Feb Log

February 10, 2021

1 1/31

1.1 Chess

Remark 1. A blunder free game with weak positional moves:

`https://www.chess.com/analysis/game/live/6409740211?tab=analysis`

Remark 2. A complicated blunder filled game:

`https://www.chess.com/a/CbAJ8Wm4XAX8`

To Analyze:

1.2 Complex Analysis

2 2/1

2.1 Chess

Remark 3. Talk about a clean game:

`https://www.chess.com/a/Gzp6PJxWXAX8`

Remark 4. My first brilliant move!:

`https://www.chess.com/a/2BfrDrz2JXAX8`

2.2 Technical Animation

Interesting 1. TA Arjun is interested in PDEs and numerical simulation.

Remark 5. Course Website:

`http://graphics.cs.cmu.edu/nsp/course/15464-s21/www/`

Computer Animation: Algorithms and Techniques is the course textbook. In drive.

Question 1. Does greater physical simulation accuarcy lead to a less palatable viewing experience?

Answer 1. Not sure but often directors will personify animations and we have different parameters to give differenter personfications. For example "angry storm".

Answer 2. It seems exaggerated motion is often more digestestible(think actors for example). Often used actors in motion capture

Interesting 2. Rig Net: automatically rigging meshes. Note: rigging is process of jointing meshes, providing structure/skeleton.

Remark 6. Beginning of rigging: find medial axis of geometry and impose some structure.

2.3 On Lp Brunn-Minkowski Type Inequalities

Tag: BrunnMinkowski

Remark 7. V is $1/n$ concave measure w.r.t Minkowsi sum. Need normalizing $1/n$ powers

Prop 1.

$$h_{K+L}(u) = h_K(u) + h_L(u)$$

Remark 8. **Brascamp-Lieb**

$\alpha \geq -1/n, t \in [0,1]$. With $f, g, h : \mathbb{R}^n \to \mathbb{R}_+$ satisfy

$h((1-t)x + ty) \geq [(1-t)f(x)^\alpha + tg(y)^\alpha]^{1/\alpha}$ then

$\int_{R^n} h(x)dx \geq [(1-t)(\int_{R^n} f(x)dx)^{\alpha/1+n\alpha} + t(\int_{R^n} g(x)dx)^{\alpha/1+n\alpha}]^{1+n\alpha/\alpha}$

Prekopa lindler is $\alpha = 0$

Prop 2.

$$(1-t)X_s 1_A \oplus_s tX_s 1_B = 1_{(1-t)A+tB}$$

Remark 9. Changing operator: minkwoski sum, to l_p variants.

Remark 10. Also some kind of interplay beteween functional inequalities and volume inequalities. Between supremal convolutions and Lp minkwoski sums.

2.4 PDEs and Data Analysis

TAG: OptimalTransport

Theme 1. The more assumptions you make on a measure the better approximation you can achieve

Interesting 3. Shimaa is interested in stochastic BDEs. Wes interested in foundations of machine learning.

Theme 2. Look at a measure as some kind of energy landscape and the transport map as the process of rearranging mass.

Remark 11. Often transportation cost is $|x-y|^p$.

Remark 12. Optimal transport minimizes transportation cost.

Theme 3. Goal is to find weaker problem which provides good solution to wider class of subproblems.

3 2/2

3.1 Goals

1. Chess: 1300 in blitz
2. Research: 3 hours worked, some progress, email tkocz
3. Thesis: 5 pages
4. Homework: Animation
5. Get glenn to agree to a time

3.2 DRL

TAG: DRL

Question 2. what is computational design?

Remark 13. Course link:

```
https://cmudeeprl.github.io/403_website/
```

Remark 14. Katerina F.

```
"My genes have strong priors from the world"
```

Remark 15. Inconsistent rewards lead to addiction.

Remark 16. For a long time large emphasis on discovering new behaiors in DRL. Now thinking we need to develop behavior repetoire and associate with some stimuli.

Remark 17. Curiosity, a desire to see new things, very intrinsically powerful.

Remark 18. Conor Igoe:

```
For a fixed known opponent, the evolution of chess is markovian from the perspective of the ma
```

In some cases(such as driving) we need multiple frames/time steps to even attempt to play. But this can also be redefined as markovian by letting states correspond to multiple time steps.

Remark 19. Model vs. non-model based. Can we learn via simulation or not.

Remark 20. Cannot use gradient based optimization often in DRL. We can if we have a model.

3.3 The Embodiment Hypothesis

Remark 21. Link:

`https://cogdev.sitehost.iu.edu/labwork/6_lessons.pdf`

Remark 22. The six lessons from child deveopment:

1. Be multimodal

3.4 Modeling Evolution

Remark 23. Selection or drift: tug of war between determinism and randomness.

3.5 Chess

Remark 24. For tactics, look for forcing moves.

Remark 25. Backrank pawns are massive!!!

`https://www.chess.com/puzzles/problem/1227605`

4 2/3 and 2/4

4.1 Goals

2/3

1. Research 3 hours
2. Thesis 5 pages
3. 1300 blitz

2/4

1. Research 3 hours
2. Thesis 5 pages
3. 1300 blitz

4.2 Technical Animation

TAG: TechnicalAnimation

Remark 26. L-systems developed to describe plant structures and generation.

Remark 27. Tools for good animation: The Anmimators Survival Kit.

Remark 28. Idea behind rigging: for easy animating want ball control points you can manipulate for convenience.

Remark 29. Cloth simulation involves a mesh... Cloth intersection problems in Pixar's Coco:

`https://www.researchgate.net/publication/326907399_Better_collisions_and_faster_cloth_for_Pixa`

Remark 30. Traditional animation: keyframing.

New variant: procedural animation. Often used for crowd animation.

Interesting 4. Interesting site:

`www.massivesoftware.com`

Interesting 5. Character controller using Motion VAEs interesting.

4.3 Complex Analysis

TAG: ComplexAnalysis

Remark 31. Cauchy riemann equations derived via simply differentiating f as a function of two varibles in real and comlex directions.

Question 3. Cauchy riemann conditions are necessary. Are they sufficient?

Answer 3. Yes.

Theorem 1. *Let $f : \Omega \to \mathbb{C}$ with $f = u + iv$ and satisfying cauchy riemann. Then f is holomorphic.*

Proof. Fix $z_0 = x_0 + iy_0$.

$$u(x_0 + h_1, y_0 + h_2) = u(x_0, y_0) + u_x h_1 + u_y h_2 + o(h)$$

and similarly for v. Then write $f(z_0 + h) - f(z_0)$ in terms of above and massage using cauchy riemann to get form $ah + o(h)$. □

Remark 32. Determinant of jacobian is really magnitude of norm of complex derivative squared.

Example 1. $f(x, y) = \sqrt{|x||y|}$ satisfies cauchy riemann but is not holomorphic.

Theorem 2. *Radius of convergence R of power series is*

$$R = \frac{1}{limsup|a_n|^{1/n}}$$

Proof. Idea is to compare to geometric series. Set R as desired. Then just compute(since we used limsup) and see that geometric series converges and or diverges in desired cases. This is also why we have problems on boundary. □

4.4 Chess

Remark 33. A beautiful positional/material tradeoff emerged:

`https://www.chess.com/a/36gbqDERtXAX8`

After analysis apparently not that good?

Remark 34. The nastiest checkmate I've ever given:

`https://www.chess.com/analysis/game/live/6441231672?tab=analysis`

Remark 35. Sharp tactic game:

`https://www.chess.com/analysis/game/live/6442135074?tab=analysis`

Remark 36. Playing more interesting games:

`https://www.chess.com/analysis/game/live/6443227668?tab=analysis`

Remark 37. Really shouldn't have resulted in a pawn structure that lead to a passed pawn for opponent

`https://www.chess.com/analysis/game/live/6443636791?tab=analysis`

4.5 PDE and Data

TAG: OptimalTransport

Remark 38. Villani's Optimal Transport: New and Old

`https://ljk.imag.fr/membres/Emmanuel.Maitre/lib/exe/fetch.php?media=b07.stflour.pdf`

Presented from a probabilistic perspective.

Remark 39. Via CoV, condition for transport map:

$$\rho(x) = \eta(T(x))|det(DT(x))|$$

Prop 3. *Change of variables formula justified via above:*

$$\int_Y f(y)d\nu(y) = \int_X f(T(x))d\mu(x)$$

Question 4. When does measure preserving map even exist between μ on X and ν on Y.

Example 2. Consider to the above the non-example $\mu(x) = 1$, $\nu(y_1) = \nu(y_2) = 1/2$.

Remark 40. Transport plan is generalization of transport map that has the source and target measures as its marginals. If we have a transport map then $(I \times T)_\sharp \mu$ is a transport plan. Effectively pushing measure onto graph of T(note I is identity).

Question 5. Why does $\mu \times \nu$ not always work?

Answer 4. It does.

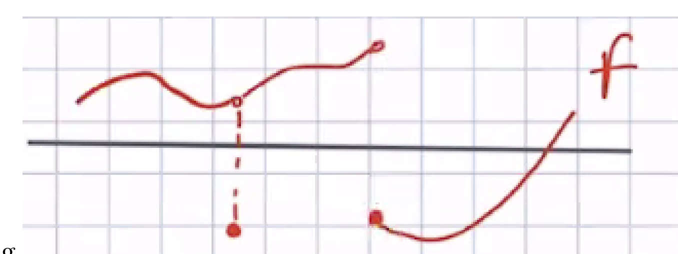
Remark 41. Optimal Transport cost in terms of plans is always well defined since a plan exists and infimum is always a min(this is the Kantorovich formulation, first one is monge formulation)

$$C(\pi) = \int_{X\times Y} c(x,y)d\pi(x,y) = \int_{X\times Y} c(x,y)d(I\times T)_\sharp \mu = \int_X c(x,T(x))d\mu(x)$$

Question 6. In kantorovich, how do we know infimum is always min?

TAG: VariationalCalculus

Prop 4. *Norms are lsc*



Example 3. Picture of lsc function: 2021/pics/lsc.png

Prop 5. *lsc functions achieve mins on compact sets*

Proof. Let m inf. Let x_n s.t. $f(x_n) \to m$. Sequential compactness gives us subsequence so that limit x stays in X. Thus $f(x)$ is minimizer since it must be smaller than all other valuations. □

Theorem 3. *If X, Y compact and $c : X \times Y \to R$ then the kantorovich formulation has a solution*

Proof. **Proof using direct method of calculus of variations**

$T_c : P(X \times Y) \to R$ is cts. w.r.t narrow convergence so $\int c\gamma = T_c(\gamma)$. Since $P(X \times Y)$ is tight and thus compact w.r.t narrow conv. Then if $\gamma_n \in \Pi(\mu, \nu)$ is a minimizing seq. then we have conv. subseq in product $P(X \times Y)$ for some γ. Need $\gamma \in \Pi$. Doable by evauating marginals directly via defintions above □

Theme 4. We study measures by looking at test functions, which all for equality. Key idea is to look for notions which allow us to gai equality.

Prop 6. $\Pi(\mu, \nu)$ *is convex.*

4.6 DRL

TAG: DRL

Remark 42. Wolfer Ted Talk:

`https://www.ted.com/talks/daniel_wolpert_the_real_reason_for_brains/transcript?language=en#t-1`

Remark 43. Bayesian inference: data + prior knowledge informs action. Bayes rule: optimal rule for combining information.

Remark 44. We are sensory predictors detecting exterior sensory and subtracting off interior prediction.

Remark 45. Plan movements to minimize negative consequences of noise.

Remark 46. Q value is expected returns, not rewards. Must be learned from experience. They are predictive.

Remark 47. Intermediate reward shaping is hard because it can conflict and lead astray actions leading to final reward.

Question 7. Why learn a model via supervised learning instead of hardcoding it in?

Answer 5. For some reason learning representation from data is very hard, even in supervised context. Representation is very important. Often times just don't generalize. This representation is more important in cases where we don't know how to hard code rules. This is a representation learning problem for the model, and representation is hard.

Question 8. Do we use supervised learning to speed up the basic manipulation aquistition action?

Answer 6. In complex cases this is how the model must learn the world, because it cannot be ”hard coded” in.

Question 9. Need more example of using supervised learning to learn dynamics model. Clearly not necessary in some cases.

Remark 48. Model free is no model, when is there is one it can be learned via supervised learning.

Question 10. Why do we need supervised learning to learn dynamics in chess? Maybe it’s just to predict the opponents move.

Answer 7. Actually no it depends on if we include the opponent in the state or not but modula that its not really required if we have a ”logical description” of the board. Sometimes neural network paramterizations of dynamics are nice though because they are less computationally expensive(than a newtonian description for example) and are differentiable(which is often useful).

Example 4. An example of both is controlling nuclear fusion power plant: there are great physics simulators that are quite accurate at predicting the next state, but they take on the order of 8 hours to solve a single second worth of real world dynamic. This is because the simulator is solving a very complex system of PDEs. In contrast, if we distill the dynamics into a neural network, it is much faster to simulate, and opens the door for novel planning and control techniques (at the cost of model bias)

Theme 5. Theory and muscle learning are two ends of an extreme. Theory is exteremely generalizable but not particularly precise. Muscle learning is very precise but not particularly generalizable. This naturally occurs because the world is complex and more precision requires more information. How do we optimize between generalization and precision? This is the overfit problem

4.7 Chess

Remark 49. Insane game with no pawn captures:

`https://www.chess.com/analysis/game/live/6448761849?tab=analysis`

The whole game I slowly let myself get backed into a corner

Remark 50. Try to use pawns to restrict play more

Remark 51. Need to exploit weakness:

`https://www.chess.com/analysis/game/live/6449777852?tab=analysis`

When opponent exposes weakness(structural) need to identify and exploit.

Remark 52. What’s better than e6 in modern? It was good in this game:

`https://www.chess.com/analysis/game/live/6451734999?tab=analysis`

It allowed me to challenge center and break open for rook without weakening pawns too much

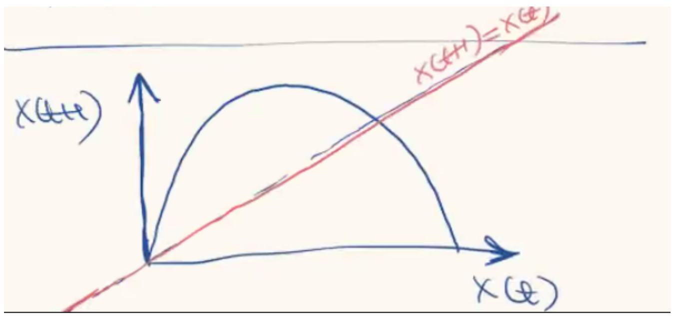
Remark 53. Complete dominance:

https://www.chess.com/analysis/game/live/6453705068?tab=analysis

4.8 Modeling Evolution

TAG: ModelingEvolution

Example 5. Reframing change in terms of dependent variable:

2021/pics/growth.png 

5 2/5

5.1 Goals

1. 1350 chess
2. Research 3 hours
3. Thesis 5 pages

5.2 Complex Analysis

TAG: ComplexAnalysis

Theorem 4. *If f power series than f' exists as $\sum na_n z^{n-1}$ with same R, since $n^{1/n} \to 1$.*

Proof. Fix $|z_0| < r < R$. Set $g(z) = \sum na_n z^{n-1}$. Write

$$\frac{f(z_0+h) - f(z_0)}{h} - g(z) = \frac{S_N(z_0+h) - S_N(z_0)}{h} - S'_N(z_0) + S'_N(z_0) - g(z_0) + \frac{E_N(z_0+h) - E_N(z_0)}{h}$$

The first term is small using triangle inequality and $(z_0+h)^n - z_0^h = h((z_0+h)^{n-1} + (z_0+h)^{n-2}z_0 + ...)$ which is bounded in norm by something still in R once h gets small enough. Thus this third term converges.

The second term is small since this is a convergent power series.

The first term is small by definition. And we are done.

Notice this proof simultaneously asserts existence and shows what it is. □

Proof. I had another proof by writing $f'(z_0 + h) = f(z_0) + ah + o(h)$ where $a = f'(z_0)$. □

Remark 54. Power series are infinitely differentiable, since we have same disk of convergence.

Question 11. Is this how we establish holomorphic functions are infinitely differentiable? Cause they're all power series? How is this done?

Answer 8. Write holomorphic function as its taylor expansion and show they are close?

Theme 6. Anytime we prove something using algebraic facts, we have a complex argument since we haven't used the changed, rigid geometry at all.

Question 12. What function is differentiable only once? Recall weierstrass cts everytwhere diff. nowhere

TAG: ComplexIntegration

Question 13. Can we allow for a countable number of cuts? Will cutting in diff. ways countably lead to diff. integrals? Does this mess up notion of equivalency? What breaks?

5.3 Set

Remark 55. For ultraset, look for clustering and try to use outliers to specify missing part of cluster

Remark 56. Don't infer cards already on desk

5.4 Chess

Remark 57. Smooth game with new sicilian: stops e4 push.

`https://www.chess.com/analysis/game/live/6460562753?tab=analysis`

Remark 58. Apparently a very accurate game against high level opponent. Feels like messed up in endgame with pawn sequences:

`https://www.chess.com/analysis/game/live/6460848543?tab=analysis`

A similar one(lower level):

`https://www.chess.com/analysis/game/live/6461016488?tab=analysis`

Remark 59. Should play good game against computer to improve "good" moves. Correct weak play. Can also learn openings this way

Remark 60. An intuitive attacking game played on throwaway:

`https://www.chess.com/analysis/game/live/6463844155?tab=analysis`

6 2/6

6.1 Goals

1. 1400 chess
2. Resarch progress: 4 hours
3. Thesis: 5 pages
4. Read some content
5. Clean up lists

7 2/7

1. Resarch progress: 4 hours
2. Thesis: 5 pages
3. Clean up lists

8 2/8

8.1 Goals

1. Resarch progress: 4 hours
2. Thesis: 3 pages
3. Clean up lists
4. Text Tessa

8.2 Complex Analysis

Theorem 5. *If f has primitive then $\int_\gamma f(z)dz = F(\omega_2) - F(\omega_1)$*

Proof. Ports from real results by construction. □

Example 6. $\int_\gamma dz/z = 2\pi i$ where γ is unit circle and hence $1/z$ must not have primitive(otherwise would be 0 on closed curve).

Theorem 6. *Gauss Lucas*

Proof. Exercise. Very algebraic. Probably write roots as convex combinations. Just compute by taking derivative

In fact bring out missing roots by taking quotient P'/P. Easier to work with these than the complement set of sums and products(often working with quotients easier to access new roots). □

Theorem 7. *When $P(z) = \sum a_k a^k$ with real coeffiecients and only real zeroes then coefficients binomial(ultra) log concave.*

Proof. One line proof via gauss-lucas. So derivatives have real roots. Reciprocal poly also has real roots. Then to get the inequality consider $z^{n-k+1}P^{(k-1)}(1/z)$ which has real roots and take $n-k-1$ derivatives resulting in two deg poly. The coefficients of this thing gives desired result by looking at the discriminant. □

Theme 7. This is common technique to show common sequences log-concave

Example 7.

- Stirling numbers of first kind
- Stirling numbers of second kind
- t_k ultra log-concave where t_k number of matchings of size k in arbitrary graph G

8.3 Technical Animation

Remark 61. 3 techniques for animation: motion capture, procedural, and keyframing.

Remark 62. CMU Panoptic Studio Dataset: Mocap data

Remark 63. Motion Capture Data Explained

8.4 Random Sign Sums and Hypercube Geometry

Question 14. Tomaszewski's Conjecture: $P(|X| > 1) \leq 1/2$ for $X = \sum a_i \epsilon_i$?

Remark 64. Random sign distributions "somewhat" log concave. Ie. log of distribution is log concave

Theme 8. Random sign sums are like gaussians.

Interesting 6. To show a probabilistic inequality find an injective map from A to B which is measure preserving.

To show $P(|X - a| \leq t) \leq P(|X| \leq t)$ we use reflection of a random walk.

8.5 DRL

TAG: DRL

8.6 Evolutionary Methods

Remark 65. Basic reinforcement naive search: randomly choose parameters and see which ones are optimal. Then sample around optimal ones to see if we can do better.

Question 15. How do we combat local optimum?

Answer 9. Restart by initially choosing start parameters uniformly.But certainly not fullproof. In general RL super prone to local optimum.

Question 16. Why is our density gaussian and not something else for mutation/parameter updates?

Answer 10. I have to assume it's just for ease of computation. Easy to learn and parameterize

Remark 66. Distributed computing for evolutionary workers has known random seeds

8.7 Modeling Evolution

TAG: ModelingEvolution

Remark 67. Why do we care about a basic model with wrong assumptions? Provides a basis for which we can compare parameters against(for example then varying popluation size and looking at divergence of effects from model).

Theme 9. Model cost(information cost to store) vs. model explanatory power. Some kind of ratio is the efficiency? Also how do we take into account the computational power necessary to answer questions.

Applying computational complexity to model theory? Often computational complexity is applied to a single question. But maybe it should be applied to a theory answering a collection of questions? Allows for ammortized analysis of collection of questions.

Remark 68. How to model? Hard to say except use common sense. very complex. Important to understand what features we want to capture and make sure our model reflects parameters that take this into account.

9 2/10

9.1 Goals

1. Finish complex homework
2. Read complex notes
3. Research : 3 hors
4. Thesis : 3 pages
5. Read Technical Animation Paper 1400 Chess

9.2 Blockers

1. Problem 3 complex

9.3 Complex Analysis

Remark 69. To show mulitiplicative/additive property of exponential show $e^z e^(c - z)$ constant.

Theorem 8. *(Goursout's Theorem): If $\Omega \subseteq \mathbb{C}$ open and $f : \Omega \to \mathbb{C}$ holomorphic with $\Delta \subseteq \Omega$ triangle then*

$$\int_{\partial\Delta} f = 0$$

Proof. Bisect Δ noting

$$\int_{\partial\Delta^{(0)}} f = \int_{\partial\Delta_1^{(1)}} f + ... + \int_{\partial\Delta_4^{(1)}} f$$

We continue recursively bisecting so that

$$|\int_{\partial\Delta^{(k)}} f| \leq 4|\int_{\Delta^{k+1}} f|$$

The diameters $d_n \to 0$ and perimeters go to 0 so $\bigcap_{n=1}^{\infty} \Delta^{(n)} = \{z_0\}$(nested compact sets with diameter going to 0).

We argue

$$4^n|\int_{\partial\Delta^{(n)}} f| \to 0$$

since f is holomorphic. Write $f(z) = f(z_0) + f'(z_0)(z - z_0) + o(z)(z - z_0)$. Clearly $f(z_0)$ and $f'(z_0)(z - z_0)$ have primitives. Thus we are only actually integrating $(z - z_0)o(z)$. Upper bounding $(z - z_0)$ by diameter and $o(z)$ by some $sup_{\Delta^{(n)}}|f(z)|$ we have the bound $p_n d_n$ sup. Recall $p_n = 1/2p_{n-1}$ and $d_n = 1/2d_{n-1}$ and so we're done. □

Prop 7. *Now since we have triangles we can also show for rectangles and by approximation most sets(tesselation).*

Theorem 9. *Let $D\mathbb{C}$ be a disc, $f : D \to \mathbb{C}$ holomorphic then f has a primitive.*

Proof. Suppose D centered at 0. Write $F(z) = \int_{\gamma_z} f(w)dw$ where γ_z is a triangular curve connecting z to 0.

Now compute with aim $F(z + h) - F(z) = hf(z) + ho(h)$

$$\int_{\gamma_{z+h}} f - \int_{\gamma_z} f = \int_{\gamma z \to z+h} f$$

by looking at paths and conidering goursout

Then we conclude by using continuity □

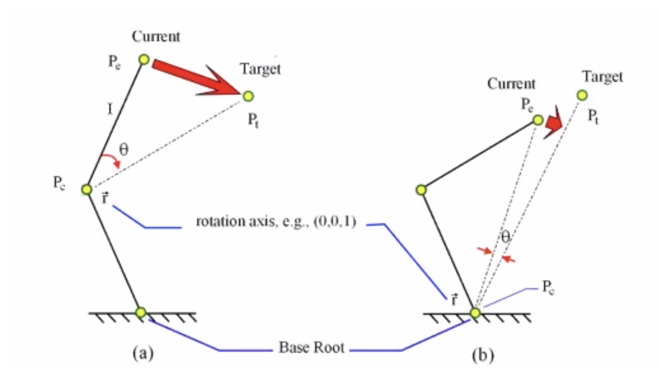
Remark 70. Notice this argument relies on convexity. The problem is with holes(like the punctured disk).

9.4 Technical Animation

TAG: TechnicalAnimation InverseKinematics

9.5 Inverse Kinematics

Remark 71. CCD Illustration: 2021/pics/ccd.png



Remark 72. Long chains of links tend to wrap up. Can also decide to iterate top to bottom affector or bottom to top. To address this can repeat recursion for every top level: whatever gets recursed more seems to bend more.

Remark 73. Basic model/fast but some limitations. Fabric better repacement

Remark 74. Alternative approach is jacobian based inverse kinematics: Introduction to Invers Kinematics with Jacobian Transpose, Pseudoinverse and Damped Leaster Squares

`http://math.ucsd.edu/~sbuss/ResearchWeb/ikmethods/iksurvey.pdf`

Remark 75. We write jacobian as:

$$\dot{s} = J(\theta)\dot{\theta}$$

for speeds on the affectors and their bending angles