UNIVERSITY OF CALIFORNIA
SANTA CRUZ

# IDENTIFYING CONVERGENCE IN GAUSSIAN PROCESS SURROGATE MODEL OPTIMIZATION, VIA STATISTICAL PROCESS CONTROL

A document submitted in partial satisfaction of the
requirements for the degree of

MASTER OF SCIENCE

in

STATISTICS AND APPLIED MATHMATICS

by

**Nicholas R. Grunloh**

July 2014

Approved by:

---

Professor Herbie Lee

---

Associate Professor John Musacchio

# Table of Contents


**1 Introduction** **1**

1.1 General . . . 1

1.2 Gaussian Process Models . . . 2

1.3 Optimization . . . 2

1.4 Convergence Criteria . . . 2

1.5 Statistical Process Control . . . 2

1.5.1 Shewhart's $\bar{x}$ Chart . . . 2

1.5.2 Exponentially Weighted Moving Average Chart . . . 3

**2 Identifying Convergence** **3**

2.1 The Control Window . . . 3

2.2 $\bar{x}$ Chart . . . 4

2.3 Model-Based Transformation . . . 4

2.4 EWMA Chart . . . 4

**3 Test Functions** **4**

3.1 Rosenbrock . . . 5

3.2 Rastringin . . . 5

3.3 Easom . . . 6

3.4 Real Data . . . 6

**4 Discussion** **6**

**5 Code Appendix** **6**

**Abstract**

```
  ----------
< Abstract Cow >
 -------------
        \   ^__^
         \  (**)\_______
            (__)\       )\/\
             U  ||----w |
                ||     ||
```


# 1 Introduction

✓ Identify convergence problem

✓ Define properties of convergence

   – Shewhart Control Defininition: A phenomenon is said to be in control "when, through the use of past experience, we can predict, at least within limits, how the phenomenon may be expected to vary in the future."

- A taste of my stuff

   – GP surrogate, with $\mathbb{E}[\,\mathrm{I}(\boldsymbol{x})\,]$ search.

   – track $\mathbb{E}[\,\mathrm{I}(\boldsymbol{x})\,]$ via SPC to identify control.

✓ Explicate road-map

## 1.1 General

- Optimization: Background and our Philosophy

   ✓ gradient free [5]

   ✓ GA [5], Simulated Annealing [5], Pattern Search [5], Trust Regions [5]

   ✓ benefits of a model based approach (ie. uncertainty measures)(i.e. General/No GP)

   ✓ uncertainty measures as convergence criteria

- Statistical process control for monitoring convergence

   – General [4] (ie. no EWMA)

   – tease EWMA

## 1.2 Gaussian Process Models

- breif explaination of simple case.
- for more explaination see [1] mountainerring? For demonstration purposes I believe that it can be helpfull to imagine the objective function as an unknown physical landscape that we are going to explore. This analogy can be considered fairly litterally at times considering the conception of Gaussian processes, initially as the kriging [].
- breif explaination of flexibility through tree partioning
- for more explaination see [1]
- implimentation via tgp; MCMC sampling; [2]; [3];
- predictive locations.

## 1.3 Optimization

✓ So and So introduced this metric. $\mathbb{E}[\,\mathrm{I}(\boldsymbol{x})\,]$ [9], [3]

✓ What is $\mathbb{E}[\,\mathrm{I}(\boldsymbol{x})\,]$?

✓ which $\mathbb{E}[\,\mathrm{I}^g(\boldsymbol{x})\,]$, ie. which $g$?[9]

✓ maximum $\mathbb{E}[\,\mathrm{I}(\boldsymbol{x})\,]$ (i.e. the mean at the predictive location that achieves the maximum mean of the samples at that location)

✓ Optimization Procedure [3]

  - code appendix, using tgp

✓ bounded at 0, stochastic decreasing function.

? advantages of model based approach for convergence sake

## 1.4 Convergence Criteria

✓ My intuition on convergence as the potential for new optima goes to zero.

✓ $\mathbb{E}[\,\mathrm{I}(\boldsymbol{x})\,]$ Behavior for convergence

## 1.5 Statistical Process Control

### 1.5.1 Shewhart's $\bar{x}$ Chart

- the notion of control (draw similarities to convergence). [4]
- how the typical charts work
- philosophy.

  - establish control (herbies book)
  - control → out-of-control

- stumbling blocks of convergence for me.
  - out-of-control → control
  - the notion of a sliding average (i.e. convergence)
  - normal assumptions are very strong for an application that strongly desired robustness in varied applications

### 1.5.2 Exponentially Weighted Moving Average Chart

- EWMA philosophy (Robustness) [7].
- How it works (derivation cite).
- look at the statistics and bounds
- Tracking slight changes (general scale and behavior of $\mathbb{E}[\ \mathrm{I}(\boldsymbol{x})\ ]$)
- weight recent data more heavily to handle the sliding average (also mention later about the window; maybe set-up here)

# 2 Identifying Convergence

- tie this stuff together and motivate the coolness factor.
- use optimization procedure outlined [3] also above and in appendix.
- recall that the maximum $\mathbb{E}[\ \mathrm{I}(\boldsymbol{x})\ ]$ each iteration is the mean at the predictive location that achieves the highest mean value.
- SPC is based on normality assumptions of the underlying sampling distribution.
- a thesis statement for the research that I did: SPC, EWMA, empirical predictive MCMC control limits, Log-Normal → model based limits.

## 2.1 The Control Window

- convergence formulated in the context of statistical process control has unique challenges since almost by definition max $\mathbb{E}[\ \mathrm{I}(\boldsymbol{x})\ ]$ starts in an out-of-control state then moves into a state of control the optimization routine approaches a state of convergence
- typical SPC goes through an initialization process, in which, initially out-of-control observation are investigated and systematically accounted for to establish an initial state of control.

- introduce the window to automate this process and thus set the current control state at a window of the most recently observed values.
- window of size $w$, tuning parameter, thus partitioning the observations into points in the control window(i.e. control training set) and points outside of the control window(i.e. control test set)
- convergence rules: out-of-control in control test set and in-control in control training set.
- choosing $w$ for difficulty of problem.

### 2.2 $\bar{x}$ Chart

- basic shewhart chart
- issues with robustness since you often see linear trends toward the lower control limit as $\max \mathbb{E}\left[\ \mathrm{I}(\boldsymbol{x})\ \right]$ seems to converge in probability to 0 from the positive direction.

### 2.3 Model-Based Transformation

- Transformation
- $\bar{x}$ Chart
- equally weighted observations leads to false positive in identifying convergence since initial $\mathbb{E}\left[\ \mathrm{I}(\boldsymbol{x})\ \right]$ may be very large.

### 2.4 EWMA Chart

- 

Example pictures above?, or keep most of the figures that I made in the below sections?, start Rosenbrok example early?:

- $\max \mathbb{E}\left[\ \mathrm{I}(\boldsymbol{x})\ \right]$, General Behavior (3 stages, not-converged(inital exploration), converging(pre-convergence) and converged(converged))
- 

## 3 Test Functions

use each example as an excuse to look at different things?

### 3.1 Rosenbrock

- write the general function down, focus on the 2-D case, code appendix (cite?)
- get a good looking window
- plot what function looks like in this window, perspective and heat plot,
- gaussian process fit perspective and heat plot (movie?), thumbnail:first converged picture
- Simple $\mathbb{E}[\,\mathrm{I}(\boldsymbol{x})\,]$ Pictures
  - $max\mathbb{E}[\,\mathrm{I}(\boldsymbol{x})\,]$/best Z (tell three stages convergence story)
  - hist of $max\mathbb{E}[\,\mathrm{I}(\boldsymbol{x})\,]$ samples
  - Q-Q plot?
  - $\bar{x}$ Chart
- Transformed Pictures
  - $max\mathbb{E}[\,\mathrm{I}(\boldsymbol{x})\,]$/best Z
  - hist of $max\mathbb{E}[\,\mathrm{I}(\boldsymbol{x})\,]$ samples
  - Q-Q plot?
- discussion of results

### 3.2 Rastringin

- write down function cite)
- several mode window
- plot what function looks like in this window, perspective and heat plot,
- gaussian process fit perspective and heat plot (movie?), thumbnail:first converged picture
- Transformed Picture $max\mathbb{E}[\,\mathrm{I}(\boldsymbol{x})\,]$/best Z
- discussion of results

### 3.3 Easom

- write down function cite
- Get reasonably flat window
- plot what function looks like in this window, perspective and heat plot,
- gaussian process fit perspective and heat plot (movie?), thumbnail:first converged picture
- Transformed Picture $max\mathbb{E}\left[\ \mathrm{I}(\boldsymbol{x})\ \right]$/best Z
- discussion of results

### 3.4 Real Data

- explore this data cite
- get good looking picture of objective function
- gaussian process fit perspective and heat plot (movie?), thumbnail:first converged picture
- Transformed Picture $max\mathbb{E}\left[\ \mathrm{I}(\boldsymbol{x})\ \right]$/best Z
- discussion of results

## 4 Discussion

- argument for a convergence criteria based on above results
- Robustness of EWMA [7]
- further research partitioned model idea.

## 5 Code Appendix

- tgp optimization [3]
- qcc EWMA SPC [6]
- example implementation (cite data)

# References

[1] R. B. Gramacy, & H. K. Lee. (2008). *Bayesian treed Gaussian process models with an application to computer modeling*. Journal of the American Statistical Association, 103, 1119-1130.

[2] R. B. Gramacy, (2007). *tgp: An R Package for Bayesian Nonstationary, Semiparametric Nonlinear Regression and Design by Treed Gaussian Process Models*. Journal of Statistical Software, 19(9), 1-46.

[3] R. B. Gramacy, & M. Taddy (2010). *Categorical Inputs, Sensitivity Analysis, Optimization and Importance Tempering with tgp Version 2, an R Package for Treed Gaussian Process Models*. Journal of Statistical Software, 33(6), 1-48.

[4] W. Shewhart (1931). *Economic Control of Quality of Manufactured Product*. New York: D. Van Nostrand Company, Inc.

[5] A. R. Conn, K. Scheinberg, & L. Vicente (2009). *Introduction to derivative-free optimization*. Philadelphia: Society for Industrial and Applied Mathematics/Mathematical Programming Society.

[6] L. Scrucca (2004). *qcc: an R package for quality control charting and statistical process control*. R News 4/1, 11-17.

[7] G. E. Box, A. Luceno, & M. Paniagua-Quiñones, (1997). *Statistical control by monitoring and feedback adjustment*. New York: Wiley.

[8] get expected improvment author

[9] M. Schonlau, D. R. Jones, & W. J. Welch (1998). *Global versus local search in constrained optimization of computer models*. In *New developments and applications in experimental design*, number 34 in IMS Lecture Notes. Monograph Series, 11-25.

[10] T. J. Santner, B. J. Williams, & W.I. Notz (2003). *The Design and analysis of computer experiments*. New York: Springer.