Last Time:

- Constrained minimization

Today:

- Regularization + Line Search with constraints
- Deterministic Optimal Control

---

* Regularization + Duality

- Given:

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

- We might like to turn this into:

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

- Practically, this is terrible. But we can get the same effect by solving:

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

- Whenever $c(x) \neq 0$, inner problem gives $+\infty$
- Similarly for inequalities:

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

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

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

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

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

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

$\Rightarrow$ When regularizing a KKT system, the lower-right block should be negative:

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

* Example

- Still have overshoot $\Rightarrow$ need line search

* Merit Function

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

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

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

- Standard Choices:

$$P(x) = \tfrac{1}{2} c(x)^T c(x) = \tfrac{1}{2} \| c(x) \|_2^2$$

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

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

$$
\begin{aligned}
&\alpha = 1 \\
&\text{while } \; P(x+\alpha \Delta x) > P(x) + b\,\alpha\, \nabla P(x)^T \Delta x \\
&\qquad \alpha \leftarrow c\,\alpha \\
&\text{end} \\
&\quad x \leftarrow x + \alpha \Delta x
\end{aligned}
$$

tolerance, step length

- How about constrained minimization?

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

- Lots of options for merit functions:

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

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

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

ρ: scalar to trade off objective vs. constraint satisfaction

$\|\cdot\|_1$: 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 penalty weights can cause problems
- AL methods come with a merit function for free

★ Deterministic Optimal Control

- Continuous time:

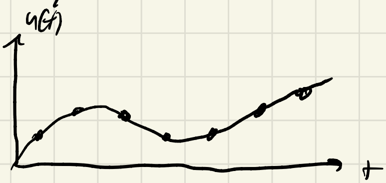
$$\min_{\substack{x(t)\\u(t)}} \underbrace{J(x(t), u(t))}_{\text{cost function}} = \int_{t_0}^{t_f} \overbrace{L(x(t), u(t))}^{\text{"stage cost"}}\, dt + \overbrace{L_F(x(t_f))}^{\text{"terminal cost"}}$$

$$\text{s.t.} \quad \dot{x}(t) = f(x(t), u(t)) \qquad \text{"dynamics constraint"}$$

(possibly other constraints)

"state/input trajectories"

- This is an "infinite dimensional" problem in the following sense:



$$u_{1:N} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \Rightarrow u(t) = \lim_{N \to \infty} u_{1:N}$$

- Solutions are open-loop trajectories

- There are a handful of problems with analytic solutions in continuous time, but not many

- We will focus on the discrete-time setting which leads to tractable algorithms

- Discrete Time:

$$\min_{\substack{x_{1:N} \\ u_{1:N-1}}} J(x_{1:N}, u_{1:N-1}) = \sum_{n=1}^{N-1} L(x_n, u_n) + L_F(x_N)$$

$$\text{s.t.} \quad x_{n+1} = f(x_n, u_n)$$

$$u_{min} \leq u_n \leq u_{max}$$ ← "torque limits"

$$c(x_n) \leq 0$$ ← obstacle/collision constraints

- This is now a finite-dimensional problem
- Samples x_n, u_n are often called "knot points"
- Convert continuous → discrete using integration methods (e.g. Runge-Kutta, etc.)
- Convert discrete → continuous using interpolation