

# Gaussian processes for Computational Biology: Theory

Peter Tonner

Computational Biology and Bioinformatics  
Duke University

October 13, 2017

# Hyperparameters change model interpretation

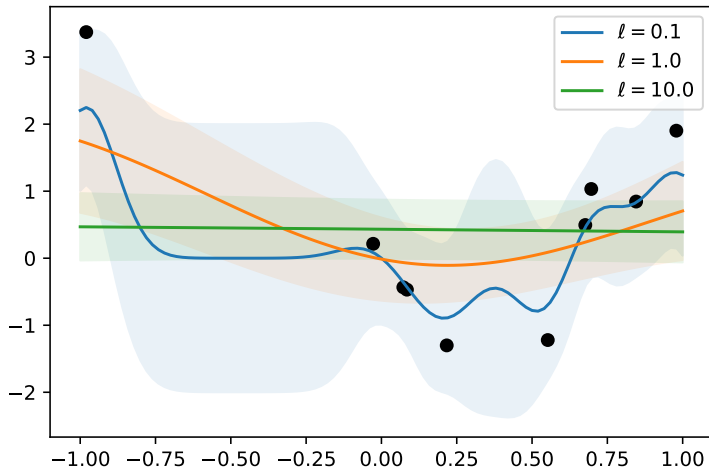
## Introduction

### Stan

Introduction  
Hierarchical Models  
Basic GP model  
Functional Linear  
Model

### Single-cell RNA-seq

Reid *et al.*



# How do we determine hyperparameters?

## 1. Maximize likelihood

- ▶ gradient optimization, point estimate
- ▶ **pros:** many existing implementations (e.g. *GPy*), often gets answer the fastest
- ▶ **cons:** requires “conjugacy” b/w  $f$  and  $y$  for analytical sol'n

## 2. Model hyperparameter uncertainty

- ▶ Posterior sampling (MCMC)
- ▶ **pros:** Doesn't care about form of  $p(y)$ , can be placed in larger model easily
- ▶ **cons:** sampling time (can be long)

## 2. Full Bayesian model

### Introduction

#### Stan

Introduction  
Hierarchical Models  
Basic GP model  
Functional Linear  
Model

#### Single-cell

#### RNA-seq

Reid *et al.*

$$\overbrace{p(\theta|y)}^{\text{posterior}} = \frac{\overbrace{p(y|\theta)}^{\text{likelihood}} \overbrace{p(\theta)}^{\text{prior}}}{\underbrace{p(y)}_{\text{evidence}}}$$

## 2. Full Bayesian model

$$\overbrace{p(\theta|y)}^{\text{posterior}} = \frac{\overbrace{p(y|\theta)}^{\text{likelihood}} \overbrace{p(\theta)}^{\text{prior}}}{\underbrace{p(y)}_{\text{evidence}}}$$

$$p(y) = \int p(y|\theta)p(\theta)d\theta$$

## 2. Full Bayesian model

$$\overbrace{p(\theta|y)}^{\text{posterior}} = \frac{\overbrace{p(y|\theta)}^{\text{likelihood}} \overbrace{p(\theta)}^{\text{prior}}}{\underbrace{p(y)}_{\text{evidence}}}$$

$$p(y) = \int p(y|\theta)p(\theta)d\theta$$

- ▶ In general, computing  $p(y)$  is quite difficult
- ▶ MCMC sampling:
  - ▶ Gibbs sampling
  - ▶ *Stan*

# Markov Chain Monte Carlo: the (extremely) short version

---

$$\overbrace{p(\theta|y)}^{\text{posterior}} \propto \overbrace{p(y|\theta)}^{\text{likelihood}} \overbrace{p(\theta)}^{\text{prior}}$$

## Introduction

### Stan

- Introduction
- Hierarchical Models
- Basic GP model
- Functional Linear Model

### Single-cell

#### RNA-seq

Reid *et al.*

# Markov Chain Monte Carlo: the (extremely) short version

---

Gaussian processes  
for Computational  
Biology: Theory

Peter Tonner

Introduction

Stan

Introduction  
Hierarchical Models  
Basic GP model  
Functional Linear  
Model

Single-cell  
RNA-seq

Reid *et al.*

$$\overbrace{p(\theta|y)}^{\text{posterior}} \propto \overbrace{p(y|\theta)}^{\text{likelihood}} \overbrace{p(\theta)}^{\text{prior}}$$

Generate samples:

$$\theta = \{\tilde{\theta} \mid \tilde{\theta} \text{ drawn from } p(\theta|y)\}$$



# Markov Chain Monte Carlo: the (extremely) short version

$$\overbrace{p(\theta|y)}^{\text{posterior}} \propto \overbrace{p(y|\theta)}^{\text{likelihood}} \overbrace{p(\theta)}^{\text{prior}}$$

Generate samples:

$$\theta = \{\tilde{\theta} \mid \tilde{\theta} \text{ drawn from } p(\theta|y)\}$$

- Quantities of interest:

$$E[\theta] \approx \frac{1}{n} \sum_{i=1}^n \tilde{\theta}_i$$

- In general (for any  $f$ ):

$$E[f(\theta)] \approx \frac{1}{n} \sum_{i=1}^n f(\tilde{\theta}_i)$$

- There are many (MCMC) algorithms to generate  $\theta$ .

## Introduction

## Stan

### Introduction

Hierarchical Models

Basic GP model

Functional Linear Model

## Single-cell RNA-seq

Reid *et al.*

## Introduction

## Stan

### Introduction

Hierarchical Models

Basic GP model

Functional Linear  
Model

## Single-cell

## RNA-seq

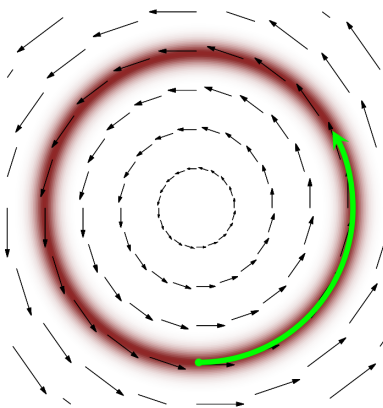
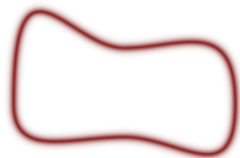
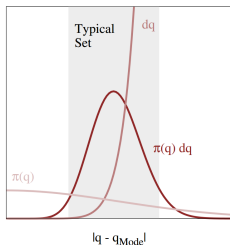
Reid *et al.*

“a state-of-the-art platform for statistical modeling and high-performance statistical computation”

<http://mc-stan.org/>

- ▶ Hamiltonian monte-carlo (HMC)
- ▶ Has other inference methods as well: maximum likelihood, variational inference
- ▶ Probabilistic programming language
- ▶ Quite fast (relatively speaking and when used properly)

$$E_{\pi}\{f\} = \int \pi(q) dq f(q)$$



# Sections of a Stan program

---

- ▶ **functions**: user defined functions
- ▶ **data**: input to the model
- ▶ **transformed data**: generated from data
- ▶ **parameters**: quantities of interest
- ▶ **transformed parameters**: similar to parameters, derived from other parameters, these can still be used in the model block
- ▶ **model**: relationship b/w data and parameters
- ▶ **generated quantities**: similar to transformed parameters, but not used in model. e.g. posterior predictive checking

## Introduction

## Stan

Introduction

**Hierarchical Models**

Basic GP model

Functional Linear Model

## Single-cell RNA-seq

Reid *et al.*

## Introduction

### Stan

Introduction

**Hierarchical Models**

Basic GP model

Functional Linear  
Model

### Single-cell

### RNA-seq

Reid *et al.*

# Note on hierarchical models

Gaussian processes  
for Computational  
Biology: Theory

Peter Tonner

Introduction

Stan

Introduction

**Hierarchical Models**

Basic GP model

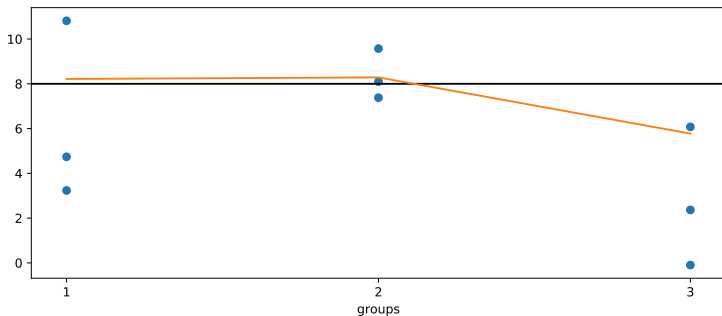
Functional Linear  
Model

Single-cell

RNA-seq

Reid *et al.*

$$y_i \sim N(\theta_i, \sigma_y^2) \quad \text{and} \quad \theta_i \sim N(\mu, \tau^2)$$



# Hierarchical models confound MCMC

Gaussian processes  
for Computational  
Biology: Theory

Peter Tonner

Introduction

Stan

Introduction

**Hierarchical Models**

Basic GP model

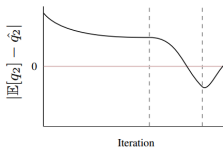
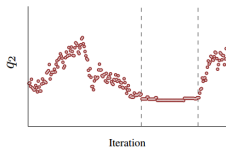
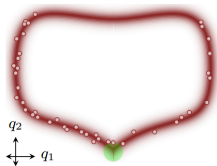
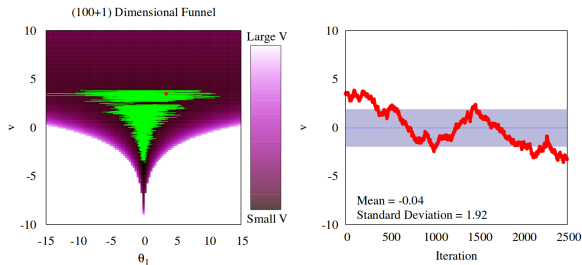
Functional Linear  
Model

Single-cell

RNA-seq

Reid *et al.*

$$y_i \sim N(\theta_i, \sigma_y^2) \quad \text{and} \quad \theta_i \sim N(\mu, \tau^2)$$





# HMC can identify problems in model

Gaussian processes  
for Computational  
Biology: Theory

Peter Tonner

Introduction

Stan

Introduction

**Hierarchical Models**

Basic GP model

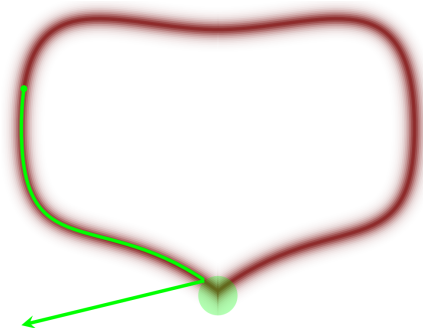
Functional Linear  
Model

Single-cell

RNA-seq

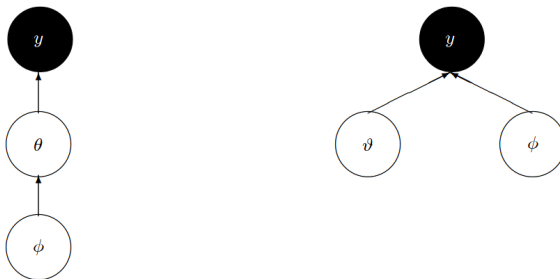
Reid *et al.*

$$y_i \sim N(\theta_i, \sigma_y^2) \quad \text{and} \quad \theta_i \sim N(\mu, \tau^2)$$



# Solution: non-centered parameterization

$$y_i \sim N(\vartheta_i \tau + \mu, \sigma_y^2) \quad \text{and} \quad \vartheta_i \sim N(0, 1)$$
$$\Rightarrow \vartheta_i \tau \sim (0, \tau^2)$$



("Hamiltonian Monte Carlo for Hierarchical Models", Michael Betancourt and Mark Girolam)

## Introduction

## Stan

Introduction

Hierarchical Models

**Basic GP model**

Functional Linear Model

## Single-cell RNA-seq

Reid *et al.*

## Introduction

### Stan

Introduction

Hierarchical Models

**Basic GP model**

Functional Linear  
Model

### Single-cell

### RNA-seq

Reid *et al.*

# A full Bayesian GP model (for Stan)

Gaussian processes  
for Computational  
Biology: Theory

Peter Tonner

Introduction

Stan

Introduction

Hierarchical Models

**Basic GP model**

Functional Linear  
Model

Single-cell

RNA-seq

Reid *et al.*

$$y(x) \sim N\left(f(x), \sigma_y^2\right)$$

$$f(x) \sim GP\left(\mu(x), \kappa(x, \alpha, \ell)\right)$$

$$\sigma_y^2 \sim \text{Ga}(\cdot, \cdot)$$

$$\alpha \sim \text{Ga}(\cdot, \cdot)$$

$$\ell \sim \text{InvGa}(\cdot, \cdot)$$

$$\Sigma = LL^T$$

- ▶  $L$ : lower triangular matrix
- ▶ allows for fast solving of  $\Sigma x = y$ , e.g. compute  $\Sigma^{-1}y$
- ▶ if  $L$  is a Cholesky factor of  $\Sigma$  and  $z \sim N(0, I)$ , then

$$x = \mu + Lz \sim N(\mu, \Sigma)$$

# Note about sampling GPs in Stan

$f(x)$ ,  $\ell$ , and  $\alpha$  are highly related - certain values of one influence the probability of the others and vice-versa.

## Non-centered implementation

$$f_{\eta}(x) \sim N(0, 1)$$

$$\kappa(x, \alpha, \ell) = LL^T \quad \text{and} \quad f(x) = L \cdot f_{\eta}(x)$$

$$\Rightarrow f(x) \sim N\left(0, \kappa(x, \alpha, \ell)\right)$$

## Introduction

## Stan

Introduction

Hierarchical Models

Basic GP model

**Functional Linear Model**

## Single-cell RNA-seq

Reid *et al.*

## Introduction

### Stan

Introduction

Hierarchical Models

Basic GP model

**Functional Linear  
Model**

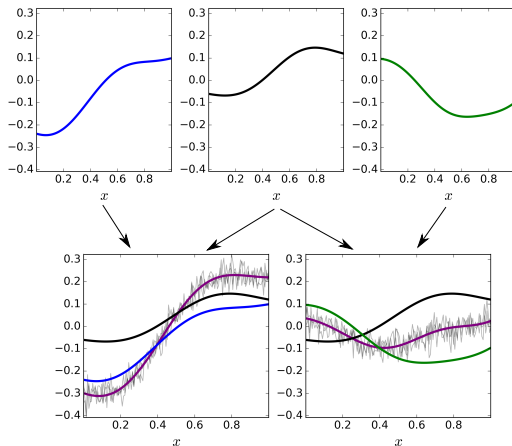
### Single-cell

### RNA-seq

Reid *et al.*

# Functional Linear Model

$$y_r(x) = \sum_{i=1}^k M_{r,i} \cdot \beta_i(x) + \epsilon(x)$$





# Functional Linear Model

---

Gaussian processes  
for Computational  
Biology: Theory

Peter Tonner

$$y_r(x) = \sum_{i=1}^k M_{r,i} \cdot \beta_i(x) + \epsilon(x)$$

Introduction

Stan

Introduction

Hierarchical Models

Basic GP model

**Functional Linear  
Model**

Single-cell

RNA-seq

Reid *et al.*

# Functional Linear Model

$$y_r(x) = \sum_{i=1}^k M_{r,i} \cdot \beta_i(x) + \epsilon(x)$$

$y(x) : n \times p$	observations
---------------------	--------------

$M : p \times k$	design matrix
------------------	---------------

$g \in \{1, \dots, L\}^k$	group assignment
---------------------------	------------------

# Functional Linear Model

$$y_r(x) = \sum_{i=1}^k M_{r,i} \cdot \beta_i(x) + \epsilon(x)$$

$y(x) : n \times p$	observations
---------------------	--------------

$M : p \times k$	design matrix
------------------	---------------

$g \in \{1, \dots, L\}^k$	group assignment
---------------------------	------------------

Each group  $j$  has a GP prior with hyperparameters  $\sigma_j^2, \ell_j$

$$\beta(\mathbf{x}) = \{\beta_1(\mathbf{x}), \dots, \beta_{m_1}(\mathbf{x}), \beta_{m_1+1}(\mathbf{x}), \dots, \beta_{m_1+m_2}(\mathbf{x}), \dots, \beta_{\sum_g m_g}(\mathbf{x})\}$$

$$\begin{array}{c} \swarrow \quad | \quad \searrow \\ GP(\mu_1(x), K_1(x_1, x_2) \mid \theta_1) \quad GP(\mu_2(x), K_g(x_1, x_2) \mid \theta_2) \quad \dots \\ \swarrow \quad \searrow \quad \quad \swarrow \quad \searrow \\ \sigma_1^2, \ell_1 \quad \quad \quad \sigma_2^2, \ell_2 \end{array}$$

## Introduction

## Stan

- Introduction
- Hierarchical Models
- Basic GP model
- Functional Linear Model

## Single-cell RNA-seq

Reid *et al.*

## Introduction

## Stan

- Introduction
- Hierarchical Models
- Basic GP model
- Functional Linear Model

## Single-cell

## RNA-seq

Reid *et al.*

# Single-cell RNA-seq and pseudotime

Gaussian processes  
for Computational  
Biology: Theory

Peter Tonner

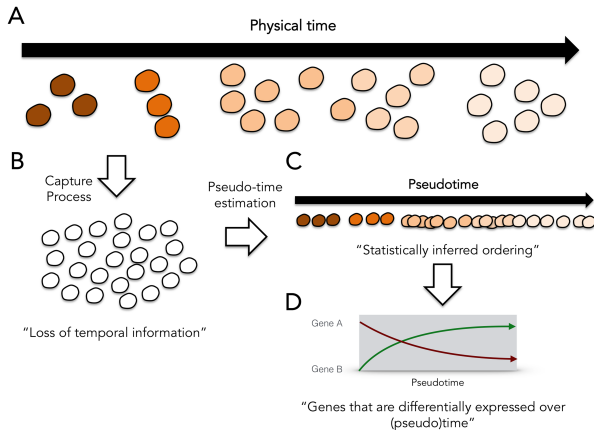
Introduction

Stan

Introduction  
Hierarchical Models  
Basic GP model  
Functional Linear  
Model

Single-cell  
RNA-seq

Reid *et al.*



"Order Under Uncertainty: Robust Differential Expression Analysis Using Probabilistic Models for Pseudotime Inference", Campbell *et al.*, 2016

$$y_g \sim N(\phi_g, \Sigma_g)$$

$$\Sigma_g(\tau_1, \tau_2) = \psi_g \kappa(\tau_1, \tau_2) + \omega_g \delta(\tau_1 = \tau_2)$$

$$\log(\psi_g) \sim N() \quad \text{and} \quad \log(\omega_g) \sim N()$$

$$\tau_i \sim N(c_i, \sigma_\tau^2)$$