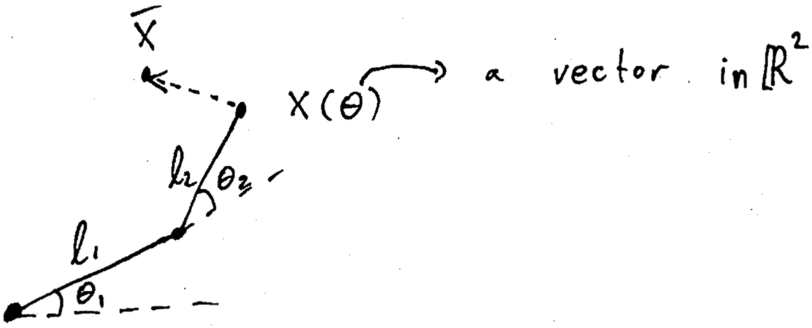
2D, 2-link Inverse Kinematics problem:



$$X(\theta) = \begin{bmatrix} l_1 \cos\theta_1 + l_2 \cos(\theta_1+\theta_2) \\ l_1 \sin\theta_1 + l_2 \sin(\theta_1+\theta_2) \end{bmatrix}$$

or

$$X(\theta) = \begin{bmatrix} l_1 \cos\theta_1 + l_2(\cos\theta_1 \cos\theta_2 - \sin\theta_1 \sin\theta_2) \\ l_1 \sin\theta_1 + l_2(\sin\theta_1 \cos\theta_2 + \cos\theta_1 \sin\theta_2) \end{bmatrix}$$

Recall:

$\cos(a+b) = \cos a \cos b - \sin a \sin b$

$\sin(a+b) = \cos a \sin b + \sin a \cos b$

$\dfrac{d \cos a}{da} = -\sin a$

$\dfrac{d \sin a}{da} = \cos a$

$$\frac{dX}{d\theta} = \begin{bmatrix} \frac{dX_x}{d\theta_1} & \frac{dX_x}{d\theta_2} \\ \frac{dX_y}{d\theta_1} & \frac{dX_y}{d\theta_2} \end{bmatrix} = \begin{bmatrix} \frac{dX}{d\theta_1} & \frac{dX}{d\theta_2} \end{bmatrix} = J$$

→ an $n \times m$ matrix, where n is dimension of X & m is dimension of θ: X is an n-dimensional vector, θ an m-dimensional vector.

$$\frac{dX}{d\theta_1} = \begin{bmatrix} -l_1 \sin\theta_1 + l_2(-\sin\theta_1 \cos\theta_2 - \cos\theta_1 \sin\theta_2) \\ l_1 \cos\theta_1 + l_2(\cos\theta_1 \cos\theta_2 - \sin\theta_1 \sin\theta_2) \end{bmatrix}$$

$$\frac{dX}{d\theta_2} = \begin{bmatrix} l_2(-\cos\theta_1 \sin\theta_2 - \sin\theta_1 \cos\theta_2) \\ l_2(-\sin\theta_1 \sin\theta_2 + \cos\theta_1 \cos\theta_2) \end{bmatrix}$$

The energy function to minimize:

$$E(\theta) = \frac{1}{2}(x(\theta)-\bar{x})^T\underbrace{(x(\theta)-\bar{x})}_{V}$$ → non-linear least squares problem.

$$\nabla E = \frac{dE}{d\theta} = \underbrace{\frac{dx}{d\theta}^T}_{m \times n} \cdot \underbrace{\frac{dE}{dx}}_{n \times 1} = \frac{dx}{d\theta}^T \cdot V = J^T V$$ (the Jacobian transpose method for IK is just gradient descent)

How can one check derivation / implementation of derivatives / gradients / Jacobians / Hessians?

· Test against finite difference estimates!

Central finite differences:

Consider a function $f(x): \mathbb{R}^n \to \mathbb{R}^m$

→ e_i a vector defined as $e_i[j] = \begin{cases} 1, & \text{if } i=j \\ 0, & \text{otherwise} \end{cases}$

$$f(x + h \cdot e_i) \doteq f(x) + \underbrace{h \cdot e_i^T \frac{df}{dx}}_{\substack{\text{isolates } i^{th} \text{ component} \\ \text{of } \frac{df}{dx} \text{ (i.e. } \frac{\partial f}{\partial x_i})}} + \frac{1}{2} h^2 e_i^T \frac{d}{dx}\frac{df}{dx} e_i + O(h^3)$$

$$f(x - h \cdot e_i) \doteq f(x) - h \cdot e_i^T \frac{df}{dx} + \frac{1}{2} h^2 e_i^T \frac{d}{dx}\frac{df}{dx} e_i + O(h^3)$$

Subtracting the second eq. from the first:

$$\frac{\partial f}{\partial x_i} = e_i^T \frac{df}{dx} \simeq \frac{f(x+h\cdot e_i) - f(x-h\cdot e_i)}{2h}$$ (approximation error $\sim O(h^3)$)

Why regularization works (and how):

Write out a vector g in basis defined by eigen vectors of H:

$$g = \sum_i \underbrace{\alpha_i}_{\text{coordinates of vector } g \text{ in basis defined by eigen vectors}} \overbrace{V_i}^{i\text{th eigen vector of } H}$$

$$H \cdot g = H \sum_i \alpha_i V_i = \sum_i \alpha_i H \cdot V_i = \sum_i \alpha_i \cdot \overbrace{\lambda_i}^{i\text{th eigenvalue of } H} \cdot V_i$$

$$g^T \cdot H \cdot g = \sum_i \alpha_i^2 \lambda_i$$ (Note $V_i^T V_j = 0$ if $i \neq j$, 1 otherwise)

$g^T H g$ can only be negative (i.e. search direction is not a descent direction) if some of the eigen values of H are negative (i.e. if g is "more aligned" with eigenvectors that have negative eigen values).

If $H \succ 0$, we always get a descent direction!

Goal is therefore to eliminate negative eigen values!

In practice, hard to isolate precisely the negative eigen values so that they can be truncated.

An alternative is to "shift" all the eigen values "up" if $H^{-1} g$ is not a descent direction:

$$\underbrace{H + I \cdot r}_{\text{shift all eigen values of } H \text{ up by "}r\text{" } (r>0).} = V \Sigma V^T + r V I V^T = V(\Sigma + rI)V^T$$

If $r \gg \lambda_i$ for all i, then $H + I \cdot r \sim I \cdot r$,

$\cancel{H^{-1}g}$ $(H + Ir)^{-1} \cdot g \sim \frac{1}{r} \cdot g$ (i.e. search direction is approaching a scaled version of the gradient as r gets larger and larger).

Another way to look at it:

$$E(\theta) = \frac{1}{2}(X(\theta) - \bar{X})^T (X(\theta) - \bar{X}) + \frac{1}{2} r \cdot \underbrace{(\theta - \bar{\theta})^T (\theta - \bar{\theta})}_{\text{don't deviate much from } \bar{\theta}.}$$

If $\bar{\theta}$ is what we are doing a Taylor expansion around, then:

$$\nabla_\theta E = J^T v + r \cdot \underbrace{(\theta - \bar{\theta})}_{0}$$

$$\nabla_\theta^2 E = \underbrace{J^T J + \frac{dJ}{d\theta} \cdot v}_{\text{the original Hessian}} + \underbrace{Ir}_{\text{regularizer contribution}}$$

Gauss Newton:
$H \sim J^T J$
$\therefore H^{-1} \cdot g = \underbrace{(J^T J)^{-1} J^T}_{\text{Jacobian pseudo-inverse method for IK. This is just Gauss-Newton}} \cdot v$

Another observation: for this problem, hessian has a special structure:

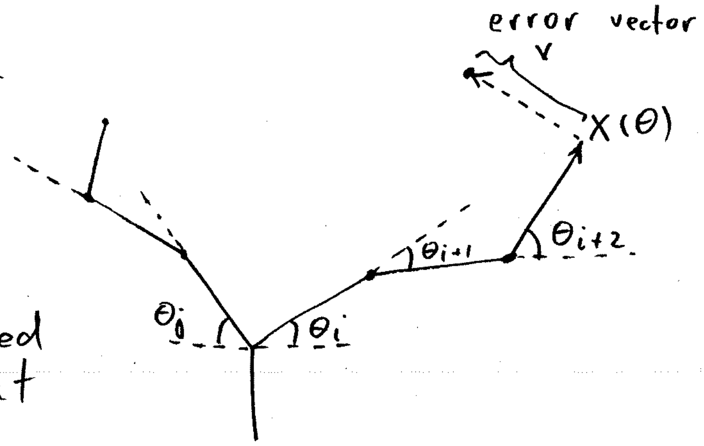
$$H = \underbrace{J^T J}_{\text{always semi-definite}} + \boxed{\frac{dJ}{d\theta} \cdot v}$$

source of indefinitness (negative eigen values) Just ignore it → Gauss Newton

always semi-definite: $\underbrace{v^T J^T}_{y^T} \cdot J \cdot v = y^T y \geq 0 \quad \forall v!$

A closer look at the Jacobian:

We have $x(\theta)$ representing world coordinates of an end-effector point, and



$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix}$$ is a vector of generalized or reduced coordinates that represent the pose.

$J \equiv \dfrac{dx}{d\theta}$ is the Jacobian that tells us how coords of $x(\theta)$ change with respect to $\theta_1, \theta_2, \ldots$

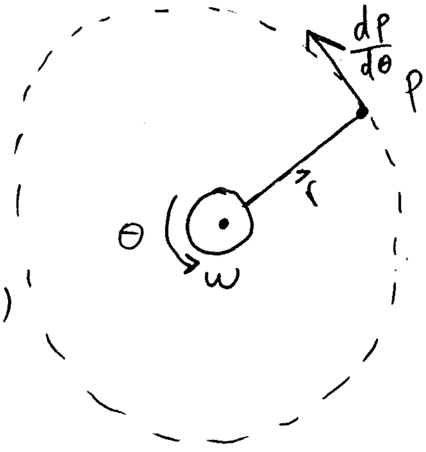
recall $$\underbrace{\nabla E}_{m \times 1} = J^T v = \underbrace{\frac{dx}{d\theta}^T}_{m \times 3} \cdot \underbrace{v}_{3 \times 1}$$

→ each element i of the gradient ∇E is a dot product between $\dfrac{dx}{d\theta_i}$ and v. It tells us how much the change you get in x by changing θ_i aligns with v. The more they align, the more θ_i should change to minimize IK objective.

$$\frac{dx}{d\theta}^T = \begin{bmatrix} \frac{dx_x}{d\theta_1} & \frac{dx_y}{d\theta_1} & \frac{dx_z}{d\theta_1} \\ \frac{dx_x}{d\theta_2} & \frac{dx_y}{d\theta_2} & \frac{dx_z}{d\theta_2} \\ & \vdots & \\ \frac{dx_x}{d\theta_m} & \frac{dx_y}{d\theta_m} & \frac{dx_z}{d\theta_m} \end{bmatrix}$$

$\dfrac{dx}{d\theta_2} \equiv$ how does position of end effector change as you "wiggle" θ_2.

Computing Jacobian entries:



Consider the following scenario: $\vec{r}$ is a vector orthogonal to (unit) rotation axis $\vec{\omega}$. How does p change as it rotates (in plane) about the origin?

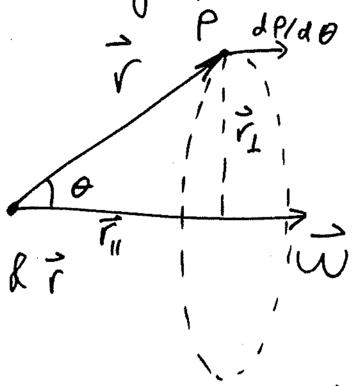
$$\frac{dp}{d\theta} \text{ is } \begin{cases} \text{- a vector} \\ \text{- orthogonal to } \vec{r} \\ \text{- orthogonal to } \vec{\omega} \end{cases}$$

Magnitude ?

if θ ~~th~~ changes by 2π, then p will have travelled a distance of $2\pi|\vec{r}|$. If θ changes by a small amount $d\theta$, then p travels by a proportional amount $dp = d\theta\,|\vec{r}|$

$$\therefore \left|\frac{dp}{d\theta}\right| = |\vec{r}|.$$

What if $\vec{r}$ is not orthogonal to $\vec{\omega}$?

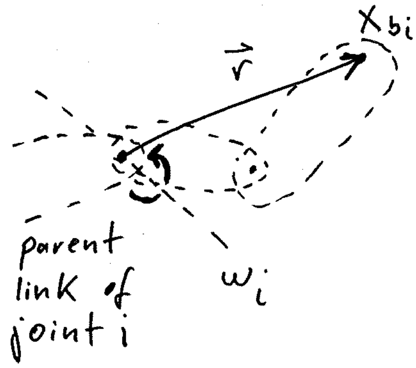


$$\vec{r} = \vec{r}_\parallel + \vec{r}_\perp$$

- $\vec{r}_\parallel$ does not change with θ.
- $\vec{r}_\perp$ has magnitude $|\vec{r}|\sin\theta$
- $\frac{dp}{d\theta}$ is orthogonal to $\vec{\omega}$, $\vec{r}_\perp$ & $\vec{r}$

Therefore, it is easy to see that $\underline{\frac{dp}{d\theta} = \vec{\omega} \times \vec{r}}$

So, for an articulated structure, compute vector from joint i to end effector x_{bi} using FK as we've seen before.



Then compute $\frac{dx_{bi}}{d\theta_i}$ using cross product. This is a vector in coordinate frame of of parent link, so bring that to world coordinates using FK to obtain $\frac{dx}{d\theta_i}$.

The Jacobian $\frac{dx}{d\theta}$ is usually sparse. Do you know why?