Computational and Theoretical Aspects of Sparsity-Constraints in Bayesian Inversion

Mini-Symposium "Sparsity-Promoting Computational Inversion"

Applied Inverse Problem Conference 2013 in Daejeon, Korea
Sparsity Constraints in Inverse Problems

Current trend in high dimensional inverse problems: **Sparsity constraints.**

- **Compressed Sensing:** High quality reconstructions from a small amount of data, if a sparse basis/dictionary is a-priori known (e.g., wavelets).
- **Total Variation (TV) imaging:** Sparsity constraints on the gradient of the unknowns.

Thank’s to Jahn Müller for these images!

(a) 20 min, EM  
(b) 5 sec, EM  
(c) 5 sec, Bregman EM-TV
Sparsity Constraints in Variational Regularization

Commonly applied formulation and analysis by means of variational regularization, mostly by incorporating L1-type norms:

$$\hat{u}_\alpha = \arg\min_{u \in \mathbb{R}^n} \left\{ \| f - K u \|_2^2 + \alpha |D u|_1 \right\}$$

assuming additive Gaussian i.i.d. noise $\sim \mathcal{N}(0, \sigma^2)$

Martin Burger

Felix Lucka (felix.lucka@wwu.de)
Sparsity Constraints in the Bayesian Approach

Sparsity as a-priori information are encoded into the prior distribution $p_{\text{prior}}(u)$:

1. Turning the functionals used in variational regularization directly into priors, e.g., L1-type priors:
   - Convenient, as prior is log-concave.
   - MAP estimate is sparse, but the prior itself is not sparse.

2. Hierarchical Bayesian modeling: Sparsity is incorporated at a higher level of the model.
   - Relies on a slightly different concept of sparsity.
   - Resulting implicit priors over unknowns are usually not log-concave.

\[
\begin{align*}
\text{(a) } & \exp \left( -\frac{1}{2} \|u\|_2^2 \right) \\
\text{(b) } & \exp \left( -|u|_1 \right) \\
\text{(c) } & \left( 1 + \frac{u^2}{3} \right)^{-2}
\end{align*}
\]
Likelihood:
\[ \exp \left( -\frac{1}{2\sigma^2} \| f - K u \|_2^2 \right) \]

Prior: \( \exp \left( -\lambda \| u \|_1 \right) \)
(\( \lambda \) via discrepancy principle)

Posterior: \( \exp \left( -\frac{1}{2\sigma^2} \| f - K u \|_2^2 - \lambda \| u \|_1 \right) \)
Bayesian Inference and Computational Techniques

Things we might want to do with the posterior:

▶ Point estimates: MAP and CM.
▶ Credible regions estimates
▶ Extreme value probabilities
▶ Conditional covariance estimates
▶ Histogram estimates
▶ Generalized Bayes estimators
▶ Marginalization of nuisance parameters & Approximation error modeling
▶ Model selection or averaging
▶ Experiment design

Computationally, this needs

▶ high-dimensional optimization

1

▶ high-dimensional integration
▶ a mix of both.


Felix Lucka (felix.lucka@wwu.de)
MAP vs. CM Estimates: Variational Regularization vs. Bayesian Inference?

Most simple Bayesian inference technique: Point estimates.

1. Maximum a-posteriori-estimate (MAP):

   \[ \hat{u}_{\text{MAP}} := \arg\max_{u \in \mathbb{R}^n} p_{\text{post}}(u | f) \]

   Practically: High-dimensional optimization problem. Direct correspondence to variational regularization.

2. Conditional mean-estimate (CM):

   \[ \hat{u}_{\text{CM}} := \mathbb{E}[u | f] = \int_{\mathbb{R}^n} u \ p_{\text{post}}(u | f) du \]

   Practically: High-dimensional integration problem.

Difference between MAP and CM estimate?
MAP vs. CM Estimates: Variational Regularization vs. Bayesian Inference?

Most simple Bayesian inference technique: Point estimates.

1. Maximum a-posteriori-estimate (MAP):
   \[ \hat{u}_{\text{MAP}} := \arg\max_{u \in \mathbb{R}^n} p_{\text{post}}(u | f) \]
   Practically: High-dimensional optimization problem.
   Direct correspondence to variational regularization.

2. Conditional mean-estimate (CM):
   \[ \hat{u}_{\text{CM}} := \mathbb{E}[u | f] = \int_{\mathbb{R}^n} u \, p_{\text{post}}(u | f) \, du \]
   Practically: High-dimensional integration problem.

Difference between MAP and CM estimate?

\[ \Rightarrow \] Most interesting question for comparing variational regularization and Bayesian inference?

Felix Lucka (felix.lucka@wwu.de)
Outline

Introduction

MAP vs. CM Estimates: The Classical View

Recent Theoretical and Computational Results
   A Fast Sampler for High-Dimensional Problems
   A 2D Deblurring Example
   The Discretization Dilemma of the TV prior
   Limited Angle CT with Besov Priors

The Rehabilitation of the MAP Estimate

Take Home Messages
MAP vs. CM Estimates: The Classical View

- CM estimate is the mean of the posterior
- MAP estimate the (highest) mode of the posterior.

![Graph showing the comparison between the mean, median, and mode for different values of σ](source: Wikimedia Commons)
MAP vs. CM Estimates: The Classical View

Hypothetical distributions to show that none is better in general.
MAP vs. CM Estimates: The Classical View

Hypothetical distributions to show that none is better in general.
MAP vs. CM Estimates: The Classical View

A theoretical argument “decides” the conflict: The Bayes cost formalism.

- An estimator is a random variable, as it relies on $f$ and $u$.
- How does it perform on average? Which estimator is “best”?
- $\rightsquigarrow$ Define a cost function $\Psi(u, \hat{u}(f))$.
- Bayes cost is the expected cost:
  $$BC(\hat{u}) = \int \int \Psi(u, \hat{u}(f)) \, p_{\text{like}}(f|u) \, df \, p_{\text{prior}}(u) \, du$$
- Bayes estimator $\hat{u}_{BC}$ for given $\Psi$ minimizes Bayes cost.

Felix Lucka (felix.lucka@wwu.de)
MAP vs. CM Estimates: The Classical View

Main classical arguments pro CM and contra MAP estimates:

▶ CM is Bayes estimator for $\Psi(u, \hat{u}) = \|u - \hat{u}\|_2^2$ (MSE).
▶ Also the minimum variance estimator.
▶ The mean value is intuitive, it is the "center of mass", the known "average".
▶ MAP estimate can be seen as an asymptotic Bayes estimator of

$$\Psi_\epsilon(u, \hat{u}) = \begin{cases} 0, & \text{if } \|u - \hat{u}\|_\infty \leq \epsilon \\ 1 & \text{otherwise}, \end{cases}$$

for $\epsilon \to 0$ (uniform cost). $\implies$ It is not a proper Bayes estimator.

Felix Lucka (felix.lucka@wwu.de)
MAP vs. CM Estimates: The Classical View

Main classical arguments pro CM and contra MAP estimates:

▶ CM is Bayes estimator for $\Psi(u, \hat{u}) = \|u - \hat{u}\|_2^2$ (MSE).
▶ Also the minimum variance estimator.
▶ The mean value is intuitive, it is the "center of mass", the known "average".
▶ MAP estimate can be seen as an asymptotic Bayes estimator of

$$\Psi_\epsilon(u, \hat{u}) = \begin{cases} 
0, & \text{if } \|u - \hat{u}\|_\infty \leq \epsilon \\
1 & \text{otherwise},
\end{cases}$$

for $\epsilon \to 0$ (uniform cost). $\implies$ It is not a proper Bayes estimator.

▶ MAP and CM seem theoretically and computationally fundamentally different $\implies$ one should decide.
▶ “A real Bayesian would not use the MAP estimate”
▶ People feel "ashamed" when they have to compute MAP estimates (even when their results are good).
Outline

Introduction

MAP vs. CM Estimates: The Classical View

Recent Theoretical and Computational Results
  A Fast Sampler for High-Dimensional Problems
  A 2D Deblurring Example
  The Discretization Dilemma of the TV prior
  Limited Angle CT with Besov Priors

The Rehabilitation of the MAP Estimate

Take Home Messages
Some Observations...

The discrimination of the MAP estimate is not intuitive.

Gaussian priors: MAP = CM. Funny coincidence?

Non-Gaussian priors:

- Theoretical considerations could often not be validated numerically
- CM as the mysterious, inaccessible estimate.
Some Observations...

The discrimination of the MAP estimate is not intuitive.

Gaussian priors: MAP = CM. Funny coincidence?

Non-Gaussian priors:
  ▶ Theoretical considerations could often not be validated numerically
  ▶ CM as the mysterious, inaccessible estimate.

Need for computational tools for CM estimation (and beyond!)

Fast Markov chain Monte Carlo sampling for sparse Bayesian inference in high-dimensional inverse problems using L1-type priors

Felix Lucka (*felix.lucka@wwu.de*)
Outline

Introduction

MAP vs. CM Estimates: The Classical View

Recent Theoretical and Computational Results
  A Fast Sampler for High-Dimensional Problems
  A 2D Deblurring Example
  The Discretization Dilemma of the TV prior
  Limited Angle CT with Besov Priors

The Rehabilitation of the MAP Estimate

Take Home Messages

Felix Lucka (felix.lucka@wwu.de)
Image Deblurring Example in 2D

Unknown function $\tilde{u}$
- Gaussian blurring + relative noise level of 10%
- Reconstruction using simple L1 prior
- $n = 1023 \times 1023 = 1046529$. 
Image Deblurring Example in 2D

(d) Unknown function \( \tilde{u} \)

(e) MAP estimate by Split Bregman
Image Deblurring Example in 2D

(a) Unknown function $\hat{u}$
(b) CM estimate by our Gibbs sampler

Felix Lucka (felix.lucka@wwu.de)
The Discretization Dilemma of the TV prior (Lassas & Siltanen, 2004)

"Can one use total variation prior for edge-preserving Bayesian inversion?"

- For $\lambda_n \propto \sqrt{n+1}$ and $n \to \infty$ the TV prior converges to a smoothness prior.
- CM converges to smooth limit.
- MAP converges to constant.
The Discretization Dilemma of the TV prior (Lassas & Siltanen, 2004)
"Can one use total variation prior for edge-preserving Bayesian inversion?"

- For $\lambda_n = \text{const.}$ and $n \to \infty$ the TV prior diverges.
- CM diverges.
- MAP converges to edge-preserving limit.

(a) CM by our Gibbs Sampler

(b) MAP by Split Bregman
The Discretization Dilemma of the TV prior (Lassas & Siltanen, 2004)
"Can one use total variation prior for edge-preserving Bayesian inversion?"

- For $\lambda_n = \text{const.}$ and $n \to \infty$ the TV prior diverges.
- CM diverges.
- MAP converges to edge-preserving limit.

(a) Zoom into CM estimates
(b) MCMC convergence check
Discretization Invariant Besov Priors

Question: Is it possible to construct discretization invariant and edge-preserving priors for Bayesian inversion?

Discretization invariant Bayesian inversion and Besov space priors.

Sparsity-promoting Bayesian inversion.

Sparse tomography.
Discretization Invariant Besov Priors

Question: Is it possible to construct discretization invariant and edge-preserving priors for Bayesian inversion?


An interesting and important scenario to implement our L1 sampler!
Computational Scenario

- CT using only 45 projection angles
- 500 measurement pixel
- 1 % relative Gaussian noise added.
Reconstructions for $\lambda = 2e4$, $n = 64 \times 64 = 4.096$

MAP estimate (by Split Bregman)  CM estimate (by our Gibbs sampler)
Reconstructions for $\lambda = 2e4, \ n = 128 \times 128 = 16.384$

MAP estimate (by Split Bregman)  
CM estimate (by our Gibbs sampler)
Reconstructions for $\lambda = 2e4, \ n = 256 \times 256 = 65.536$

MAP estimate (by Split Bregman)  
CM estimate (by our Gibbs sampler)
Reconstructions for $\lambda = 2e4$, $n = 512 \times 512 = 262.144$

MAP estimate (by Split Bregman)  
CM estimate (by our Gibbs sampler)
Reconstructions for $\lambda = 2\times10^4$, $n = 1024 \times 1024 = 1,048,576$

MAP estimate (by Split Bregman)  
CM estimate (by our Gibbs sampler)
Posterior Samples for $\lambda = 2e4$, $n = 1024 \times 1024 = 1.048.576$

Abbildung: Sample 1
Posterior Samples for $\lambda = 2e4$, $n = 1024 \times 1024 = 1.048.576$

Abbildung: Sample 2
Posterior Samples for $\lambda = 2e4$, $n = 1024 \times 1024 = 1.048.576$

Abbildung: Sample 3
Posterior Samples for $\lambda = 2e4$, $n = 1024 \times 1024 = 1.048.576$
Posterior Samples for $\lambda = 2 \times 10^4$, $n = 1024 \times 1024 = 1.048.576$

**Abbildung:** Sample 5
First Results for Sample-Based Tomography with Besov Priors

In line with former results, we have a sampler that works for $n > 10^6$

First reconstructions supports former results of:

  - discretization invariant.
  - MAP and CM coincide for large $\lambda$.

A lot of future work to do!

Felix Lucka (felix.lucka@wwu.de)
Outline

Introduction

MAP vs. CM Estimates: The Classical View

Recent Theoretical and Computational Results
  A Fast Sampler for High-Dimensional Problems
  A 2D Deblurring Example
  The Discretization Dilemma of the TV prior
  Limited Angle CT with Besov Priors

The Rehabilitation of the MAP Estimate

Take Home Messages
Summary of Observations and Discussions

- Gaussian priors: MAP = CM. Funny coincidence?

- For reasonable priors, CM and MAP look quite similar. Fundamentally different?

- If a CM estimate looks good, it looks like the MAP estimate.

- MAP estimates are sparser, sharper, look and perform better,…

- Gribonval, 2011: CM are MAP estimates for different priors.
Bayesian Inversion from a Bregman Distance Perspective

Assume

- Linear $K$
- Additive Gaussian noise: $\mathcal{N}(0, \Sigma_\varepsilon)$
- Log-concave prior, i.e., $p_{prior}(u) \propto \exp(-\lambda J(u))$, where $J(u)$ is convex.

Martin Burger developed several ideas (joint paper in preparation) to shed new light on the issue.

He uses Bregman distances as a main tool.

I will report some key results here.
Excursus: Bregman Distances

\[ D^q_J(u, v) = J(u) - J(v) - \langle q, u - v \rangle, \quad q \in \partial J(v) \]

- Basically: difference between \( J(u) \) and its linearization.
- Proven useful in variational regularization.
A False Conclusion

“A real Bayesian would not use the MAP estimate as it is not a proper Bayes estimator”.

“MAP estimate can be seen as an asymptotic Bayes estimator of

$$\Psi_\epsilon(u, \hat{u}) = \begin{cases} 0, & \text{if } \|u - \hat{u}\|_\infty < \epsilon \\ 1 & \text{otherwise,} \end{cases}$$

for $\epsilon \to 0$.

???>?? It is not a proper Bayes estimator.”

“MAP estimator is asymptotic Bayes estimator for some degenerate $\Psi$”
\implies “MAP can’t be Bayes estimator for some proper $\Psi$” !!!!
Two New Bayes Cost Functions

Define

(a) $\Psi_{LS}(u, \hat{u}) := \|K(\hat{u} - u)\|^2_{\Sigma^{-1}} + \beta \|L(\hat{u} - u)\|^2$

(b) $\Psi_{Brg}(u, \hat{u}) := \|K(\hat{u} - u)\|^2_{\Sigma^{-1}} + \lambda D_J(\hat{u}, u)$

for a regular $L$ and $\beta > 0$.

Properties:

- Proper, convex cost functions
- For $J(u) = \beta/\lambda \|Lu\|^2$ we have $\lambda D_J(\hat{u}, u) = \beta \|L(\hat{u} - u)\|^2$, and $\Psi_{LS}(u, \hat{u}) = \Psi_{Brg}(u, \hat{u})$!
Two New Bayes Cost Functions

Define

(a) $\Psi_{LS}(u, \hat{u}) := \|K(\hat{u} - u)\|_{\Sigma^{-1}}^2 + \beta \|L(\hat{u} - u)\|^2_2$

(b) $\Psi_{Brg}(u, \hat{u}) := \|K(\hat{u} - u)\|_{\Sigma^{-1}}^2 + \lambda D_{\mathcal{J}}(\hat{u}, u)$

for a regular $L$ and $\beta > 0$.

Properties:

- Proper, convex cost functions
- For $\mathcal{J}(u) = \beta/\lambda \|Lu\|_2^2$ we have $\lambda D_{\mathcal{J}}(\hat{u}, u) = \beta \|L(\hat{u} - u)\|^2_2$, and $\Psi_{LS}(u, \hat{u}) = \Psi_{Brg}(u, \hat{u})$.

Theorems:

(I) The CM estimate is the Bayes estimator for $\Psi_{LS}(u, \hat{u})$

(II) The MAP estimate is the Bayes estimator for $\Psi_{Brg}(u, \hat{u})$.

Felix Lucka (felix.lucka@wwu.de)
The Posterior is Well Centered around the MAP Estimate

"The posterior is well centered around the CM but not around the MAP estimate"

\[
\hat{u}_{\text{MAP}} \in \arg\min_u \left\{ \frac{1}{2} \| f - K(u) \|_{\Sigma^{-1}}^2 + \lambda J(u) \right\}
\]

Use optimality condition

\[
K^* \Sigma^{-1} (K \hat{u}_{\text{MAP}} - f) + \lambda \hat{p}_{\text{MAP}} = 0, \quad \hat{p}_{\text{MAP}} \in \partial J(\hat{u}_{\text{MAP}}).
\]

to rewrite posterior in terms of \( \hat{u}_{\text{MAP}} \):

\[
p_{\text{post}}(u|f) \propto \exp \left( -\frac{1}{2} \| K(u - \hat{u}_{\text{MAP}}) \|_{\Sigma^{-1}}^2 - \lambda D_{J}(u, \hat{u}_{\text{MAP}}) \right)
\]

Posterior energy is sum of two convex functionals both minimized by \( \hat{u}_{\text{MAP}} \).
Average Optimality of the CM Estimate

You can show an “average optimality condition” for the CM estimate:

\[
\mathbb{E}(u|f)[K^* \Sigma^{-1}_\epsilon (Ku - f) + \lambda J'(u)] = K^* (K \Sigma^{-1}_\epsilon \mathbb{E}(u|f)[u] - f) + \lambda \mathbb{E}(u|f)[J'(u)] \\
= K^* \Sigma^{-1}_\epsilon (K \hat{u}_{CM} - f) + \lambda \hat{p}_{CM} = 0
\]

where \( \hat{p}_{CM} = \int J'(u)p_{post}(u|f)du \) is the CM estimate for the gradient of \( J \).
Average Optimality of the CM Estimate

You can show an “average optimality condition” for the CM estimate:
\[
\mathbb{E}_{(u|f)}[K^* \Sigma^{-1}_\varepsilon (Ku - f) + \lambda J'(u)] = K^* (K \Sigma^{-1}_\varepsilon \mathbb{E}_{(u|f)}[u] - f) + \lambda \mathbb{E}_{(u|f)}[J'(u)]
\]
\[
= K^* \Sigma^{-1}_\varepsilon (K \hat{u}_{CM} - f) + \lambda \hat{p}_{CM} = 0
\]
where \( \hat{p}_{CM} = \int J'(u) p_{post}(u|f) du \) is the CM estimate for the gradient of \( J \).

Compare it to optimality condition for MAP estimate:
\[
K^* \Sigma^{-1}_\varepsilon (K \hat{u}_{MAP} - f) + \lambda \hat{p}_{MAP} = 0
\]
Difference: \( J'(\mathbb{E}_{(u|f)}[u]) \neq \mathbb{E}_{(u|f)}[J'(u)] \) (except for Gaussian case).

Furthermore:
\[
\mathbb{E}_{(u|f)} \| L(\hat{u}_{CM} - u) \|_2^2 \leq \mathbb{E}_{(u|f)} \| L(\hat{u}_{MAP} - u) \|_2^2
\]
\[
\mathbb{E}_{(u|f)} D_J(\hat{u}_{MAP}, u) \leq \mathbb{E}_{(u|f)} D_J(\hat{u}_{CM}, u)
\]
Take Home Messages

▶ Sample-based Bayesian inversion with sparsity constraints is feasible in high dimensions.

▶ Computing CM estimates is NOT the only use of it.

▶ MAP estimates are proper Bayes estimates for a proper, convex cost function, and the posterior is well-centered around them.

▶ A "real Bayesian" can use them without feeling ashamed.

▶ Bregman distances are also an interesting tool to analyze Bayesian inversion.

▶ "MAP vs. CM" is NOT the most interesting question for comparing variational regularization and Bayesian inference.
Thank you for your attention!

Work was part of the Chinese-Finnish-German project "Sparsity-constrained inversion with tomographic applications" ("Inverse Problems Initiative" of the DFG).

Coordination by Samuli Siltanen (Helsinki); four teams:
- Bremen (Germany), PI: Professor Peter Maass
- Helsinki (Finland), PI: Professor Matti Lassas
- Münster (Germany), PI: Professor Martin Burger
- Shanghai (China), PI: Professor Jianguo Huang
Single Component Gibbs Sampling

Basic idea:

1. Choose component to update $s \in \{1, \ldots, n\}$ (random or systematic).
2. Update $u_s$ by sample from the cond., 1-dim density $p(\cdot | u_{[-s]})$.

To be fast one needs:

a) fast and explicit comp. of the 1-dim densities.
b) fast, robust and exact sampling from 1-dim densities.
Single Component Gibbs Sampling

Basic idea:

1. Choose component to update $s \in \{1, \ldots, n\}$ (random or systematic).

2. Update $u_s$ by sample from the cond., 1-dim density $p(\cdot | u_{[-s]})$.

To be fast one needs:

a) fast and explicit comp. of the 1-dim densities.

b) fast, robust and exact sampling from 1-dim densities.

Nasty, involved and time consuming to implement for L1-type priors
Sketch of Gibbs Sampler Implementation

\[ p_{post}(u|f) \propto \exp \left( -\frac{1}{2\sigma^2} \| f - Ku \|_2^2 - \lambda \| Wu \|_1 \right) \]

\[ p_{post}(u|f) \propto \exp \left( -\frac{1}{2\sigma^2} \| f - Kw^{-1}\xi \|_2^2 - \lambda \| \xi \|_1 \right) \]

- **K**: Radon transform of object integrated into measurement sensors.
- **W**: Haar-Wavelet transform in 2D, \( W = [v_1, \ldots, v_n]^T \)
- **\( \xi = Du \)**: Wavelet coefficients.

Fast sampling needs fast setup-up of \( Kvi \), and projection of \( Kvi \) on current residual \( (f - Kw^{-1}\xi) \):
- Haar wavelets consist of 1,2 or 4 rectangles.
- The projection of a rectangle is a symmetric trapezoid.
- Design fast scheme to integrate this into measurement grid.
- Loop over projection angles.

Felix Lucka (felix.lucka@wwu.de)
Haar Wavelets & Radon Transforms: $j = 0$, $l = 0$, $k_1 = 0$, $k_2 = 0$
Haar Wavelets & Radon Transforms: $j = 0, l = 1, 2, 3, k_1 = 0, k_2 = 0$
Haar Wavelets & Radon Transforms: $j = 1$, $l = 1, 2, 3$, $k_1 = 0$, $k_2 = 0$
Haar Wavelets & Radon Transforms: $j = 1$, $l = 1, 2, 3$, $k_1 = 0$, $k_2 = 1$
Haar Wavelets & Radon Transforms: $j = 1, l = 1, 2, 3, k_1 = 1, k_2 = 0$

(a)  
(b)  
(c)  
(d)  
(e)  
(f)
Haar Wavelets & Radon Transforms: $j = 1, \, l = 1, 2, 3, \, k_1 = 1, \, k_2 = 1$
Radon Integration Matrices

For computing MAP estimates we need a fast way to compute $K \cdot u$ and $K^* \cdot v$

Way 1: Matlab’s `radon.m`. Turn’s out to be problematic:
- `iradon.m` is not exact adjoint
- Strange offset
- Only radon transform, not integrated
- Fixed output image size.
- Differs from implementation of $K$ used in sampler.

Way 2: Use code to compute integrated radon transform of pixel basis to build $K$ as a sparse matrix.
- ✓ Fast: 3 min vs. 2h with `radon.m`
- ✓ Size: 400 MB
- ✓ Compatible with sampler implementation
- ✓ Choose offset and output size freely
- ✓ Application of $K \cdot u$ about 2.5 times faster.
- ✓ Code on my website (soon)

Felix Lucka (felix.lucka@wwu.de)
Future Work

What happens to the posterior?
- Why do MAP and CM coincide in strongly non-Gaussian situation?
- Role of $\lambda$, $\sigma^2$: Phase transition?
- Does the covariance concentrate?
- Use Wasserstein distances via embedding?

How can we make more use of the sampler?
- More elaborate inference task.
- Real data.

How to further improve the sampler?
- Single component adaptive Gibbs: Construct Markovian transition kernel from sample history.
- Rao-Blackwellization

Felix Lucka (felix.lucka@wwu.de)