2/10: The Gauss-Newton method

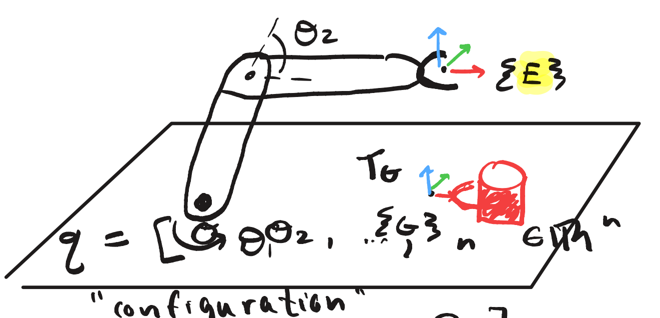
Announcements:

- HW 1 due 2/12 (Thurs)
- Project proposal due 2/19 (Thurs)
  - Groups of ≤ 3 students
  - 1p. (max.) proposal of topics
  - e.g., apply to your research, implement paper

Today:

- nonlinear least-squares
- Gauss-Newton
- matrix-free GN via autodiff

Problem: Inverse Kinematics (IK)



$q = [\theta_1, \theta_2, \dots, \theta_n] \in \mathbb{R}^n$

"configuration"

$$x \in SE(3) = \begin{bmatrix} R & p \\ 0 & 1 \end{bmatrix}$$

"forward kinematics"

$FK(q): \mathbb{R}^n \to SE(3)$

$$\min_{q \in \mathbb{R}^n} \| FK(q) \ominus x_{des} \|^2$$

$q_0 \to q_1 \to \dots \to q_n$

$FK(q_n) = x_{des}$

$f: \mathbb{R}^n \to \mathbb{R}$

Recall: Descent methods

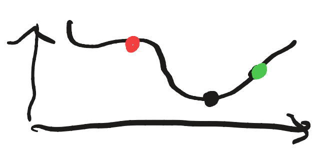
$$q_{k+1} \leftarrow q_k + \alpha_k d_k$$

step length, descent direction

i) Gradient descent: $d_k = -\nabla f(q_k)$

ii) Newton's method $\nabla^2 f(q_k) d_k = -\nabla f(q_k)$

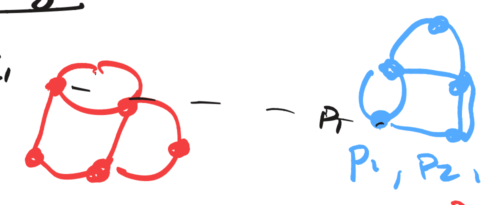
iff $\nabla^2 f(q_k) \succ 0$, d_k is a descent



Today: Nonlinear least-squares

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \sum r_i(x)^2 \qquad r_i: \mathbb{R}^n \to \mathbb{R}$$

e.g.) (Point cloud alignment)



$$\min_{\substack{R \in SO(3) \\ t \in \mathbb{R}^3}} \frac{1}{2} \sum_{i=1}^{N} \| R q_i + t - p_i \|^2$$

$\sum_i r_i(x)^2$

e.g.) (Inverse Kinematics)

$$\min_{q \in \mathbb{R}^n} \| FK(q) \ominus x_{des} \|^2 = \frac{1}{2} \vec{r}(q)^T \vec{r}(q)$$

$$\begin{bmatrix} p(q) - p_{des} \\ \log(R(q)^T R_{des})^\vee \end{bmatrix} = \vec{r}(q)$$

Suppose we want to solve this via Newton's

$$H(x_k) d_k = -g(x) \qquad H = \nabla^2 f(x_k), \quad g = \nabla f(x_k)$$

Two issues:

i) $H \nsucc 0 \Rightarrow d_k$ is not a descent dir

ii) $d_k = -H_k^{-1} g_k$ (assume H_k^{-1} exists, $H_k \succ 0$)

inverses are expensive $O(n^3)$ flops

$d_k = -H_k / g_k$ np.linalg.solve

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \sum_i r_i(x)^2 = \frac{1}{2} r(x)^T r(x)$$

$$r(x) = \begin{bmatrix} r_1(x) \\ r_2(x) \\ \vdots \\ r_N(x) \end{bmatrix} : \mathbb{R}^n \to \mathbb{R}^N$$

$$\nabla f(x) = \nabla_x \frac{1}{2} r(x)^T r(x) = J(x)^T r(x)$$

$$\nabla^2 f(x) = J(x)^T J(x) + \sum_{i=1}^{N} r_i(x) \nabla^2 r_i(x)$$

≈ 0

Assume $r_i(x) \cong 0$ when near an opt.

$$\nabla^2 f(x) \approx J(x)^T J(x)$$

$$\boxed{J^T(x) J(x) d = -J(x)^T r(x)}$$

Gauss-Newton update

Improvement: Levenberg-Marquardt (LM)

Add a small λD, D diagonal to make the system solveable

$$(J^T J + \lambda D) d = -J^T r$$

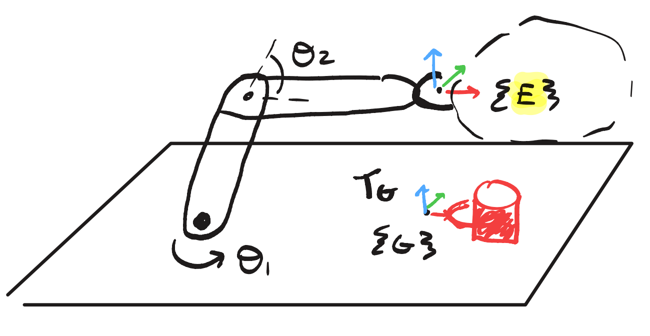
choices: $D = I$, $D = \mathrm{diag}(\mathrm{diag}(J^T J))$

$$v^T (J^T J + \lambda I) v = \| J v \|^2 + \underbrace{\lambda v^T v}_{} > 0$$

A nice interp: LM interpolates between GD and Gauss-Newton

as $\lambda \to \infty$, $(\cancel{J^T J} + \lambda I) d = -J^T r$ (with $\cancel{J^T J} \to 0$)

$$d \approx \frac{1}{\lambda}(-J^T r) = \frac{1}{\lambda} - \nabla f(x)$$



$$\min_{q_{1:T}} \sum_t \| FK(q_t) \ominus x_{des}^t \|^2 + \lambda\, \mathrm{lim}(q_t)^2 + \mathrm{vel}(q_t)^2$$

$q_t \in \mathbb{R}^{n \cdot T}$