Variational Inference over Non-differentiable Cardiac Simulators using Bayesian Optimization

Adam McCarthy[1], Blanca Rodriguez[1], and Ana Mincholé[1]

[1] Department of Computer Science, University of Oxford*

Abstract

Performing inference over simulators is generally intractable as their runtime means we cannot compute a marginal likelihood. We develop a likelihood-free inference method to infer parameters for a cardiac simulator, which replicates electrical flow through the heart to the body surface. We improve the fit of a state-of-the-art simulator to an electrocardiogram *(ECG)* recorded from a real patient.

1 Introduction

The heart contracts due to the propagation of electricity through the heart wall. This electricity then propagates through the torso and we can record an electrocardiogram *(ECG)*. Cardiac simulators replicate this propagation, outputting an ECG mirroring that recorded from a real patient (see Cardone-Noott et al., 2016, for a review). They are slow, require significant compute power - generally supercomputers, and are non-differentiable (we cannot compute its jacobian analytically). We must fit the parameters of these simulators efficiently so that they replicate the recording taken from real patients and provide clinicians with deep, interpretable understandings of the sources of risk. These insights also aid basic research and the development of new treatments.

This combination of slow run time and lack of differentiability make parameter inference and uncertainty propagation significant challenges and active areas of research. We cannot compute a likelihood due to the simulator run time and these problems frequently have many minima.

An approach to parameter inference is variational inference, which provides a parametric approximation of the posterior of a probabilistic model. Early work on variational inference was limited to forms in which gradient descent steps could be computed analytically (e.g. Jordan et al., 1999). Nowadays it is possible to take steps where the closed form is not known, for example using the *reparameterisation trick* (Kingma and Welling, 2014). This uses a Monte Carlo estimate of the likelihood to approximate the gradient of the optimisation function, but requires that the simulator is differentiable. Ranganath et al. (2014) drop this requirement with the *score function estimate* but gradient estimates are noisy. We focus on the optimiser instead of the gradients by using Bayesian optimisation, which being a global optimisation method makes converging to local minima less likely.

This problem of parameter inference over non-differentiable simulators applies to many fields in which simulators are used. We need to be able to fit weather simulators to predict hurricanes (Rappaport et al., 2009), simulators of the formation of the universe (Boylan-Kolchin, 2014), and simulators to predict the spread of wildland fires (Sullivan, 2009), to give a few examples. We also need to understand how changes in these parameters affect the output, and how certain the simulators are that events will happen.

Contribution We develop a method to infer parameters for a non-differentiable simulator using Variational Inference combined with Bayesian Optimisation, allowing us to fit a posterior distribution with a minimal number of trials and avoid local minima where possible. We show that we can use the inferred parameters to improve the fit of a state-of-the-art cardiac simulator.
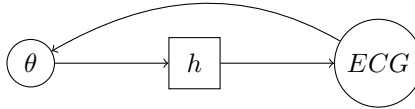
---

*`firstname.lastname@cs.ox.ac.uk`

Figure 1: The forward problem of the ECG involves taking the anatomy as a prior and defining parameters representing biological features, then propagating electricity through the heart, diffusing the charge through the torso and producing a simulated ECG. The inverse problem involves performing inference over the forward problem. θ are the model parameters, h is a heart simulator, and ECG is the output ECG trace.

2 Methodology

Simulator The state-of-the-art cardiac simulator implements the bidomain equations (see Trayanova, 2011, for a review). The bidomain-based simulator, whilst physiologically very accurate, takes several hours to run on a supercomputer. We developed a fast cardiac simulator which includes an anatomical heart model derived from MRI as a prior, simulates electrical propagation through the heart to the body surface, and outputs a 12-lead ECG (see Figure1 for a visualisation). Whilst this simulator is much faster, it is still too slow to compute a likelihood. It has a 17 dimensional parameter space consisting of locations at which electricity enters the heart wall and propagation velocities through the inner and outer heart wall, which have previously been measured in-vivo. For a biophysical meaning of these parameters, we refer the reader to Cardone-Noott et al. (2016).

The simulator we use will be presented in future work, so for the purposes of this paper can be a function $f(\theta)$ which has parameters $\theta \in \mathbb{R}^p$ and returns output $X \in \mathbb{R}^d$. It is a deterministic, black box function, with no simple closed form, and that it can be evaluated at any θ within the domain of interest. We have observations $ECG \in \mathbb{R}^d$ recorded from the patient and the aim is to compute a posterior $p(\theta|ECG)$. We also test the inferred parameters on the bidomain-based simulator, which can be defined similarly, to demonstrate that we have improved the fit of the state-of-the-art.

Variational inference Given that our simulator is non-differentiable we cannot use methods which use second-order information (e.g. Regier et al., 2017) or reduce the variance of the gradients (e.g. Miller et al., 2017; Roeder et al., 2017). We construct our variational objective following the *variational autoencoder* (Kingma and Welling, 2014) and optimise over the marginals of our Gaussian approximating distribution: μ and σ. Variational autoencoders define a parametric generative model with likelihood $p(X|\theta)$ for each dimension, prior $p(\theta)$ over the parameters and a model $q(\theta|X)$ which approximates the posterior. It can be shown that

$$\log p(x) \geq -KL\left(q(\theta|X), p(\theta)\right) + \mathbb{E}_{q(\theta|X)}\left[\log p(X|\theta)\right] \tag{1}$$

where the right hand term in Equation 1 is termed the *variational lower bound* or *evidence based lower bound objective (ELBO)*. In variational inference we maximise the ELBO by optimising over the marginal parameters of Q.

Our approach shares many similarities with the variational autoencoder, but the neural network is replaced by a cardiac simulator. We avoid more recent work on improving the posterior distributions (e.g. Ranganath et al., 2016; Liu and Wang, 2016; Huszár, 2017) and we will explain the reasons for this in Section 3.

Optimisation of the variational objective Our simulator is non-differentiable, so we cannot perform gradient-based optimisation without estimating gradients in some way. We could approximate the jacobian using finite differences, but this requires n evaluations of the model, where n is the number of dimensions over which we are optimising. There is also work on approximating gradients in variational inference, but they tend to be noisy. Instead, we perform global optimisation using Bayesian Optimisation to make our method applicable to all non-differentiable simulators.

It is common with local optimisation to grow the KL-divergence over time such that the log likelihood, which fits the mean of Q, dominates initially. Global optimisation routines have memory, however, so we fit the log likelihood first and then the variance via the KL-divergence as independent optimisations to prevent the KL-divergence dominating initially.

When performing Bayesian optimisation, we found a Square Exponential kernel was optimal, likely because the optimisation surface should be relatively smooth. We initially used Expected Improvement as our acquisition function, but found that the noisy samples meant Augmented EI (Huang et al., 2006) had superior performance. We sample our initial design matrix using Latin Hypercube Sampling and define our parameter space θ as the endocardial conductivities representing the Purkinje system, three conductivities for anisotropic conductivity in the myo- and epi-cardium, and the position of the stimulus locations, which includes all of the nodes in the inner heart wall *(the endocardium)*. We allow the conductivities to vary $\pm 50\%$. For a biophysical meaning of these parameters we refer the reader to Trayanova (2011). We set bounds on the parameters to represent the underlying physiological variability (Britton et al., 2013), and define inequality constraints for:

1. The order of three of the conductivities representing the direction of muscle fibres. We know a priori that they are ordered.

2. The stimulus positions are position invariant, i.e. $a, b, c \equiv c, b, a$, so we specify that the first dimension of the latent space representing the positions on a manifold through the heart (which we will discuss in the following paragraph) is ordered.

Dimensionality reduction So far we have assumed that our simulator is a black box. In this section, we move to a white box setting and use our knowledge of its internal dynamics to reduce dimensionality. We take magnetic resonance imaging (MRI) scans of the patient's heart and use this to derive an anatomical mesh. When we use this mesh within the simulator by converting it to a graph representation of electrical propagation, we are enforcing a strong prior on its dynamics. The heart itself lies on a lower dimensional manifold within Cartesian space due to the hollow ventricles, which makes optimisation over Cartesian coordinates discontinuous. We reduce the dimensionality of the graph vertices using an *isomap* (Tenenbaum et al., 2000). This works by computing geodesic distances between the vertices, and applying PCA to produce an embedding. We then use this embedding as a 2D manifold over which to optimise and will show in Section 3 that this significantly improves convergence time. Bayesian optimisation can struggle in large dimensions because the acquisition function, Expected Improvement, becomes flat near minima. We found that reducing the dimensionality improved the accuracy of the inferred parameters.

Isomaps are injective, meaning there is no inverse mapping from the 2D latent space back into Cartesian space. There are, however, a finite number of nodes in the mesh. We compute a lookup table between the nodes in Cartesian space and those in the latent space. We use a kd-tree to take a node in latent space and find its nearest neighbour in Cartesian space using the lookup table. This is necessary to transfer parameters inferred to the bidomain-based simulator, which uses a finer mesh.

3   Results and discussion

In this section we will report our results and then discuss how they improve fitting these simulators.

Dimensionality reduction Figure 2a shows the heart mesh reduced to 2D using an isomap over 16 neighbours for each point, which produces a mean reconstruction error of only 0.0016mm when projecting back into Cartesian space. Reducing dimensionality from three to two will have a minimal effect on the speed of fitting of a Gaussian Process, so the improvement in convergence speed and performance (see Figure 2b) is likely related to the increased smoothness of the optimisation surface created by following the manifold.

Parameter fitting Figure 3 shows an ECG recorded from a real patient and the associated bidomain simulation - the current state-of-the-art. Whilst the bidomain equations are physiologically accurate, their long runtime makes parameter fitting difficult. We take the parameters inferred on the simpler simulator and show that they improve the fit of the bidomain-based simulator which is considered the gold standard in cardiac modelling.

Structural modelling of the anatomy enforces a strong prior on the model which is essential in a small data regime such as medicine. Rather than treating the inverse problem as one of regression (e.g. Intini et al., 2005), we perform inference over a simulator representing the forward problem which provides this prior.

In-vivo measurements vary widely in the literature, due to noisy biological readings. Whilst maintaining uncertainty is important, it should be regarded with some scepticism due to the inaccuracy of the underlying readings on which these models are based. Choosing simple approximating distributions is important for two reasons: firstly they simplify the problem, allowing us to fit based on limited data and secondly they provide simpler, actionable decision support to clinicians. Projecting a Gaussian onto the heart surface is far more interpretable for a surgeon, for example, than a more complex distribution which might span multiple parts of the heart. Enforcing simplicity is an important consideration in the medical domain.

Finally, simulators of this form often present many local minima, so whereas variational inference usually relies on forms of stochastic gradient descent, we have used global optimisation which is more effective for providing accurate effective decision support.

Related work   Approximate Bayesian Computation (ABC) is a form of likelihood-free inference commonly used for these forms of problem. Tavaré et al. (1997) were the first to apply ABC, but it was named later by Beaumont et al. (2002). The simplest form of ABC, *rejection ABC*, is inefficient because θ is arbitrarily sampled from its prior distribution. leading to numerate rejections. *MCMC ABC* builds a markov chain through the prior, creating a proposal distribution from which to obtain new query points (see Marjoram et al., 2003, for a review). Whilst this is more efficient than rejection ABC, it would still be intractable in our case due to the number of samples required.

Papamakarios and Murray (2016) avoid rejecting samples by directly approximating the posterior as a mixture of Gaussians: a proposal prior $q(\theta)$. They draw samples from this prior and run the corresponding simulations, iteratively updating the proposal prior using a *mixture density network* (MDN) (Bishop, 1994), until it matches the target posterior.

Another approach is to approximate the likelihood function using a parametric model rather than avoiding its computation entirely, a *synthetic likelihood*. Wood (2010) used a single Gaussian distribution and Fan et al. (2012) used a mixture of Gaussians by learning multiple models based on repeated evaluations for fixed values of θ. Meeds and Welling (2014) used a Gaussian process to combine the likelihood approximations for each θ. Finally, we note a relationship with Louppe and Cranmer (2017), which has very similar motivations but relies on local optimisation using Variational Inference combined with a GAN.

In the modelling literature, van Dam et al. (2016) provide a method for non-invasive mapping of the heart from the 12-lead ECG. They enforce a strong prior by using a cardiac simulator as we do, but use local optimisation which may fall into local minima, whereas we use global optimisation. They find a point estimate of the PVC origin, whereas we provide a distribution over possible PVC origins. This is useful clinically, because it gives the clinician a measure of how accurate the model believes the localisation to be.

Lastly, Giffard-Roisin et al. (2017) approach the problem by learning a regression. They also provide uncertainty estimates and perform dimensionality reduction. In contrast to our method, they use 205 sensors across the torso (ECGI body surface potential maps Ramanathan et al., 2004), whereas we take standard 12-lead ECG as input. We can use less input data because the simulator enforces a strong prior. Their optimisation is local whereas our is global and they use Relevance Vector Machines, which are patented by Microsoft, so this may be a consideration when applying their work.

4   Conclusion

We have demonstrated an efficient, generic method for parameter inference over cardiac simulators by using Variational Inference and Bayesian Optimization. We have used a cardiac simulator to infer parameters to reproduce the ECG of a patient and used these parameters to improve the fit of a simulator based on the gold standard for cardiac modelling: the bidomain equations.

The ability to do efficient parameter inference makes it possible to personalise simulators and capture much more broadly the underlying electrophysiological variability. We plan to use this work to improve our cardiac simulators by capturing new forms of variability, which will represent a significant improvement in the accuracy of these simulators. We hope researchers from other fields will find similar benefits.

a) MSE cost curves for optimising over cartesian space and a manifold derived from an embedding of the anatomical mesh using an isomap

b) A 2D embedding of the inner heart wall, *the endocardium*, using an isomap over 16 neighbours
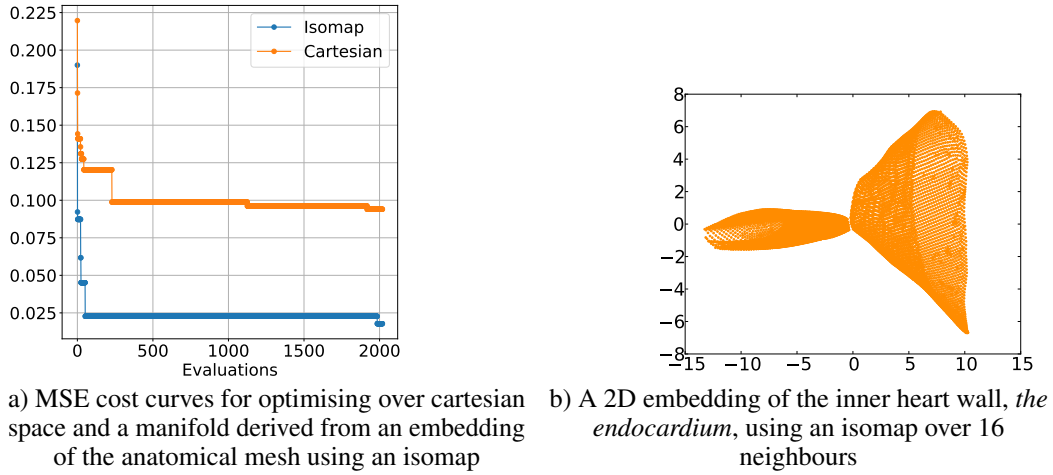
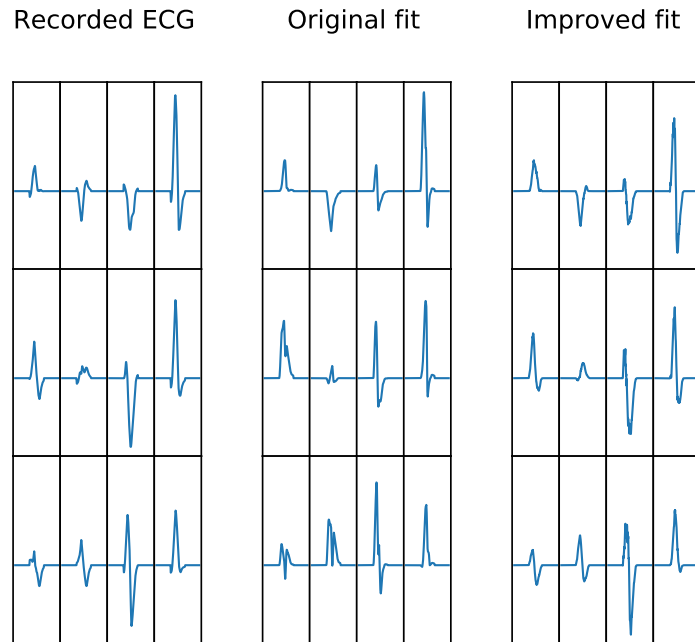Figure 2: Plots to demonstrate dimensionality reduction

Figure 3: Comparison between the recording from the real patient, the baseline simulation - the current state-of-the-art, and our improved fit. Our inference method allowed us to find parameters which better fit the dynamics of the real patient's heart. The recorded ECG is subject to electrical noise and noise from lead placement.

5

References

M. A. Beaumont, W. Zhang, and D. J. Balding. Approximate Bayesian Computation in Population Genetics. *Genetics*, 162(4):2025–2035, 2002. ISSN 00166731. doi: GeneticsDecember1,2002vol.162no.42025-2035.

C. M. Bishop. Mixture density networks. Technical report, Aston University, 1994.

M. Boylan-Kolchin. Cosmology: A virtual Universe. *Nature*, 509(7499):170–171, 2014. ISSN 0028-0836. doi: 10.1038/509170a.

O. J. Britton, A. Bueno-Orovio, K. Van Ammel, H. R. Lu, R. Towart, D. J. Gallacher, and B. Rodriguez. Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology. *Proceedings of the National Academy of Sciences*, 110(23):E2098–E2105, 2013. ISSN 0027-8424. doi: 10.1073/pnas.1304382110.

L. Cardone-Noott, A. Bueno-Orovio, A. Mincholé, N. Zemzemi, and B. Rodriguez. Human ventricular activation sequence and the simulation of the electrocardiographic QRS complex and its variability in healthy and intraventricular block conditions. *Europace*, 18(suppl_4):iv4–iv15, dec 2016. doi: 10.1093/europace/euw346.

Y. Fan, D. J. Nott, and S. A. Sisson. Approximate Bayesian Computation via Regression Density Estimation. *Statistics*, pages 1–17, 2012.

S. Giffard-Roisin, H. Delingette, T. Jackson, L. Fovargue, J. Lee, A. Rinaldi, N. Ayache, R. Razavi, and M. Sermesant. Sparse Bayesian Non-linear Regression for Multiple Onsets Estimation in Non-invasive Cardiac Electrophysiology. In *Functional Imaging and Modelling of the Heart*, 2017.

D. Huang, T. T. Allen, W. I. Notz, and N. Zeng. Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of Global Optimization*, 34(3):441–466, 2006. ISSN 09255001. doi: 10.1007/s10898-005-2454-3.

F. Huszár. Variational Inference using Implicit Distributions. *arXiv:1702.08235*, 2017.

A. Intini, R. N. Goldstein, P. Jia, C. Ramanathan, K. Ryu, B. Giannattasio, R. Gilkeson, B. S. Stambler, P. Brugada, W. G. Stevenson, Y. Rudy, and A. L. Waldo. Electrocardiographic imaging (ECGI), a novel diagnostic modality used for mapping of focal left ventricular tachycardia in a young athlete. *Heart Rhythm*, 2(11):1250–1252, 2005. ISSN 15475271. doi: 10.1016/j.hrthm.2005.08.019.

M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. Introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, 1999. ISSN 08856125. doi: 10.1023/A:1007665907178.

D. P. Kingma and M. Welling. Auto-Encoding Variational Bayes. In *International Conference on Learning Representations*, 2014.

Q. Liu and D. Wang. Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm. *arXiv:1608.04471*, aug 2016.

G. Louppe and K. Cranmer. Adversarial Variational Optimization of Non-Differentiable Simulators. *arXiv:1707.07113*, jul 2017.

P. Marjoram, J. Molitor, V. Plagnol, and S. Tavare. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, 100(26):15324–8, dec 2003. ISSN 0027-8424. doi: 10.1073/pnas.0306899100.

E. Meeds and M. Welling. GPS-ABC: Gaussian Process Surrogate Approximate Bayesian Computation. *arXiv preprint arXiv:1401.2838*, pages 1–17, 2014. ISSN <null>.

A. C. Miller, N. J. Foti, A. D'Amour, and R. P. Adams. Reducing Reparameterization Gradient Variance. may 2017.

G. Papamakarios and I. Murray. Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation. In *Advances in Neural Information Processing Systems*, pages 1028–1036, 2016.

C. Ramanathan, R. N. Ghanem, P. Jia, K. Ryu, and Y. Rudy. Noninvasive electrocardiographic imaging for cardiac electrophysiology and arrhythmia. *Nature medicine*, 10(4):422–428, 2004. ISSN 1078-8956. doi: 10.1038/nm1011.

R. Ranganath, S. Gerrish, and D. M. Blei. Black Box Variational Inference. *AISTATS*, 33:814–822, 2014. ISSN 15337928.

R. Ranganath, J. Altosaar, D. Tran, and D. M. Blei. Operator Variational Inference. In *Advances in Neural Information Processing*, oct 2016.

E. N. Rappaport, J. L. Franklin, L. a. Avila, S. R. Baig, J. L. Beven, E. S. Blake, C. a. Burr, J.-G. Jiing, C. a. Juckins, R. D. Knabb, C. W. Landsea, M. Mainelli, M. Mayfield, C. J. McAdie, R. J. Pasch, C. Sisko, S. R. Stewart, and A. N. Tribble. Advances and Challenges at the National Hurricane Center. *Weather and Forecasting*, 24(2):395–419, 2009. ISSN 0882-8156. doi: 10.1175/2008WAF2222128.1.

J. Regier, M. I. Jordan, and J. McAuliffe. Fast Black-box Variational Inference through Stochastic Trust-Region Optimization. In *Advances in Neural Information Processing Systems*, 2017.

G. Roeder, Y. Wu, and D. Duvenaud. Sticking the Landing: Simple, Lower-Variance Gradient Estimators for Variational Inference. *arXiv:1703.09194*, 2017.

A. L. Sullivan. Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models. *International Journal of Wildland Fire*, 18(4):349–368, 2009. ISSN 1049-8001. doi: 10.1071/WF06144.

S. Tavaré, D. J. Balding, R. C. Griffiths, and P. Donnelly. Inferring coalescence times from DNA sequence data, 1997. ISSN 00166731.

J. B. Tenenbaum, V. de Silva, and J. C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, dec 2000. doi: 10.1126/science.290.5500.2319.

N. A. Trayanova. Whole-heart modeling : Applications to cardiac electrophysiology and electromechanics, 2011. ISSN 00097330.

P. M. van Dam, N. G. Boyle, M. M. Laks, and R. Tung. Localization of premature ventricular contractions from the papillary muscles using the standard 12-lead electrocardiogram: a feasibility study using a novel cardiac isochrone positioning system. *Europace*, pii(5):euw099, 2016. doi: 10.1093/europace/euw347.

S. N. Wood. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(August): 1102–1104, 2010. ISSN 0028-0836. doi: 10.1038/nature09319.