Decoupling Noise and Features via Weighted ℓ_1-Analysis Compressed Sensing (補足資料)

藤堂 英樹

2014/07/26

1 Generalized Cross-Validation (GCV) の導出について

$$\mathrm{GCV}_n(\lambda) = \frac{\frac{1}{n}\|P - \hat{S}_n(\lambda)\|_F^2}{(1 - \frac{1}{n}\mathrm{tr}[A_n(\lambda)])^2} \tag{1}$$

近似解法について，元論文では，$\mathrm{tr}[A_n(\lambda)]$ を固有値で近似する方法しか明記されていない．$A_n(\lambda) = (I_n + \lambda M)^{-1}$ であり，μ_i を $0 \leq \mu_1 \leq \cdots \leq \mu_n$ となる行列 M の固有値とすると，$\mathrm{tr}[A_n(\lambda)] = \sum_{i=1}^n \frac{1}{1+\lambda\mu_i}$ の形で近似できる．

しかし，$\|P - \hat{S}_n(\lambda)\|_F^2 = \|P(I_n - A_n(\lambda))\|_F^2$ についても，$A_n(\lambda)$ を含むため，高速化のためには近似が必要となる．

$\|P(I_n - A_n(\lambda))\|_F^2$ の近似については，supplemental material の式 (8) にそれらしい式変形がある．

$$\mathbf{f}^T(A_n(\lambda) - I_n)^2\mathbf{f} = \sum_{j=1}^n \frac{\lambda^2\mu_j^2}{(1+\lambda\mu_j)^2} e_j^2 \tag{2}$$

ここで，$e_j = \mathbf{u}_j^T\mathbf{f}$ である．現状の実装では，$\|P(I_n - A_n(\lambda))\|_F^2 = \sum_{j=1}^n \frac{\lambda^2\mu_j^2}{(1+\lambda\mu_j)^2}\mathbf{u}_j^T P$ として計算している．

この式の導出については，完全には理解できていない．$A_n(\lambda) = (I_n + \lambda M)^{-1}$ より，

$$\begin{aligned} \mathbf{f}^T(\frac{I_n}{I_n + \lambda M} - I_n)^2\mathbf{f} &= \mathbf{f}^T(-\frac{\lambda M}{I_n + \lambda M})^2\mathbf{f} & (3) \\ &= \mathbf{f}^T\frac{(\lambda M)^2}{(I_n + \lambda M)^2}\mathbf{f} & (4) \end{aligned}$$

$M = \mu u$ と書けるとすると，$\mathbf{f}^T\frac{u\mu^2u^T}{(I_n+\lambda\mu u)^2}\mathbf{f}$ となり，式 (2) の形がおぼろげながら見えてくる．より正しい導出は今後の課題である．

また，現状の実装では，固有値，固有ベクトルの計算にかなりの時間がかかっているのも課題である．例えば，3070 頂点の 8 の字メッシュでは，固有値，固有ベクトルの計算に 400 秒も時間がかかる．論文では，固有値を計算すれば GCV を計算できると書いてあるので，固有ベクトルを使わない $\|P(I_n - A_n(\lambda))\|_F^2$ の近似法が分かれば，もう少し早く計算ができる可能性がある．

2 ℓ_1 最適化の補足

論文で実装されている手法では，以下の制約を内点法 [1] で解いているが，現状の実装では，より簡易に解きやすい形に問題を変えて ℓ_1 解析を行っている ([2] の ℓ_0 の定式化を ℓ_1 に置き換えた方法)．

$$\min_{\mathbf{h}} \|\mathbf{h} - \mathbf{b}\|_2^2 \tag{5}$$

$$\text{s.t. } \|WL\mathbf{h}\|_1 = \textstyle\sum_{i=1}^n w_i|L_i\mathbf{h}| \leq \tau \tag{6}$$

まず，式 (6) を Laglange 未定乗数を導入して以下のように書き換える．

$$\min_{\mathbf{h}} \|\mathbf{h} - \mathbf{b}\|_2^2 + \tau\|WL\mathbf{h}\|_1 \tag{7}$$

さらに，これを最適化するため，補助変数 α を導入し，以下の形に書き換える．

$$\min_{\mathbf{h}} \|\mathbf{h} - \mathbf{b}\|_2^2 + \alpha\|WL\mathbf{h} - \mathbf{y}\|_2^2 + \tau\|\mathbf{y}\|_1 \tag{8}$$

式 (8) を最適化するため，以下の部分問題に切り分ける．まず，$\mathbf{y}$ を固定し，$\mathbf{h}$ について解く．

$$\min_{\mathbf{h}} \|\mathbf{h} - \mathbf{b}\|_2^2 + \alpha\|WL\mathbf{h} - \mathbf{y}\|_2^2 \tag{9}$$

$$(E + \alpha(WL)^T WL)\mathbf{h} = \mathbf{b} + \alpha(WL)^T\mathbf{y} \tag{10}$$

次に，$\mathbf{h}$ を固定し，$\mathbf{y}$ について解く．

$$\min_{\mathbf{y}} \alpha\|WL\mathbf{h} - \mathbf{y}\|_2^2 + \tau\|\mathbf{y}\|_1 \tag{11}$$

$$y_i = \max(|x_i| - \tau, 0)\frac{x_i}{|x_i|} \tag{12}$$

$$\mathbf{x} = WL\mathbf{h} \tag{13}$$

α をだんだん大きくしながら ($\alpha_{n+1} = 2\alpha_n$)，$\mathbf{y}, \mathbf{h}$ を解く処理をそれぞれ交互に繰り返す．現状の実験では，ノイズ付の 1 次元信号を入力とし，完全ではないものの疎な特徴を解析により検出できている．今後は，ノイズに対してよりロバストな ℓ_1 最適化の手法について検討していく必要がある．

参考文献

[1] Boyd, S. and Vandenberghe, L.: *Convex Optimization*, Cambridge University Press, New York, NY, USA (2004).

[2] Xu, L., Lu, C., Xu, Y. and Jia, J.: Image Smoothing via L0 Gradient Minimization, *ACM Trans. Graph.*, Vol. 30, No. 6, pp. 174:1–174:12 (2011).