

Proximal algorithms with mysterious “plug-and-play” operators.

Dr. Marcelo Pereyra

<http://www.macs.hw.ac.uk/~mp71/>

Maxwell Institute for Mathematical Sciences, Heriot-Watt University

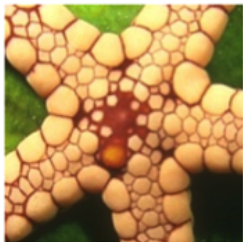
March 2019, New York.



Outline

- 1 Imaging inverse problems
- 2 Bayesian statistical modelling

Forward imaging problem



True scene

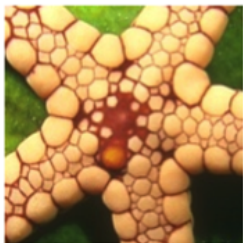


Imaging device



Observed image

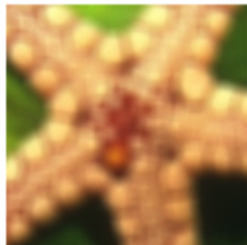
Inverse imaging problem



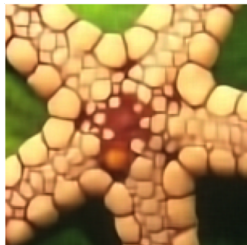
True scene



Imaging device



Observed image



Restored image

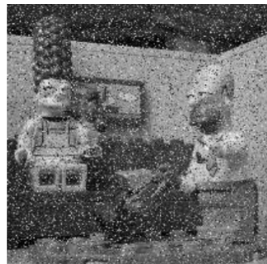
Short-exposure imaging



True scene



Imaging device



Observed data

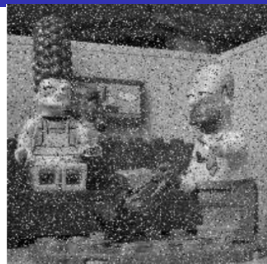
Short-exposure imaging



True scene



Imaging device



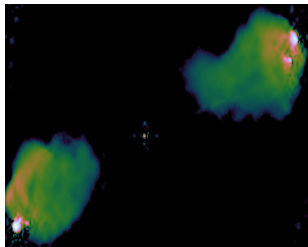
Observed data



Restored image

(J. Delon and A. Desolneux (2013))

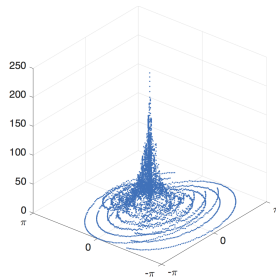
Radio-astronomy imaging



True scene

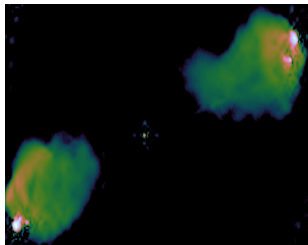


Imaging device



Observed data

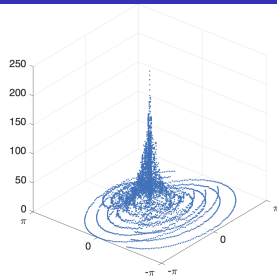
Radio-astronomy imaging



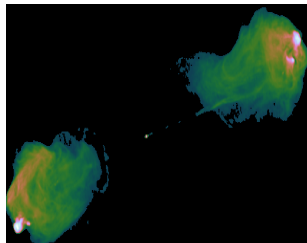
True scene



Imaging device



Observed data



Restored image
(X. Cai et al. (2018))

Outline

- 1 Imaging inverse problems
- 2 Bayesian statistical modelling

Imaging inverse problems

- We are interested in an unknown image $x \in \mathbb{R}^d$.
- We measure y , related to x by some mathematical model.
- For example, in many imaging problems

$$y = Ax + w,$$

for some operator A that is poorly conditioned or rank deficient, and an unknown perturbation or “noise” w .

- The recovery of x from y is often ill-posed or ill-conditioned, so we regularise the problem to make it well posed.

The Bayesian framework

Bayesian statistics is a mathematical framework for deriving inferences about x , from some observed data y and prior knowledge available.

Adopting a **subjective probability** approach, we represent x as a random quantity and use probability distributions to model expected properties.

To derive inferences about x from y we postulate a joint statistical model $p(x, y)$; typically specified via the decomposition $p(x, y) = p(y|x)p(x)$.

The Bayesian framework

The decomposition $p(x, y) = p(y|x)p(x)$ has two key ingredients:

The **likelihood** function: the conditional distribution $p(y|x)$ that models the data observation process (forward model).

The **prior** function: the marginal distribution $p(x) = \int p(x, y)dx$ that models our knowledge about x “before observing y ”.

For example, for $y = Ax + w$, with $w \sim \mathcal{N}(0, \sigma^2 \mathbb{I})$, we have

$$y \sim \mathcal{N}(Ax, \sigma^2 \mathbb{I}),$$

or equivalently

$$p(y|x) \propto \exp\{-\|y - Ax\|_2^2 / 2\sigma^2\}.$$

Posterior distribution

We base our inferences on the **posterior** distribution $p(x|y)$.

We derive $p(x|y)$ from the likelihood $p(y|x)$ and the prior $p(x)$ by using

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

where $p(y) = \int p(y|x)p(x)dx$ measures model-fit-to-data.

The conditional $p(x|y)$ models our knowledge about x after observing y .

Maximum-a-posteriori (MAP) estimation

The predominant Bayesian approach in imaging to derive a “solution” from $p(x|y)$ is MAP estimation

$$\begin{aligned}\hat{x}_{MAP} &= \operatorname{argmax}_{x \in \mathbb{R}^d} p(x|y), \\ &= \operatorname{argmin}_{x \in \mathbb{R}^d} -\log p(y|x) - \log p(x) + \log p(y).\end{aligned}\tag{1}$$

When $p(x|y)$ is log-concave, then $\hat{x}_{MAP}$ is a convex optimisation problem and can be efficiently solved (Chambolle and Pock, 2016).

Let $f(x) = -\log p(y|x)$ and $g(x) = -\log p(x)$. Often $f, g \in \Gamma_0(\mathbb{R})$, f is L_f -Lipschitz differentiable, and g has a computable proximal operator.

MAP estimation by proximal optimisation

For example, we could use a **proximal gradient** iteration

$$x^{m+1} = \text{prox}_g^{L_f^{-1}} \{x^m + L_f^{-1} \nabla f(x^m)\},$$

converges to $\hat{x}_{MAP}$ at rate $O(1/m)$, with poss. acceleration to $O(1/m^2)$.

Alternatively, we could **draw samples from $p(x|y)$** with a **proximal Markov chain Monte Carlo** algorithm, e.g.,

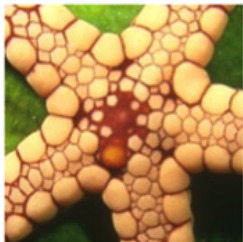
$$X^{m+1} = \text{prox}_g^\delta \{X^m + \delta \nabla f(X^m)\} + \sqrt{2\delta} Z^{m+1},$$

with $0 < \delta < L_f^{-1}$ and $Z^{m+1} \sim \mathcal{N}(0, I_d)$.

Definition The proximal operator of g is defined as (Moreau, 1962)

$$\text{prox}_g^\lambda(x) \triangleq \underset{u \in \mathbb{R}^N}{\text{argmin}} \ g(u) + \frac{1}{2\lambda} \|u - x\|^2.$$

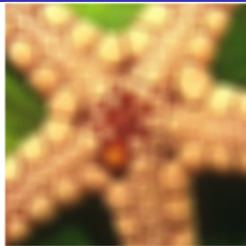
Image super-resolution



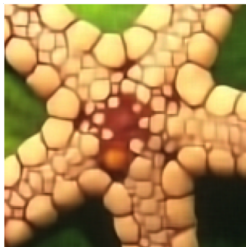
True scene



Imaging device



Observed image



Restored image

(Y. Romano, M. Elad, P. Milanfar (2017))

Bibliography:

- Chambolle, A. and Pock, T. (2016). An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319.
- Moreau, J.-J. (1962). Fonctions convexes duales et points proximaux dans un espace Hilbertien. *C. R. Acad. Sci. Paris Sér. A Math.*, 255:2897–2899.
- Romano, Yaniv and Elad, Michael and Milanfar, Peyman, The Little Engine that Could: Regularization by Denoising (RED), *SIAM J. Imaging Sciences*, Vol. 10, No. 4, pp. 1804–1844, 2017.