Today: Advanced Topics

1. Coordinate Descent
2. Equality-constrained Newton
3. Interior-Point Methods (out of scope)

Coordinate Descent

- Sequential version:

$$F(\vec{x}) = g(\vec{x}) + \sum_{i=1}^{n} h_i(x_i)$$

g: convex, diffable

h_i: convex

LASSO: $F(\vec{x}) = \frac{1}{2} \| A\vec{x} - \vec{b} \|_2^2 + \underbrace{\lambda \|\vec{x}\|_1}$

↳ non-diffable, but seperable ie, $\sum |x_i|$

↳ LASSO is a type of problem where this sequential version of coordinate descent works really well

Coordinate Descent:

Initial guess: $\vec{x}^{(0)} = [x_1^{(0)} \; x_2^{(0)} \cdots x_n^{(0)}]$

K^{th} guess for x_1:

$$x_1^{(K)} = \underset{x_1}{\text{argmin}} \; F(x_1, x_2^{(K-1)}, \ldots, x_n^{(K-1)})$$

← minimization variable

fixed terms

doing a minimization over scalar x_1

eg 1^{st} guess for x_1:

$$x_1^{(1)} = \underset{x_1}{\text{argmin}} \; F(x_1, x_2^{(0)}, x_3^{(0)}, \ldots x_n^{(0)})$$

↳ get this from previous step

Algorithm:

1. $x_1^{(K)} = \underset{y}{\text{argmin}} \; F(y, x_2^{(K-1)}, \ldots, x_n^{(K-1)})$
2. $x_2^{(K)} = \underset{y}{\text{argmin}} \; F(x_1^{(K)}, y, x_3^{(K-1)}, \ldots, x_n^{(K-1)})$

$\vdots$

n. $x_n^{(K)} = \underset{y}{\text{argmin}} \; F(x_1^{(K)}, x_2^{(K)}, \ldots, x_{n-1}^{(K)}, y)$

LASSO outline w/coordinate descent:

$$f(\vec{x}) = \underbrace{\tfrac{1}{2}\|A\vec{x} - \vec{b}\|_2^2}_{\text{quadratic Fcn of } x_1} + \underbrace{\lambda\|x\|_1}_{\lambda|x_i|}$$

$\nabla_{x_1} F$

↳ cases $x_1 > 0$ (compute derivative) $x_2 < 0$, $x_2 = 0$

↳ compute derivative

↳ do soft-thresholding coordinate by coordinate

↳ converges as $k \to \infty$ to optimum point

- Note: a lot of times, convergence is determined by looking at the duality gap.

→ ie, solve primal & corresponding dual → see gap btwn dual feasible & point we just found

Newton's Method for Equality-constrained QPs

- Equality-constrained QPs → unconstrained QPs

$$\min \tfrac{1}{2}\vec{x}^T H \vec{x} + \vec{c}^T\vec{x} + d \qquad H \text{ PD} \to \text{strong duality}$$
$$\text{s.t. } A\vec{x} = \vec{b}$$

$$L(\vec{x},\vec{\nu}) = \tfrac{1}{2}\vec{x}^T H\vec{x} + \vec{c}^T\vec{x} + d + \vec{\nu}^T(A\vec{x} - \vec{b})$$

KKT conditions: (necessary & sufficient bc convex & strong duality)

↳ no λ^* conds bc no inequalities

$$\begin{cases} A\vec{x}^* = \vec{b} \\ H\vec{x}^* + \vec{c} + A^T\vec{\nu}^* = 0 \end{cases}$$

→ linear:

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \vec{x}^* \\ \vec{\nu}^* \end{bmatrix} = \begin{bmatrix} -\vec{c} \\ \vec{b} \end{bmatrix}$$

if $\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}$ is non-singular/invertible, we have a unique solution

$$\min \; F(\vec{x}) \quad \text{s.t.} \quad A\vec{x} = \vec{b}$$

[Say I start with some $\vec{x}_0$ s.t. $A\vec{x}_0 = \vec{b}$]

$$\min \; \hat{F}(\vec{x}_0 + \vec{v}) = \underbrace{F(\vec{x}_0)}_{\text{const}} + \underbrace{\nabla F(\vec{x}_0)^T \vec{v}}_{\text{linear}} + \underbrace{\frac{1}{2} v^T \nabla^2 F(\vec{x}_0) \vec{v}}_{\text{quadratic}}$$

$$\text{s.t.} \quad A(\vec{x}_0 + \vec{v}) = \vec{b}$$

$$A\vec{v} = \vec{0}$$

→ want to find that $\vec{v}$ that minimizes the quadratic

solution: $\begin{bmatrix} \nabla^2 F(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \vec{v} \\ \vec{\nu} \end{bmatrix} = \begin{bmatrix} -\nabla F(x_0)^T \\ 0 \end{bmatrix}$

→ bc you've started in the feasible set & you're requiring $A\vec{v} = \vec{0}$, we always stay in the feasible set

→ the step remains feasible every time