Last Time:

- Inequality Constraints
- Interior-Point Methods

Today:

- Revisit regularization + line search
- Control History

---

* Duality + Regularization

- Given:

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

- We can theoretically turn this into:

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

- 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

- Similar for inequalities:

$$\left.\begin{aligned} &\min_x \; f(x) \\ &\text{s.t.} \; c(x) \geq 0 \end{aligned}\right\} \rightarrow \min_x \; f(x) + P_\infty^-(c(x))$$

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

$$\Rightarrow \min_x \max_{\lambda \geq 0} \; f(x) - \lambda^T c(x)$$

- Whenever $c(x) < 0$, inner problem blows up.
- Interpretation: KKT conditions define a saddle point in (x, λ)
- KKT system should have $\dim(x)$ positive eigenvalues $\dim(\lambda)$ negative eigenvalues at an optimum. (Called "quasi-definite")

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

* Example:

  - Still overshoot $\Rightarrow$ need line search

---

* Merit Function

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

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

  - Define scalar "merit function" $P(x)$ that measures distance to 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 $P(x)$:

$$\begin{aligned}
&\alpha = 1 \\
&\text{while } P(x + \alpha \Delta x) > P(x) + \overset{\text{tolerance}}{b}\,\underset{\text{step length}}{\alpha} \nabla P(x)^T \Delta x \\
&\qquad \alpha \leftarrow \theta \alpha \\
&\text{end}
\end{aligned}$$

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

- How about constrained optimization?

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

- Lots of options for merit functions!

$$P(x, \lambda, \mu) = \frac{1}{2} \| \underbrace{r_{kkt}(x, \lambda, \mu)}_{\text{KKT Residual } \begin{bmatrix} \nabla_x L \\ \min(0, c(x)) \\ d(x) \end{bmatrix}} \|_2^2$$

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

any norm works