# Sparse projections onto the simplex

Tasos Kyrillidis[†], Stephen Becker[‡], Volkan Cevher[†], Christoph Koch[§]
{anastasios.kyrillidis,volkan.cevher,christoph.koch}@epfl.ch,
stephen.becker@upmc.fr

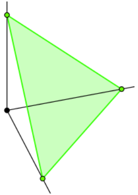
[†]Laboratory for Information and Inference Systems (EPFL)
[‡]Laboratory Jacques Louis-Lions (Université Pierre et Marie Curie)
[§]Data Analysis, Theory and Applications Laboratory (EPFL)

http://lions.epfl.ch/ http://data.epfl.ch/

# Sparse projections with simplex and simplex-type constraints

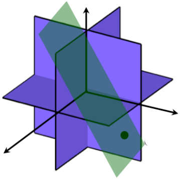
- **Sparse projection onto the simplex hyperplane:**

$$(\mathcal{P}^{\mathcal{S}}): \qquad \boldsymbol{\beta}^* \in \underset{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_0 \leqslant s, \boldsymbol{\beta} \in \Delta_\lambda^+}{\operatorname{argmin}} \quad \|\boldsymbol{\beta} - \mathbf{w}\|_2^2$$



- **Sparse projection onto simplex-type hyperplane:**

$$(\mathcal{P}^{\mathcal{H}}): \qquad \boldsymbol{\beta}^* \in \underset{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_0 \leqslant s, \boldsymbol{\beta} \in \Delta_\lambda}{\operatorname{argmin}} \quad \|\boldsymbol{\beta} - \mathbf{w}\|_2^2$$



- Subtle difference:

$$\Delta_\lambda^+ = \left\{ \boldsymbol{\beta} \in \mathbb{R}^n : \beta_i \geqslant 0, \sum_i \beta_i = \lambda \right\} \quad \text{Vs.} \quad \Delta_\lambda = \left\{ \boldsymbol{\beta} \in \mathbb{R}^n : \sum_i \beta_i = \lambda \right\}$$

## Why are these projections interesting?

# When convex relaxations conflict with problem constraints...

- **Sparse constrained optimization:**

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^n : \|\boldsymbol{\beta}\|_0 \leqslant s} f(\boldsymbol{\beta}) \quad \xrightarrow{\text{Convexify...}} \quad \min_{\boldsymbol{\beta} \in \mathbb{R}^n : \|\boldsymbol{\beta}\|_1 \leqslant \tau} f(\boldsymbol{\beta})$$

  (i) Specific instances of $f(\boldsymbol{\beta})$ in the poster session...
  (ii) $\|\boldsymbol{\beta}\|_0$: $\ell_0$-"norm" where its convex relaxation its $\|\boldsymbol{\beta}\|_1$.

- **The power of convex relaxations:** polynomial solvability and provable recovery guarantees.

# When convex relaxations conflict with problem constraints...

- **Sparse constrained optimization:**

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^n : \|\boldsymbol{\beta}\|_0 \leqslant s} f(\boldsymbol{\beta}) \quad \xrightarrow{\text{Convexify...}} \quad \min_{\boldsymbol{\beta} \in \mathbb{R}^n : \|\boldsymbol{\beta}\|_1 \leqslant \tau} f(\boldsymbol{\beta})$$

 $(i)$ Specific instances of $f(\boldsymbol{\beta})$ in the poster session...
 $(ii)$ $\|\boldsymbol{\beta}\|_0$: $\ell_0$-"norm" where its convex relaxation its $\|\boldsymbol{\beta}\|_1$.

- **The power of convex relaxations:** polynomial solvability and provable recovery guarantees.
- In many cases, **true constraints result in fixed convex metric:**

E.g.: simplex constraint $\Delta_1^+$:

$$\left.\begin{array}{l} \sum_i \beta_i = 1 \\ \beta_i \geqslant 0 \end{array}\right\} \Longrightarrow \|\boldsymbol{\beta}\|_1 = 1.$$

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^n : \|\boldsymbol{\beta}\|_0 \leqslant s, \|\boldsymbol{\beta}\|_1 = 1} f(\boldsymbol{\beta}) \quad \xrightarrow{\text{Convexify...}} \quad ???$$

# Sparse projections onto the simplex

- **Sparse projection onto the simplex:**

$$(\mathcal{P}^{\mathcal{S}}): \qquad \boldsymbol{\beta}^* \in \underset{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_0 \leqslant s, \boldsymbol{\beta} \in \Delta_\lambda^+}{\operatorname{argmin}} \|\boldsymbol{\beta} - \mathbf{w}\|_2^2 \tag{1}$$

- The problem (1) is equivalent to the nested minimization problem:

$$\{\mathcal{S}^*, \boldsymbol{\beta}^*_{\mathcal{S}^*}\} = \underset{\mathcal{S}:\mathcal{S}\in\Sigma_s}{\operatorname{argmin}} \left[ \min_{\substack{\boldsymbol{\beta}_{\mathcal{S}} \in \Delta_\lambda^+, \\ \boldsymbol{\beta}_{\setminus \mathcal{S}} = 0}} \|(\boldsymbol{\beta} - \mathbf{w})_{\mathcal{S}}\|_2^2 + \|(\mathbf{w})_{\setminus \mathcal{S}}\|_2^2 \right]$$

# Sparse projections onto the simplex

- **Sparse projection onto the simplex:**

$$(\mathcal{P}^{\mathcal{S}}): \quad \boldsymbol{\beta}^* \in \underset{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_0 \leqslant s, \boldsymbol{\beta} \in \Delta_\lambda^+}{\operatorname{argmin}} \|\boldsymbol{\beta} - \mathbf{w}\|_2^2 \qquad (1)$$

- The problem (1) is equivalent to the nested minimization problem:

$$\{\mathcal{S}^*, \boldsymbol{\beta}^*_{\mathcal{S}^*}\} = \underset{\mathcal{S}:\mathcal{S}\in\Sigma_s}{\operatorname{argmin}} \left[ \min_{\substack{\boldsymbol{\beta}_{\mathcal{S}} \in \Delta_\lambda^+, \\ \boldsymbol{\beta}_{\setminus \mathcal{S}} = 0}} \|(\boldsymbol{\beta} - \mathbf{w})_{\mathcal{S}}\|_2^2 + \|(\mathbf{w})_{\setminus \mathcal{S}}\|_2^2 \right]$$

- Given $\mathcal{S}^*$, $(\boldsymbol{\beta}^*)_{\mathcal{S}^*} = \left[ w_i + \frac{1}{|\mathcal{S}^*|} \left( \lambda - \sum_{i \in \mathcal{S}^*} w_i \right) \right]_+ \qquad (1\alpha)$

# Sparse projections onto the simplex

- **Sparse projection onto the simplex:**

$$(\mathcal{P}^{\mathcal{S}}): \quad \boldsymbol{\beta}^* \in \operatorname*{argmin}_{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_0 \leqslant s, \boldsymbol{\beta} \in \Delta_\lambda^+} \|\boldsymbol{\beta} - \mathbf{w}\|_2^2 \tag{1}$$

- The problem (1) is equivalent to the nested minimization problem:

$$\{\mathcal{S}^*, \boldsymbol{\beta}^*_{\mathcal{S}^*}\} = \operatorname*{argmin}_{\mathcal{S}:\mathcal{S}\in\Sigma_s} \left[ \min_{\substack{\boldsymbol{\beta}_{\mathcal{S}} \in \Delta_\lambda^+, \\ \boldsymbol{\beta}_{\setminus \mathcal{S}} = 0}} \|(\boldsymbol{\beta} - \mathbf{w})_{\mathcal{S}}\|_2^2 + \|(\mathbf{w})_{\setminus \mathcal{S}}\|_2^2 \right]$$

- Given $\mathcal{S}^*$, $(\boldsymbol{\beta}^*)_{\mathcal{S}^*} = \left[ w_i + \frac{1}{|\mathcal{S}^*|} \left( \lambda - \sum_{i \in \mathcal{S}^*} w_i \right) \right]_+ \qquad (1\alpha)$
- But how to project *jointly* onto $\Delta_\lambda^+ \cap \Sigma_s$?

# Sparse projections onto the simplex

- **Sparse projection onto the simplex:**

$$(\mathcal{P}^{\mathcal{S}}): \qquad \boldsymbol{\beta}^* \in \underset{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_0 \leqslant s, \boldsymbol{\beta} \in \Delta_\lambda^+}{\operatorname{argmin}} \|\boldsymbol{\beta} - \mathbf{w}\|_2^2 \qquad (1)$$

- The problem (1) is equivalent to the nested minimization problem:

$$\{\mathcal{S}^*, \boldsymbol{\beta}^*_{\mathcal{S}^*}\} = \underset{\mathcal{S}:\mathcal{S}\in\Sigma_s}{\operatorname{argmin}} \left[ \min_{\substack{\boldsymbol{\beta}_{\mathcal{S}} \in \Delta_\lambda^+, \\ \boldsymbol{\beta}_{\setminus \mathcal{S}} = 0}} \|(\boldsymbol{\beta} - \mathbf{w})_{\mathcal{S}}\|_2^2 + \|(\mathbf{w})_{\setminus \mathcal{S}}\|_2^2 \right]$$

- Given $\mathcal{S}^*$, $(\boldsymbol{\beta}^*)_{\mathcal{S}^*} = \left[ w_i + \frac{1}{|\mathcal{S}^*|} \left( \lambda - \sum_{i \in \mathcal{S}^*} w_i \right) \right]_+ \qquad (1\alpha)$
- But how to project *jointly* onto $\Delta_\lambda^+ \cap \Sigma_s$?

**Greedy selector and simplex projector (GSSP)**

$$\mathcal{S}^* = \operatorname{supp}\left(\mathcal{P}_{L_s}(\mathbf{w})\right), \ (\boldsymbol{\beta}^*)_{\mathcal{S}^*} \text{ given by } (1\alpha) \text{ and, } \ (\boldsymbol{\beta}^*)_{\setminus \mathcal{S}^*} = 0.$$

$\mathcal{P}_{L_s}$ keeps the $s$-largest entries (**not in magnitude**).

- Complexity: $\mathcal{O}(n \log n)$.

# Sparse projections onto simplex-type constraints

- **Sparse projection onto simplex-type constraints:**

$$(\mathcal{P}^{\mathcal{H}}): \qquad \boldsymbol{\beta}^* \in \underset{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_0 \leqslant s, \boldsymbol{\beta} \in \Delta_\lambda}{\operatorname{argmin}} \|\boldsymbol{\beta} - \mathbf{w}\|_2^2 \tag{2}$$

- Given $\mathcal{S}^*$, $(\boldsymbol{\beta}^*)_{\mathcal{S}^*} = w_i + \frac{1}{|\mathcal{S}^*|}\left(\lambda - \sum_{i \in \mathcal{S}^*} w_i\right).$ $\quad (2\alpha)$
- GSSP procedure fails in this case...

# Sparse projections onto simplex-type constraints

- **Sparse projection onto simplex-type constraints:**

$$(\mathcal{P}^{\mathcal{H}}): \qquad \boldsymbol{\beta}^* \in \underset{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_0 \leqslant s, \boldsymbol{\beta} \in \Delta_\lambda}{\operatorname{argmin}} \|\boldsymbol{\beta} - \mathbf{w}\|_2^2 \tag{2}$$

- Given $\mathcal{S}^*$, $(\boldsymbol{\beta}^*)_{\mathcal{S}^*} = w_i + \frac{1}{|\mathcal{S}^*|}\left(\lambda - \sum_{i \in \mathcal{S}^*} w_i\right).$ $\quad (2\alpha)$
- GSSP procedure fails in this case...

**Greedy selector and hyperplane projector (GSHP)**

1. $\ell = 1,\ \mathcal{S} = j, \quad j \in \arg\max_i (\lambda w_i).$
2. Repeat: $\ell \leftarrow \ell + 1, \mathcal{S} \leftarrow \mathcal{S} \cup j$, where

$$j \in \arg\max_{i \in \mathcal{N} \setminus \mathcal{S}} \left| w_i - \frac{\sum_{j \in \mathcal{S}} w_j - \lambda}{\ell - 1} \right|,$$

   until $\ell = k$.
3. Set $\mathcal{S}^* \leftarrow \mathcal{S}$ and solve $(2\alpha)$

- GSHP selects the index of the largest element with the same sign as $\lambda$ (Step 1). It then grows the index set one at a time by finding the farthest element from the current mean, as adjusted by $\lambda$ (Step 2).

- Theorem. *GSHP Algorithm provably solves (2).*
- Complexity: $\mathcal{O}(n \log_2(n))$.

## More information and applications in the poster session...

Thank you.