Extrema of Gaussian processes

Global maximum of a Gaussian process

Let us consider a stationary Gaussian process $(Z(\boldsymbol{x}))_{\boldsymbol{x}\in\mathbb{R}^d}$ with mean zero and covariance function $c(\boldsymbol{x})$, such that $c(\mathbf{0}) = 1$.

When $|\Omega| \gg l_c^d$, the number N_I^Ω of local maxima of $Z(\boldsymbol{x})^2$ in a domain Ω that have peak values larger than I is

$$N_I^\Omega \simeq \frac{2|\Omega|}{V_c} I^{(d-1)/2} e^{-I/2}$$

Intuitively, the global maximum $I_\Omega = \max_{\boldsymbol{x}\in\Omega} Z(\boldsymbol{x})^2$ should be such that $N_{I_\Omega}^\Omega \sim 1$.

Cf. R. Adler's book:

$$I_\Omega = 2\ln\Big(\frac{2|\Omega|}{V_c}\Big) + (d-1)\ln\Big[2\ln\Big(\frac{2|\Omega|}{V_c}\Big)\Big] + 2Z_g$$

where Z_g follows a Gumbel distribution $\mathbb{P}(Z_g \leq z) = \exp(-e^{-z})$.

Array imaging with additive noise

Source imaging

- In the presence of a point source at $\boldsymbol{z}_{\rm so}$ the response vector is

$$\big(\hat{u}_0(\omega)\big)_r = \hat{G}(\omega, \boldsymbol{x}_r, \boldsymbol{z}_{\rm so})\sigma_{\rm so}(\omega), \qquad r = 1, \dots, N$$

or equivalently:

$$\hat{\boldsymbol{u}}_0(\omega) = \tau_{\rm so}(\omega)\boldsymbol{g}(\omega, \boldsymbol{z}_{\rm so}), \text{ with } \quad \tau_{\rm so}(\omega) = \sigma_{\rm so}(\omega)\Big(\sum_{l=1}^N |\hat{G}(\omega, \boldsymbol{z}_{\rm so}, \boldsymbol{x}_l)|^2\Big)^{1/2}$$

- $\tau_{\rm so}(\omega)$ is proportional to the emission power,
- $\boldsymbol{g}(\omega, \boldsymbol{x})$ is the normalized vector of Green's functions from the array to the point $\boldsymbol{x}$:

$$\boldsymbol{g}(\omega, \boldsymbol{x}) = \frac{1}{\big(\sum_{l=1}^N |\hat{G}(\omega, \boldsymbol{x}, \boldsymbol{x}_l)|^2\big)^{1/2}}\big(\hat{G}(\omega, \boldsymbol{x}, \boldsymbol{x}_j)\big)_{j=1,\dots,N}$$

- With an additive Gaussian white noise (with variance δ^2), the data set is

$$\boldsymbol{u}(\omega) = \boldsymbol{u}_0(\omega) + \boldsymbol{W}(\omega)$$

where $(W_r(\omega))_{r=1,\dots,N}$ are complex-valued, independent and identically distributed Gaussian $\mathcal{N}(0, \delta^2)$ ($W_r = W_{r1} + iW_{r2}$ with W_{r1} and W_{r2} real-valued, independent and identically distributed Gaussian $\mathcal{N}(0, \delta^2/2)$).

- We want to estimate $\boldsymbol{z}_{\mathrm{so}}$ from the measured vector $\boldsymbol{u}(\omega)$:

$$\boldsymbol{u}(\omega) = \boldsymbol{u}_0(\omega) + \boldsymbol{W}(\omega)$$

- RT imaging functional:

$$\mathcal{I}_{\mathrm{RT}}(\boldsymbol{z}^S) = \overline{\boldsymbol{g}(\omega, \boldsymbol{z}^S)}^T \boldsymbol{u}(\omega)$$

where $\boldsymbol{g}$ is the normalized vector of Green's functions:

$$\boldsymbol{g}(\omega, \boldsymbol{x}) = \frac{1}{\left(\sum_{l=1}^N |\hat{G}(\omega, \boldsymbol{x}, \boldsymbol{x}_l)|^2\right)^{1/2}} \left(\hat{G}(\omega, \boldsymbol{x}, \boldsymbol{x}_j)\right)_{j=1,\ldots,N}$$

- KM imaging functional:

$$\mathcal{I}_{\mathrm{KM}}(\boldsymbol{z}^S) = \overline{\boldsymbol{d}(\omega, \boldsymbol{z}^S)}^T \boldsymbol{u}(\omega)$$

where

$$\boldsymbol{d}(\omega, \boldsymbol{z}) = \frac{1}{\sqrt{N}} \Big(\exp\left(i\omega \mathcal{T}(\boldsymbol{x}_j, \boldsymbol{z})\right) \Big)_{j=1,\ldots,N}$$

Complement: Bayes' theorem

Bayes' theorem shows how to determine inverse probabilities: knowing the conditional probability of B given A, what is the conditional probability of A given B ?

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

For probability densities: Assume $(\boldsymbol{X}, \boldsymbol{Y})$ is a random vector with pdf $p_{\boldsymbol{X},\boldsymbol{Y}}(x, y)$. Then the distribution of $\boldsymbol{X}$ knowing $\boldsymbol{Y} = \boldsymbol{y}$ is:

$$p_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{Y} = \boldsymbol{y}) = \frac{p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y})}{p_{\boldsymbol{Y}}(\boldsymbol{y})} = \frac{p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{X} = \boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x})}{p_{\boldsymbol{Y}}(\boldsymbol{y})}$$

Here:
- $p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y})$ is the joint pdf (probability density function) of $(\boldsymbol{X}, \boldsymbol{Y})$.
- $p_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{Y} = \boldsymbol{y})$ is the posterior pdf of $\boldsymbol{X}$ given $\boldsymbol{Y} = \boldsymbol{y}$.
- $p_{\boldsymbol{X}}(\boldsymbol{x})$ and $p_{\boldsymbol{Y}}(\boldsymbol{y})$ are the marginal pdfs of $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively ($p_{\boldsymbol{X}}(\boldsymbol{x})$ is the prior pdf of $\boldsymbol{X}$).

- Bayesian analysis.

Given $\boldsymbol{z}$, τ, and δ, the data $\boldsymbol{u}$ has the probability density function

$$p\left(\boldsymbol{u} \mid \boldsymbol{z}, \tau, \delta\right) = \frac{1}{(\pi\delta^2)^N} \exp\left(-\frac{\|\boldsymbol{u} - \tau \boldsymbol{g}(\boldsymbol{z})\|^2}{\delta^2}\right)$$

(with respect to the Lebesgue measure over the space of complex vectors).

Maximum likelihood estimator: We look for $\hat{\boldsymbol{z}}$ that maximizes the pdf $\boldsymbol{z} \to p(\boldsymbol{z}|\boldsymbol{u})$ given the data $\boldsymbol{u}$.

Using Bayes theorem with the Jeffreys prior for the parameters δ, τ (a non-informative prior distribution proportional to δ^{-1}) and the uniform distribution in the search domain for $\boldsymbol{z}$, we find that, given the observations $\boldsymbol{u}$, the likelihood function of the parameters $\boldsymbol{z}$, τ, and δ is proportional to

$$l_0\left(\boldsymbol{z}, \tau, \delta \mid \boldsymbol{u}\right) = \frac{1}{\delta^{2N+1}} \exp\left(-\frac{\|\boldsymbol{u} - \tau \boldsymbol{g}(\boldsymbol{z})\|^2}{\delta^2}\right)$$

Complement: Jeffreys prior

In Bayesian analysis, the Jeffreys prior $\pi(\boldsymbol{\theta})$ is a non-informative prior distribution on parameter space Θ that is proportional to :

$$\pi(\boldsymbol{\theta}) \propto \sqrt{\det \mathbf{I}(\boldsymbol{\theta})}$$

where $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher information matrix:

$$I_{jl}(\boldsymbol{\theta}) = \int_{\Theta} \partial_{\theta_j} \left[\ln p(\boldsymbol{\theta}) \right] \partial_{\theta_j} \left[\ln p(\boldsymbol{\theta}) \right] p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

It is invariant under reparameterization of the parameter vector $\boldsymbol{\theta}$.

Examples:
the Jeffreys prior for the mean of a Gaussian variable is uniform over the entire real line.
the Jeffreys prior for the variance δ^2 of a Gaussian variable is $1/\delta$.

The maximum likelihood estimate of $\boldsymbol{z}$ and the nuisance parameters δ and τ are found by maximizing the likelihood function with respect to these:

$$\left(\hat{\boldsymbol{z}}, \hat{\tau}, \hat{\delta}\right) = \underset{\boldsymbol{z},\tau,\delta}{\operatorname{argmax}}\, l_0\left(\boldsymbol{z}, \tau, \delta \mid \boldsymbol{u}\right)$$

We first eliminate δ by requiring

$$\frac{\partial l_0\left(\boldsymbol{z}, \tau, \delta \mid \boldsymbol{u}\right)}{\partial \delta} = 0$$

which gives

$$\hat{\delta} = \frac{\|\boldsymbol{u} - \tau \boldsymbol{g}(\boldsymbol{z})\|}{\sqrt{2N+1}},$$

and then the likelihood ratio is proportional to

$$l_0\left(\boldsymbol{z}, \tau, \hat{\delta} \mid \boldsymbol{u}\right) \sim \|\boldsymbol{u} - \tau \boldsymbol{g}(\boldsymbol{z})\|^{-N-1/2}$$

Since $\|\boldsymbol{g}(\boldsymbol{z})\| = 1$, we have

$$\hat{\tau} = \underset{\tau}{\operatorname{argmin}} \|\boldsymbol{u} - \tau \boldsymbol{g}(\boldsymbol{z})\|^2 = \langle \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{u} \rangle$$

where $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \overline{\boldsymbol{a}}^T \boldsymbol{b}$.

Therefore the estimate $\hat{\boldsymbol{z}}$ derives from maximizing the function

$$\hat{\boldsymbol{z}} = \underset{\boldsymbol{z}}{\operatorname{argmin}} \left\| \boldsymbol{u} - \langle \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{u} \rangle \boldsymbol{g}(\boldsymbol{z}) \right\|^2$$

We find

$$\begin{aligned} \left\| \boldsymbol{u} - \langle \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{u} \rangle \boldsymbol{g}(\boldsymbol{z}) \right\|^2 &= \|\boldsymbol{u}\|^2 - \left| \langle \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{u} \rangle \right|^2 \\ &= \|\boldsymbol{u}\|^2 - |\mathcal{I}_{\mathrm{RT}}(\boldsymbol{z})|^2 \end{aligned}$$

The maximum likelihood estimation of the source location is

$$\hat{\boldsymbol{z}} = \underset{\boldsymbol{z}}{\operatorname{argmax}} \left| \mathcal{I}_{\mathrm{RT}}(\boldsymbol{z}) \right|$$

Active array imaging

• In the presence of a point reflector at $\boldsymbol{z}_{\rm ref}$ and in the Born approximation, the response matrix is

$$\big(u_0(\omega)\big)_{rs} = \hat{G}(\omega, \boldsymbol{x}_r, \boldsymbol{z}_{\rm ref}) \frac{\omega^2}{c_0^2} \sigma_{\rm ref} l_{\rm ref}^3 \hat{G}(\omega, \boldsymbol{z}_{\rm ref}, \boldsymbol{x}_s), \qquad r, s = 1, \ldots, N$$

or equivalently:

$$\mathbf{u}_0(\omega) = \tau_{\rm ref} \boldsymbol{g}(\omega, \boldsymbol{z}_{\rm ref}) \boldsymbol{g}(\omega, \boldsymbol{z}_{\rm ref})^T, \ \text{ with } \quad \tau_{\rm ref} = \frac{\omega^2}{c_0^2} \sigma_{\rm ref} l_{\rm ref}^3 \Big(\sum_{l=1}^N |\hat{G}(\omega, \boldsymbol{z}_{\rm ref}, \boldsymbol{x}_l)|^2 \Big)$$

- $\tau_{\rm ref}$ is the scattering amplitude of the reflector,
- $l_{\rm ref}^3$ is the volume of the reflector,
- $\boldsymbol{g}(\omega, \boldsymbol{x})$ is the normalized vector of Green's functions from the array to the point $\boldsymbol{x}$:

$$\boldsymbol{g}(\omega, \boldsymbol{x}) = \frac{1}{\big(\sum_{l=1}^N |\hat{G}(\omega, \boldsymbol{x}, \boldsymbol{x}_l)|^2\big)^{1/2}} \big(\hat{G}(\omega, \boldsymbol{x}, \boldsymbol{x}_j)\big)_{j=1,\ldots,N}$$

The matrix $\mathbf{u}_0$ is symmetric and it has rank 1.

• With an additive Gaussian white noise (with variance δ^2), the data set is

$$\mathbf{u}(\omega) = \mathbf{u}_0(\omega) + \mathbf{W}(\omega)$$

- We want to estimate $\boldsymbol{z}_{\rm ref}$ from the symmetrized matrix $\mathbf{u}^s(\omega) = \frac{1}{2}\big(\mathbf{u}(\omega) + \mathbf{u}(\omega)^T\big)$.
- RT imaging functional:

$$\mathcal{I}_{\rm RT}(\boldsymbol{z}^S) = \overline{\boldsymbol{g}(\omega, \boldsymbol{z}^S)}^T \mathbf{u}^s(\omega) \overline{\boldsymbol{g}(\omega, \boldsymbol{z}^S)}$$

where $\boldsymbol{g}$ is the normalized vector of Green's functions.

- KM imaging functional:

$$\mathcal{I}_{\rm KM}(\boldsymbol{z}^S) = \overline{\boldsymbol{d}(\omega, \boldsymbol{z}^S)}^T \mathbf{u}^s(\omega) \overline{\boldsymbol{d}(\omega, \boldsymbol{z}^S)}$$

where

$$\boldsymbol{d}(\omega, \boldsymbol{z}) = \frac{1}{\sqrt{N}} \Big(\exp\big(i\omega\mathcal{T}(\boldsymbol{x}_j, \boldsymbol{z})\big)\Big)_{j=1,\dots,N}$$

- MUSIC imaging functional:

$$\mathcal{I}_{\rm MU}(\boldsymbol{z}^S) = \frac{1}{\Big|\boldsymbol{g}(\omega, \boldsymbol{z}^S) - \big\langle \boldsymbol{v}_1(\omega), \boldsymbol{g}(\omega, \boldsymbol{z}^S)\big\rangle \boldsymbol{v}_1(\omega)\Big|^2}$$

where $\boldsymbol{v}_1(\omega)$ is the first vector of the symmetric SVD (Takagi factorization) of $\mathbf{u}^s(\omega) = \mathbf{V}(\omega)\boldsymbol{\Sigma}(\omega)\mathbf{V}(\omega)^T$.

- MUSIC imaging functional:

$$\mathcal{I}_{\rm MU}(\boldsymbol{z}^S) = \frac{1}{\big|\boldsymbol{g}(\omega,\boldsymbol{z}^S) - \big\langle \boldsymbol{v}_1(\omega), \boldsymbol{g}(\omega,\boldsymbol{z}^S)\big\rangle \boldsymbol{v}_1(\omega)\big|^2}$$

where $\boldsymbol{v}_1(\omega)$ is the first vector of the symmetric SVD of $\mathbf{u}^s(\omega) = \mathbf{V}(\omega)\boldsymbol{\Sigma}(\omega)\mathbf{V}(\omega)^T$.

Here:

- the unperturbed matrix $\mathbf{u}_0(\omega)$ has rank one. Its image space is spanned by its first singular vector $\boldsymbol{g}(\omega,\boldsymbol{z}_{\rm ref})$.
- $\big\langle \boldsymbol{v}_1(\omega), \boldsymbol{g}(\omega,\boldsymbol{z}^S)\big\rangle \boldsymbol{v}_1(\omega)$ is the projection of $\boldsymbol{g}(\omega,\boldsymbol{z}^S)$ on the "image space" of $\mathbf{u}^s(\omega)$.
- $\boldsymbol{g}(\omega,\boldsymbol{z}^S) - \big\langle \boldsymbol{v}_1(\omega), \boldsymbol{g}(\omega,\boldsymbol{z}^S)\big\rangle \boldsymbol{v}_1(\omega)$ is the projection of $\boldsymbol{g}(\omega,\boldsymbol{z}^S)$ on the "noise space" of $\mathbf{u}^s(\omega)$.

If there is no noise, then $\boldsymbol{v}_1(\omega) = \boldsymbol{g}(\omega,\boldsymbol{z}_{\rm ref})$ up to a phase term, and the projection on the noise space is zero if $\boldsymbol{z}^S = \boldsymbol{z}_{\rm ref}$.

Note: in the absence of noise:

$$\begin{aligned}
\mathcal{I}_{\rm MU}(\boldsymbol{z}^S)^{-1} &= 1 - \big|\big\langle \boldsymbol{g}(\omega,\boldsymbol{z}_{\rm ref}), \boldsymbol{g}(\omega,\boldsymbol{z}^S)\big\rangle\big|^2 = 1 - \frac{1}{\tau_{\rm ref}} \overline{\boldsymbol{g}(\omega,\boldsymbol{z}^S)}^T \mathbf{u}_0 \overline{\boldsymbol{g}(\omega,\boldsymbol{z}^S)} \\
&= 1 - \frac{1}{\tau_{\rm ref}} \mathcal{I}_{\rm RT}(\boldsymbol{z}^S)
\end{aligned}$$

- The noise matrix $\mathbf{W}^s = \frac{1}{2}(\mathbf{W} + \mathbf{W}^T)$ has entries such that $\mathbb{E}[|W^s_{jj}|^2] = \delta^2$ and $\mathbb{E}[|W^s_{jl}|^2] = \delta^2/2$.

- Bayesian analysis.

Given $\boldsymbol{z}$, τ, and δ, the symmetrized response matrix $\mathbf{u}^s$ has the probability density function

$$p\left(\mathbf{u}^s \mid \boldsymbol{z}, \tau, \delta\right) = \frac{1}{\pi^{\frac{N^2+N}{2}} \delta^{N^2+N}} \exp\left(-\frac{\|\mathbf{u}^s - \tau \boldsymbol{g}(\boldsymbol{z})\boldsymbol{g}(\boldsymbol{z})^T\|^2}{\delta^2}\right)$$

(with respect to the Lebesgue measure over the space of complex symmetric matrices)

Using Bayes theorem with the Jeffreys prior for the parameters τ, δ (non-informative prior distributions) and the uniform distribution in the search domain for $\boldsymbol{z}$, we find that, given the observations $\mathbf{u}^s$, the likelihood function of the parameters $\boldsymbol{z}$, τ, and δ is proportional to

$$l_0\left(\boldsymbol{z}, \tau, \delta \mid \mathbf{u}^s\right) = \frac{1}{\delta^{N^2+N+1}} \exp\left(-\frac{\|\mathbf{u}^s - \tau \boldsymbol{g}(\boldsymbol{z})\boldsymbol{g}(\boldsymbol{z})^T\|^2}{\delta^2}\right)$$

The maximum likelihood estimate of $\boldsymbol{z}$ and the nuisance parameters δ and τ are found by maximizing the likelihood function with respect to these:

$$\left(\hat{\boldsymbol{z}}, \hat{\tau}, \hat{\delta}\right) = \underset{\boldsymbol{z},\tau,\delta}{\operatorname{argmax}}\, l_0\left(\boldsymbol{z}, \tau, \delta \mid \mathbf{u}^s\right)$$

We first eliminate δ by requiring

$$\frac{\partial l_0\left(\boldsymbol{z}, \tau, \delta \mid \mathbf{u}^s\right)}{\partial \delta} = 0$$

which gives

$$\hat{\delta} = \frac{\|\mathbf{u}^s - \tau \boldsymbol{g}(\boldsymbol{z})\boldsymbol{g}(\boldsymbol{z})^T\|}{\sqrt{N^2+N+1}},$$

and then the likelihood ratio is proportional to

$$l_0\big(\boldsymbol{z}, \tau, \hat{\delta} \mid \mathbf{u}^s\big) \sim \|\mathbf{u}^s - \tau \boldsymbol{g}(\boldsymbol{z})\boldsymbol{g}(\boldsymbol{z})^T\|^{-(N^2+N+1)}$$

Since $\mathbf{u}^s$ is complex symmetric it admits a symmetric SVD: there exist unitary vectors $\boldsymbol{v}_j$ and nonnegative numbers σ_j (the singular values) such that

$$\mathbf{u}^s = \sum_{j=1}^{N} \sigma_j \boldsymbol{v}_j \boldsymbol{v}_j^T$$

Since $\|\boldsymbol{g}(\boldsymbol{z})\| = 1$, we have $\|\boldsymbol{g}(\boldsymbol{z})\boldsymbol{g}(\boldsymbol{z})^T\|_2 = 1$ and we then find that

$$\hat{\tau} = \underset{\tau}{\text{argmin}} \|\mathbf{u}^s - \tau \boldsymbol{g}(\boldsymbol{z})\boldsymbol{g}(\boldsymbol{z})^T\|^2 = \text{tr}\left(\mathbf{u}^s \overline{\boldsymbol{g}(\boldsymbol{z})\boldsymbol{g}(\boldsymbol{z})}^T\right) = \overline{\boldsymbol{g}(\boldsymbol{z})}^T \mathbf{u}^s \overline{\boldsymbol{g}(\boldsymbol{z})} = \sum_{j=1}^{N} \sigma_j \langle \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{v}_j \rangle^2$$

Therefore the estimate $\hat{\boldsymbol{z}}$ derives from maximizing the function

$$\hat{\boldsymbol{z}} = \underset{\boldsymbol{z}}{\text{argmin}} \left\| \mathbf{u}^s - \hat{\tau} \boldsymbol{g}(\boldsymbol{z})\boldsymbol{g}(\boldsymbol{z})^T \right\|^2$$

We have in fact

$$\begin{aligned} \left\| \mathbf{u}^s - \hat{\tau} \boldsymbol{g}(\boldsymbol{z})\boldsymbol{g}(\boldsymbol{z})^T \right\|^2 &= \left\| \mathbf{u}^s \right\|^2 - \Big| \sum_{j=1}^{N} \sigma_j \langle \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{v}_j \rangle^2 \Big|^2 \\ &= \left\| \mathbf{u}^s \right\|^2 - \left| \overline{\boldsymbol{g}(\boldsymbol{z})}^T \mathbf{u}^s \overline{\boldsymbol{g}(\boldsymbol{z})} \right|^2 \\ &= \left\| \mathbf{u}^s \right\|^2 - \left| \mathcal{I}_{\text{RT}}(\boldsymbol{z}) \right|^2 \end{aligned}$$

The maximum likelihood estimation of the reflector location is

$$\hat{\boldsymbol{z}} = \underset{\boldsymbol{z}}{\text{argmax}} \left| \mathcal{I}_{\text{RT}}(\boldsymbol{z}) \right|$$

- Conclusion: In the presence of additive noise (measurement noise): Reverse-Time imaging (or Kirchhoff migration) is optimal.

- Warning: Bayesian analysis is powerful but depends on the prior.
Here the prior is: there exists a reflector.