Sparse Maximin Aggregation of Neuronal Activity

Søren Wengel Mogensen
Department of Mathematical Sciences
University of Copenhagen
Email: swengel@math.ku.dk

Adam Lund
Department of Mathematical Sciences
University of Copenhagen

Niels Richard Hansen
Department of Mathematical Sciences
University of Copenhagen

***Abstract*—When analyzing large and inhomogeneous data sets it is of interest to obtain a robust estimate of an underlying signal. We consider a large data set describing neuronal activity in which systematic noise components are present. We propose the use of (soft) maximin aggregation and L_1-penalization to obtain a robust and sparse signal from this noisy data. An approximative computational method and an exact LARS-type method giving the entire solution path are presented.**

I. Introduction

Let X and B be random vectors taking values in $\mathbb{R}^p$ and ε be a zero-mean real random variable. Assume

$$Y = X^t B + \varepsilon.$$

We think of X as a vector of predictor values and of B as a vector of coefficients. If the distribution of B is degenerate, we have a standard linear regression. In the general case, we could ask for a single $\beta \in \mathbb{R}^p$ to capture some feature of the data. For this purpose, define the *maximin effects* [1]

$$\arg\max_{\beta \in \mathbb{R}^p} \min_{b \in F}(V_{\beta,b}), \quad \text{where} \quad V_{\beta,b} = 2\beta^t \Sigma b - \beta^t \Sigma \beta, \tag{1}$$

Σ is the population Gram matrix of X, and F is the support of the distribution of B. The maximin effects maximize minimal (over F) explained variance, $V_{\beta,b}$, when compared to the constant prediction.

Increasing F will only bring the maximin effects closer to the origin which corresponds to the constant prediction. This robustness feature is attractive when estimating F from noisy and inhomogeneous data sets as argued in [1].

We consider the case with observations $Y_1, \dots, Y_n$ having *known groups*, meaning that B is constant within each of G groups,

$$Y_i = X_i^t B_{g(i)} + \varepsilon_i$$

for a known labeling function $g : \{1, \dots, n\} \to \{1, \dots, G\}$.

II. Soft maximin Effects

For $x \in \mathbb{R}^G$ and $\zeta > 0$ consider the *soft maximum function*

$$s_\zeta(x) = \zeta^{-1} \log\left(\sum_{g=1}^{G} \exp(\zeta x_g)\right).$$

Define the L_1-penalized *soft maximin* problem,

$$\min_{\beta \in \mathbb{R}^p} s_\zeta(-\hat{V}_\beta) + \lambda \|\beta\|_1 \tag{2}$$

where $\hat{V}_\beta = (\hat{V}_\beta^1, \dots, \hat{V}_\beta^G)$ are empirical, group-specific explained variances and λ is a non-negative tuning parameter. The soft maximin problem can either be seen as an empirical approximation to (1), or as a problem in its own right, noting that depending on ζ the problem interpolates between mean aggregation and maximin aggregation, and thus balances the properties of the two.

III. Computations

We solved (2) using a proximal gradient based algorithm that iteratively applies a forward-backward type operator (essentially a proximal operator) to an initial point in the solution space [2]. By showing that s_ζ is strongly convex and C^∞ it follows that (2) can be solved using the non-monotone proximal algorithm from [2]. This algorithm extends the standard proximal gradient algorithm to problems with a locally Lipschitz continuous loss, by finding fixed points for the then locally firmly non-expansive forward-backward operator. These fixed points constitute solutions to the problem (2). Using warm starting, we solved (2) for a finite sequence of λs. The algorithm is implemented in the R CRAN package SMMA [3] with the further option of using a SCAD instead of an L_1-penalty.

For our data example the design matrix is the same for all groups and one can exploit this to obtain the complete (hard) maximin solution path $\beta(\lambda)$ which will be piecewise linear in $\mathbb{R}^p$ as a function of λ [4], see Figure 4. This is analogous to the LARS algorithm in the standard regression setting [5], [6]. This method was not applied to the example data as it does not scale well with the size of the data.

IV. Data Example

Our example data was obtained using *voltage-sensitive dye imaging* on the visual cortices of ferrets under a stimulus. The observations are spatio-temporal measurements of light intensity (two spatial dimensions and time) and stem from a total of 275 recordings of 13 ferrets. We treat the recordings as the known groups in (2).

Due to the delicate nature of the method many observations are highly irregular (Figures 1 and 3), prompting the original authors to discard some data [7]. The measurements suffer from both a low signal-to-noise ratio and large, systematic noise components. Due to the size of data, design-matrix free methods come in handy [8].

Neuroscientists expected to observe a temporally (and possibly spatially) sparse signal after the stimulus. Using a L_1-penalty, we estimated the spatio-temporal domain in which the signal is non-zero, exploiting the robust nature of the maximin effects.

V. Conclusion

Maximin aggregation combined with penalization offers an attractive way of obtaining a sparse signal or signal localization from extremely noisy and inhomogeneous data. In our data example it allows the analyst to obtain meaningful results using the entire data set and localizes a spatio-temporally sparse signal (Figure 2).

The estimation of penalized maximin effects is a computational challenge but is feasible for a sequence of penalty parameter values using e.g. a proximal gradient algorithm. Considering the soft maximin problem allows the analyst to strike a balance between (hard) maximin aggregation and mean aggregation.

Acknowledgments

This work was supported by a research grant (13358) from VILLUM FONDEN.

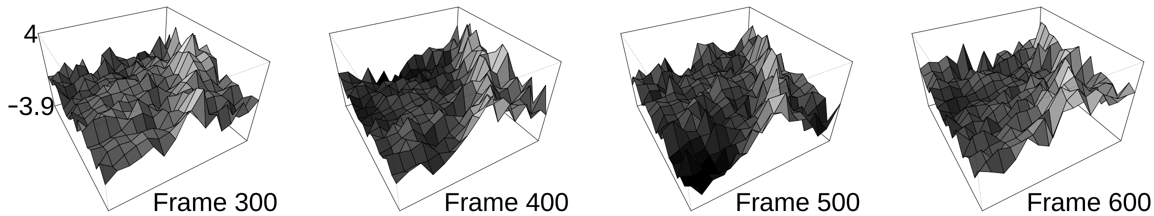

Fig. 1. Snapshots of a single recording (raw data). The recording exhibits systematic noise components.

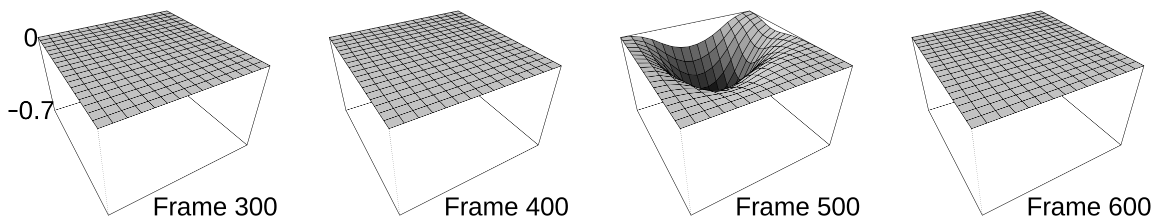

Fig. 2. Snapshots of the fitted maximin effects showing a temporally sparse signal after stimulus onset, despite the irregularity of some recordings (see above).

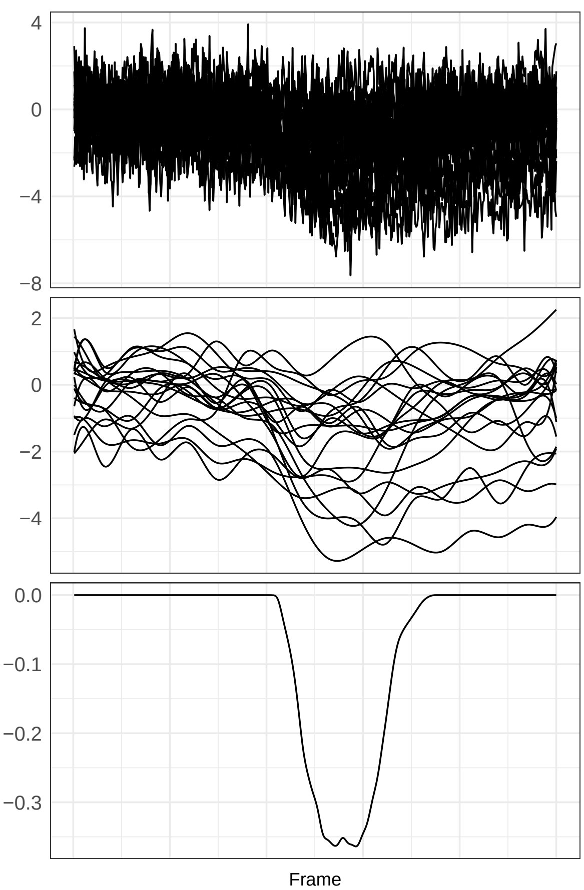

Fig. 3. Measurements from a single recording device during several recordings. Top: 20 randomly selected tracks as they evolve over time. Middle: smoothed version of the 20 tracks above. Bottom: prediction by the maximin effects estimated from the full data set. Note the different scales of the y-axes due to the implicit shrinkage of the estimator. The maximin effects localize a temporally sparse signal.

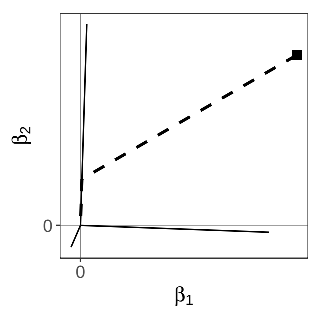

Fig. 4. Simple example of a solution path in $\mathbb{R}^2$. The three black line segments indicate the sets of points in which the loss is not differentiable. The square is the unpenalized (hard) maximin effects and the dashed line is the solution path. For large enough values of λ the solution is the zero vector.

References

[1] N. Meinshausen and P. Bühlmann, "Maximin effects in inhomogeneous large-scale data," *The Annals of Statistics*, vol. 43, no. 4, pp. 1801–1830, 2015.

[2] X. Chen, Z. Lu, and T. K. Pong, "Penalty methods for a class of non-Lipschitz optimization problems," *SIAM Journal on Optimization*, vol. 26, no. 3, pp. 1465–1492, 2016.

[3] A. Lund, "SMMA: Soft maximin estimation for large scale array-tensor models," 2017, R package, version 1.0.1. [Online]. Available: https://CRAN.R-project.org/package=SMMA

[4] J. Roll, "Piecewise linear solution paths with application to direct weight optimization," *Automatica*, vol. 44, no. 11, pp. 2732–2737, 2008.

[5] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–451, 2004.

[6] S. Rosset and J. Zhu, "Piecewise linear regularized solution paths," *The Annals of Statistics*, vol. 35, no. 3, pp. 1012–1030, 2007.

[7] P. E. Roland, A. Hanazawa, C. Undeman, D. Eriksson, T. Tompa, H. Nakamura, S. Valentiniene, and B. Ahmed, "Cortical feedback depolarization waves: A mechanism of top-down influence on early visual areas," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 12 586–12 591, 2006.

[8] A. Lund, M. Vincent, and N. R. Hansen, "Penalized estimation in large-scale generalized linear array models," *Journal of Computational and Graphical Statistics*, 2017.