Bilevel learning with applications to accelerated MRI

Ferdia Sherry
with Martin Benning, Matthias J. Ehrhardt, Martin Graves, Georg Maierhofer, Juan Carlos De los Reyes, Carola-Bibiane Schönlieb and Guy Williams

fs436@cam.ac.uk

SIAM UQ18
April 18, 2018

Learning parts of a variational model

In practice, F is often based on a training set of clean images and corresponding measurements $\{u_i^*, y_i\}_{i=1}^n$:

$$F(u_1, \ldots, u_n, p) = \frac{1}{n}\sum_{i=1}^{n} \ell(u_i, u_i^*) + g(p).$$

Bilevel optimisation problems are hard to solve: regardless of whether the objective function is convex, the problem is usually nonconvex as the constraint set is nonconvex.

Some previous work on bilevel learning

De los Reyes, Schönlieb, and Valkonen 2017: Learning optimal denoising parameters $p = (\alpha, \beta)$ for higher order regularisation methods, with theoretical results in the function space setting.

(a) Clean image (b) Noisy image (c) Denoised image

A different but related approach

Rather than look at the reduced problem, we could look at the following problem (Domke 2012; Ochs et al. 2016)

$$\min_{p} F(u^{\mathrm{alg}}(p), p).$$

Tools for automatic differentiation such as Tensorflow allow for easy computation of derivatives of u^{alg}.

- Hammernik et al. 2017: Variational networks, inspired by the fields of experts model, for accelerated MRI reconstruction,
- Adler and Öktem 2017: learned gradient descent and learned primal-dual algorithms for inverse problems.

Magnetic Resonance Imaging

(a) An MRI machine in operation

(b) An image of a knee acquired by MRI

We model the (fully sampled) measurements y taken by the MRI machine as

$$y = \mathscr{F} u^* + \varepsilon$$

and reconstruct the image from measurements $\operatorname{diag}(S)y$ by minimising a TV-regularised least squares functional:

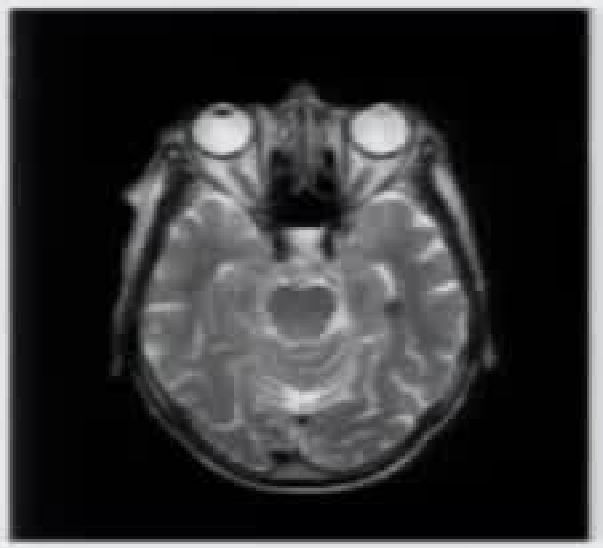
$$\hat{u} = \operatorname*{argmin}_{u \geqslant 0} E_{\mathrm{TV}}(u, y, S, \alpha) := \frac{1}{2} \| \operatorname{diag}(S)(\mathscr{F} u - y)\|_2^2 + \alpha \|\nabla u\|_1.$$

Total variation regularisation

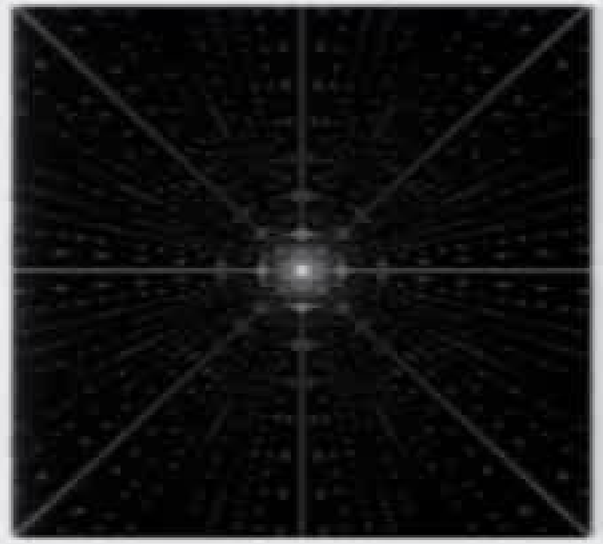
Total variation regularisation (Rudin, Osher, and Fatemi 1992) gives a smoothing effect, while preserving edges.

The effect of the sampling pattern

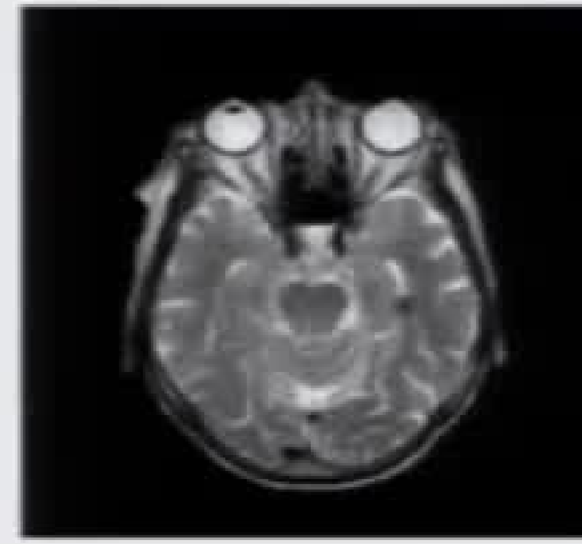
Ground truth



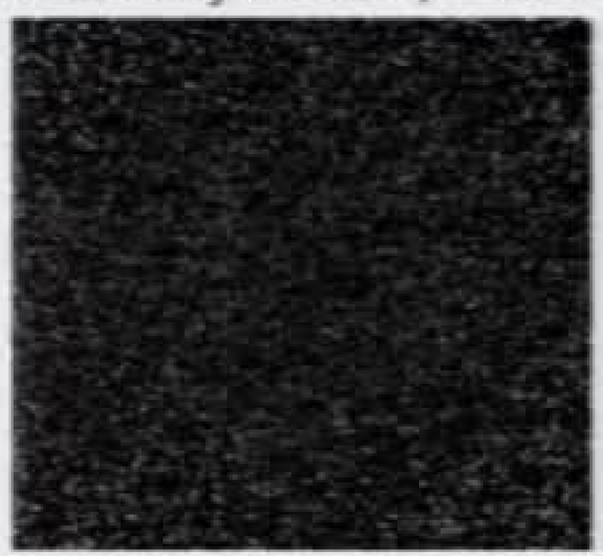
Radial lines, ~3%



TV Reconstruction



Uniformly random, ~10%



TV Reconstruction

(Krahmer and Ward 2014; Poon 2016)

Solving the bilevel problem

We can now put together the parts that we saw before to get an algorithm to learn a sampling pattern:

- The LBFGS-B algorithm can be applied since the problem is smooth with box constraints.
- To compute the required gradients, we solve a large linear system, so we use an iterative solver such as GMRES.
- The objective function splits as sum over the training set, so computations can easily be parallellised or randomised.

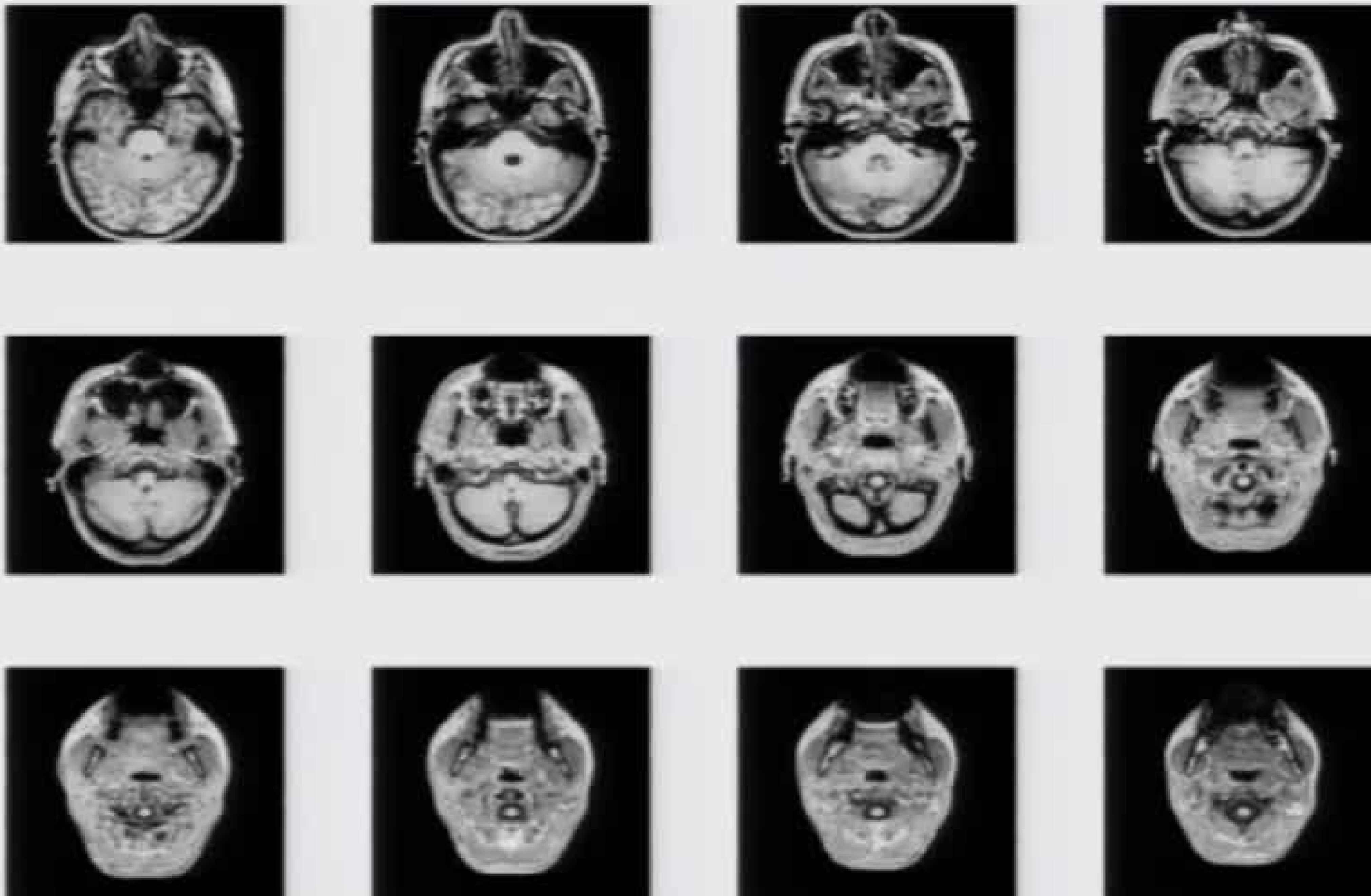
Some preliminary results

A training set consisting of a single image of a square with resolution 64^2.

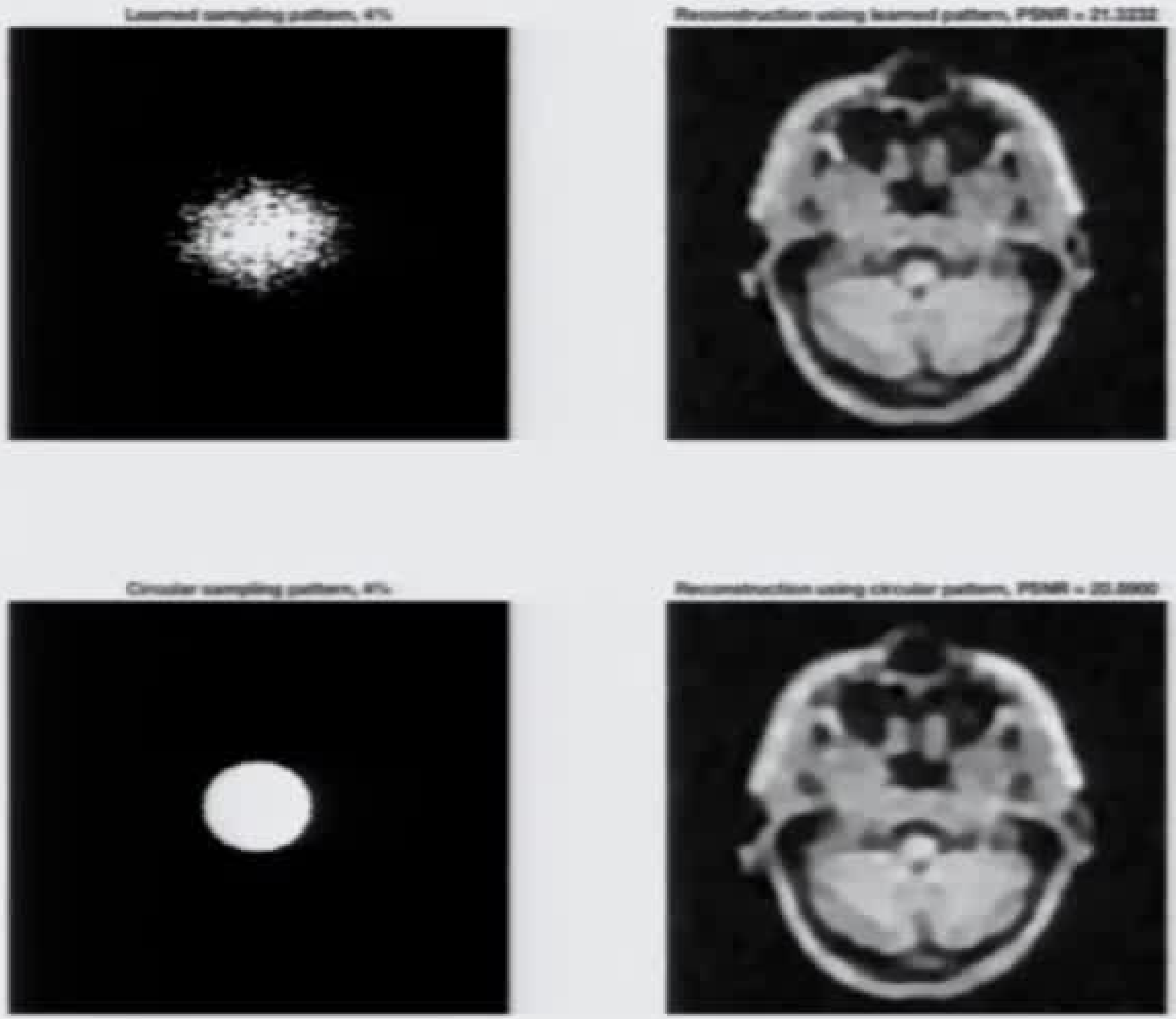
(a) Clean image (b) Learned pattern (c) Reconstruction

Some preliminary results

We use a training set of size 12, consisting of brain MRI images of resolution 192^2 taken at Addenbrooke's Hospital in Cambridge



Some preliminary results



Better comparisons should be possible with less sparse sampling patterns.

Summary and future work

Bilevel optimisation problems pop up naturally when learning parts of a model. We have proposed a bilevel optimisation problem to learn a sampling pattern for MRI and an approach to try to solve it. We still need to

- Compare to other methods for choosing sampling patterns,
- Study the effect of different regularisations in lower level problem,
- Study sensitivity of the learned sampling pattern to changes in the training set,
- Investigate the use of stochastic optimisation methods to scale up to large training sets,
- Investigate ways to incorporate physical constraints imposed by the MRI machine.