Supplement to "Convex Hierarchical Testing of Interactions"

Jacob Bien*, Noah Simon †, and Robert Tibshirani‡

October 26, 2014

1 A detailed look at the optimization problem (2.3) in the main paper

1.1 Roadmap for supplement

Our procedure is based on the following optimization problem:

$$\begin{aligned}\text{Minimize}\quad & \frac{1}{2}\sum_{j=1}^{p}(w_j-(\beta_j^+-\beta_j^-))^2+\frac{1}{2}\sum_{j=1}^{p}\sum_{j\neq k}(z_{jk}-\theta_{jk})^2+\lambda_1\sum_{j=1}^{p}[\beta_j^++\beta_j^-]+\lambda_2\sum_{j=1}^{p}\sum_{k\neq j}|\theta_{jk}|\\ & \text{s.t. }\ \beta_j^{\pm}\geq 0,\sum_{k\neq j}|\theta_{jk}|\leq\beta_j^++\beta_j^-\text{ for } j=1,\ldots,p.\end{aligned}$$

Although we take $\lambda_1=\lambda_2=\lambda$ for our testing procedure, we study this slightly more general version of the optimization problem in the supplement. Observe that this problem decouples into p separate problems, one for each j, involving the variables $(\beta_j^+,\beta_j^-,\theta_{j\cdot})$. For notational simplicity, let $w\equiv w_j$, $z\equiv(z_{j1},\ldots,z_{j,j-1},z_{j,j+1},\ldots,z_{jp})^T$, $\beta^{\pm}\equiv\beta_j^{\pm}$, and $\theta\equiv(\theta_{j1},\ldots,\theta_{j,j-1},\theta_{j,j+1},\ldots,\theta_{jp})^T$. We adhere to this notation for the rest of Section 1, and then for clarity in Section 2 we rephrase our results in terms of the notation of the original problem. The jth problem, whose solution is $(\hat\beta_j^+,\hat\beta_j^-,\hat\theta_{j\cdot})$, is

$$\begin{aligned}\text{Minimize}_{\beta^{\pm}\in\mathbb{R},\ \theta\in\mathbb{R}^{p-1}}\quad & \frac{1}{2}(w-(\beta^+-\beta^-))^2+\frac{1}{2}\|z-\theta\|^2+\lambda_1(\beta^++\beta^-)+\lambda_2\|\theta\|_1\\ & \text{s.t. }\ \beta^{\pm}\geq 0,\|\theta\|_1\leq\beta^++\beta^-.\end{aligned}\tag{1}$$

Because our original problem decouples into p separate problems involving the jth main effect and the associated $p-1$ interactions with that variable, we will study problem (1) throughout Section 1. Once this "jth-row" problem is well-understood, we will by direct

*Departments of Biological Statistics and Computational Biology and Statistical Science, Cornell University, jbien@cornell.edu

†Department of Biostatistics, University of Washington, nrsimon@uw.edu

‡Departments of Health, Research & Policy, and Statistics, Stanford University, tibs@stanford.edu

extension have studied the original problem. Therefore, for the rest of Section 1 the "inputs" to the problem will be thought of as $w \in \mathbb{R}$ and $z \in \mathbb{R}^{p-1}$ and the solution is denoted by $(\hat{\beta}^+, \hat{\beta}^-, \hat{\theta}) \in \mathbb{R}^{2+(p-1)}$, which in the rest of the paper is denoted by $(\hat{\beta}_j^+, \hat{\beta}_j^-, \hat{\theta}_{j\cdot})$. In Section 2, we apply the results of Section 1 to the paper's main problem. Here is an overview of the main elements of Section 1:

- We write out the Karush-Kuhn Tucker (KKT) conditions for problem (1) and observe that this leads to a very simple characterization of the (primal) solution in terms of the optimal dual variables $(\hat{\gamma}^+, \hat{\gamma}^-, \hat{\alpha})$.
- We prove that the solution is unique (Lemma 2). This is important for establishing that our test statistics are well-defined (since they are defined as the largest λ for which a primal variable becomes nonzero).
- We characterize the solution path in Proposition 1. It turns out that the path is conveniently described by distinguishing among three cases: the first is the "Big main effect" case, in which $w > \|z\|_1$; the second is the "Moderate main effect" case, in which $\|z\|_\infty \leq w \leq \|z\|_1$; the third is the "Big interaction" case, in which $w < \|z\|_\infty$. The path is described in terms of some special values of λ that are defined in Lemma 3. This Lemma breaks down by case whether these values are finite and if so the relative size of these values. The description of the path in Proposition 1 is not completely closed-form, but it turns out that our characterization of the path is precise enough for our purposes.
- In particular, our test statistics only depend on when the solution paths become nonzero. Proposition 3 describes these values in closed form and Corollary 1, which is the ultimate goal of Section 1, provides a very simple expression for the points at which the solution path becomes nonzero.

1.2 KKT conditions, a primal-dual relation, and uniqueness

The Lagrangian associated with problem (1) is

$$L(\beta^\pm, \theta; \gamma^\pm, \alpha) = \frac{1}{2}(w - (\beta^+ - \beta^-))^2 + \frac{1}{2}\|z - \theta\|^2 + (\lambda_2 + \alpha)\|\theta\|_1 \\ + (\lambda_1 - \gamma^+ - \alpha)\beta^+ + (\lambda_1 - \gamma^- - \alpha)\beta^-,$$

where $\gamma^{\pm}, \alpha \geq 0$ are dual variables corresponding to the constraints. The KKT conditions are

$$\begin{aligned}
(\beta^+ - \beta^- - w) + \lambda_1 - \gamma^+ - \alpha &= 0 \\
-(\beta^+ - \beta^- - w) + \lambda_1 - \gamma^- - \alpha &= 0 \\
\theta - z + (\lambda_2 + \alpha)s &= 0 \\
\gamma^+ \beta^+ &= 0 \\
\gamma^- \beta^- &= 0 \\
\alpha(\|\theta\|_1 - \beta^+ - \beta^-) &= 0 \\
\|\theta\|_1 \leq \beta^+ + \beta^-; \quad \beta^+, \beta^- &\geq 0 \\
\gamma^+, \gamma^-, \alpha &\geq 0
\end{aligned}$$

where s_k is a subgradient of $|\theta_k|$ evaluated at the solution.

Lemma 1. *The solution to* (1) *may be written as*

$$\hat{\beta}^+ - \hat{\beta}^- = w + (\hat{\gamma}^+ - \hat{\gamma}^-)/2$$
$$\hat{\theta} = \mathcal{S}(z, \ \lambda_2 + \hat{\alpha})$$

Furthermore, the first expression can be written as

$$\hat{\beta}^+ - \hat{\beta}^- = \mathcal{S}(w, \lambda_1 - \hat{\alpha}).$$

Proof. The stationarity conditions (first three lines of the optimality conditions) imply the characterization of $\hat{\theta}$ and the first characterization of $\hat{\beta}^+ - \hat{\beta}^-$. The second characterization of $\hat{\beta}^+ - \hat{\beta}^-$ follows from verifying four cases:

- $\hat{\beta}^+ > 0, \hat{\beta}^- > 0$. In this case, $\hat{\gamma}^+ = \hat{\gamma}^- = 0$ and thus $\hat{\alpha} = \lambda_1$. Therefore, $\hat{\beta}^+ - \hat{\beta}^- = w = \mathcal{S}(w, \lambda_1 - \hat{\alpha})$.
- $\hat{\beta}^+ > 0, \hat{\beta}^- = 0$. Then $\hat{\gamma}^+ = 0$ and $\hat{\beta}^+ - w + \lambda_1 - \hat{\alpha} = 0$. That is, $\hat{\beta}^+ - \hat{\beta}^- = w - (\lambda_1 - \hat{\alpha}) = \mathcal{S}(w, \lambda_1 - \hat{\alpha})$. The last equality follows from $\hat{\beta}^+ - (w - \lambda_1 - \hat{\alpha}) = \hat{\gamma}^- \geq 0$.
- $\hat{\beta}^- > 0, \hat{\beta}^+ = 0$. Identical argument as above. $\hat{\beta}^+ - \hat{\beta}^- = w + (\lambda_1 - \hat{\alpha}) = \mathcal{S}(w, \lambda_1 - \hat{\alpha})$.
- $\hat{\beta}^+ = \hat{\beta}^- = 0$. Then $\pm w + \lambda_1 - \hat{\alpha} \geq 0$ or $|w| \leq \lambda_1 - \hat{\alpha}$. So $\mathcal{S}(w, \lambda_1 - \hat{\alpha}) = 0 = \hat{\beta}^+ - \hat{\beta}^-$.

□

Lemma 2. *Assume $\lambda_1 > 0$. Then, the solution to* (1) *is unique.*

Proof. Let $(\hat{\beta}^+, \hat{\beta}^-, \hat{\theta})$ be a solution to (1) with associated dual variables $(\hat{\alpha}, \hat{\gamma}^{\pm})$. Since (1) is convex, it suffices to show that the solution is unique in a neighborhood around this point. Noting that the objective function is strongly convex in all directions except for $(\beta^+, \beta^-, \theta)$ for which $\beta^+ - \beta^- = \hat{\beta}^+ - \hat{\beta}^-$, it remains to consider perturbations of the optimal point of the form $(\beta^+, \beta^-, \theta) = (\hat{\beta}^+ + \epsilon, \hat{\beta}^- + \epsilon, \hat{\theta} + \delta)$. If this new point is a solution, it must have a corresponding set of dual variables $(\gamma^+, \gamma^-, \alpha)$ satisfying the KKT conditions. We begin by showing in all cases that $\epsilon \neq 0$ implies $\theta = \hat{\theta}$:

- $\hat{\beta}^+ > 0, \hat{\beta}^- > 0$: In this case, $\hat{\gamma}^+ = \hat{\gamma}^- = 0$ and $\hat{\alpha} = \lambda_1$. Choosing $|\epsilon|$ small enough such that $\beta^+, \beta^- > 0$, we must have $\gamma^+ = \gamma^- = 0$ and so $\alpha = \lambda_1$. It follows that $\theta = \hat{\theta}$.

- $\hat{\beta}^+ > 0, \hat{\beta}^- = 0$: In this case, $\hat{\gamma}^+ = 0$, and $\epsilon \geq 0$ for our perturbation to be feasible. If $\epsilon > 0$, then $\beta^+, \beta^- > 0$ and so $\gamma^+ = \gamma^- = 0$ and $\alpha = \lambda_1$. Now, $\beta^+ - \beta^- = \hat{\beta}^+$, so the first KKT condition for this new point is $(\hat{\beta}^+ - w) + \lambda_1 - \alpha = 0$; however, the first KKT condition for our original point is $(\hat{\beta}^+ - w) + \lambda_1 - \hat{\alpha} = 0$. This implies that $\hat{\alpha} = \alpha$, and so $\hat{\theta} = \theta$.

- $\hat{\beta}^+ = 0, \hat{\beta}^- > 0$: Identical argument to previous case.

- $\hat{\beta}^+ = \hat{\beta}^- = 0$: Again, $\epsilon \geq 0$ is required for new point to be feasible. If $\epsilon > 0$, then $\gamma^+ = \gamma^- = 0$ and $\alpha = \lambda_1$. On the other hand, the KKT conditions for the original point implies $\hat{\alpha} \leq \lambda_1 - |w|$ (since $\pm w + \lambda_1 - \hat{\alpha} = \hat{\gamma}^\pm \geq 0$). Thus, $\alpha \geq \hat{\alpha}$. Now, $\hat{\theta} = 0$ means that $\|z\|_\infty \leq \lambda_2 + \hat{\alpha} \leq \lambda_2 + \alpha$, and so $\theta = 0$ as well.

Thus, our new point is $(\beta^+, \beta^-, \theta) = (\hat{\beta}^+ + \epsilon, \hat{\beta}^- + \epsilon, \hat{\theta})$, and so the objective value at the two points differs by $2\lambda_1\epsilon$. It follows that if $\epsilon \neq 0$, then one of the points is not optimal. Therefore, (1) has a unique solution. □

1.3 Characterizing the path

We take $\lambda_1 = \lambda_2 = \lambda$, and now consider the path of solutions generated by varying $\lambda > 0$. Note that by Lemma 2, it makes sense to speak of "the" path. We assume, without loss of generality, that $w \geq 0$. In studying the path, as we do in the proof of Proposition 1, we find that there are three "special" values of λ at which the behavior of the path changes. These are the following:

$$\begin{aligned}
\tilde{\lambda}_1 &= \min\{\lambda \geq 0 : \|\mathcal{S}(z, \lambda)\|_1 + \lambda \leq w\}, \\
\tilde{\lambda}_2 &= \max\{\lambda \geq 0 : \|\mathcal{S}(z, \lambda)\|_1 + \lambda \leq w\}, \\
\tilde{\lambda}_3 &= \max\{\lambda \geq 0 : \|\mathcal{S}(z, 2\lambda)\|_1 \geq w\}.
\end{aligned}$$

In Lemma 3, we collect some facts about $\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3$ that will be useful in the rest of the supplement; however, most readers may prefer to skip directly to Proposition 1, which describes the solution path.

Lemma 3. *Defining $|z_1| \geq |z_2| \geq \cdots$, the following statements hold:*

1. *$\|z\|_\infty \leq w$ iff. $\tilde{\lambda}_1, \tilde{\lambda}_2$ are finite, in which case $\tilde{\lambda}_1 \leq |z_2|$ and $\tilde{\lambda}_2 = w$.*

2. *$\|z\|_1 \geq w > 0$ iff. $\tilde{\lambda}_3$ finite, in which case $\tilde{\lambda}_3 \leq (1 - w/\|z\|_1)\|z\|_\infty/2$.*

3. *If $\|z\|_\infty \leq w \leq \|z\|_1$, then $\tilde{\lambda}_3 \leq \tilde{\lambda}_1 \leq \tilde{\lambda}_2$.*

Proof. The function $f_1(\lambda) = \|\mathcal{S}(z,\lambda)\|_1 + \lambda$ is piecewise-linear, convex, and minimized on $\lambda \in [|z_2|, |z_1|]$ with minimal value $|z_1|$ (where $|z_1| \geq |z_2| \geq \cdots$). Thus, $[\tilde{\lambda}_1, \tilde{\lambda}_2]$ defines a nonempty interval iff. $\|z\|_\infty \leq w$. If this holds, then $\tilde{\lambda}_1 \leq |z_2|$ and $\tilde{\lambda}_2 = w$. Likewise, $\tilde{\lambda}_3$ is finite iff. $\|z\|_1 \geq w$.

The function $f_2(\lambda) = \|\mathcal{S}(z, 2\lambda)\|_1$ is a piecewise linear, decreasing convex function with $f_2(0) = \|z\|_1$ and $f_2(\|z\|_\infty/2) = 0$. Since f_2 is convex, it lies beneath the line $L(\lambda) = \|z\|_1 - 2(\|z\|_1/\|z\|_\infty)\lambda$ on the interval $\lambda \in [0, \|z\|_\infty/2]$ and so $\tilde{\lambda}_3 \leq \tilde{\lambda}$ where $L(\tilde{\lambda}) = w$, i.e. $\tilde{\lambda} = (1 - w/\|z\|_1)\|z\|_\infty/2$.

Finally, observe that $f_1(\lambda) - f_2(\lambda) = \|\mathcal{S}(z,\lambda)\|_1 - \|\mathcal{S}(z,2\lambda)\|_1 + \lambda \geq \lambda$ since $\|\mathcal{S}(z,\cdot)\|_1$ is a non-increasing function and $2\lambda \geq \lambda$. Thus, for $\lambda > 0$, $f_1(\lambda) > f_2(\lambda)$ from which it follows that $\tilde{\lambda}_3 < \tilde{\lambda}_1$. We can only have equality if $\tilde{\lambda}_1 = \tilde{\lambda}_3 = 0$, in which case $\|z\|_1 = \|z\|_\infty = w$ (i.e. z has no more than one nonzero value), which is a 0 probability event under most reasonable models. $\square$

Proposition 1. *Let $\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3$ be as defined above, let $\tilde{\lambda}_4 = (w + \|z\|_\infty)/2$, and write $\hat{\beta} = \hat{\beta}^+ - \hat{\beta}^-$. The solution path of (1) depends on the relative size of the main effects and interactions and is given by the following:*

1. *Big main effect: $\|z\|_\infty \leq \|z\|_1 < w$.*

$$\begin{aligned}
\lambda \geq w &\implies \hat{\beta} = 0, & \hat{\theta} &= 0 & &\text{[Case III]}\\
\tilde{\lambda}_1 \leq \lambda < w &\implies \hat{\beta} = w - \lambda, & \hat{\theta} &= \mathcal{S}(z,\lambda) & &\text{[Case I(i)]}\\
\lambda < \tilde{\lambda}_1 &\implies \hat{\beta} = w - \lambda + \hat{\alpha}(\lambda), & \hat{\theta} &= \mathcal{S}(z, \lambda + \hat{\alpha}(\lambda)),\ \hat{\alpha}(\lambda) > 0 & &\text{[Case I(ii)(a)C.]}
\end{aligned}$$

2. *Moderate main effect: $\|z\|_\infty \leq w \leq \|z\|_1$.*

$$\begin{aligned}
\lambda \geq w &\implies \hat{\beta} = 0, & \hat{\theta} &= 0 & &\text{[Case III]}\\
\tilde{\lambda}_1 \leq \lambda < w &\implies \hat{\beta} = w - \lambda, & \hat{\theta} &= \mathcal{S}(z,\lambda) & &\text{[Case I(i)]}\\
\tilde{\lambda}_3 \leq \lambda < \tilde{\lambda}_1 &\implies \hat{\beta} = w - \lambda + \hat{\alpha}(\lambda), & \hat{\theta} &= \mathcal{S}(z, \lambda + \hat{\alpha}(\lambda)),\ \hat{\alpha}(\lambda) > 0 & &\text{[Case I(ii)(a)B.]}\\
\lambda < \tilde{\lambda}_3 &\implies \hat{\beta} = w, & \hat{\theta} &= \mathcal{S}(z, 2\lambda) & &\text{[Case II]}
\end{aligned}$$

3. *Big interaction: $w < \|z\|_\infty \leq \|z\|_1$.*

$$\begin{aligned}
\lambda \geq \tilde{\lambda}_4 &\implies \hat{\beta} = 0, & \hat{\theta} &= 0 & &\text{[Case III]}\\
\tilde{\lambda}_3 \leq \lambda < \tilde{\lambda}_4 &\implies \hat{\beta} = w - \lambda + \hat{\alpha}(\lambda), & \hat{\theta} &= \mathcal{S}(z, \lambda + \hat{\alpha}(\lambda)),\ \hat{\alpha}(\lambda) > 0 & &\text{[Case I(ii)(b) + (a)A.]}\\
\lambda < \tilde{\lambda}_3 &\implies \hat{\beta} = w, & \hat{\theta} &= \mathcal{S}(z, 2\lambda) & &\text{[Case II]}
\end{aligned}$$

Proof. We partition the solution path as follows:

Case I. $\hat{\beta}^+ > 0, \hat{\beta}^- = 0$.

Case II. $\hat{\beta}^+ > 0, \hat{\beta}^- > 0$.

Case III. $\hat{\beta}^+ = 0, \hat{\beta}^- = 0$.

Case IV. $\hat{\beta}^+ = 0, \hat{\beta}^- > 0$.

By Lemma 2, this is a well-defined partition of the path (i.e., for a given $\lambda > 0$, only one of the following cases holds).

Case I. $\hat{\beta}^+ > 0, \hat{\beta}^- = 0$.

The KKT conditions imply that $\hat{\gamma}^+ = 0$, $\hat{\beta}^+ = w - \lambda + \hat{\alpha}$, and $\hat{\gamma}^- = 2(\lambda - \hat{\alpha})$. The conditions $\hat{\gamma}^-, \hat{\alpha} \geq 0$ and $\hat{\beta}^+ > 0$ require that

$$0 \leq \hat{\alpha} \leq \lambda \qquad \hat{\alpha} > \lambda - w$$

and, defining $f_\lambda(\alpha) = \|\mathcal{S}(z, \lambda + \alpha)\|_1 - w + \lambda - \alpha$, they also require that

$$f_\lambda(\hat{\alpha}) \leq 0 \text{ with } \hat{\alpha} f_\lambda(\hat{\alpha}) = 0.$$

Thus, this case occurs iff. ("if" direction by Lemma 2) there exists $\hat{\alpha}$ satisfying these constraints. We consider two subcases:

(i) $\hat{\alpha} = 0$. Requires $\lambda - w < 0$ and $\|\mathcal{S}(z, \lambda)\|_1 + \lambda \leq w$. In light of Lemma 3, this subcase happens iff. $\|z\|_\infty \leq w$ and $\lambda \in [\tilde{\lambda}_1, w)$.

(ii) $\hat{\alpha} > 0$. Requires $f_\lambda(\hat{\alpha}) = 0$. Now, f_λ is a strictly decreasing function, so to show that such an $\hat{\alpha}$ exists with $[\lambda - w]_+ < \hat{\alpha} \leq \lambda$, it suffices to check that

$$\begin{cases} \text{(a) } f_\lambda(0) > 0 \geq f_\lambda(\lambda) & \text{if } \lambda - w < 0 \\ \text{(b) } f_\lambda(\lambda - w) > 0 \geq f_\lambda(\lambda) & \text{if } \lambda - w \geq 0 \end{cases}$$

Now, $0 \geq f_\lambda(\lambda) = \|\mathcal{S}(z, 2\lambda)\|_1 - w$, so by Lemma 3, this constraint is satisfied for all λ if $\|z\|_1 < w$ and for $\lambda \geq \tilde{\lambda}_3$ when $\|z\|_1 \geq w$.

(a) $f_\lambda(0) > 0 \iff \|\mathcal{S}(z, \lambda)\|_1 + \lambda > w$, which (in light of Lemma 3) is satisfied for all λ when $\|z\|_\infty > w$ and for $\lambda \notin [\tilde{\lambda}_1, \tilde{\lambda}_2]$ when $\|z\|_\infty \leq w$. Incorporating this with the constraints from $0 \geq f_\lambda(\lambda)$ and $\lambda < w$ (and using that $\|z\|_\infty \leq \|z\|_1$ and $\tilde{\lambda}_2 = w$) gives the following cases in which (a) holds:

A. $\|z\|_\infty > w$, $\tilde{\lambda}_3 \leq \lambda < w$,

B. $\|z\|_\infty \leq w \leq \|z\|_1$, $\tilde{\lambda}_3 \leq \lambda < \tilde{\lambda}_1$

C. $\|z\|_1 < w$, $\lambda < \tilde{\lambda}_1$

(b) We have $\lambda \geq w$ and $0 < f_\lambda(\lambda - w) = \|\mathcal{S}(z, 2\lambda - w)\|_1$, which holds as long as $\|z\|_\infty > 2\lambda - w$, i.e. $\lambda < (\|z\|_\infty + w)/2$.

A. $\|z\|_1 < w \leq \lambda < (\|z\|_\infty + w)/2$. But this case can never occur since it would imply $\max\{w, \|z\|_1\} < (\|z\|_\infty + w)/2 \leq (\|z\|_1 + w)/2$.

B. $\|z\|_1 \geq w$, $\lambda \geq \tilde{\lambda}_3$, $w \leq \lambda < (\|z\|_\infty + w)/2$, Case II. $\hat\beta^+ > 0, \hat\beta^- > 0$.

In this case, $\hat\gamma^+ = \hat\gamma^- = 0$, so we have $\hat\beta^+ - \hat\beta^- = w$ and $\hat\alpha = \lambda$. Now, $\hat\beta^+ = w + \hat\beta^- > w$ and so $\hat\beta^+ + \hat\beta^- > w$. The only remaining requirement of the KKT conditions is that $\|\mathcal{S}(z, 2\lambda)\|_1 = \hat\beta^+ + \hat\beta^- > w$. This case occurs iff. (again, we use Lemma 2 for the "if" direction) $\|\mathcal{S}(z, 2\lambda)\|_1 > w$, which by Lemma 3 is equivalent to $\|z\|_1 \geq w$ and $\lambda < \tilde\lambda_3$ (note that $\lambda = \tilde\lambda_3$ could only occur when $w = 0$, a case we exclude).

Case III. $\hat\beta^+ = 0, \hat\beta^- = 0$.

In this case, $\hat\gamma^\pm = \lambda \pm w - \hat\alpha \geq 0$, and so we require $0 \leq \hat\alpha \leq \lambda - w$ with $\|\mathcal{S}(z, \lambda + \hat\alpha)\|_1 \leq 0$. Putting this together gives

$$[\|z\|_\infty - \lambda]_+ \leq \hat\alpha \leq \lambda - w.$$

This interval is nonempty iff. $[\|z\|_\infty - \lambda]_+ \leq \lambda - w$, i.e.

$$\lambda \geq \max\{w, (\|z\|_\infty + w)/2\}.$$

It is interesting to break this $\hat\beta^+ = \hat\beta^- = 0$ case into two subcases: the case in which the everything is zero even without the hierarchy constraint and the case in which the hierarchy constraint is the "reason" for everything being zero (in other words $\hat\alpha > 0$ and we wouldn't be in this case if $\hat\alpha = 0$). The hierarchy-active case occurs when $\|z\|_\infty - \lambda > 0$. In light of the lower bound on λ given above, active hierarchy sets everything to zero specifically when $\|z\|_\infty > w$ and $\lambda \geq (\|z\|_\infty + w)/2$.

Case IV. $\hat\beta^+ = 0, \hat\beta^- > 0$.

The KKT conditions imply that $\hat\gamma^- = 0$, $\hat\beta^- = \hat\alpha - (w + \lambda)$ and $\hat\gamma^+ = 2(\lambda - \hat\alpha) \geq 0$. Putting these together gives $\hat\alpha \leq \lambda$ and $\hat\beta^- < -w \leq 0$. In other words, this case does not occur!

We summarize these cases more succinctly:

Case I. $\hat\beta^+ > 0, \hat\beta^- = 0$.

- (i) $\hat\alpha = 0$ iff. $\|z\|_\infty \leq w$ and $\tilde\lambda_1 \leq \lambda < w$.
- (ii) $\hat\alpha > 0$ iff.
  - (a) A. $\|z\|_\infty > w$, $\tilde\lambda_3 \leq \lambda < w$,
    B. $\|z\|_\infty \leq w \leq \|z\|_1$, $\tilde\lambda_3 \leq \lambda < \tilde\lambda_1$
    C. $\|z\|_1 < w$, $\lambda < \tilde\lambda_1$
  - (b) $\|z\|_1 \geq w$, $\max\{w, \tilde\lambda_3\} \leq \lambda < (\|z\|_\infty + w)/2$,

Case II. $\hat{\beta}^+ > 0, \hat{\beta}^- > 0$.

iff. $\|z\|_1 \geq w$ and $\lambda < \tilde{\lambda}_3$ (note that $\lambda = \tilde{\lambda}_3$ could only occur when $w = 0$, a case we exclude).

Case III. $\hat{\beta}^+ = 0, \hat{\beta}^- = 0$.

iff.

$$\lambda \geq \max\{w, (\|z\|_\infty + w)/2\}.$$

The hierarchy-active case occurs iff. $\|z\|_\infty > w$ and $\lambda \geq (\|z\|_\infty + w)/2$.

Case IV. $\hat{\beta}^+ = 0, \hat{\beta}^- > 0$.

Does not occur!

By rearranging these cases depending on the relative sizes of w, $\|z\|_\infty$, and $\|z\|_1$, we get the paths given in the Proposition statement. $\square$

Proposition 2. *The solution to* (1) *has* $|\hat{\theta}_k(\lambda)|$ *non-increasing in* λ *for each* k*.*

Proof. Based on Proposition 1, this statement is immediate except when $\hat{\alpha} > 0$, i.e. Case I(ii). Referring back to this case in the proof of Proposition 1, let $\hat{\alpha}(\lambda)$ be the unique[1] point for which $f_\lambda(\hat{\alpha}(\lambda)) = 0$. Defining $t(\lambda) = \lambda + \hat{\alpha}(\lambda)$, we have

$$f_\lambda(\hat{\alpha}(\lambda)) = 0 \iff \|\mathcal{S}(z, t(\lambda)\|_1 - t(\lambda) + \lambda = w.$$

This must hold for λ over the range in which $\hat{\alpha}(\lambda) > 0$ (as specified in Proposition 1). Since $\|\mathcal{S}(z, t)\|_1 - t$ is a strictly decreasing function of t, it follows that increasing λ requires an increase in $t(\lambda)$ for λ. Since $\hat{\theta} = \mathcal{S}(z, t(\lambda))$, this proves that $|\hat{\theta}_j|$ is nonincreasing in λ. $\square$

1.4 Where do the paths become nonzero?

The results from the previous section provide sufficient information for us to derive an exact, closed-form expression for the point in the path at which each coefficient becomes nonzero.

Proposition 3. *Recall that* $\tilde{\lambda}_1 = \min\{\lambda \geq 0 : \|\mathcal{S}(z, \lambda)\|_1 + \lambda \leq w\}$*. The main effects and interactions become nonzero at*

$$\hat{\nu} = \sup\{\lambda : |\hat{\beta}| \neq 0\}$$
$$\hat{\nu}_k = \sup\{\lambda : |\hat{\theta}_k| \neq 0\},$$

which have the following "closed form" values:

[1]Since f_λ is strictly decreasing, it has a unique root.

1. *Big main effect:* $\|z\|_\infty \le \|z\|_1 < w$.

$$\hat{\nu} = w$$
$$\hat{\nu}_k = \begin{cases} |z_k| & \text{for } |z_k| \ge \tilde{\lambda}_1 \\ (|z_k| + w - \|\mathcal{S}(z,|z_k|)\|_1)/2 & \text{for } |z_k| < \tilde{\lambda}_1 \end{cases}$$

2. *Moderate main effect:* $\|z\|_\infty \le w \le \|z\|_1$.

$$\hat{\nu} = w$$
$$\hat{\nu}_k = \begin{cases} |z_k| & \text{for } |z_k| \ge \tilde{\lambda}_1 \\ |z_k|/2 + [w - \|\mathcal{S}(z,|z_k|)\|_1]_+/2 & \text{for } |z_k| < \tilde{\lambda}_1 \end{cases}$$

3. *Big interaction:* $w < \|z\|_\infty \le \|z\|_1$.

$$\hat{\nu} = (w + \|z\|_\infty)/2$$
$$\hat{\nu}_k = |z_k|/2 + [w - \|\mathcal{S}(z,|z_k|)\|_1]_+/2$$

Proof. The expressions for $\hat{\nu}$ and $\hat{\nu}_k$ follow from Proposition 1. The only parts that are not immediate are the cases in which $\hat{\alpha}(\lambda) > 0$.

We begin by considering the "Big main effect" when $\lambda < \tilde{\lambda}_1$. By definition of $\hat{\nu}_k$ and by the expression for $\hat{\theta}$ in this case, we see that $|z_k| = \hat{\nu}_k + \hat{\alpha}(\hat{\nu}_k)$. We also know that $f_{\hat{\nu}_k}(\hat{\alpha}(\hat{\nu}_k)) = 0$ (since $\hat{\alpha}(\hat{\nu}_k) > 0$). Putting these together gives $\|\mathcal{S}(z,|z_k|)\|_1 - w + 2\hat{\nu}_k - |z_k| = 0$, which we solve for $\hat{\nu}_k$. We also require that $\hat{\alpha}(\hat{\nu}_k) > 0$ and $\hat{\nu}_k \ge 0$, i.e. that $0 \le \hat{\nu}_k < |z_k|$. Notice that $|z_k| < \tilde{\lambda}_1$ implies that $|z_k| > w - \|\mathcal{S}(z,|z_k|)\|_1$, establishing that $\hat{\nu}_k < |z_k|$ for $|z_k| < \tilde{\lambda}_1$. In the "Big main effect" case, it is easy to see that $(|z_k| + w - \|\mathcal{S}(z,|z_k|)\|_1)/2 > 0$ (which implies that $\hat{\nu}_k \ge 0$, as required). This completes the proof for the "Big main effect" case.

In the "Moderate main effect" case, if $|z_k| < \tilde{\lambda}_1$, we know that one of two cases can occur. The logic proceeds identically to the "Big main effect" case, except that we are not guaranteed that $(|z_k| + w - \|\mathcal{S}(z,|z_k|)\|_1)/2 \ge 0$, which must hold for us to have $\hat{\nu}_k \ge 0$. If this does hold, we're done. Assume that instead $(|z_k| + w - \|\mathcal{S}(z,|z_k|)\|_1)/2 < 0$. This implies that $\|\mathcal{S}(z,|z_k|)\|_1 > w$ or equivalently that $|z_k|/2 < \tilde{\lambda}_3$. This means we're in Case II and so $\hat{\nu}_k = |z_k|/2$.

Finally, we consider the "Big interaction" case. By Proposition 1, it is clear that $\hat{\nu}_k \le \tilde{\lambda}_4$ and the expression for $\hat{\nu}_k$ will depend on whether $\hat{\nu}_k \ge \tilde{\lambda}_3$. As above, if $\hat{\nu}_k \ge \tilde{\lambda}_3$, then we would have $\hat{\nu}_k = (|z_k| + w - \|\mathcal{S}(z,|z_k|)\|_1)/2$, which to be valid must fall within $[0,|z_k|)$ (to ensure that $\hat{\nu}_k \ge 0$ and $\hat{\alpha}(\hat{\nu}_k) > 0$). These requirements simplify to $|z_k| > w - \|\mathcal{S}(z,|z_k|)\|_1$ and $|z_k| \ge -(w - \|\mathcal{S}(z,|z_k|)\|_1)$. This first inequality always holds in this case since, as noted in the proof of Proposition 1, $f_1(\lambda) = \|\mathcal{S}(z,\lambda)\|_1 + \lambda \ge \|z\|_\infty > w$. If the second inequality holds, then $\hat{\nu}_k = (|z_k| + w - \|\mathcal{S}(z,|z_k|)\|_1)/2$. Otherwise, we have $\|\mathcal{S}(z,|z_k|)\|_1 > w$, implying that $|z_k|/2 < \tilde{\lambda}_3$ and we're in Case II, i.e. $\hat{\nu}_k = |z_k|/2$. □

Remarks:

- In the "Big main effect" case, only small interactions are modified. In particular, the test statistic of these smallest interactions are made smaller.
- In the "Moderate main effect" case, interactions below a certain threshold are modified, with the very smallest ones being reduced to half their size.
- In the "Big interaction" case, the main effect statistic receives a positive boost and the interactions are reduced. Notice that for $|z_k| = \|z\|_\infty$, we have $\hat{\nu}_k = \hat{\nu}$, i.e. the largest interaction enters at the same time as the main effect.

Corollary 1. *The $\hat{\nu}$ and $\hat{\nu}_k$ defined as in Proposition 3 are given by*

$$\hat{\nu} = \max\left\{w, \frac{w + \|z\|_\infty}{2}\right\}$$
$$\hat{\nu}_k = \min\left\{|z_k|, \frac{|z_k|}{2} + \frac{[w - \|\mathcal{S}(z, |z_k|)\|_1]_+}{2}\right\}$$

Proof. For the "Big main effect" and "Moderate main effect" cases, observe that $|z_k| > |z_k|/2 + [w - \|\mathcal{S}(z, |z_k|)\|_1]_+/2$ is equivalent to $\|\mathcal{S}(z, |z_k|)\|_1 + |z_k| > w$ (assuming $z_k \neq 0$), which can be related directly to $|z_k| < \tilde{\lambda}_1$. For the "Big interaction" case, $\|\mathcal{S}(z, |z_k|)\|_1 + |z_k| \geq \|z\|_\infty > w$ and so $|z_k| > |z_k|/2 + [w - \|\mathcal{S}(z, |z_k|)\|_1]_+/2$ holds automatically. □

Remarks:

- The main effect statistic is boosted when there's at least one large interaction.
- Interaction statistics get reduced by at most a factor of 2. The size of the reduction for $\hat{\nu}_k$ depends only on the size of the main effect and on those interactions that are *larger* (in absolute value) than $|z_k|$. In particular, $\|\mathcal{S}(z, |z_k|)\|_1 = \sum_{\ell:|z_\ell|>|z_k|}(|z_\ell| - |z_k|)$ measures how much larger the other interactions are compared to $|z_k|$, and the larger this value, the more $\hat{\nu}_k$ is reduced.

2 Test statistics

In Section 1, we studied what happens to the jth variable (i.e. its main effect and $p-1$ interactions). In this section, we use the result of Section 1 to obtain a closed-form expression for our test statistics. To do so, we return to the main paper's notation, in which w_j (rather than w) represents the jth main effect, z_{jk} (rather than z_k) represents the jkth interaction, $\hat{\lambda}_j$ (rather than ν) represents the test statistic for the jth main effect and $\hat{\lambda}_{jk}$ (rather than ν_k) represents the test statistics for the jkth interaction.

Proposition 4. *Problem (2.3) of the main paper has a unique solution path,* $\left(\hat{\beta}^{+}(\lambda), \hat{\beta}^{-}(\lambda), \hat{\theta}(\lambda)\right) \in \mathbb{R}^{2p+p(p-1/2)}$ *for* $\lambda > 0$*. The values*

$$\begin{aligned}\hat{\lambda}_j &= \sup\{\lambda \geq 0 : \ \hat{\beta}_j^{+}(\lambda) - \hat{\beta}_j^{-}(\lambda) \neq 0\} \\ \hat{\lambda}_{jk} &= \sup\{\lambda \geq 0 : \ \hat{\theta}_{jk}(\lambda) \neq 0\}\end{aligned}$$

can be expressed in closed-form as

$$\begin{aligned}\hat{\lambda}_j &= \max\left\{|w_j|, \frac{|w_j| + \|z_{j\cdot}\|_\infty}{2}\right\} \\ \hat{\lambda}_{jk} &= \min\left\{|z_{jk}|, \frac{|z_{jk}|}{2} + \frac{[|w_j| - \|\mathcal{S}(z_{j\cdot}, |z_{jk}|)\|_1]_+}{2}\right\},\end{aligned}$$

where $z_{j\cdot} \in \mathbb{R}^{p-1}$ *is the vector of interaction contrasts involving the* j*th variable and* w_j *is the main effect contrast.*

Proof. This follows immediately from Corollary 1. □