# 1 3/1/2014

This note analyzes SCAM where we penalize $\|f\|_\infty$ instead of $\|\partial f\|_\infty$.

## 1.1 Changes to Optimization

The optimization program must change. We modify two formulations of the optimization. The first is the formulation in which we use both $f$ and $\beta$ as variables.

$$\begin{aligned}
\min_{h_k,\beta_k,\gamma_k} \frac{1}{2n}\sum_{i=1}^n \Big(y_i - \sum_{k'\neq k} h_{k'i} - h_{ki}\Big)^2 + \lambda \sum_k \|h_k\|_\infty \\
\text{s.t. } h_{k(i+1)} = h_{k(i)} + \beta_{k(i)}(x_{k(i+1)} - x_{k(i)}) \\
\beta_{k(i+1)} \geq \beta_{k(i)}, \mathbf{1} h_k = 0
\end{aligned}$$

Of course, the $\lambda\|h_k\|_\infty$ penalty can be replaced by $\lambda\gamma_k$ and linear inequalities involving $\gamma_k$.

Now we reformulate the optimization program that is in term of the discretized second derivative $d_k$.

$$\begin{aligned}
\min_{d_k} \frac{1}{2n}\|Y - \sum_{k=1}^p \Delta_k d_k\|_2^2 + \lambda \sum_{k=1}^p \|\Delta_k d_k\|_\infty \\
\text{s.t. } d_{k(2)}, ..., d_{k(n-1)} \geq 0 \\
\mathbf{1}_n^\mathsf{T} \Delta_k d_k = 0 \quad \forall k
\end{aligned}$$

## 1.2 Subgradient analysis

We take the subgradient of the optimization program with $d_k$.

First, we note that the subgradient of the sup-norm $\partial\|x\|_\infty$ at $x=0$ is $\{z : \sum_i z_i = 1\}$.

We fix one dimension $k$, let $Y' = Y - \sum_{k'\neq k} \Delta_{k'} d_{k'}$.

The Lagrangian form of the optimization is

$$\mathcal{L}(d_k, v, \gamma) = \frac{1}{2n}\|Y' - \Delta_k d_k\|_2^2 + \lambda\|\Delta_k d_k\|_\infty - \sum_{i=2}^{n-1} v_i d_{ki} + \gamma \mathbf{1}_n^\mathsf{T}\Delta_k d_k$$

with the constraint that $v_i \geq 0$ for all $i$.

We want to find under what conditions does the solution $d_k = 0$ satisfy the KKT conditions.

$$\partial_{d_k}\mathcal{L} = \frac{1}{n}\Delta_k^\mathsf{T}(Y' - \Delta_k d_k) + \lambda\Delta_k^\mathsf{T} z - v + \gamma\mathbf{1}_n^\mathsf{T}\Delta_k$$

where $z \in \partial\|\Delta_k d_k\|_\infty$.

If we evaluate the subgradient at $d_k = 0$, then we have that

$$\partial_{d_k}\mathcal{L}\Big|_{d_k=0} = \frac{1}{n}\Delta_k^\mathsf{T} Y' + \lambda\Delta_k^\mathsf{T} z - v + \gamma\mathbf{1}_n^\mathsf{T}\Delta_k$$

where $\|z\|_1 \leq 1$.

We want to argue that $d_k = 0$ is an optimal solution. If $d_k = 0$, then complementary slackness and primal feasibility are obvious satisfied. We need only verify stationarity and dual feasibility then.

Let us take a brief digression and understand the $\Delta_k$ matrix a little bit more. $\Delta_k$ is $n \times n-1$. Each row corresponds to sample $i$; each column corresponds to an order $(j)$. Each entry $(i,j)$ is $[X_{ki} - X_{k(j)}]_+$. Let us reorder the samples so that the $i$-th sample is the $i$-smallest sample.

We will construct $\gamma = 0$, and $z = (0, 0, ..., a)$ for some $0 < a < 1$. (coordinates of $z$ correspond to the new sample ordering) We then just need to show that

$$\begin{aligned}
\frac{1}{n}\Delta_k^\mathsf{T} Y' + \lambda \Delta_k^\mathsf{T} z &\geq 0 \\
\frac{1}{n}\sum_{i>j}(X_{ki} - X_{kj})Y'_i + \lambda(X_{kn} - X_{kj})a &\geq 0 \quad \text{for each } j \\
\frac{1}{n}\sum_{i>j}\sum_{j<i'\leq i}\mathsf{gap}_{i'} Y'_i + \lambda(X_{kn} - X_{kj})a &\geq 0 \\
\frac{1}{n}\sum_{i'>j}\mathsf{gap}_{i'}\sum_{i\geq i'} Y'_i + \lambda(X_{kn} - X_{kj})a &\geq 0 \\
\frac{1}{n}\sum_{i'>j}\mathsf{gap}_{i'}\mathbf{1}_{i':n}^\mathsf{T} Y' + \lambda(X_{kn} - X_{kj})a &\geq 0
\end{aligned}$$

Where $\mathsf{gap}_i = X_{ki} - X_{k,i-1}$. (with respect to ordered indices) The notation $\mathbf{1}_{i':n}$ is a vector that is one on the last $i'$ to $n$ coordinates and zero elsewhere.

Suppose we show that $\frac{1}{n}\mathbf{1}_{i:n}^\mathsf{T} Y'$ is on the order of $O(\frac{1}{\sqrt{n}})$, then we would be done. Here, we will need to bound $\|Y'\|_\infty$.

## 1.3 False Negative Analysis

Before we begin, we keep in mind that the analysis should be flexible and should be easy to modify to accomodate the following:

1. choice of norm in the penalty
2. with or without a Lipschitz constraint, boundedness constraint

### 1.3.1 Preliminary

We start with the definitions. Let $y = (y_1, ..., y_n)$ and $y_i = f_0(x_i) + w_i$. We assume $f_0$ to be convex. [1] We assume that all $x_i \in [-b, b]^s$

Let $\mathcal{C}^1$ denote the set of univariate convex functions supported on $[-b, b]$. Let $\mathcal{C}^1_L \equiv \{f \in \mathcal{C}^1 : \|\partial f\|_\infty \leq L\}$ denote the set of $L$-Lipschitz univariate convex functions.

[1] We can probably relax this assumption to be a little bit more general.

Define $\mathcal{C}^s$ as the set of convex additive functions

$$\mathcal{C}^s \equiv \{f \,:\, f = \sum_{k=1}^{s} f_k,\, f_k \in \mathcal{C}^1\}$$

We will also define classes of bounded and Lipschitz convex additive functions.

$$\mathcal{C}^s_B = \{f \in \mathcal{C}^s \,:\, \|f\|_\infty \leq B\}$$
$$\mathcal{C}^s_L = \{f \in \mathcal{C}^s \,:\, f = \sum_{k=1}^{s} f_k,\, \|\partial f_k\|_\infty \leq L\}$$

Let $f^*(x) = \sum_{k=1}^s f^*_k(x_k)$ be the population risk minimizer:

$$f^* = \arg\min_{f \in \mathcal{C}^s} \mathbb{E}\Big(f_0(X) - f^*(X)\Big)^2$$

We let $B$ be an upper bound on $\|f_0\|_\infty$ and $\|f^*\|_\infty$ and let $L$ be an upper bound on $\|\partial_{x_k} f_0\|_\infty$ and $\|\partial f^*_k\|_\infty$.

In the finite sample context where we have $n$ samples $x_1, ..., x_n$, we will abuse notation and let $y, f_0, f^*, w$ denote their de-meaned counterpart, that is, $f_0$ is an $n$-dimensional vector whose coordinate $i$ is $f_0(x_i) - \frac{1}{n}\sum_i f_0(x_i)$.

We define $\widehat{f}$ as the empirical risk minimizer:

$$\widehat{f} = \arg\min\{\|y - f\|_n^2 + \lambda \sum_{k=1}^{s} \|f_k\|_\infty \,:\, f \in \mathcal{C}^s_L, \mathbf{1}_n^\mathsf{T} f_k = 0\}$$

We also define $\widehat{f}^*$ as the noiseless empirical risk minimizer:

$$\widehat{f}^* = \arg\min\{\|f_0 - f\|_n^2 \,:\, f \in \mathcal{C}^s_L, \mathbf{1}_n^\mathsf{T} f_k = 0\}$$

### 1.3.2 Proof

We start from the definition.

$$\|y - \widehat{f}\|_n^2 + \lambda \sum_{k=1}^{s} \|\widehat{f}_k\|_\infty \leq \|y - \widehat{f}^*\|_n^2 + \lambda \sum_{k=1}^{s} \|\widehat{f}^*_k\|_\infty$$
$$\|f_0 + w - \widehat{f}\|_n^2 + \lambda \sum_{k=1}^{s} \Big(\|\widehat{f}_k\|_\infty - \|\widehat{f}^*_k\|_\infty\Big) \leq \|f_0 + w - \widehat{f}^*\|_n^2$$
$$\|f_0 - \widehat{f}\|_n^2 + 2\langle w, f_0 - \widehat{f}\rangle_n + \lambda \sum_{k=1}^{s} \Big(\|\widehat{f}_k\|_\infty - \|\widehat{f}^*_k\|_\infty\Big) \leq \|f_0 - \widehat{f}^*\|_n^2 + 2\langle w, f_0 - \widehat{f}^*\rangle$$
$$\|f_0 - \widehat{f}\|_n^2 - \|f_0 - \widehat{f}^*\|_n^2 + \lambda \sum_{k=1}^{s} \Big(\|\widehat{f}_k\|_\infty - \|\widehat{f}^*_k\|_\infty\Big) \leq 2\langle w, \widehat{f} - \widehat{f}^*\rangle$$

# 2 2/15/2014

The two step procedure for variable selection:

1. Jointly convex approximation:
   $\{f_k^c\}_{k=1,..,p} = \arg\min_{f_k \in \mathcal{C}_x} \mathbb{E}\Big(f(X) - \sum_{k=1}^p f_k(X_k)\Big)^2$
2. Separate concave post-processing, for each $k = 1, ..., p$:
   $f_k^v = \arg\min_{f_k \in \mathcal{C}_v} \mathbb{E}\Big(f(X) - \sum_{k' \neq k} f_{k'}^c(X_{k'}) - f_k(X_k)\Big)^2$

We consider as irrelevant any $k$ where $f_k^c = 0$ and $f_k^v = 0$.

**Claim:** Let $f : C \to \mathbb{R}$ where $C = [0,1]^p$. Suppose $p(x)$ is a positive density on $C$ and that it satisfies the boundary-points condition.

Suppose $f(x)$ and $\nabla f(x)$ are continuous and for all $k$, that $\partial_{x_k} p(x \,|\, x_k)$ and $\partial_{x_k}^2 p(x \,|\, x_k)$ are also continuous as functions on $C$.

Then, the two step procedure is faithful, that is, $f_k^c = 0$ and $f_k^v = 0$ implies that $\partial_{x_k} f(x) = 0$.

*Proof.* We focus on a specific $k$ and suppose that $f_k^c$ and $f_k^v$ are both 0.

We now invoke proposition 2.1 by letting $g(x) = f(x) - \sum_{k' \neq k} f_{k'}^c(x_{k'})$. It is easy to verify that derivative continuity requirements on $g$ and $p(x \,|\, x_k)$ are all met. We can therefore conclude that $g_k^*(x) = \mathbb{E}[f(X) - \sum_{k' \neq k} f_{k'}^c(X_{k'}) \,|\, x_k] = 0$.

We now use corollary 2.1 by letting $\phi(x_{-k}) = \sum_{k' \neq k} f_{k'}^c(x_{k'})$. We therefore derive that $\partial_{x_k} f(x) = 0$ and thus prove faithfulness.

□

## 2.1 Analysis

To prove the faithfulness of this procedure, we use the following extension of additive faithfulness.

**Corollary 2.1.** *Suppose $p(x)$ is a positive density on $[0,1]^p$ and it satisfies the boundary-points condition.*

*For any function $\phi(X_{-k})$ that does not depend on $X_k$:*

$$f_k^*(x_k) = \arg\min_{f_k} \mathbb{E}\Big(f(X) - \phi(X_{-k}) - f_k(X_k)\Big)^2 = \mathbb{E}\Big[f(X) - \phi(X_{-k}) \,|\, x_k\Big]$$

*We have that $f_k^* = 0 \Rightarrow \partial_{x_k} f(x) = 0$.*

*Proof.* Identical to the proof of additive faithfulness. □

We also use the following form of the shape-constrained projection theorem.

**Proposition 2.1.** *Let $C \subset \mathbb{R}^p$ be a compact set and let $g : C \to \mathbb{R}$. Let $p(x)$ be a positive density on $C$ and suppose $\mathbb{E}g(X) = 0$.*

*Suppose that $\partial_{x_k} g(x)$, $\partial_{x_k} p(x \,|\, x_k)$, and $\partial_{x_k}^2 p(x \,|\, x_k)$ are all continuous as functions on $C$. Suppose that $\partial_{x_k}^2 g(x) \geq 0$.*

*Let* $f_k^c(x_k) = \arg\min_{f_k \in \mathcal{C}_x} \mathbb{E}\Big(g(X) - f_k(X_k)\Big)^2$ *and* $f_k^v(x_k) = \arg\min_{f_k \in \mathcal{C}_v} \mathbb{E}\Big(g(X) - f_k(X_k)\Big)^2$ *be the best convex and concave univariate approximation respectively.*

*Then,* $f_k^c = 0$ *and* $f_k^v = 0$ *iff* $g_k^*(x_k) = \mathbb{E}[g(X) \,|\, x_k] = 0$.

*Proof.* First, we will establish that $g_k^*(x_k)$ is twice differentiable and that $\partial_{x_k}^2 g_k^*(x_k)$ is lower bounded.

$$\begin{aligned}
g_k^*(x_k) &= \mathbb{E}[g(X) \,|\, x_k] \\
&= \int_{x_{-k}} g(x) p(x \,|\, x_k) \\
\partial_{x_k}^2 g_k^*(x_k) &= \int_{x_{-k}} g''(x) p(x \,|\, x_k) + 2g'(x) p'(x \,|\, x_k) + g(x) p''(x \,|\, x_k) dx_{-k}
\end{aligned}$$

The first term $g''(x)p(x \,|\, x_k)$ is strictly positive. By assumption, the remaining terms are continuous and hence bounded on a compact set. $\partial_{x_k}^2 g_k^*(x_k)$ is therefore lower bounded.

Before proceeding, we also note that because $\mathbb{E}g(X) = 0$, it must be that $\mathbb{E}g_k^*(X_k) = 0$.

Now suppose $f_k^x = 0$ and $f_k^v = 0$. Let $\sigma_k^2$ denote $\mathbb{E}X_k^2$. Then

$$\arg\min_{c \in \mathbb{R}} \mathbb{E}\Big(g(X) - c(X_k^2 - \sigma_k^2)\Big)^2 = 0$$

Since optimal $c^* = \frac{\mathbb{E}[g(X)(X_k^2 - \sigma_k^2)]}{\mathbb{E}[X_k^2]}$, we know $\mathbb{E}[g(X)X_k^2] = \mathbb{E}\Big[\mathbb{E}[g(X) \,|\, X_k] X_k^2\Big] = 0$.

Because $\partial_{x_k}^2 g_k^*(x_k)$ is lower bounded, for large enough $\alpha$, $g_k^*(x_k) + \alpha(x_k^2 - \sigma_k^2)$ has a non-negative second derivative and thus is convex. Then

$$\arg\min_{c \in \mathbb{R}} \mathbb{E}\Big(g(X) - c(g_k^*(X_k) + \alpha(X_k^2 - \sigma_k^2))\Big)^2 = 0$$

Again, $c^* = \frac{\mathbb{E}[g(X)\big(g_k^*(X_k) + \alpha(X_k^2 - \sigma_k^2)\big)]}{\mathbb{E}\big(g_k^*(X_k) + \alpha(X_k^2 - \sigma_k^2)\big)^2} = 0$, so

$$\begin{aligned}
\mathbb{E}[g(X)\big(g_k^*(X_k) + \alpha X_k^2\big)] &= \mathbb{E}[g(X) g_k^*(X_k)] \\
&= \mathbb{E}\Big[\mathbb{E}[g(X)|X_k] g_k^*(X_k)\Big] \\
&= \mathbb{E}g_k^*(X_k)^2 = 0
\end{aligned}$$

Therefore, $g_k^*(x_k) = 0$.

□

# 3 2/7/2014

## 3.1 Best Shape-Constrained Projection

Even if a function $f_0$ is non-zero, it may be possible that the best convex approximation is zero. But, it cannot be that the best convex approximation and the best concave approximation are both zero.

**Theorem 3.1.** *Let $f_0 : C \to \mathbb{R}$ be a function with a bounded Hessian; let $p$ be some distribution on $C$. Suppose*

$$\operatorname*{arg\,min}_{f \in \mathcal{C}_x \cup \mathcal{C}_v} \mathbb{E}(f_0(x) - f(x))^2 = 0$$

*then $f_0 = 0$ necessarily.*

*Proof.* Suppose that the $\operatorname{arg\,min}_{f \in \mathcal{C}_x \cup \mathcal{C}_v} \mathbb{E}(f_0(x) - f(x))^2 = 0$, then it must be that

$$\operatorname*{arg\,min}_{c \in \mathbb{R}} \mathbb{E}\Big(f_0(x) - cx^\mathsf{T} x\Big)^2 = 0$$

Since the optimal $c$ in the above optimization is $c^* = \frac{\mathbb{E}[f_0(x)x^\mathsf{T} x]}{\mathbb{E}(x^\mathsf{T} x)^2}$, we have that $\mathbb{E}[f_0(x)x^\mathsf{T} x] = 0$.

We know that there exists a convex function $f'$ such that $f' = f_0 + rx^\mathsf{T} c$ for some large $r > 0$.

By assumption and convexity of $f'$, it must be that

$$\operatorname*{arg\,min}_{c \in \mathbb{R}} \mathbb{E}\Big(f_0(x) - cf'(x)\Big)^2 = 0$$

We know by similar argument then that $\mathbb{E}[f_0(x)f'(x)] = 0$.

However, $\mathbb{E}[f_0(x)f'(x)] = \mathbb{E}[f_0(x)^2 + rf_0(x)x^\mathsf{T} x] = \mathbb{E}[f_0(x)^2]$. This is a contradiction.

□

Thus, if we tried both convex and concave projection and the solution is zero, then we can be sure that the original function is identically zero.

The same argument applies if we consider the best single component approximation:

$$\operatorname*{arg\,min}_{f_k \in \mathcal{C}_x \cup \mathcal{C}_v} \mathbb{E}\Big(f_0(x) - f_k(x_k)\Big)^2$$

Because $\mathbb{E}[f_0(x)x_k^2] = \mathbb{E}\Big[\mathbb{E}[f_0(x) \,|\, x_k]x_k^2\Big] = \mathbb{E}[f_k^*(x_k)x_k^2]$. Thus, we can apply the same argument and use $f_k^*(x_k)$ where needed.

# 4 1/23/2014

## 4.1 Additive Faithfulness Case Study with Quadratic Function and Gaussian Distribution

We consider a quadratic function $f(x) = x^\top Hx + c^\top x$ and a Gaussian distribution $X \sim N(0, \Sigma)$.

We then have a closed form for the additive approximation.

- If $f$ does not depend on $x_j$, then $f_j^*(x_j) = 0$.
- If $f$ depends on $x_j$, then, letting $H_j$ be the $j$-th row of $H$ and $\Sigma_j$ be the $j$-th row of $\Sigma$:

$$f_j^*(x_j) = H_j^\top \Sigma_j \frac{1}{\Sigma_{jj}} x_j^2 + c_j x_j$$

Let us assume $\Sigma_{jj} = 1$ for all $j$ and that $c = 0$ for convenience. We then have two direct corollaries:

**Corollary:** We have additive convexity if and only if $\text{diag}(H\Sigma) \geq 0$.

**Corollary:** We have additive faithfulness if and only if $\text{diag}(H\Sigma) \neq 0$.

As an example where additive convexity and additive faithfulness are violated. Let $H = [1, 2; 2, 5]$ and $\Sigma = [1, -c; -c, 1]$. For $c = 0.5$, additive faithfulness is violated; for $c > 0.5$, additive convexity is violated.

*Proof.* We will show that the $f_j^*(x_j)$'s, so described, satisfy the KKT stationarity equations.

$$f_j^*(x_j) = \mathbb{E}[f(x) - \sum_{k \neq j} f_k^*(x_k) \,|\, x_j] \qquad \text{for all } j$$

To prove this, we use the following conditional mean and conditional covariance property of the multivariate Gaussian distribution.

$$\begin{aligned}
&\mathbb{E}[x_k \,|\, x_j] = \Sigma_{jk} \Sigma_{jj}^{-1} x_j \\
&\mathbb{E}[x_k x_{k'} \,|\, x_j] \text{ is some constant for all } x_j \\
&\mathbb{E}[x_k^2 \,|\, x_j] \text{ is some constant for all } x_j
\end{aligned}$$

□

Why can Gaussian distribution violate additive faithfulness? Because $\frac{\partial p(x_{-j} \,|\, x_j)}{\partial x_j}$ is always large for some values of $x_{-j}$.

# 5 1/2/2014

## 5.1 Convex-minus-Quadratic Estimation

Instead of estimating a convex-plus-concave function, it is theoretically sufficient to estimate a convex-minus-quadratic function.

**Theorem 5.1.** *Any function $h : \mathbb{R}^d \to \mathbb{R}$ with a bounded Hessian can be decomposed as $h(x) = f(x) - q(x)$ where $f(x)$ is convex and $q(x)$ is $cx^\mathsf{T}x$ for some $c \geq 0$.*

*Proof.* For large enough $c$, the Hessian of $h + c$ is a positive semidefinite. The function $f$ is thus convex. □

Convex-minus-quadratic functions are faster to learn. The optimization program has about twice as few variables. It is also possibly easier to analyze.

### 5.1.1 Implementation

For additive modeling, I use the following optimization program. There are two parameters $\lambda$ and $L$.

$$\begin{aligned} \min_{f_1,\ldots,f_p} \frac{1}{n} \sum_i \Big( y_i - \sum_j f_j(x_{ij}) + c_j x_{ij}^2 \Big)^2 + \lambda \sum_j \|\partial f_j - 2c_j x_j^2\|_\infty \\ \text{s.t. } f_j \text{ is convex} \\ c_j \geq 0 \quad \text{and} \quad c_j \leq L \end{aligned}$$

The second $L$ parameter is necessary. A similar parameter is required in convex-plus-concave estimation as well, which I will explain in the next section. $L$ can be interpreted as a lower bound on the second derivative of the estimated regression function. We set $L = 200$ in experiments.

**Demonstration:** We estimate a one-dimensional function so that we can visualize the behavior of convex-minus-quadratic functions. We set $lambda = 0$. The result is in figure 5.1.1.

Here, $n = 300$, the $SNR$ is about 0.6 (high noise).

## 5.2 Tuning Parameters for Convex-plus-Concave Estimation

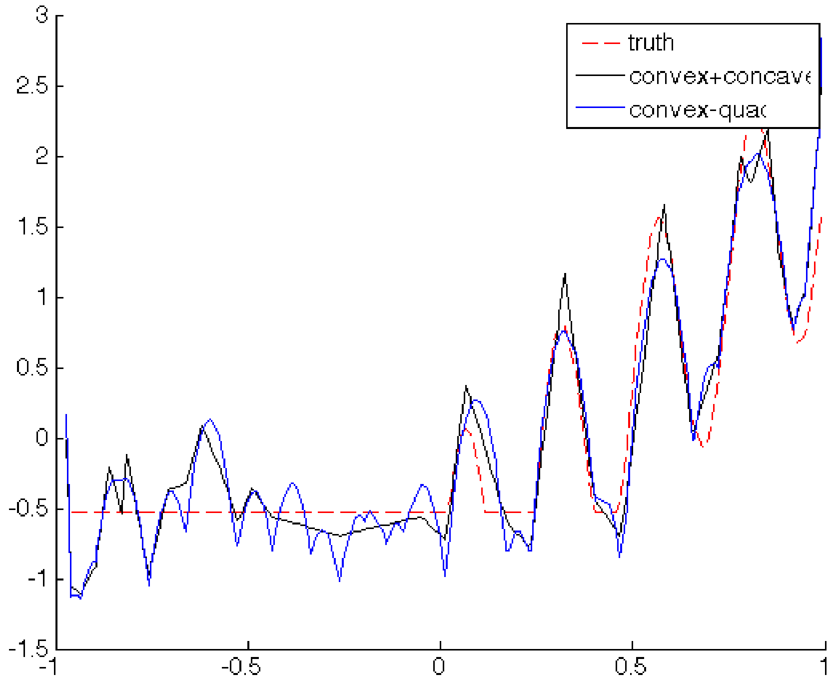
Convex-plus-concave estimation is not free of tuning parameters. Arbitrary sum of convex and concave functions can represent any function with a bounded Hessian.

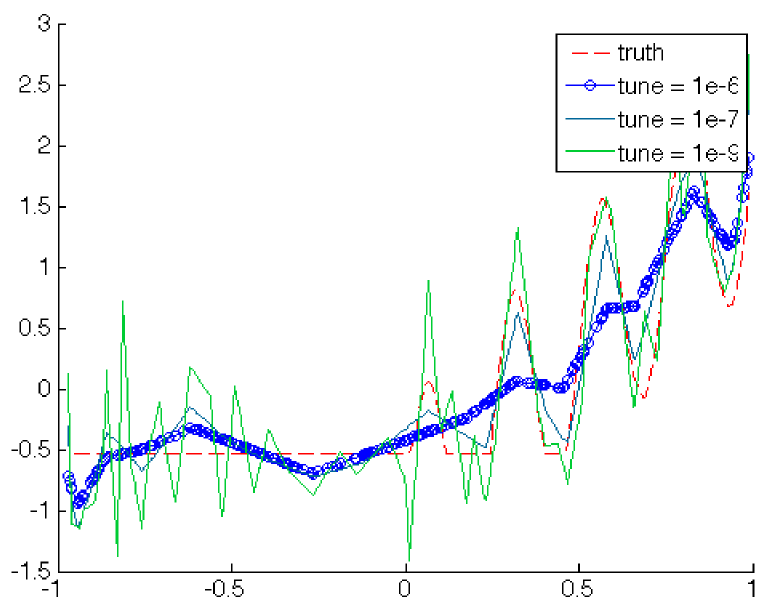
In Minhua's original SCCAM implementation, the objective is augmented with a square-penalty on the value of the gradients of $f$ and $g$, the convex and the concave functions. The penalty looks like $10^{-6}\|\text{gradient}\|_2^2$. This penalty is not just for numerical stability; its magnitude affects the estimation.

**Demonstration:**

The experimental set-up is the same as before. The result is in figure 5.2

As one can see, when the tuning parameter (the coefficient for the $\|\text{gradient}\|_2^2$ penalty) is too small, the estimated function is fitting noise. When the tuning parameter is too large, the estimated function is too smooth.


truth
convex+concave
convex-quad
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-1
-0.5
0
0.5
1


truth
tune = 1e-6
tune = 1e-7
tune = 1e-9
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-1
-0.5
0
0.5
1

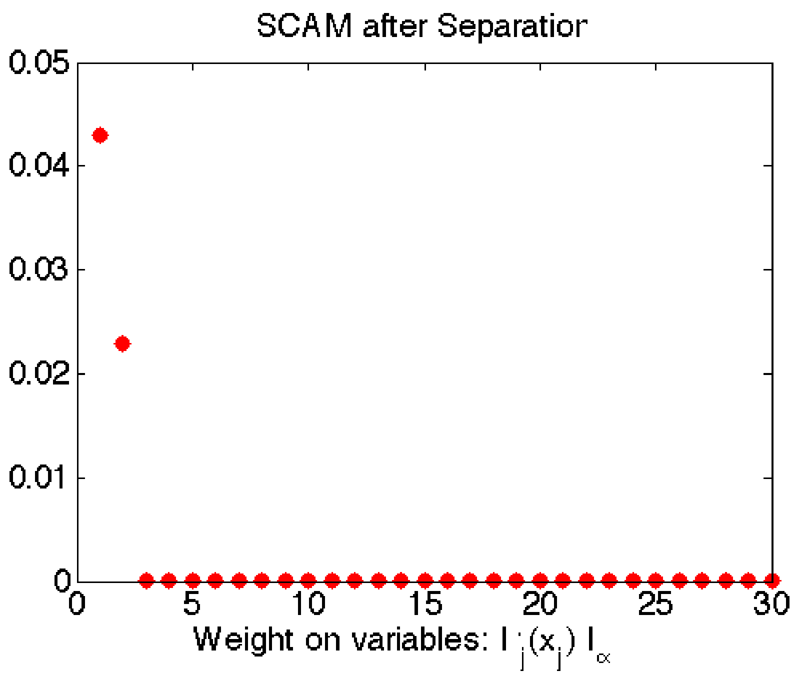
## 5.3 Convex-Concave Separation

We study the problem of variable selection on a convex-plus-concave function $h$. The approach has two steps:

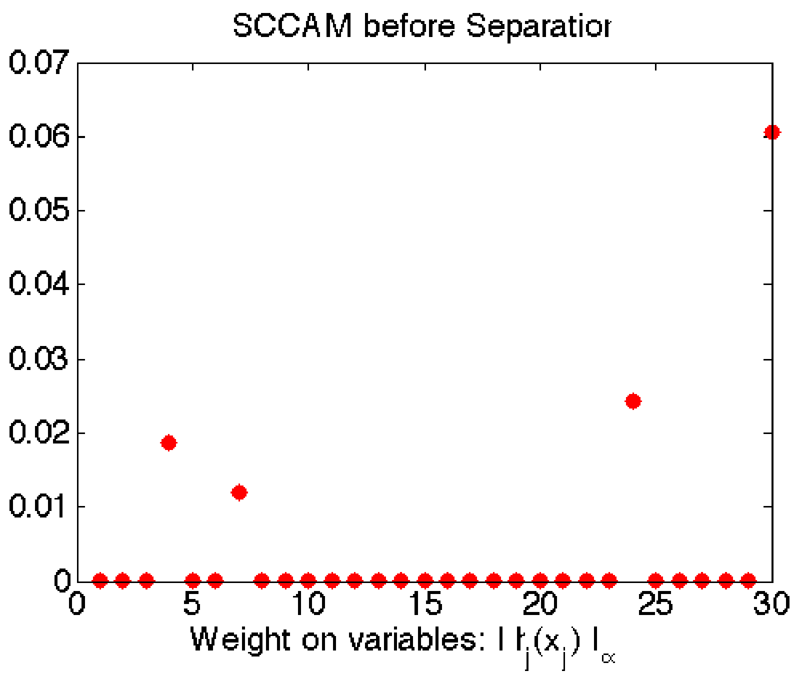
1. Learn **sparse non-additive** convex function $f$ and concave function $g$ such that $h = f + g$. This is the separation stage.
2. Apply sparse additive model on $f$ and $g$.

Preliminary results are favorable. In our experiment, we let $p = 30, n = 300$. The true function is $h(x) = 2x_1x_2$ and uses only the first two variables. The SNR is 4. $h$ is an example of a function that cannot be consistently estimated by additive modeling.

We can indeed identify the correct variables by applying additive model on $f$ after separation.




Without separation however, the additive model indeed fails.


SCCAM before Separatior
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
5
10
15
20
25
30
Weight on variables: I hj(xj) Iα