# Holonomic extended *least angle regression*

**Marc Härkönen** Georgia Institute of Technology, Atlanta, GA, USA
**Tomonari Sei** The University of Tokyo, Tokyo, Japan
**Yoshihiro Hirose** Hokkaido University, Sapporo, Japan

## 1 Introduction

Let $\boldsymbol{y} = (y_1, \dots, y_n)$ be a set of independent observations from a statistical model. Consider exponential families of the form

$$p(\boldsymbol{y} \mid \boldsymbol{\xi}) = \exp\left( \sum_{a=1}^{n} y_a \xi^a + \sum_{b=1}^{r} u_b(\boldsymbol{y}) \xi^{b+n} - \psi^*(\boldsymbol{\xi}) \right), \tag{1}$$

for some functions $u_b \colon \mathbb{R}^n \to \mathbb{R}$. We make the following interpretation

- $\xi^1, \dots, \xi^n$ are natural parameters associated to $y_i$.
- $\xi^{n+1}, \dots, \xi^{n+r}$ are natural parameters common to all observations (e.g. variance, if assumed the same for all $y_i$).
- $\psi^*(\boldsymbol{\xi})$ is the logarithm of the normalizing constant.

If for $i = 1, \dots, n$ we have $\xi_i = \boldsymbol{x}_i \cdot \boldsymbol{\theta}'$ for a covariate vector $\boldsymbol{x}_i \in \mathbb{R}^d$, we can view this exponential family as a generalized linear model with canonical link function. If we set $\theta^{d+b} = \xi^{n+b}$ for all $b = 1, \dots, r$ we can write (1) as

$$p(\boldsymbol{y} \mid \boldsymbol{\theta}) = \exp\left( \boldsymbol{Y} \cdot \tilde{\boldsymbol{X}} \boldsymbol{\theta} - \psi(\boldsymbol{\theta}) \right). \tag{2}$$

by setting $\boldsymbol{Y}(\boldsymbol{y}) = (y_1, \dots, y_n, u_1(\boldsymbol{y}), \dots, u_r(\boldsymbol{y}))$ and for some matrix $\tilde{\boldsymbol{X}}$.

## 2 Information geometry

We can view the model (2) as a manifold with a coordinate system $\boldsymbol{\theta}$, where each point corresponds to a probability distribution. Exponential families give rise to dually flat manifolds [1]. The function $\psi(\boldsymbol{\theta})$ is the potential function of $\boldsymbol{\theta}$. From this we get a dual coordinate system $\boldsymbol{\eta}$, the expectation parameter, given by

$$\boldsymbol{\eta} = \mathrm{E}[\boldsymbol{Y} \mid \boldsymbol{\theta}] = \left( \frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T. \tag{3}$$

## 3 Dually flat manifolds

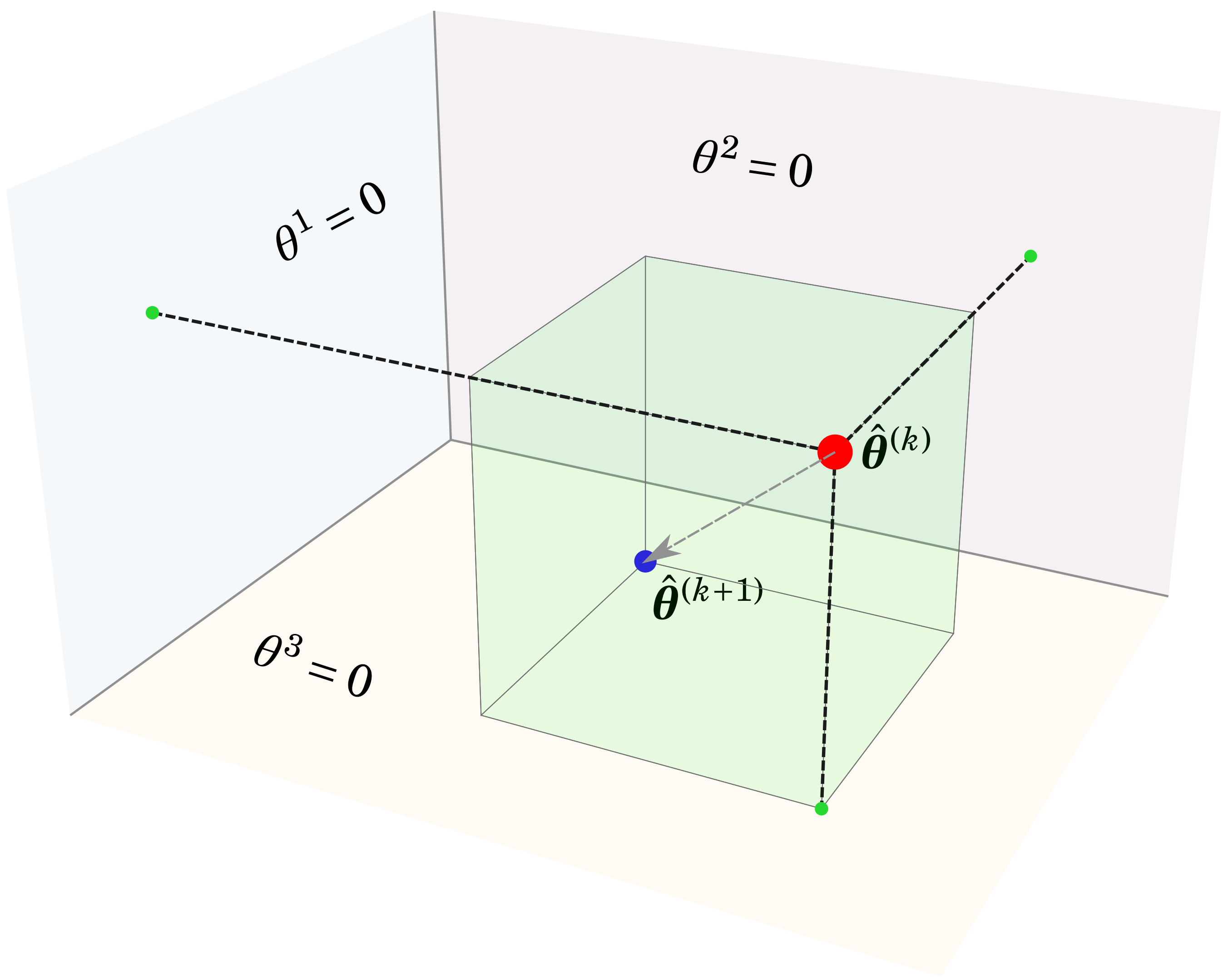
The $\boldsymbol{\theta}$ coordinate system is said to be *exponential flat*, or *e-flat*, and the $\boldsymbol{\eta}$ coordinate system is *mixture flat*, or *m-flat*. We can also obtain the Fisher information matrix from the potential function

$$\boldsymbol{G} = \mathrm{Hess}(\psi(\boldsymbol{\theta})) = \frac{\partial^2 \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}.$$

## 4 Extended LARS

The extended LARS algorithm [2] is a modification of LARS (Least Angle Regression) to exponential families of the form (2). The algorithm ranks the covariates by order of importance. We start with the MLE of the model containing all covariates, and at each step we find and eliminate the least impactful covariate:

- m-project the current estimate $\hat{\boldsymbol{\theta}}^{(k)}$ onto each plane $\theta^i = 0$.
- identify the covariate $i^*$ whose m-projection is the closest (in terms of Kullback–Leibler divergence).
- move along the diagonal of a cube until $\theta^{i^*} = 0$.



$\theta^1 = 0$
$\theta^2 = 0$
$\hat{\boldsymbol{\theta}}^{(k)}$
$\hat{\boldsymbol{\theta}}^{(k+1)}$
$\theta^3 = 0$

## 5 Adding holonomicity

There are cases where $\psi(\boldsymbol{\theta})$ has no closed form expression, and has to be evaluated by numerical integration. In this case the coordinate conversion $\boldsymbol{\theta} \mapsto \boldsymbol{\eta}$ in eq. (3) and the inverse conversion $\boldsymbol{\eta} \mapsto \boldsymbol{\theta}$ have to be computed numerically. Our main result is the holonomic extended LARS, or HELARS. We use a method inspired from the holonomic gradient method [3] to avoid numerical integration. Instead, we will use computationally efficient numerical ODE solvers.

If we have a holonomic ideal in the Weyl algebra annihilating $\psi(\boldsymbol{\theta})$, we can construct a Pfaffian system

$$\frac{\partial \boldsymbol{Q}}{\partial \theta^i} = \boldsymbol{A}_i(\boldsymbol{\theta}) \boldsymbol{Q},$$

where $\psi(\boldsymbol{\theta})$ can be recovered from $\boldsymbol{Q}$. Such an ideal can be either computed by hand or algorithmically [4]. Using the Pfaffian system, we can easily recover the conversion $\boldsymbol{\theta} \mapsto \boldsymbol{\eta}$ and the Fisher information matrix.

In the first step of the algorithm, we use the holonomic gradient method to compute the MLE of the model including all covariates. Assume that at each step $k$ we have computed the value $\boldsymbol{Q}(\hat{\boldsymbol{\theta}}^{(k)})$. We then need to compute m-projections, which we do using the Newton-Raphson method. Since we have the derivative and Hessian of $\psi(\boldsymbol{\theta})$ for free from the Pfaffian system, we can avoid numerical integration. From the m-projections we deduce the next point $\hat{\boldsymbol{\theta}}^{(k+1)}$, and numerically solve an ODE to obtain $\boldsymbol{Q}(\hat{\boldsymbol{\theta}}^{(k+1)})$.

## References

[1] Amari, S.i., Nagaoka, H.: Methods of information geometry, *Translations of Mathematical Monographs*, vol. 191. American Mathematical Society, Providence, RI; Oxford University Press, Oxford (2000).

[2] Hirose, Y., Komaki, F.: An extension of least angle regression based on the information geometry of dually flat spaces. J. Comput. Graph. Statist. **19**(4), 1007–1023 (2010).

[3] Nakayama, H., Nishiyama, K., Noro, M., Ohara, K., Sei, T., Takayama, N., Takemura, A.: Holonomic gradient descent and its application to the Fisher-Bingham integral. Adv. in Appl. Math. **47**(3), 639–658 (2011).

[4] Oaku, T.: Algorithms for $b$-functions, restrictions, and algebraic local cohomology groups of $D$-modules. Adv. in Appl. Math. **19**(1), 61–105 (1997).

**Georgia Institute of Technology**
**Atlanta, GA, USA**

**Contact information: Marc Härkönen**
**Email: harkonen@gatech.edu**