THE UNIVERSITY *of* EDINBURGH

Afonso Eduardo

November 2022

1 Likelihood-Free Inference with Informative Surrogates

1.1 Motivation

Likelihood-Free Inference (LFI), which can be framed as a conditional density estimation problem [see, e.g., Sections 2.4 and 2.6 1, and Appendix A], presents at least two challenges.[1]

A first challenge is related to the accuracy of approximations given unlimited computational power. Even under this ideal condition, poor summary statistics lead to a large approximation error. On the other hand, sufficient statistics yield zero approximation error. In general, the more informative are the summary statistics, the lower is the approximation error. Automatic discovery of informative summary statistics is an important topic, but for now it falls beyond the scope of this project.

The other challenge is conditional density estimation on a budget. At least two problems need to be solved efficiently. One is that of identifying regions of non-negligible density and the other of correctly estimating regions of interest, typically those of highest density. In this sense, informative models that express preferences for these regions can lead to more sample-efficient LFI. For example, in `BOLFI` [2], high density regions correspond to subspaces where a discrepancy is small. In turn, this discrepancy is assumed to be conditionally normally distributed with a Gaussian process prior. Then, more informative GP priors, such as those proposed in `BOIC` [3], are directly applicable. In fact, as I wrote in [Section 6 1], "finding the region near the minimum (small discrepancies) appears to be the major difficulty in high-dimensional problems. For this reason, many of the recent advances in high-dimensional Bayesian optimization, including high-dimensional regression, seem to be directly applicable to high-dimensional likelihood-free inference."

1.2 Research Questions

Q1: Can more informative surrogates increase the efficiency of LFI methods?

While the motivation can be traced back to `BOLFI` [2, 1], the significance of more informative surrogates extends to other amortized LFI methods [e.g., 4, 5, 6]. The project focuses first on the application of GP surrogates with informative covariance functions [3] to existing LFI methods.

Later, there is the possibility to broaden the scope of the project to include, e.g., other space transformations [e.g., 7, Appendix C], informative priors over functions with low effective dimensionality, neural models and the design of new or modifications to existing LFI methods [see, e.g., Appendix A].

Q2 (optional): Is it possible to design more efficient LFI methods?

1.3 KL-UCB with Informative Covariance

In `KL-UCB` [6], the goal is to approximate the posterior $p(\boldsymbol{\theta}|\boldsymbol{x}) \propto p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$, where, in addition to the normalizing constant being unknown, the (log-)likelihood $\ell(\boldsymbol{\theta}) = \log p(\boldsymbol{x}|\boldsymbol{\theta})$ is an expensive function, only available via noisy

[1] This is a restatement of some challenges I identified in [Section 3 1].

estimates.[2] To this end, KL-UCB minimizes the KL divergence between an approximation q and the true posterior $p(\boldsymbol{\theta}|\boldsymbol{x})$.[3] This is equivalent to the maximization of the evidence lower bound (ELBO),

$$q^\star = \arg\max_q \mathbb{E}_{\boldsymbol{\theta}\sim q}[\ell(\boldsymbol{\theta})] - \mathrm{KL}(q(\boldsymbol{\theta}) \parallel p(\boldsymbol{\theta})), \tag{1.1}$$

where $p(\boldsymbol{\theta})$ is the prior. Then, a surrogate GP model approximates the log-likelihood ℓ by assuming it belongs to a RKHS and that the observation model is additive with zero mean and sub-Gaussian. In more detail, it replaces ℓ by an upper confidence bound (UCB) $\ell_n^u(\boldsymbol{\theta}) = m_n(\boldsymbol{\theta}) + \beta_n\sqrt{v_n(\boldsymbol{\theta})}$, where m_n and v_n are the posterior predictive mean and variance given by the surrogate. The resulting objective, essentially a UCB on ELBO, is maximized when $q_{n+1}(\boldsymbol{\theta}) \propto p(\boldsymbol{\theta})\exp(\ell_n^u(\boldsymbol{\theta}))$, for which MCMC provides a sample-based approximation $q_{n+1}(\boldsymbol{\theta}) \approx 1/S\sum_i \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_{n+1,i})$. At each step, S samples are evaluated $\hat{\ell}_{n+1,i} \sim \ell(\boldsymbol{\theta}_{n+1,i})$, and added to the training set $\mathcal{D}_{n+1} = \mathcal{D}_n \cup \{(\boldsymbol{\theta}_{n+1,i}, \hat{\ell}_{n+1,i})\}_{i\leq S}$. The final posterior approximation uses the posterior predictive mean as log-likelihood.

Limitations. Since KL-UCB uses a standard GP model and UCB, it has similar limitations as those found in standard BO with UCB/LCB. For instance, regret shows exponential growth with dimensionality [see, e.g., Section 6.2 and Appendix D 6]. There are also limitations when it comes to the estimation of β_n and theoretical guarantees. It assumes that 1) a concentration bound holds with high probability, 2) ℓ belongs to a known RKHS, 3) an upper bound on the RKHS norm is known. Beyond the toy problem with known RKHS [Section 7.1 6], $\beta_n = 3$ is used, and theoretical results no longer hold. The code by Oliveira et al. [6] only supports GP surrogates with fixed parameters.

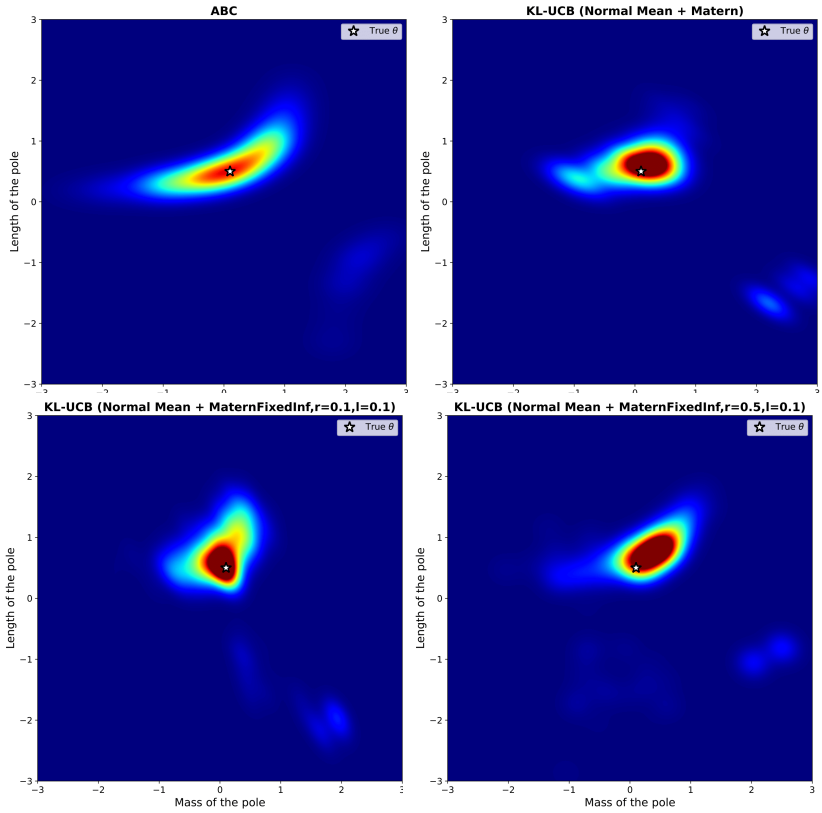

Figure 1.1. Cart-pole posterior approximations. The reference posterior is obtained by ABC Rejection [Figure 4 6]. Other posteriors are estimated via KL-UCB with the same seed, but different GP surrogates. The posterior approximations on the bottom row use log-likelihood surrogates with informative covariances (fixed parameters) [3]. Even in low-dimensional problems, informative surrogates appear to be useful.

[2] Unlike some previous works on LFI, direct access to noisy estimates of the log-likelihood is assumed. This simplifies the methodology because there is no longer the need to address problems typically found in Approximate Bayesian Computation (ABC), i.e., on how to approximate the likelihood function by simulations (simulated data). For instance, there is no need to specify or estimate summary statistics, a discrepancy function and a tolerance.

[3] It differs from objectives in Appendix A by reversing KL.

1.4 Other Methods

In addition to `KL-UCB` and `BOLFI`, there are other LFI methods that use GP surrogates. For instance, in Variational Bayesian Monte Carlo (`VBMC`) [5], posterior approximations are found by maximizing a lower confidence bound (LCB) on ELBO,

$$h_{\text{VBMC}}(q|\mathcal{D}_n) = \mathbb{E}_q[m_n] - \text{KL}(q \parallel p) - \beta\sigma_n(q), \text{ with } \sigma_n^2(q) = \iint C_n(\boldsymbol{\theta}, \boldsymbol{\theta}')q(\boldsymbol{\theta})q(\boldsymbol{\theta}')d\boldsymbol{\theta}d\boldsymbol{\theta}', \tag{1.2}$$

fixed β and posterior covariance C_n. Naturally, the choice of covariance function C influences q and acquisitions, chosen according to $\boldsymbol{\theta}^\star = \arg\max \sigma_n^2(\boldsymbol{\theta})q_{n+1}(\boldsymbol{\theta})\exp(m_n(\boldsymbol{\theta}))$. Another method is (Ellipsoidal) Robust Optimization Monte Carlo [`ROMC` 4, Algorithm 3]. Here, informative surrogates may be particularly useful because the method must solve many different but related BO problems, i.e., information from one problem can accelerate optimization of related problems by transfer learning.

References

[1] A. Eduardo. "High-dimensional Bayesian Optimization for Learning in Generative Models". Master's thesis. University of Edinburgh, 2018. URL: https://project-archive.inf.ed.ac.uk/msc/20182903/msc_proj.pdf.

[2] M. U. Gutmann, J. Corander, et al. "Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models". In: *JMLR* (2016).

[3] A. Eduardo and M. U. Gutmann. "Bayesian Optimization with Informative Covariance". In: *arXiv preprint arXiv:2208.02704* (2022).

[4] B. Ikonomov and M. U. Gutmann. "Robust Optimisation Monte Carlo". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pages 2819–2829.

[5] L. Acerbi. "Variational Bayesian Monte Carlo with Noisy Likelihoods". In: *Advances in Neural Information Processing Systems* 33 (2020), pages 8211–8222.

[6] R. Oliveira et al. "No-regret Approximate Inference via Bayesian Optimisation". In: *Uncertainty in Artificial Intelligence*. PMLR. 2021, pages 2082–2092.

[7] C. Oh et al. "BOCK: Bayesian optimization with cylindrical kernels". In: *ICML*. PMLR. 2018, pages 3868–3877.

[8] D. Greenberg et al. "Automatic Posterior Transformation for Likelihood-Free Inference". In: *International Conference on Machine Learning*. PMLR. 2019, pages 2404–2414.

A LFI and Density Estimation

In Likelihood-Free Inference (LFI), we have access to a simulator $\boldsymbol{x} \sim f(\cdot \mid \boldsymbol{\theta})$, but cannot directly evaluate the likelihood function $f(\boldsymbol{x}_o \mid \boldsymbol{\theta})$, where $\boldsymbol{x}_o$ is the observed data. One approach to solving this problem is to assume we have a flexible density estimator $q_{\boldsymbol{\phi}}$ parametrized by $\boldsymbol{\phi}$, where $\boldsymbol{\phi}$ may be the output and parameters of a neural network. We then try to minimize a divergence between the joint density $p(\boldsymbol{\theta}, \boldsymbol{x}) = f(\boldsymbol{x} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})$ and our approximation $q_{\boldsymbol{\phi}}(\boldsymbol{\theta}, \boldsymbol{x})$. Note that we cannot evaluate $p(\boldsymbol{\theta}, \boldsymbol{x})$, but if we choose to minimize the (forward) KL divergence, we can write

$$\mathrm{D}_{\mathrm{KL}}(p \,||\, q_{\boldsymbol{\phi}}) = \mathbb{E}_{p(\boldsymbol{\theta},\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{\theta}, \boldsymbol{x})}{q_{\boldsymbol{\phi}}(\boldsymbol{\theta}, \boldsymbol{x})}\right] \tag{A.1}$$

$$= -\mathbb{E}_{p(\boldsymbol{\theta},\boldsymbol{x})}\left[\log q_{\boldsymbol{\phi}}(\boldsymbol{\theta}, \boldsymbol{x})\right] + C, \tag{A.2}$$

where C is a constant that does not depend on $\boldsymbol{\phi}$. Then, we can define our loss as the first term in Equation (A.2) and use a Monte Carlo approximation,

$$\mathcal{L}(\boldsymbol{\phi}) = -\mathbb{E}_{p(\boldsymbol{\theta},\boldsymbol{x})}\left[\log q_{\boldsymbol{\phi}}(\boldsymbol{\theta}, \boldsymbol{x})\right] \tag{A.3}$$

$$\approx -\frac{1}{N}\sum_{i=1}^{N} \log q_{\boldsymbol{\phi}}(\boldsymbol{\theta}^{(i)}, \boldsymbol{x}^{(i)}), \quad \boldsymbol{\theta}^{(i)} \sim p(\boldsymbol{\theta}),\ \boldsymbol{x}^{(i)} \sim f(\cdot \mid \boldsymbol{\theta}^{(i)}). \tag{A.4}$$

Minimization of the loss in Equation (A.4) amounts to maximum likelihood estimation, allowing $\boldsymbol{\phi}$ to be learned. The likelihood can be approximated as $\hat{f}(\boldsymbol{x}_o \mid \boldsymbol{\theta}) = q_{\boldsymbol{\phi}^\star}(\boldsymbol{\theta}, \boldsymbol{x}_o)/p(\boldsymbol{\theta})$, with $\boldsymbol{\phi}^\star$ denoting the learned parameters.

A.1 Direct Posterior Estimation

If the interest lies in the posterior $p(\boldsymbol{\theta} \mid \boldsymbol{x}_o)$, it is possible to estimate it directly using $q_{\boldsymbol{\phi}}(\boldsymbol{\theta}, \boldsymbol{x}) = q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x})q(\boldsymbol{x})$. In this case, the loss becomes

$$\mathcal{L}(\boldsymbol{\phi}) = -\mathbb{E}_{f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}\left[\log q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x})\right] \tag{A.5}$$

$$\approx -\frac{1}{N}\sum_{i=1}^{N} \log q_{\boldsymbol{\phi}}(\boldsymbol{\theta}^{(i)} \mid \boldsymbol{x}^{(i)}), \quad \boldsymbol{\theta}^{(i)} \sim p(\boldsymbol{\theta}),\ \boldsymbol{x}^{(i)} \sim f(\cdot \mid \boldsymbol{\theta}^{(i)}). \tag{A.6}$$

Note that the KL divergence is only 0 when $f(\boldsymbol{x} \mid \boldsymbol{\theta})p(\boldsymbol{\theta}) = q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x})q(\boldsymbol{x})$ which, in turn, implies that $q(\boldsymbol{x}) = \int_\Theta f(\boldsymbol{x} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$. It is also possible to use importance sampling, in which case the objective is

$$\mathcal{L}(\boldsymbol{\phi}) = -\mathbb{E}_{f(\boldsymbol{x}|\boldsymbol{\theta})q(\boldsymbol{\theta})}\left[\frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\log q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x})\right] \tag{A.7}$$

$$\approx -\frac{1}{N}\sum_{i=1}^{N} \frac{p(\boldsymbol{\theta}^{(i)})}{q(\boldsymbol{\theta}^{(i)})}\log q_{\boldsymbol{\phi}}(\boldsymbol{\theta}^{(i)} \mid \boldsymbol{x}^{(i)}), \quad \boldsymbol{\theta}^{(i)} \sim q(\boldsymbol{\theta}),\ \boldsymbol{x}^{(i)} \sim f(\cdot \mid \boldsymbol{\theta}^{(i)}). \tag{A.8}$$

A.1.1 Some Extensions[4]

By adopting an adaptive sampling strategy, the proposal can be set as $q^{(k)}(\boldsymbol{\theta}) = q^{(k-1)}(\boldsymbol{\theta} \mid \boldsymbol{x}_o)$, with $q^{(0)}(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$. Alternatively, one may sample from $q(\boldsymbol{\theta})$, but not correct the bias with importance weights during training. The reason is that these weights increase the variance of the $\boldsymbol{\phi}$ updates, leading to potential problems such as slow inference [8]. By not correcting the introduced bias, the learned (proposal) posterior density is

$$\tilde{q}_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{f(\boldsymbol{x} \mid \boldsymbol{\theta})q(\boldsymbol{\theta})}{\int_\Theta f(\boldsymbol{x} \mid \boldsymbol{\theta})q(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{A.9}$$

It is then possible to adjust this proposal posterior in a post-hoc fashion to obtain the correct posterior density. In this case, $q(\boldsymbol{x}) = \int_\Theta f(\boldsymbol{x} \mid \boldsymbol{\theta})q(\boldsymbol{\theta})d\boldsymbol{\theta}$, and we have

$$q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x}) = \tilde{q}_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x})\frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\frac{q(\boldsymbol{x})}{p(\boldsymbol{x})}. \tag{A.10}$$

[4]This section is based on a technical report I wrote in 2019 that describes the method by Greenberg et al. [8]. It may be worth revisiting in the future.

The problem with this approach is that it is only possible to obtain a closed-form expression under a restricted choice of density estimators and priors. In a more general setting, it is necessary to estimate the partition function,

$$Z_{\boldsymbol{\phi}}(\boldsymbol{x}) = \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} = \int_{\Theta} \tilde{q}_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x}) \frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}. \tag{A.11}$$

Following Greenberg et al. [8], the proposal posterior density during training can be given by

$$\tilde{q}_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x}) = q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x}) \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} \Big/ \tilde{Z}_{\boldsymbol{\phi}}(\boldsymbol{x}) \ , \quad \text{with } \tilde{Z}_{\boldsymbol{\phi}}(\boldsymbol{x}) = \int_{\Theta} q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x}) \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} d\boldsymbol{\theta}, \tag{A.12}$$

where we also have that $\tilde{Z}_{\boldsymbol{\phi}}(\boldsymbol{x}) = 1/Z_{\boldsymbol{\phi}}(\boldsymbol{x})$. Then, an empirical density can be used as proposal,

$$q(\boldsymbol{\theta} \mid \Theta^{(k)}) = \frac{1}{N} \sum_{i=1}^{N} \delta(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k,i)}), \tag{A.13}$$

where δ is the Dirac delta function and $\{\theta^{(k,i)}\}_{i=1}^{N} = \Theta^{(k)} \subset \Theta$. The partition function $\tilde{Z}_{\boldsymbol{\phi}}(\boldsymbol{x})$ at each round k can be computed as

$$\tilde{Z}_{\boldsymbol{\phi}}^{(k)}(\boldsymbol{x}) = \int_{\Theta} \frac{1}{N} \sum_{i=1}^{N} q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x}) \frac{\delta(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k,i)})}{p(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{A.14}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \frac{q_{\boldsymbol{\phi}}(\boldsymbol{\theta}^{(k,i)} \mid \boldsymbol{x})}{p(\boldsymbol{\theta}^{(k,i)})}, \qquad \theta^{(k,i)} \in \Theta^{(k)}, \tag{A.15}$$

and the proposal posterior at round k is given by

$$\tilde{q}_{\boldsymbol{\phi}}^{(k)}(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{x})}{p(\boldsymbol{\theta})} \Bigg/ \sum_{i=1}^{N} \frac{q_{\boldsymbol{\phi}}(\boldsymbol{\theta}^{(k,i)} \mid \boldsymbol{x})}{p(\boldsymbol{\theta}^{(k,i)})} \ , \qquad \boldsymbol{\theta}, \boldsymbol{\theta}^{(k,i)} \in \Theta^{(k)}. \tag{A.16}$$

While $\tilde{q}_{\boldsymbol{\phi}}^{(k)}(\boldsymbol{\theta} \mid \boldsymbol{x})$ follows a categorical distribution over $\Theta^{(k)}$, the bias-free posterior estimate $q_{\boldsymbol{\phi}}(\boldsymbol{\theta}^{(k,i)} \mid \boldsymbol{x})$ does not. In addition, $\forall \boldsymbol{\theta} \in \text{supp}(q(\boldsymbol{\theta} \mid \Theta^{(k)})) = \Theta^{(k)}$ and $\mathrm{D_{KL}} = 0$, we know that

$$\tilde{q}_{\boldsymbol{\phi}^\star}(\boldsymbol{\theta} \mid \boldsymbol{x}) q(\boldsymbol{x}) = f(\boldsymbol{x} \mid \boldsymbol{\theta}) q(\boldsymbol{\theta}) \iff \tilde{q}_{\boldsymbol{\phi}^\star}(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{f(\boldsymbol{x} \mid \boldsymbol{\theta}) q(\boldsymbol{\theta})}{q(\boldsymbol{x})} \tag{A.17}$$

$$\implies q_{\boldsymbol{\phi}^\star}(\boldsymbol{\theta} \mid \boldsymbol{x}) \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} \propto f(\boldsymbol{x} \mid \boldsymbol{\theta}) q(\boldsymbol{\theta}) \tag{A.18}$$

$$\implies q_{\boldsymbol{\phi}^\star}(\boldsymbol{\theta} \mid \boldsymbol{x}) \propto p(\boldsymbol{\theta} \mid \boldsymbol{x}), \tag{A.19}$$

which means we find an unnormalized estimate of the true posterior by plugging $\boldsymbol{x}_o$. As a result, we cannot directly evaluate the density, but can sample from $q_{\boldsymbol{\phi}^\star}(\boldsymbol{\theta} \mid \boldsymbol{x}_o)$. Finally, in order to draw the sets $\Theta^{(k)}$, the authors consider using a density that can be interpreted as a mixture of all previous (normalized) posterior density estimates,

$$\Theta^{(k)} = \{\boldsymbol{\theta}^{(k,i)}\}_{i=1}^{N} \overset{iid}{\sim} g^{(k)}(\boldsymbol{\theta} \mid \boldsymbol{x}_o) = \frac{1}{k} \sum_{j=0}^{k-1} q_n^{(j)}(\boldsymbol{\theta} \mid \boldsymbol{x}_o), \quad \text{with } q_n^{(0)}(\boldsymbol{\theta} \mid \boldsymbol{x}_o) = p(\boldsymbol{\theta}). \tag{A.20}$$

B BO for LFI

In [BOLFI 2, Section 2.6 1], the discrepancy between generated and observed data, $\Delta(\boldsymbol{\theta}) = d(\boldsymbol{s}_{\boldsymbol{\theta}}, \boldsymbol{s}_{\text{obs}})$, is assumed to be conditionally normally distributed, $\Delta|\mathcal{D} \sim \mathcal{GP}(m_\pi, C_\pi + \sigma_n^2\delta)$ with posterior mean m_π, posterior covariance C_π (of the latent process) and noise variance σ_n^2. For a uniform kernel with bandwidth ε, the likelihood surrogate is computed as

$$f_{\text{BOLFI}}(\boldsymbol{s}_{\text{obs}} \mid \boldsymbol{\theta}) \mid \mathcal{D} = \mathbb{P}(\Delta(\boldsymbol{\theta}) \leq \varepsilon \mid \mathcal{D}) \tag{B.1}$$

$$= F_{\mathcal{N}}\left(\frac{\varepsilon - m_\pi(\boldsymbol{\theta})}{\sqrt{C_\pi(\boldsymbol{\theta}, \boldsymbol{\theta}) + \sigma_n^2}}\right), \tag{B.2}$$

where $F_{\mathcal{N}}$ is the standard Normal cumulative distribution function.

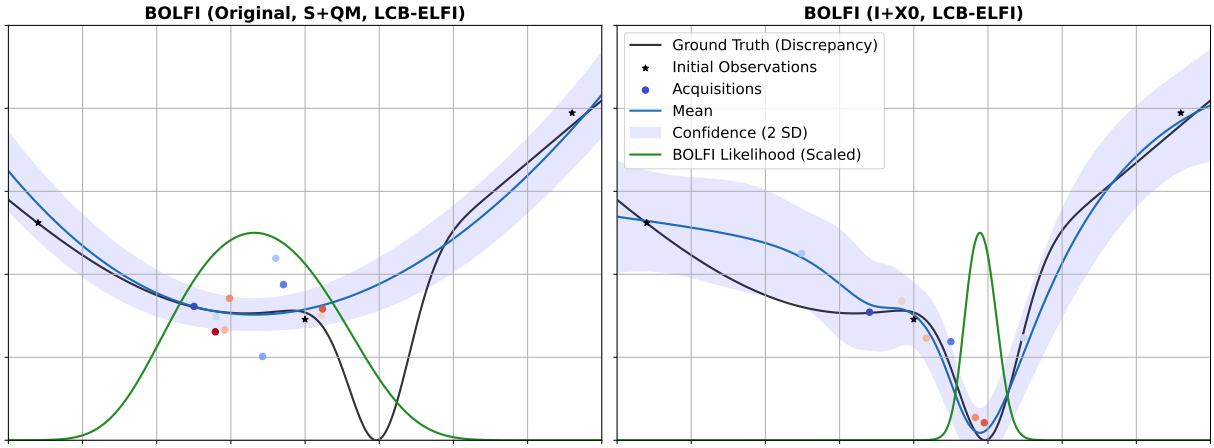

Figure B.1. In [BOLFI 2], the GP prior is characterized by an axis-aligned quadratic mean function and a stationary covariance function (S+QM, left). However, as in the example from [Figure 1 3], this prior leads again to an overconfident surrogate that is unable to capture the ground truth, specifically the region of small discrepancies. The region of high likelihood values does not match the region of small discrepancies, as given by the ground truth. This example reveals that the choice of GP prior is important for both optimization and (likelihood-free) inference. In fact, efficient exploration of modes for inference hinges on their efficient identification, i.e., optimization.

C Informative Cylindrical Covariance

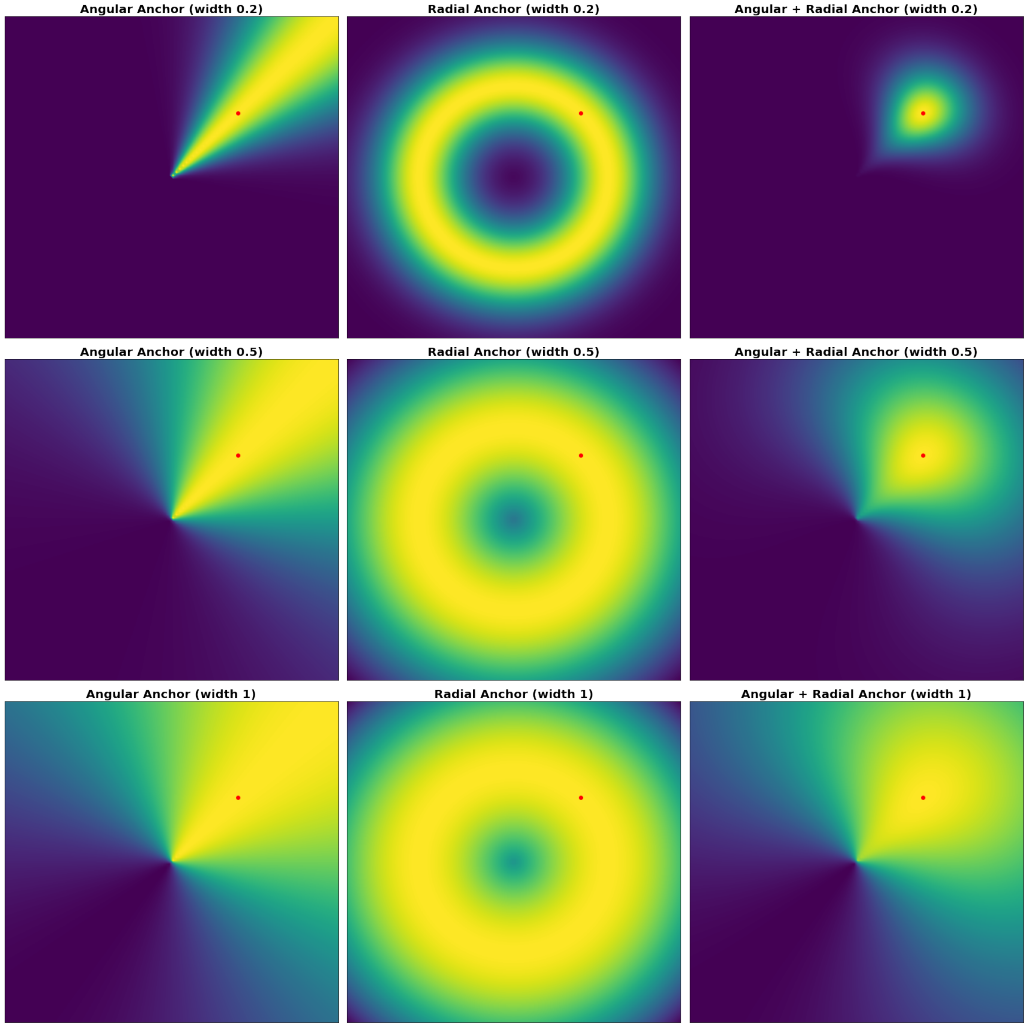

Figure C.1. Cylindrical covariance function [7] with spatially-varying prior/signal variance. The anchor is represented by the red dot.