Large Scale Sequential Experimental Design for Signal Acquisition Optimization

Matthias Seeger

mseeger@gmail.com

08/05/2015

Outline

Magnetic Resonance Imaging

- ⊕ Extremely versatile
- ⊕ Noninvasive, no ionizing radiation
- ⊖ Very expensive
- ⊖ Long scan times: Major limiting factor

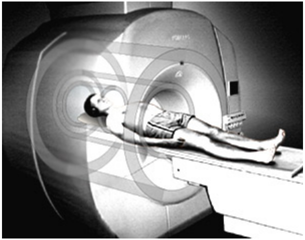
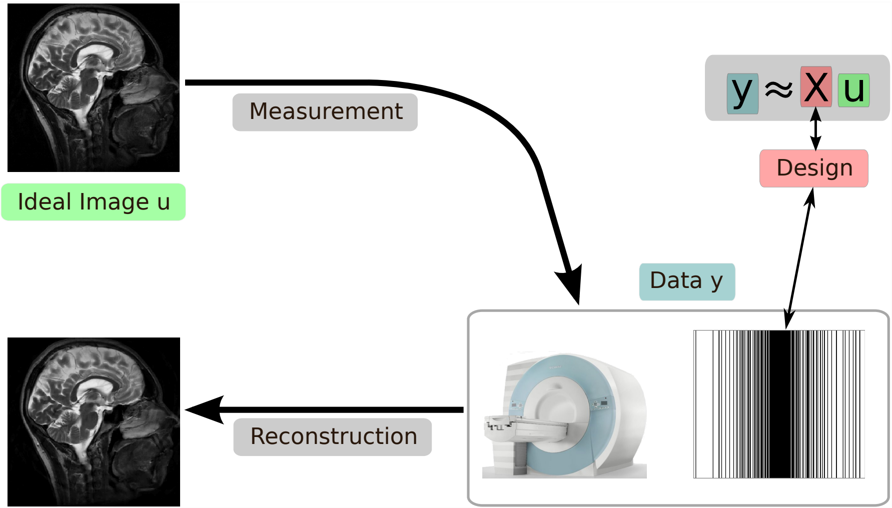
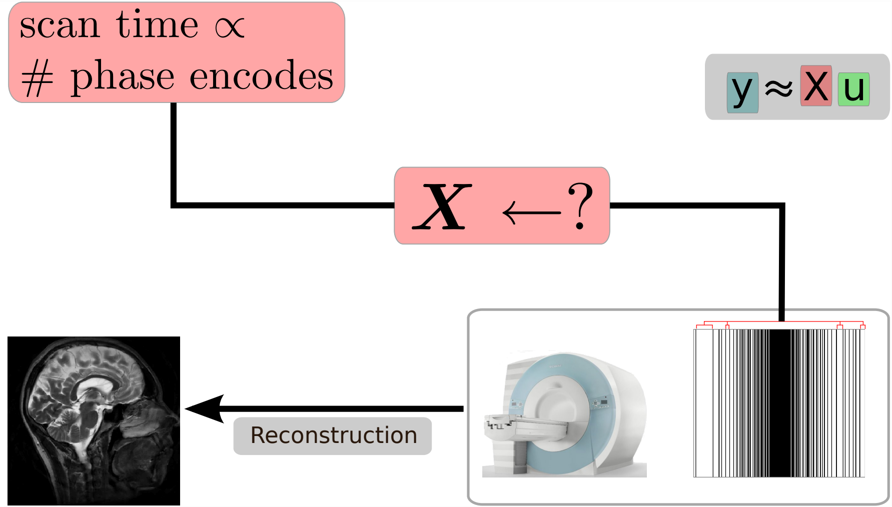


Image Reconstruction



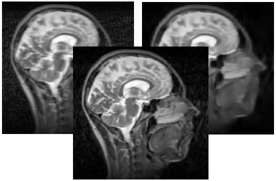
Sampling Optimization

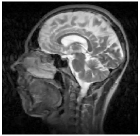


Posterior Distribution

- Likelihood $P(\boldsymbol{y}|\boldsymbol{u})$: Data fit
- Prior $P(\boldsymbol{u})$: Signal properties
- Posterior distribution $P(\boldsymbol{u}|\boldsymbol{y})$: Consistent information summary

$P(\boldsymbol{y}|\boldsymbol{u})$

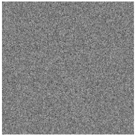


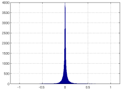


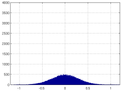
Posterior Distribution

- Likelihood $P(\boldsymbol{y}|\boldsymbol{u})$: Data fit
- Prior $P(\boldsymbol{u})$: Signal properties
- Posterior distribution $P(\boldsymbol{u}|\boldsymbol{y})$: Consistent information summary

$$P(\boldsymbol{y}|\boldsymbol{u}) \times P(\boldsymbol{u})$$



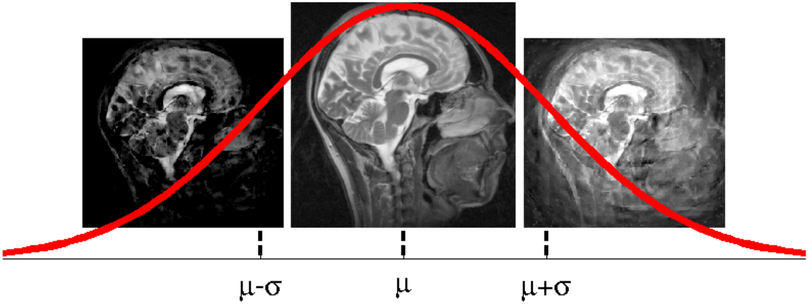




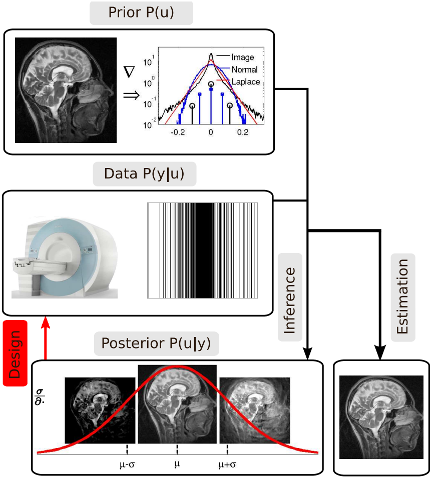
Posterior Distribution

- Likelihood $P(\boldsymbol{y}|\boldsymbol{u})$: Data fit
- Prior $P(\boldsymbol{u})$: Signal properties
- Posterior distribution $P(\boldsymbol{u}|\boldsymbol{y})$: Consistent information summary

$$P(\boldsymbol{u}|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|\boldsymbol{u}) \times P(\boldsymbol{u})}{P(\boldsymbol{y})}$$

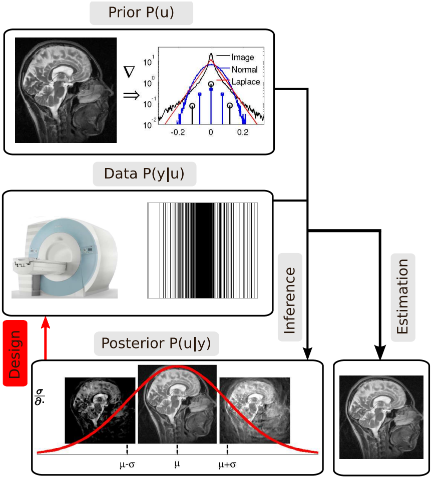


Bayesian Experimental Design

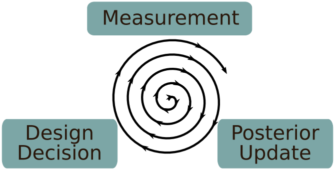


- Posterior: Uncertainty in reconstruction
- Experimental design: Find poorly determined directions
- Sequential search with interjacent partial measurements

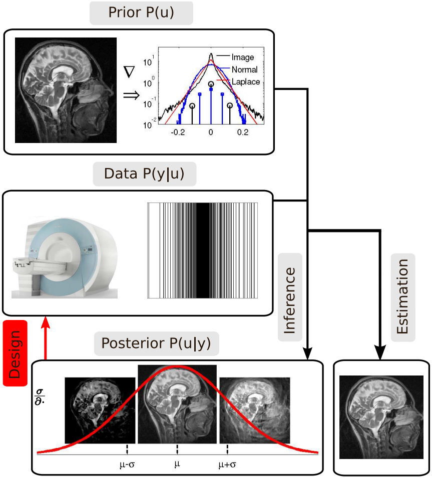
Bayesian Experimental Design



- Posterior: Uncertainty in reconstruction
- Experimental design: Find poorly determined directions
- Sequential search with interjacent partial measurements

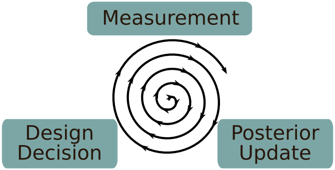


Bayesian Experimental Design

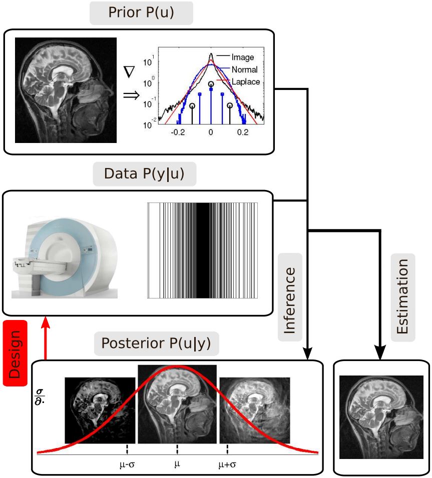


Bayesian Sequential Design

- How much do I know?
  $P(\boldsymbol{u}|\boldsymbol{y}), \mathcal{H} = \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})]$
- How much would $(\boldsymbol{x}_*, y_*)$ help?
  $I(\boldsymbol{x}_*, y_*) = \mathcal{H} - \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y}, y_*)]$
- How much will $\boldsymbol{x}_*$ help?
  $I(\boldsymbol{x}_*) = \mathrm{E}_{P(y_*|\boldsymbol{y})}[I(\boldsymbol{x}_*, y_*)]$

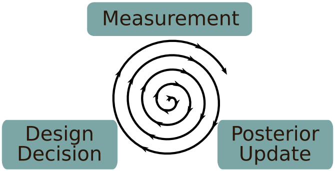


Bayesian Experimental Design

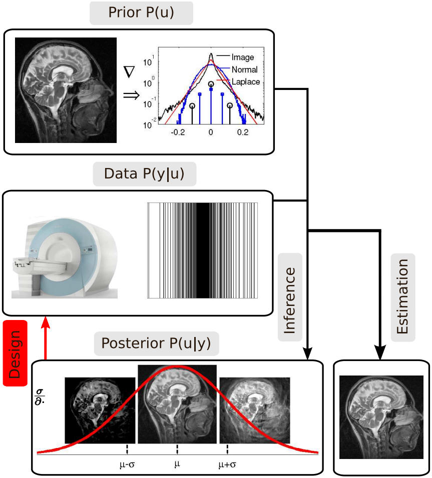


Bayesian Sequential Design

- How much do I know?
  $P(\boldsymbol{u}|\boldsymbol{y})$, $\mathcal{H} = \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})]$
- How much would $(\boldsymbol{x}_*, y_*)$ help?
  $I(\boldsymbol{x}_*, y_*) = \mathcal{H} - \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y}, y_*)]$
- How much will $\boldsymbol{x}_*$ help?
  $I(\boldsymbol{x}_*) = \mathrm{E}_{P(y_*|\boldsymbol{y})}[I(\boldsymbol{x}_*, y_*)]$

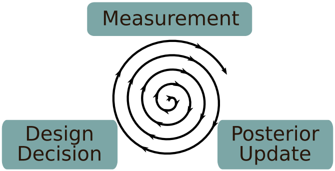


Bayesian Experimental Design



Bayesian Sequential Design

- How much do I know?
  $P(\boldsymbol{u}|\boldsymbol{y})$, $\mathcal{H} = \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})]$
- How much would $(\boldsymbol{x}_*, y_*)$ help?
  $I(\boldsymbol{x}_*, y_*) = \mathcal{H} - \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y}, y_*)]$
- How much will $\boldsymbol{x}_*$ help?
  $I(\boldsymbol{x}_*) = \mathrm{E}_{P(y_*|\boldsymbol{y})}[I(\boldsymbol{x}_*, y_*)]$



Maximizing Information Gain

- Score design extension $\boldsymbol{X}_*$ by information gain:

$$\mathrm{I}(\boldsymbol{X}_*) = \mathrm{I}(\boldsymbol{y}_*, \boldsymbol{u}|\boldsymbol{y}) = \underbrace{\mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})]}_{\text{Before}} - \underbrace{\mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y}_*, \boldsymbol{y})]}_{\text{After}}$$

- Standard approach in literature:
  - Assume: $\mathrm{I}(\boldsymbol{X}_*)$ tractable to compute
  - Assume: $\mathrm{I}(\boldsymbol{X}_*)$ cheap to compute (many $\boldsymbol{X}_*$)
  - Concentrate on combinatorial optimization aspects

Maximizing Information Gain

- Score design extension $\boldsymbol{X}_*$ by information gain:

$$\mathrm{I}(\boldsymbol{X}_*) = \mathrm{I}(\boldsymbol{y}_*, \boldsymbol{u} | \boldsymbol{y}) = \underbrace{\mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})]}_{\text{Before}} - \underbrace{\mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y}_*, \boldsymbol{y})]}_{\text{After}}$$

- Standard approach in literature:
  - Assume: $\mathrm{I}(\boldsymbol{X}_*)$ tractable to compute
  - Assume: $\mathrm{I}(\boldsymbol{X}_*)$ cheap to compute (many $\boldsymbol{X}_*$)
  - Concentrate on combinatorial optimization aspects

Maximizing Information Gain

- Score design extension $\boldsymbol{X}_*$ by information gain:

$$\mathrm{I}(\boldsymbol{X}_*) = \mathrm{I}(\boldsymbol{y}_*, \boldsymbol{u} | \boldsymbol{y}) = \underbrace{\mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})]}_{\text{Before}} - \underbrace{\mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y}_*, \boldsymbol{y})]}_{\text{After}}$$

- Standard approach in literature:
  - Assume: $\mathrm{I}(\boldsymbol{X}_*)$ tractable to compute
  - Assume: $\mathrm{I}(\boldsymbol{X}_*)$ cheap to compute (many $\boldsymbol{X}_*$)
  - Concentrate on combinatorial optimization aspects

So is it . . . ?

Challenges

$$\mathrm{I}(\boldsymbol{X}_*) = \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})] - \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y}_*, \boldsymbol{y})]$$

- Assume: $\mathrm{I}(\boldsymbol{X}_*)$ tractable to compute.
  Only if $P(\boldsymbol{u}|\boldsymbol{y})$ Gaussian ...

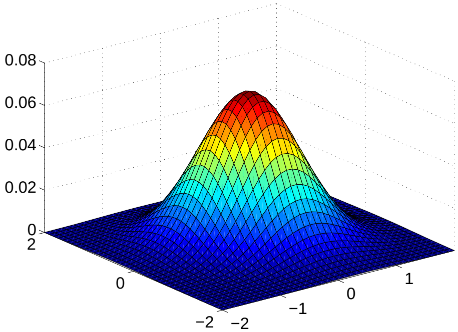
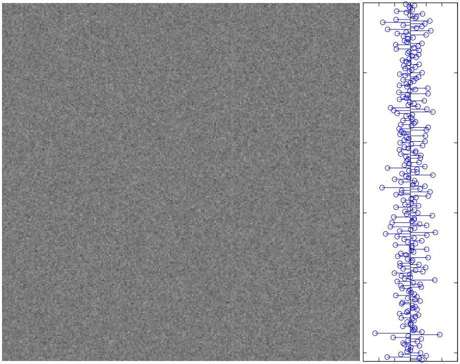
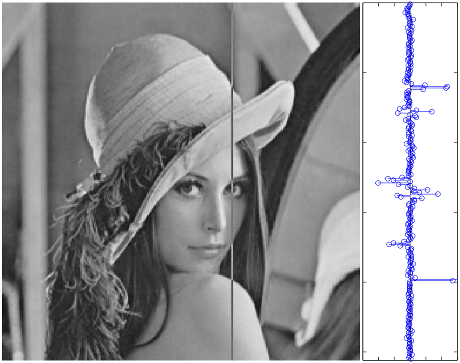


Image Statistics

Whatever images are . . .

they are not Gaussian!





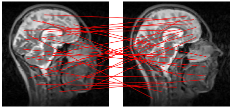
Challenges

$$\mathrm{I}(\boldsymbol{X}_*) = \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})] - \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y}_*, \boldsymbol{y})]$$

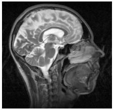
- Assume: $\mathrm{I}(\boldsymbol{X}_*)$ tractable to compute?
  No: Needs approximate inference
- Assume: $\mathrm{I}(\boldsymbol{X}_*)$ cheap to compute (many $\boldsymbol{X}_*$).

Size Does Matter

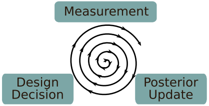
1. **Global covariances**
   Scores $\mathrm{I}(\boldsymbol{X}_*)$ need full $\mathrm{Cov}_P[\boldsymbol{u}|\boldsymbol{y}]$



2. **Massive scale**
   $\mathbb{R}^{131072}$ (just one slice).



3. **Many times**
   Posterior after each design extension



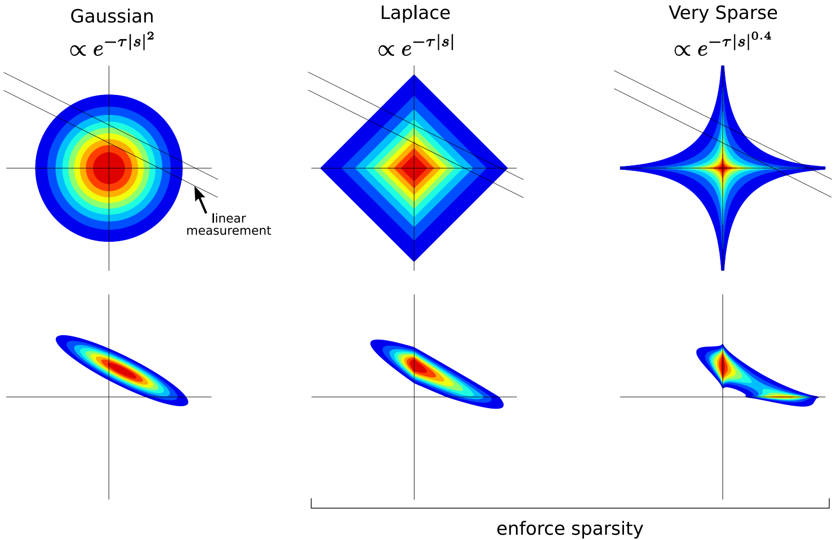
Challenges

$$\mathrm{I}(\boldsymbol{X}_*) = \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})] - \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y}_*, \boldsymbol{y})]$$

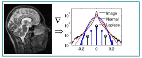
- Assume: $\mathrm{I}(\boldsymbol{X}_*)$ tractable to compute?
  No: Needs approximate inference
- Assume: $\mathrm{I}(\boldsymbol{X}_*)$ cheap to compute (many $\boldsymbol{X}_*$)?
  No: Needs new algorithms and high performance computing

Sparsity Priors

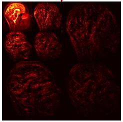
courtesy Florian Steinke



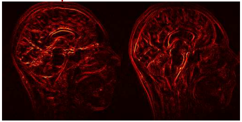
Sparse Linear Model



$$P(\mathbf{u}) \propto \prod_{i=1}^{q} t_i(s_i) = e^{-\tau_w \|\mathbf{B}_w \mathbf{u}\|_1} \times e^{-\tau_{tv} \|\mathbf{B}_{tv} \mathbf{u}\|_1}, \quad \mathbf{s} = \mathbf{B}\mathbf{u}$$

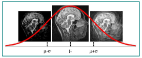


wavelet



gradient

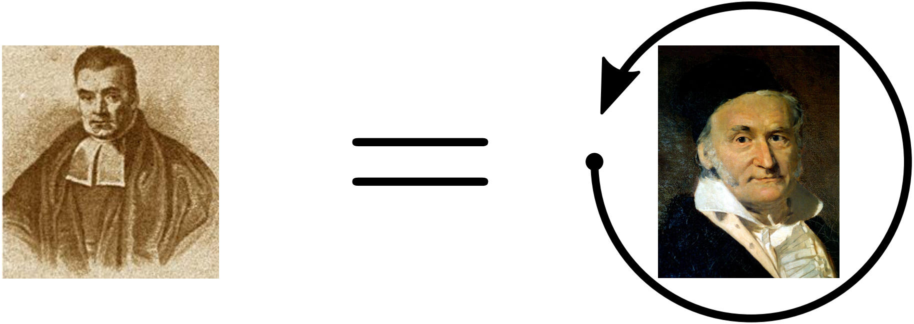
$$P(\mathbf{y}|\mathbf{u}) = \mathcal{N}(\mathbf{y}|\mathbf{X}\mathbf{u}, \sigma^2 \mathbf{I})$$



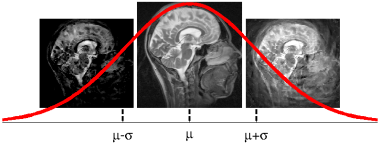
$$P(\mathbf{u}|\mathbf{y}) \propto P(\mathbf{u}) P(\mathbf{y}|\mathbf{u})$$

Variational Bayesian Inference

- Approximate inference for non-Gaussian models
- Computations driven by Gaussian inference



Variational Bayesian Inference

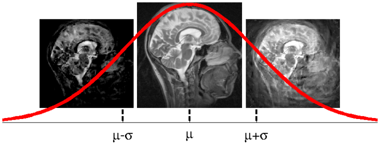


$$P(\boldsymbol{u}|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|\boldsymbol{u}) \times P(\boldsymbol{u})}{P(\boldsymbol{y})}$$

Variational Inference Approximation

- Write intractable integration as optimization
- Relax to tractable optimization problem

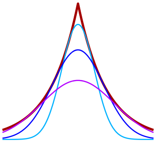
Variational Bayesian Inference



$$P(\boldsymbol{u}|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|\boldsymbol{u}) \times P(\boldsymbol{u})}{P(\boldsymbol{y})}$$

- Variational Relaxation: Bound the master function

$$-\log P(\boldsymbol{y}) = -\log \int P(\boldsymbol{u}, \boldsymbol{y})\, d\boldsymbol{u} \leq \frac{1}{2} \min_{\gamma} \min_{\boldsymbol{u}_*} \phi(\boldsymbol{u}_*, \gamma)$$



  - Approximate posterior $P(\boldsymbol{u}|\boldsymbol{y})$ by Gaussian
  - Integration $\Rightarrow$ Convex optimization

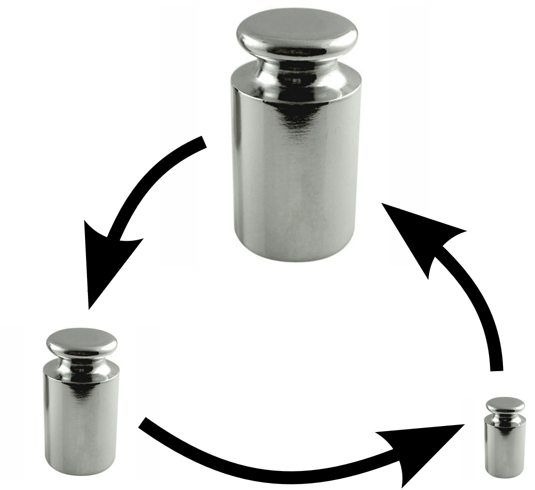
No Inference Without . . .

Gaussian Variances

Linear Systems

Peanuts $O(q)$

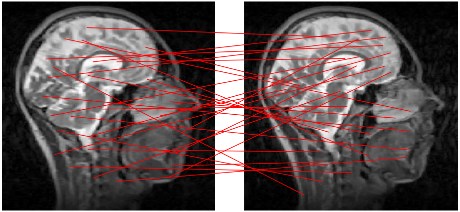
Why Inference Algorithms Are Slow



Decoupling by Concavity

$$-\log \int P(\boldsymbol{u}, \boldsymbol{y})\, d\boldsymbol{u} \le \min_{\gamma} \min_{\boldsymbol{u}} \phi(\boldsymbol{u}, \gamma)$$

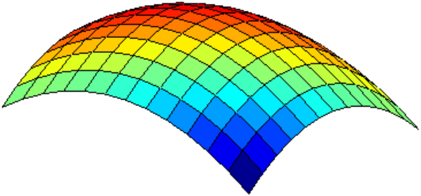
- Dependencies in posterior $P(\boldsymbol{u}|\boldsymbol{y})$
  $\Rightarrow$ Difficult couplings in criterion ϕ
- Critically coupled part is concave
- Upper bound by tangent plane



Decoupling by Concavity

$$-\log \int P(\boldsymbol{u}, \boldsymbol{y})\, d\boldsymbol{u} \le \min_{\gamma} \min_{\boldsymbol{u}} \phi(\boldsymbol{u}, \gamma)$$

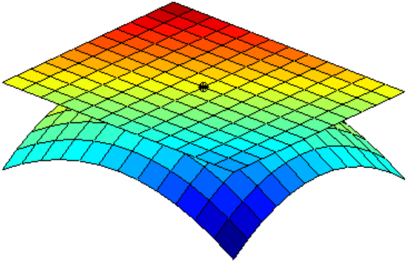
- Dependencies in posterior $P(\boldsymbol{u}|\boldsymbol{y})$
  $\Rightarrow$ Difficult couplings in criterion ϕ
- Critically coupled part is concave
- Upper bound by tangent plane



Decoupling by Concavity

$$-\log \int P(\boldsymbol{u}, \boldsymbol{y})\, d\boldsymbol{u} \leq \min_{\gamma} \min_{\boldsymbol{u}} \phi(\boldsymbol{u}, \gamma) = \min_{\boldsymbol{z}} \underbrace{\min_{\gamma} \min_{\boldsymbol{u}} \phi_{\boldsymbol{z}}(\boldsymbol{u}, \gamma)}_{\text{Decoupled problem}}$$

- Dependencies in posterior $P(\boldsymbol{u}|\boldsymbol{y})$
  $\Rightarrow$ Difficult couplings in criterion ϕ
- Critically coupled part is concave
- Upper bound by tangent plane

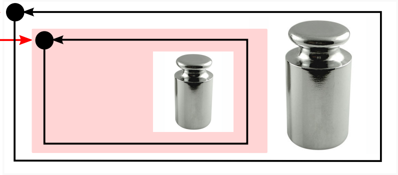


Double Loop Algorithm

Double loop algorithm

- Inner loop optimization:
  Standard MAP Estimation
- Outer loop update:
  Gaussian Variances

$$\min_{\gamma} \phi \quad = \min_{\gamma} \quad \min_{z} \min_{u_*} \phi$$
$$\stackrel{!}{=} \min_{z}$$
$$\min_{u_*} \min_{\gamma} \phi$$



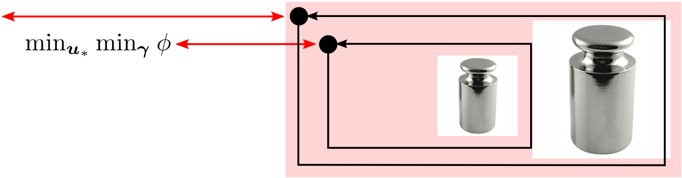
Double Loop Algorithm

Double loop algorithm

- Inner loop optimization:
  Standard MAP Estimation
- Outer loop update:
  Gaussian Variances

$$\min_{\boldsymbol{\gamma}} \phi \;=\; \min_{\boldsymbol{\gamma}} \;\; \min_{\boldsymbol{z}} \min_{\boldsymbol{u}_*} \phi$$

$$\stackrel{!}{=} \min_{\boldsymbol{z}} \;\; \min_{\boldsymbol{u}_*} \min_{\boldsymbol{\gamma}} \phi$$



Closing the Loop

$$I(\boldsymbol{X}_*) = \mathrm{E}_{P(\boldsymbol{y}_*|\boldsymbol{y})}\left[\mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})] - \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{y}_*)]\right]$$

Gaussian prior $P(\boldsymbol{u})$? Things are simple

- Posterior $P(\boldsymbol{u}|\boldsymbol{y})$ Gaussian as well
- Information gain tractable

$$I(\boldsymbol{X}_*) = \frac{1}{2}\log\left|\boldsymbol{I} + \boldsymbol{X}_*\boldsymbol{A}^{-1}\boldsymbol{X}_*^T\right|, \quad \boldsymbol{A} = \mathrm{Cov}_P[\boldsymbol{u}|\boldsymbol{y}]^{-1}$$

$\Rightarrow$ Posterior covariance

Closing the Loop

$$I(\boldsymbol{X}_*) = \mathrm{E}_{P(\boldsymbol{y}_*|\boldsymbol{y})}\left[\mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})] - \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{y}_*)]\right]$$

Gaussian prior $P(\boldsymbol{u})$? Things are simple

- Posterior $P(\boldsymbol{u}|\boldsymbol{y})$ Gaussian as well
- Information gain tractable

$$I(\boldsymbol{X}_*) = \frac{1}{2}\log\left|\boldsymbol{I} + \boldsymbol{X}_*\boldsymbol{A}^{-1}\boldsymbol{X}_*^T\right|, \quad \boldsymbol{A} = \mathrm{Cov}_Q[\boldsymbol{u}|\boldsymbol{y}]^{-1}$$

$\Rightarrow$ Posterior covariance

Image priors $P(\boldsymbol{u})$ are not Gaussian . . .

- Variational approximation:
  Gaussian posterior $Q(\boldsymbol{u}|\boldsymbol{y}) \approx P(\boldsymbol{u}|\boldsymbol{y})$

Variational Bayesian Experimental Design

$$I(\boldsymbol{X}_*) = \mathrm{E}_{P(\boldsymbol{y}_*|\boldsymbol{y})}\left[\mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y})] - \mathrm{H}[P(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{y}_*)]\right]$$

Given initial $\boldsymbol{X}$, $\boldsymbol{y}$.

repeat

Variational inference: Update γ s.t. $Q(\boldsymbol{u}|\boldsymbol{y}) \approx P(\boldsymbol{u}|\boldsymbol{y})$.

for $\boldsymbol{X}_* \in \mathcal{X}_{\text{cand}}$ **do**

Compute approximate information gain:

$I(\boldsymbol{X}_*) = \frac{1}{2}\log|\boldsymbol{I} + \boldsymbol{X}_*\boldsymbol{A}^{-1}\boldsymbol{X}_*^T|, \quad \boldsymbol{A} = \mathrm{Cov}_Q[\boldsymbol{u}|\boldsymbol{y}]^{-1}$.

end for

Pick score maximizer $\boldsymbol{X}_*$. Acquire data $\boldsymbol{y}_*$.

Append $\boldsymbol{X}_*$ to $\boldsymbol{X}$, $\boldsymbol{y}_*$ to $\boldsymbol{y}$.

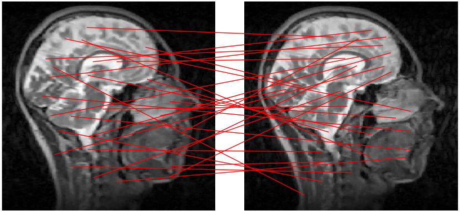
until $\boldsymbol{X}$ has desired size

Gaussian Covariances

Elephant in the room . . .

Covariance matrix

$$\mathrm{Cov}_Q[\boldsymbol{u}|\boldsymbol{y}] = \boldsymbol{A}^{-1}$$
$$\boldsymbol{A} = \sigma^{-2}\boldsymbol{X}^H\boldsymbol{X} + \boldsymbol{B}^T\Gamma^{-1}\boldsymbol{B}$$



1. Outer loop updates: $\boldsymbol{z} \leftarrow \mathrm{diag}(\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T)$
2. Experimental design scores: $\mathrm{I}(\boldsymbol{X}_*) = \frac{1}{2}\log|\boldsymbol{I} + \sigma^{-2}\boldsymbol{X}_*\boldsymbol{A}^{-1}\boldsymbol{X}_*^H|$

Gaussian Covariances

$$\mathrm{Cov}_Q[\boldsymbol{u}|\boldsymbol{y}]^{-1} = \boldsymbol{A} = \sigma^{-2}\boldsymbol{X}^H\boldsymbol{X} + \boldsymbol{B}^T\Gamma^{-1}\boldsymbol{B}$$

- Some fields care about them
  - Electronic structure calculations
  - Uncertainty quantifications for PDEs
  - Gaussian MRFs for remote sensing
- Simple trick: Perturb&MAP — Papandreou, Yuille, NIPS 2010

$$\boldsymbol{w}_l \sim N(\boldsymbol{0}, \boldsymbol{A}) \;\; [\text{simple}], \quad \boldsymbol{q}_l = \boldsymbol{A}^{-1}\boldsymbol{w}_l \sim N(\boldsymbol{0}, \boldsymbol{A}^{-1}),$$
$$\mathrm{Cov}_Q[\boldsymbol{u}|\boldsymbol{y}] \approx \sum_l \boldsymbol{q}_l(\boldsymbol{q}_l)^H$$

  - One linear system per sample (parallelizable)
  - Noisy, but unbiased

Gaussian Covariances

$$\mathrm{Cov}_Q[\boldsymbol{u}|\boldsymbol{y}]^{-1} = \boldsymbol{A} = \sigma^{-2}\boldsymbol{X}^H\boldsymbol{X} + \boldsymbol{B}^T\Gamma^{-1}\boldsymbol{B}$$

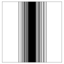
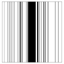
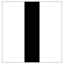
- Some fields care about them
  - Electronic structure calculations
  - Uncertainty quantifications for PDEs
  - Gaussian MRFs for remote sensing
- Simple trick: Perturb&MAP

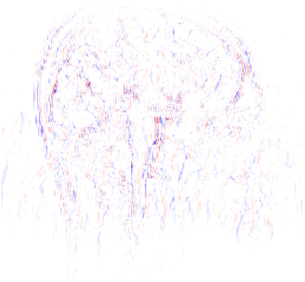
Papandreou, Yuille, NIPS 2010

$$\boldsymbol{w}_l \sim N(\boldsymbol{0}, \boldsymbol{A}) \text{ [simple]}, \quad \boldsymbol{q}_l = \boldsymbol{A}^{-1}\boldsymbol{w}_l \sim N(\boldsymbol{0}, \boldsymbol{A}^{-1}),$$
$$\mathrm{Cov}_Q[\boldsymbol{u}|\boldsymbol{y}] \approx \sum_l \boldsymbol{q}_l(\boldsymbol{q}_l)^H$$

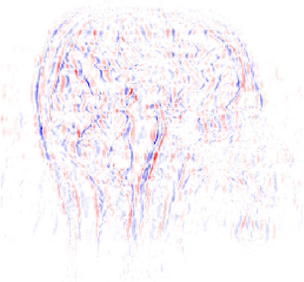
  - One linear system per sample (parallelizable)
  - Noisy, but unbiased

Optimizing Cartesian MRI

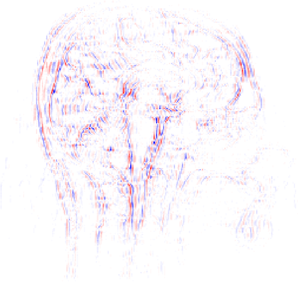
Bayes Optim.  VD Random  Low Pass 



Seeger *et.al.*, MRM 63(1), 2010

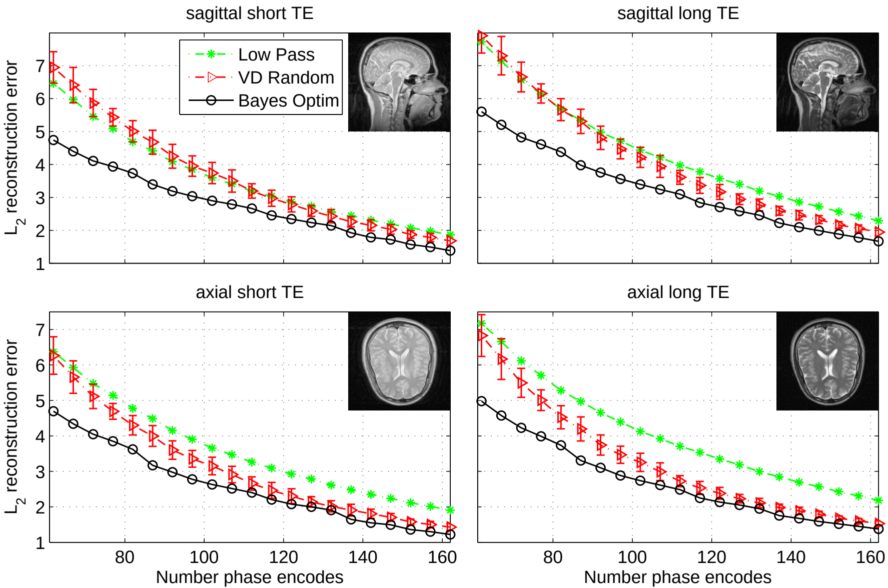


Lustig, Donoho, Pauli, MRM 58(6), 2007



Common MRI practice

Experimental Results: Test Set Errors



Large Scale Bayesian Inference

- Computer Vision
  - Hierarchically structured image priors — Ko, Seeger, ICML 2012
  - Learning Image Models (fields of experts, . . .)
  - Bayesian dictionary learning
  - Intelligent user interfaces (Bayesian active learning)
- Advanced variational inference
  - Speeding up expectation propagation — Seeger, Nickisch, AISTATS 2011
- Generic framework
  - You can do penalized estimation efficiently?
    You can do variational Bayesian inference!

Large Scale Bayesian Inference

- Computer Vision
  - Hierarchically structured image priors — Ko, Seeger, ICML 2012
  - Learning Image Models (fields of experts, . . .)
  - Bayesian dictionary learning
  - Intelligent user interfaces (Bayesian active learning)
- Advanced variational inference
  - Speeding up expectation propagation — Seeger, Nickisch, AISTATS 2011
- Generic framework
  - You can do penalized estimation efficiently?
    You can do variational Bayesian inference!

Large Scale Bayesian Inference

- Computer Vision
  - Hierarchically structured image priors — Ko, Seeger, ICML 2012
  - Learning Image Models (fields of experts, . . .)
  - Bayesian dictionary learning
  - Intelligent user interfaces (Bayesian active learning)
- Advanced variational inference
  - Speeding up expectation propagation — Seeger, Nickisch, AISTATS 2011
- Generic framework
  - You can do penalized estimation efficiently?
    You can do variational Bayesian inference!

People& Code

glm-ie: Toolbox by Hannes Nickisch

mloss.org/software/view/269/

- Generalized sparse linear models
- MAP reconstruction and variational Bayesian inference (double loop algorithm for super-Gaussian bounding)
- Matlab 7.x, GNU Octave 3.2.x

- Hannes Nickisch (now Philips Research, Hamburg)
- Rolf Pohmann, Bernhard Schölkopf (MPI Tübingen)
- Young Jun Ko (EPFL)
- Emtiyaz Khan (EPFL)

References

- Seeger, Nickisch
  Large Scale Bayesian Inference and Experimental Design for Sparse Linear Models
  SIAM Journal on Imaging Science 4(1), 2011
- Seeger, Nickisch, Pohmann, Schölkopf
  Optimization of k-Space Trajectories for Compressed Sensing by Bayesian Experimental Design
  Magnetic Resonance in Medicine 63(1), 2010
- Seeger, Wipf
  Variational Bayesian Inference Techniques
  IEEE Signal Processing Magazine 27(6), 2010
- Seeger, Nickisch
  Compressed Sensing and Bayesian Experimental Design
  International Conference on Machine Learning (ICML), 2008