Sequential design for Bayesian inverse problems

Mascot-Num 2024

P. Lartaud (paul.lartaud@polytechnique.edu)
CEA DAM
Ecole polytechnique, CMAP

Outline

Outline

Some notations

- Consider $\mathbf{y} = (y^{(k)})_{1 \leq k \leq N}$ noisy observations of the direct model f:

$$y^{(k)} = f(x) + \varepsilon^{(k)} \in \mathbb{R} \text{ with } \varepsilon^{(k)} \sim \mathcal{N}\left(0, \sigma_m^2\right) \text{ iid} \tag{1}$$

- The direct model f is too expensive (Monte-Carlo simulation of neutron transport).
- We want the posterior distribution $p(x|\mathbf{y}) \propto L(\mathbf{y}|x)p(x)$ for $x \in \mathcal{X}$. The prior is chosen uniform on a compact.

Aleatoric and epistemic UQ in inverse problems

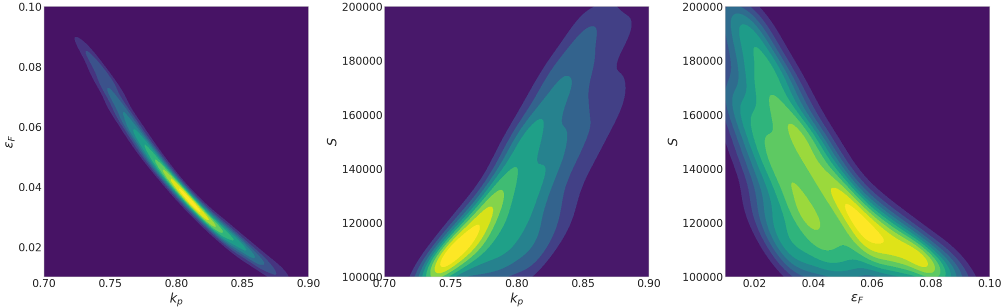
Let f_s be a GP surrogate with predictive distribution $f_s(x) \sim \mathcal{N}\left(\overline{f_s(x)}, k_s(x)\right)$.
Objective: include epistemic and aleatoric uncertainty in the posterior distribution, with a global covariance $\mathbf{C}_{\mathrm{tot}}(x) = k_s(x)\mathcal{A}_N + \sigma_m^2\mathcal{I}_N$.

$$\begin{aligned} p(x|\mathbf{y}) &\propto \frac{1}{\sqrt{(2\pi)^N|\mathbf{C}_{\mathrm{tot}}(x)|}} \exp\left[-\frac{1}{2}\left((\mathbf{y}-\overline{\mathbf{f_s}(x)})^T\mathbf{C}_{\mathrm{tot}}(x)^{-1}(\mathbf{y}-\overline{\mathbf{f_s}(x)})\right)\right] \\ &\propto (\sigma_m^2 + N k_s(x))^{-1/2} \exp\left[-\frac{1}{2}\left(\frac{\left(\overline{y}-\overline{f_s(x)}\right)^2}{k_s(x)+\frac{\sigma_m^2}{N}}\right)\right] \end{aligned} \tag{2}$$

where $\overline{\mathbf{f_s}(x)} = \left(\overline{f_s(x)}, \ldots, \overline{f_s(x)}\right)^T$ and $\mathcal{A}_N$ is the matrix of ones of size N.

An example

Problem statement: how do we choose new design points to enrich the surrogate ?



Posterior distribution sampled with HMC-NUTS with 10^5 samples

Outline

Some optimal designs

Consider a GP $f_s^{(n)}$ with predictive distribution $\mathcal{N}\left(m_n(x), k_n(x)\right)$.

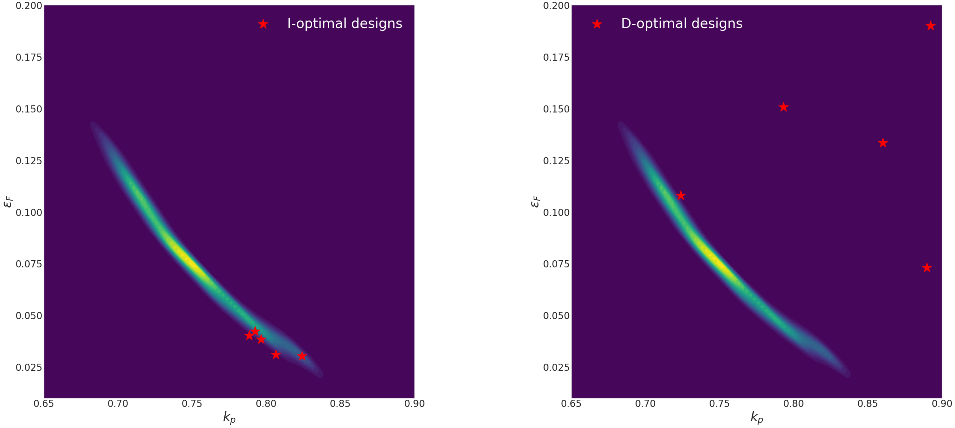
- D-optimal design: maximize the determinant of the predictive variance

$$x_{n+1} \in \mathrm{argmax}_{x \in \mathcal{X}} |k_n(x)|$$

- I-optimal design: minimize the integrated (updated) predictive variance

$$x_{n+1} \in \mathrm{argmin}_{x \in \mathcal{X}} \int_{\mathcal{X}} |k_{n+1}(\tilde{x}|x)| d\tilde{x}$$

Back to the inverse problem



Gaussian processes and Gaussian measures

Let $\mathbb{M}$ be the set of Gaussian measures on $\mathbb{F} = \mathcal{C}(\mathcal{X})$.

- For any $\nu \in \mathbb{M}$, there exists a GP $f_s \sim \mathcal{GP}(m_\nu(x), k_\nu(x, x'))$ with **continuous sample paths** whose probability distribution is ν [Vaart, Zanten, et al. 2008].
- The probability distribution P^f of any given GP f_s is a Gaussian measure on $\mathbb{F}$ i.e. $P^f \in \mathbb{M}$.
- Let f_s be a continuous GP with and probability distribution P^f. We define $P_n^f = P^f|(x_1, z_1, \dots, x_n, z_n) \in \mathbb{M}$ as the probability of f_s given $\mathcal{F}_n = \sigma(x_1, z_1, \dots, x_n, z_n)$.

SUR methods

Define a functional $\mathcal{H}: \mathbb{M} \to \mathbb{R}_+$ and denote by $\mathbb{E}_{n,x}$ the expectation given $\mathcal{F}_n$. The SUR strategy is:

$$\boxed{x_{n+1} \in \underset{x \in \mathcal{X}}{\operatorname{argmin}} \left\{ \mathbb{E}_{n,x} \left[\mathcal{H}(P^f_{n+1}) \right] \right\}} \tag{3}$$

Convergence results can be obtained provided the functional has the **supermartingale property** [Bect, Bachoc, and Ginsbourger 2019], i.e. for all $x \in \mathcal{X}$:

$$\mathbb{E}_{n,x} \left[\mathcal{H}(P^f_{n+1}) \right] \leq \mathcal{H}(P^f_n) \tag{4}$$

Outline

Constrained D-optimal design

Adapt the D-optimal design strategy to a well-chosen subset $\mathcal{B} \subset \mathcal{X}$.

$$x_{n+1} \in \underset{x \in \mathcal{B}}{\operatorname{argmax}} \; k_n(x)$$

Choose $\mathcal{B}$ to be close to the MAP $x_n^{(m)}$:

$$\mathcal{B}_h^{(n)} = \left\{ x \in \mathcal{X} \middle| \log p_n \left(x_m^{(n)} | \mathbf{y} \right) - \log p_n(x|\mathbf{y}) \leq h \right\}. \tag{5}$$

This defines a constraint set query (CSQ) design:

$$\boxed{x_{n+1} \in \operatorname{argmax}_{x \in \mathcal{B}_h^{(n)}} \; k_n(x)} \tag{6}$$

Outline

Metric of uncertainty for IPSUR design

Our metric is derived from *I*-optimal designs:

$$\mathcal{H}(\nu) = \mathbb{E}_{p_\nu}\left[k_\nu(\tilde{x})\right] = \int_{\mathcal{X}} k_\nu(\tilde{x}) p_\nu(\tilde{x}|\mathbf{y}) d\tilde{x} \tag{7}$$

The SUR criteria derived from this posterior-weighted predictive variance is **tractable** [Lartaud, Humbert, and Garnier 2024] and the metric can be evaluated with an ergodic Markov chain $(X_l)_{1\leq l\leq L}$. The IPSUR (Inverse Problem SUR) design is:

$$\boxed{x_{n+1} \in \underset{x\in\mathcal{X}}{\operatorname{argmin}} \left\{\mathbb{E}_{n,x}\left[\mathcal{H}(P^f_{n+1})\right]\right\}} \tag{8}$$

Convergence (1)

Supermartingale property

The functional $\mathcal{D}$ has the supermartingale property $\mathbb{E}_{n,x}\left[\mathcal{D}(P^f_{n+1})\right] \leq \mathcal{D}(P^f_n)$ where:

$$\mathcal{D}(\nu) = C_\nu \mathcal{H}(\nu) \tag{9}$$

$$C_\nu = \int_{\mathcal{X}} p(\tilde{x}) L_\nu(\mathbf{y}|\tilde{x}) d\tilde{x} \tag{10}$$

Remark: $\mathcal{H}$ does not have the supermartingale property.

Convergence (2)

Convergence of $\mathcal{H}(P_n^f)$

Under some regularity conditions, the uncertainty metric $H_n = \mathcal{H}(P_n^f)$ converges almost-surely to 0.

$$\boxed{\mathcal{H}(P_n^f) \xrightarrow[n \to +\infty]{a.s.} 0.} \tag{11}$$

Remark: the result holds for multi-output GP.

Two-step proof

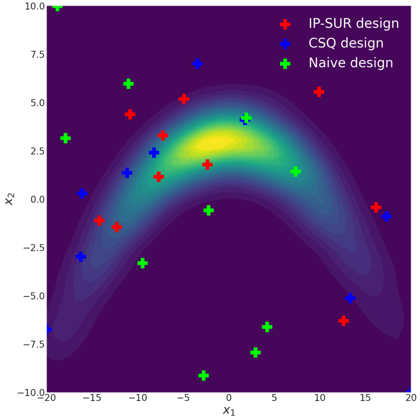
- First, show that $D_n = \mathcal{D}(P_n^f)$ converges a.s. to 0 using main theorem in [Bect, Bachoc, and Ginsbourger 2019].
  - Supermartingale property for $\mathcal{D}$.
  - Show that the zeros of $\mathcal{D}$ and zeros of $\mathcal{G}$ are the same where:

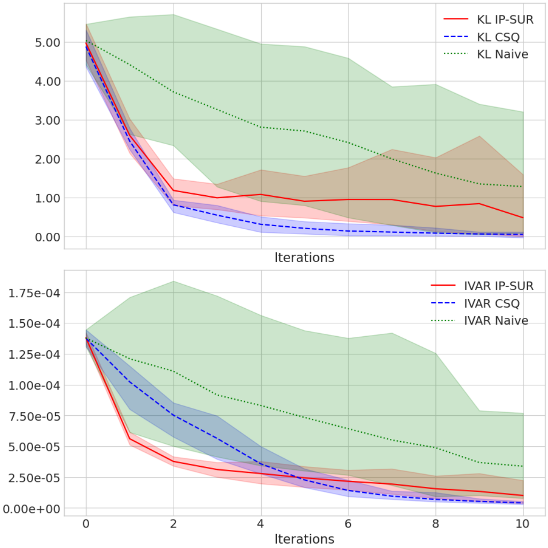
$$\mathcal{G}(\nu) = \sup_{x \in \mathcal{X}} \left(\mathcal{D}(\nu) - \mathbb{E}_z \left[\mathcal{D}(\nu | (x, z)) \right] \right) \tag{12}$$

- Prove that $H_n = \frac{D_n}{C_n}$ converges to 0 by showing that $C_n \xrightarrow[n \to +\infty]{} C_\infty > 0$

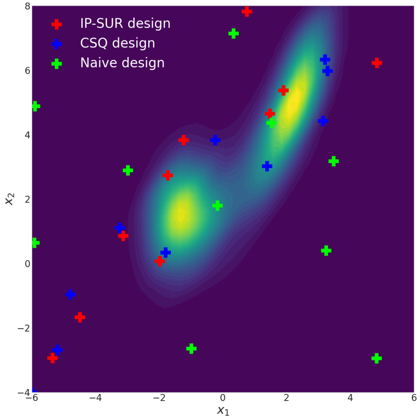
Outline

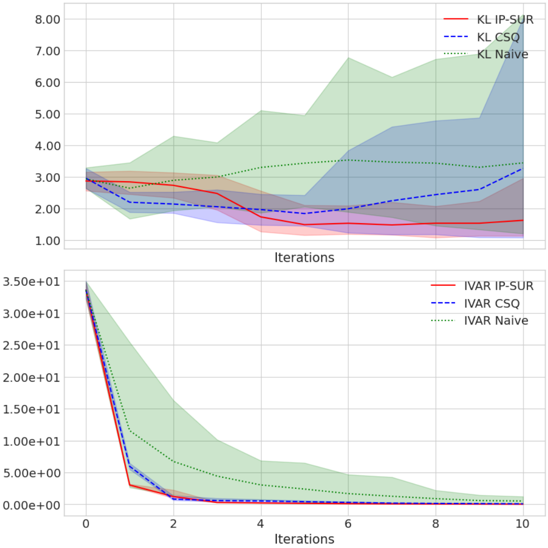
Test case 1


IP-SUR design
CSQ design
Naive design
x_1
x_2
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−20
−15
−10
−5
0
5
10
15
20

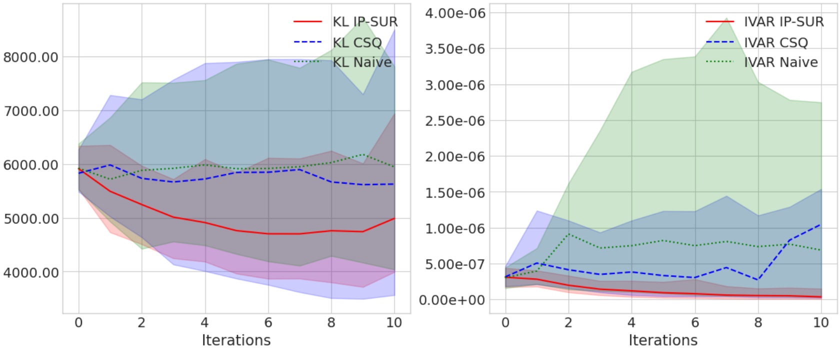

KL IP-SUR
KL CSQ
KL Naive
5.00
4.00
3.00
2.00
1.00
0.00
Iterations
IVAR IP-SUR
IVAR CSQ
IVAR Naive
1.75e-04
1.50e-04
1.25e-04
1.00e-04
7.50e-05
5.00e-05
2.50e-05
0.00e+00
0
2
4
6
8
10
Iterations

Test case 2


IP-SUR design
CSQ design
Naive design
x_1
x_2
8
6
4
2
0
−2
−4
−6
−4
−2
0
2
4
6


KL IP-SUR
KL CSQ
KL Naive
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
Iterations
IVAR IP-SUR
IVAR CSQ
IVAR Naive
3.50e+01
3.00e+01
2.50e+01
2.00e+01
1.50e+01
1.00e+01
5.00e+00
0.00e+00
0
2
4
6
8
10
Iterations

Application to neutron correlations



Conclusion

- Sequential design strategies specifically suited for Bayesian inverse problems.
- Tractability and guarantee of convergence for the IPSUR design.
- Possible extension to tempered posteriors with the same guarantee of convergence:

$$\mathcal{H}_{\beta}(\nu) = \frac{1}{C_{\nu,\beta}} \int_{\mathcal{X}} k_{\nu}(\tilde{x}) \left(L_{\nu}(\tilde{x}|\mathbf{y})\right)^{\beta} p(\tilde{x}) d\tilde{x} \tag{13}$$

$$C_{\nu,\beta} = \int_{\mathcal{X}} \left(L_{\nu}(\tilde{x}|\mathbf{y})\right)^{\beta} p(\tilde{x}) d\tilde{x} \text{ for } \beta \in [0,1] \tag{14}$$

References I

Bect, Julien, François Bachoc, and David Ginsbourger (2019). "A supermartingale approach to Gaussian process based sequential design of experiments". In: *Bernoulli* 25.4A, pp. 2883–2919. DOI: 10.3150/18-BEJ1074. URL: https://doi.org/10.3150/18-BEJ1074.

Lartaud, Paul, Philippe Humbert, and Josselin Garnier (2024). *Sequential design for surrogate modeling in Bayesian inverse problems*. arXiv: 2402.16520 [stat.ME].

Vaart, Aad W van der, J Harry van Zanten, et al. (2008). "Reproducing kernel Hilbert spaces of Gaussian priors". In: *IMS Collections* 3, pp. 200–222.

Thank you for your attention. Any questions ?

Appendix 1: SUR criteria

The SUR criteria is given by $x_{n+1} \in \text{argmin}_{x \in \mathcal{X}} \widetilde{J_n(x)}$ where $J_n(x) = \mathbb{E}_{z(x)}\left[H_{n+1}(x,z)\right]$:

$$\widetilde{J_n(x)} = \frac{C_{n+1}}{C_n} J_n(x) = \int_{\mathcal{X}} p_n(\tilde{x}|\mathbf{y}) h(\tilde{x},x) I(\tilde{x},x) d\tilde{x}.$$

$$h(\tilde{x},x) = k_{n+1}(\tilde{x},|x) \frac{|\Sigma_n(\tilde{x})|^{1/2}}{|\Sigma_{n+1}(\tilde{x}|x)|^{1/2}} \exp\left[-\frac{1}{2}\left(\|\mathbf{y}-\mathbf{m_n}(\tilde{x})\|^2_{\Sigma_{n+1}} - \|\mathbf{y}-\mathbf{m_n}(\tilde{x})\|^2_{\Sigma_n}\right)\right]$$

$$I(\tilde{x},x) = \frac{1}{\sqrt{\lambda_n(x,\tilde{x})\|\mathbf{u}\|^2_{\Sigma_{n+1}} + 1}} \exp\left[\frac{\lambda_n(x,\tilde{x})\langle \mathbf{y}-\mathbf{m_n}(\tilde{x})|\mathbf{u}\rangle^2_{\Sigma_{n+1}}}{2(\lambda_n(x,\tilde{x})\|\mathbf{u}\|^2_{\Sigma_{n+1}} + 1)}\right]$$

where $\lambda_n(x,\tilde{x}) = \frac{k_n(x,\tilde{x})^2}{k_n(x)}$ and $\mathbf{u} = (1,\dots,1)^T$.
We also introduced $\mathbf{m_n}(\tilde{x}) = m_n(x)\mathbf{u}$ where $m_n(\tilde{x})$, and $\Sigma_n(\tilde{x}) = k_n(x)\mathcal{A}_N + \sigma_m^2 \mathcal{I}_N$.