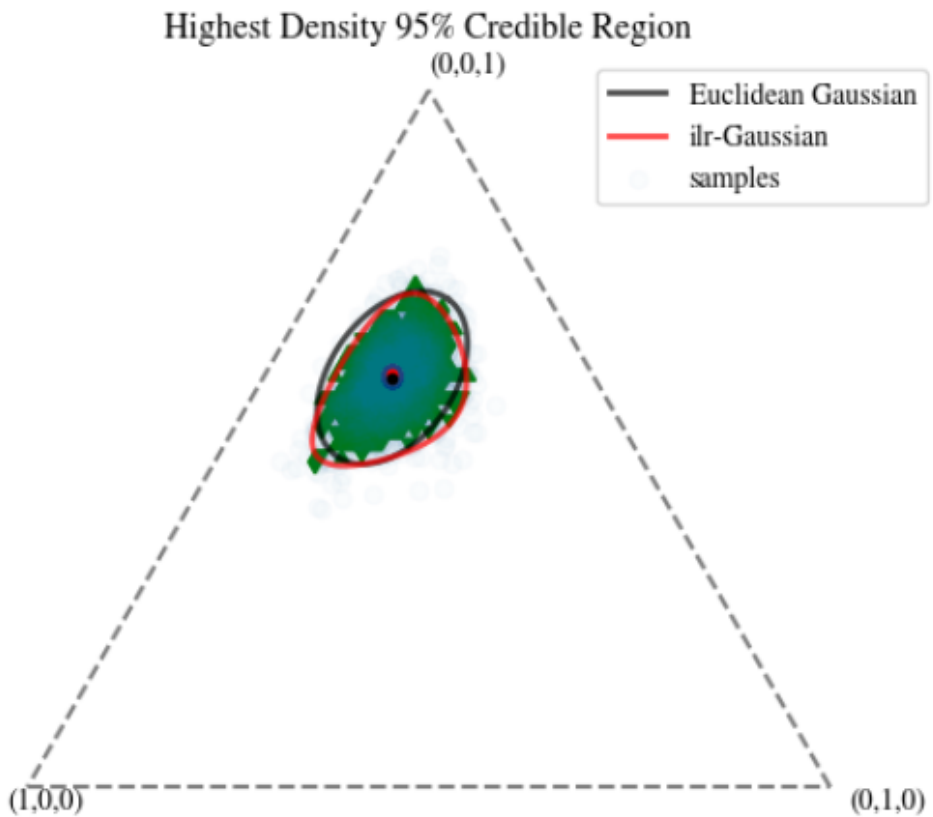
with $t$ chosen such that

$$\int_{H_t} p_{X|Y}(\mathbf{x}|\mathbf{y})d\mathbf{x} = \alpha$$

# Pixel-wise Uncertainty

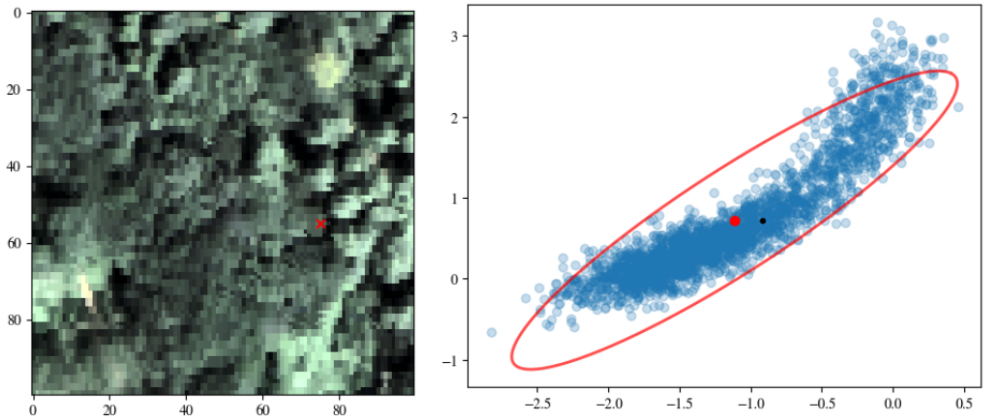
$(1, 0, 0)$: Broadleaf, $(0, 1, 0)$: Conifer, $(0, 0, 1)$: Grass

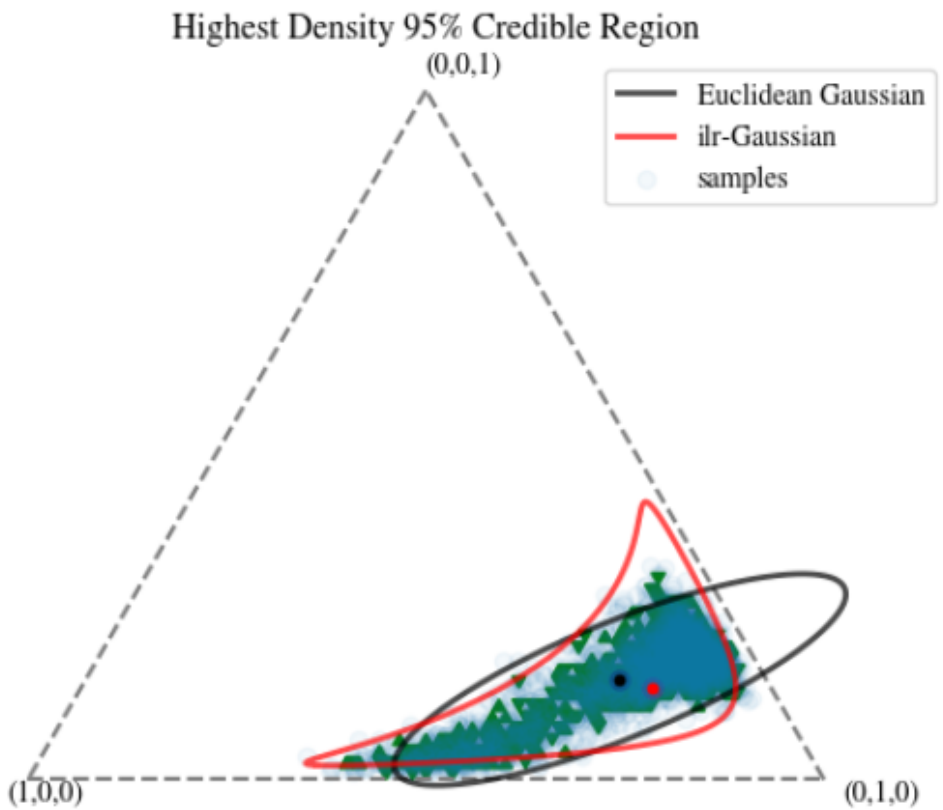


# Pixel-wise Uncertainty

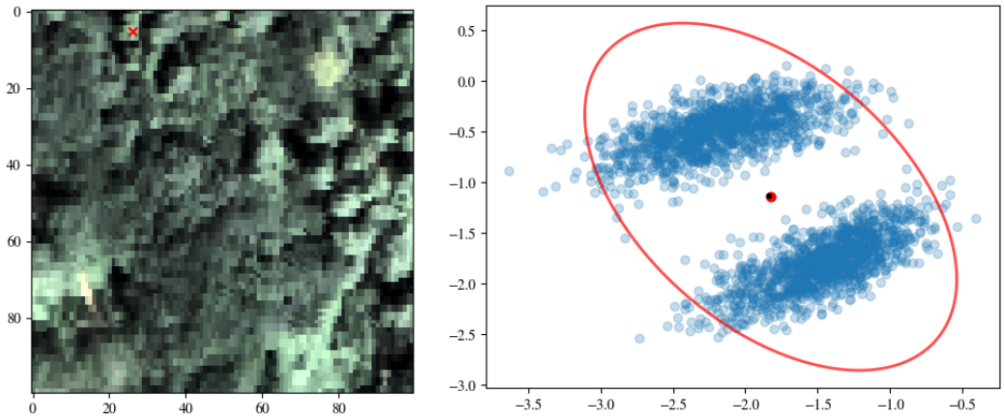
$(1, 0, 0)$: Broadleaf, $(0, 1, 0)$: Conifer, $(0, 0, 1)$: Grass

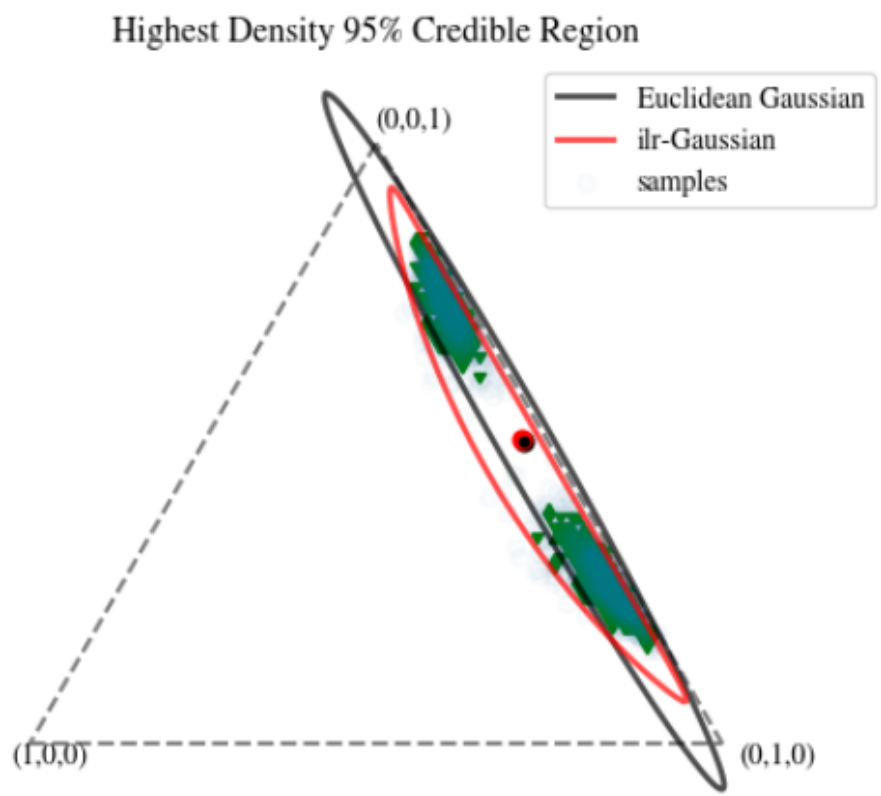




# Pixel-wise Uncertainty

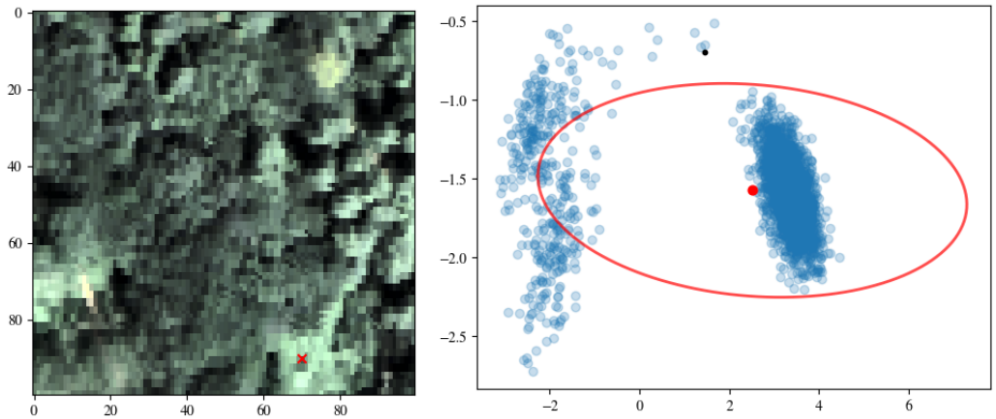
$(1, 0, 0)$: Broadleaf, $(0, 1, 0)$: Conifer, $(0, 0, 1)$: Grass

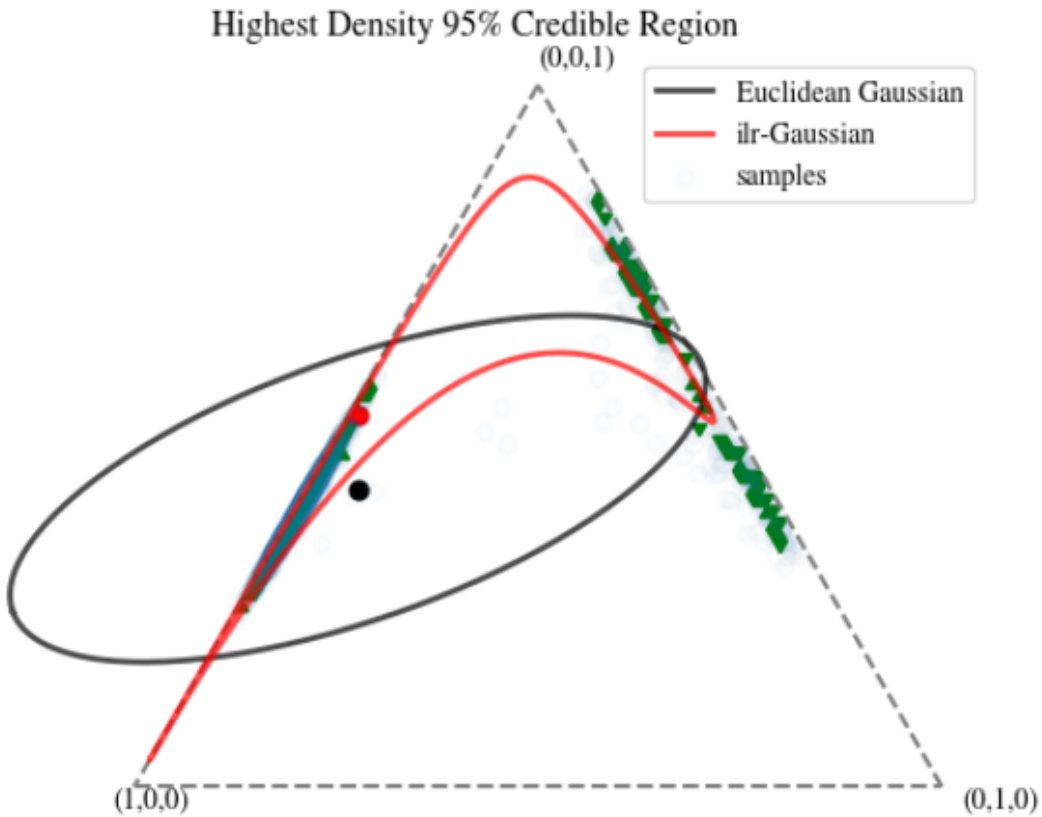




# Pixel-wise Uncertainty

$(1, 0, 0)$: Broadleaf, $(0, 1, 0)$: Conifer, $(0, 0, 1)$: Grass



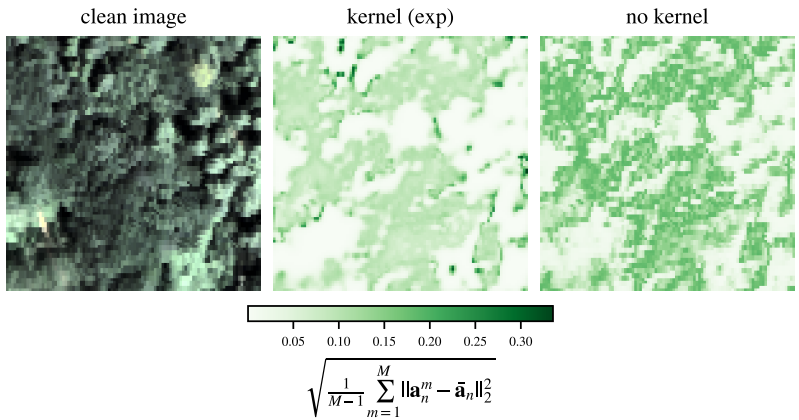


# Spatializing Uncertainty in C

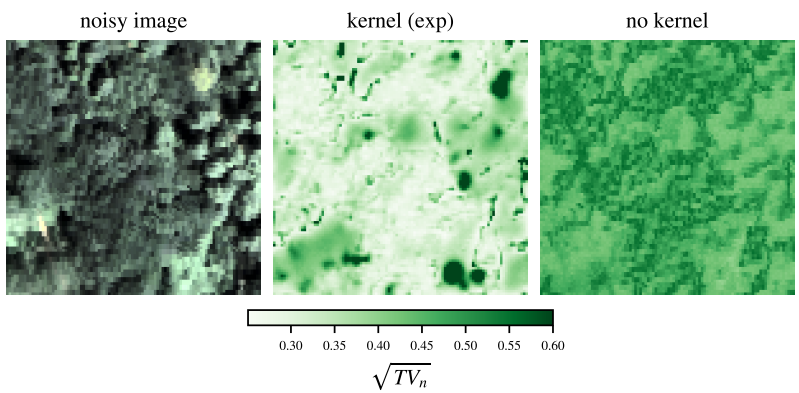
So far we can only do UQ at the pixel level. We need a scalar in each pixel to be able to spatialize the uncertainty.
Total variance is one candidate, either in Euclidean or ilr space:

$$\mathrm{Var}_X = \mathrm{tr}(\mathbf{\Sigma}_X), \quad \mathrm{Var}_Z = \mathrm{tr}(\mathbf{\Sigma}_Z)$$

$\mathrm{Var}_X$ is sensitive to large coordinate-wise changes:




$\mathrm{Var}_Z$ is sensitive to large fold changes (in the log ratios between components):



# Nonlinearity/Multimodality diagnostics

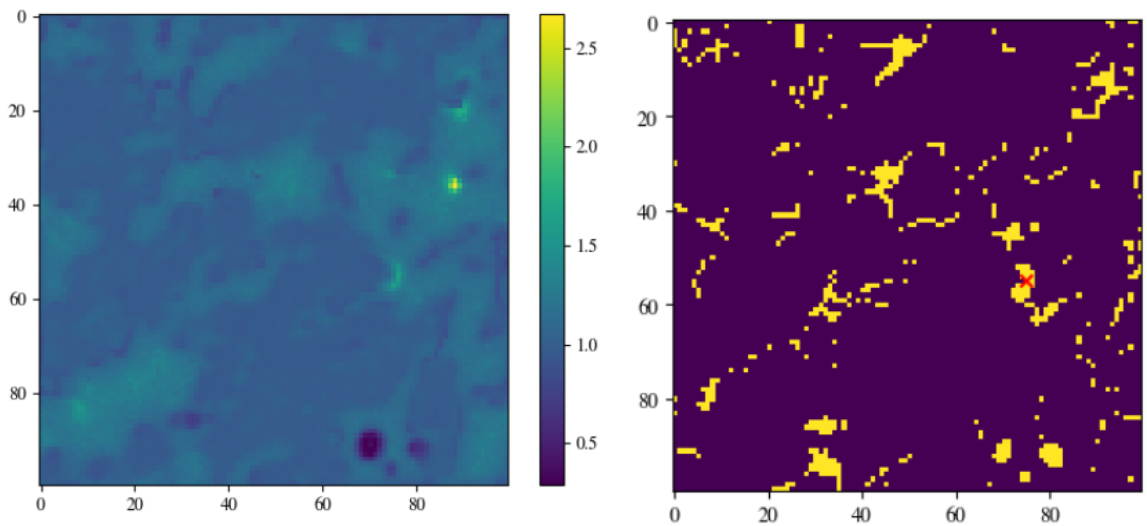
Most pixels have unimodal, similar looking densities in both spaces. This happens when softmax behaves almost linearly (e.g. for concentrated densities not too far from the center)

$\rightarrow$ Detect nonlinearity/multimodality by comparing Euclidean total variance and its "prediction" by propagating the ilr-covariance by a linearized $\mathrm{ilr}^{-1}$ around the mean value:

$$\mathbf{\Sigma}_{X_{lin}} = \mathbf{J}_{\boldsymbol{\mu}_Z} \mathbf{\Sigma}_Z \mathbf{J}_{\boldsymbol{\mu}_Z}^T$$

where $\mathbf{J}_{\boldsymbol{\mu}_Z}$ is the Jacobian of $\mathrm{ilr}^{-1}$ calculated around the posterior ilr-mean.

$\rightarrow$ A ratio $\frac{\mathrm{tr}(\mathbf{\Sigma}_X)}{\mathrm{tr}(\mathbf{\Sigma}_{X_{lin}})}$ far away from 1 means nonlinearity, possibly multimodality.




Trace ratio (left). Binarized version (large deviations from 1 only), removing pixels too close to the boundary

# Conclusion

- Natural class of prior distributions on constrained domains diffeomorphic to Euclidean spaces
- Allows to define constrained Gaussian processes for spatialized priors
- Also allows for pixel and image level UQ metrics
- Application to spectral unmixing

# Perspectives

- Hypothesis testing on variance ratio
- Define better UQ diagnostics for multimodality, confidence, model misspecification...
- Construction allows to do more than just GP on $\mathcal{C}$: transport the Euclidean vector space structure
- Integrate into more realistic unmixing pipelines and analyze results with practitioners
- Same framework also applies to separate epistemic/aleatoric uncertainty in Bayesian Classification (e.g. Bayesian) logistic regression.
- Extend to other spaces and exploit applications