Independent hypothesis weighting with total variation regularization via ADMM

Baichen Tan baichen@uchicago.edu

Wolfgang Huber whuber@embl.de

Nikolaos Ignatiadis ignat@uchicago.edu

June, 2023

1 Introduction

Groups $\{1, \ldots, G\}$ with m_g tests, $m = \sum_g m_g$ and $q_g = m_g \pi_g / m$. Per group distribution function (estimated by Grenander in this case) F_g.

Let

$$\mathrm{Fdr}(\boldsymbol{t}) := \frac{\sum_{g=1}^{G} q_g t_g}{\sum_{g=1}^{G} q_g F_g(t_g)}.$$

Also:

$$\mathrm{LinFdr}(\boldsymbol{t}; \alpha) := \sum_{g=1}^{G} q_g (t_g - \alpha F_g(t_g)).$$

This is defined such that LinFdr ≤ 0 if and only if Fdr $\leq \alpha$.

2 Algorithm without regularization

To begin with, the weighted BH procedure seeks to maximize the number of discoveries, with a constraint on the False discovery rate FDR. Therefore, our maximization problem becomes

$$\begin{aligned} \underset{\boldsymbol{t} \in [0,1]^G}{\text{maximize}} \quad & \sum_{g=1}^{G} q_g F_g(t_g) \\ \text{subject to} \quad & \mathrm{Fdr}(\boldsymbol{t}) \leq \alpha. \end{aligned} \tag{1}$$

Our goal is to find the optimal t_g^* that solves this maximization problem. We propose the following algorithm to solve for the optimal t_g^*. We first define

Definition 2.1. Let $\mathcal{J}_g$ be the index set for the p-values in the group g. For each F_g, it is a linear function that has the form

$$F_g(t) = \min_{j \in \mathcal{J}_g} a_j^g + b_j^g t$$

Our set-up: Let $b_1^g > b_2^g > \cdots > b_{|\mathcal{J}_g|}^g$ be the estimated slopes in group g and let $s_j^g, j \in \mathcal{J}_g$ be the points at which the slope changes, i.e., the slope is equal to b_j^g in the interval (s_{j-1}^g, s_j^g). We also define the set

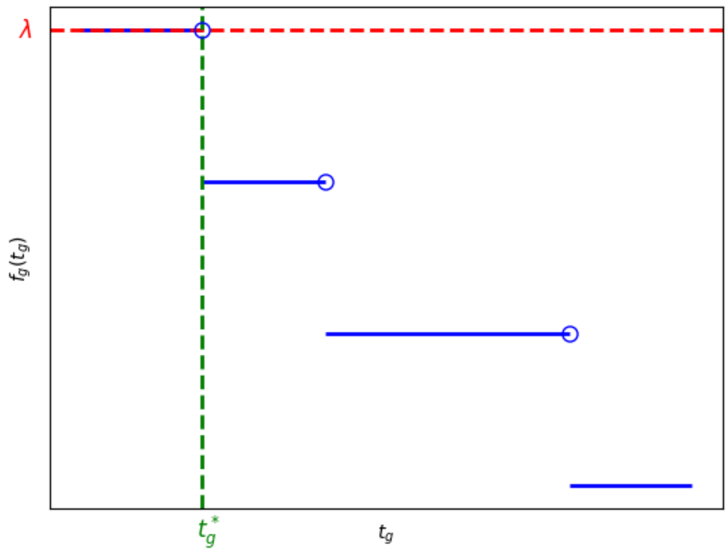
$$\mathcal{B} = \{b_j^g : j \in \mathcal{J}_g, g \in G\}$$

. Then the optimization algorithm is

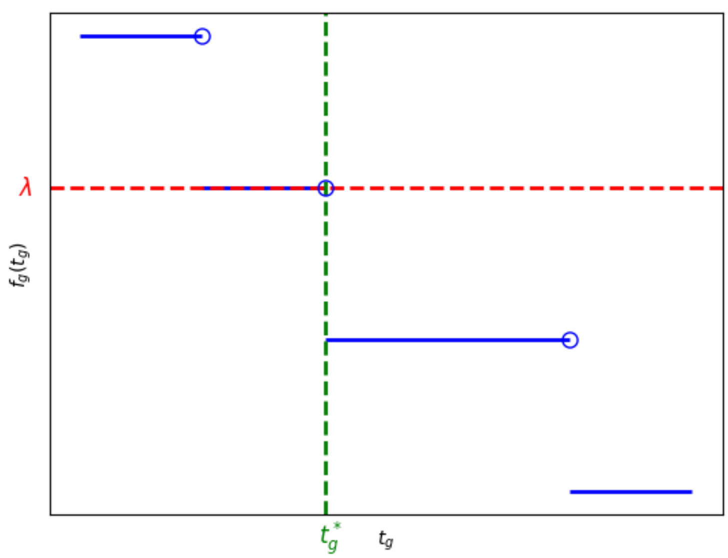
Algorithm 1: Optimization of weighted BH procedure

Data: the nominal level α
the total number of test m
the number of tests within fold l and each group $g \in G$
$m_g =| \{i \in I_l : X_i = g\} |$
the fitted Grenander estimator with slop change points s_j^g and slopes
b_j^g
the set $\mathcal{B}$ defined as in k-Bonferroni Sort the set $\mathcal{B}$
while $|\mathcal{B}| > 1$ **do**
  Let λ be the median in $\mathcal{B}$
  for $g \in G$ **do**
    if *there exists* $j \in \mathcal{J}_g$ *such that* $\lambda = b_j^g$ **then**
      let $l_g(\lambda) = s_{j-1}^g$, $u_g(\lambda) = s_j^g$
    end
    else if *there exists* $j \in \mathcal{J}_\}$ *such that* $\lambda = (b_{j+1}^g, b_j^g)$ **then**
      let $l_g(\lambda) = u_g(\lambda) = s_j^g$
    end
    else if $\lambda > b_j^g)$ *for all* $j \in \mathcal{J}_g$ **then**
      let $l_g(\lambda) = u_g(\lambda) = 0$
    end
  end
  Compute: $\mathrm{WightBudget_l}(\lambda) = \sum_{g=1}^G m_g l_g(\lambda)\pi_g - \alpha \sum_{g=1}^G F_g(l_g(\lambda))m_g \leq 0$ and $\mathrm{WightBudget_u}(\lambda) = \sum_{g=1}^G m_g u_g(\lambda)\pi_g - \alpha \sum_{g=1}^G F_g(u_g(\lambda))m_g \geq 0$
  if $\mathrm{WightBudget_l}(\lambda) > 0$ **then**
    $\mathcal{B} \leftarrow b \in \mathcal{B} : b > \lambda$
  end
  else if $\mathrm{WightBudget_u}(\lambda) < 0$ **then**
    $\mathcal{B} \leftarrow b \in \mathcal{B} : b < \lambda$
  end
  else if $1 \in (\mathrm{WightBudget_l}(\lambda), \mathrm{WightBudget_u}(\lambda)$ **then**
    $\mathcal{B} \leftarrow \{\lambda\}$
  end
end
Let λ^* be the unique element remaining in $\mathcal{B}$
Let $c = (\mathrm{WightBudget_u}(\lambda) - 0) \div (\mathrm{WightBudget_u}(\lambda) - \mathrm{WightBudget_l}(\lambda))$
Let $t_g = cl_g(\lambda^*) + (1-c)u_g(\lambda^*), g \in G$
return the weights $w_g^* = \frac{t_g|I_l|}{\sum_{g=1}^G |i \in I_l : X_i = g| t_g}$

When F_g is a continuous function, we can directly write out the form of the lagrangian and obtained a closed form for our λ in the lagrangian. However, the tricky case is when F_g are piece-wise linear. To gain an intuition when the F_g are piece-wise linear, let's consider a simple case.



(a) Case 1



(b) Case 2

Figure 1: Shifting a dataset

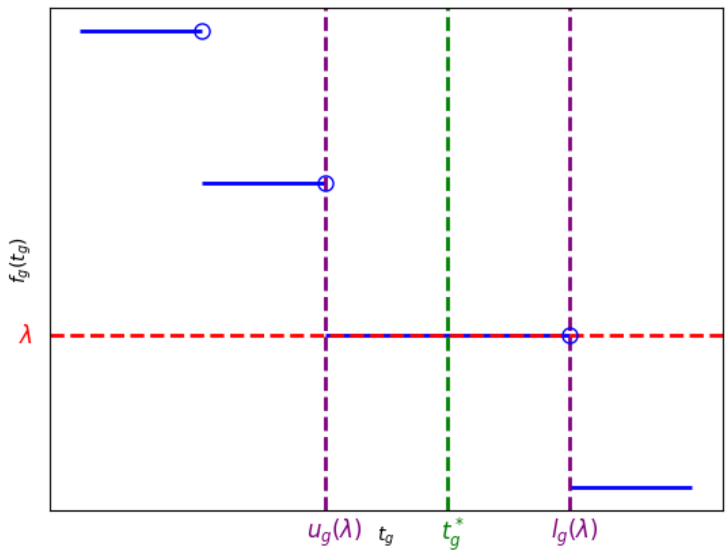


Figure 2: Case 3

Remember that in lagrangian method, the value of λ is the derivative of our desired maximizing function in respect to the constraint α. Therefore the value of $\lambda = F'(t_g) = f_g(t_g)$ where f_g is the density functions for the p-values (or equivalently the rejection threshold t_g). In the case of λ in figure 1a, it is apparent that the rejection threshold should be t_g^*. Similarly, in the case of λ in figure 1b, the rejection threshold should be the new t_g^* as shown in the picture. However, in case 3 when λ lands in a wide range of interval, it is hard to tell where to draw the rejection threshold between the lower bound l_g and the upper bound u_g. Therefore, algorithm seeks to, for each group g, capture the lower bound and upper bound for the optimal rejection threshold t_g and use that information to determine the value of t_g that satisfies the FDR control.

THen intuitively, we want the lower bound and upper bound to capture the controlled FDR rate α. Namely,

$$\mathrm{WightBudget_l}(\lambda) = \frac{\sum_{g=1}^{G} m_g l_g(\lambda)\pi_g}{\sum_{g=1}^{G} F_g(l_g(\lambda))m_g} \leq \alpha \leq \frac{\sum_{g=1}^{G} m_g u_g(\lambda)\pi_g}{\sum_{g=1}^{G} F_g(u_g(\lambda))m_g} = \mathrm{WightBudget_u}(\lambda) \tag{2}$$

where F_g are estimated distribution functions and π_g is the estimated null proportion for each group g. Therefore, we start with an initial λ (we usually set it to be the median in the set $\mathcal{B}$ so as to improve computational efficiency). Then, at each iteration in the while loop, we check whether the chosen λ can bound the lower bound of LinFdr below 0 and the upper bound of LinFdr above 0. During the last step when calculating the weights, we define

Definition 2.2. The weights for each group g is

$$w_g^* = \frac{t_g|I_l|}{\sum_{g=1}^{G} |i \in I_l : X_i = g|t_g}$$

where I_l is one of the folds $I_1, I_2, \cdots, I_K$ that partition the m hypothesis.

To check the validity of this method that it gives the optimal solution, we need to check the KKT conditions.

We first introduce the notion of **local FDR**.

Definition 2.3. We define the local FDR at a threshold t to be

$$\text{LFDR} = \Pr(H \text{ is null } \mid t_g) = \frac{\pi_0}{f(t)}$$

where $f(t)$ is the density function of the p values at $t > 0$.

Remark 1. Specially, if $t = 0$, we define the local FDR by the right limit of $f(0)$, namely,

$$\text{LFDR} = \frac{\pi_0}{\tilde{f}(0)}$$

where $\tilde{f}(0) = \lim_{t \to 0^+} f(t)$

We now proceed to prove the validity of the algorithm 5 for weighted BH in two cases based on local FDR:

Case 1: if all local FDR at 0 are larger than α, i.e.,

$$\frac{\pi_g}{f_g(t_g)} > \alpha, \forall g \in G$$

In this case we let the optimal solution $t^* = [0, 0, \cdots, 0] \in \mathbb{R}^G$.

We want to show that in the case when all local FDR at 0 are larger than α, $t^* = [0, 0, \cdots, 0] \in \mathbb{R}^G$ is the only solution that satisfies the constraint **??**.

Because $F_g(t_g)$ is a piece-wise linear function that is concave. We know that $F_g(t_g) \leq f_g(0)t_g, t_g = 0$ for all $g \in G$. We know that

$$\frac{\pi_g}{f_g(t_g)} > \alpha \longrightarrow f_g(t_g) < \frac{\pi_g}{\alpha}, t_g = 0 \forall g \in G$$

and therefore

$$F_g(t_g) \leq f_g(0)t_g < \frac{\pi_g}{\alpha} t_g$$

Thus we obtain an upper bound for $F_g(t_g)$ and we have

$$\begin{aligned}
\sum_{g=1}^{G} m_g t_g \pi_g - \alpha \sum_{g=1}^{G} F_g(t_g) m_g &\geq \sum_{g=1}^{G} m_g t_g \pi_g - \alpha \sum_{g=1}^{G} \frac{\pi_g}{\alpha} t_g m_g \\
&= \sum_{g=1}^{G} m_g t_g \pi_g - \sum_{g=1}^{G} m_g t_g \pi_g \\
&= 0
\end{aligned}$$

We need both the above condition and the condition **??** to hold. Namely, we have both

$$\sum_{g=1}^{G} m_g t_g \pi_g - \alpha \sum_{g=1}^{G} F_g(t_g) m_g \geq 0$$

and

$$\sum_{g=1}^{G} m_g t_g \pi_g - \alpha \sum_{g=1}^{G} F_g(t_g) m_g \leq 0$$

The only way to achieve this is to set $t_g = 0$ for all $g \in G$.

Intuitively, because increasing the threshold value t_g could only increase local FDR, if the local FDR at 0 is larger than α, this means that for all values of t_g, its local FDR is larger than α so the algorithm cannot find the optimal t^* to bound the local FDR and is force to make no rejections by setting $t^* = 0$.

Case 2: if at least 1 local FDR at 0 is smaller than or equal to α.

First, we want to show that if Case 2 occurs, this means that at least one $t_g^* > 0$. Recall that the local FDR at 0 is defined as the right limit of $f_g(t_g)$ as $t_g \to 0$ and $f_g(t_g)$ is a piece-wise constant function. This means that there exists some $t_g > 0$ such that its local FDR is smaller than α. In other words, the optimal solution exists. In this case, we can then use the algorithm described in section 2 to find the optimal t^*.

To prove the KKT conditions of the algorithm, we first write out the formula for the Lagrangian.

$$\mathcal{L} = \sum_{g=1}^{G} m_g F_g(t_g) + v^T t + \lambda(\alpha \sum_{g=1}^{G} m_g F_g(t_g) - \sum_{g=1}^{G} m_g t_g \pi_g) \tag{3}$$

where $v^T = [v_1, \cdots, v_G]^T$ are the dual variables corresponding to the non-negativity constraint of $t = [t_1, \cdots, t_G]$ and λ the dual variable corresponding to the non-negativity constraint.

Let's take a pause here to summarize the relationship betwen local FDR and λ. Let $t_{g,1}$ be the first point at which the density function f_g jumps for the gth group. If LFDR $< \alpha + \frac{1}{\lambda}$ at 0, then it must be the case that t_g^* is at least $t_{g,1}$ (at least for that lambda). If LFDR $> \alpha + \frac{1}{\lambda}$, then it must be that $t_g^* = 0$. Finally, if LFDR $= \alpha + \frac{1}{\lambda}$, then we would have an interval for t_g^* that would range from 0 to $t_{g,1}$.

We need to prove that the following three conditions hold:

- Complementary slackness: $v^T t = 0$
- Dual feasibility: $v_g \geq 0$ for all $g \in G$
- Primal feasibility: $t_g \geq 0$ and $\alpha \sum_{g=1}^{G} m_g F_g(t_g) - \sum_{g=1}^{G} m_g t_g \pi_g = 0$
- Stationary condition: $0 \in \partial_g \mathcal{L}$ for all $g \in G$

Complementary slackness We set $v_g = 0$ if $t_g^* > 0$ and $v_g = m_g(\lambda \pi_g - (1 + \lambda\alpha) f_g(0))$ if $t_g^* = 0$. Then, $v^T t^* = 0$ and complementary slackness holds by construction.

Dual feasibility Notice that when $t_g^* = 0$, this means that the local fdr $\frac{\pi_g}{f_g(t_g^*)} > \alpha$ for all values of $t_g^* \in [0, 1]$. Therefore if we rewrite

$$\lambda \pi_g - (1 + \lambda\alpha)) f_g(0) \geq 0$$

to be

$$\frac{\pi_g}{f_g(0)} \geq \frac{1}{\lambda} + \alpha > \alpha$$

is exactly the case of $\frac{\pi_g}{f_g(t_g^*)} > \alpha$ when $t_g^* = 0$.

Primal feasibility $t_g^* \geq 0$ and $\alpha \sum_{g=1}^G m_g F_g(t_g) - \sum_{g=1}^G m_g t_g^* \pi_g$ holds by construction.

Stationarity Finally, to check that stationarity holds, we take the superdifferential with respect to t_g, then we get

$$\begin{aligned}\partial_g \mathcal{L} &= m_g \partial F_g(t_g) + v_g + \lambda(\alpha m_g \partial F_g(t_g) - m_g \pi_g) \\ &= m_g(\partial F_g(t_g) + \lambda(\alpha \partial F_g(t_g) - \pi_g)) + v_g\end{aligned}$$

If $t_g^* = 0$, then we know that all the local FDR is larger than α. Therefore $f_g(0) = \partial_g F_g(t_g)$ for $t_g = 0$. Plugging $f_g(0)$ back to the subdifferential and assert $v_g = m_g(\lambda \pi_g - (1 + \lambda\alpha))f_g(0)$ we will obtain $0 \in \partial_g \mathcal{L}$.

If $t_g^* > 0$, then this means that the algorithm can find a corresponding λ that minimizes the problem.

$$\frac{\pi_g}{f_g(g)} \leq \alpha + \frac{1}{\lambda}$$

and therefore $0 \in \partial_g \mathcal{L}$.

3 Algorithm With regularization

$$\begin{aligned}\underset{\boldsymbol{t} \in [0,1]^G}{\text{maximize}} \quad & \sum_{g=1}^G q_g F_g(t_g) \ - \ \lambda \sum_{g=2}^G |t_g - t_{g-1}| \\ \text{subject to} \quad & \text{LinFdr}(\boldsymbol{t}; \alpha) \leq 0.\end{aligned} \tag{4}$$

We optimize this by ADMM. To derive the ADMM [?] steps, we introduce additional variables $\boldsymbol{y} = (y_1, \ldots, y_G) \in [0,1]^G$ subject to the additional constraint, $\boldsymbol{t} = \boldsymbol{y}$.

$$\begin{aligned}\underset{\boldsymbol{t} \in [0,1]^G}{\text{maximize}} \quad & \sum_{g=1}^G q_g F_g(t_g) \ - \ \lambda \sum_{g=2}^G |y_g - y_{g-1}| \\ \text{subject to} \quad & \text{LinFdr}(\boldsymbol{t}; \alpha) \leq 0, \ \ \boldsymbol{t} = \boldsymbol{y}.\end{aligned} \tag{5}$$

We compute the augmented Lagrangian with respect to only the total variation constraint. The augmented Lagrangian with scaled dual variables $\boldsymbol{u}$ has the form:

$$L(\boldsymbol{t}, \boldsymbol{y}, \boldsymbol{u}) := \sum_{g=1}^G q_g F_g(t_g) - \lambda \sum_{g=2}^G |y_g - y_{g-1}| - \frac{\rho}{2} \|\boldsymbol{t} - \boldsymbol{y} + u\|^2 - \delta\left(\text{LinFdr}(\boldsymbol{t}; \alpha) \leq 0\right). \tag{6}$$

By $\delta(\cdot)$ we denote the convex analysis indicator function, which is zero if the argument is true and infinity otherwise (and so encodes hard constraints).

The upshot of splitting $\boldsymbol{t}$ and $\boldsymbol{y}$ is that the total variation constraint becomes separable. We can optimize each step separately, while suitably accounting for the updates on the dual variable.

Each ADMM iteration then consists of the following steps:

Initialization: Set $\boldsymbol{t}^0, \boldsymbol{y}^0, \boldsymbol{u}^0$ to some initial values

Iteration:

$$\boldsymbol{t}^{j+1} \quad \leftarrow \underset{\boldsymbol{t}}{\operatorname{argmax}} \left\{ \sum_{g=1}^{G} q_g F_g(t_g) - \frac{\rho}{2} \left\| \boldsymbol{t} - \boldsymbol{y}^j + \boldsymbol{u}^j \right\|^2 - \delta \left(\mathrm{LinFdr}(\boldsymbol{t}; \alpha) \leq 0 \right) \right\} \tag{7a}$$

$$\boldsymbol{y}^{j+1} \quad \leftarrow \underset{\boldsymbol{y}}{\operatorname{argmin}} \left\{ \frac{\rho}{2} \left\| \boldsymbol{t}^{j+1} - \boldsymbol{y} + \boldsymbol{u}^j \right\|^2 + \lambda \sum_{g=2}^{G} |y_g - y_{g-1}| \right\} \tag{7b}$$

$$\boldsymbol{u}^{j+1} \quad \leftarrow \boldsymbol{u}^j + \boldsymbol{t}^{j+1} - \boldsymbol{y}^{j+1}. \tag{7c}$$

Stopping criterion: Need to think about this, but basically until the iteration above stabilizes.

3.1 On the subproblems

Let us provide some detail for each step.

Fused lasso step: (7b) is merely a Fused Lasso update, which can be solved efficiently (nearly linear time), for example by dynamic programming [**?**].

"FDR" step: The update in (7a) is a bit more involved. However, it is quite similar to the algorithm without regularization. Our goal amounts to solving optimization problems of the following form:

$$\begin{aligned} \underset{\boldsymbol{t} \in [0,1]^G}{\text{maximize}} \quad & \sum_{g=1}^{G} q_g F_g(t_g) \; - \; \frac{\rho}{2} \left\| \boldsymbol{t} - \boldsymbol{b} \right\|^2. \\ \text{subject to} \quad & \mathrm{LinFdr}(\boldsymbol{t}; \alpha) \leq 0. \end{aligned} \tag{8}$$

Above, $\boldsymbol{b}$ in the $j+1$-iteration is given by $\boldsymbol{y}^j - \boldsymbol{u}^j$. We can solve (8) by a bisection search over an additional Lagrangian parameter. To be concrete, let

$$\begin{aligned} L_{\mathrm{Fdr}}(\boldsymbol{t}, \mu; \alpha) &:= \sum_{g=1}^{G} q_g F_g(t_g) \; - \; \frac{\rho}{2} \left\| \boldsymbol{t} - \boldsymbol{b} \right\|^2 - \mu \, \mathrm{LinFdr}(\boldsymbol{t}; \alpha) \\ &= (1 + \alpha\mu) \sum_{g=1}^{G} q_g F_g(t_g) \; - \; \frac{\rho}{2} \left\| \boldsymbol{t} - \boldsymbol{b} \right\|^2 - \mu \sum_{g=1}^{G} q_g t_g. \end{aligned}$$

For fixed μ, we can solve this by taking derivatives (or superdifferentials in the case of the Grenander estimator). Let $f_g(\cdot)$ denote the derivative or an element of the superdifferential of $F_g(\cdot)$. Then, the solution to the maximization problem with respect to t_g is given by $t_g(\mu)$ such that:

$$(1 + \alpha\mu) q_g f_g(t_g(\mu)) - \rho t_g(\mu) = q_g \mu - \rho b_g. \tag{9}$$

In case there is no solution to the above, then we set $t_g(\mu) = 0$. If there exists a solution, then is it unique we can find it efficiently by noting that $f_g(\cdot)$ for the Grenander estimator is piecewise constant, left-continuous, and non-increasing. At any point of discontinuity, u of f_g, we are allowed to take $f_g(u)$ equal to any value in the interval $[f_g(u_+), f_g(u_-)]$ (the superdifferential is a set-valued mapping) when checking (9).

In any case, for any value of μ, we thus get $\boldsymbol{t}(\mu) = (t_1(\mu), \ldots, t_G(\mu))$. We then need to check whether for that value of μ, it also holds that $\mathrm{Fdr}(\boldsymbol{t}(\mu)) = \alpha$. By bisecting over $\mu \geq 0$, we can find μ^* such that $\mathrm{Fdr}(\boldsymbol{t}(\mu^*)) = \alpha$, and then $\boldsymbol{t}(\mu^*)$ furnishes the solution to (8).

3.2 Path following and ρ updates

In practice we want to solve the above problem for a range of λ values. We proceed as follows with warm starts.

1. Compute $\lambda_{\max}$, the smallest value of $\lambda > 0$ such that the solution to (4) collapses to the solution of the problem without groupping, i.e., at optimality it holds that $t_1 = \ldots = t_G$ (this can be done efficiently, TODO: add explanation). Let t^* denote the common value of the optimal solution. Set $\boldsymbol{u}^0 = \boldsymbol{y}^0 = 0$ and $\boldsymbol{t}^0 = (t^*, \ldots, t^*)$.

2. Decrease λ to $\lambda = \eta\lambda_{\max}$ for some pre-specified $\eta < 1$. Run the ADMM algorithm with $\rho = \lambda$ and initialize $\boldsymbol{u}^0, \boldsymbol{t}^0, \boldsymbol{y}^0$ as in the previous step.

3. Decrease λ further and also set $\rho = \lambda$. Initialize by the values of $\boldsymbol{u}^j, \boldsymbol{t}^j, \boldsymbol{y}^j$ computed at the last iteration of the ADMM algorithm in the previous step.

4. …

Computing $\lambda_{\max}$. First let (t^*, μ^*) be the (primal,dual) solution to the primal-dual problem without grouping, i.e.,

$$t^* \in \operatorname*{argmax}_t \left\{ \sum_{g=1}^G q_g F_g(t) - \mu^* \,\mathrm{LinFdr}(t\mathbf{1}; \alpha) \right\},$$

and such that $\mathrm{Fdr}(t^*\mathbf{1}) = \alpha$.

The above has superdifferential:

Next consider the total-variation penalized problem:

$$\max_{\boldsymbol{t}} \left\{ \sum_{g=1}^G q_g F_g(t_g) - \mu^* \,\mathrm{LinFdr}(\mathbf{t}; \alpha) - \lambda \sum_{g=2}^G |t_g - t_{g-1}| \right\}.$$

Now also introduce $\boldsymbol{u} = D\boldsymbol{t}$. We get:

$$\max_{\boldsymbol{t},\boldsymbol{u}} \left\{ \sum_{g=1}^G q_g F_g(t_g) - \mu^* \,\mathrm{LinFdr}(\mathbf{t}; \alpha) - \lambda \left\|\boldsymbol{u}\right\|_1 + \nu^\intercal(\boldsymbol{u} - D\boldsymbol{t}) \right\}.$$

Take the subdifferential with respect to u_k;

$$\nu_k/\lambda \in \begin{cases} \{1\} & \text{if } u_k > 0 \\ \{-1\} & \text{if } u_k < 0 \\ [-1, 1] & \text{if } u_k = 0. \end{cases},$$

i.e., $\lambda \geq |\nu_k|$. Take the subdifferential with respect to t_g;

$$(1 + \alpha\mu^*) q_g \partial F_g(t_g) - \mu^* q_g - (D^\intercal \nu)_g.$$

We would like this to include zero at $t_g = t^*$. Let $f_g(t) \in \partial F_g(t)$ be the g-th density at t and $y_g = (1 + \alpha\mu^*) q_g f_g(t) - \mu^* q_g$. Then we would like dual variables ν such that:

$$D^\intercal \nu = \boldsymbol{y}.$$

One such solution is given by:

$$\nu = (DD^\intercal)^{-1} D\boldsymbol{y}.$$

Hence for $\lambda_{\text{mas}} = \|\nu\|_\infty$, indeed the solution collapses to the ungrouped solution.

References