Last Time:

- Constrained Optimization
- KKT Conditions
- Line Search

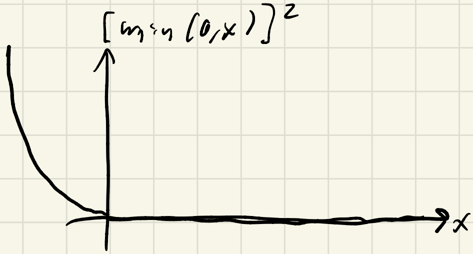
Today:

- Augmented Lagrangian
- Quadratic Programs
- More one Regularization/Line Search

Penalty Method:

- Replace inequalities with objective term that penalizes violations:

$$\left.\begin{aligned} &\min_x \ f(x) \\ &\text{s.t. } c(x) \geq 0 \end{aligned}\right\} \longrightarrow \min_x \ f(x) + \frac{\rho}{2}\left[\min(0, c(x))\right]^2$$



- Easy to implement
- Large penalties → ill-conditioning
- Difficult to achieve high accuracy

* Augmented Lagrangian

- Add Lagrange multiplier estimate to penalty method:

$$\min_x \underbrace{f(x) - \tilde{\lambda}^T c(x) + \frac{\rho}{2}[\min(0, c(x))]^2}_{L_\rho(x, \tilde{\lambda}) \text{ "Augmented Lagrangian"}}$$

- Update $\tilde{\lambda}$ by "offloading" penalty into multiplier at each iteration:

$$\frac{\partial f}{\partial x} - \tilde{\lambda}^T \frac{\partial c}{\partial x} + \rho c(x)^T \frac{\partial c}{\partial x} = \frac{\partial f}{\partial x} - \underbrace{[\tilde{\lambda} - \rho c(x)]}^T \frac{\partial c}{\partial x} = 0$$

$$\Rightarrow \tilde{\lambda} \leftarrow \tilde{\lambda} - \rho c(x)$$

(for active constraints)

- Repeat until convergence:

1) $\min_x L_\rho(x, \tilde{\lambda})$

2) $\tilde{\lambda} \leftarrow \max(0, \tilde{\lambda} - \rho c(x))$ ← clamp to guarantee non-negativity

3) $\rho \leftarrow \alpha \rho$ (optional)

← Typically ≈ 10

- Fixes ill-conditioning of penalty method
- Converges fast (super linear)
- Works well on non-convex problems

* Quadratic Programs

$$\min_x \tfrac{1}{2}x^T Q x + q^T x \quad , \quad Q > 0$$

$$\text{s.t.} \quad Ax \le b$$
$$\quad Cx = d$$

- Very useful in control
- Can be solved very fast (e.g. ~KHz)

* Regularization + Duality

- Given:

$$\min_x f(x)$$
$$\text{s.t.} \quad c(x) = 0$$

- We might like to turn this into:

$$\min_x f(x) + P_\infty(c(x)), \quad P_\infty(x) = \begin{cases} 0, & x = 0 \\ +\infty, & x \neq 0 \end{cases}$$

- Practically terrible, but we can get the same effect by solving:

$$\min_x \max_\lambda f(x) + \lambda^T c(x)$$

- Whenever $c(x) \neq 0$, inner problem blows up
- Similarly for inequalities:

$$\left.\begin{array}{l} \min f(x) \\ \text{s.t. } c(x) \geq 0 \end{array}\right\} \Rightarrow \min_x f(x) + P_\infty^+(c(x))$$

$$P_\infty^+(x) = \begin{cases} 0, & x \geq 0 \\ +\infty, & x < 0 \end{cases}$$

$$\Rightarrow \min_x \max_{\lambda \geq 0} \underbrace{f(x) - \lambda^T c(x)}_{L(x,\lambda)}$$

- Aside: for convex problems I can switch the order of min $\leftrightarrow$ max and get the same answer (duality). Not true in general.

- Interpretation: KKT conditions define a saddle point in (x, λ)

- KKT system should have dim(x) positive eigenvalues and dim(λ) negative eigenvalues at an optimum. Called "Quasi-definite"

⇒ When regularizing a KKT system, the lower-right block should be negative:

$$\begin{bmatrix} \boxed{H+\alpha I} & C^T \\ C & \boxed{-\alpha I} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} -\nabla_x L \\ -C(x) \end{bmatrix}, \quad \alpha > 0$$

* Example:

- Still have overshoot ⇒ need line search

* Merit Function

- How do we do a line search on a root-finding problem?

$$\text{find } x^* \quad \text{s.t. } C(x^*) = 0$$

- Define a scalar "merit function" $P(x)$ that measures distance from a solution

- Standard Choices:

$$P(x) = \frac{1}{2} C(x)^T C(x) = \frac{1}{2} \|C(x)\|_2^2$$

$$P(x) = \|C(x)\|_1 \quad \text{(any norm works)}$$

- Now just do Armijo on $P(x)$:

$\alpha = 1$

while $P(x+\alpha\Delta x) > P(x) + b\,\alpha\,\nabla P(x)^T \Delta x$

step length

tol

$\alpha \leftarrow c\,\alpha$

end

$0 < c < 1$

$x \leftarrow x + \alpha \Delta x$

- How about constrained minimization?

$$\left.\begin{aligned} &\min_x \ f(x) \\ &\text{s.t.} \ \ c(x) \geq 0 \\ &\qquad d(x) = 0 \end{aligned}\right\} \quad L(x,\lambda,\mu) = f(x) - \lambda^T c(x) + \mu^T d(x)$$

- Lot's of options for merit functions:

$$P(x,\lambda,\mu) = \frac{1}{2}\| \underbrace{\nabla L(x,\lambda,\mu)}_{\text{KKT residual}} \|_2^2$$

$$\text{KKT residual} = \begin{bmatrix} \nabla_x L(x,\lambda,\mu) \\ \min(0, c(x)) \\ d(x) \end{bmatrix}$$

$$P(x,\lambda,\mu) = f(x) + \rho \left\| \begin{bmatrix} \min(0, c(x)) \\ d(x) \end{bmatrix} \right\|_1$$

scalar to trade objective vs. constraints

any norm works

$$P(x, \lambda, \mu) = f(x) - \tilde{\lambda}^T c(x) + \tilde{\mu}^T d(x) + \frac{\rho}{2} \| \min(0, c(x)) \|_2^2 + \frac{\rho}{2} \| d(x) \|_2^2$$

(augmented Lagrangian)

* Example Take Away Messages:

- $P(x)$ based on KKT residual is expensive
- Excessively large constraint penalties can cause problems
- AL methods come with a merit function for free.