# Novel information criteria in Bayesian experimental design

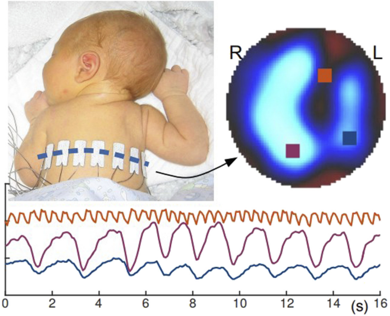

**Tapio Helin**
tapio.helin@lut.fi
School of Engineering Science
LUT University

LUT University

Research Council of Finland

October 28, 2025

# Motivating example: Electrical Impedance Tomography




- Identify electric conductivity inside body by boundary measurements
- Non-invasive and non-costly procedure
- Electric potential satisfies

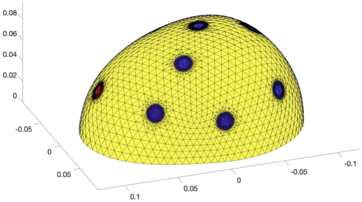
$$
\begin{aligned}
-\nabla \cdot \sigma \nabla u &= 0 \quad \text{in } \Omega \\
\sigma \mathbf{n} \cdot \nabla u &= g \quad \text{on } \partial\Omega
\end{aligned}
$$

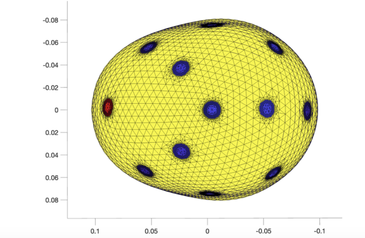
giving rise to the N-to-D mapping

$$\Lambda = \Lambda(\sigma) : g \mapsto u|_{\partial\Omega}$$

**Inverse problem:** given the Neumann-to-Dirichlet mapping $\Lambda$, recover conductivity $\sigma$

# Motivating example: Head imaging with EIT




- EIT is sensitive to conductivity changes when imaging a brain stroke

**Hyvönen N**, **Maaninen, J** and **Puska J-P**, *Bayesian experimental design for head imaging by EIT*, SIAM J. Appl. Math. 84 (4), 2024.

# Bayesian inference in inverse problems

Identify the unknown $x$ from noisy data $y$
where $y = \mathcal{G}(x; \theta) + \xi$.

# Bayesian inference in inverse problems

Identify the unknown $x$ from noisy data $y$
where $y = \mathcal{G}(x;\theta) + \xi$.

**Parameter-to-observable map:**

$$\mathcal{G}(\cdot;\theta) = E(\theta) \circ \mathcal{F} : \mathcal{X} \to \mathcal{Y},$$

with the forward map $\mathcal{F} : \mathcal{X} \to \mathcal{X}'$ and an
evaluation map $E : \mathcal{X}' \to \mathcal{Y} := \mathbb{R}^d$.

# Bayesian inference in inverse problems

Identify the unknown $x$ from noisy data $y$ where $y = \mathcal{G}(x;\theta) + \xi$.

**Parameter-to-observable map:**

$$\mathcal{G}(\cdot;\theta) = E(\theta) \circ \mathcal{F} : \mathcal{X} \to \mathcal{Y},$$

with the forward map $\mathcal{F} : \mathcal{X} \to \mathcal{X}'$ and an evaluation map $E : \mathcal{X}' \to \mathcal{Y} := \mathbb{R}^d$.

**Bayes formula:** $\mu_0 =$ prior, $\mu^y =$ posterior and

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(\cdot|\theta) = \frac{\pi(y|\cdot;\theta)}{\pi(y;\theta)}, \quad \text{where} \quad .$$

$$\pi(y;\theta) = \int_{\mathcal{X}} \pi(y|x;\theta)\mu_0(\mathrm{d}x)$$

# Bayesian inference in inverse problems

Identify the unknown $x$ from noisy data $y$ where $y = \mathcal{G}(x;\theta) + \xi$.

**Parameter-to-observable map:**

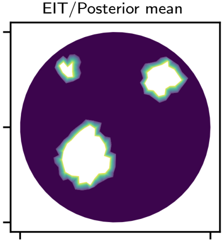
$$\mathcal{G}(\cdot;\theta) = E(\theta) \circ \mathcal{F} : \mathcal{X} \to \mathcal{Y},$$

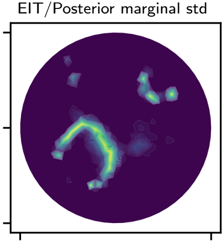
with the forward map $\mathcal{F} : \mathcal{X} \to \mathcal{X}'$ and an evaluation map $E : \mathcal{X}' \to \mathcal{Y} := \mathbb{R}^d$.

**Bayes formula:** $\mu_0 =$ prior, $\mu^y =$ posterior and

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(\cdot|\theta) = \frac{\pi(y|\cdot;\theta)}{\pi(y;\theta)}, \quad \text{where} \quad .$$

$$\pi(y;\theta) = \int_{\mathcal{X}} \pi(y|x;\theta)\mu_0(\mathrm{d}x)$$




@Alghamdi et al. 2024 (CUQIpy)

# Bayesian optimal experimental design

How to choose $\theta$? First pick a utility function[1]

$$U(\theta) = \mathbb{E}\, u(X, Y; \theta)$$

An optimal design $\theta_* \in \mathcal{D}$ maximizes $U$, that is

$$\theta_* \in \arg\max_{\theta \in \mathcal{D}} U(\theta).$$

Typically used utility: expected information gain

$$U(\theta) = \mathrm{EIG}(\theta) = \mathbb{E}\, \mathsf{KL}(\mu^Y, \mu).$$

[1] **Huan X et al.**, *Optimal experimental design: Formulations and computations*, Acta Numerica 33, 2024

# Computational challenges

Consider e.g. EIG:

$$\begin{aligned}
\theta_* &\in \operatorname*{arg\,max}_{\theta \in \mathcal{D}} \mathrm{EIG}(\theta) \\
&= \operatorname*{arg\,max}_{\theta \in \mathcal{D}} \left[ \iint_{\mathcal{Y} \times \mathcal{X}} \log \left( \frac{\pi(y|x;\theta)}{\pi(y;\theta)} \right) \pi(y|x;\theta) \mu_0(\mathrm{d}x)\mathrm{d}y \right]
\end{aligned}$$

$\Longrightarrow$ Computational cost prohibitive (double expectation combined with optimization)

# Computational challenges

Consider e.g. EIG:

$$\begin{aligned} \theta_* &\in \arg\max_{\theta \in \mathcal{D}} \mathrm{EIG}(\theta) \\ &= \arg\max_{\theta \in \mathcal{D}} \left[ \iint_{\mathcal{Y} \times \mathcal{X}} \log \left( \frac{\pi(y|x;\theta)}{\pi(y;\theta)} \right) \pi(y|x;\theta) \mu_0(\mathrm{d}x) \mathrm{d}y \right] \end{aligned}$$

$\Longrightarrow$ Computational cost prohibitive (double expectation combined with optimization)

---

Stable numerical approximation schemes necessary to enable study of large-scale problems!

# Stability of expected utility

Now let

$$\mathrm{EIG}_N(\theta) = \mathbb{E}\,\mathrm{KL}(\mu_N^Y \parallel \mu)$$

where

$$\frac{\mathrm{d}\mu_N^y}{\mathrm{d}\mu_0}(\cdot|\theta) = \frac{\pi_N(y|\cdot;\theta)}{\pi_N(y;\theta)}.$$

**Theorem (Duong-H-Rojo Garcia 2023)**

*Under suitable moment bounds, there exists $C > 0$ such that for all $N$ sufficiently large,*

$$\sup_{\theta\in\mathcal{D}} |\mathrm{EIG}(\theta) - \mathrm{EIG}_N(\theta)| \leq C\sqrt{\mathbb{E}^{\mu_0}\left[\mathrm{KL}\left(\pi_N(\cdot|X;\theta) \parallel \pi(\cdot|X;\theta)\right)\right]}.$$

# Special case: Gaussian likelihood

Consider the inverse problem

$$y = \mathcal{G}(x;\theta) + \xi$$

with noise $\xi \sim \mathcal{N}(0, \Gamma)$.

**Theorem (Duong-H-Rojo Garcia 2023)**

*Suppose* $\mathcal{G}, \mathcal{G}_N \in L^\infty_\theta(\mathcal{D}, L^4_{\mu_0}(\mathcal{X}, \mathbb{R}^d))$. *Then*

$$\sup_{\theta\in\mathcal{D}} |\mathrm{EIG}(\theta) - \mathrm{EIG}_N(\theta)| \leq C\sqrt{\mathbb{E}^{\mu_0} \|\mathcal{G}(X;\theta) - \mathcal{G}_N(X;\theta)\|^2_\Gamma},$$

*for all* $N$ *sufficiently large.*

# Stability of optimal design

**Theorem (Duong-H-Rojo Garcia 2023)**

*On top of previous assumptions add continuity of*

$$\tilde{\theta} \mapsto \mathbb{E}^{\mu_0}\left[\mathrm{KL}\left(\pi(\cdot|X;\tilde{\theta}) \parallel \pi(\cdot|X;\theta)\right)\right]$$

*around some neighbourhood of any $\theta \in \mathcal{D}$. Now suppose*

$$\theta_N^* \in \arg\max_{\theta \in \mathcal{D}} \mathrm{EIG}_N(\theta).$$

*Then, the limit $\theta^*$ of any converging subsequence of $\{\theta_N^*\}_{N=1}^{\infty}$ is a maximizer of $\mathrm{EIG}$, that is,*

$$\theta^* \in \arg\max_{\theta \in \mathcal{D}} \mathrm{EIG}(\theta).$$

*Moreover,*

$$\limsup_{N \to \infty} \mathrm{EIG}_N(\theta_N^*) = \mathrm{EIG}(\theta^*).$$

# EIG/stability - Key takeaways

- Numerical evaluation of EIG has received considerable attention, whereas the impact of model approximations has been less explored
- EIG and corresponding optimal designs are stable wrt likelihood perturbations
- What about mispecified/empirical/data-driven priors?

# Wasserstein information criterion

Define $\mathcal{P}_p(\mathcal{X}) = \{\mu \text{ Borel prob. measure on } \mathcal{X} \mid M_p(\mu) < \infty\}$ with $M_p(\mu) = \int_{\mathcal{X}} \|x\|_{\mathcal{X}}^p \, \mu(dx)$.

For $\mu_1, \mu_2 \in \mathcal{P}(\mathcal{X})$ and $p \in [1, \infty)$, we define

$$W_p(\mu_1, \mu_2) = \left( \inf_{\rho \in \Xi(\mu_1, \mu_2)} \int_{\mathcal{X} \times \mathcal{X}} \|x - w\|^p \, \mathrm{d}\rho(x, w) \right)^{1/p},$$

where $\Xi(\mu_1, \mu_2)$ is the set of all couplings between $\mu_1$ and $\mu_2$.

Wasserstein information criterion:[2]

$$U_p(\theta) = \mathbb{E}^{\pi(y;\theta)} W_p^p(\mu, \mu^y(\cdot;\theta)).$$

[2] **Helin, Marzouk and Rojo-Garcia**, *Bayesian optimal experimental design with Wasserstein information criteria*, arxiv:2504.10092

## WOED - what properties

- Well-defined for infinite-dimensional $\mathcal{X}$
- Bayesian utility[3]; depends on $\mu_{post}$
- Information measure according to Ginebra[4]: (*i*) real-valued, (*ii*) yields zero for non-informative experiment and (*iii*) satisfies sufficiency ordering

**Statistical sufficiency:** We say $Y|X, \theta_1$ is sufficient for $Y|X, \theta_2$ if there exists a random variable $\eta$ with a known probability distribution and a function $W$ such that

$$W(Y, \eta)|x, \theta_1 \quad = \quad Y|x, \theta_2$$

in distribution for all $x$.

[3] **Ryan E et al.**, *A Review of Modern Computational Algorithms for Bayesian Optimal Design*, ISR 84(1) 2016

[4] **Ginebra J**, *On the measure of the information in a statistical experiment*, Bayesian Analysis 2(1) 2007

# W-OED: Linear and Gaussian case

Recall that we consider

$$y = \mathcal{G}(x;\theta) + \xi, \quad \xi \sim \mathcal{N}(0, \Gamma).$$

**Theorem (H-Marzouk-Rojo-Garcia 25+)**

*Suppose $\mathcal{G}$ is linear and the prior $\mu_0$ is a centred Gaussian with covariance $C_0$ We have that the expected utility $U_2$ satisfies*

$$U_2(\theta) = 2\mathrm{tr}_{\mathcal{X}}\left(C_0\right) - 2\mathrm{tr}_{\mathcal{X}}\left(\left(C_0^{\frac{1}{2}} C_{post}(\theta) C_0^{\frac{1}{2}}\right)^{\frac{1}{2}}\right),$$

*where $C_{post} = \left(\mathcal{G}^*\Gamma^{-1}\mathcal{G} + C_0^{-1}\right)^{-1}$.*

# W-OED: Stability

**Assumption (A)**

*The following conditions hold for the mapping $\mathcal{G} : \mathcal{X} \to \mathbb{R}^d$ and the Borel probability measure $\mu$ on $\mathcal{X}$:*

(i) *(Lipschitz) There exists $L_1 > 0$ such that*

$$\left\|\mathcal{G}(x) - \mathcal{G}(x')\right\|_\Gamma \leq L_1 \left\|x - x'\right\|$$

*for all $x, x' \in \mathcal{X}$.*

(ii) *(sub-Gaussian prior) There exists $L_2 > 0$ such that*

$$\mathbb{E}^\mu \exp\left(L_2 \left\|x\right\|^2\right) < \infty.$$

(iii) *($\mathcal{G}$ is proper) There exists $R, L_3 > 0$ such that $\mu(B(0, R)) > 0$ and $\sup_{x \in B(0,R)} \left\|\mathcal{G}(x)\right\|_\Gamma < L_3$.*

# W-OED: Likelihood stability $p = 1$

**Theorem (H-Marzouk-Rojo-Garcia 25+)**

*Suppose $\mathcal{G}$ and $\mathcal{G}_*$ satisfy Assumption A with probability measure $\mu$ on $\mathcal{X}$ and with same constants. Moreover, we assume $L_1^2 < \frac{\sqrt{2}-1}{2} L_2$. Then there exist constants $K_1$ and $K_2$ such that:*

(i) *The evidence averaged posterior perturbation is bounded by*

$$\mathbb{E}^{\pi(y)} W_1(\mu^y, \mu_*^y) \leq K_1 \left( \mathbb{E}^{\mu} \|\mathcal{G}(x) - \mathcal{G}_*(x)\|_\Gamma^2 \right)^{\frac{1}{2}}.$$

(ii) *The perturbation of the expected $U_1$-utility satisfies*

$$|U_1 - U_1^*| \leq K_2 \left( \mathbb{E}^{\mu} \|\mathcal{G}(x) - \mathcal{G}_*(x)\|_\Gamma^2 \right)^{\frac{1}{2}}.$$

# W-OED: Prior stability $p = 1$

**Theorem (H-Marzouk-Rojo-Garcia 25+)**

*Suppose* $\mathcal{G}$ *satisfies Assumption (A) with two probability measures* $\mu$ *and* $\tilde{\mu}$ *on* $\mathcal{X}$*, and* $L_1^2 < \frac{\sqrt{3}-1}{2} L_2$*. Then there exist constants* $K_1$ *and* $K_2$ *such that*

(i) *The evidence averaged posterior perturbation is bounded by*

$$\mathbb{E}^{\pi(y)} W_1(\mu^y, \tilde{\mu}^y) \leq K_1 W_2(\mu, \tilde{\mu}),$$

(ii) *The perturbation of the expected* $U_1$*-utility satisfies*

$$|U_1 - \widetilde{U}_1| \leq W_1(\mu, \tilde{\mu}) + K_2 W_2(\mu, \tilde{\mu}).$$

# W-OED: Prior stability $p = 2$

**Theorem (H-Marzouk-Rojo-Garcia 25+)**

*Suppose $\mathcal{G}$ satisfies Assumption A with two probability measures $\mu$ and $\tilde{\mu}$ on $\mathcal{X}$, and $L_1^2 < \frac{\sqrt{3}-1}{2} L_2$. Then there exist constants $K_1$ and $K_2$ such that*

$$|U_2 - \widetilde{U}_2| \leq K_1 W_2(\mu, \tilde{\mu}) + K_2 \sqrt{\mathbb{E}^{\pi(y)} W_2^2(\mu^y, \tilde{\mu}^y)}.$$

To-do: establish bound for the posterior perturbation!

# Wasserstein information criterion - Key takeaways

- Novel information criterion with explicit formula for Gaussian distributions
- WOED ($p = 1$) is stable with respect to prior perturbations
- Kerrigan et al. (`arXiv:2510.14848`) extend the idea to general transport distances

# Back to EIG: maximum entropy sampling

Recall that EIG satisfies

$$\begin{aligned}
\mathrm{EIG}(\theta) &= \mathbb{E}\,\mathrm{KL}(\mu^{Y}, \mu) \\
&= -\int_{\mathbb{R}^d} \log\left(\pi\left(y;\theta\right)\right)\pi(y;\theta)\,\mathrm{d}y \\
&\qquad + \iint_{\mathbb{R}^d\times\mathcal{X}} \log\left(\pi\left(y \mid x;\theta\right)\right)\pi(y \mid x;\theta)\,\mathrm{d}y\,\mu(\mathrm{d}x) \\
&= \mathrm{Ent}(\pi(\,\cdot\,;\theta)) - \mathbb{E}^{\mu}\,\mathrm{Ent}(\pi(\,\cdot \mid X;\theta)),
\end{aligned}$$

Suppose our likelihood emerges from

$$y = \mathcal{G}(x;\theta) + \epsilon,$$

where $\epsilon \sim \eta$. Since differential entropy is translation-invariant with respect to the density, we have

$$\mathbb{E}^{\mu}\,\mathrm{Ent}(\pi(\,\cdot \mid X;\theta)) = \mathrm{Ent}(\eta).$$

# EIG with maximum entropy sampling

We have

$$\mathrm{EIG}(\theta) = \mathsf{Ent}(\pi(\,\cdot\,;\theta)) + \mathsf{const}$$

Divide the OED task into two sub-tasks:

1. Density estimation of the evidence $\pi(y) \approx \pi_M^K(y)$ based on prior samples of the unknown $x$, push-forwarded through an approximate version of $\mathcal{G}$ that can be obtained, e.g., via discretization of the underlying PDE
2. Estimation of the differential entropy of the approximated evidence $\rightarrow$ can be efficiently evaluated with off-the-shelf kernel density estimation software packages

# Evidence approximations with Gaussian mixture model

- Likelihood
$$y = \mathcal{G}(x;\theta) + \epsilon,$$
where $\epsilon \sim N(0, \Gamma)$
- Let $\{x_m\}_{m=1}^M \subset \mathcal{X}$ be unsupervised training data
- $\mathcal{G}_K$ approximates $\mathcal{G}$
- Set
$$\pi_M^K(y) = \frac{1}{M}\sum_{m=1}^{M} \mathcal{N}(\mathcal{G}(x_m), \Gamma)$$

We utilize estimator[5]

$$\widehat{J}_{M,N}^K(\theta) = -\frac{1}{N}\sum_{n=1}^{N} \log\left(\pi_M^K\left(Y_n;\theta\right)\right)$$

---

[5] **Chen, Helin, Hyvönen and Suzuki**, *Approximation of differential entropy in Bayesian optimal experimental design*, arXiv:2510.00734

# Monte Carlo based Gaussian mixture models

Define

$$\delta_K := \sqrt{\mathbb{E}^{\mu} \left\| \mathcal{G}(X;\theta) - \mathcal{G}_K(X;\theta) \right\|_{\Gamma}^2}$$

**Theorem (Chen-Helin-Hyvönen-Suzuki 25+)**

*Suppose Assumption (A) holds for $\mathcal{G}$ and $\mathcal{G}_K$ with $L_1^2 < \frac{1}{12} L_2$ and let $x_m \sim \mu$, $m = 1, \ldots, M$ be i.i.d. We have*

$$\mathbb{E}^{\otimes \mu} \mathbb{E}^{\otimes \pi_M^K} \left| J - \widehat{J}_{M,N}^K \right|^2 \leq C \left( \delta_K^2 + \frac{1}{M} + \frac{1}{N} \right) \tag{1}$$

*for some constant $C$ and all $K, M, N > 0$.*

- The core part of the proof relies on Goldfeld et al. 2020 (also minimax!)

# QMC-based Gaussian mixture models

**Theorem (Chen-Helin-Hyvönen-Suzuki 25+)**

- *$\mathcal{G}$ and $\mu$ satisfy Assumption (A)*
- *$\mathcal{X}_K \subset \mathcal{X}$, $K \in \mathbb{N}$ subspace isomorphic to $\mathbb{R}^K$*
- *projection of $\mu$ to $\mathcal{X}_K$ is the uniform measure over $[0,1]^K$*
- *$\mathcal{G}_K \in W^{1,\infty}_{\mathrm{mix}}([0,1]^K)^d$, extended canonically to $\mathcal{X}$*

*Let $\{X_m\}_{m=1}^M$ be the randomized lattice points with the generating vector z constructed by the component-by-component algorithm (Kuo 2003). Then, for any $\gamma > 0$,*

$$\mathbb{E}^{\Delta}\mathbb{E}^{\otimes \pi_M^K}\left|J - \widehat{J}_{M,N}^K\right|^2 = C\left(\delta_K^2 + \frac{1}{M^{2-\gamma}} + \frac{1}{N}\right), \tag{2}$$

*where the constant $C > 0$ depends on $K$ and $\gamma$.*

# Maximum entropy sampling - key takeaways

- Applicable when the noise model is well-understood
- Evidence approximation through push-forward samples and estimation of the differential entropy decouple
- We applied GMM approximation to evidence motivated by convergence results from Goldfeld et al. 2020 - other density estimation methods apply as well
- Being able to design the training points for density estimation (QMC) improves convergence rate

# Conclusions

- Bayesian OED in large-scale inverse problems requires significant computational effort
- Stability under numerical approximations is key for efficient algorithms
- Wasserstein information criterion provides a novel utility for Bayesian OED
- For EIG, maximum entropy sampling is a potential alternative for nested methods when applicable

**Helin T, Marzouk Y and Rojo-Garcia R**,
*Bayesian optimal experimental design with Wasserstein information criteria*, arxiv:2504.10092

**Duong D-L**, **Helin T** and **Rojo-Garcia R**,
*Stability estimates for the expected utility in Bayesian optimal experimental design*, Inverse Problems 39 (12), 2023.

**Chen C**, **Helin T**, **Hyvönen N** and **Suzuki Y**,
*Approximation of differential entropy in Bayesian optimal experimental design*, arXiv:2510.00734

# If time permits: Numerical tests for WOED

- Approximate prior $\mu$ by empirical measure: $\mu_M = \frac{1}{M}\sum_{m=1}^{M}\delta(x - x^m), \quad x^m \sim \mu$ i.i.d.
- Corresponding posterior:

$$\mu_M^y = \frac{1}{M}\sum_{m=1}^{M} w_m^y \delta(x - x^m), \quad w_m^y = \frac{\exp(-\Phi(x^m, y))}{\sum_k \exp(-\Phi(x^k, y))}.$$

- It is well-known that empirical approximations of one-dimensional Gaussian distributions satisfy

$$\mathbb{E}^{\otimes\mu} W_1(\mu, \mu_M) \lesssim \frac{1}{\sqrt{M}} \quad \text{and} \quad \mathbb{E}^{\otimes\mu} W_2(\mu, \mu_M) \lesssim \sqrt{\frac{\log\log M}{M}},$$

  where $\mathbb{E}^{\otimes\mu}$ stands for the ensemble average and the proportionality constants are universal.
- **Toy example:** $\mathcal{G}(x;\theta) = 5\theta^6 x$ on $\mathbb{R} \times [-1, 1]$, prior $\mu = \mathcal{N}(0, 1)$, noise $\epsilon \sim \mathcal{N}(0, 0.05^2)$. Posterior is Gaussian $\Rightarrow$ exact utility evaluation.

# Numerical tests: Stability in 1d

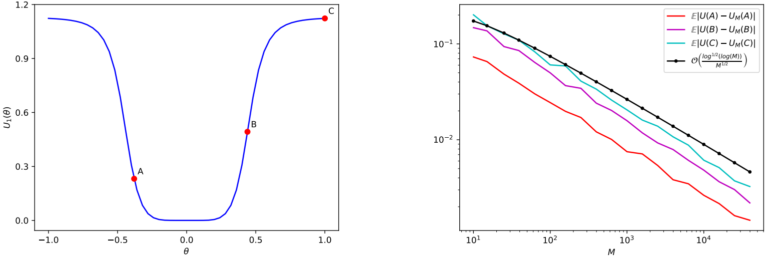



Figure: Expected utility in the one-dimensional linear problem: (a) true expected utility $U_1(\theta)$; (b) convergence behavior under empirical prior approximations.

# Numerical tests: Heat diffusion

**Theorem (Brenier's theorem)**

*Let $\mu_1, \mu_2 \in \mathcal{P}_p(\mathcal{X})$ be probability measures. There exist an optimal coupling $\gamma^* \in \Gamma(\mu_1, \mu_2)$, such that*

$$W_p^p(\mu_1, \mu_2) = \int_{\mathcal{X}\times\mathcal{X}} \|x_1 - x_2\|^p \gamma^*(dx_1, dx_2).$$

*Moreover, if $\mathcal{X} = \mathbb{R}^n$, $p = 2$ and $\mu_1$ is absolutely continuous respect to the Lebesgue measure, then there exists a unique (up to a constant) convex and almost everywhere differentiable potential function $\varphi$ such that $T(x) = \nabla\varphi(x)$ and*

$$W_2^2(\mu_1, \mu_2) = \int_{\mathbb{R}^n} \|x - T(x)\|^2 \mu_1(dx).$$

We solve $\phi$ through the Monge-Ampere equations on $\mathbb{R}^2$.

# Numerical tests: Heat diffusion

- Consider heat diffusion in the unit square $\Omega$ with source $S$ and Neumann boundary conditions:

$$\frac{\partial v}{\partial t} = \Delta v + S(z, t, x), \quad \nabla v \cdot n = 0, \quad v(z, 0) = 0,$$

where the source term satisfies

$$S(z, t, x) = \begin{cases} \frac{1}{\pi h^2} \exp\left(-\frac{\|z-x\|^2}{2h^2}\right), & 0 \leq t < \tau \\ 0, & t \geq \tau \end{cases}, \quad h = 0.05,\ \tau = 0.3.$$

- Design goal: find optimal sensor location $\theta \in \Omega$ to recover the source $x$ from observations $y = [v(\theta, t_1), \ldots, v(\theta, t_5)] \in \mathbb{R}^5$.
- Forward map: $\mathcal{G} : \Omega \times \Omega \to \mathbb{R}^5$, where $y = \mathcal{G}(x; \theta)$.

# Numerical tests: Heat diffusion

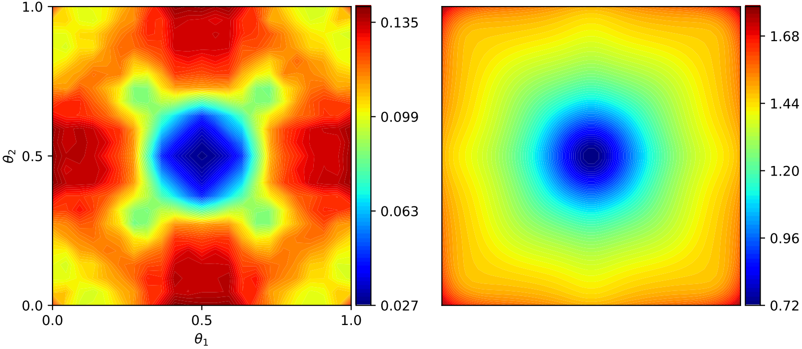



Figure: Expected utility functions $U(\theta)$ for heat diffusion. Left: Wasserstein-2 information criterion. Right: EIG criterion.

# Numerical tests: Heat diffusion

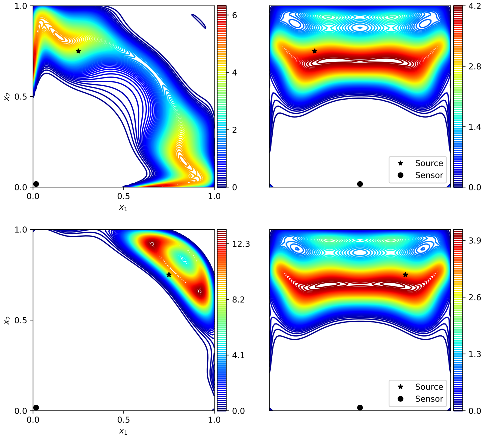



Figure: Posterior densities generated at optimal design nodes. Each row shows posteriors for different ground truth $x$: (a) $x = (0.25, 0.75)$ and (b) $x = (0.75, 0.75)$. In each subfigure, the left panel corresponds to the EIG optimal design and the right to the Wasserstein-2 distance optimal design.