Optimal Design under Heteroscedasticity for Gaussian Process Emulators with replicated observations

Alexis Boukouvalas, Dan Cornford, Milan Stehlík

boukouva@aston.ac.uk

NCRG, Aston University, Birmingham, UK

http://wiki.aston.ac.uk/AlexisBoukouvalas

September 2, 2011

Overview

- Introduction to Emulation.
- Heteroscedastic Gaussian Process models.
- Optimal Design Theory.
- Experiments.
    - Simulation Experiments on Synthetic data.
    - On the monotonicity of the Fisher design criterion.
    - Application to Prokaryotic Autoregulatory network systems biology simulator.
- Conclusions.

Motivation

Emulation

- A statistical model to the computer code simulator.
- Typically modelled as a Gaussian Process.
- The use of Gaussian Process Emulators to approximate deterministic computer simulators is well known.

Design

- Standard optimal design theory assumes independent homoscedastic errors.
- Extended to heteroscedastic independent errors - (Tack et al (2002)).
- Correlated homoscedastic errors, e.g. Zhu & Stein (2005). However GET, additivity of information matrix do not hold.

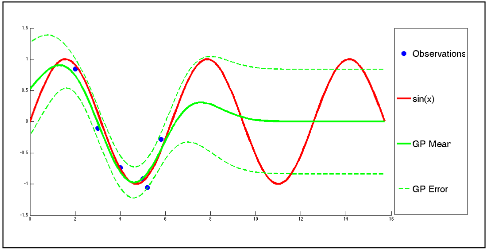
Contribution

Optimal design for heteroscedastic GPs

- A method for emulating a stochastic, or random output, simulator: heteroscedastic GPs.
- A method to optimally learn the parameters of such models based on Fisher information.
  - Extend Zhu & Stein (2005) to heteroscedastic case.
- Explicitly consider replicated observations in both stages.

Gaussian Processes

- A collection of random variables, any finite subset of which has a joint Gaussian distribution.
- Defined by a mean and a covariance function, the specification of which allows the incorporation of prior knowledge in the emulation analysis such as the smoothness and differentiability of the approximated function.



$$E[f(\mathbf{x})] = 0,\ Cov(\mathbf{x}_i, \mathbf{x}_j) = \sigma^2 e^{\frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{\lambda^2}} + \delta_{ij}\tau^2.$$

Emulation of deterministic simulators

Deterministic Simulator

$\mathbf{y} = \mathbf{f}(\mathbf{x})$ where $\mathbf{x}$ represents the inputs to the simulator, $\mathbf{y}$ represents the outputs of the simulator, or some summary of these, and $\mathbf{f}$ represents the mapping imposed by the simulator evaluation.

Gaussian Process Approximation

The probabilistic nature of the Gaussian Process emulator arises from the approximation of the simulator due to having a finite number of simulator runs.

Random Output Simulators

Stochastic Simulator

A mapping that produces random output given a fixed set of inputs.

Gaussian Process Approximation

In addition to having a finite number of simulator runs, uncertainty due to stochastic simulator. Our assumed observational model is:

$$y_i(x_i) = t_i(x_i) + \epsilon(x_i)$$

Approaches to Heteroscedastic Modelling

Coupled system of GPs

- Model heteroscedastic variance using a coupled system of GPs.
- MCMC inference, Goldberg et al (1998)
- Most Likely value, Kersting et al (2007),
- Variational, Lázaro and Titsias (2011).
- Extended to utilise repeated observations (replicates).

Joint Likelihood Model

- Coupled model too complex for design calculations.
- Use parametric deterministic variance model.
- Optimisation of the mean and variance model parameters proceeds jointly $\rightarrow$ tractable optimal design calculations.
- Efficient inference with replicated observations.

Should we always evaluate simulator at different inputs to gain maximum information?
Two clear advantages when evaluating the simulator repeatedly at same location:

- Simulator evaluations at input locations much closer than length scale can cause numeric difficulties.
- For moderate number of simulator evaluations, inference time can become impractical. Utilising replicate observations allows for much quicker inference.

Joint Likelihood Model

Crucial simplification: consideration of only deterministic variance models. The heteroscedastic GP prior is thus:

$$p(\mu|\theta,\beta) = N\left(0, K_\theta + \mathrm{diag}(\exp(f_{\sigma^2}(x,\beta))P^{-1}\right),$$

where $f_{\sigma^2}(x,\beta)$ is the deterministic variance model.
The joint log likelihood of the sample mean $\hat{\mu}$ and variance s^2 for N observations:

$$\log p(\hat{\mu}, s^2|\mathbf{X},\theta,\beta) = \left(\sum_{i=1}^{N} \log p(s_i^2|\beta, x_i, n_i)\right) + \log N\left(\hat{\mu}|0, K_\theta + RP^{-1}\right),$$

where K_θ the GP covariance function with parameters θ, R the diagonal matrix with elements $\exp\left(f_{\sigma^2}(x_i,\beta)\right)$.

Deterministic Variance Models

Fixed Basis

Fixed Basis variance model, the log variance function is modelled as a linear in parameters regression using a set of fixed basis functions:

$$f_{\sigma^2}(x, \beta) = \exp\left(H(x)^T \beta\right),$$

where $H(x)$ is the set of fixed basis functions with known parameters.

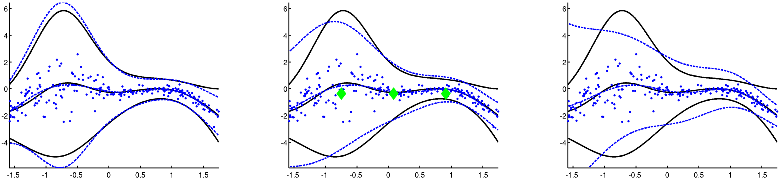
Latent Kernel

In high dimensional cases a non-parametric method could be considered using an additional 'variance kernel'.

$$f_{\sigma^2}(x, \beta) = k_\Sigma^T (K_\Sigma + \sigma_n^2)^{-1} \beta,$$

where $K_\Sigma = k(X_z, X_z)$ and $k_\Sigma = k(X_z, X_t)$ are the variance kernel functions, depending on parameters θ_Σ and σ_n^2 a nugget term.

Example of three variance models



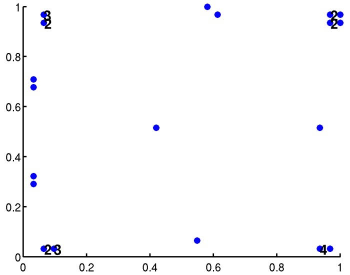
(a) Coupled Model (b) Latent Kernel (c) Quadratic Polynomial

Comparison of the Coupled, Latent Kernel and Quadratic polynomial variance models. Training set consists of using 200 design points with 4 replicate observations at each site. Dots are the empirical means of the samples. The black solid lines are the true function mean and standard deviation and the blue dashed lines the GP predictions.

Experimental Design For Correlated Processes

- Design to minimise kernel parameter uncertainty $\rightarrow$ D-optimality.
- Why not minimise predictive variance instead?
- Ans: All such methods either assume parameters known or use approximate values.
- D-optimal design used as preliminary design or as part of hybrid criterion (e.g. see Zhu and Stein (2006)).

Optimal Design for Heteroscedastic Gaussian Process Regression with replicated observations



- Design to minimise parameter uncertainty $\rightarrow$ D-optimality
- Minimise Fisher information of design ξ:

$$\mathcal{F}(\xi) = E\left[\frac{\partial^2}{\partial \theta^2} \ln L(X|\theta, \xi)\right]$$

- Analytic solution derived for GP with parametric variance model.

Joint Likelihood: Fisher Information

The FIM for a design ξ is defined as:

$$\mathcal{F}(\xi) = \int \left(\frac{\partial^2}{\partial \theta^2} \ln \left[L(X|\theta, \xi) \right] \right) L(X|\theta, \xi) \, \mathrm{d}X,$$

where $L(X|\theta, \xi)$ is the likelihood function.

For Joint Likelihood model FIM can be calculated analytically:

$$\boxed{\mathcal{F}_{ij} = \sum_{m=1}^{M} F_{ij}^{s} + F_{ij}^{N},} \tag{1}$$

where

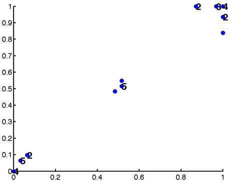
- M the number of design points.
- $F_{ij}^{s} = \frac{n_i - 1}{2} \frac{\partial f}{\partial \theta_i} \frac{\partial f}{\partial \theta_j}$ where n_i the number of replicate observations at design point i and $\frac{\partial f}{\partial \theta_j}$ the derivative of the variance model $f(\theta)$ with respect to parameter θ_j.
- $F_{ij}^{N} = \frac{1}{2} \mathrm{tr}(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j})$.

Experimental Setup

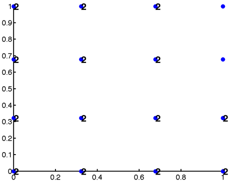
Synthetic Experiment

- Sample from GP with known parameters.
- GP Maximum Likelihood Inference with same covariance using different designs.
- Compute parameter errors.
- 500 realisations.

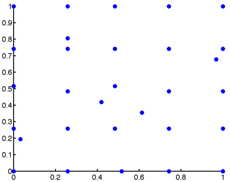
Local Design: Latent Kernel Variance model



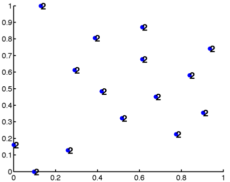
(a) Greedy



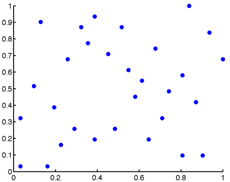
(b) Replicate Grid



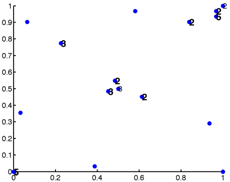
(c) Grid



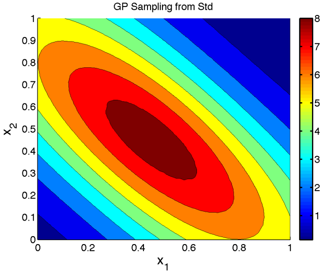
(d) Latin Hypercube Rep



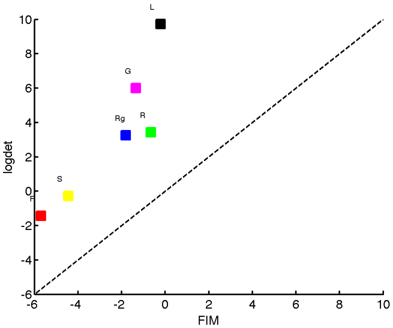
(e) Latin Hypercube



(f) Sim Annealing



Variance surface.

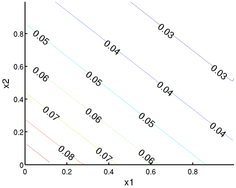


FIM (x axis) and LDM (y axis).

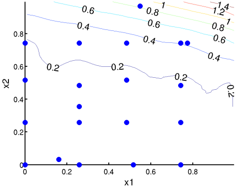
Variance model parameter errors

Greedy	Replicate Grid	Grid	Latin Rep	Latin	Sim Ann
0.22	0.46	0.66	0.49	0.82	0.25

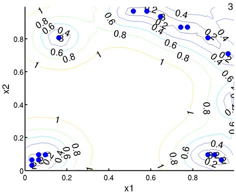
Synthetic Example: Log Linear variance model Individual example



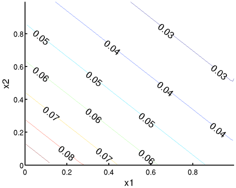
(a) True Std



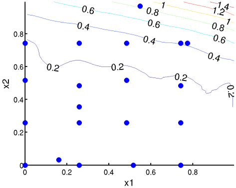
(b) Std for Grid



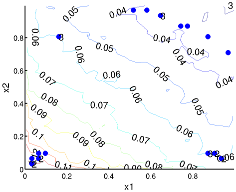
(c) Std for Simulated Ann



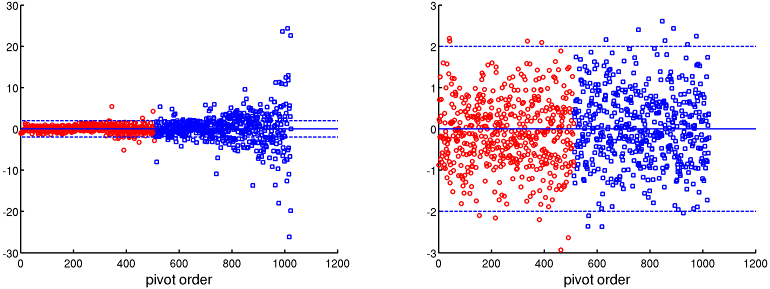
(d) Reference Set: True Std



(e) Reference Set: Std for Grid



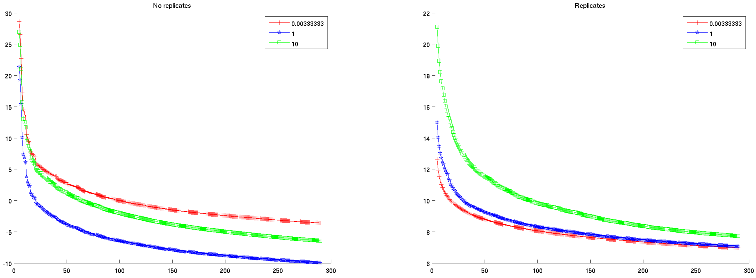
(f) Reference Set: Std for Simulated Ann



(a) Grid

(b) Simulated Annealing

Uncorrelated Errors vs Pivot Order for the Grid and Simulated Annealing designs (Bastos and O'Hagan (2006)).

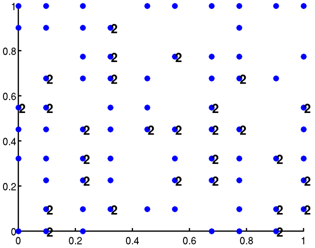


(a) No replicates

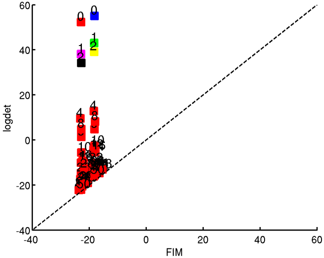
(b) 4 Point Replicate design

Monotonicity of Fisher information with regards to design size. X axis is design size, Y axis is Fisher information. Results shown for three different nugget values.

Monotonicity of Fisher to LDM



Example Grid design.



FIM (x axis) and LDM (y axis)

Prokaryotic Autoregulatory network

- Minimal in terms of biological detail included but contains many of the interesting features of an auto-regulatory feedback network (see Wilkinson (2006)).
- Restrict our attention to k_6 and k_7 reaction parameters.

Reactant species: gene g, protein P and its dimer P_2, and the mRNA molecule. Eight reactions:

$g + P_2 \xrightarrow{k_1} g.P_2$		Repression
$g.P_2 \xrightarrow{k_2} g + P_2$		Reverse Repression
$g \xrightarrow{k_3} g + r$		Transcription
$r \xrightarrow{k_4} r + P$		Translation
$2P \xrightarrow{k_5} P_2$		Dimerisation
$P_2 \xrightarrow{k_6} 2P$		Dissociation
$r \xrightarrow{k_7} \emptyset$		mRNA degradation
$P \xrightarrow{k_8} \emptyset$		Protein degradation

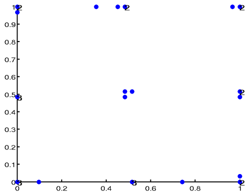
Experiment Setup

Parameter Ranges

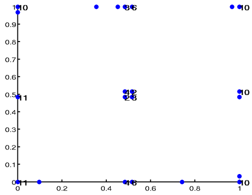
- Domain region $k_6 \in [0, 7]$ and $k_7 \in [0.05, 0.4]$.
- Other parameters are set to the nominal values $\{1, 10, 0.01, 10, 1, k_6, k_7, 0.01\}$.
- Initial number of molecules were set to $\{g.P_2, g, r, P, P_2\} = \{100, 0, 0, 0, 0\}$.
- The response we have selected to emulate is the number of bounded molecules $g.P2$ at time step $T = 18$.

Emulator and Design

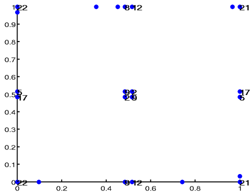
- Zero mean Matérn covariance with fixed differentiability 5/2 GP prior.
- Variance model: nine point latent kernel arranged on a grid.
- Local Design assuming short length scale process ($\lambda = 0.6$) and high noise to signal ratio $z_{1,\ldots,9} = 3.5, \sigma_p^2 = 0.36$



(a) 30 pts



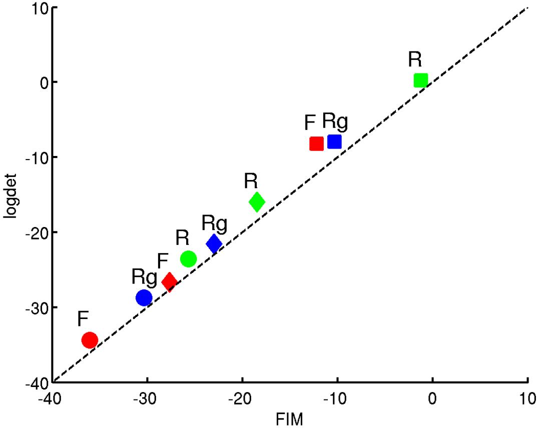
(b) 100 pts



(c) 200 pts

Points correspond to the locations of the latent points of the latent kernel variance model.

Fisher Information (x) compared to the empirical log determinant of the maximum likelihood covariance (y).



30 point design (square), 100 point design (diamond), 200 point design (circle).

Conclusions

Model

- Heteroscedastic Emulation: Variety of model complexity.
- Simple heteroscedastic GP model allows for optimal design calculation.
- In combination with a discrepancy model and real-world observations, this method could facilitate the efficient statistical calibration of stochastic simulators.

Experimental Design

- Fisher Designs minimise kernel parameter estimation variance.
- Utilising Replicated observations beneficial for stochastic emulation.
- Mahalanobis based validation reflects lower parameter estimation error.

Open Questions

Model

Model flexibility. Modelling first two moments may be too restrictive in some applications.

- Non-parametric density methods such as Indicator Kriging or Copulas.

Experimental Design

- Use of replicated observations improves FIM approximation. Because of nugget parameters better identified?
- Effect of Fisher Designs on Parameter posterior (Bayesian Inference).
- Discrete optimisation : Curse of Dimensionality.
- Incorporate trend parameter estimation.
- Extensions to adaptive context, and other models (e.g. quantile regression).

References

Gaussian Processes

- K. Kersting, C. Plagemann, P. Pfaff, and W. Burgard. "Most likely heteroscedastic gaussian process regression". In Proc. 24th International Conf. on Machine Learning, 2007.
- Bastos, L. S. and O'Hagan, A. "Diagnostics for Gaussian process emulators". Technometrics, 2008.
- M. Lázaro-Gredilla and M. K. Titsias. Variational Heteroscedastic Gaussian Process Regression. International Conference on Machine Learning (ICML), 2011.
- Paul W. Goldberg and Christopher K. I. Williams and Christopher M. Bishop. "Regression with Input-dependent Noise: A Gaussian Process Treatment". Advances in Neural Information Processing Systems. The MIT Press, 1998.

Design

- Boukouvalas, A., Cornford, D., Singer, A., Managing Uncertainty in Complex Stochastic Models: Design and Emulation of a Rabies Model. In St. Petersburg Workshop on Simulation (2009).
- Z. Zhu and M. L. Stein. Spatial sampling design for parameter estimation of the covariance function. Journal of Statistical Planning and Inference, 134 (2): 583-603, 2005.
- Z. Zhu and M. L. Stein. Spatial sampling de- sign for prediction with estimated parameters. Journal of Agricultural, Biological, and Environmental Statistics, 11(1):24-44, March 2006.
- L. Tack, P. Goos, and M. Vandebroek (2002). Efficient bayesian designs under heteroscedasticity. Journal of Statistical Planning and Inference, 104(2):469-483.
- Müller W.G. and Stehlík M. (2009). Issues in the Optimal Design of Computer Simulation Experiments, Applied Stochastic Models in Business and Industry, 25:163-177.

Systems Biology

- D. J. Wilkinson. Stochastic Modelling for Systems Biology. Chapman & Hall/CRC, 1 edition, 2006.