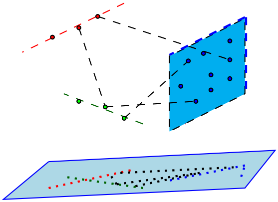
Too many secants: a hierarchical approach to secant-based dimensionality reduction on large data sets.

Henry Kvinge* Elin Farnell Michael Kirby Chris Peterson

Colorado State University, Fort Collins, CO
*On job market this fall.



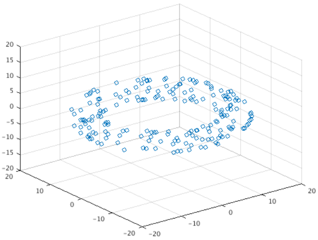
Dimensionality reduction

Dimensionality reduction is a key tool for extracting information from high dimensional data sets.

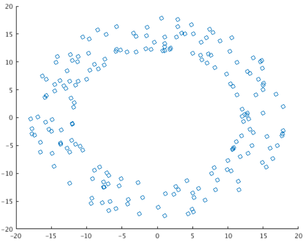
When our data points correspond to elements of $\mathbb{R}^n$ we can think of dimensionality reduction as the process of mapping:

$$\text{points in } \mathbb{R}^n \quad \mapsto \quad \text{points in } \mathbb{R}^m$$

for $m < n$.



$\longrightarrow$



Dimensionality reduction

Dimensionality reduction allows us to:

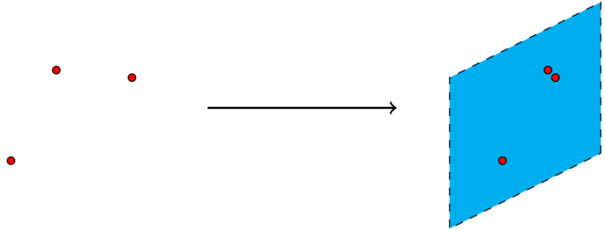
- Create a new "approximately equivalent" data set of significantly smaller dimension for storage and analysis.
- Extract features for machine learning applications.
- Better understand the geometry of the data set in question and answer questions such as

"how is my data set changing over time?"

Dimensionality reduction

Generally, a good dimensionality reduction algorithm should preserve structure in the data set.

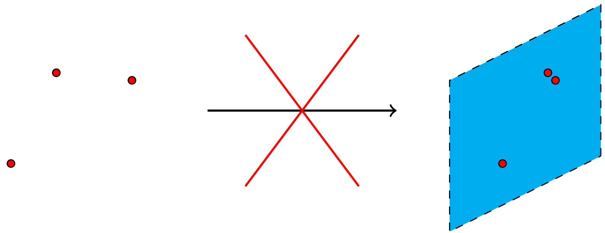
Example: In many cases we do not want to collapse points in $\mathbb{R}^n$ onto each other in $\mathbb{R}^m$.



Dimensionality reduction

Generally, a good dimensionality reduction algorithm should preserve structure in the data set.

Example: In many cases we do not want to collapse points in $\mathbb{R}^n$ onto each other in $\mathbb{R}^m$.



Secant-based dimensionality reduction

What is another way of saying that we want to preserve the distances between points?

For a data set $D \in \mathbb{R}^n$ the normalized *secant set* S is

$$S := \left\{ \frac{x-y}{||x-y||} \mid x, y \in D, \text{ with } x \neq y \right\}.$$

Secant-based dimensionality reduction algorithms work under the principle that we should look for dimension reducing transformations which **preserve the secant set of our data set**.

SAP algorithm outline

We developed the *secant-avoidance projection (SAP) algorithm* to solve the optimization problem

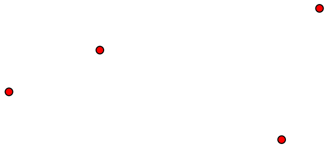
$$\underset{P \in \text{Proj}(\mathbb{R}^n, \mathbb{R}^m)}{\arg\max} \left(\min_{s \in S} ||P^T s|| \right)$$

where $\text{Proj}(\mathbb{R}^n, \mathbb{R}^m)$ consists of all $n \times m$ matrices whose columns are orthonormal vectors in $\mathbb{R}^n$.

SAP searches for the projection such that the most shrunken secant is maximized.

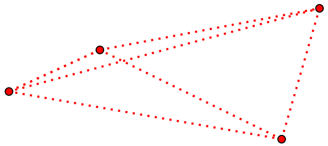
SAP algorithm outline

Given a data set in $\mathbb{R}^n$ we calculate the secant set.



SAP algorithm outline

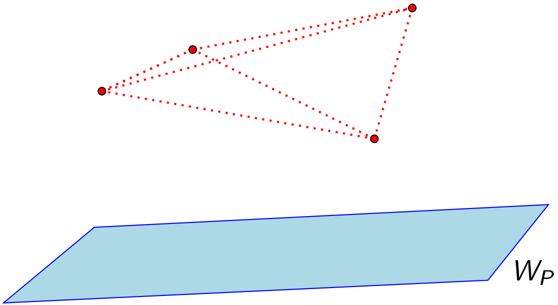
Given a data set in $\mathbb{R}^n$ we calculate the secant set.



Note: To ensure that each secant is given equal weight, we normalize the secants to have unit length.

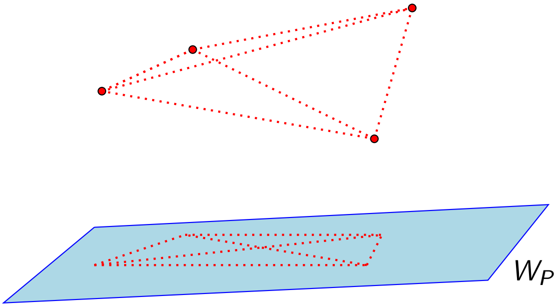
SAP algorithm outline

At each iteration of the algorithm we project these secants onto our current projection subspace W_P (corresponding to projection P).



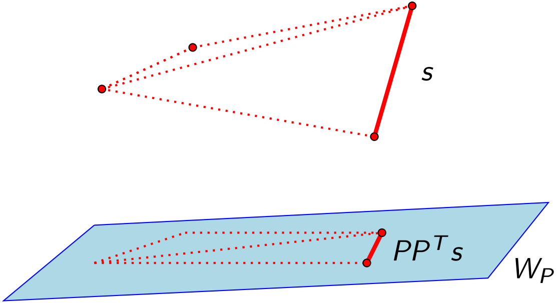
SAP algorithm outline

At each iteration of the algorithm we project these secants onto our current projection subspace W_P (corresponding to projection P).



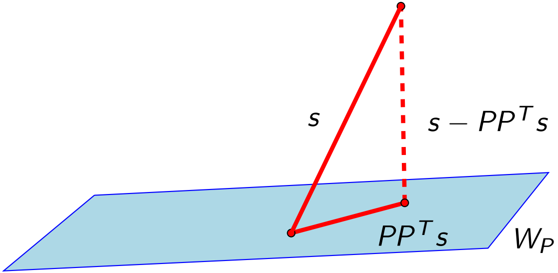
SAP algorithm outline

We choose the secant s that is most diminished by projection onto W_P, i.e. we choose s that gives the smallest value $||PP^T s||_{\ell_2}$.



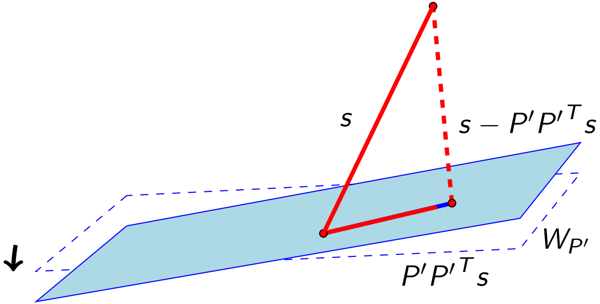
SAP algorithm outline

We rotate W_P (respectively P) to $W_{P'}$ (resp. P') to better capture s. In particular we rotate W_P toward the orthogonal complement of the projection of s onto W_P, $s - PP^Ts$.



SAP algorithm outline

We rotate W_P (respectively P) to $W_{P'}$ (resp. P') to better capture s. In particular we rotate W_P toward the orthogonal complement of the projection of s onto W_P, $s - PP^T s$.



SAP algorithm outline

Input: Data set, projection dimension m, iterations, step size α

Calculate and normalize secant set S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
    Find secant $s_i \in S$ with smallest projection under $P^{(i-1)}$
    $P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \dots, p_m^{(i-1)}]$
    $P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

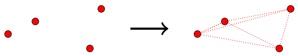
$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

    $i \leftarrow i + 1$
end for

* Orthonormalization is currently implemented using a QR decomposition

SAP algorithm outline

Input: Data set, projection dimension m,
iterations, step size α



Calculate and normalize secant set S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
  Find secant $s_i \in S$ with smallest projection under $P^{(i-1)}$
  $P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \dots, p_m^{(i-1)}]$
  $P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

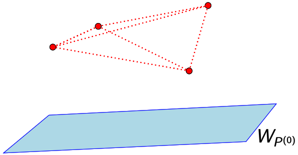
$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

  $i \leftarrow i+1$
end for

* Orthonormalization is currently implemented using a QR decomposition

SAP algorithm outline

Input: Data set, projection dimension m, iterations, step size α



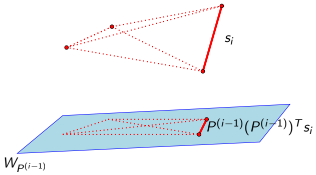
Calculate and normalize secant set S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
  Find secant $s_i \in S$ with smallest projection under $P^{(i-1)}$
  $P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \ldots, p_m^{(i-1)}]$
  $P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

  $i \leftarrow i+1$
end for

* Orthonormalization is currently implemented using a QR decomposition

SAP algorithm outline



Input: Data set, projection dimension m, iterations, step size α

Calculate and normalize secant set S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
  Find secant $s_i \in S$ with smallest projection under $P^{(i-1)}$
  $P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \dots, p_m^{(i-1)}]$
  $P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

  $i \leftarrow i+1$
end for

* Orthonormalization is currently implemented using a QR decomposition

SAP algorithm outline

Input: Data set, projection dimension m,
iterations, step size α

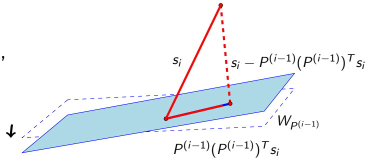
Calculate and normalize secant set S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
Find secant $s_i \in S$ with smallest projection under $P^{(i-1)}$
$P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \dots, p_m^{(i-1)}]$
$P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

$i \leftarrow i+1$
end for

* Orthonormalization is currently implemented using a QR decomposition

SAP algorithm outline



Input: Data set, projection dimension m, iterations, step size α

Calculate and normalize secant set S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
  Find secant $s_i \in S$ with smallest projection under $P^{(i-1)}$
  $P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \ldots, p_m^{(i-1)}]$
  $P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

  $i \leftarrow i + 1$
end for

* Orthonormalization is currently implemented using a QR decomposition

A comparison to other dim. reduction algorithms

SAP solves a different optimization problem in terms of secants than PCA

- PCA solves

$$\underset{P \in \text{Proj}(\mathbb{R}^n, \mathbb{R}^m)}{\arg\max} \sum_{s \in S} ||P^T s||$$

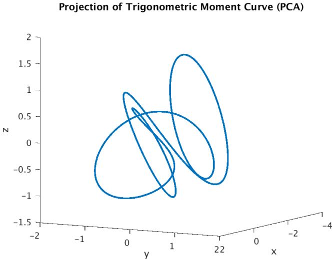
- SAP solves

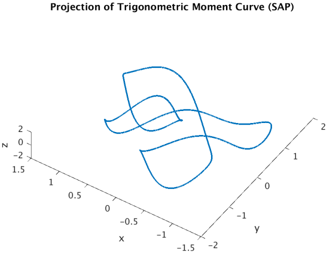
$$\underset{P \in \text{Proj}(\mathbb{R}^n, \mathbb{R}^m)}{\arg\max} \left(\min_{s \in S} ||P^T s|| \right)$$

A comparison to other dim. reduction algorithms

As a consequence:

- PCA "tries" to minimize the extent to which all secants are shrunk in projection,
- while SAP "focuses" on making sure no particular secant is shrunk too much.





Trigonometric moment curve $\phi : \mathbb{R} \to \mathbb{R}^{10}$,
$\phi(t) := (\cos(t), \sin(t), \cos(2t), \ldots, \cos(5t), \sin(5t))$.

SAP algorithm outline

A problem:
The algorithm works well when $|D| < 5,000$ but when $|D| = p$ then

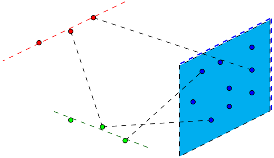
$$\# \text{ of secants of } D = \frac{p(p-1)}{2}.$$

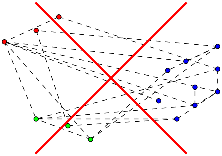
There are just **too many secants** (and probably a lot of redundant information).

SAP algorithm outline

Solution: Use hierarchical structure in the data and approximation to reduce number of secants to consider.

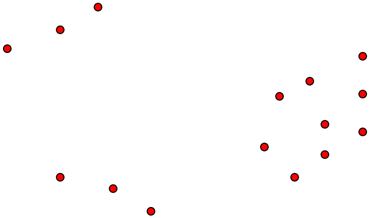
We developed a new algorithm, the **hierarchical secant avoidance projection (HSAP) algorithm** to do this.





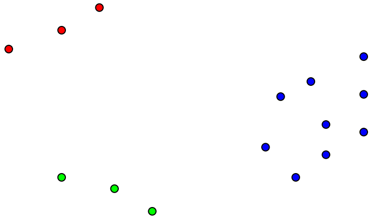
HSAP algorithm outline

Given a data set in $\mathbb{R}^n$ we start by clustering it.



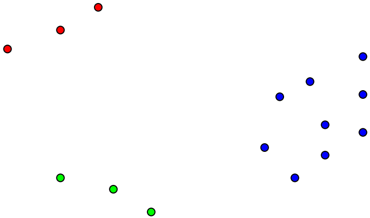
HSAP algorithm outline

Given a data set in $\mathbb{R}^n$ we start by clustering it.



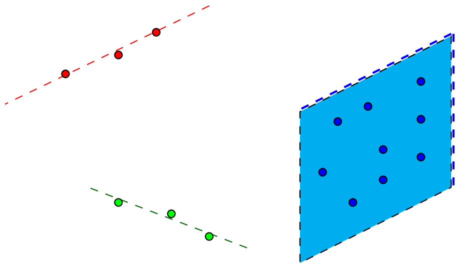
HSAP algorithm outline

We construct linear approximations of each cluster and select a few (normalized) secants between points in different clusters.



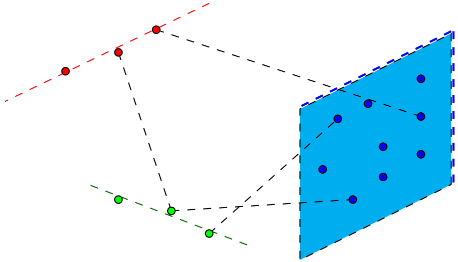
HSAP algorithm outline

We construct linear approximations of each cluster and select a few (normalized) secants between points in different clusters.



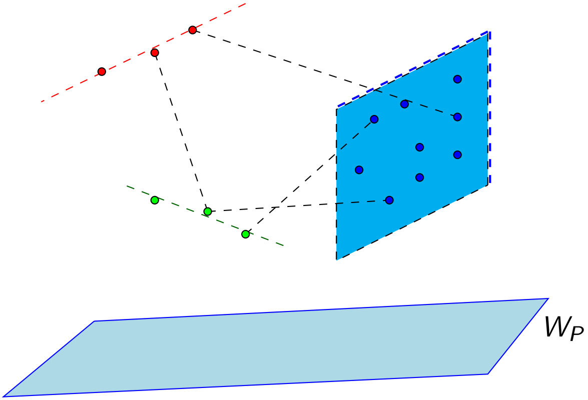
HSAP algorithm outline

We construct linear approximations of each cluster and select a few (normalized) secants between points in different clusters.



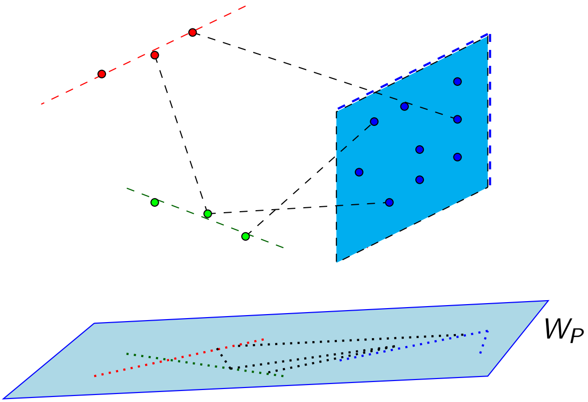
HSAP algorithm outline

At each iteration of the algorithm we project the selected secants and vectors representing the linear approximations onto our current projection subspace W_P (corresponding to projection P).



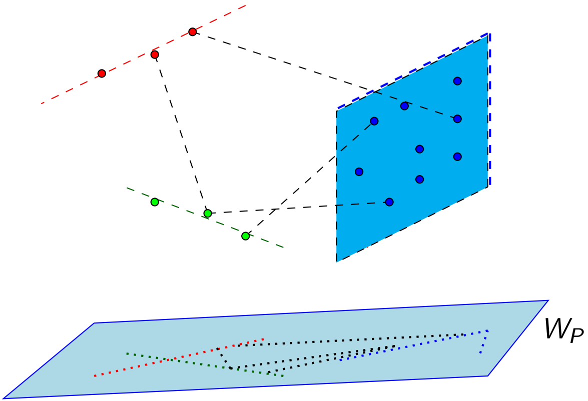
HSAP algorithm outline

At each iteration of the algorithm we project the selected secants and vectors representing the linear approximations onto our current projection subspace W_P (corresponding to projection P).



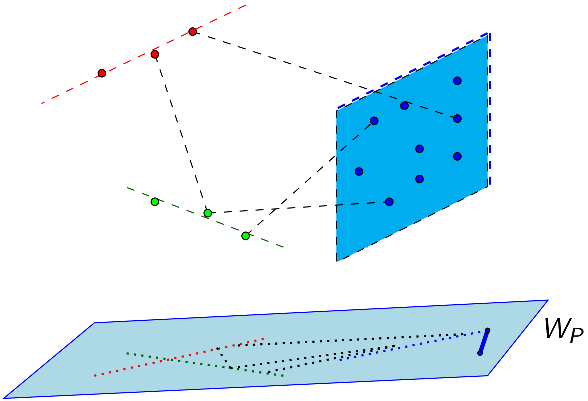
HSAP algorithm outline

Analogously to before, we choose the secant or linear approximation vector s that is most diminished by projection onto W_P.



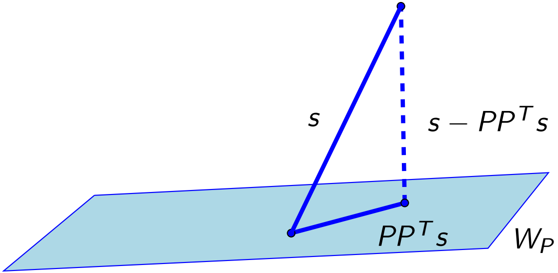
HSAP algorithm outline

Analogously to before, we choose the secant or linear approximation vector s that is most diminished by projection onto W_P.



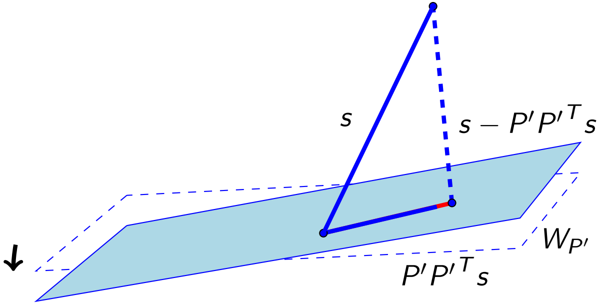
HSAP algorithm outline

Then we shift W_P (respectively P) as in the previous algorithm.



HSAP algorithm outline

Then we shift W_P (respectively P) as in the previous algorithm.



HSAP algorithm outline

Input: Data set, projection dimension m,
$\#$ of clusters k, iterations, step size α

Cluster the dataset in k clusters
For all i, calculate a linear approximation V_i for the ith cluster
Calculate a subset of normalized secants S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
Calculate the set L_1 of lengths for each projected secant with respect to $P^{(i-1)}$
Calculate the set L_2 of singular values of $(P^{(i-1)})^T V_j$ for each $1 \leq j \leq k$
Choose the smallest secant/singular vector s_i corresponding to $\min L_1 \cup L_2$
$P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \ldots, p_m^{(i-1)}]$
$P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

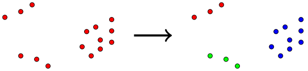
$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

$i \leftarrow i+1$
end for

* Orthonormalization is currently implemented using a QR decomposition

HSAP algorithm outline

Input: Data set, projection dimension m,
$\#$ of clusters k, iterations, step size α



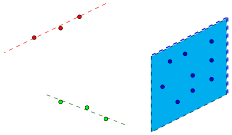
Cluster the dataset in k clusters
For all i, calculate a linear approximation V_i for the ith cluster
Calculate a subset of normalized secants S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
Calculate the set L_1 of lengths for each projected secant with respect to $P^{(i-1)}$
Calculate the set L_2 of singular values of $(P^{(i-1)})^T V_j$ for each $1 \leq j \leq k$
Choose the smallest secant/singular vector s_i corresponding to $\min L_1 \cup L_2$
$P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \ldots, p_m^{(i-1)}]$
$P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

$i \leftarrow i+1$
end for

* Orthonormalization is currently implemented using a QR decomposition

HSAP algorithm outline



Input: Data set, projection dimension m,
$\#$ of clusters k, iterations, step size α

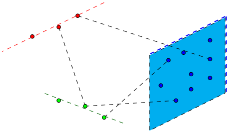
Cluster the dataset in k clusters
For all i, calculate a linear approximation V_i for the ith cluster
Calculate a subset of normalized secants S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
Calculate the set L_1 of lengths for each projected secant with respect to $P^{(i-1)}$
Calculate the set L_2 of singular values of $(P^{(i-1)})^T V_j$ for each $1 \leq j \leq k$
Choose the smallest secant/singular vector s_i corresponding to $\min L_1 \cup L_2$
$P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \dots, p_m^{(i-1)}]$
$P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

$i \leftarrow i+1$
end for

* Orthonormalization is currently implemented using a QR decomposition

HSAP algorithm outline



Input: Data set, projection dimension m,
$\#$ of clusters k, iterations, step size α

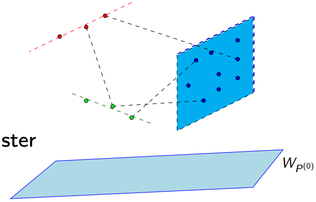
Cluster the dataset in k clusters
For all i, calculate a linear approximation V_i for the ith cluster
Calculate a subset of normalized secants S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
  Calculate the set L_1 of lengths for each projected secant with respect to $P^{(i-1)}$
  Calculate the set L_2 of singular values of $(P^{(i-1)})^T V_j$ for each $1 \leq j \leq k$
  Choose the smallest secant/singular vector s_i corresponding to $\min L_1 \cup L_2$
  $P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \dots, p_m^{(i-1)}]$
  $P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

  $i \leftarrow i + 1$
end for

* Orthonormalization is currently implemented using a QR decomposition

HSAP algorithm outline



Input: Data set, projection dimension m,
# of clusters k, iterations, step size α

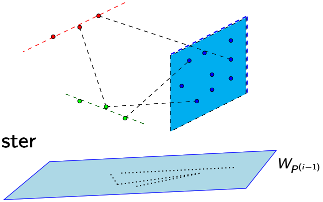
Cluster the dataset in k clusters
For all i, calculate a linear approximation V_i for the ith cluster
Calculate a subset of normalized secants S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
Calculate the set L_1 of lengths for each projected secant with respect to $P^{(i-1)}$
Calculate the set L_2 of singular values of $(P^{(i-1)})^T V_j$ for each $1 \leq j \leq k$
Choose the smallest secant/singular vector s_i corresponding to $\min L_1 \cup L_2$
$P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \ldots, p_m^{(i-1)}]$
$P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

$i \leftarrow i+1$
end for

* Orthonormalization is currently implemented using a QR decomposition

HSAP algorithm outline



Input: Data set, projection dimension m,
# of clusters k, iterations, step size α

Cluster the dataset in k clusters
For all i, calculate a linear approximation V_i for the ith cluster
Calculate a subset of normalized secants S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
  Calculate the set L_1 of lengths for each projected secant with respect to $P^{(i-1)}$
  Calculate the set L_2 of singular values of $(P^{(i-1)})^T V_j$ for each $1 \leq j \leq k$
  Choose the smallest secant/singular vector s_i corresponding to $\min L_1 \cup L_2$
  $P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \ldots, p_m^{(i-1)}]$
  $P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

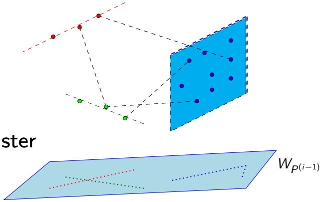
$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

  $i \leftarrow i + 1$
end for

* Orthonormalization is currently implemented using a QR decomposition

HSAP algorithm outline

Input: Data set, projection dimension m,
$\#$ of clusters k, iterations, step size α



Cluster the dataset in k clusters
For all i, calculate a linear approximation V_i for the ith cluster
Calculate a subset of normalized secants S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
Calculate the set L_1 of lengths for each projected secant with respect to $P^{(i-1)}$
Calculate the set L_2 of singular values of $(P^{(i-1)})^T V_j$ for each $1 \leq j \leq k$
Choose the smallest secant/singular vector s_i corresponding to min $L_1 \cup L_2$
$P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \dots, p_m^{(i-1)}]$
$P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

$i \leftarrow i + 1$
end for

* Orthonormalization is currently implemented using a QR decomposition

HSAP algorithm outline

Input: Data set, projection dimension m,
$\#$ of clusters k, iterations, step size α

Cluster the dataset in k clusters
For all i, calculate a linear approximation V_i for the ith cluster
Calculate a subset of normalized secants S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
  Calculate the set L_1 of lengths for each projected secant with respect to $P^{(i-1)}$
  Calculate the set L_2 of singular values of $(P^{(i-1)})^T V_j$ for each $1 \leq j \leq k$
  Choose the smallest secant/singular vector s_i corresponding to $\min L_1 \cup L_2$
  $P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \ldots, p_m^{(i-1)}]$
  $P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

  $i \leftarrow i + 1$
end for

* Orthonormalization is currently implemented using a QR decomposition

HSAP algorithm outline

Input: Data set, projection dimension m,
$\#$ of clusters k, iterations, step size α

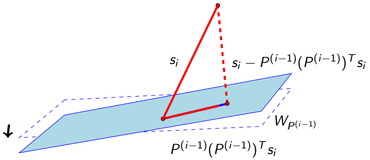
Cluster the dataset in k clusters
For all i, calculate a linear approximation V_i for the ith cluster
Calculate a subset of normalized secants S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
Calculate the set L_1 of lengths for each projected secant with respect to $P^{(i-1)}$
Calculate the set L_2 of singular values of $(P^{(i-1)})^T V_j$ for each $1 \leq j \leq k$
Choose the smallest secant/singular vector s_i corresponding to $\min L_1 \cup L_2$
$P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \ldots, p_m^{(i-1)}]$
$P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

$i \leftarrow i+1$
end for

* Orthonormalization is currently implemented using a QR decomposition

HSAP algorithm outline



Input: Data set, projection dimension m,
# of clusters k, iterations, step size α

Cluster the dataset in k clusters
For all i, calculate a linear approximation V_i for the ith cluster
Calculate a subset of normalized secants S
Choose an initial projection $P^{(0)}$
$i \leftarrow 1$
for $i \leq$ iterations **do**
    Calculate the set L_1 of lengths for each projected secant with respect to $P^{(i-1)}$
    Calculate the set L_2 of singular values of $(P^{(i-1)})^T V_j$ for each $1 \leq j \leq k$
    Choose the smallest secant/singular vector s_i corresponding to $\min L_1 \cup L_2$
    $P^{(i-1)} \leftarrow$ orthonormalization* of $[P^{(i-1)}(P^{(i-1)})^T s_i, p_2^{(i-1)}, \dots, p_m^{(i-1)}]$
    $P^{(i)} \leftarrow P^{(i-1)}$ but with first column the normalized vector

$$(1-\alpha)P^{(i-1)}(P^{(i-1)})^T s_i + \alpha(s_i - P^{(i-1)}(P^{(i-1)})^T s_i)$$

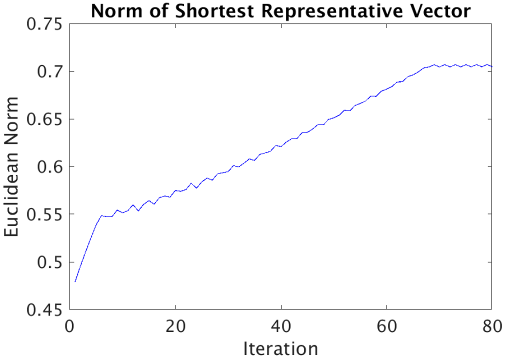
    $i \leftarrow i + 1$
end for

* Orthonormalization is currently implemented using a QR decomposition

Convergence of HSAP

Conditions and proofs for convergence of HSAP are work in progress.

Empirically, convergence is fairly consistent when the projected dimension is not too small and step-size not too big.



The κ-profile

Besides producing a projection $P^T : \mathbb{R}^n \to \mathbb{R}^m$, we also get a measure the secant **least well preserved by the projection produced by SAP or HSAP**:

$$\kappa_m := \min_{s \in S} ||P^T s||.$$

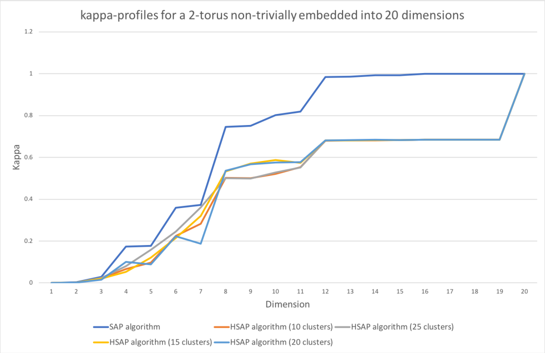
We define the κ-**profile** to be

$$(\kappa_1, \kappa_2, \ldots, \kappa_n).$$

The κ-profile gives a measure of how well the data can be projected into dimensions $1, 2, \ldots, n$.

The κ-profile

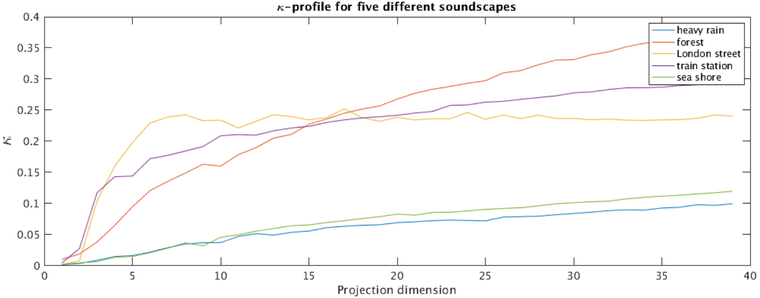
The projections obtained from HSAP generally preserve **all** secants in S (even those that it didn't see) only slightly less well than SAP.



However, depending on the parameters chosen, HSAP stores far fewer secants.

The κ-profile

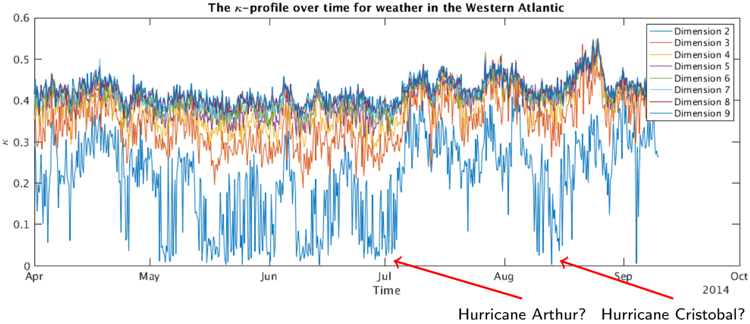
The κ-profile for soundscapes.



The κ-profile suggests that soundscapes with more random, incoherent noise are unsurprisingly higher dimensional.

The κ-profile

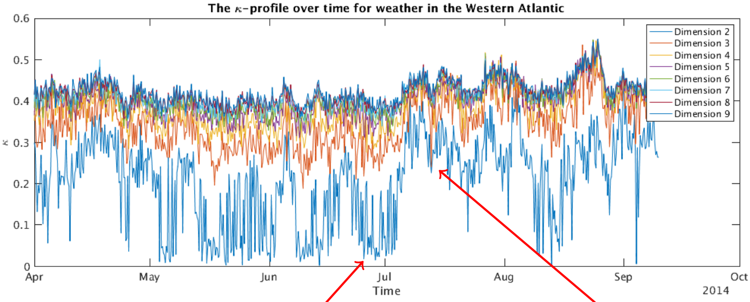
The κ-profile for dimensions $2, 3, \ldots, 9$ as a function of time, for a large 2015 weather data set taken from a grid in the Western Atlantic.

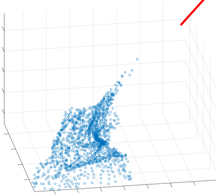


Storm activity seems to roughly correlate with jumps in κ_2.

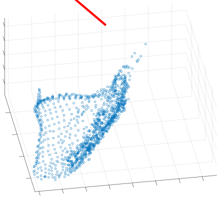
Stormy weather corresponds to the data becoming more 3-dimensional?

The κ-profile





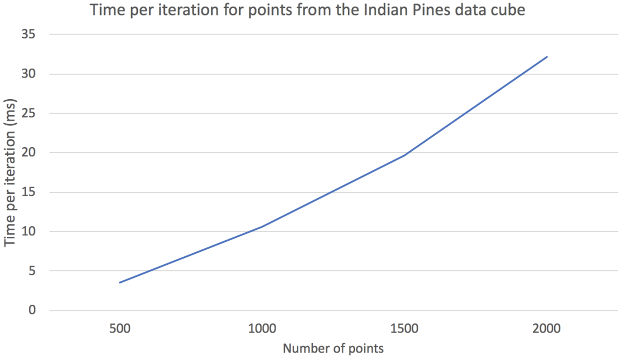
projection into $\mathbb{R}^3$



projection into $\mathbb{R}^3$

Thank you!

Time per iteration



- Code written in CUDA, using cublas library.
- Run on two Nvidia K80s.

HSAP complexity

HSAP is an

$$O\Big(\max\{n^3N, n^2N^2(\max|A_i|^2)\}\Big)$$

algorithm, where

- $n =$ original dimension of data,
- $N =$ number of clusters,
- $\max|A_i| \sim$ the largest number of sample secants between two clusters.

We generally expect $N, \max|A_i| \ll n$.