

Simulation and optimisation in imaging inverse problems: Part 2.

Marcelo Pereyra

<http://www.stats.bris.ac.uk/~mp12320/>

University of Bristol

8th of July 2016, Peyresq, France.



IMAGES ARE CHALLENGING
PHYSICAL MEASUREMENTS,
NOT PICTURES!

- 1 Maximum-a-posteriori estimation with Bayesian confidence regions
- 2 Maximum-a-posteriori estimation with unknown regularisation parameters

- 1 Maximum-a-posteriori estimation with Bayesian confidence regions
 - Bayesian uncertainty quantification in imaging inverse problems
 - Approximating Bayesian confidence regions by convex optimisation
 - Proposed approximation
 - Approximation error analysis
 - Applications to tomography and microscopy
 - Tomographic image reconstruction with a total-variation prior
 - Sparse image deblurring with an ℓ_1 prior
 - Conclusion
- 2 Maximum-a-posteriori estimation with unknown regularisation parameters
 - Hierarchical maximum-a-posteriori estimation
 - Proposed Bayesian inference methods
 - Applications to image processing
 - App. 1: Compressive sensing reconstruction with ℓ_1 -wavelet analysis prior
 - App. 2: Image resolution enhancement with a total-variation prior
 - Conclusion

- 1 **Maximum-a-posteriori estimation with Bayesian confidence regions**
 - Bayesian uncertainty quantification in imaging inverse problems
 - Approximating Bayesian confidence regions by convex optimisation
 - Proposed approximation
 - Approximation error analysis
 - Applications to tomography and microscopy
 - Tomographic image reconstruction with a total-variation prior
 - Sparse image deblurring with an ℓ_1 prior
 - Conclusion
- 2 **Maximum-a-posteriori estimation with unknown regularisation parameters**
 - Hierarchical maximum-a-posteriori estimation
 - Proposed Bayesian inference methods
 - Applications to image processing
 - App. 1: Compressive sensing reconstruction with ℓ_1 -wavelet analysis prior
 - App. 2: Image resolution enhancement with a total-variation prior
 - Conclusion

Imaging inverse problems

- We are interested in an unknown image $\mathbf{x} \in \mathbb{R}^n$.
- We measure $\mathbf{y} \in \mathbb{C}^p$, related to \mathbf{x} by a statistical model $p(\mathbf{y}|\mathbf{x})$.
- The recovery of \mathbf{x} from \mathbf{y} is ill-posed or ill-conditioned, **resulting in significant uncertainty about \mathbf{x}** .
- For example, linear imaging problems of the form

$$\mathbf{y} = A\mathbf{x} + \mathbf{w},$$

for some linear operator A with $\text{rank}(A) < \dim(\mathbf{x})$.

The Bayesian framework

- We use priors to reduce uncertainty and deliver accurate results.
- Given the prior $p(\mathbf{x})$, the posterior distribution of \mathbf{x} given \mathbf{y}

$$p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})/p(\mathbf{y})$$

models our knowledge about \mathbf{x} after observing \mathbf{y} .

- In this talk we consider that $p(\mathbf{x}|\mathbf{y})$ is log-concave; i.e.,

$$p(\mathbf{x}|\mathbf{y}) = \exp\{-g_{\mathbf{y}}(\mathbf{x})\}/Z_{\mathbf{y}},$$

where $g_{\mathbf{y}}(\mathbf{x})$ is a convex function and $Z_{\mathbf{y}} = \int \exp\{-g_{\mathbf{y}}(\mathbf{x})\}d\mathbf{x}$.

Convex formulations and log-concavity

For example, imaging inverse problems of the form

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\{-\psi(\mathbf{x}) - \phi(\mathbf{x})\} \quad (1)$$

where $g_{\mathbf{y}} = \psi + \phi$ is a convex function from $\mathbb{R}^n \rightarrow (-\infty, +\infty]$. Typically

$$\psi(\mathbf{x}) = \frac{1}{2\sigma^2} \|\mathbf{y} - A\mathbf{x}\|_2^2$$

for some linear operator $A \in \mathbb{C}^{p \times n}$, and

$$\phi(\mathbf{x}) = \alpha \|B\mathbf{x}\|_{\dagger} + \mathbf{1}_{\mathcal{S}}(\mathbf{x})$$

for some norm $\|\cdot\|_{\dagger}$, dictionary $B \in \mathbb{R}^{n \times n}$, and convex set \mathcal{S} .

Maximum-a-posteriori (MAP) estimation

The predominant Bayesian approach in imaging is MAP estimation

$$\begin{aligned}\hat{\mathbf{x}}_{MAP} &= \operatorname{argmax}_{\mathbf{x} \in \mathbb{R}^n} p(\mathbf{x}|\mathbf{y}), \\ &= \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} g_{\mathbf{y}}(\mathbf{x}),\end{aligned}\tag{2}$$

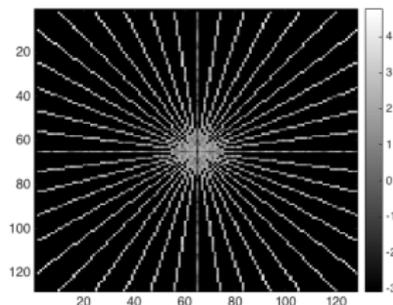
which can be **computed very efficiently by convex optimisation** (Combettes & Pesquet 2011, Parikh & Boyd 2014).

Limitations

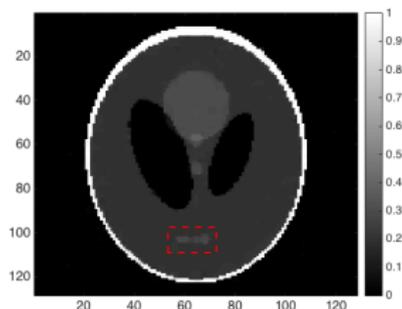
- **Raw MAP estimation fails to deliver basic elements of Bayesian paradigm ($\hat{\mathbf{x}}_{MAP}$ provides very little information about $p(\mathbf{x}|\mathbf{y})$).**
- However, more advanced analyses require other tools (e.g. MCMC) that are often very computationally expensive (Green et al. 2015).

Illustrative example

Tomographic reconstruction of the Shepp-Logan phantom image from noisy tomographic data (computing time 0.75 seconds).



Tomographic data



Bayesian MAP estimate

*Impressive results! but how **confident** are we about this result?*

Where does the posterior probability mass of \mathbf{x} lie?

- A set C_α is a posterior credible region of confidence level $(1 - \alpha)\%$ if

$$P[\mathbf{x} \in C_\alpha | \mathbf{y}] = 1 - \alpha.$$

- The *highest posterior density* (HPD) region is decision-theoretically optimal (Robert 2001)

$$C_\alpha^* = \{\mathbf{x} : g_{\mathbf{y}}(\mathbf{x}) \leq \gamma_\alpha\}$$

with $\gamma_\alpha \in \mathbb{R}$ chosen such that $\int_{C_\alpha^*} p(\mathbf{x} | \mathbf{y}) d\mathbf{x} = 1 - \alpha$ holds.

- However, computing any C_α becomes intractable as n increases.

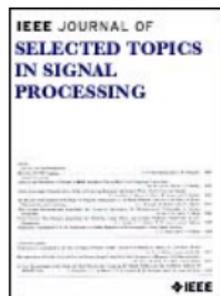
Recent surveys on Bayesian computation

State-of-the art Bayesian computation - The big picture:



25th anniversary special issue on Bayesian computation

P. Green, K. Latuszynski, M. Pereyra, C. P. Robert, "Bayesian computation: a perspective on the current state, and sampling backwards and forwards", *Statistics and Computing*, vol. 25, no. 4, pp 835-862, Jul. 2015.



Special issue on "Stochastic simulation and optimisation in signal processing"

M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tournernet, A. Hero, and S. McLaughlin, "A Survey of Stochastic Simulation and Optimization Methods in Signal Processing" *IEEE Sel. Topics in Signal Processing*, in press.

- 1 **Maximum-a-posteriori estimation with Bayesian confidence regions**
 - Bayesian uncertainty quantification in imaging inverse problems
 - **Approximating Bayesian confidence regions by convex optimisation**
 - Proposed approximation
 - Approximation error analysis
 - Applications to tomography and microscopy
 - Tomographic image reconstruction with a total-variation prior
 - Sparse image deblurring with an ℓ_1 prior
 - Conclusion
- 2 **Maximum-a-posteriori estimation with unknown regularisation parameters**
 - Hierarchical maximum-a-posteriori estimation
 - Proposed Bayesian inference methods
 - Applications to image processing
 - App. 1: Compressive sensing reconstruction with ℓ_1 -wavelet analysis prior
 - App. 2: Image resolution enhancement with a total-variation prior
 - Conclusion

Key observation: when n is large, if we use MYULA or P_X -MALA to generate a chain targeting $p(\mathbf{x}|\mathbf{y})$, all samples score similarly w.r.t. $g_{\mathbf{y}}!$.

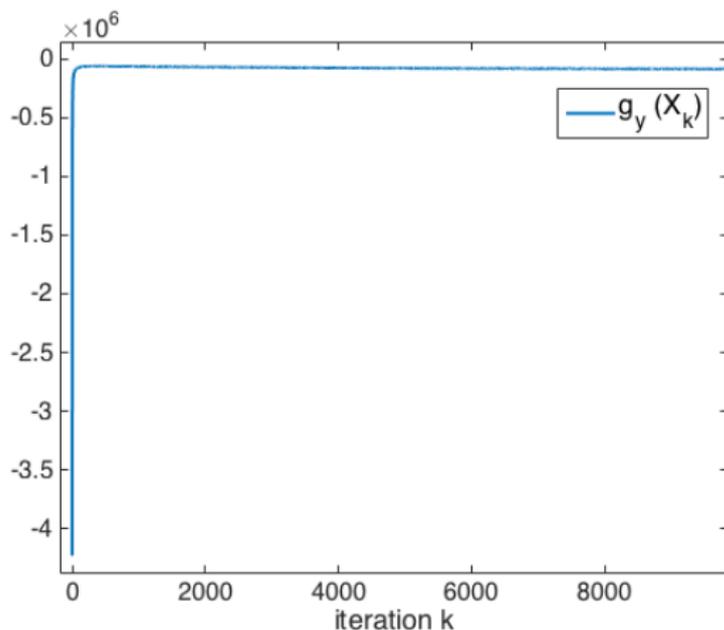


Figure : Trace of $g_{\mathbf{y}}(X_k)$ for a Markov chain $\{X_k\}_{k=1}^K$ related to a sparse regression problem of dimension $n = 10^4$.

Theorem 1.1 (Pereyra (2016))

Suppose that the posterior $p(\mathbf{x}|\mathbf{y}) = \exp\{-g_{\mathbf{y}}(\mathbf{x})\}/Z_{\mathbf{y}}$ is log-concave on \mathbb{R}^n . Then, for any $\alpha \in (4 \exp(-n/3), 1)$, the HPD region C_α^* is contained by

$$\tilde{C}_\alpha = \{\mathbf{x} : g_{\mathbf{y}}(\mathbf{x}) \leq g_{\mathbf{y}}(\hat{\mathbf{x}}_{MAP}) + \sqrt{n\tau_\alpha + n}\},$$

with positive constant $\tau_\alpha = \sqrt{16 \log(3/\alpha)}$ independent of $p(\mathbf{x}|\mathbf{y})$, and where $\hat{\mathbf{x}}_{MAP} = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} g_{\mathbf{y}}(\mathbf{x})$ is the maximum-a-posteriori estimator of \mathbf{x} .

Remark 1: \tilde{C}_α is a conservative approximation of C_α^* , i.e.,

$$\mathbf{x} \notin \tilde{C}_\alpha \implies \mathbf{x} \notin C_\alpha^*.$$

Remark 2: \tilde{C}_α is available as a by-product in any convex inverse problem that is solved by MAP estimation!

Proof sketch (Theorem 1.1)

The proof is based on the following two information theory results:

Lemma 1

Suppose that $p(\mathbf{x}|\mathbf{y}) = \exp\{-g_{\mathbf{y}}(\mathbf{x})\}/Z_{\mathbf{y}}$ is log-concave on \mathbb{R}^n , then

$$P[|g_{\mathbf{y}}(\mathbf{x}) - E\{g_{\mathbf{y}}(\mathbf{x})\}| \geq \tau n] \leq 3 \exp(-\tau^2 n/16),$$

for any $\tau \in [0, 2]$, and where $E\{g_{\mathbf{y}}(\mathbf{x})\} = \int_{\mathbb{R}^n} g_{\mathbf{y}}(\mathbf{x}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$.

Lemma 2

Suppose that $p(\mathbf{x}|\mathbf{y}) = \exp\{-g_{\mathbf{y}}(\mathbf{x})\}/Z_{\mathbf{y}}$ is log-concave on \mathbb{R}^n , then

$$g_{\mathbf{y}}(\hat{\mathbf{x}}_{MAP}) \leq E\{g_{\mathbf{y}}(\mathbf{x})\} \leq g_{\mathbf{y}}(\hat{\mathbf{x}}_{MAP}) + n,$$

where $\hat{\mathbf{x}}_{MAP} = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} g_{\mathbf{y}}(\mathbf{x})$ is the maximum-a-posteriori estimator of \mathbf{x} .

Approximation error bounds

Is \tilde{C}_α a “good” approximation of C_α^* ?

Let γ_α and $\tilde{\gamma}_\alpha = g_{\mathbf{y}}(\hat{\mathbf{x}}_{MAP}) + n(\tau_\alpha + 1)$ be the thresholds defining the HDP region $C_\alpha^* = \{\mathbf{x} : g_{\mathbf{y}}(\mathbf{x}) \leq \gamma_\alpha\}$ and the approximation $\tilde{C}_\alpha = \{\mathbf{x} : g_{\mathbf{y}}(\mathbf{x}) \leq \tilde{\gamma}_\alpha\}$.

Theorem 1.2 (Finite-dimensional error bound (Pereyra 2016))

Suppose that $p(\mathbf{x}|\mathbf{y}) = \exp\{-g_{\mathbf{y}}(\mathbf{x})\}/Z_{\mathbf{y}}$ is log-concave on \mathbb{R}^n , then

$$0 \leq \tilde{\gamma}_\alpha - \gamma_\alpha \leq \eta_\alpha \sqrt{n} + n,$$

with positive constant $\eta_\alpha = \sqrt{16 \log(3/\alpha)} + \sqrt{1/\alpha}$ independent of $p(\mathbf{x}|\mathbf{y})$.

Remark 3: \tilde{C}_α is stable (as n becomes large, the error $\tilde{\gamma}_\alpha - \gamma_\alpha \lesssim n$).

Approximation error bounds

Are the bounds of Theorem 1.2 tight?

Let $\mathbb{X} = \{x_n, n \in \mathbb{N}\}$ be discrete-time stochastic process. Assume that for each $n \in \mathbb{N}$ the random vector $\mathbf{x}^{(n)} = (x_1, \dots, x_n)$ has marginal distribution $p_n(\mathbf{x}^{(n)}) = \exp\{-\lambda \sum_{i=1}^n |x_i|^q\} / \lambda^{-n/q}$ with $q \in [1, \infty)$ and $\lambda \in \mathbb{R}^+$.

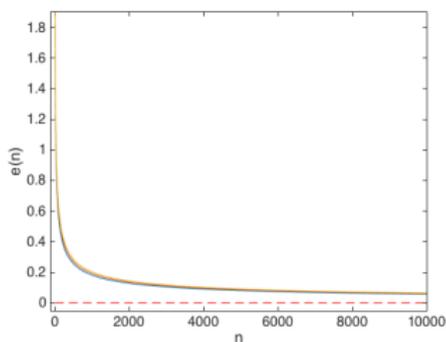
Corollary 3

For each $n \in \mathbb{N}$, let $\gamma_\alpha^{(n)}$ and $\tilde{\gamma}_\alpha^{(n)}$ denote the threshold values of the HDP region $C_\alpha^{(n)}$ and the approximation $\tilde{C}_\alpha^{(n)}$ associated with $p_n(\mathbf{x}^{(n)})$. Then,*

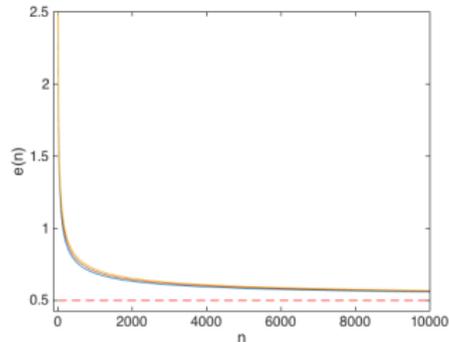
$$\lim_{n \rightarrow \infty} \frac{\tilde{\gamma}_\alpha^{(n)} - \gamma_\alpha^{(n)}}{n} = 1 - 1/q.$$

Remark 4: The lower and upper error bounds of Theorem 1.2 are attained by $q = 1$ and $q \rightarrow \infty$ (support constraint) when $n \rightarrow \infty$.

Are the asymptotics of Corollary 3 relevant?



Laplace ($q = 1$)



Gaussian ($q = 2$)

Figure : Normalised error $e(n) = \frac{\tilde{\gamma}_\alpha^{(n)} - \gamma_\alpha^{(n)}}{n}$ and asymptotics for $q = 1$ and $q = 2$, and $\alpha = 0.2, 0.1, 0.05$.

- 1 **Maximum-a-posteriori estimation with Bayesian confidence regions**
 - Bayesian uncertainty quantification in imaging inverse problems
 - Approximating Bayesian confidence regions by convex optimisation
 - Proposed approximation
 - Approximation error analysis
 - **Applications to tomography and microscopy**
 - Tomographic image reconstruction with a total-variation prior
 - Sparse image deblurring with an ℓ_1 prior
 - Conclusion
- 2 **Maximum-a-posteriori estimation with unknown regularisation parameters**
 - Hierarchical maximum-a-posteriori estimation
 - Proposed Bayesian inference methods
 - Applications to image processing
 - App. 1: Compressive sensing reconstruction with ℓ_1 -wavelet analysis prior
 - App. 2: Image resolution enhancement with a total-variation prior
 - Conclusion

Experiment 1: Tomographic reconstruction with TV prior

Recover $\mathbf{x} \in \mathbb{R}^n$ from partially observed and noisy Fourier measurements

$$\mathbf{y} = \Phi \mathcal{F} \mathbf{x} + \mathbf{w},$$

where Φ is a (tomographic) mask and \mathcal{F} is the 2D Fourier operator.

We use the Bayesian model

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\|\mathbf{y} - \Phi \mathcal{F} \mathbf{x}\|^2 / 2\sigma^2 - \lambda TV(\mathbf{x})\right), \quad (3)$$

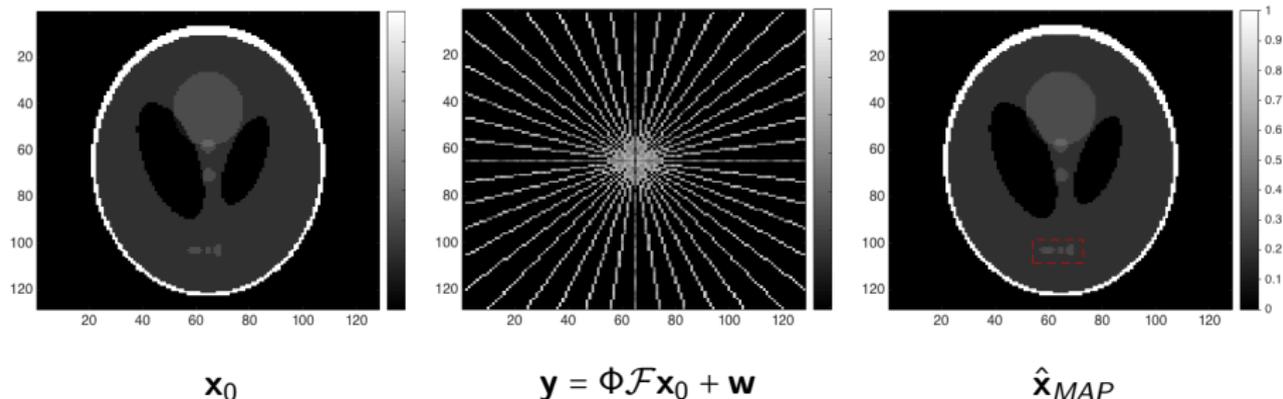
where $TV(\mathbf{x}) = \|\nabla_d \mathbf{x}\|_{1-2}$ is the total-variation norm of \mathbf{x} .

We compute the MAP estimator $\hat{\mathbf{x}}_{MAP}$ by convex optimisation.

$$\hat{\mathbf{x}}_{MAP} = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \|\mathbf{y} - \Phi \mathcal{F} \mathbf{x}\|^2 / 2\sigma^2 - \lambda TV(\mathbf{x}).$$

MRI reconstruction of the Shepp Logan phantom image

MAP estimation (Case 1: High SNR - $\sigma = 7 \times 10^{-3}$)



MRI experiment (high SNR): (a) Shepp-Logan phantom image (128×128 pixels), (b) tomographic observation \mathbf{y} (amplitude of Fourier coefficients in logarithmic scale, $\sigma = 7 \times 10^{-3}$), (c) MAP estimate $\hat{\mathbf{x}}_{MAP}$.

Suppose that the structure highlighted in red is clinically important (e.g., lesion).

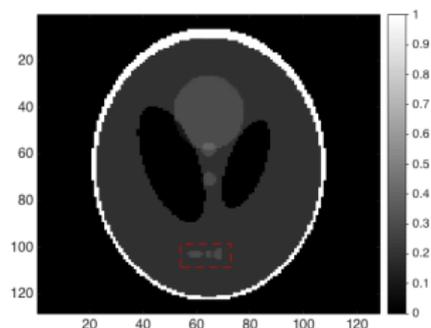
- **Are we confident about this structure** (its presence, intensity values, etc.)?
- **Idea:** use \tilde{C}_α to explore/quantify the uncertainty about this structure.

Proposed “knockout” test: double negation approach - assume that the structure is NOT present in the image and seek to REJECT the hypothesis.

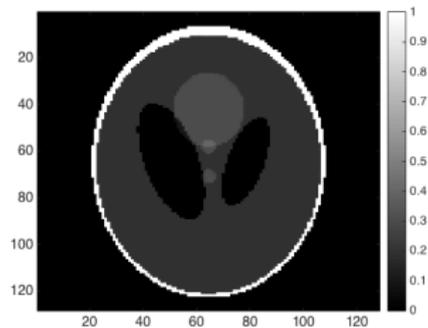
Test procedure:

- 1 Generate a surrogate test image \mathbf{x}_\dagger by modifying $\hat{\mathbf{x}}_{MAP}$ to **remove the structure of interest** (in best agreement with prior).
- 2 If $\mathbf{x}_\dagger \notin \tilde{\mathcal{C}}_\alpha$ the model rejects \mathbf{x}_\dagger with probability $(1 - \alpha)$, suggesting that the structure is present in the true image with high probability.
- 3 Otherwise, if $\mathbf{x}_\dagger \in \tilde{\mathcal{C}}_\alpha$ the posterior uncertainty about the structure is too high to draw conclusions \rightarrow increase measurements / reduce noise.

MRI experiment - Knockout test (high SNR)



$\hat{\mathbf{x}}_{MAP}$



Test image \mathbf{x}_{\dagger}

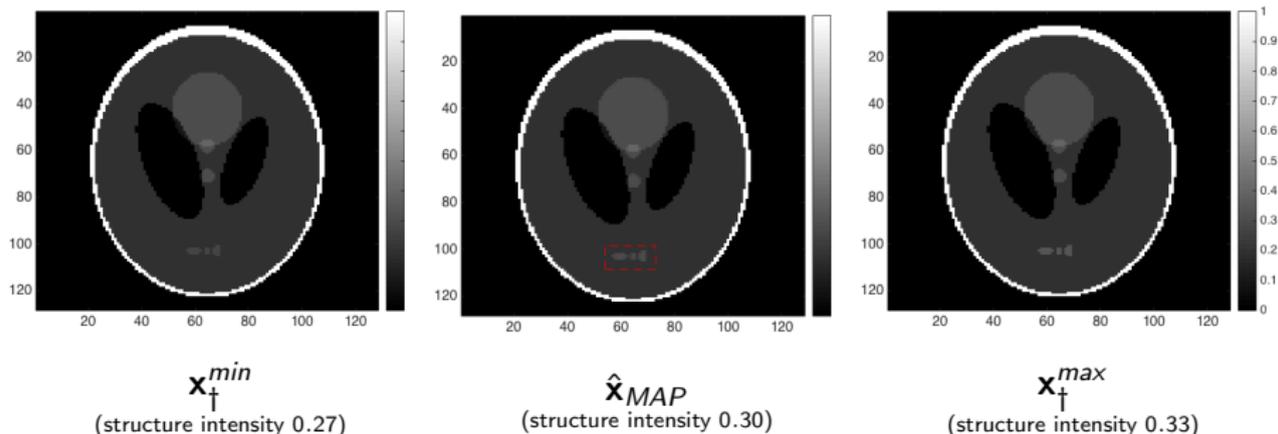
Knockout test result:

- 1 Score $g_{\mathbf{y}}(\mathbf{x}_{\dagger}) = 2.91 \times 10^5$.
- 2 The 99% threshold $\tilde{\gamma}_{0.01} = g_{\mathbf{y}}(\hat{\mathbf{x}}_{MAP}) + n(\tau_{0.01} + 1) = 1.53 \times 10^5$.
- 3 Therefore $\mathbf{x}_{\dagger} \notin \tilde{\mathcal{C}}_{\alpha}$, **rejecting the knockout hypothesis** and providing **evidence in favour** of the structure considered.

Note: computing $\tilde{\gamma}_{0.01}$ to perform this test required 75 milliseconds.

Intensity uncertainty quantification

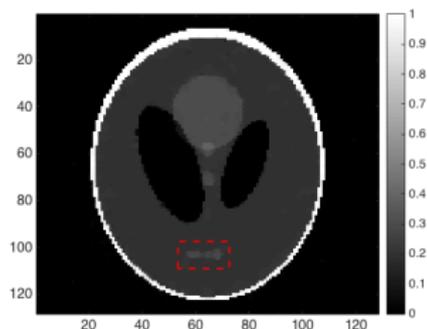
Find minimum and maximum structure intensity values within $\tilde{\mathcal{C}}_\alpha$:



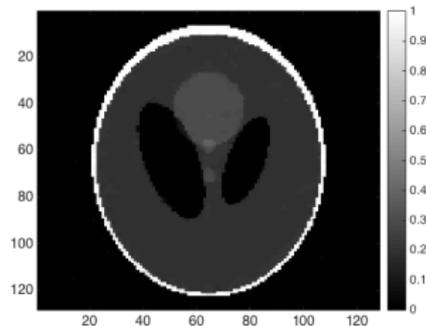
- Intensity of the structure in $\hat{\mathbf{x}}_{MAP}$ is 0.30 (surrounding intensity 0.20).
- Surrogates \mathbf{x}_\dagger^{min} and \mathbf{x}_\dagger^{max} indicate uncertainty of the order of 10%.

MRI experiment - Knockout test (low SNR)

MAP estimation (Case 2: Low SNR - $\sigma = 7 \times 10^{-2}$)



$\hat{\mathbf{x}}'_{MAP}$



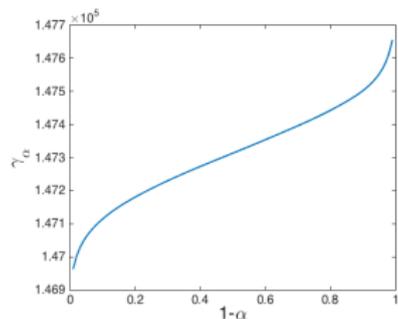
Test image \mathbf{x}'_{\dagger}

Knockout test result:

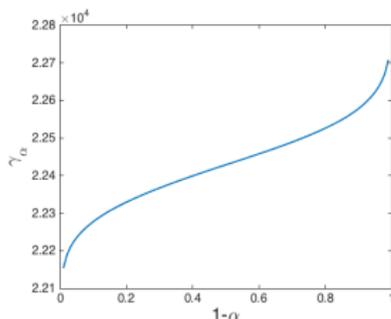
- 1 The 80% threshold $\tilde{\gamma}_{0.80} = g_y(\hat{\mathbf{x}}'_{MAP}) + n(\tau_{0.2} + 1) = 2.85 \times 10^4$.
- 2 Score $g_y(\mathbf{x}'_{\dagger}) = 1.27 \times 10^4$, therefore \mathbf{x}'_{\dagger} is a potential solution.
- 3 We conclude that, because of the lower SNR, it is not possible to assert confidently that the structure is present in the image.

MRI experiment - Approximation error analysis

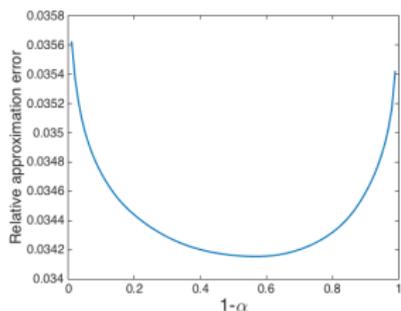
To assess the approximation error we compute the exact HPD thresholds γ_α by proximal Markov chain Monte Carlo integration (Pereyra 2015).



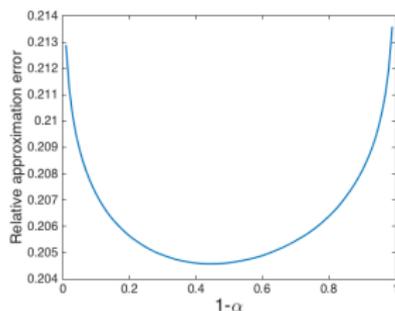
HDP threshold γ_α (high SNR)



HDP threshold γ_α (low SNR)



Rel. approx. error (high SNR)



Rel. approx. error (low SNR)

Sparse image deblurring with an ℓ_1 prior

Recover a sparse high-resolution image $\mathbf{x} \in \mathbb{R}^n$ from a blurred and noisy observation

$$\mathbf{y} = A\mathbf{x} + \mathbf{w},$$

where A is a linear blur operator and \mathbf{w} is Gaussian noise.

We use the Bayesian model

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\|\mathbf{y} - A\mathbf{x}\|^2/2\sigma^2 - \lambda\|\mathbf{x}\|_1\right). \quad (4)$$

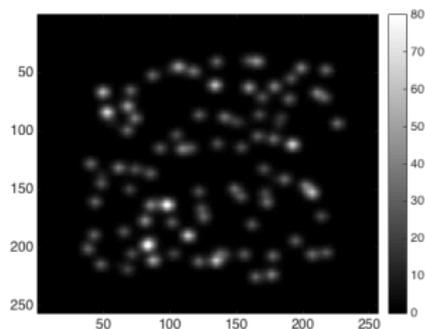
with a Laplace or ℓ_1 -norm prior for \mathbf{x} promoting soft sparsity.

We compute the MAP estimator $\hat{\mathbf{x}}_{MAP}$ by convex optimisation

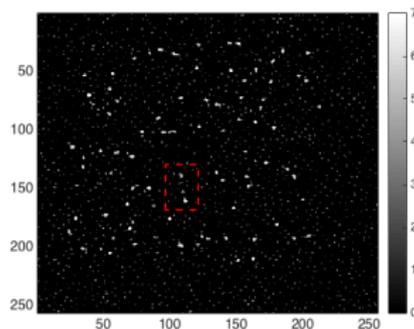
$$\hat{\mathbf{x}}_{MAP} = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \|\mathbf{y} - A\mathbf{x}\|^2/2\sigma^2 - \lambda\|\mathbf{x}\|_1.$$

Microscopy experiment

Deblurring experiment - live cell microscopy dataset (Zhu et al. 2012):



Microscopic image y



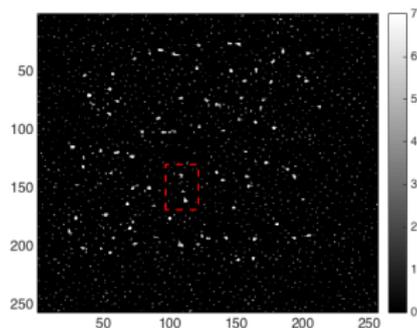
$\hat{x}MAP$ (log-scale)

Consider the molecular structure in the highlighted region:

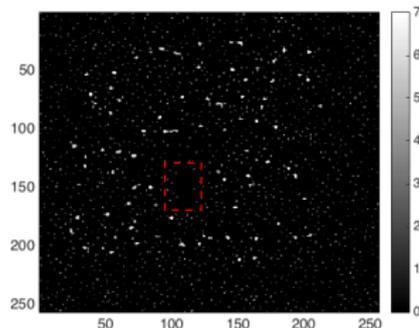
- **Are we confident about this structure** (its presence, position, etc.)?
- **Idea:** use \tilde{C}_α to explore/quantify the uncertainty about this structure.

Microscopy experiment - Knockout test

Knockout test:



$\hat{\mathbf{x}}_{MAP}$ (log-scale)



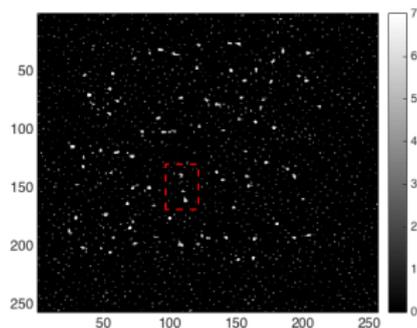
Test image \mathbf{x}_\dagger (log-scale)

- 1 Score $g_{\mathbf{y}}(\mathbf{x}_\dagger) = 1.19 \times 10^5$.
- 2 The 99% threshold $\tilde{\gamma}_{0.01} = g_{\mathbf{y}}(\hat{\mathbf{x}}_{MAP}) + n(\tau_{0.01} + 1) = 1.03 \times 10^5$.
- 3 Therefore $\mathbf{x}_\dagger \notin \tilde{\mathcal{C}}_\alpha$, **rejecting the knockout hypothesis** and providing **evidence in favour** of the structure considered.

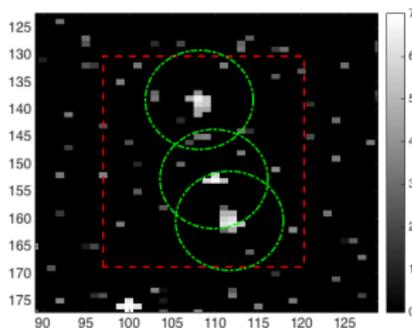
Microscopy experiment - uncertainty quantification

Position uncertainty quantification

Find maximum molecule displacement within \tilde{C}_α :



$\hat{x}MAP$
(log-scale)

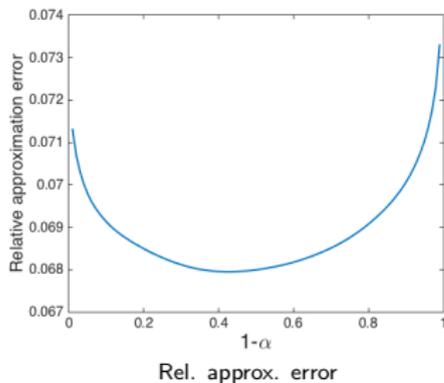
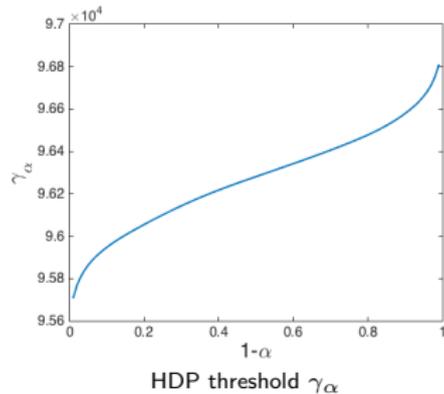


Molecule position uncertainty
($\pm 93nm \times \pm 140nm$)

Note: Uncertainty analysis ($\pm 93nm \times \pm 140nm$) in close agreement with the experimental results (average precision $80nm$) of Zhu et al. (2012).

Microscopy experiment - Approximation error analysis

To assess the approximation error we compute the exact HPD thresholds γ_α by proximal Markov chain Monte Carlo integration (Pereyra 2015).



- 1 **Maximum-a-posteriori estimation with Bayesian confidence regions**
 - Bayesian uncertainty quantification in imaging inverse problems
 - Approximating Bayesian confidence regions by convex optimisation
 - Proposed approximation
 - Approximation error analysis
 - Applications to tomography and microscopy
 - Tomographic image reconstruction with a total-variation prior
 - Sparse image deblurring with an ℓ_1 prior
 - Conclusion
- 2 **Maximum-a-posteriori estimation with unknown regularisation parameters**
 - Hierarchical maximum-a-posteriori estimation
 - Proposed Bayesian inference methods
 - Applications to image processing
 - App. 1: Compressive sensing reconstruction with ℓ_1 -wavelet analysis prior
 - App. 2: Image resolution enhancement with a total-variation prior
 - Conclusion

- New and general methodology to compute approximate HPD regions for high-dimensional inverse problems that are convex.
- Remarkable theoretical and computational properties: conservative, stable, and available as by-product of MAP estimation.
- Enables uncertainty exploration and quantification in imaging inverse problems (e.g., knockout hypothesis testing).
- Great potential for scientific imaging applications, particularly medical, biological, and remote sensing.

Beyond point estimation: **MCMC methods improve our understanding and aid discovery.**

- 1 Maximum-a-posteriori estimation with Bayesian confidence regions
 - Bayesian uncertainty quantification in imaging inverse problems
 - Approximating Bayesian confidence regions by convex optimisation
 - Proposed approximation
 - Approximation error analysis
 - Applications to tomography and microscopy
 - Tomographic image reconstruction with a total-variation prior
 - Sparse image deblurring with an ℓ_1 prior
 - Conclusion
- 2 Maximum-a-posteriori estimation with unknown regularisation parameters
 - Hierarchical maximum-a-posteriori estimation
 - Proposed Bayesian inference methods
 - Applications to image processing
 - App. 1: Compressive sensing reconstruction with ℓ_1 -wavelet analysis prior
 - App. 2: Image resolution enhancement with a total-variation prior
 - Conclusion

- 1 **Maximum-a-posteriori estimation with Bayesian confidence regions**
 - Bayesian uncertainty quantification in imaging inverse problems
 - Approximating Bayesian confidence regions by convex optimisation
 - Proposed approximation
 - Approximation error analysis
 - Applications to tomography and microscopy
 - Tomographic image reconstruction with a total-variation prior
 - Sparse image deblurring with an ℓ_1 prior
 - Conclusion
- 2 **Maximum-a-posteriori estimation with unknown regularisation parameters**
 - **Hierarchical maximum-a-posteriori estimation**
 - Proposed Bayesian inference methods
 - Applications to image processing
 - App. 1: Compressive sensing reconstruction with ℓ_1 -wavelet analysis prior
 - App. 2: Image resolution enhancement with a total-variation prior
 - Conclusion

- We are interested in an unknown image $\mathbf{x} \in \mathbb{R}^n$.
- We observe $\mathbf{y} \in \mathbb{R}^p$, related to \mathbf{x} by $p(\mathbf{y}|\mathbf{x}) = \exp\{-\ell_{\mathbf{y}}(\mathbf{x})\}$.
- The recovery of \mathbf{x} from \mathbf{y} is ill-posed or ill-conditioned.
- We address this difficulty by using a prior distribution

$$p(\mathbf{x}|\lambda) = \exp\{-\lambda h(\mathbf{x})\} / \mathcal{C}(\lambda)$$

with $h: \mathbb{R}^n \rightarrow [0, \infty]$ promoting expected properties of \mathbf{x} .

- $\lambda \in \mathbb{R}^+$ is a “regularisation” (hyper-) parameter that controls the delicate balance between observed and prior information.

Once $p(\mathbf{x}, \mathbf{y}|\lambda) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\lambda)$ is properly specified, \mathbf{x} is typically estimated by computing the MAP estimator

$$\hat{\mathbf{x}}_\lambda = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \ell_{\mathbf{y}}(\mathbf{x}) + \lambda h(\mathbf{x}) - \log C(\lambda) - \log p(\mathbf{y}), \quad (5)$$

which we assume computationally tractable and unique for a given λ .

We consider the infamous problem of (not) specifying λ .

Bayesian treatment of unknown λ

- The Bayesian framework allows estimating \mathbf{x} without specifying λ .
- We incorporate λ to the model by assigning it a gamma hyper-prior

$$p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp\{-\beta\lambda\} \mathbf{1}_{\mathbb{R}^+}(\lambda),$$

with fixed parameters α and β .

- The extended model is

$$p(\mathbf{x}, \lambda | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x}) p(\mathbf{x} | \lambda) p(\lambda)}{p(\mathbf{y})} \propto \frac{\exp\{-\ell_{\mathbf{y}}(\mathbf{x}) - \lambda h(\mathbf{x}) - \log p(\lambda)\}}{C(\lambda)}$$

but $C(\lambda) = \int_{\mathbb{R}^n} \exp\{-\lambda h(\mathbf{x})\} d\mathbf{x}$ is typically intractable!

- If we had access to $C(\lambda)$ we could either **estimate \mathbf{x} and λ jointly**, or alternatively **marginalise λ followed by inference on \mathbf{x}** .

Idea: Use MYULA or Px-MALA to estimate $E[h(\mathbf{x})|\lambda]$ over a λ -grid, and then approximate $\log C(\lambda)$ via the identity $\frac{d}{d\lambda} \log C(\lambda) = E[h(\mathbf{x})|\lambda]$.

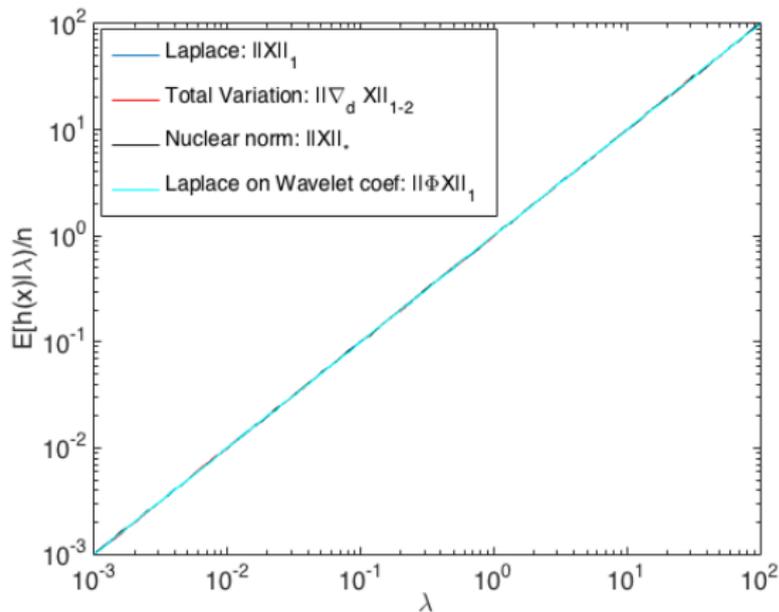


Figure : Monte Carlo approximations of $E[h(\mathbf{x})|\lambda]$ for 4 widely used prior distributions and for $\lambda \in [10^{-3}, 10^2]$. **Surprise:** they all coincide!

- 1 **Maximum-a-posteriori estimation with Bayesian confidence regions**
 - Bayesian uncertainty quantification in imaging inverse problems
 - Approximating Bayesian confidence regions by convex optimisation
 - Proposed approximation
 - Approximation error analysis
 - Applications to tomography and microscopy
 - Tomographic image reconstruction with a total-variation prior
 - Sparse image deblurring with an ℓ_1 prior
 - Conclusion
- 2 **Maximum-a-posteriori estimation with unknown regularisation parameters**
 - Hierarchical maximum-a-posteriori estimation
 - **Proposed Bayesian inference methods**
 - Applications to image processing
 - App. 1: Compressive sensing reconstruction with ℓ_1 -wavelet analysis prior
 - App. 2: Image resolution enhancement with a total-variation prior
 - Conclusion

Definition 2.1

k-homogeneity

The regulariser h is a k -homogeneous function if $\exists k \in \mathbb{R}^+$ such that

$$h(\eta \mathbf{x}) = \eta^k h(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n, \forall \eta > 0. \quad (6)$$

Note: Property (6) holds for most models used in modern image processing. In particular, all norms (e.g., ℓ_1 , ℓ_2 , total-variation, nuclear, etc.), composite norms (e.g., $\ell_1 - \ell_2$), and compositions of norms with linear operators (e.g., analysis terms of the form $\|\Psi \mathbf{x}\|_1$) are homogenous.

A central contribution of this talk is to show that Pereyra et al. (2015):

Proposition 2.1

Suppose that h , the sufficient statistic of $p(\mathbf{x}|\lambda)$, is k -homogenous. Then the normalisation factor has the form

$$C(\lambda) = D\lambda^{-n/k},$$

with (generally intractable) constant $D = C(1)$ independent of λ .

The proof follows straightforwardly by using the change of variables $\mathbf{u} = \lambda^{1/k}\mathbf{x}$ and (6) to express $C(\lambda)$ as a product of a function of λ and the generally intractable constant $D = \int_{\mathbb{R}^n} \exp\{-h(\mathbf{u})\}d\mathbf{u}$.

Joint MAP estimation:

$$\hat{\mathbf{x}}^*, \lambda^* = \operatorname{argmax}_{\mathbf{x}, \lambda} \log p(\mathbf{x}, \lambda | \mathbf{y}),$$

Then $\mathbf{0}_{n+1} \in \partial_{\mathbf{x}, \lambda} \log p(\hat{\mathbf{x}}^*, \lambda^* | \mathbf{y})$ which implies that

$$\hat{\mathbf{x}}^* = \hat{\mathbf{x}}_{\lambda^*} = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \ell_{\mathbf{y}}(\mathbf{x}) + \lambda^* h(\mathbf{x}),$$

and, together with Proposition 2.1, that

$$\lambda^* = \frac{n/k + \alpha - 1}{h(\hat{\mathbf{x}}_{\lambda^*}) + \beta}. \quad (7)$$

Joint maximum-a-posteriori estimation

The values λ^* can be identified by one-dimensional root-finding, and are guaranteed to exist because $t(\lambda) = h(\hat{\mathbf{x}}_\lambda)$ is non-increasing.

In all our experiments $p(\mathbf{x}, \lambda | \mathbf{y})$ is unimodal and λ^* is unique, and can be computed by alternating maximisation of $\log p(\mathbf{x}, \lambda | \mathbf{y})$

$$\begin{aligned}\mathbf{x}^{(t)} &= \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \ell_{\mathbf{y}}(\mathbf{x}) + \lambda^{(t-1)} h(\mathbf{x}), \\ \lambda^{(t)} &= \frac{n/k + \alpha - 1}{h(\mathbf{x}^{(t)}) + \beta},\end{aligned}\tag{8}$$

which in our experiments converged within 5 to 10 iterations.

The theoretical conditions for uniqueness are currently under investigation.

Marginal MAP estimation:

$$\begin{aligned}\hat{\mathbf{x}}^\dagger &= \operatorname{argmax}_{\mathbf{x} \in \mathbb{R}^n} \int_0^\infty p(\mathbf{x}, \lambda | \mathbf{y}) d\lambda, \\ &= \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \ell_{\mathbf{y}}(\mathbf{x}) + (n/k + \alpha) \log\{h(\mathbf{x}) + \beta\},\end{aligned}\tag{9}$$

which incorporates the uncertainty about λ in the inferences.

We compute $\hat{\mathbf{x}}^\dagger$ by *majorisation-minimisation* with the convex majorant

$$\ell_{\mathbf{y}}(\mathbf{x}) + (\alpha + n/k)q(\mathbf{x}|\mathbf{x}^{(t)}) \geq \ell_{\mathbf{y}}(\mathbf{x}) + (n/k + \alpha) \log\{h(\mathbf{x}) + \beta\},$$

with

$$q(\mathbf{x}|\mathbf{x}^{(t)}) \triangleq \log\{h(\mathbf{x}^{(t)}) + \beta\} + \frac{h(\mathbf{x}) - h(\mathbf{x}^{(t)})}{h(\mathbf{x}^{(t)}) + \beta} \geq \log\{h(\mathbf{x}) + \beta\}.$$

Marginal maximum-a-posteriori estimation

The resulting iterative scheme is

$$\begin{aligned}\mathbf{x}^{(t)} &= \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \ell_{\mathbf{y}}(\mathbf{x}) + \lambda^{(t-1)} h(\mathbf{x}), \\ \lambda^{(t)} &= \frac{n/k + \alpha}{h(\mathbf{x}^{(t)}) + \beta}.\end{aligned}\tag{10}$$

which is also an *expectation-maximisation* algorithm. Note that

$$\hat{\mathbf{x}}^\dagger = \hat{\mathbf{x}}_{\lambda^\dagger} = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \ell_{\mathbf{y}}(\mathbf{x}) + \lambda^\dagger h(\mathbf{x}), \quad \lambda^\dagger = (n/k + \alpha) / (h(\mathbf{x}^\dagger) + \beta).$$

Because $n/k \gg 1$ we can expect $\hat{\mathbf{x}}^*$ and $\hat{\mathbf{x}}^\dagger$ to be practically equivalent.

Again, the values λ^\dagger are guaranteed to exist and can be identified by one-dimensional root-finding. In all our experiments λ^\dagger is unique.

- 1 **Maximum-a-posteriori estimation with Bayesian confidence regions**
 - Bayesian uncertainty quantification in imaging inverse problems
 - Approximating Bayesian confidence regions by convex optimisation
 - Proposed approximation
 - Approximation error analysis
 - Applications to tomography and microscopy
 - Tomographic image reconstruction with a total-variation prior
 - Sparse image deblurring with an ℓ_1 prior
 - Conclusion
- 2 **Maximum-a-posteriori estimation with unknown regularisation parameters**
 - Hierarchical maximum-a-posteriori estimation
 - Proposed Bayesian inference methods
 - **Applications to image processing**
 - App. 1: Compressive sensing reconstruction with ℓ_1 -wavelet analysis prior
 - App. 2: Image resolution enhancement with a total-variation prior
 - Conclusion

Application 1: CS with ℓ_1 -wavelet analysis prior

Recover an original image $\mathbf{x} \in \mathbb{R}^n$ of size $n = 512 \times 512$ from a compressed and noisy measurement

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{w},$$

of size $p = n/2$, where $\Phi \in \mathbb{R}^{p \times n}$ is a compressive sensing random matrix and $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_p)$ is Gaussian noise with $\sigma^2 = 10$.

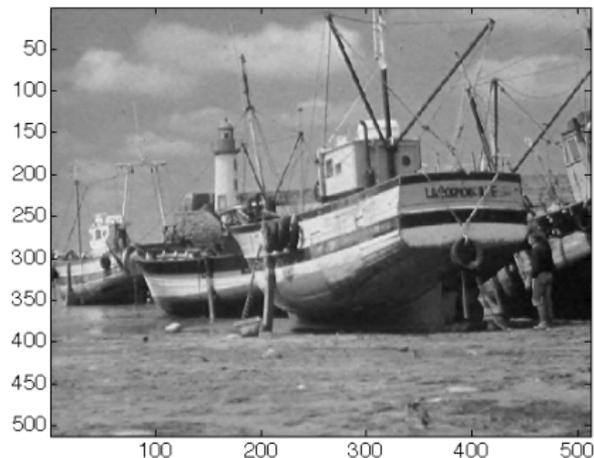
We use the *analysis* prior

$$p(\mathbf{x}|\lambda) = \exp\{-\lambda \|\Psi \mathbf{x}\|_1\} / C(\lambda)$$

where Ψ is a *Daubechies 4* wavelet frame.

Note: $\|\Psi(\mathbf{x})\|_1$ is k -homogenous with $k = 1$.

Experiment 1: Boat



Joint MAP \mathbf{x}^*

($\lambda^* = 56.4$, PSNR=33.4)



Marg. MAP \mathbf{x}^\dagger

($\lambda^\dagger = 56.4$, PSNR=33.4)

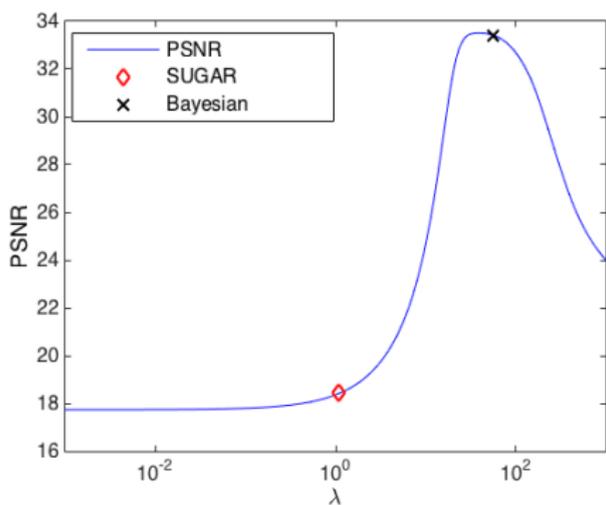
Figure : Compressive sensing experiment with the Boat image. [Left:] Bayesian joint MAP estimate (8). [Right:] Bayesian marginal MAP estimate (10).

We compare the Bayesian methods (8) and (10) with the SURE-type technique SUGAR Deledalle et al. (2014) and with the MSE oracle.

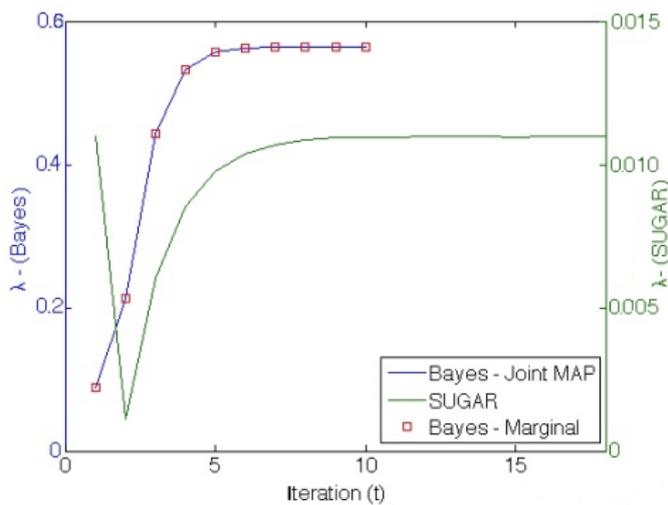
Experiment 1: Boat

Table : Values of λ , estimation accuracy (PSNR and SSIM), and computing times for the Boat experiment.

	λ	PSNR	SSIM	time [sec]
Joint MAP (8)	56.4	33.4	0.96	299
Marginal MAP (10)	56.4	33.4	0.96	299
SUGAR	1.10	18.4	0.55	1137
MSE Oracle	38.2	33.5	0.96	n/a
Least-squares	n/a	17.7	0.52	0.04



PSNR vs λ



Iterates $\lambda^{(t)}$

Figure : Compressive sensing experiment with the Boat image. [Left] Estimation PSNR as a function of λ . [Right] Evolution of the iterates $\lambda^{(t)}$ for the proposed Bayesian methods (8) and (10) (left axis) and for SUGAR (right axis).

Experiment 2: Mandrill

We compare the Bayesian methods (8) and (10) with the SURE-type technique SUGAR and with the MSE oracle.

Table : Values of λ , estimation accuracy (PSNR and SSIM), and computing times for the Mandrill experiment.

	λ	PSNR	SSIM	time [sec]
Joint MAP (8)	2.04	25.3	0.87	229
Marginal MAP (10)	2.04	25.3	0.87	229
SUGAR	0.95	22.9	0.80	984
MSE Oracle	4.65	26.0	0.90	n/a
Least-squares	n/a	18.6	0.22	0.04

Application 2: image deblurring with a total-variation prior

Recover an original image $\mathbf{x} \in \mathbb{R}^n$ from a blurred and noisy observation

$$\mathbf{y} = H\mathbf{x} + \mathbf{w},$$

where H is a 9×9 blur operator and \mathbf{w} is Gaussian noise (BSNR = 40dB).

Many image processing methods use the convex model

$$\pi(\mathbf{x}|\mathbf{y}, \lambda) \propto \exp\left(-\|\mathbf{y} - H\mathbf{x}\|^2/2\sigma^2 - \lambda TV(\mathbf{x})\right), \quad (11)$$

where $TV(\mathbf{x}) = \|\nabla_d \mathbf{x}\|_{1-2}$ is the total-variation pseudo-norm.

Note: $TV(\mathbf{x})$ is k -homogenous with $k = 1$!



Cameraman



Boat



House



Man

Figure : Deblurring experiment using the proposed Bayesian method (10).

Table : Values of λ , PSNR and computing times [secs] for the Cameraman and Boat experiments.

	Cameraman			Boat		
	λ	PSNR	time	λ	PSNR	time
Joint MAP (8)	0.04	26.6	261	0.02	30.1	1118
Marg. MAP (10)	0.04	26.6	261	0.02	30.1	1118
SUGAR	0.01	26.5	1120	0.004	30.0	4790
MSE Oracle	0.03	26.6	37	0.02	30.1	160
Bayesian Oracle	0.02	26.6	37	0.01	30.1	160
Least-squares	n/a	23.0	0.02	n/a	25.8	0.02

Table : Values of λ , PSNR and computing times [secs] for the House and Man experiments.

	House			Man		
	λ	PSNR	time	λ	PSNR	time
Joint MAP (8)	0.03	33.6	221	0.03	30.2	1136
Marg. MAP (10)	0.03	33.6	221	0.03	30.2	1136
SUGAR	0.009	33.0	221	0.005	30.1	4870
MSE Oracle	0.03	33.6	37	0.015	30.2	162
Bayesian Oracle	0.02	33.5	37	0.016	30.1	162
Least-squares	n/a	27.5	0.02	n/a	26.9	0.04

- 1 **Maximum-a-posteriori estimation with Bayesian confidence regions**
 - Bayesian uncertainty quantification in imaging inverse problems
 - Approximating Bayesian confidence regions by convex optimisation
 - Proposed approximation
 - Approximation error analysis
 - Applications to tomography and microscopy
 - Tomographic image reconstruction with a total-variation prior
 - Sparse image deblurring with an ℓ_1 prior
 - Conclusion
- 2 **Maximum-a-posteriori estimation with unknown regularisation parameters**
 - Hierarchical maximum-a-posteriori estimation
 - Proposed Bayesian inference methods
 - Applications to image processing
 - App. 1: Compressive sensing reconstruction with ℓ_1 -wavelet analysis prior
 - App. 2: Image resolution enhancement with a total-variation prior
 - Conclusion

Conclusions

- We proposed two new hierarchical Bayesian methods for MAP inference with unknown regularisation parameters.
- When $p(\mathbf{x}|\lambda) = \exp\{-\lambda h(\mathbf{x})\}/C(\lambda)$ with h k -homogenous, then

$$C(\lambda) = D\lambda^{-n/k}.$$

- λ is estimated by one-dimensional root-finding, or by iterative optimisation (convergence properties under investigation).
- Promising performance on image compressive-sensing and deblurring with analysis and total-variation priors.

Beyond point estimation: **MCMC methods improve our understanding and aid discovery.**

Thank you!

Bibliography I

- Combettes, P. L. & Pesquet, J.-C. (2011), Proximal splitting methods in signal processing, *in* H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke & H. Wolkowicz, eds, 'Fixed-Point Algorithms for Inverse Problems in Science and Engineering', Springer New-York, pp. 185–212.
- Deledalle, C., Vaiter, S., Peyré, G. & Fadili, J. (2014), 'Stein Unbiased Gradient estimator of the Risk (SUGAR) for multiple parameter selection', *SIAM J. Imaging Sci.* **7**(4), 2448–2487.
- Green, P. J., Łatuszyński, K., Pereyra, M. & Robert, C. P. (2015), 'Bayesian computation: a summary of the current state, and samples backwards and forwards', *Statistics and Computing* **25**(4), 835–862.
- Parikh, N. & Boyd, S. (2014), 'Proximal algorithms', *Found. Trends Optim.* **1**(3), 127–239.
URL: <http://dx.doi.org/10.1561/24000000003>
- Pereyra, M. (2015), 'Proximal Markov chain Monte Carlo algorithms', *Statistics and Computing* . open access paper, <http://dx.doi.org/10.1007/s11222-015-9567-4>.
- Pereyra, M. (2016), 'Maximum-a-posteriori estimation with Bayesian confidence regions', *SIAM J. Imaging Sci.* . submitted.
URL: <http://arxiv.org/abs/1602.08590>

- Pereyra, M., Bioucas-Dias, J. M. & Figueiredo, M. A. T. (2015), Maximum-a-posteriori estimation with unknown regularisation parameters, *in* 'Proc. European Signal Proc. Conf. (EUSIPCO), Nice, France, Sep. 2015.', pp. 230–234.
- Robert, C. P. (2001), *The Bayesian Choice (second edition)*, Springer Verlag, New-York.
- Zhu, L., Zhang, W., Elnatan, D. & Huang, B. (2012), 'Faster STORM using compressed sensing', *Nat. Meth.* **9**(7), 721–723.