Lasso における FISTA, ADMM の最適化方法

武山尚生

2020 年 3 月 9 日

1 ISTA

FISTA の説明をする前に，まず ISTA の説明を簡潔にする．まず，以下の最適化問題を考える．

$$\mathrm{argmin}_{\mathbf{x}}\Big(\frac{1}{2}\big(\mathbf{y}-\mathrm{A}\mathbf{x}\big)^2+\alpha|\mathbf{x}|\Big) \tag{1}$$

これを最適化するために，微分可能な近接写像を定義することで，$\frac{1}{2}(y-Ax)^2$ を分離性のある形にする．

$$\mathcal{L}(\mathbf{x})=\frac{1}{2}\big(\mathbf{y}-\mathrm{A}\mathbf{x}\big)^2 \tag{2}$$

と $\mathcal{L}(\mathbf{x})$ をおくと，

$$\mathcal{G}(\mathbf{x};\mathbf{x}^{(t)})=\mathcal{L}(\mathbf{x}^{(t)})+\frac{d\mathcal{L}(\mathbf{x}^{(t)})}{d\mathbf{x}}(\mathbf{x}-\mathbf{x}^{(t)})+\frac{\rho}{2}\big|\mathbf{x}-\mathbf{x}^{(t)}\big|^2 \tag{3}$$

$$=\frac{\rho}{2}\Big|\big(\mathbf{x}^{(t)}-\frac{1}{\rho}\frac{d\mathcal{L}(\mathbf{x}^{(t)})}{d\mathbf{x}}\big)-\mathbf{x}\Big|^2+\mathrm{const.} \tag{4}$$

但し，$\mathcal{G}(\mathbf{x};\mathbf{x}^{(t)})>\mathcal{L}(\mathbf{x})$ を全ての x において満たす ρ は以下の式で求まる．

$$\rho=\max_j\sum_i\big|\mathrm{A}^\top\mathrm{A}_{ij}\big| \tag{5}$$

これより，$\mathcal{L}(\mathbf{x}$ の最適値は，

$$\mathbf{x}^{(t+1)}=\mathrm{argmin}_{\mathbf{x}}\Big(\mathcal{G}(\mathbf{x};\mathbf{x}^{(t)})+\alpha|\mathbf{x}|\Big) \tag{6}$$

に (4) で得られた極値を代入して，

$$\mathbf{x}^{(t+1)}=S_{\alpha/\rho}\Big(\mathbf{x}^{(t)}-\frac{1}{\rho}\frac{d\mathcal{L}(\mathbf{x}^{(t)})}{d\mathbf{x}}\Big) \tag{7}$$

$$=S_{\alpha/\rho}\Big(\mathbf{x}^{(t)}-\frac{1}{\rho}A^T(y-Ax_t)\Big) \tag{8}$$

という漸化式で表せる．以上が ISTA の導出となる．

2 FISTA

次に，FISTA の導出について説明する．FISTA は，ISTA に Nesterov の加速法を適用したものである．Nesterov の加速法は以下の式によって x を更新する．

$$\boldsymbol{x}_{k+1} = \bar{\boldsymbol{x}}_k - \eta \bigtriangledown f(\bar{\boldsymbol{x}}_k) \tag{9}$$

$$\bar{\boldsymbol{x}}_k = \boldsymbol{x}_k + \gamma_k \Delta \boldsymbol{x}_k \tag{10}$$

$$\gamma_k = \frac{\rho_{k-1} - 1}{\rho_k} \tag{11}$$

$$\rho_k = \frac{1 + \sqrt{1 + 4\rho_{k-1}^2}}{2} \tag{12}$$

勾配法にはこの他に Momentum 法を中心として様々な加速法がある．単純に勾配を利用して最適化するだけでなく，それ以前の勾配の重みを利用することで，勾配の降下に対して慣性をつけたり，到達地点の予測をすることなどによって最適化を加速させることが出来る．例えば Momentum 法の場合，以下の式に従って x を更新する．

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \eta \bigtriangledown f(\boldsymbol{x}_k) + \alpha \Delta \boldsymbol{x}_k \tag{13}$$

この式を過去の勾配の値を用いて表す．

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \eta \sum_{l=0}^{k} \alpha^l \bigtriangledown f(\boldsymbol{x}_{k-l}) \tag{14}$$

勾配の値を指数減衰させた，総和で表すことが出来る．よって，Momentum 法では過去の勾配の記録を指数減衰させた値を，慣性として x の更新式に加えていることが分かる．
Nesterov の加速法でも同様の方法を行っており，(8) を式変形すると，以下のようになる．

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \eta \bigtriangledown f(\bar{\boldsymbol{x}}_k) + \gamma_k \Delta \boldsymbol{x}_k \tag{15}$$

但し，$\bar{\boldsymbol{x}}_k = \boldsymbol{x}_k + \gamma_k \Delta \boldsymbol{x}_k$ であり，$\bar{\boldsymbol{x}}_k$ は x の予測された移動地点を示している．これより，Nesterov の加速法では Momentum 法と比較して，予測された地点での勾配を求めている点や，γ_k の項が次第に増加していくことから，Momentum 法や他の勾配法よりも早く最適化が進むことが期待される．

Nesterov の加速法を ISTA に適用すると，以下のようになる．これが FISTA のアルゴリズムとなる．

$$\mathbf{x}^{(t+1)} = S_{\alpha/\rho}\Big(\mathbf{z}^{(t)} - \frac{1}{\rho} A^T (y - A z_t)\Big) \tag{16}$$

$$\beta_{t+1} = \frac{1}{2}\sqrt{1 + 4{\beta_t}^2} \tag{17}$$

$$z_{t+1} = x_{t+1} + \frac{\beta_t - 1}{\beta_{t+1}}(x_{t+1} - x_t) \tag{18}$$