现在我们来解释 EEG 源的重构问题

$$\min_{u}\left\{\frac{1}{2}\|Au-b\|_2^2+\lambda\, TV_\alpha(u)\right\} \tag{2}$$

即

$$\min_{u,v}\left\{\frac{1}{2}\|Au-b\|_2^2+\lambda\|v\|_1\right\},\quad \text{subject to } D_\alpha u=v.$$

的 ADMM 算法(3) 及另一种采用 Pock - Chambolle 算法思想去处理的加速迭代方法 (4).

$$\min_{u}\left\{\frac{1}{2}\|Au-b\|_2^2+\lambda\|D_\alpha u\|_1\right\}$$

$$\Updownarrow$$

$$\min_{u,v}\left\{\frac{1}{2}\|Au-b\|_2^2+\lambda\|v\|_1\right\},\ \text{subject to}\ \ v=D_\alpha u$$

即约束条件为 $D_\alpha u-v=0$

~~选择~~ ADMM方法：(Alternating direction method of multipliers)

采用 Lagrange 乘数法，将目标函数更改为

$$\cancel{\min_{u,v}}\ \frac{1}{2}\|Au-b\|_2^2+\lambda\|v\|_1+\rho\langle\tilde{v},D_\alpha u-v\rangle+\frac{\rho}{2}\|D_\alpha u-v\|_2^2$$

Lagrange 乘子　　　　添上这一项

交替迭代

对 v 而言：

$$\min_{v}\left\{\lambda\|v\|_1-\rho\langle\tilde{v},v\rangle+\frac{\rho}{2}\|D_\alpha u-v\|_2^2\right\}$$

$$\Leftrightarrow\min_{v}\left\{\lambda\|v\|_1+\frac{\rho}{2}\|v-D_\alpha u-\tilde{v}\|_2^2-\frac{\rho}{2}\|\tilde{v}\|_2^2\right\}$$

$$\Leftrightarrow\min_{v}\left\{\|v\|_1+\frac{1}{2\cdot\frac{\lambda}{\rho}}\|v-(D_\alpha u+\tilde{v})\|_2^2\right\}$$

$$\Leftrightarrow v=\mathrm{Prox}_{\frac{\lambda}{\rho}}^{\|\cdot\|_1}(D_\alpha u+\tilde{v})$$

$$\Leftrightarrow v=\mathrm{Shrink}(D_\alpha u+\tilde{v})$$，这里 $\mathrm{Shrink}(u,\mu)_i=\mathrm{sign}(u_i)\max\{|u_i|-\mu,0\}$.

对 u 而言：

$$\min_{u}\left\{\frac{1}{2}\|Au-b\|_2^2+\rho\langle\tilde{v},D_\alpha u\rangle+\frac{\rho}{2}\|D_\alpha u-v\|_2^2\right\}$$

$$\cancel{\Leftrightarrow\min_{u}\left\{\frac{1}{2}\|Au-b\|_2^2+\rho\langle D_\alpha^T\tilde{v},u\rangle+\frac{\rho}{2}\|D_\alpha u-v\|_2^2\right\}}$$

$$\Leftrightarrow\min_{u}\left\{\frac{1}{2}\|Au-b\|_2^2+\frac{\rho}{2}\|D_\alpha u-(v-\tilde{v})\|_2^2\right\}$$

∴ 对 u 求导，得到

$$A^T(Au-b)+\rho D_\alpha^T\left(D_\alpha u-(v-\tilde{v})\right)=0$$

$$\therefore\ (A^TA+\rho D_\alpha^TD_\alpha)u=A^Tb+\rho D_\alpha^T(v-\tilde{v})$$

$$\therefore\ u=(A^TA+\rho D_\alpha^TD_\alpha)^{-1}\left(A^Tb+\rho D_\alpha^T(v-\tilde{v})\right)$$

对 $\tilde{v}$ 而言：对 $\tilde{v}$ 求导，~~且令其为0，对~~ 得到 $\rho(D_\alpha u-v)$

$$\therefore\ \tilde{v}\leftarrow v+\gamma(D_\alpha u-v)$$，可取 $\gamma\in\left(0,\frac{\sqrt{5}+1}{2}\right)$.

或者说 讨论 $G(x)+\frac{1}{2\tau}\|x-\hat{x}\|_2^2$ 的极小化问题.

定义: $\cancel{\mathrm{Prox}_\tau^G(\hat{x})}$ $x^* = \underset{x}{\arg\min}\ G(x)+\frac{1}{2\tau}\|x-\hat{x}\|_2^2$

$$0 \in \partial G(x)+\frac{1}{\tau}(x-\hat{x})$$

$$\therefore\quad 0 = \tau\partial G(x^*)+(x^*-\hat{x})$$

$$\therefore\quad x^*+\tau\partial G(x^*) = \hat{x}$$

$$\therefore\quad (1+\tau\partial G)(x^*) = \hat{x},$$

于是 $x^* = (1+\tau\partial G)^{-1}\hat{x}$

也将其 x^* 记为 $x^* = \mathrm{Prox}_\tau^G(\hat{x})$

现在讨论

$$\min_x F(Kx)+G(x)$$

$\Longleftrightarrow$ ~~min~~

利用 Fenchel 对偶函数的概念:

令 $$F^*(t) = \max_x \left(\langle t, x\rangle - F(x)\right)$$

且有 $(F^*)^* = F$.

$$\therefore\quad F(y) = \max_t \left\{\langle y, t\rangle - F^*(t)\right\}$$

$\therefore$ 问题变成

$$\min_x \left\{ \max_y \left(\langle Kx, y\rangle - F^*(y)\right) + G(x) \right\}$$

所以除了原始的变量 x 外, 增加了辅助变量 y.
(的对偶)

迭代了k步后，已得到 x^k, y^k 的值.

现在进行 k+1 步迭代：

对 x^{k+1}：

$$\min_x \left\{ G(x) + \langle Kx, y^k \rangle \right\}$$

添加一项：

$$x^{k+1} = \arg\min_x \left\{ G(x) + \underline{\overset{\overset{\langle x, K^T y^k \rangle}{\|}}{\langle Kx, y^k \rangle}} + \underline{\frac{1}{2\tau} \| x - x^k \|_2^2} \right\}$$

↑ 要求 x^{k+1} 在 x^k 附近.

$$= \arg\min_x \left\{ G(x) + \frac{1}{2\tau} \| x - (x^k - \tau K^T y^k) \|_2^2 \right\}$$

$$= \mathrm{Prox}_\tau^G \left(x^k - \tau K^T y^k \right)$$

$$= (1 + \tau \partial G)^{-1} \left(x^k - \tau K^T y^k \right).$$

对 y^{k+1}：

$$\max_y \left\{ \langle K x^{k+1}, y \rangle - F^*(y) \right\} \Longleftrightarrow \min_y \left\{ F^*(y) - \langle K x^{k+1}, y \rangle \right\}$$

添加一项：

$$y^{k+1} = \arg\min_y \left\{ F^*(y) - \langle y, K x^{k+1} \rangle + \underline{\frac{1}{2\sigma} \| y - y^k \|^2} \right\}$$

$$= \arg\min_y \left\{ F^*(y) + \frac{1}{2\sigma} \| y - (y^k + \sigma K x^{k+1}) \|^2 \right\}$$

$$= \mathrm{Prox}_\sigma^{F^*} \left(y^k + \sigma K x^{k+1} \right)$$

$$= (1 + \sigma \partial F^*)^{-1} \left(y^k + \sigma K \underset{\uparrow}{x^{k+1}} \right)$$

可以用 $x^{k+1} + \theta (x^{k+1} - x^k)$ 来替代

当 $\theta = 1$ 时，为 $2x^{k+1} - x^k$.

所以

$$x^{k+1} = (1 + \tau \partial G)^{-1} \left(x^k - \tau K^T y^k \right)$$

$$y^{k+1} = (1 + \sigma \partial F^*)^{-1} \left(y^k + \sigma K \left(x^{k+1} + \theta (x^{k+1} - x^k) \right) \right)$$

τ, σ 分别与 x^k, y^k 的迭代步长有关. 但对向量 x^k 的各个分量 $(x^k)_i$ 而言，迭代步长是一致的.

$$\begin{cases} x^{k+1} = (I + T\partial G)^{-1}(x^k - TK^T y^k) \\ y^{k+1} = (I + \Sigma\partial F^*)^{-1}(y^k + \Sigma K(x^{k+1} + \theta(x^{k+1} - x^k))) \end{cases}$$

$\langle x_1, x_2\rangle_X = \langle T^{-1}x_1, x_2\rangle$

$\langle y_1, y_2\rangle_Y = \langle \Sigma^{-1}y_1, y_2\rangle$

$x^* = \arg\min_x G(x) + \frac{1}{2}\|x - \hat{x}\|_X^2$

$0 \in \partial G(x) + T^{-1}(x - \hat{x})$

$\therefore 0 \in \partial G(x^*) + T^{-1}(x^* - \hat{x})$

$\therefore x^* = (I + T\partial G)^{-1}(\hat{x})$

由 G 及 F^* 的凸性知:

$\forall (x,y) \in X \times Y$, 有变分不等式:

$$\left\langle \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}, F\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} + M\begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\rangle \geq 0$$

这里

$$F\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} \partial G(x^{k+1}) + K^T y^{k+1} \\ \partial F^*(y^{k+1}) - Kx^{k+1} \end{pmatrix}$$

$$M = \begin{pmatrix} T^{-1} & -K^T \\ -\theta K & \Sigma^{-1} \end{pmatrix}$$

这~~是~~写成了 proximal point algorithm 的形式（可参看 Pock-Chambolle 2011年文章）

对 θ, T, Σ 加上什么条件 才能保证 算法的收敛性？

如 $\theta = 1$, T, Σ 满足

$$\|\Sigma^{\frac{1}{2}} K T^{\frac{1}{2}}\|^2 < 1$$

则可推出 M 为对称正定，即 $\forall (x,y)^T \in X \times Y$, $(x,y)^T \neq 0$, 有 $\left\langle \begin{pmatrix} x \\ y \end{pmatrix}, M\begin{pmatrix} x \\ y \end{pmatrix} \right\rangle > 0$

于是 迭代序列 $\begin{pmatrix} x^n \\ y^n \end{pmatrix}$ 将收敛于 鞍点问题(1) 的一个最优解 $\begin{pmatrix} x^* \\ y^* \end{pmatrix}$.

如果 G, F^* 是可分的，则 T, Σ 可取为对角阵.

令 $T = \mathrm{diag}(\tau)$， $\tau = (\tau_1, \cdots, \tau_n)$， $\tau_j = \dfrac{1}{\sum_{i=1}^{m} |K_{i,j}|^{2-\alpha}}$

$\Sigma = \mathrm{diag}(\sigma)$， $\sigma = (\sigma_1, \cdots, \sigma_m)$

$\sigma_i = \dfrac{1}{\sum_{j=1}^{n} |K_{i,j}|^{\alpha}}$

这里，$\forall \alpha \in [0, 2]$

~~择此~~ 上述对角阵满足

$$\|\Sigma^{1/2} K T^{1/2}\|^2 = \sup_{x \in X, x \neq 0} \frac{\|\Sigma^{1/2} K T^{1/2} x\|^2}{\|x\|^2} \leq 1$$

为使"<1"，

可将 $\tau \to \mu\tau$，$\sigma \to \nu\sigma$，且 $\mu\nu < 1$.

∴ 迭代收敛性得以保证.

现在再回到原来的在弯曲曲面上的TV问题：

$$\min_u \left\{ \frac{1}{2}\|Au-b\|_2^2 + \lambda \|D_\alpha u\|_1 \right\}$$

令

$$F(\cdot) = \frac{1}{2}\|\cdot\|_2^2 \quad , \quad F_2(\cdot) = \lambda\|\cdot\|_1$$

则问题可重写为

$$\min_u \max_{s,t} \left\{ \left(\langle Au-b, t \rangle - F_1^*(t) \right) + \left(\langle D_\alpha u, s \rangle - F_2^*(s) \right) \right\}$$

现在来计算 F_1^* 及 F_2^*：

(1) 对 F_1，有

$$F_1^*(t) = \max_x \left\{ \langle t, x \rangle - F_1(x) \right\} = \max_x \left\{ \langle t, x \rangle - \frac{1}{2}\|x\|_2^2 \right\}$$

$$= \max_x \left\{ -\frac{1}{2}\left(\|x-t\|_2^2 + \|t\|_2^2 \right) \right\}$$

$$= \min_x \frac{1}{2}\left(\|x-t\|_2^2 + \|t\|_2^2 \right)$$

因为当 $x=t$ 时达到极小，所以极小值为 $\frac{1}{2}\|t\|_2^2$，即

$$F_1^*(t) = \frac{1}{2}\|t\|_2^2$$

(2) 对 F_2，有

$$F_2^*(s) = \max_p \left(\langle s, p \rangle - \lambda\|p\|_1 \right)$$ ，用分量形式可表为

$$= \lambda \max_p \left\{ \frac{1}{\lambda}\sum_i s_i p_i - \sum_i |p_i| \right\}$$

$$= \lambda \sum_i \max_{p_i} \left(\frac{s_i}{\lambda} \cdot p_i - |p_i| \right)$$

因为 $\left|\frac{s_i}{\lambda}\right| \le$ 时，$\max_{p_i}\left(\frac{s_i}{\lambda} p_i - |p_i| \right) = 0$，否则为 $+\infty$，所以

$$F_2^*(s) = \begin{cases} 0 , & \|s\|_\infty \le \lambda \\ +\infty , & \text{否则} \end{cases}$$

所以(*)式中，

关于 x^{k+1} 的迭代式中的 $(I+\Sigma\partial G)^{-1}$，由于 $G=0$，所以退化为 I.

关于 t^{k+1} 的迭代式中的 $(I+\Gamma_2\partial F_1^*)^{-1}$，由于 $F_1^*(t)=\frac{1}{2}\|t\|_2^2$，所以

$\partial F_1^*=I$，因此算视为 $(I+\Gamma_2)^{-1}$

而关于 s^{k+1} 的迭代式中的 $(I+\Gamma_1\partial_1 F_2^*)^{-1}(\tilde{s})$

我们换一种表述方法：

$$s^{k+1}=\mathrm{Prox}_1^{F_2^*}(\tilde{s})$$

这里 $\tilde{s}=s^k+\Gamma_1 D_x\left(x^k-2\Sigma\left(D_x^T s^k+A^T t^k\right)\right)$

$$=\underset{s}{\arg\min}\left\{F_2^*(s)+\frac{1}{2}\|s-\tilde{s}\|^2\right\}$$

$\underbrace{\quad}_{\text{因为这里忽略了 }\Gamma_1\text{ 前置的 ~~参数~~ 参数}}$

由于

$$F_2^*(s)=\begin{cases}0, & \|s\|_\infty\le\lambda\\ +\infty, & \text{否则}\end{cases},$$

所以求极小值只能在 ~~Box~~ $\underbrace{\|s\|_\infty\le\lambda}_{\text{Box 区域}}$ 中进行，在此区域中找一点与 $\tilde{s}$ 最接近，

最后就得出了

$$s^{k+1}=\mathrm{Proj}_{\{\|\cdot\|_\infty\le\lambda\}}(\tilde{s}),$$

即 ~~■~~ 把 $\tilde{s}$ 投影到 Box 区域中去，投影点即为 s^{k+1}

(投影点与 Γ_1 的选择无关)

前面说过