STORM imaging

Gustavo Lacerda
Department of Statistics
Columbia University
gusl@stat.columbia.edu

Abstract

We implement a spatial image model based on Ziniel and Schniter's "Dynamic Compressive Sensing of Time-Varying Signals via Approximate Message Passing", in which the image is filtered through an unknown series of binary masks, blurred, and finally observed as Poisson counts at each pixel. We initialize L1-regularized Poisson regression, with (a) an interior-point method, and (b) the FISTA algorithm. The next step is to infer the masks, via Belief Propagation.

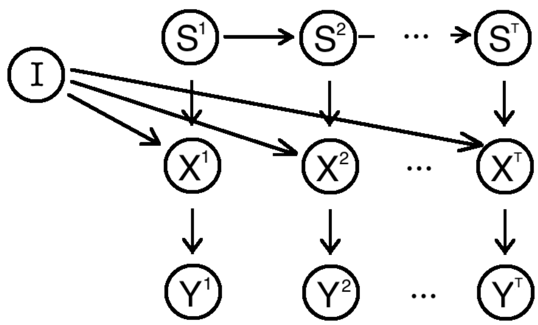
1 Introduction

The diffraction limit is a fundamental property of a lens: it gives an upper bound on the resolution of the images that can be obtained. It is proportional to the wavelength of the light, and to the baseline noise level. "Super-resolution" refers to ways of defeating this limit, using techniques such as masking: a mask blocks selected pixels of the image, and by taking multiple pictures with complementary masks, one may put them together to achieve super-resolution.

With telescopes, one can carefully design a set of masks; however, with microscopes, the masks would need to be tiny, and the imager may not be able to implement masks at this level of precision. One workaround is to introduce random masks, such as gas particles that move around, randomly blocking different sets of pixels. This is not going to be quite as informative as if we could have designed the masks, and the analysis is going to be harder, but it still allows us to achieve super-resolution.

2 The model

- I is the original images, encoded as a vector of U^2 pixels. This is what we are ultimately interested in.
- $S_i^{t+1} \sim Q(S_i^t)$, i.e. the masks follow a two-state Markov chain with state-space equal to $\{0, 1\}$, and in particular each pixel evolves independently.
- Let $X^t = I \odot S_t$, or in other words: $X_i^t = I_i S_i^t$ i.e. X^t is the image after masking with S^t.
- Let $Z^t = AX^t$, where A is a Gaussian blur matrix, i.e. Z^t is a blurred version of X^t.
- $Y_i^t \sim Poi(\gamma Z_i^t)$ for each pixel i. We can interpret Y_i^t as the number of photons observed at pixel i and time t. γ can be interpreted as exposure time.



3 Simulations

We simulate a 5x5 image, in which the Gaussian blur has a standard deviation equal to 0.5.

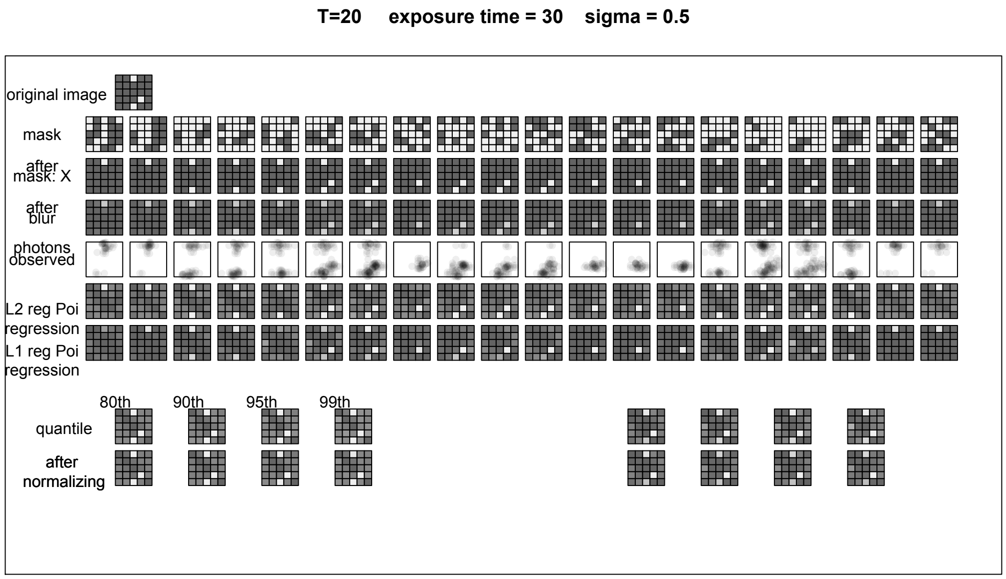
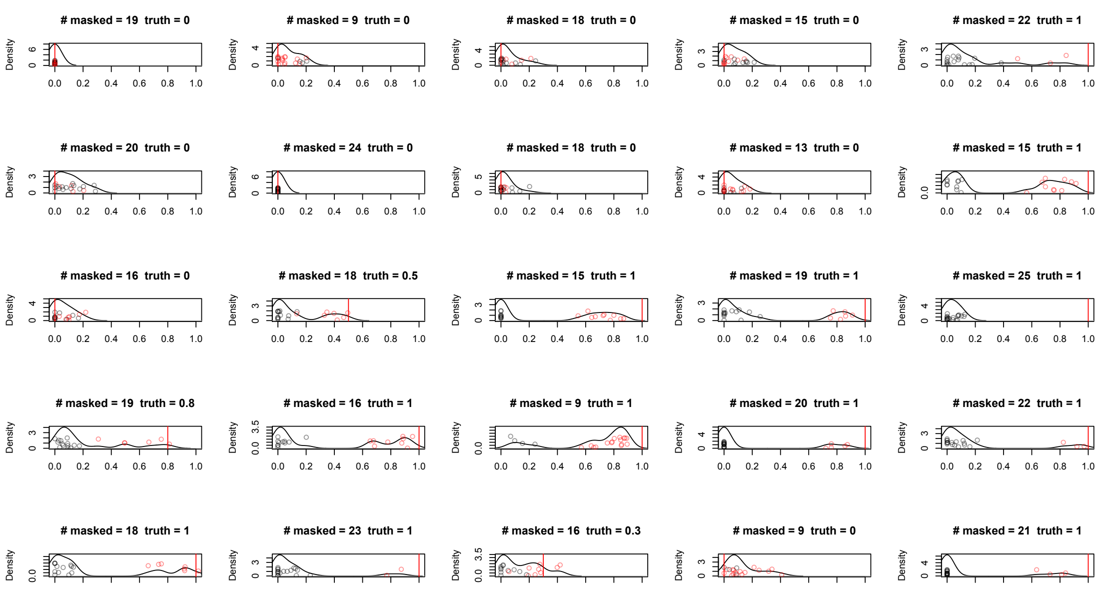


Figure 1: A
At the bottom of the figure, on the left side, we see the quantiles from the Ridge estimates; on the right, we see quantiles from the Lasso estimates.

Now, for each pixel, the Ridge estimate has a bimodal distribution: one bump near 0 for times in which the mask is hiding the pixel (black dots), one bump near the true value for times in which the mask is open (red dots). The value in the true image is shown as a red line.



4 Inference

4.1 General considerations

As a matter of identifiability, we can only expect to have information about pixels that become **active** during the experiment, i.e. $S_i^t = 1$ for some $t \in \{1, ..., T\}$. This can fail if T is not large enough, or if the Markov chain mixes too slowly.

4.2 Maximum Likelihood

The simplest approach to inference would be to maximize the marginal likelihood:
$\hat{I} = \arg\max_I P(Y|I)$. However, this approach quickly becomes intractable because it involves integrating out S, which has dimension equal to U^2T. Even exploiting the conditional independences, this is too difficult to do directly.

4.3 Initialization

For each t, we can estimate X^t by only considering Y^t, i.e. without smoothing it with the neighboring X^t. If the chain mixes quickly enough, the X^t become approximately exchangeable, and smoothing becomes unnecessary.

We use two objective functions:

- with a ridge penalty: $\widehat{X}^t = \arg\min_{X^t} -logP(Y^t|X^t) + \lambda\,||X^t||_2^2$
- with a lasso penalty: $\widetilde{X}^t = \arg\min_{X^t} -logP(Y^t|X^t) + \lambda\,||X^t||_1$

The Lasso estimate tends to be sparse, and is more appropriate in settings where I tends to be sparse.

Having obtained these estimated "masked images" for each t, we could now try extract the upper bump of the bimodal distribution, to initialize, or to estimate the image. Instead, we take a high quantile of the bimodal distribution, with the hope that it will be on the upper bump.

5 Optimization

Both objectives, ridge and Lasso, are convex. The ridge objective is differentiable, and thus easy to optimize. Therefore this section focuses on the Lasso problem.

5.1 Interior-point method

We smoothe the non-differentiability of the Lasso penalty with a quadratic in the range $[-\epsilon, \epsilon]$, and iteratively shrink ϵ, initializing with the optimum for the previous value of ϵ.

The 25 runs of the interior point method take 26 seconds.

5.2 ISTA

Since we desire a faster algorithm, we decided to look into ISTA and FISTA (Beck and Taboulle, 2009). The ISTA algorithm is based on iteratively optimizing a quadratic approximation of the objective around the current point x_k.

$Q_L(x, y) = f(y) + < x - y, \nabla f(y) > + \frac{L}{2} \, ||x - y||_2^2$

The iteration is: $x_k = p_L(x_{k-1})$

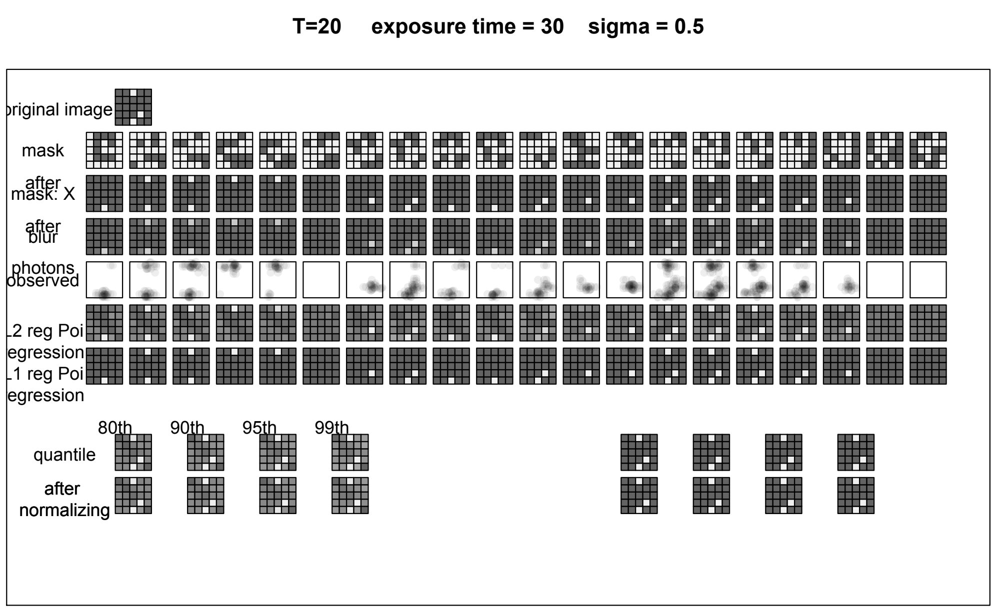
In the case of the Lasso penalty, Beck and Taboulle derive an analytic form for p_L.

$p_L(y) = \arg\min_x \{Q_L(x, y)\}$

$x_k = \mathcal{T}_{\lambda t_k}(x_{k-1} - t_k \nabla f(x_{k-1}))$, where $\mathcal{T}_\alpha$ is the thresholding operator, a continuous piecewise linear function that maps small values to 0.

FISTA modifies ISTA by keeping track of two variables, x and y, and accelerating the optimization by centering the quadratic around an extrapolated step.

Using $\lambda = 3$, and step size of 0.001, the 25 runs of ISTA with analytical updates take 4.3 seconds, and produce the following figure.



5.3 Future work

The positivity constraint for the Poisson parameter is a linear constraint, since the blur can be represented by a matrix. However, the constraint that all pixels in X^t be positive is non-linear, and base R does not support optimization with non-linear constraints.

Acknowledgements

Thanks to Liam Paninski and Eftychios Pnevmatikakis.

References

- Beck and Teboulle (2009) - "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems"
- Ziniel and Schniter (2012) - "Dynamic Compressive Sensing of Time-Varying Signals via Approximate Message Passing"