1. # Tighter Bounds for Random Projections of Manifolds

Ken Clarkson
IBM Almaden

2. # Dimensionality Reduction

- Given a high-dimensional dataset, in $\mathbb{R}^m$, map it to a lower-dimensional space
- One approach: carefully pick which coordinates to keep
  - Some dimensions are *features*, others are not
- Or: carefully rotate the data, then carefully pick which coordinates to keep, or do something even more complicated
  - SVD (PCA, LSI, EigenFace*), SDP, ICA, MDS, ETC
  - *See also: EigenEyebrow, EigenEye, EigenNose, EigenMouth, EigenHead....
  - EigenHand, EigenBody, EigenHeart...
  - EigenSign, EigenImage, EigenFish, EigenForm, EigenTracking, EigenWindow, EigenGait, EigenLightField, EigenSurface, EigenFeature, Eigen Lightfield, Eigen-Scale-Space, Eigen Nodule, Eigen-Prosody, EigenShape, EigenTree, EigenEdge, EigenEdginess, EigenHills, Eigen (grapefruit) stems, EigenCharacter, EigenSignature, EigenWord, EigenSign, EigenLetter, EigenScrabble**
  - **Not: EigenCluster, EigenMonkey

3. # Random projection

- Instead of picking a rotation carefully,
  pick one at random
- Instead of picking from the new coordinates carefully,
  pick the first k

4. # Random projection, more specifically

- Again:
  - Apply a random rotation to $v \in \mathbb{R}^m$
  - Drop all but k coordinates
  - Scale (multiply by a constant) so that new vector v' has $E[\|v'\|] = \|v\|$
- Equivalently: pick a random subspace of dimension k, project v onto it, then scale
- Johnson-Lindenstrauss (JL) Lemma: with high probability, this preserves length, approximately:
  - Let a *k-map* P be a random projection from $\mathbb{R}^m$ to $\mathbb{R}^k$, as above
  - If $k \geq \varepsilon^{-2} C \log(1/\delta)$, then with probability at least $1 - \delta$,

$$(1-\varepsilon)\|v\| \leq \|Pv\| \leq (1+\varepsilon)\|v\|$$

- Since P is linear, $\|\alpha P v\| = \alpha \|Pv\|$ for $\alpha \geq 0$, so WLOG $\|v\| = 1$

5. (Random projection : why?)

- Existence proof: if a random projection gives good results, what if we work harder?
- There are many similar algorithms with the same properties
  - Multiply by a $k \times m$ matrix of random ± 1, or of Gaussians
  - Use a matrix with a fast multiply [AC]
- Obliviousness: the random projection is chosen without looking at the data at all
  - ...and so is called "universal feature reduction"
  - Feature reduction without "feedback": no loops
  - Brain may work this way; a recent model of the brain [SOP]:
    - Is a "feedforward" neural network
    - Uses randomness for feature reduction in a similar way

6. From one point to many

- Point *isometrizing*: for one vector (point) v, the probability of failure is

  $$\delta \leq \exp(-k\varepsilon^2 / C)$$

- Finite set *isometrizing*: for set S of n points, probability of failure for all points is

  $$\delta \leq n\exp(-k\varepsilon^2 / C)$$

- Finite set *embedding*: for $S - S := \{x - y \mid x, y \in S\}$,

  $$\delta \leq n^2\exp(-k\varepsilon^2 / C)$$

  - $k = O(\varepsilon^{-2}\log(n / \delta))$
  - That is, preserving distances

7. From many to infinite

- Subspace JL [M][Sar]: for d-dimensional linear subspace F,

  $$\delta = O(1)^d\exp(-k\varepsilon^2 / C)$$

- Hint:
  - There is a finite subset of F so that isometrizing it $\Rightarrow$ isometrizing F
  - It helps that if $x, y \in F$, so is $x - y$, and so is αx
- "Doubling" JL [AHY][IN]: Embedding bounds for sets in $\mathbb{R}^m$ of bounded doubling dimension
  - Mostly, additive approximation bounds on distance approximation, not relative
  - Doubling dimension [L67][A83] is a kind of "intrinsic dimensionality"; applied e.g. to NN searching [C99][KL04]
- Manifold JL [BW], here: embedding a *(smooth, connected)* d-dimensional manifold,

  $$\delta = O(1 / \varepsilon^d)\exp(-k\varepsilon^2 / C)$$

8. (When is the input to a program infinite?)

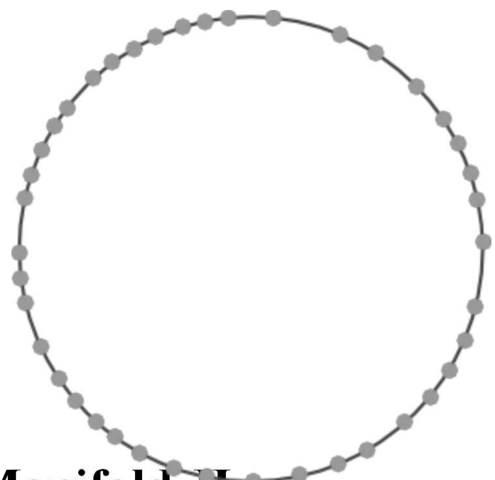
- It isn't
- Peta-Shmeta: "uncountably infinite" will always be "massive"
- And: the bounds hold for any finite subset of the infinite set
- So: for a set of n points on a manifold, better bound when n is very large

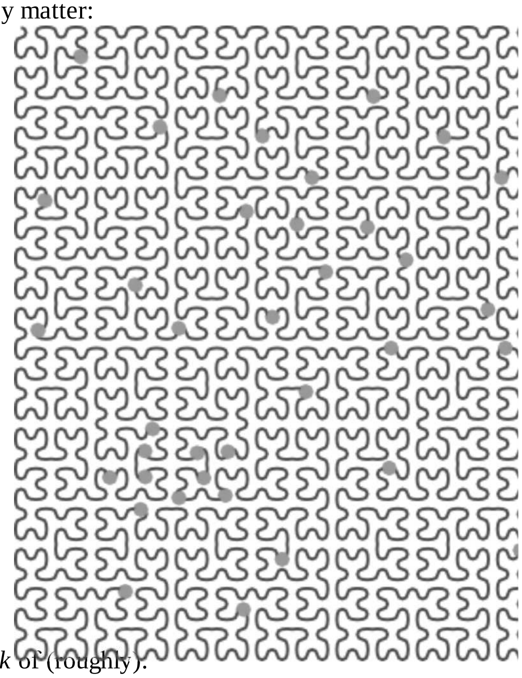
9. NB: Embedding and embedding

- Embedding a manifold here means preserving *Euclidean* distances
- This implies also preserving geodesic distances and other local properties
- If only geodesic distances are of interest, results here simplify a bit

10. "All d-manifolds are not the same"

- The leading term for k, here and [BW], is $k = O(\varepsilon^{-2}(d \log(1/\varepsilon) + \log(1/\delta)))$
- Improvement here is for lower-order terms, but they matter:





11. Manifold JL

- Baraniuk and Wakin result has additional term for k of (roughly):
  $O(\varepsilon^{-2}(d \log(m \mu_I(M)/\rho)))$
  is enough for failure probability δ, where:
  - m is (as before) the ambient dimension
  - $\mu_I(M)$ is the surface area of M
  - ρ is the *reach* [F59], the minimum distance of any point of M to its medial axis, and $1/\rho$ is an upper bound for curvature at any point of M
- My result has additional term (roughly):
  $O(\varepsilon^{-2}(\log(\mu_I(M)/\tau^d + \mu_{III}(M))))$
  where:
  - $\mu_{III}(M)$ is the total absolute curvature of M
  - $\tau(M)$ is a low-torsion-path threshold: if $a, b \in M$ have $\|a - b\| \leq \tau$ then there is a low-curvature *or* low-torsion path between them
  - If a path has zero torsion, it is planar; if very low total torsion, $\approx$ planar

12. Why is this an improvement or interesting?

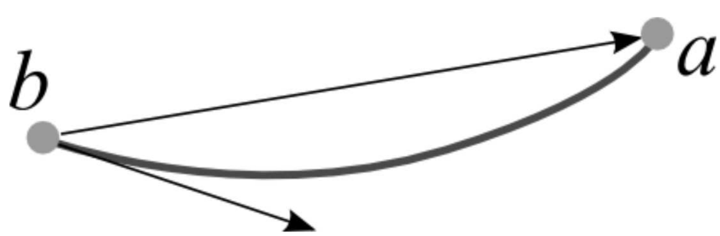
- Removed dependence on ambient dimension m entirely
  - Sometimes $m = \infty$
- $1/\tau$ plays a role similar to $1/\rho$, but can be much more smaller
  - If M is a pure quadric, then $1/\tau$ is zero
- Also showed: can use curvature measure $\mu_{II}(M)$ instead of surface area $\mu_I(M)$
  - $\mu_{II}(M)$ can be $\ll \mu_I(M)$
- Places "JL complexity" among other properties of M bounded by integral measures $\mu_X(M)$

13. The General Approach : Long Chords

- As in prior work [IN][AHY], approximate the infinite set of all $(a-b)/\|a-b\|$, for $a, b \in M$ by a sequence of finite sets, and then apply JL Lemma to all the finite sets
- "Long chords", from a, b that are far apart, are easy to handle, because a' close to a and b' close to $b \Rightarrow$ normalized differences are close

14. The General Approach : Short Chords

- For short chords, the smoothness of the manifold is helpful: if $a, b \in M$ are very close together, then $a - b \approx$ a tangent vector of M

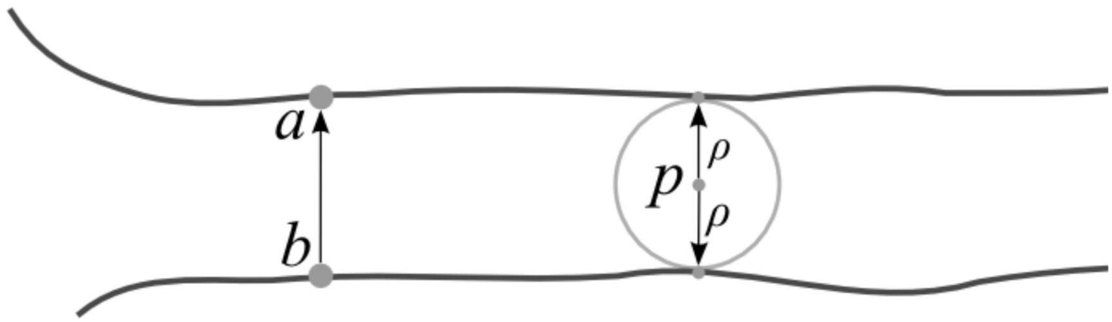


- If the max curvature is small, chords need not be very short for this to be good

- Approximation for short chords becomes approximation of tangent vectors, which have total complexity $\mu_{III}(M)$

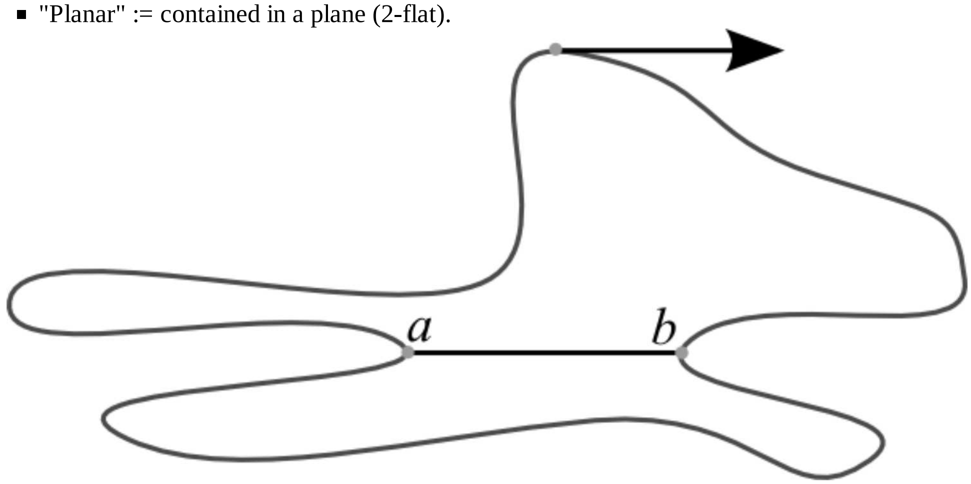
15. Short chords, tangents, reach

- Suppose $a, b \in M$ very close in Euclidean distance, but very far in arc length
- Then tangent at a or b has nothing to do with $a - b$
- This can happen when the reach ρ of M is small
  - As mentioned, the reach is the minimum distance of a point of M to the medial axis of M
  - Smallest distance of point $p \in \mathbb{R}^m$ to M, when p has two nearest neighbors in M
  - A.K.A., reciprocal *condition number* of M
- Reach is a key property, but very "local" and "worst case"



16. Short Chords via Planar Tangents

- How to avoid max curvature / reach?
- When $a, b \in M$ are connected by a *planar* curve in M, that curve has a tangent vector parallel to $a - b$
  - "Planar" := contained in a plane (2-flat).



- M is a pure quadric $\Rightarrow$ $a, b \in M$ connected by a planar curve
- Low-torsion == approximately planar

17. Concluding Remarks

- Results here give a relation of projection dimension k to standard measures
  - May not be "news you can use": projection dimension guarantee relies on quantities that may not be available
  - Like many results, gives an unverifiable sufficient condition
  - Test for the right k statistically?
- OK for Manifold + (Gaussian) noise
- Relation to linear compression [Thurs, 4:30]
  - Both: multiply by $k \times m$ matrix, $k \ll m$
  - There: x is sparse $\Rightarrow x$ is recovered approximately
  - Here: x's in a manifold, preserve (only) distances
  - (Could apply [S] to all d-flats of d-sparse vectors)
- Probably extendible to polyhedral manfolds

Thank you for your attention