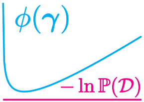
Convex Variational Bayesian Inference for Large Scale Generalized Linear Models

Hannes Nickisch[1] and Matthias W. Seeger[2]

[1]Max Planck Institute for Biological Cybernetics Tübingen, Germany
[2]Saarland University, Saarbrücken, Germany

June 16, 2009

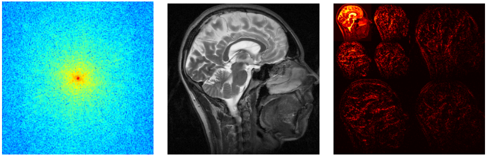


MAX-PLANCK-GESELLSCHAFT

MPI FOR BIOLOGICAL CYBERNETICS

Motivation

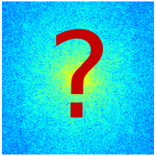
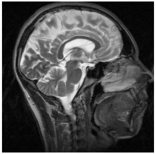
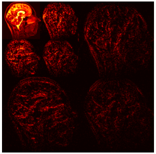
- Image acquisition in MRI



- Binary classification

Motivation

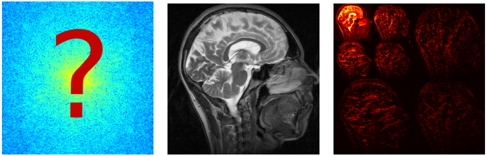
- Image acquisition in MRI

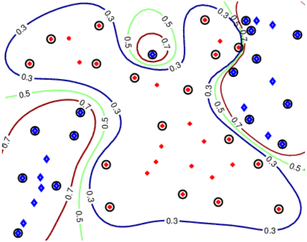
- Binary classification

Motivation

- Image acquisition in MRI

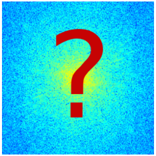
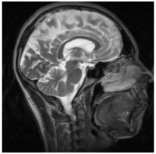
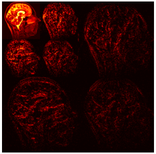


- Binary classification

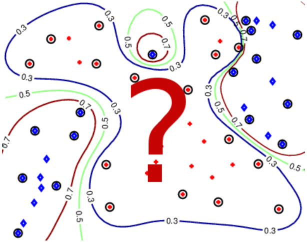


Motivation

- Image acquisition in MRI

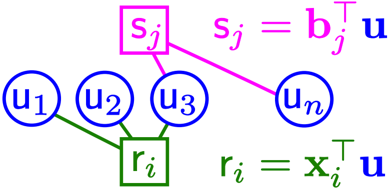
- Binary classification



Generalized Linear Model and Experimental Design

- Generalized Linear Model of $\mathbf{y} = \mathbf{X}\mathbf{u} + \varepsilon$, $\mathbf{s} = \mathbf{B}\mathbf{u}$
- Gaussian $\mathcal{N}(r_i|y_i, \sigma^2)$ and non-Gaussian potentials $t_j(s_j)$

$$\mathbb{P}(\mathbf{u}|\mathcal{D}) \propto \mathcal{N}(\mathbf{y}|\mathbf{X}\mathbf{u}, \sigma^2\mathbf{I}) \prod_{j=1}^{n} t_j(s_j)$$



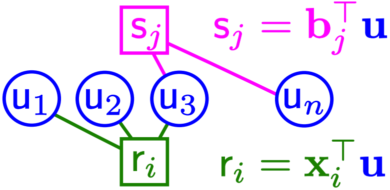
Experimental Design / Measurement Optimization

Along which $\mathbf{x}_i$ or $\mathbf{b}_j$ shall I measure? $\Rightarrow$ Needs posterior covariance info!

Generalized Linear Model and Experimental Design

- Generalized Linear Model of $\mathbf{y} = \mathbf{X}\mathbf{u} + \varepsilon$, $\mathbf{s} = \mathbf{B}\mathbf{u}$
- Gaussian $\mathcal{N}(r_i|y_i, \sigma^2)$ and non-Gaussian potentials $t_j(s_j)$

$$\mathbb{P}(\mathbf{u}|\mathcal{D}) \propto \mathcal{N}(\mathbf{y}|\mathbf{X}\mathbf{u}, \sigma^2\mathbf{I}) \prod_{j=1}^{n} t_j(s_j)$$



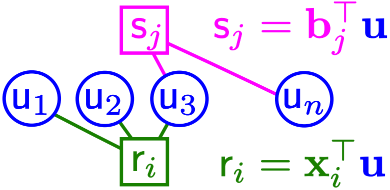
Experimental Design / Measurement Optimization

Along which $\mathbf{x}_i$ or $\mathbf{b}_j$ shall I measure? $\Rightarrow$ Needs posterior covariance info!

Generalized Linear Model and Experimental Design

- Generalized Linear Model of $\mathbf{y} = \mathbf{X}\mathbf{u} + \varepsilon$, $\mathbf{s} = \mathbf{B}\mathbf{u}$
- Gaussian $\mathcal{N}(r_i | y_i, \sigma^2)$ and non-Gaussian potentials $t_j(s_j)$

$$\mathbb{P}(\mathbf{u}|\mathcal{D}) \propto \mathcal{N}(\mathbf{y}|\mathbf{X}\mathbf{u}, \sigma^2\mathbf{I}) \prod_{j=1}^{n} t_j(s_j)$$



Experimental Design / Measurement Optimization

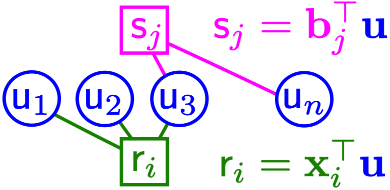
Along which $\mathbf{x}_i$ or $\mathbf{b}_j$ shall I measure? $\Rightarrow$ Needs posterior covariance info!

- MRI: $\mathbf{u}$ unknown image
  - scanner output $\mathbf{y}$, measurement design $\mathbf{X}$
  - sparsity prior $t_j(s_j)$ on multi scale gradients $\mathbf{B}\mathbf{u}$

Generalized Linear Model and Experimental Design

- Generalized Linear Model of $\mathbf{y} = \mathbf{X}\mathbf{u} + \varepsilon$, $\mathbf{s} = \mathbf{B}\mathbf{u}$
- Gaussian $\mathcal{N}(r_i|y_i, \sigma^2)$ and non-Gaussian potentials $t_j(s_j)$

$$\mathbb{P}(\mathbf{u}|\mathcal{D}) \propto \mathcal{N}(\mathbf{y}|\mathbf{X}\mathbf{u}, \sigma^2\mathbf{I}) \prod_{j=1}^{n} t_j(s_j)$$



Experimental Design / Measurement Optimization

Along which $\mathbf{x}_i$ or $\mathbf{b}_j$ shall I measure? $\Rightarrow$ Needs posterior covariance info!

- Classification: $\mathbf{u}$ classifier weights
  - Bernoulli potentials and sparsity prior $t_j(s_j)$ or Gaussian prior $\mathcal{N}$ on $\mathbf{u}$

Posterior I: Site Bounding

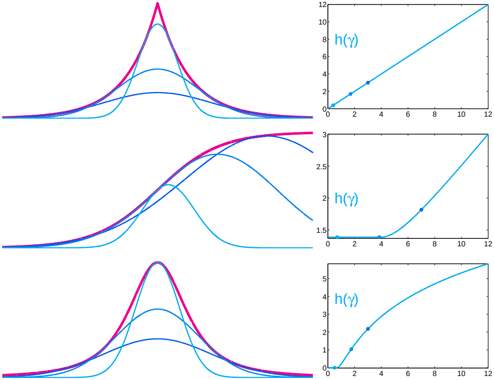
- Legendre-Fenchel (super Gaussian) site bounding:

$$t_j(s_j) \geq \exp\left(\beta_i s_j - \frac{1}{2}s_j^2/\gamma_j - \frac{1}{2}h(\gamma_j)\right) =: \tilde{t}(s_j, \gamma_j)$$

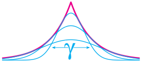
Laplace

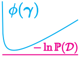
Bernoulli

Student's t



Posterior II: Variational Inference Problem

- site bounds 

$\Longrightarrow$ partition function bound 

$$\begin{aligned}
\mathbb{P}(\mathcal{D}) &= \int \mathcal{N}(\mathbf{y}|\mathbf{X}\mathbf{u}, \sigma^2\mathbf{I}) \prod_{i=1}^{q} t_j(s_j) d\mathbf{u} \\
&\geq \int \mathcal{N}(\mathbf{y}|\mathbf{X}\mathbf{u}, \sigma^2\mathbf{I}) \prod_{i=1}^{q} \tilde{t}(s_j, \gamma_j) d\mathbf{u} = C \cdot \exp\left(-\frac{1}{2}\phi(\boldsymbol{\gamma})\right) \\
\phi(\boldsymbol{\gamma}) &= \ln |\mathbf{A}_{\boldsymbol{\gamma}}| + \sum_j h_j(\gamma_j) + \min_{\mathbf{u}} R(\mathbf{u}, \boldsymbol{\gamma})
\end{aligned}$$

$$\mathbf{A}_{\boldsymbol{\gamma}} = \mathbf{X}^\top\mathbf{X} + \mathbf{B}^\top\boldsymbol{\Gamma}^{-1}\mathbf{B}, \quad R(\mathbf{u}, \boldsymbol{\gamma}) = \|\mathbf{X}\mathbf{u} - \mathbf{y}\|^2 + \mathbf{u}^\top\mathbf{B}^\top\boldsymbol{\Gamma}^{-1}\mathbf{B}\mathbf{u} - 2\boldsymbol{\beta}^\top\mathbf{B}\mathbf{u}$$

Convexity

$$\phi(\gamma) \;=\; \overbrace{\ln|\mathbf{A}_\gamma|}^{2.} + \overbrace{\sum_j h_j(\gamma_j)}^{3.} + \overbrace{\min_{\mathbf{u}} R(\mathbf{u},\gamma)}^{1.}$$

$$\mathbf{A}_\gamma = \mathbf{X}^\top\mathbf{X} + \mathbf{B}^\top\mathbf{\Gamma}^{-1}\mathbf{B}, \quad R(\mathbf{u},\gamma) = \|\mathbf{X}\mathbf{u}-\mathbf{y}\|^2 + \mathbf{u}^\top\mathbf{B}^\top\mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{u} - 2\beta^\top\mathbf{B}\mathbf{u}$$

- approximate posterior $\mathbb{Q}(\mathbf{u}|\mathcal{D}) = \mathcal{N}(\mathbf{u}^*, \mathbf{A}_\gamma^{-1})$

1. $R(\mathbf{u},\gamma)$ jointly convex $\Rightarrow \min_{\mathbf{u}} R(\mathbf{u},\gamma)$ convex
2. $\ln|\mathbf{A}_\gamma|$ convex in γ
3. $h_j(\gamma_j)$ convex in $\gamma_j \Leftrightarrow \ln t_j(s_j)$ concave

Convexity

$$\phi(\gamma) \quad = \quad \overbrace{\ln|\mathbf{A}_\gamma|}^{2.} + \overbrace{\sum_j h_j(\gamma_j)}^{3.} + \overbrace{\min_{\mathbf{u}} R(\mathbf{u},\gamma)}^{[1.]}$$

$$\mathbf{A}_\gamma = \mathbf{X}^\top\mathbf{X} + \mathbf{B}^\top\mathbf{\Gamma}^{-1}\mathbf{B}, \quad R(\mathbf{u},\gamma) = \|\mathbf{X}\mathbf{u} - \mathbf{y}\|^2 + \mathbf{u}^\top\mathbf{B}^\top\mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{u} - 2\beta^\top\mathbf{B}\mathbf{u}$$

- approximate posterior $\mathbb{Q}(\mathbf{u}|\mathcal{D}) = \mathcal{N}(\mathbf{u}^*, \mathbf{A}_\gamma^{-1})$

1. $R(\mathbf{u},\gamma)$ jointly convex $\Rightarrow \min_{\mathbf{u}} R(\mathbf{u},\gamma)$ convex
2. $\ln|\mathbf{A}_\gamma|$ convex in γ
3. $h_j(\gamma_j)$ convex in $\gamma_j \Leftrightarrow \ln t_j(s_j)$ concave

Convexity

$$\phi(\gamma) = \overbrace{\ln|\mathbf{A}_\gamma|}^{[2.]} + \overbrace{\sum_j h_j(\gamma_j)}^{3.} + \overbrace{\min_{\mathbf{u}} R(\mathbf{u},\gamma)}^{1.}$$

$$\mathbf{A}_\gamma = \mathbf{X}^\top\mathbf{X} + \mathbf{B}^\top\mathbf{\Gamma}^{-1}\mathbf{B}, \quad R(\mathbf{u},\gamma) = \|\mathbf{X}\mathbf{u} - \mathbf{y}\|^2 + \mathbf{u}^\top\mathbf{B}^\top\mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{u} - 2\beta^\top\mathbf{B}\mathbf{u}$$

- approximate posterior $\mathbb{Q}(\mathbf{u}|\mathcal{D}) = \mathcal{N}(\mathbf{u}^*, \mathbf{A}_\gamma^{-1})$

1. $R(\mathbf{u},\gamma)$ jointly convex $\Rightarrow$ $\min_{\mathbf{u}} R(\mathbf{u},\gamma)$ convex
2. $\ln|\mathbf{A}_\gamma|$ convex in γ
3. $h_j(\gamma_j)$ convex in γ_j $\Leftrightarrow$ $\ln t_j(s_j)$ concave

Convexity

$$\phi(\gamma) = \overbrace{\ln|\mathbf{A}_\gamma|}^{2.} + \overbrace{\sum_j h_j(\gamma_j)}^{[3.]} + \overbrace{\min_{\mathbf{u}} R(\mathbf{u},\gamma)}^{1.}$$

$$\mathbf{A}_\gamma = \mathbf{X}^\top\mathbf{X} + \mathbf{B}^\top\boldsymbol{\Gamma}^{-1}\mathbf{B}, \quad R(\mathbf{u},\gamma) = \|\mathbf{X}\mathbf{u}-\mathbf{y}\|^2 + \mathbf{u}^\top\mathbf{B}^\top\boldsymbol{\Gamma}^{-1}\mathbf{B}\mathbf{u} - 2\beta^\top\mathbf{B}\mathbf{u}$$

- approximate posterior $\mathbb{Q}(\mathbf{u}|\mathcal{D}) = \mathcal{N}(\mathbf{u}^*, \mathbf{A}_\gamma^{-1})$

1. $R(\mathbf{u},\gamma)$ jointly convex $\Rightarrow \min_{\mathbf{u}} R(\mathbf{u},\gamma)$ convex
2. $\ln|\mathbf{A}_\gamma|$ convex in γ
3. $h_j(\gamma_j)$ convex in $\gamma_j \Leftrightarrow \ln t_j(s_j)$ concave

Optimization

$$\phi(\gamma) = \ln|\mathbf{A}| + h(\gamma) + \min_{\mathbf{u}} \left(\|\mathbf{X}\mathbf{u} - \mathbf{y}\|^2 + \mathbf{s}^\top \mathbf{\Gamma}^{-1}\mathbf{s} - 2\beta^\top \mathbf{s} \right)$$

$$\mathbf{A} = \mathbf{X}^\top\mathbf{X} + \mathbf{B}^\top\mathbf{\Gamma}^{-1}\mathbf{B}, \quad \mathbf{s} = \mathbf{B}\mathbf{u}$$

- $\phi(\gamma)$ is convex, so things are easy, right?
- gradient hard: $\nabla_{\gamma^{-1}} \ln|\mathbf{A}| = \mathrm{diag}(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\top)$
- coordinate descent hard: n linear systems
- coupling term $\ln|\mathbf{A}|$ causing trouble $\Rightarrow$ decouple
- $\gamma^{-1} \mapsto \ln|\mathbf{A}|$ concave
- Legendre duality:
  $\ln|\mathbf{A}| \le \underbrace{\mathbf{z}^\top(\gamma^{-1}) - g^*(\mathbf{z})}_{\text{convex in } \gamma}$

Optimization

$$\phi(\gamma) \quad = \quad \ln|\mathbf{A}| + h(\gamma) + \min_{\mathbf{u}} \left(\|\mathbf{X}\mathbf{u} - \mathbf{y}\|^2 + \mathbf{s}^\top \mathbf{\Gamma}^{-1}\mathbf{s} - 2\beta^\top \mathbf{s} \right)$$

$$\mathbf{A} = \mathbf{X}^\top\mathbf{X} + \mathbf{B}^\top\mathbf{\Gamma}^{-1}\mathbf{B}, \quad \mathbf{s} = \mathbf{B}\mathbf{u}$$

- $\phi(\gamma)$ is convex, so things are easy, right?
- gradient hard: $\nabla_{\gamma^{-1}} \ln|\mathbf{A}| = \operatorname{diag}(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\top)$
- coordinate descent hard: n linear systems
- coupling term $\ln|\mathbf{A}|$ causing trouble $\Rightarrow$ decouple

- $\gamma^{-1} \mapsto \ln|\mathbf{A}|$ concave
- Legendre duality:
  $\ln|\mathbf{A}| \leq \underbrace{\mathbf{z}^\top(\gamma^{-1}) - g^*(\mathbf{z})}_{\text{convex in } \gamma}$

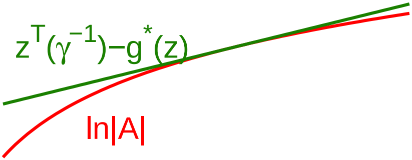
Optimization

$$\phi(\gamma) \;=\; \ln|\mathbf{A}| + h(\gamma) + \min_{\mathbf{u}} \left(\|\mathbf{X}\mathbf{u} - \mathbf{y}\|^2 + \mathbf{s}^\top \boldsymbol{\Gamma}^{-1}\mathbf{s} - 2\beta^\top \mathbf{s} \right)$$

$$\mathbf{A} = \mathbf{X}^\top\mathbf{X} + \mathbf{B}^\top\boldsymbol{\Gamma}^{-1}\mathbf{B}, \quad \mathbf{s} = \mathbf{B}\mathbf{u}$$

- $\phi(\gamma)$ is convex, so things are easy, right?
- gradient hard: $\nabla_{\gamma^{-1}} \ln|\mathbf{A}| = \mathrm{diag}(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\top)$
- coordinate descent hard: n linear systems
- coupling term $\ln|\mathbf{A}|$ causing trouble $\Rightarrow$ decouple

- $\gamma^{-1} \mapsto \ln|\mathbf{A}|$ concave
- Legendre duality:
  $\ln|\mathbf{A}| \le \underbrace{\mathbf{z}^\top(\gamma^{-1}) - g^*(\mathbf{z})}_{\text{convex in } \gamma}$

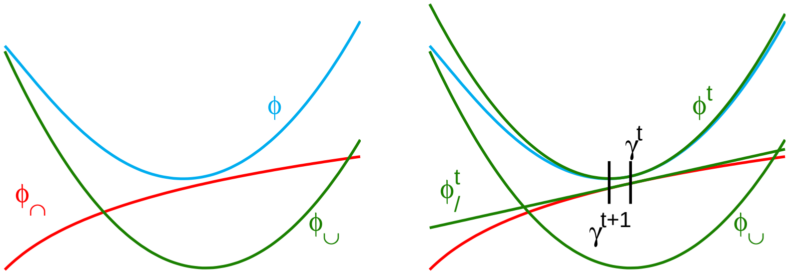


Double Loop Details

$$\phi(\gamma) \quad = \quad \ln|\mathbf{A}| + h(\gamma) + \min_{\mathbf{u}} \left(\|\mathbf{X}\mathbf{u} - \mathbf{y}\|^2 + \mathbf{s}^\top \mathbf{\Gamma}^{-1}\mathbf{s} - 2\beta^\top \mathbf{s} \right)$$

$$\mathbf{A} = \mathbf{X}^\top \mathbf{X} + \mathbf{B}^\top \mathbf{\Gamma}^{-1} \mathbf{B}, \quad \mathbf{s} = \mathbf{B}\mathbf{u}$$

- outer: $\phi = \phi_\cap + \phi_\cup \leq \phi_/ + \phi_\cup$, $\quad \ln|\mathbf{A}| \leq \mathbf{z}^\top(\gamma^{-1}) - g^*(\mathbf{z})$
- inner: $\min_\gamma \phi_/ + \phi_\cup$



Double Loop Summary

$$\phi(\gamma) \quad = \quad \ln|\mathbf{A}| + h(\gamma) + \min_{\mathbf{u}} \left(\|\mathbf{X}\mathbf{u} - \mathbf{y}\|^2 + \mathbf{s}^\top \mathbf{\Gamma}^{-1}\mathbf{s} - 2\beta^\top \mathbf{s} \right)$$

$$\mathbf{A} = \mathbf{X}^\top \mathbf{X} + \mathbf{B}^\top \mathbf{\Gamma}^{-1}\mathbf{B}, \quad \mathbf{s} = \mathbf{B}\mathbf{u}$$

Gradient Based Algorithm

- `for` $t = 1..\mathcal{O}(n)$
  - `compute gradient` $\nabla_\gamma \phi$
  - `get descent direction` $\boldsymbol{\delta}$
  - `compute stepsize` λ
  - $\gamma^{t+1} \leftarrow \gamma^t + \lambda\boldsymbol{\delta}$

Double Loop Algorithm

- `for` $t = 1..\mathcal{O}(1)$
  - `outer loop update:` `compute gradient` $\nabla_\gamma \phi$
  - `inner loop optim:` $\gamma^{t+1} \leftarrow \mathtt{IRLS}(\gamma^t)$

- efficient use of expensive computations

Experimental Design

- approximate posterior $\mathbb{Q}(\mathbf{u}) = \mathcal{N}(\mathbf{u}^*, \mathbf{A}^{-1})$, conditioned on $\mathcal{D}$
- relative entropy $\mathrm{KL}[\mathbb{Q}'||\mathbb{Q}] = \mathcal{H}[\mathbb{Q}'||\mathbb{Q}] - \mathcal{H}[\mathbb{Q}'] \geq 0$
- information gain score
- MRI: continuous Gaussian sites

$$S_{IG}(\mathbf{x}_i) = \int \mathbb{Q}(y_i)\mathrm{KL}[\mathbb{Q}'(\mathbf{u}|y_i)||\mathbb{Q}(\mathbf{u})]\mathrm{d}y_i$$

- Classification: binary Bernoulli sites

$$S_{IG}(\mathbf{b}_j) = \sum_{c_j=\pm 1} \mathbb{Q}(c_j)\mathrm{KL}[\mathbb{Q}'(\mathbf{u}|c_j)||\mathbb{Q}(\mathbf{u})]$$

- design loop: (1) update $\mathbb{Q}$, (2) design decision, (3) measurement

Experimental Design

- approximate posterior $\mathbb{Q}(\mathbf{u}) = \mathcal{N}(\mathbf{u}^*, \mathbf{A}^{-1})$, conditioned on $\mathcal{D}$
- relative entropy $\mathrm{KL}[\mathbb{Q}'||\mathbb{Q}] = \mathcal{H}[\mathbb{Q}'||\mathbb{Q}] - \mathcal{H}[\mathbb{Q}'] \geq 0$
- information gain score
- MRI: continuous Gaussian sites

$$S_{IG}(\mathbf{x}_i) = \int \mathbb{Q}(y_i)\mathrm{KL}[\mathbb{Q}'(\mathbf{u}|y_i)||\mathbb{Q}(\mathbf{u})]\mathrm{d}y_i$$

- Classification: binary Bernoulli sites

$$S_{IG}(\mathbf{b}_j) = \sum_{c_j=\pm 1} \mathbb{Q}(c_j)\mathrm{KL}[\mathbb{Q}'(\mathbf{u}|c_j)||\mathbb{Q}(\mathbf{u})]$$

- design loop: (1) update $\mathbb{Q}$, (2) design decision, (3) measurement

Experimental Design

- approximate posterior $\mathbb{Q}(\mathbf{u}) = \mathcal{N}(\mathbf{u}^*, \mathbf{A}^{-1})$, conditioned on $\mathcal{D}$
- relative entropy $\mathrm{KL}[\mathbb{Q}'||\mathbb{Q}] = \mathcal{H}[\mathbb{Q}'||\mathbb{Q}] - \mathcal{H}[\mathbb{Q}'] \geq 0$
- information gain score
- MRI: continuous Gaussian sites

$$S_{IG}(\mathbf{x}_i) = \int \mathbb{Q}(y_i)\mathrm{KL}[\mathbb{Q}'(\mathbf{u}|y_i)||\mathbb{Q}(\mathbf{u})]\mathrm{d}y_i$$

- Classification: binary Bernoulli sites

$$S_{IG}(\mathbf{b}_j) = \sum_{c_j=\pm 1} \mathbb{Q}(c_j)\mathrm{KL}[\mathbb{Q}'(\mathbf{u}|c_j)||\mathbb{Q}(\mathbf{u})]$$

- design loop: (1) update $\mathbb{Q}$, (2) design decision, (3) measurement

Experimental Design

- approximate posterior $\mathbb{Q}(\mathbf{u}) = \mathcal{N}(\mathbf{u}^*, \mathbf{A}^{-1})$, conditioned on $\mathcal{D}$
- relative entropy $\mathrm{KL}[\mathbb{Q}'||\mathbb{Q}] = \mathcal{H}[\mathbb{Q}'||\mathbb{Q}] - \mathcal{H}[\mathbb{Q}'] \geq 0$
- information gain score
- MRI: continuous Gaussian sites

$$S_{IG}(\mathbf{x}_i) = \int \mathbb{Q}(y_i)\mathrm{KL}[\mathbb{Q}'(\mathbf{u}|y_i)||\mathbb{Q}(\mathbf{u})]\mathrm{d}y_i$$

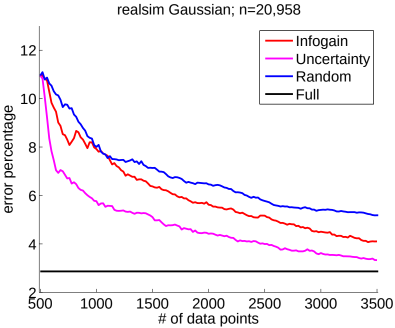
- Classification: binary Bernoulli sites

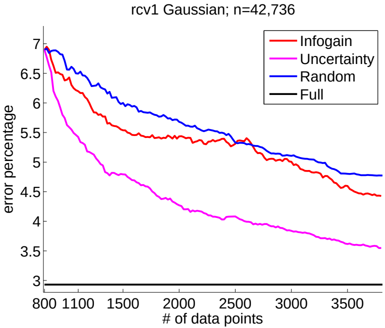
$$S_{IG}(\mathbf{b}_j) = \sum_{c_j=\pm 1} \mathbb{Q}(c_j)\mathrm{KL}[\mathbb{Q}'(\mathbf{u}|c_j)||\mathbb{Q}(\mathbf{u})]$$

- design loop: (1) update $\mathbb{Q}$, (2) design decision, (3) measurement

Experiments

- large scale active learning using approximate inference





Conclusions

- variational relaxation convex iff. MAP estimation convex
- posterior approximation successfully drives experimental design
- fully scalable and generic double loop algorithm
- computational primitives: CG (means), Lanczos (variances)
- complexity = speed of MVMs with $\mathbf{X}$ and $\mathbf{B}$

Appendix: Convexity of log determinant

Theorem

$\gamma \mapsto \ln|\mathbf{A}_\gamma|$ *is convex.* $\mathbf{A}_\gamma = \mathbf{X}^\top\mathbf{X} + \mathbf{B}^\top\mathbf{\Gamma}^{-1}\mathbf{B}$

- $(\mathbf{u}, \gamma) \mapsto \mathbf{u}^\top\left(\mathbf{X}^\top\mathbf{X} + \mathbf{B}^\top\mathbf{\Gamma}^{-1}\mathbf{B}\right)\mathbf{u}$ jointly convex
- $(\mathbf{u}, \gamma) \mapsto \exp(-\frac{1}{2}\mathbf{u}^\top\mathbf{A}_\gamma\mathbf{u})$ jointly log-concave
- marginalization theorem (Prékopa)
  - Log-concave functions are closed under marginalization.
- **Ga**ussian integral: $\ln|\mathbf{A}_\gamma| = n\ln 2\pi - 2\ln\int\exp(-\frac{1}{2}\mathbf{u}^\top\mathbf{A}_\gamma\mathbf{u})d\mathbf{u}$

Theorem

$\gamma \mapsto \ln|\mathbf{X}^\top\mathbf{X} + \mathbf{B}^\top f(\mathbf{\Gamma})\mathbf{B}|$ *is convex iff.* $\ln f(\gamma)$ *is convex.*

Appendix: Convexity of individual height functions

Theorem

$h(\gamma)$ is convex iff. $g(s) = \ln t(s)$ is concave in s and convex in $x = s^2$.

1. $t(s) = \max_{\gamma \geq 0} \exp\left(-\frac{s^2}{2\gamma} - h(\gamma)\right)$, $h(\gamma) = g^*(-\frac{1}{2\gamma})$
2. $h(\gamma) = \max_{s \geq 0} f(s, \gamma) = \max_{s \geq 0} \frac{-1}{2\gamma} s^2 - g(s) = \max_{x \geq 0} \frac{-1}{2\gamma} x - g(x)$
3. $0 \stackrel{!}{=} \frac{\partial}{\partial x} f(x, \gamma) \Rightarrow g'(x_*) = -\frac{1}{2\gamma}$ implicitly defining $x_*(\gamma)$ and $h(\gamma) = f(x_*(\gamma), \gamma)$
4. $\frac{\mathrm{d}x_*}{\mathrm{d}\gamma} = \frac{1}{2\gamma^2 g''(x_*)} \wedge x \mapsto g(x)$ convex $\Rightarrow \gamma \mapsto x_*$ increasing $\Rightarrow \gamma \mapsto s_*$ increasing
5. $0 \stackrel{!}{=} \frac{\partial}{\partial s} f(s, \gamma) \Rightarrow g'(s_*) = -\frac{s_*}{\gamma}$, $\quad g(s)$ concave $\Rightarrow g'(s)$ decreasing
6. $h'(\gamma) = \frac{\partial}{\partial \gamma} f(s_*, \gamma) = \frac{1}{2}\left(g'(s_*)\right)^2 \wedge (5) \Rightarrow s_* \mapsto h'(\gamma)$ increasing due to square
7. $(4,6) \Rightarrow \gamma \mapsto h'(\gamma)$ increasing $\Rightarrow h(\gamma)$ convex