May 7, 2021

$\underline{f(x_{k+1}) = f(x_k + p)}$

① Why sparsity
② non-smoothness

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

$f(x) = g(x) + h(x)$

smooth, cvx   not-smooth, cvx

GD: $f(x_{k+1}) \approx f(x_k) + \nabla f(x_k)^T p \quad p = -\alpha \nabla f(x_k) \quad x_{k+1} = x_k - \alpha \nabla f(x_k)$

$f(x_k + p) \approx g(x_k) + \nabla g(x_k)^T p + h(x_k + p)$ (for p small)

Idea: $x_{k+1} = \arg\min_u \left(\nabla g(x_k)^T (u - x_k) + h(u) + \frac{1}{2\alpha}\|u - x_k\|_2^2 \right)$

regularization

$$\frac{1}{2\alpha}\|u - x_k + \alpha \nabla g(x_k)\|_2^2 = \frac{1}{2\alpha}\left[\|u - x_k\|_2^2 + \alpha^2 \cancel{\|\nabla g(x_k)\|_2^2} + 2\alpha \nabla g(x_k)^T (u - x_k)\right]$$

$$x_{k+1} = \arg\min_u h(u) + \frac{1}{2\alpha}\|u - z_k\|_2^2 \qquad z_k = \underbrace{x_k - \alpha \nabla g(x_k)}_{\text{grad. step w/r/t } g}$$

$$\star = \arg\min_u \alpha h(u) + \frac{1}{2}\|u - z_k\|_2^2$$

$$= \operatorname{prox}_{\alpha h}(z_k) \quad \text{(proximal gradient methods)}$$

$$x_{k+1} = \text{prox}_{\alpha h}\left(x_k - \alpha \nabla g(x_k)\right) = \arg\min_{u} \ \alpha h(u) + \frac{1}{2}\|u - z_k\|_2^2$$

Subgradient $\partial h(x) \ \{ v \mid h(w) \geq h(x) + v^T(w - x) \ \forall \ w \in \mathbb{R}^n \}$

$$= g(x) + \nabla g(x)^T(w - x) + \bar{w}^T H(x) w$$

repeat