Convex f & g_i

$$\min f(x) \quad \text{s.t. } g_i(x) \leq 0$$

option 1

$$F(x) = f(x) + \sum_i I_{g_i}(x)$$

option 2

$$F(x) = f(x) + \max_i \min_{u: g_i(u) \leq 0} \|x - u\|_2$$

$$\min_x F(x)$$

Either obtain solution by setting $g_F(x) = 0$ & solving for x <u>OR</u> applying a descent algorithm

Convex f, g_i

$$\min f(x) \quad \text{s.t. } g_i(x) \le 0$$

option 1

$$F(x) = f(x) + \sum_i \underbrace{I_{g_i}(x)}_{\text{subgradient = normal cone (on bndry)}}$$

option 2

$$F(x) = f(x) + \max_i \underbrace{\min_{u : g_i(u) \le 0} \|x-u\|_2}_{\|x - P_{g_i}(x)\|}$$

$$\min_x F(x)$$

Either obtain solution by setting $g_F(x) = 0$ & solving for x OR applying a descent algorithm.

option 3

$$F_t(x) = f(x) + \left(-\frac{1}{t}\right) \sum_i \log(-g_i(x))$$

$$x^*(t) = \underset{x}{\operatorname{argmin}}\ F_t(x)$$

Barrier method

It turns out that analysing Barrier method (or analysing convergence of prox/projected gradient descent) becomes meaningful if we understand conditions for optimality for constrained opt. . .

option 4

$$F_k(x) = f_{Q_k}(x) + \sum_i I_{g_i}(x)$$

$$x^{k+1} = \underset{x}{\operatorname{argmin}}\ F_k(x)$$

Projected gradient method

Recall: $f_{Q_k}(x)$ = Quadratic approx to f around x^k

$$= f(x^k) + \nabla^T f(x^k)(x - x^k) + \frac{\|x - x^k\|^2}{2t}$$

$$\therefore x^{k+1} = \underset{x}{\operatorname{argmin}}\ \frac{1}{2t}\left\|x - (x^k - t\nabla f(x^k))\right\|^2 + \sum_i I_{g_i}(x)$$

$$= \underset{x : g_i(x) \le 0}{\operatorname{argmin}} \|x - \hat{x}^{k+1}\|^2$$

$$= P_{C_1 \cap C_2 .. \cap C_m}(\hat{x}^{k+1})$$

More generally, the 4th option: called projected gradient descent

Proximal gradient descent.

$\min f(x) + r(x)$

A = features
x = feature wts
in Lasso

f is differentiable
$\|Ax-y\|^2$

r is not differentiable
eg: $\lambda\|x\|_1$

Iteratively solve: $x^{(0)}$

$$x^{(k+1)} = \min_x f(x^{(k)}) + \nabla f(x^{(k)})(x - x^k) + \frac{1}{2t}\|x - x^k\|^2 + r(x)$$

until convergence

For our problem: $x^{(k+1)} = \min_x \|x-y\|^2 \ldots \|x\|_1$

H/w: complete & reduce to known problem

Recall: $\min_x \|y-x\|^2 + \lambda\|x\|_1$ had a closed form optimal soln:

$$x^*_i = \begin{cases} y_i + \lambda & \text{if } y_i < -\lambda \\ -y_i + \lambda & \text{if } y_i > \lambda \\ 0 & \text{o/w} \end{cases}$$

Obtained by setting a subgradient to 0.

Projected gradient descent is prox gradient descent when $r(x) = I_\Omega(x)$