FORUM STIC

Paris-Saclay 2014

DOHMATOB, GRAMFORT, THIRION, VAROQUAUX

N° 4.16

BENCHMARKING SOLVERS FOR TV-ℓ_1 PENALIZED LEAST-SQUARE AND LOGISTIC REGRESSION IN BRAIN IMAGING

PROBLEM STATEMENT

$$\textbf{Minimize} : \mathcal{L}(X, y, w) + \alpha\left(\rho\|w\|_1 + (1-\rho)\,TV(w)\right) \qquad (1)$$
$$\textbf{For} : w \in \mathbb{R}^p$$

where :

- $\mathcal{L}(X, y, w)$ is the *loss* term = the loss incurred by using the *loadings vector* w to predict a sample $y \in \mathbb{R}^n$ of n responses from a sample $X \in \mathbb{R}^{n\times p}$ of n corresponding brain images.
  - X is commonly called the *design matrix* whilst y is the *response variate*.
  - $\mathcal{L}(X, y, w) \equiv \frac{1}{2}\|y - Xw\|_2^2$ for linear regression, etc.
  - $n \ll p$ for brain data (high-dimensional problem) $\Longrightarrow$ need for regularization
    - Typically, $n \sim 10 - 10^2$ brain images and $p \sim 10^4 - 10^6$ voxels
- $\alpha\left(\rho\|w\|_1 + (1-\rho)\,TV(w)\right)$ is the *regularization* (aka *penalty*) term (Baldassarre et al., 2012; Michel et al., 2011; Abraham et al., 2013) :
  - It encodes "sparsity + structure" prior on the optimal loadings vector $\hat{w}$.
  - $\alpha \geq 0$ is the overall amount of regularization (tradeoff between data and regu.).
  - $\|w\|_1$ is the ℓ_1-norm of w defined $\|w\|_1 := \sum_{j\in voxels} |w_j|$
  - TV(w) is the *isotropic* Total-variation of w defined by $TV(w) := \|\nabla(w)\|_{21} := \sum_{j\in voxels}\sqrt{(\nabla^x w)_j^2 + (\nabla^y w)_j^2 + (\nabla^z w)_j^2}$, and $\nabla := [\nabla^x, \nabla^y, \nabla^z]^T \in \mathbb{R}^{3p\times p}$ is the 3D discrete spatial gradient operator.
  - $\rho \in [0, 1]$ controls the tradeoff between *sparsity* (enforced by the minimizing ℓ_1-norm of w) and *spatial structure* (enforced by minimizing the Total-Variation of w).

NEED FOR FAST SOLVERS

Problem (1) is a high-dimensional non-smooth convex optimization problem and calls for novel optimization techniques. Dohmatob et al. (2014)

- We focus on convergence time for small tolerance
- Lack of good solver and explicit control of tolerance can lead to brain maps and conclusions that reflect properties of the solver more than of the TV-ℓ_1 solution !

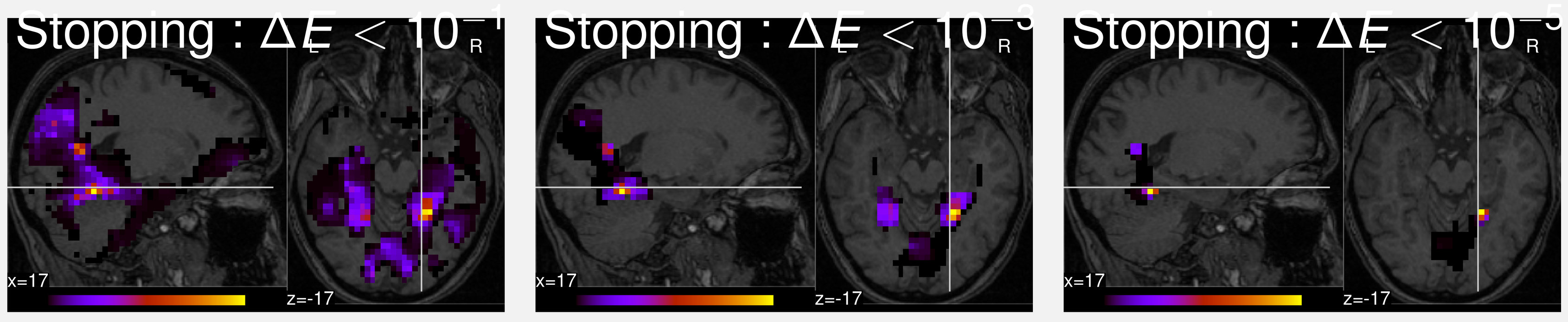

FIGURE : TV$-\ell_1$ maps ($\hat{w}$) for the face-house discrimination task on the HAXBY2001 dataset, for different levels of tolerance. Dohmatob et al. (2014)

BENCHMARKS

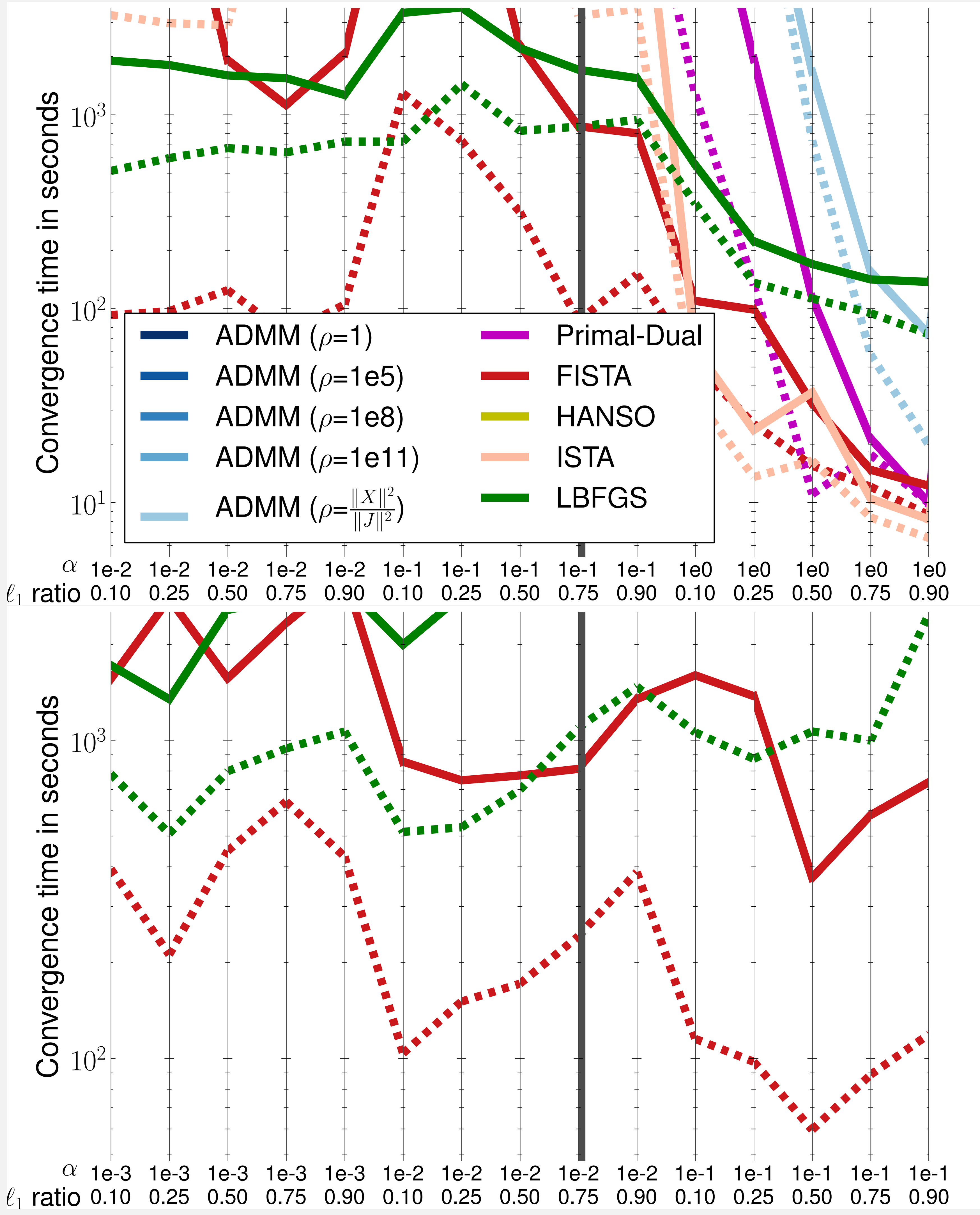

FIGURE : TV$-\ell_1$ penalized Least-Squares Regression. **Top** : on the visual recognition face-house discrimination task ; **Bottom** : on the Mixed gambles dataset. Broken lines correspond to a tolerance of 10^0, whilst full-lines correspond to 10^{-2}. The thick vertical line indicates the best model selected by cross-validation. Dohmatob et al. (2014)

Alexandre Abraham, Elvis Dohmatob, Bertrand Thirion, Dimitris Samaras, and Gael Varoquaux. Extracting brain regions from rest fMRI with total-variation constrained dictionary learning. *MICCAI*, 2013.

Luca Baldassarre, Janaina Mourao-Miranda, and Massimiliano Pontil. Structured sparsity models for brain decoding from fMRI data. In *PRNI*, page 5, 2012.

Elvis Dohmatob, Alexandre Gramfort, Bertrand Thirion, Gaël Varoquaux, et al. Benchmarking solvers for tv-l1 least-squares and logistic regression in brain imaging. *Pattern Recognition in Neuroimaging (PRNI)*, 2014.

Vincent Michel, Alexandre Gramfort, Gael Varoquaux, Evelyn Eger, and Bertrand Thirion. Total variation regularization for fMRI-based prediction of behaviour. *IEEE Transactions on Medical Imaging*, 30 :1328, 2011.