Graph trend filtering

- local adaptivity
- computational efficiency
- analysis regularization

Univariate trend filtering: $\hat{\beta} = \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2}\|y-\beta\|_2^2 + \lambda \| D^{(k+1)} \beta \|_1$

$$\| D^{(1)} \cdot \beta \|_1 = \sum_{i=1}^{n-1} |\beta_{i+1} - \beta_i|$$

$$D^{(k+1)} = D^{(1)} \cdot D^{(k)}$$

Trend filtering on graphs: $G = (V, E)$; $V = \{1, 2, \dots, n\}$;

$E = \{e_1, e_2, \dots, e_m\}$ undirected

$y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$

GTF estimate, $\hat{\beta}$ &

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2}\|y-\beta\|_2^2 + \lambda \| \Delta^{(k+1)} \beta \|_1$$

$\Delta^{k+1} = k+1^{th}$ order Graph Difference Operator (GDO)

$$\| \Delta^{(1)} \cdot \beta \|_1 = \sum_{i,j \in E} |\beta_i - \beta_j|$$

If $e_l = (i, j)$ then,

l^{th} row of $\Delta^{(1)}$ i.e, $\Delta^{(1)}_l = (0, \dots, \underset{i}{-1}, 0, \dots, \underset{j}{1}, 0, \dots 0)$

$$\Delta^{(k+1)} = \begin{cases} (\Delta^{(1)})^T \Delta^{(k)} = L^{(k+1)/2} ; & k = 2m+1 \\ \Delta^{(1)} \cdot \Delta^{(k)} = D \cdot L^{k/2} ; & k = 2m \end{cases}$$

$L = D^T D$ and,

$\Delta^{(k+1)} \in \mathbb{R}^{n \times n}$; $k = 2m+1$

$\Delta^{(k+1)} \in \mathbb{R}^{m \times n}$; $k = 2m$

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \ \frac{1}{2} \|y - \beta\|_2^2 + \lambda \|\Delta^{(k+1)} \beta\|_1$$

Fast ADMM algorithm :

Reparameterization for odd & even ks,

$$\min_{\beta, z \in \mathbb{R}^n} \ \frac{1}{2} \|y - \beta\|_2^2 + \lambda \|Dz\|_1 \quad \text{s.t,} \quad z = L^{k/2} x \qquad \text{(even)}$$

$$\min_{\beta, z \in \mathbb{R}^n} \ \frac{1}{2} \|y - \beta\|_2^2 + \lambda \|z\|_1 \quad \text{s.t,} \quad z = L^{(k+1)/2} x \qquad \text{(odd)}$$

Augmented Lagrangian, $\mathcal{L} =$

$$\frac{1}{2} \|y - \beta\|_2^2 + \lambda \|Sz\|_1 + \frac{\rho}{2} \|z - L^{a}\beta + u\|_2^2 - \frac{\rho}{2} \|u\|_2^2$$

$$S = \begin{cases} D ; & k \in \text{even} \\ I ; & k \in \text{odd} \end{cases} \qquad a = \begin{cases} k/2 ; & k \in \text{even} \\ (k+1)/2 ; & k \in \text{odd} \end{cases}$$

ADMM based updates, for some b; $\Rightarrow$

$$\beta \longleftarrow (I + \rho L^{2a})^{-1} \cdot b$$

$$z \longleftarrow \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \ \frac{1}{2} \|b - x\|_2^2 + \frac{\lambda}{\rho} \|Sx\|_1$$

Newton's method :

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \ \frac{1}{2} \|y - \beta\|_2^2 + \lambda \|\Delta^{(k+1)} \beta\|_1$$

Dual form :

$$\hat{v} = \underset{v \in \mathbb{R}^r}{\operatorname{argmin}} \ \|y - (\Delta^{(k+1)})^T v\|_2^2 \quad \text{s.t,} \quad \|v\|_\infty \leq \lambda$$

and, $\hat{\beta} = y - (\Delta^{(k+1)})^T \hat{v}$

Update rule,

$$v \longleftarrow a + ((\Delta_I^{(k+1)})^T)^{+} b$$

for some a, b & set of indices I

Lap SVM trained in the primal :

$$\min_{\alpha \in \mathbb{R}^n, \xi \in \mathbb{R}^l} \sum_{i=1}^{l} \xi_i + r_A \alpha^T K \alpha + r_I \alpha^T K L K \alpha$$

s.t, $$y_i \left(\sum_{j=1}^{n} \alpha_j \cdot K(x_i, x_j) + b \right) + \xi_i \geq 1$$

$$\xi_i \geq 0 \longrightarrow \text{Dual} \longrightarrow \underline{\underline{\text{QPP}}}$$

$$\forall\, i = 1, 2, \cdots, l$$

$$\min_{f \in H_K} \sum_{i=1}^{l} \max(1 - y_i \cdot f(x_i), 0) + r_A \|f_A\|^2 + r_I \|f_I\|^2$$

$$\equiv \min_{\alpha \in \mathbb{R}^n, b \in \mathbb{R}} \sum_{i=1}^{l} \underbrace{V(x_i, y_i, K_i^T \alpha + b)}_{\text{loss-}f^n} + r_A \alpha^T K \alpha + \underbrace{r_I (\alpha^T K + 1^T b) L (K\alpha + 1 b)}_{\text{bias included}}$$

loss-f^n ↓ Squared Hinge loss

$$\equiv \min_{\alpha \in \mathbb{R}^n, b \in \mathbb{R}} \frac{1}{2} \left[\sum_{i=1}^{l} \max(1 - y_i (K_i^T \alpha + b), 0)^2 + r_A \alpha^T K \alpha + \underbrace{r_I (\alpha^T K + 1^T b) L (K\alpha + b)}_{\text{piece wise quadratic}} \right] \quad \text{— (1)}$$

Newton's method :

$$z = [b, \alpha^T]^T$$

H – Hessian of (1) wrt. z

∇ – Gradient of (1) wrt. z

$$z^t = z^{t-1} - s H^{-1} \nabla$$

$$\nabla = \begin{bmatrix} \nabla_b \\ \nabla_\alpha \end{bmatrix}$$

$$\nabla^2 = H = \begin{bmatrix} \nabla_b^2 & \nabla_b(\nabla_\alpha) \\ \nabla_\alpha(\nabla_b) & \nabla_\alpha^2 \end{bmatrix}$$

TFMCM :

$$\min_{f \in H_K} \frac{1}{l}\sum_{i=1}^{l} V(x_i, y_i, f) + r_A \|f\|_A^2 + \frac{r_I}{(u+l)^2}\|\Delta f\|_1$$

$L = \Delta^T \Delta$

$$\min_{h, \alpha, b, \lambda} \quad h + \frac{1}{l}\sum_{i=1}^{l} \alpha_i + \frac{r_I}{(u+l)}\|\Delta K \lambda\|_1$$

Such that,

$$h \geq y_i \left(\sum_{j=1}^{l+u} \lambda_j \cdot K_{ij} + b\right)$$

$$y_i \left(\sum_{j=1}^{l+u} \lambda_j \cdot K_{ij} + b\right) + \alpha_i \geq 1$$

$$\alpha_i \geq 0, \quad h \geq 1$$

$$\forall\ i = 1, 2, \ldots, l$$

$$y_i \left(\sum_{j=1}^{l+u} \lambda_j \cdot K_{ij} + b\right) + \alpha_i \geq 1$$

objective,

$$\frac{1}{l}\sum_{i=1}^{l} V(x_i, y_i, f) \equiv \sum_{i=1}^{l} V(x_i, y_i, K_i^T \lambda + b)$$

$$\sum_{i=1}^{l} \max\left(1 - y_i \left(\sum_{j=1}^{l+u} \lambda_j \cdot K_{ij} + b\right), 0\right)^2 =$$

$$\equiv \min \left[\max\left(1 - y_i \left(\sum_{j=1}^{l+u} \lambda_j \cdot K_{ij} + b\right), 0\right)^2 + \max\left(y_i \left(\sum_{j=1}^{l+u} \lambda_j \cdot K_{ij} + b\right) h, 0\right)^2 + \underline{\gamma_I} \cdot \|\Delta K \lambda\|_1 \right]$$

$$= \min_{\substack{\lambda \in \mathbb{R}^n \\ b \in \mathbb{R} \\ h \in \mathbb{R}}} \frac{1}{2} \left[\max\left(1 - y_i \left(\sum_{j=1}^{n} \lambda_j \cdot K_{ij} + b\right), 0\right)^2 + \max\left(y_i \left(\sum_{j=1}^{l+u} \lambda_j K_{ij} + b\right) - h, 0\right)^2 + \gamma_I \|\Delta K \lambda\|_1 \right]$$

$$z = [h, b, \lambda^T]^T$$

$$z^t = z^{t-1} - s\, H^{-1} \nabla$$

$$\nabla = \begin{bmatrix} \nabla_h \\ \nabla_b \\ \nabla_\lambda \end{bmatrix} ; \quad H = \begin{bmatrix} \nabla_h^2 & \nabla_h \nabla_b & \nabla_h \nabla_\lambda \\ \nabla_b \nabla_h & \nabla_b^2 & \nabla_b \nabla_\lambda \\ \nabla_\lambda \nabla_h & \nabla_\lambda \nabla_b & \nabla_\lambda^2 \end{bmatrix}$$

Newton's method

ADMM (Alternating direction method of multipliers):

$$\min f(x) + g(z)$$
$$\text{s.t.}, \quad Ax + Bz = c$$

$$\mathcal{L}(x, y, z) = f(x) + g(z) + y^T (Ax + Bz - c) + \frac{\rho}{2} \|Ax + Bz - c\|_2^2$$

ADMM:

$$x^{k+1} := \underset{x}{\operatorname{argmin}}\ \mathcal{L}(x, z^k, y^k)$$
$$z^{k+1} := \underset{z}{\operatorname{argmin}}\ \mathcal{L}(x^{k+1}, z, y^k)$$

Gauss-Seidel method

$$y^{k+1} := \cancel{\text{argmin}}\ y^k + \rho (Ax^{k+1} + Bz^{k+1} - c)$$

Scaled up ADMM :
$$\min_{x,z} \quad f(x)+g(z)$$
$$\text{s.t,} \quad Ax+Bz=c$$

Augmented Lagrangian,

$$\mathcal{L}(x,z,y) = f(x)+g(z)+y^T(Ax+Bz-c)+\frac{\rho}{2}\|Ax+Bz-c\|_2^2$$
$$= f(x)+g(z)+\frac{\rho}{2}\|Ax+Bz-c+u\|_2^2 + \text{const.}$$

with, $u^k = \frac{1}{\rho}y^k$

Updates :

$$x^{k+1} := \underset{x}{\operatorname{argmin}}\left(f(x)+\rho/2\,\|Ax+Bz^k-c+u^k\|_2^2\right)$$
$$z^{k+1} := \underset{z}{\operatorname{argmin}}\left(g(z)+\frac{\rho}{2}\|Ax^{k+1}+Bz-c+u^k\|_2^2\right)$$
$$y^{k+1} := u^k+(Ax^{k+1}+Bz^{k+1}-c)$$

Ex: Lasso problem, $\min \frac{1}{2}\|Ax-b\|_2^2+\lambda\|x\|_1$

$$\equiv \min \frac{1}{2}\|Ax-b\|_2^2+\lambda\|z\|_1$$
$$\text{s.t,} \quad x-z=0$$

ADMM updates :

$x^{k+1} := ?$ $\quad z^{k+1} := ?$ $\quad u^{k+1} := ?$

$$\mathcal{L}(x,z,y) = \frac{1}{2}\|Ax-b\|_2^2+\lambda\|z\|_1+y^T(x-z)+\frac{\rho}{2}\|x-z\|_2^2$$
$$= \frac{1}{2}\|Ax-b\|_2^2+\lambda\|z\|_1+\frac{\rho}{2}\|x-z+u\|_2^2+\text{const}$$

$u^k = y^k/\rho$

For x^{k+1}; $\quad \underset{x}{\operatorname{argmin}}\left(\frac{1}{2}\|Ax-b\|_2^2+\underbrace{\frac{\rho}{2}\|x-z^k-u^k\|_2^2}_{\theta}\right)$

$$\frac{\partial\theta}{\partial x} =$$

:

Lasso problem : $\min \frac{1}{2}\|Ax-b\|_2^2 + \lambda\|x\|_1$

ADMM form : $\min \frac{1}{2}\|Ax-b\|_2^2 + \lambda\|z\|_1$

s.t, $x-z=0$

Augmented Lagrangian with penalty term $\rho = \frac{1}{\tau} > 0$;

$$\mathcal{L}_{\frac{1}{\tau}}(x,y,z) = \frac{1}{2}\|Ax-b\|_2^2 + \lambda\|z\|_1 + \frac{1}{\tau}\langle y, x-z\rangle + \frac{1}{2\tau}\|x-z\|_2^2$$

Update rule for x :

$$x_k = \underset{x}{\operatorname{argmin}}\ \mathcal{L}_{\frac{1}{\tau}}(x, z_{k-1}, y_{k-1})$$

$$= \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2}\|Ax-b\|_2^2 + \lambda\|z_{k-1}\|_1 + \frac{1}{\tau}\langle y_{k-1}, x-z_{k-1}\rangle + \frac{1}{2\tau}\|x-z_{k-1}\|_2^2 \right\}$$

$$= \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2}\left\langle x, \left(A^TA + \frac{1}{\tau}I\right)x\right\rangle - \left\langle x, A^Tb + \frac{1}{\tau}(z_{k-1}-y_{k-1})\right\rangle \right\}$$

$$= \left(A^TA + \frac{1}{\tau}I\right)^{-1} \cdot \left(A^Tb + \frac{1}{\tau}(z_{k-1}-y_{k-1})\right)$$

Update rule for z:

$$z_k = \underset{z}{\operatorname{argmin}}\ \mathcal{L}_{\frac{1}{\tau}}(x_k, z, y_{k-1})$$

$$= \underset{z}{\operatorname{argmin}} \left\{ \frac{1}{2}\|Ax_k-b\|_2^2 + \lambda\|z\|_1 + \frac{1}{\tau}\langle y_{k-1}, x_k-z\rangle + \frac{1}{2\tau}\|x_k-z\|_2^2 \right\}$$

$$= \underset{z}{\operatorname{argmin}} \left\{ \frac{1}{2}\|x_k+y_{k-1}-z\|_2^2 + \lambda\|z\|_1 \right\}$$

$$= S_{\lambda T}(x_k+y_{k-1})$$

where, $S_{\lambda T}$ is the soft-thresholding operator

$$S_{\lambda T}(x) = \operatorname{sgn}(x)\,(|x|-\lambda)_+$$

Update rule for y:

$$y_k = y_{k-1} + \frac{1}{\tau}(x_k - z_k)$$

$$\frac{1}{2}(Ax-b)^T(Ax-b) + \frac{1}{\tau}\langle y_{k-1}, x - z_{k-1}\rangle + \frac{1}{2\tau}\|x - z_{k-1}\|_2^2$$

$$\cancel{\frac{1}{2}Ax} = \frac{1}{2}(x^TA^T - b^T)(Ax-b) + \frac{1}{\tau}y_{k-1}^T(x - z_{k-1})$$

$$+ \frac{1}{2\tau}(x^T - z_{k-1}^T)(x - z_{k-1})$$

$$= \frac{1}{2}x^TA^TAx - \frac{1}{2}x^TA^Tb - \frac{1}{2}b^TAx + \cancel{\frac{b^Tb}{2}}$$

$$+ \frac{1}{\tau}y_{k-1}^Tx - \cancel{\frac{1}{\tau}y_{k-1}^Tz_{k-1}}$$

$$+ \frac{1}{2\tau}x^Tx - \frac{1}{2\tau}x^Tz_{k-1} - \frac{1}{2\tau}z_{k-1}^Tx + \cancel{\frac{1}{2\tau}z_{k-1}^Tz_{k-1}}$$

∂ wrt. x

$$= A^TA\cdot x + \frac{1}{\tau}\cdot I\cdot x - \frac{1}{2}A^Tb + \frac{1}{\tau}y_{k-1}^T - \frac{1}{2\tau}z_{k-1} - \frac{1}{2\tau}z_{k-1}^T$$

$$\Rightarrow \left(A^TA + \frac{I}{\tau}\right)x = \frac{1}{2}A^Tb + \frac{z_{k-1}}{\tau} - \frac{y_{k-1}}{\tau}$$

$$x = \left(A^TA + \frac{I}{\tau}\right)^{-1}\left(A^Tb + \frac{1}{\tau}(z_{k-1} - y_{k-1})\right)$$

$$\cancel{\frac{1}{2}\|Ax_k - b\|_2^2} + \lambda\|z\|_1 + \frac{1}{\tau}\langle y_{k-1}, x_k - z\rangle + \frac{1}{2\tau}\|x_k - z\|_2^2$$

$$\cancel{\frac{1}{\tau}y_{k-1}^T\cdot x_k} - \frac{1}{\tau}y_{k-1}^Tz + \frac{1}{2\tau}(x_k - z)^T(x_k - z)$$

$$\lambda\|z\|_1 - \frac{1}{\tau}y_{k-1}^Tz + \cancel{\frac{1}{2\tau}x_k^Tx_k} - \frac{1}{2\tau}x_k^Tz - \frac{1}{2\tau}z^Tx_k$$

$$+ \frac{1}{2\tau}z^Tz$$

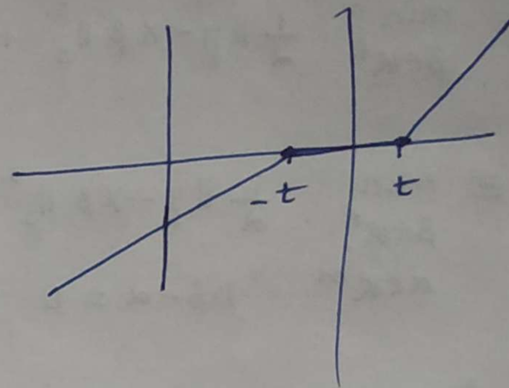
$$\lambda\|z\|_1^{\cdot} - \frac{1}{\tau}y_{k-1} - \frac{1}{2\tau}x_k - \frac{1}{2\tau}x_k + \frac{1}{\tau}z = 0$$

$$\lambda\|z\|_1^{\cdot} - \frac{1}{\tau}(x_k + y_{k-1}) + \frac{z}{\tau} = 0$$

$$\frac{z}{t} + \lambda \dot{} \|z\|_1 = \frac{1}{\tau}(x_k + y_{k-1})$$

$$z = -\tau\lambda \dot{} \|z\|_1 + \frac{\cancel{t}}{\cancel{t}}(x_k + y_{k-1})$$

$$z = S_{\lambda\tau}(x_k + y_{k-1})$$



$$[S_t(x)]_j = \begin{cases} x_j - t, & x > t \\ 0, & -t < x < t \\ x_j + t, & x < t \end{cases}$$

$j = 1, 2, \dots p$

$$z = S_{\lambda\tau}(x_k + y_{k-1})$$

$$[S_{\lambda\tau}(x_k + y_{k-1})]_j = \begin{cases} (x_k + y_{k-1} - \lambda\tau); & x_k + y_{k-1} > \lambda\tau \\ 0; & -\lambda\tau < x_k + y_{k-1} < \lambda\tau \\ (x_k + y_{k-1} + \lambda\tau); & x_k + y_{k-1} < -\lambda\tau \end{cases}$$

Soft threshold f^n

$$y_k = y_{k-1} + \frac{1}{\tau}(x_k - z_k)$$

Fused Lasso regression using ADMM :

$$\min_{\beta \in \mathbb{R}^p} \quad \frac{1}{2}\|y - X\beta\|_2^2 + \lambda \|D\beta\|_1$$

$$\equiv \min_{\substack{\beta \in \mathbb{R}^p, \\ \alpha \in \mathbb{R}^m}} \quad \frac{1}{2}\|y - X\beta\|_2^2 + \lambda \|\alpha\|_1$$

$$D\beta - \alpha = 0$$

Augmented Langrangian with penalty $\rho > 0$,

$$\mathcal{L}_\rho(\beta, \alpha, w) = \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\alpha\|_1 + w^T(D\beta - \alpha) + \frac{\rho}{2}\|D\beta - \alpha\|_2^2$$

Updates :

$$\beta^k = (X^TX + \rho D^TD)^{-1}(X^Ty + \rho D^T(\alpha^{k-1} - w^{k-1}))$$

$$\alpha^k = S_{\lambda/\rho}\cdot(D\beta^k + w^{k-1})$$

$$w^k = w^{k-1} + D\beta^k - \alpha^k$$

LapMCM using Trend Filtering solved in Primal using ADMM / Newton's method :

TFMCM :

$$\min_{h, b, \lambda, q} \quad h + \sum_{i=1}^{l} q_i + r_I \|\Delta K \lambda\|_1$$

s.t,

$$h \geq y_i \cdot \left(\sum_{j=1}^{l+u} \lambda_j \cdot K_{ij} + b\right)$$

$$y_i \cdot \left(\sum_{j=1}^{l+u} \lambda_j \cdot K_{ij} + b\right) + q_i \geq 1$$

$$q_i \geq 0 \; ; \; (h \geq 1)$$

$$\forall \; i = 1, 2, 3, \ldots l$$

l = labelled samples

u = unlabelled "

$n = l + u$

Primal with no constraints :

$$\min_{h, b, \lambda} \frac{1}{2} \sum_{i=1}^{l} \Big\{ \max\Big(1 - y_i\Big(\sum_{j=1}^{l+u} \lambda_j \cdot K_{ij} + b\Big), 0\Big)^2 + \max\Big(y_i\Big(\sum_{j=1}^{l+u} \lambda_j \cdot K_{ij} + b\Big) - h, 0\Big)^2 + r_I \|\Delta K \lambda\|_1 \Big\}$$

Newton's method: (using squared hinge loss)

$$\min_{h,b,\lambda} \frac{1}{2}\left\{\sum_{i=1}^{l}\left[\max\Big(1-y_i\big(\sum_{j=1}^{l+u}\lambda_j\cdot k_{ij}+b\big),0\Big)^2+\max\Big(y_i\big(\sum_{j=1}^{l+u}\lambda_j\cdot k_{ij}+b\big)-h,0\Big)^2\right]\right\}$$

$$+\ r_I \,\|\Delta K\lambda\|_1$$

let $\underline{z=[h,b,\alpha^T]^T}$ then, $\underline{z^t = z^{t-1} - s\,h^{-1}\nabla}$

$$\nabla = \begin{bmatrix}\nabla h\\ \nabla b\\ \nabla\lambda\end{bmatrix};\quad h = \begin{bmatrix}\nabla h^2 & \nabla h\nabla b & \nabla h\nabla\lambda\\ \nabla b\nabla h & \nabla b^2 & \nabla b\nabla\lambda\\ \nabla\lambda\nabla h & \nabla\lambda\nabla b & \nabla\lambda^2\end{bmatrix}$$

Using ADMM: use linear hinge loss

$$\min_{h,b,\lambda} \frac{1}{2}\sum_{i=1}^{l}\left\{\max\Big(1-y_i\big(\sum_{j=1}^{l+u}\lambda_j\cdot k_{ij}+b\big),0\Big)+\max\Big(y_i\big(\sum_{j=1}^{l+u}\lambda_j\cdot k_{ij}+b\big)-h,0\Big)\right\}$$

$$+\ r_I\,\|\Delta K\lambda\|_1$$

Incorporating bias into λ,

let, $\lambda_j \in \mathbb{R}^{l+u+1}$ and; $\tilde{K} = \begin{bmatrix} K & 1 \\ \cancel{1} & \cancel{1}\end{bmatrix}_{l+u\times l+u+1}$ ~~$l+u \times l+u+1$~~ ~~$l+u+1$~~

$$\equiv \min_{h,\lambda} \frac{1}{2}\sum_{i=1}^{l}\left\{\max\Big(1-y_i\big(\sum_{j=1}^{l+u+1}\lambda_j\cdot\tilde{k}_{ij}\big),0\Big)+\max\Big(y_i\big(\sum_{j=1}^{l+u+1}\lambda_j\cdot\tilde{k}_{ij}\big)-h,0\Big)\right\}$$

$$+\ r_I\,\|\Delta\tilde{K}\lambda\|_1 \quad \ldots$$

~~$Y = f$~~

$Y = \mathrm{diag}(y_1, y_2, \ldots, y_l) \in \mathbb{R}^{l\times l}$

$f = K\lambda \in \mathbb{R}^{l+u\times 1}$

~~Y~~ $\qquad Y_{l\times l} \qquad (K\lambda)_{l+u\times 1}$

$$\begin{bmatrix} Y \end{bmatrix}\begin{bmatrix}1 & 0\\ \_ & J\end{bmatrix}\begin{bmatrix} f_l \\ \hline f_u\end{bmatrix}$$

$$= \begin{bmatrix} Y & 0\end{bmatrix}$$

$\underline{YJ\tilde{K}\tilde{\lambda}}$

$$\cancel{\min_{h,\lambda} \frac{1}{2} \sum_{i=1}^{l} \{\max(1-4JK\lambda, 0) + \max(4JK\lambda-h, 0)}$$

$$\min_{h,\lambda} \frac{1}{2} \sum_{i=1}^{l} \left\{ \max(1-4JK\lambda, 0) + \max(4JK\lambda - h, 0) \right\} + r_I \|\Delta K\lambda\|_1$$

$$\equiv \min_{x,y} \frac{1}{2} \sum_{i=1}^{l} \max(1-4JK\lambda, 0) + r_I \|\Delta K\lambda\|_1 + \frac{1}{2} \sum_{i=1}^{l} \max(z, 0)$$

$$\text{s.t.,} \quad z + h - 4JK\lambda = 0$$

Extra variable to be removed.

Now, using ADMM,

Augmented Lagrangian with penalty, $\rho > 0$:

$$\mathcal{L}_\rho(x, y, z) = \frac{1}{2} \sum \max(1-4JK\lambda, 0) + r_I \|\Delta K\lambda\|_1 + \frac{1}{2} \sum \max(z, 0) + \langle y, z+h-4JK\lambda \rangle + \frac{\rho}{2} \|z+h-4JK\lambda\|_2^2$$

Updates:

$$x^{k+1} := \underset{x}{\operatorname{argmin}}\ \mathcal{L}_\rho(x, y^k, z^k)$$

$$z^{k+1} := \underset{z}{\operatorname{argmin}}\ \mathcal{L}_\rho(x^{k+1}, y^k, z)$$

$$y^{k+1} := y^k + \rho(z^{k+1}$$

Trend filtered MCM :

$$\min_{h,b,\lambda} \quad h + r_I \|\Delta K \lambda\|_1 + \sum_{i=1}^{l} q_i$$

$$\text{s.t.} \quad y_i \left[\sum_{j=1}^{l+u} \lambda_j \cdot k_{ij} + b \right] + q_i \geq 1$$

$$h \geq y_i \left[\sum_{j=1}^{l+u} \lambda_j \cdot k_{ij} + b \right] \qquad \forall\ i = 1, 2, 3, \ldots, l$$

$$q_i \geq 0, \quad h \geq 1$$

using squared Hinge loss

$$\equiv \min_{h,b,\lambda} \frac{1}{2} \left\{ \sum_{i=1}^{l} \left[\underbrace{\max\left(1 - y_i \left\{ \sum_{j=1}^{l+u} \lambda_j \cdot k_{ij} + b \right\}, 0\right)^2}_{\text{penalty if } f_i < 1} + \underbrace{\max\left(y_i \left\{ \sum_{j=1}^{l+u} \lambda_j \cdot k_{ij} + b \right\} - h_j, 0\right)^2}_{\text{penalty if } f_i > h} \right] + \underbrace{h}_{\text{minimize } h} + \underbrace{r_I \|\Delta K \lambda\|_1}_{\text{regularization}} \right\}$$

regularizaⁿ for smoothness of conditional over the geodesics of the marginal through TF.

Solution using Newton's method :

Problem

$$\equiv \min_{h,b,\lambda} \frac{1}{2} \left\{ \sum_{i=1}^{l} \left[\max\left(1 - y_i (k_i^T \lambda + b)\right)_+^2 + \max\left(y_i (k_i^T \lambda + b) - h\right)_+^2 \right] + h + r_I \|\Delta K \lambda\|_1 \right\}$$

$\left(\max(a)_+ = \max\{a, 0\} \right)$

Let $z = [h, b, \lambda^T]^T$

then, $z^t = z^{t-1} - s H^{-1} \nabla$ where,

$\nabla = [\nabla_h, \nabla_b, \nabla_\lambda]^T$

$$H = \begin{bmatrix} \nabla_h^2 & \nabla_h \nabla_b & \nabla_h \nabla_\lambda \\ \nabla_b \nabla_h & \nabla_b^2 & \nabla_b \nabla_\lambda \\ \nabla_\lambda \nabla_h & \nabla_\lambda \nabla_b & \nabla_\lambda^2 \end{bmatrix}$$

$$\nabla_h = 1 + \frac{\partial}{\partial h}\left\{ h^2 - 2h y_i (k_i^T \lambda + b) \right\}$$

$$\nabla_h = 1 + 2h - 2y_i (k_i \lambda + b)$$

$$\Rightarrow \boxed{\nabla_h = 1 + \sum_{i=1}^{l} \left(2\left[h - y_i (k_i \lambda + b)\right] \right)}$$

$$\nabla_b = \frac{\partial}{\partial b}\left\{ y_i^2 (k_i \lambda + b)^2 - 2y_i (k_i \lambda + b) \right\} + \frac{\partial}{\partial b}\left\{ y_i^2 (k_i \lambda + b)^2 - 2h y_i (k_i \lambda + b) \right\}$$

$$= y_i^2 \, 2 \cdot (k_i \lambda + b) - 2y_i + 2y_i^2 (k_i \lambda + b) - 2h y_i$$

$$\nabla_b = 4y_i^2 (k_i \lambda + b) - 2y_i (1 + h)$$

$$\boxed{\nabla_b = \sum_{i=1}^{l} 2y_i \left(2y_i (k_i \lambda + b) - 1 - h \right)}$$

$$\nabla_\lambda = \frac{\partial}{\partial \lambda}\left\{ y_i^2 (k_i \lambda + b)^2 - 2y_i (k_i \lambda + b) \right\} + \frac{\partial}{\partial \lambda}\left\{ y_i^2 (k_i \lambda + b)^2 - 2y_i h (k_i \lambda + b) \right\}$$
$$+ \frac{\partial}{\partial \lambda} \| \Delta K \lambda \|_1$$

$$= 2y_i^2 \cdot k_i (k_i \lambda + b) - 2y_i \cdot k_i + 2y_i^2 \cdot k_i (k_i \lambda + b) - 2y_i \cdot h \cdot k_i$$
$$+ \frac{\partial}{\partial \lambda} \| \Delta K \lambda \|_1$$

$$= 4y_i^2 k_i \cdot (k_i \lambda + b) - 2y_i \cdot k_i (1 + h) + \frac{\partial}{\partial \lambda} \| \Delta K \lambda \|_1$$

$$\nabla_\lambda = \sum_{i=1}^{l} k_i 2y_i \left(2y_i (k_i \lambda + b) - 1 - h \right) + \frac{\partial}{\partial \lambda} \| \Delta K \lambda \|_1$$

for, $\frac{\partial}{\partial \lambda} \| \Delta K \lambda \|_1 = ?$; $y = \Delta K \lambda$

so, $dy = \Delta K \cdot d\lambda$ $\qquad A : B = tr(A^T B)$

$\Rightarrow \theta = \|y\|_1 = sgn(y) : y$

so, $d\theta = sgn(y) : dy = sgn(y) : \Delta K \, d\lambda = (\Delta K)^T sgn(y) : d\lambda$

so, $\frac{d\theta}{d\lambda} = \frac{d}{d\lambda} \| \Delta K \lambda \|_1 = K^T \Delta^T sgn(\Delta K \lambda)$

$$\boxed{\nabla_\lambda = 2 \sum_{i=1}^{l} y_i \cdot k_i \left(2y_i (k_i \lambda + b) - 1 - h \right) + K^T \Delta^T sgn(\Delta K \lambda)}$$

so;

$$\nabla = \begin{bmatrix} \nabla_h \\ \nabla_b \\ \nabla_\lambda \end{bmatrix} = \begin{bmatrix} 1 + 2\sum_{i=1}^{l}[h - y_i(k_i\lambda + b)] \\ 2\sum_{i=1}^{l}[y_i(2y_i(k_i\lambda+b) - 1 - h)] \\ 2\sum_{i=1}^{l}[k_iy_i(2y_i(k_i\lambda+b)-1-h)] + k^T\Delta^T \operatorname{sgn}(\Delta k\lambda) \end{bmatrix}$$

For the Hessian,

$$\nabla_h^2 = 2l$$

$$\nabla_b^2 = 2\sum_{i=1}^{l} y_i^2$$

$$\nabla_\lambda^2 = 2\sum_{i=1}^{l} 2k_i^2y_i^2 + (k^T\Delta^T)^2 \overbrace{\delta(\Delta k\lambda)}^{0} = 4\sum_{i=1}^{l} k_i^2 y_i^2$$

$$\nabla_h\nabla_b = -2\sum_{i=1}^{l} y_i^2$$

$$\nabla_h\nabla_\lambda = -2\sum_{i=1}^{l} k_iy_i$$

$$\nabla_b\nabla_h = -2\sum_{i=1}^{l} y_i$$

$$\nabla_b\nabla_\lambda = 2\sum_{i=1}^{l} 2k_iy_i^2$$

$$\nabla_\lambda\nabla_h = \cancel{-2\sum_{i=1}^{l} k_iy_i} \; -2\sum_{i=1}^{l} y_i \cdot k_i$$

$$\nabla_\lambda\nabla_b = 2\sum_{i=1}^{l} 2k_iy_i^2$$

So;

$$H = \begin{bmatrix} 2l & -2\sum_{i=1}^{l} y_i^2 & -2\sum_{i=1}^{l} k_i \cdot y_i \\ -2\sum_{i=1}^{l} y_i & 2\sum_{i=1}^{l} y_i^2 & 4\sum_{i=1}^{l} k_iy_i^2 \\ -2\sum_{i=1}^{l} k_iy_i & 4\sum_{i=1}^{l} k_iy_i^2 & 4\sum_{i=1}^{l} k_i^2y_i^2 + (k^T\Delta^T)^2\delta(\Delta k\lambda) \end{bmatrix}$$

TFSVM using ADMM :

formulation;

$$\min_{\lambda, b} \frac{1}{2} \left\{ \sum_{i=1}^{l} \max\left(1 - y_i[k_i^T \lambda + b]; 0\right) + r_A \lambda^T K \lambda + r_I \|\Delta K \lambda\|_1 \right\}$$

$$\equiv \min_{\lambda, b} \sum_{i=1}^{l} a_{i+} + r_A \lambda^T K \lambda + r_I \|\Delta K \lambda\|_1$$

s.t, $\cancel{a = 1 - YJK\lambda}$ $a = 1 - YJK\lambda - Yb1$

$$\equiv \min \sum_{i=1}^{l} \max(a_i, 0) + r_A \cancel{\lambda}^T K \lambda + r_I \|\Delta K \lambda\|_1$$

s.t, $a = 1 - Y(JK\lambda + b1)$

Augmented Lagrangian where, $\rho > 0$

$$\mathcal{L}_\rho(\lambda, b, a, u) = \sum_{i=1}^{l} \max(a_i, 0) + r_A \lambda^T K \lambda + r_I \|\Delta K \lambda\|_1 + \frac{1}{2}\langle u, 1 - Y(JK\lambda + b1) - a\rangle + \frac{\rho}{2}\|1 - Y(JK\lambda + b1) - a\|_2^2$$

updates,

$$(\lambda^{k+1}, b^{k+1}) = \underset{\lambda, b}{\operatorname{argmin}} \cancel{\mathcal{L}_\rho(\lambda^k,} \mathcal{L}_\rho(\lambda, b, a^k, u^k)$$

$$a^{k+1} = \underset{a}{\operatorname{argmin}} \mathcal{L}_\rho(\lambda^{k+1}, b^{k+1}, a, u^k)$$

$$u^{k+1} = u^k + \rho\left(1 - Y(JK\lambda^{(k+1)} + b^{(k+1)} \cdot 1) - a^{(k+1)}\right)$$