Our contributions:
- A sparse convex additive model (SCAM) to estimate convex (and monotonic) component functions in high dimensional additive modeling
- A sparse difference of convex additive model (SDCAM) to address potential robustness issue of SCAM, e.g., convex functions are mistakenly believed to be concave
- An efficient backfitting algorithm with linear per-iteration complexity

Sparse Convex Additive Model (SCAM)

Given a set of data samples \( \{ (x_i, y_i) : x_i \in \mathbb{R}^m, y_i \in \mathbb{R} \} \), for \( i = 1, \ldots, n \), solve

\[
\min_{\mathbf{y}, \mathbf{z} \in \mathbb{R}^m} \sum_{i=1}^n \left( y_i - \sum_{j=1}^m f_j(x_{ij}) \right)^2 + \lambda \sum_{j=1}^m \left\| f_j \right\|_2^2
\]

where \( f_j \in \mathbb{R} \) are the component fits on the observed values: \( y_j = f_j(x_{ij}), i = 1, \ldots, n \).

- \( f_j \) is convex if \( \| f_j \|_2^2 = \sqrt{\mathbb{E}(f_j^2)} \) is the L2 norm of component function \( f_j \).
- Can reduce to an equivalent finite-dimensional optimization problem:

\[
\min_{\mathbf{y}, \mathbf{z} \in \mathbb{R}^m} \sum_{i=1}^n \left( y_i - \sum_{j=1}^m z_{ij} \right)^2 + \lambda \sum_{j=1}^m \| z_{i, :} \|_2^2
\]

where for \( z \in \mathbb{R}^m \) we define

\[
\| z \|_{\Sigma^{-1} C} := \sum_{i=1}^n \left( \frac{z_i}{\Sigma_{ii}} - \frac{\Sigma_{ii} z_i}{\Sigma_{i,i-1}} \right),
\]

Modified Backfitting Algorithm

In each iteration, fix all component fits except for one \( z_j \) and solve the resulting subproblem:

\[
\min_{\mathbf{y}, \mathbf{z} \in \mathbb{R}^m} \frac{1}{2} \| \mathbf{f} - \mathbf{z} \|_2^2 + \lambda \| \mathbf{z} \|_{\Sigma^{-1} C} + \lambda \| \mathbf{z} \|_2
\]

where \( \mathbf{f} \in \mathbb{R}^m \) is the partial residual that removes the contribution of \( z_j \).

Theorem: The solution can be characterized as

\[
P_{\lambda}(\mathbf{z} \mid \Sigma^{-1} C, \| \cdot \|_2) = P_{\lambda}(P_{\Sigma^{-1} C}(\mathbf{z} \mid \| \cdot \|_2) - \lambda \mathbf{f}).
\]

where \( P_{\Sigma^{-1} C} \) is the proximal operator associated with a convex function \( f \).

\[
P_{\lambda}(\mathbf{r}) = \arg\min_{\mathbf{z} \in \mathbb{R}^m} \frac{1}{2} \| \mathbf{z} - \mathbf{r} \|_2^2 + \lambda \| \mathbf{z} \|_2.
\]

\( P_{\lambda} \) amounts to subtracting the average

\[
P_{\lambda}(\mathbf{r} \mid \lambda) = (1 - \lambda^{-1}) \mathbf{r}
\]

is the block soft thresholding operator

With a suitable change of variables, computing \( P_{\lambda}(\mathbf{y} \mid \| \cdot \|_1) \) is equivalent to

\[
\min_{\mathbf{x} \in \mathbb{R}^m} \frac{1}{2} \| \mathbf{x} - \mathbf{y} \|_2^2 + \lambda \| \mathbf{x} \|_1
\]

for certain lower triangular matrix \( A \).

Using the linear-time algorithm in (Davies and Kovac 2001) to compute the proximal operator \( P_{\lambda}(\mathbf{y} \mid \| \cdot \|_1) \), we are able to compute \( P_{\lambda}(\mathbf{y} \mid \| \cdot \|_1) \) iteratively using the accelerated proximal gradient algorithm.

References