Sparse PCA Asymptotics & Analysis of Tree Data

Dan Shen
Department of Statistics & Operations Research
University of North Carolina at Chapel Hill

dshen@live.unc.edu

October 11, 2012

Outline

- Motivation & Background
- PCA Asymptotics
- Spike Covariance Models
- Theoretical Results of PCA
- Sparse PCA
- Summary
- Analysis of Tree Data

Outline

- **Motivation & Background**
- PCA Asymptotics
- Spike Covariance Models
- Theoretical Results of PCA
- Sparse PCA
- Summary
- Analysis of Tree Data

Modern Dataset Features

- High Dimensionality
  - Microarray, image, …
  - Dimension reduction techniques
    - Principle component analysis (PCA) — Pearson (1901)
    - Partial least squares — Wold (1985)
    - Canonical correlation analysis — Hotelling (1936)
    - …

- Sparsity
  - Signal sparse … most signal dimensions insignificant
  - Sparsity constraints
    - Sparse PCA
    - …

Modern Dataset Features

- High Dimensionality
  - Microarray, image, …
  - Dimension reduction techniques
    - **Principle component analysis (PCA) — Pearson (1901)**
    - Partial least squares — Wold (1985)
    - Canonical correlation analysis — Hotelling (1936)
    - …
- Sparsity
  - Signal sparse … most signal dimensions insignificant
  - Sparsity constraints
    - Sparse PCA
    - …

UNC, Stat & OR

Data Matrix

one data object

$$X = \begin{pmatrix} X_{1,1} & \cdots & X_{1,j} & \cdots & X_{1,n} \\ X_{2,1} & \cdots & X_{2,j} & \cdots & X_{2,n} \\ & & \vdots & & \\ X_{d,1} & \cdots & X_{d,j} & \cdots & X_{d,n} \end{pmatrix}$$

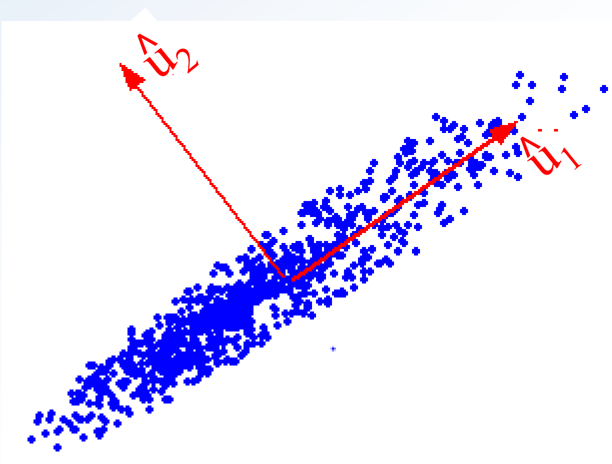
dimension of data objects is d

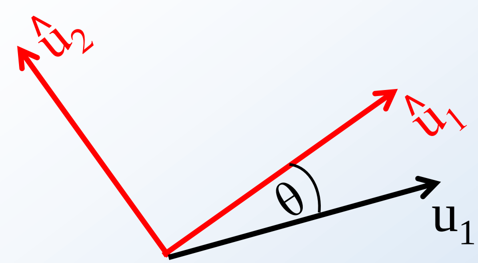
data matrix is a sample of n objects

Principle Component Analysis

Principle Component Analysis (PCA):

- Purpose: dimension reduction & visualization
- Goal: few linear combinations of the raw variables to explain majority of the data variation
- Calculation: eigen-decomposition of sample covariance matrix





As $n \to \infty$, $d \to \infty$, or $d \& n \to \infty$

- Consistency: $\theta \to 0$
- Strong Inconsistency: $\theta \to \pi/2$

Outline

- Motivation & Background
- **PCA Asymptotics**
- Spike Covariance Models
- Theoretical Results of PCA
- Sparse PCA
- Summary
- Analysis of Tree Data

PCA Asymptotics

- PCA – very popular tool
  - Offers useful insights
  - Reveals simple low-dimensional structure in high-dimensional data

- Important to understand asymptotic properties of PCA
  - Consistency
  - Strong inconsistency
  - Subspace consistency
  - Studied through mathematical statistics

Asymptotic Settings

Sample size n, dimension (# of variables) d

- Classical asymptotics:

  d fixed and $n \to \infty$

- Random matrix asymptotics:

  $d/n \to c$, as $n \to \infty$

- High Dimension, Low Sample Size (HDLSS) asymptotics:

  n fixed and $d \to \infty$

Eigen-Decomposition

Assume that $X_1,\ldots, X_n \sim N(0, \Sigma_d)$

Eigen-Decomposition

Assume that $X_1, \ldots, X_n \sim N(0, \Sigma_d)$

- $\Sigma_d = U \Lambda U^T$

Eigen-Decomposition

Assume that $X_1,\ldots, X_n \sim N(0, \Sigma_d)$

- $\Sigma_d = U \Lambda U^T$
- $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)$

Eigen-Decomposition

Assume that $X_1, \dots, X_n \sim N(0, \Sigma_d)$

- $\Sigma_d = U \Lambda U^T$
- $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_d)$
- $U = [u_1, \dots, u_d]$

Eigen-Decomposition

Assume that $X_1,\ldots, X_n \sim N(0, \Sigma_d)$

- $\Sigma_d = U \Lambda U^T$
- $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$
- $U = [u_1, \ldots, u_d]$

Denote $\hat{\Sigma}_d = n^{-1} X X^T$, where $X = [X_1,\ldots, X_n]$

Eigen-Decomposition

Assume that $X_1, \ldots, X_n \sim N(0, \Sigma_d)$

- $\Sigma_d = U \Lambda U^T$
- $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)$
- $U = [u_1, \ldots, u_d]$

Denote $\hat{\Sigma}_d = n^{-1} X X^T$, where $X = [X_1, \ldots, X_n]$

- $\hat{\Sigma}_d = \hat{U} \hat{\Lambda} \hat{U}^T$

Eigen-Decomposition

Assume that $X_1,\dots, X_n \sim N(0, \Sigma_d)$

- $\Sigma_d = U \Lambda U^T$
- $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d)$
- $U = [u_1, \dots, u_d]$

Denote $\hat{\Sigma}_d = n^{-1} X X^T$, where $X = [X_1, \dots, X_n]$

- $\hat{\Sigma}_d = \hat{U} \hat{\Lambda} \hat{U}^T$
- $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_d)$

Eigen-Decomposition

Assume that $X_1, \ldots, X_n \sim N(0, \Sigma_d)$

- $\Sigma_d = U \Lambda U^T$
- $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$
- $U = [u_1, \ldots, u_d]$

Denote $\hat{\Sigma}_d = n^{-1} X X^T$, where $X = [X_1, \ldots, X_n]$

- $\hat{\Sigma}_d = \hat{U} \hat{\Lambda} \hat{U}^T$
- $\hat{\Lambda} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_d)$
- $\hat{U} = [\hat{u}_1, \ldots, \hat{u}_d]$

Eigen-Decomposition

Assume that $X_1, \ldots, X_n \sim N(0, \Sigma_d)$

- $\Sigma_d = U \Lambda U^T$
- $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$
- $U = [u_1, \ldots, u_d]$

Denote $\hat{\Sigma}_d = n^{-1} X X^T$, where $X = [X_1, \ldots, X_n]$

- $\hat{\Sigma}_d = \hat{U} \hat{\Lambda} \hat{U}^T$
- $\hat{\Lambda} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_d)$
- $\hat{U} = [\hat{u}_1, \ldots, \hat{u}_d]$

Study angle $(\hat{u}_j, u_j)$

Information Contribution

Contribution to consistency

- n: positive
- d: negative
- Spike size (e.g. λ_1 / λ_2) : positive
  - relative sizes of the leading eigenvalues

Information Contribution

Contribution to consistency

- n: positive
- d: negative
- Spike size (e.g. λ_1/λ_2) : positive
  - relative sizes of the leading eigenvalues

Question:

- Interaction among the three informations ←→ Consistency of PCA???

Outline

- Motivation & Background
- PCA Asymptotics
- **Spike Covariance Models**
- Theoretical Results of PCA
- Sparse PCA
- Summary
- Analysis of Tree Data

Spike Covariance Model

- Johnstone (2001)
- General math description of m-component spike model

- Examples:
  - m=1: single component spike model
    - $\lambda_1 >> \lambda_2 \sim \ldots \sim \lambda_d \sim 1$
  - m>1: multi-component spike model
    - $\lambda_1 > \ldots > \lambda_m >> \lambda_{m+1} \sim \ldots \sim \lambda_d \sim 1$
  - multi-component with tiered eigen-values
    - $\lambda_1 \geq \ldots \geq \lambda_m >> \lambda_{m+1} \sim \ldots \sim \lambda_d \sim 1$

Single-Component Spike Model

Example 1:

- $\lambda_1 \sim d^{\alpha}$, $\lambda_2 = \ldots = \lambda_d = 1$
- $n \sim d^{\gamma}$
- Sample index: γ and Spike index: α

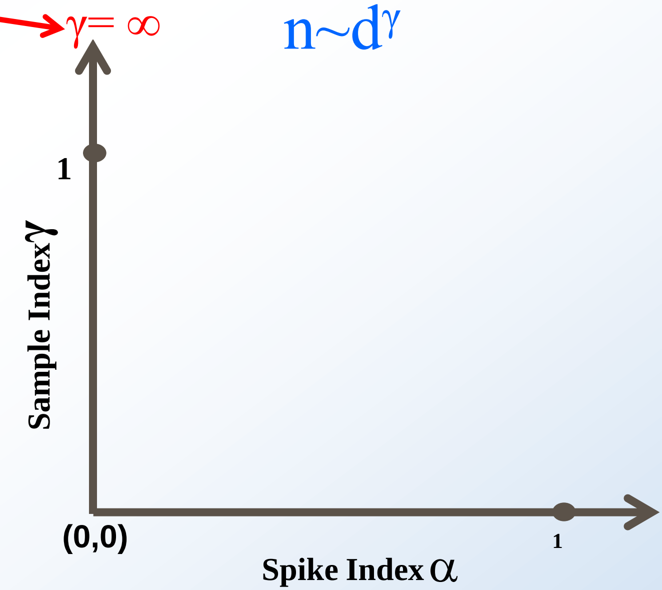
Single Spike General Framework

Classical asymptotics

- Anderson (1963): consistent when d fixed and $n \rightarrow \infty$

$\lambda_1 \sim d^{\alpha}$

$n \sim d^{\gamma}$



Single Spike General Framework

Classical asymptotics

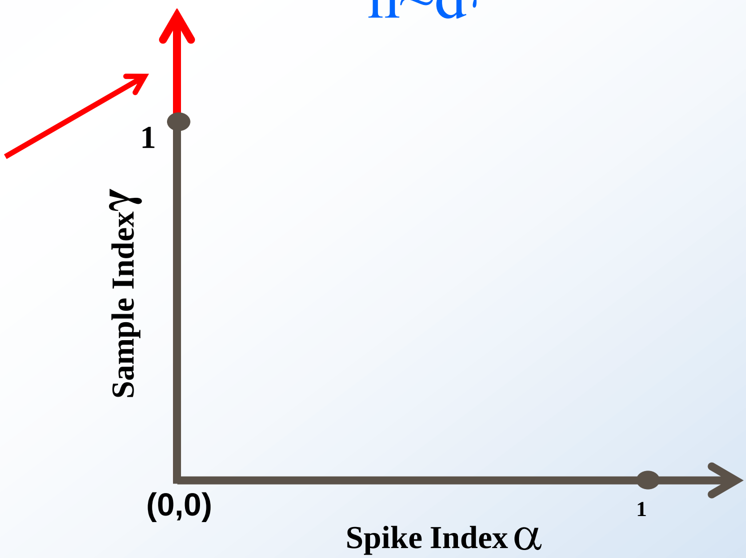
- Anderson (1963): consistent when d fixed and $n \rightarrow \infty$

Random matrix asymptotics

- Johnstone and Lu (2009): consistent when $\alpha=0$, $\gamma > 1$

$\lambda_1 \sim d^{\alpha}$

$n \sim d^{\gamma}$



Single Spike General Framework

Classical asymptotics

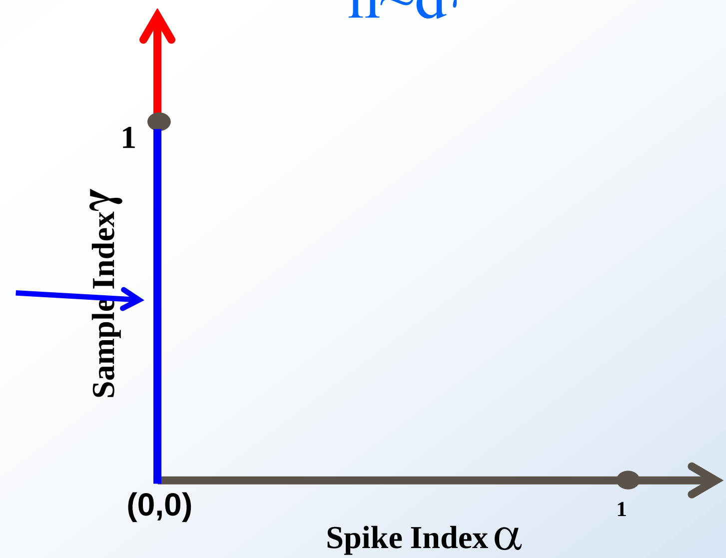
- Anderson (1963): consistent when d fixed and n→ ∞

Random matrix asymptotics

- Johnstone and Lu (2009): consistent when α=0, γ> 1

- Johnstone and Lu (2009): str. incon. when α=0, γ< 1

$\lambda_1 \sim d^\alpha$
$n \sim d^\gamma$



Single Spike General Framework

Classical asymptotics

• Anderson (1963): consistent when d fixed and n → ∞

Random matrix asymptotics

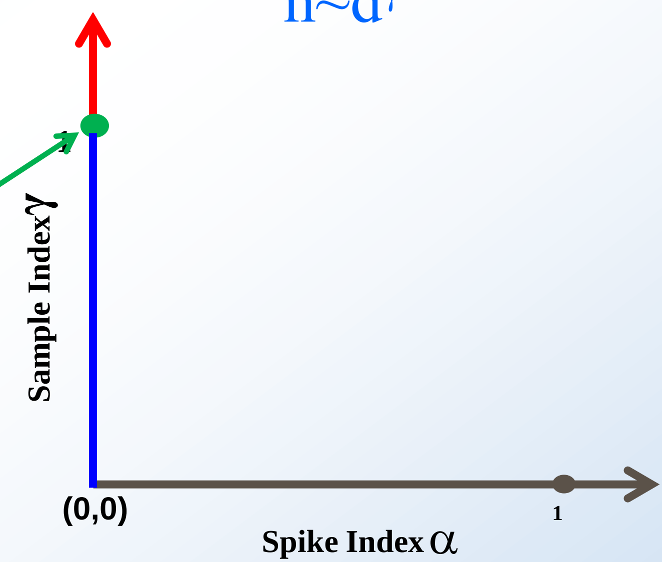
• Johnstone and Lu (2009): consistent when α=0, γ> 1

• Johnstone and Lu (2009): str. incon. when α=0, γ< 1

• Nadler (2008) : boundary case α=0, γ=1

$\lambda_1 \sim d^{\alpha}$

$n \sim d^{\gamma}$



Single Spike General Framework

Classical asymptotics

- Anderson (1963): consistent when d fixed and n → ∞

Random matrix asymptotics

- Johnstone and Lu (2009): consistent when α=0, γ> 1

- Johnstone and Lu (2009): str. incon. when α=0, γ< 1

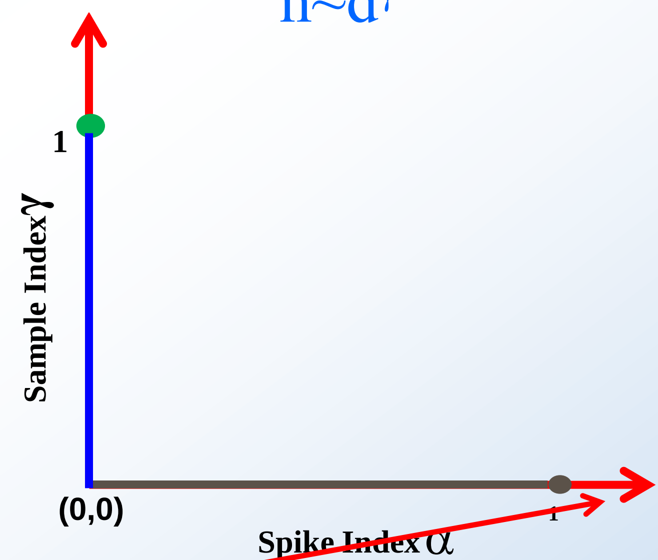
- Nadler (2008) : boundary case α=0, γ=1

HDLSS asymptotics

- Jung and Marron (2009): consistent when α >1, γ=0

$\lambda_1 \sim d^{\alpha}$

$n \sim d^{\gamma}$



Single Spike General Framework

Classical asymptotics

• Anderson (1963): consistent when d fixed and n $\rightarrow \infty$

Random matrix asymptotics

• Johnstone and Lu (2009): consistent when α=0, γ> 1

• Johnstone and Lu (2009): str. incon. when α=0, γ< 1

• Nadler (2008) : boundary case α=0, γ=1

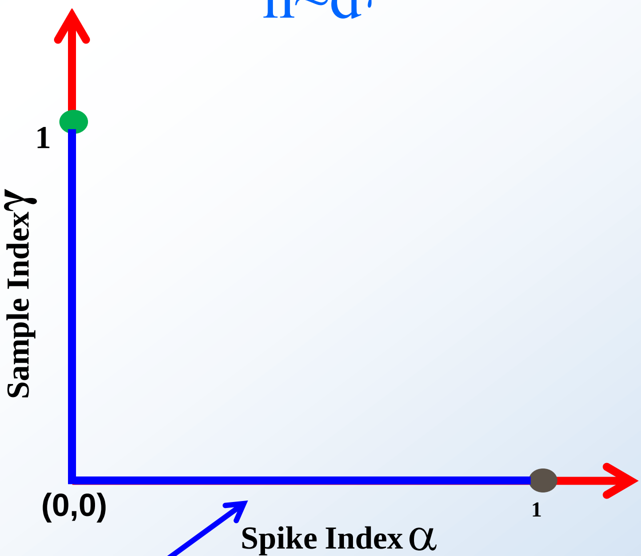
HDLSS asymptotics

• Jung and Marron (2009): consistent when α >1, γ=0

• Jung and Marron (2009): strongly inconsistent when α <1, γ=0

$\lambda_1 \sim d^{\alpha}$

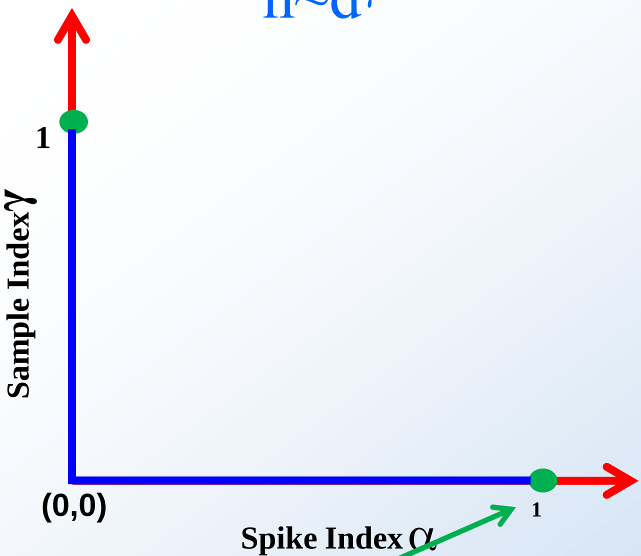
$n \sim d^{\gamma}$



UNC, Stat & OR

Single Spike General Framework

$\lambda_1 \sim d^{\alpha}$
$n \sim d^{\gamma}$



Classical asymptotics
- Anderson (1963): consistent when d fixed and n → ∞

Random matrix asymptotics
- Johnstone and Lu (2009): consistent when $\alpha=0$, $\gamma > 1$
- Johnstone and Lu (2009): str. incon. when $\alpha=0$, $\gamma < 1$
- Nadler (2008) : boundary case $\alpha=0$, $\gamma=1$

HDLSS asymptotics
- Jung and Marron (2009): consistent when $\alpha > 1$, $\gamma=0$
- Jung and Marron (2009): strongly inconsistent when $\alpha < 1$, $\gamma=0$
- Jung et al. (2010): boundary case $\alpha=1$, $\gamma=0$

Single Spike General Framework

Classical asymptotics

- Anderson (1963): consistent when d fixed and n→ ∞

Random matrix asymptotics

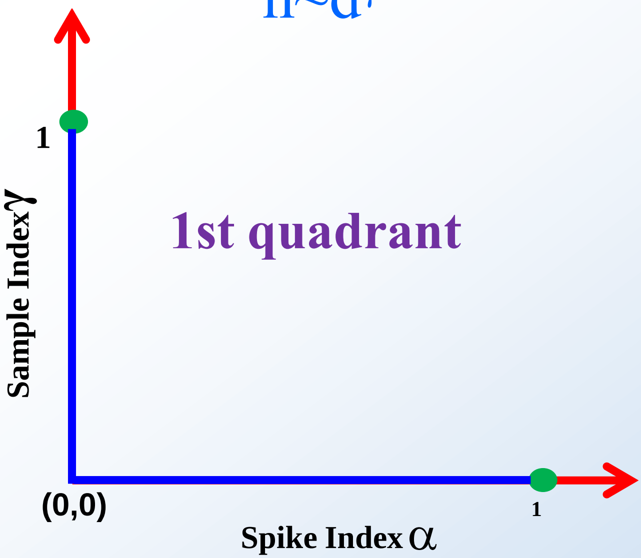
- Johnstone and Lu (2009): consistent when $\alpha=0$, $\gamma> 1$
- Johnstone and Lu (2009): str. incon. when $\alpha=0$, $\gamma< 1$
- Nadler (2008) : boundary case $\alpha=0$, $\gamma=1$

HDLSS asymptotics

- Jung and Marron (2009): consistent when $\alpha >1$, $\gamma=0$
- Jung and Marron (2009): strongly inconsistent when $\alpha <1$, $\gamma=0$
- Jung et al. (2010): boundary case $\alpha=1$, $\gamma=0$

$\lambda_1 \sim d^{\alpha}$

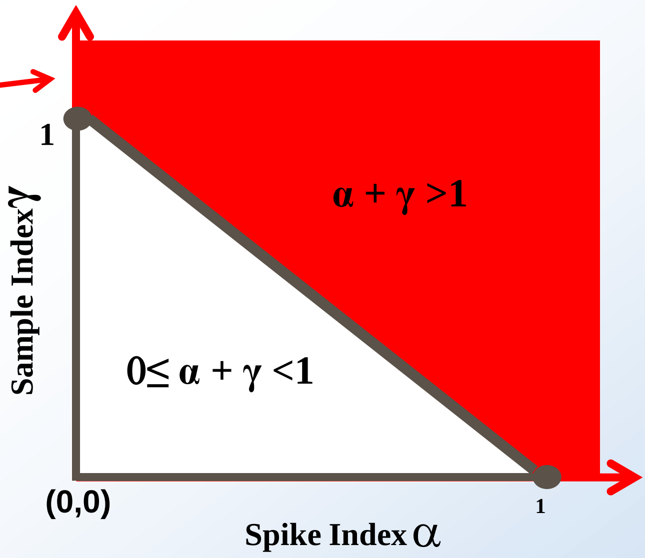
$n \sim d^{\gamma}$



Single Spike General Framework

Our result : bridge between settings

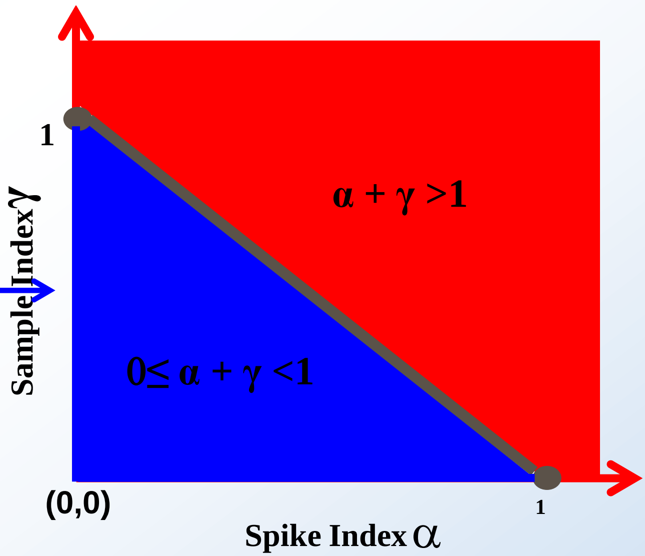
- consistent when α +γ> 1



Single Spike General Framework

Our result : bridge between settings

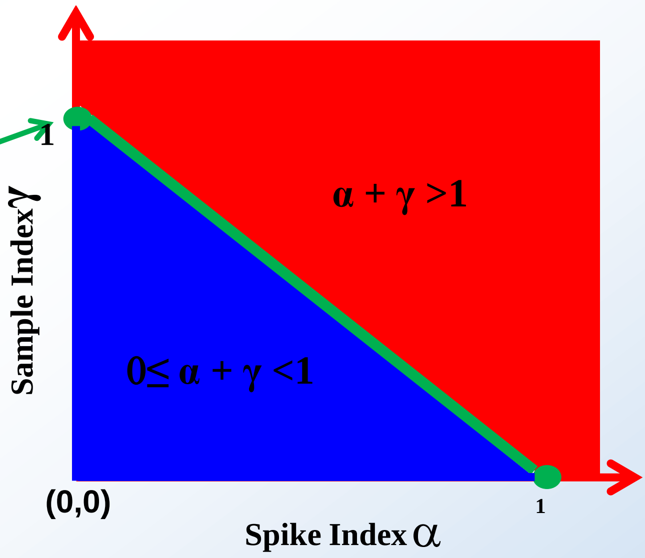
- consistent when α +γ> 1
- strongly inconsistent when α +γ<1



Single Spike General Framework

Our result : bridge between settings

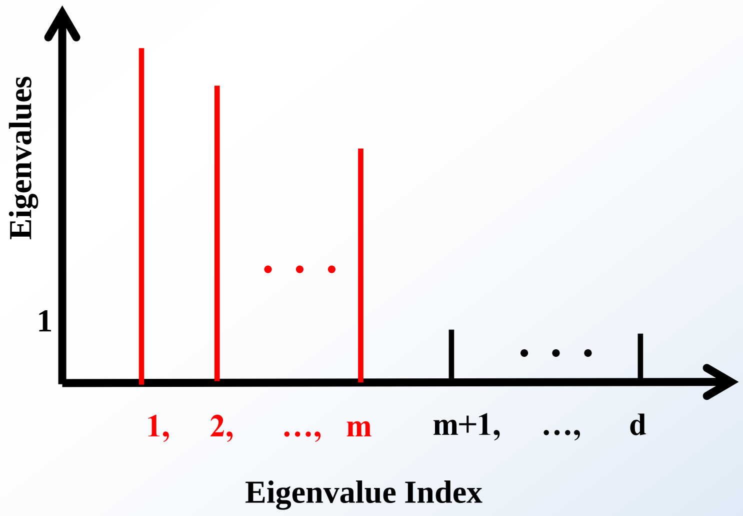
- consistent when $\alpha + \gamma > 1$
- strongly inconsistent when $\alpha + \gamma < 1$
- boundary case $\alpha + \gamma = 1$



Multi-Component Spike Model

Example 2:

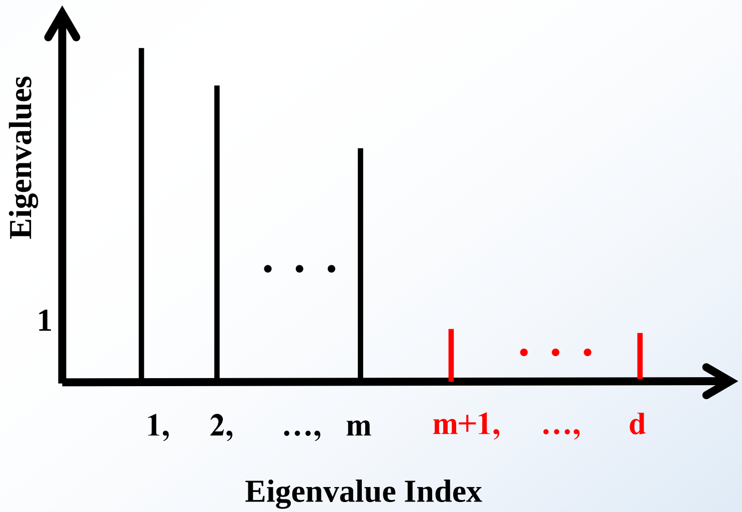
- $\lambda_j = c_j d^{\alpha}$, $j \le m$, where $c_j > c_{j-1} > 0$



Multi-Component Spike Model

Example 2:

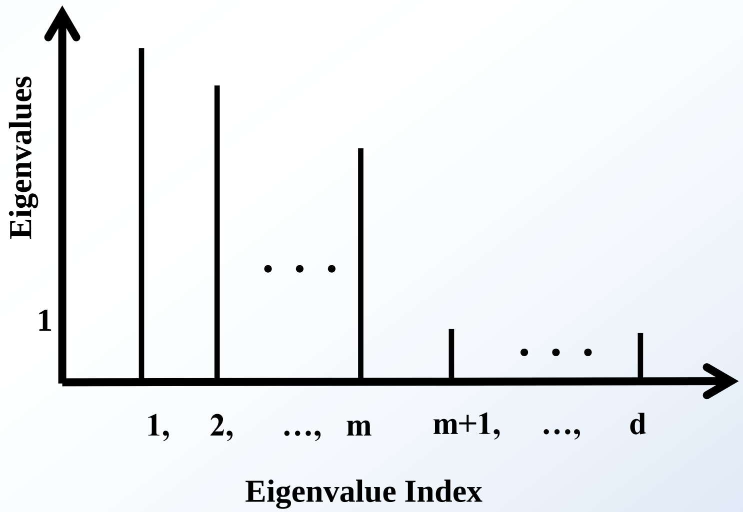
- $\lambda_j = c_j d^{\alpha}$, $j \le m$, where $c_j > c_{j-1} > 0$
- $\lambda_{m+1} = \ldots = \lambda_d = 1$



Multi-Component Spike Model

Example 2:

- $\lambda_j = c_j d^{\alpha}$, $j \le m$, where $c_j > c_{j-1} > 0$
- $\lambda_{m+1} = \ldots = \lambda_d = 1$
- $n \sim d^{\gamma}$

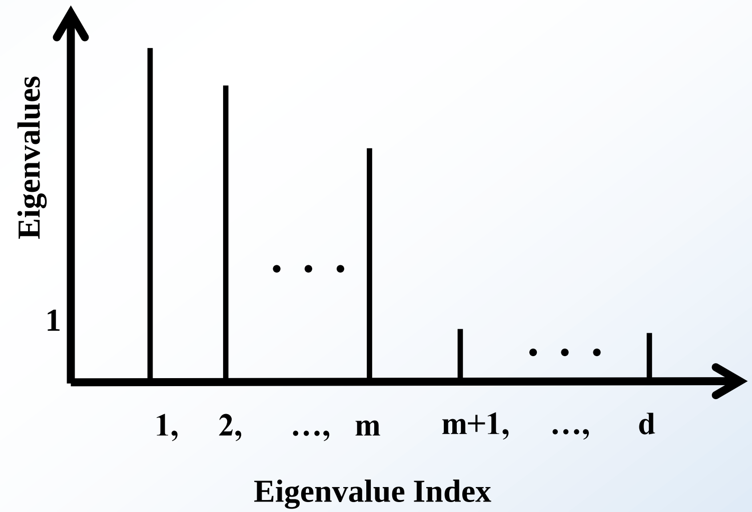


Multi-Component Spike Model

UNC, Stat & OR

Example 2:

- $\lambda_j = c_j d^\alpha$, $j \le m$, where $c_j > c_{j-1} > 0$
- $\lambda_{m+1} = \ldots = \lambda_d = 1$
- $n \sim d^\gamma$
- Sample index: γ and Spike index: α
  - Common α for λ_j, $j=1,\ldots,m$

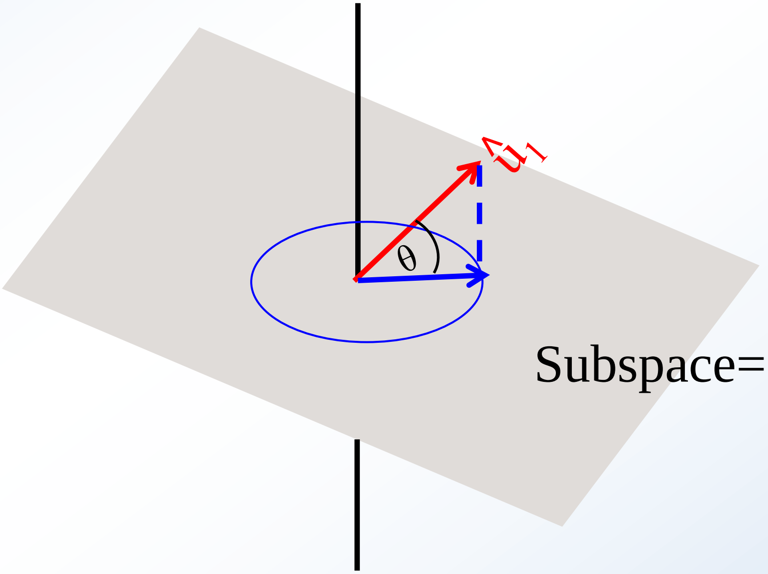


Subspace Consistency

Introduced by Jung and Marron (2009) under HDLSS asymptotics

Similar eigenvalues:

- Eigen-direction not identified
- Focus on subspace (generated)



Subspace=span$\{u_1, \ldots, u_m\}$

As $n \to \infty$, $d \to \infty$, or $d\&n \to \infty$

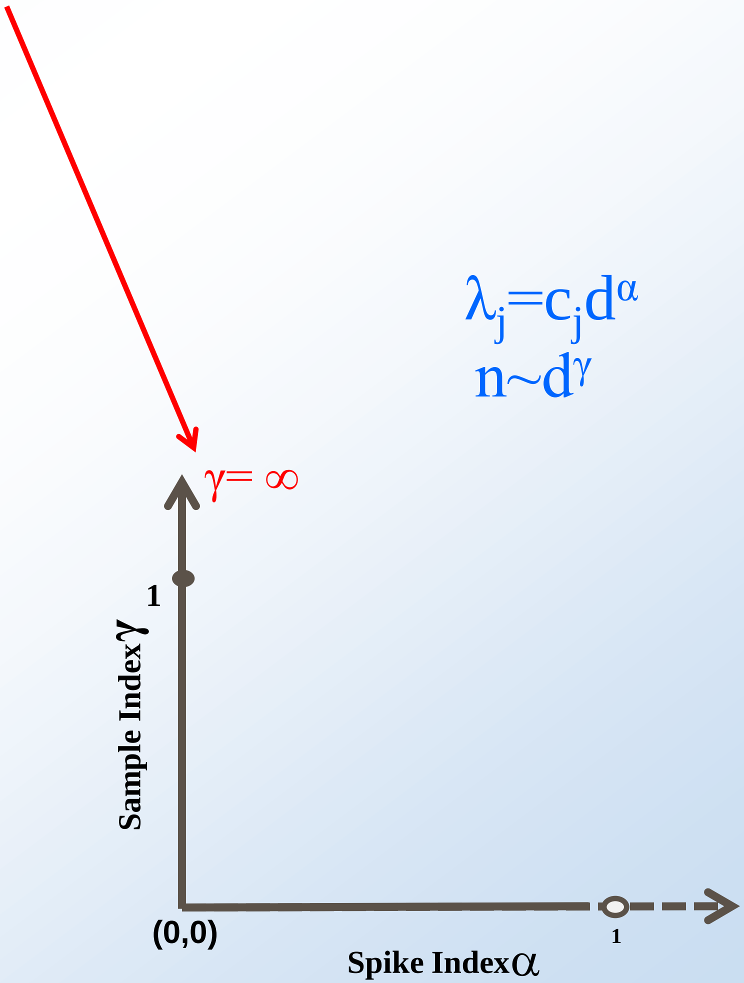
- Subspace consistency: $\theta \to 0$

Multi-Spike General Framework

Classical asymptotics

- Anderson (1963): consistent when d fixed and n→ ∞

$\lambda_j = c_j d^{\alpha}$
$n \sim d^{\gamma}$



Multi-Spike General Framework

Classical asymptotics

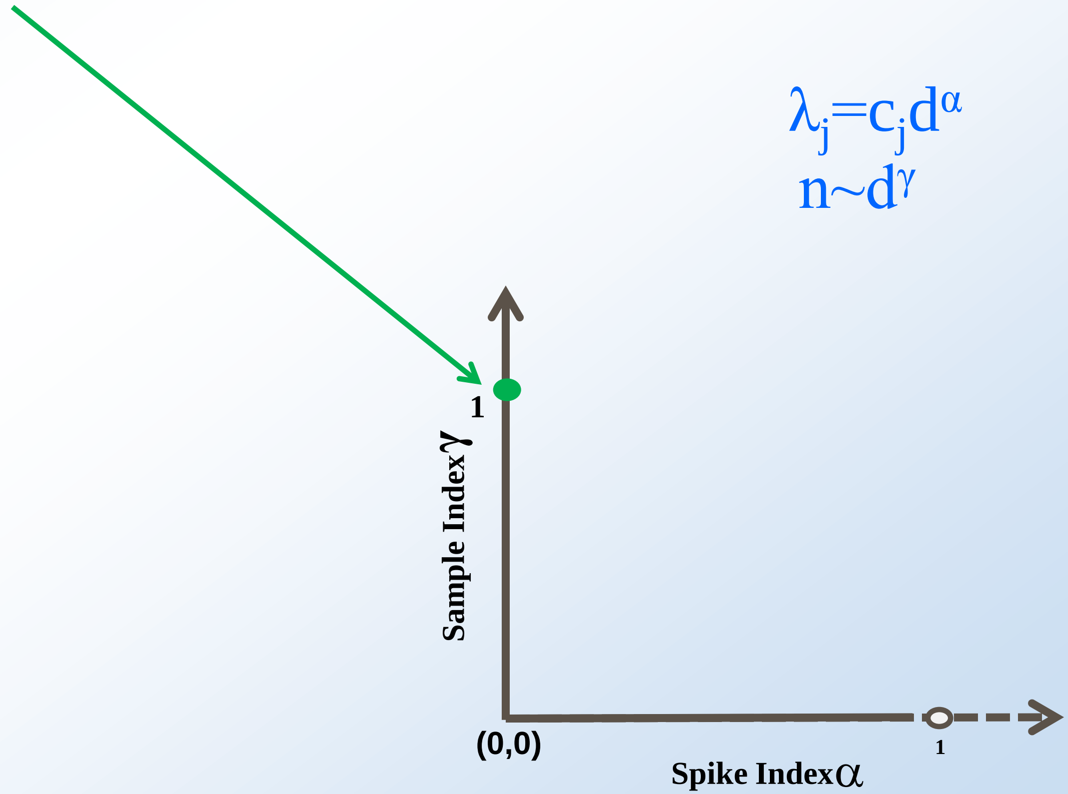
- Anderson (1963): consistent when d fixed and n$\rightarrow \infty$

Random matrix asymptotics

- Paul (2007) : boundary case α=0, γ=1

$\lambda_j = c_j d^\alpha$

$n \sim d^\gamma$



Multi-Spike General Framework

UNC, Stat & OR

Classical asymptotics

- Anderson (1963): consistent when d fixed and n→ ∞

Random matrix asymptotics

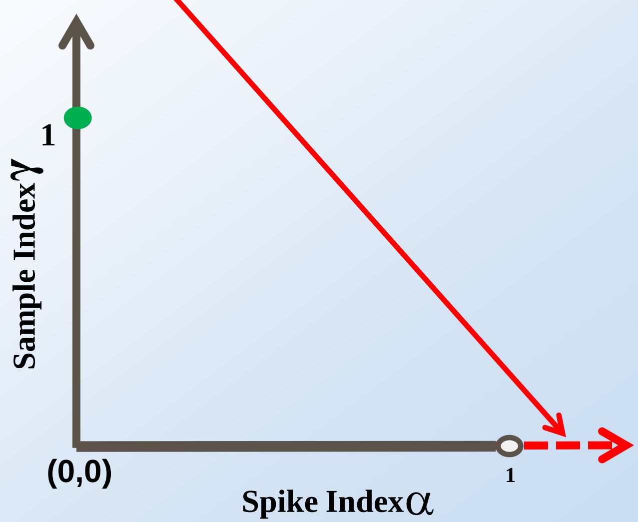
- Paul (2007) : boundary case α=0, γ=1

HDLSS asymptotics

- Jung and Marron (2009): subspace consistent when α >1, γ=0

$\lambda_j = c_j d^{\alpha}$

$n \sim d^{\gamma}$



Multi-Spike General Framework

Classical asymptotics

- Anderson (1963): consistent when d fixed and n→ ∞

Random matrix asymptotics

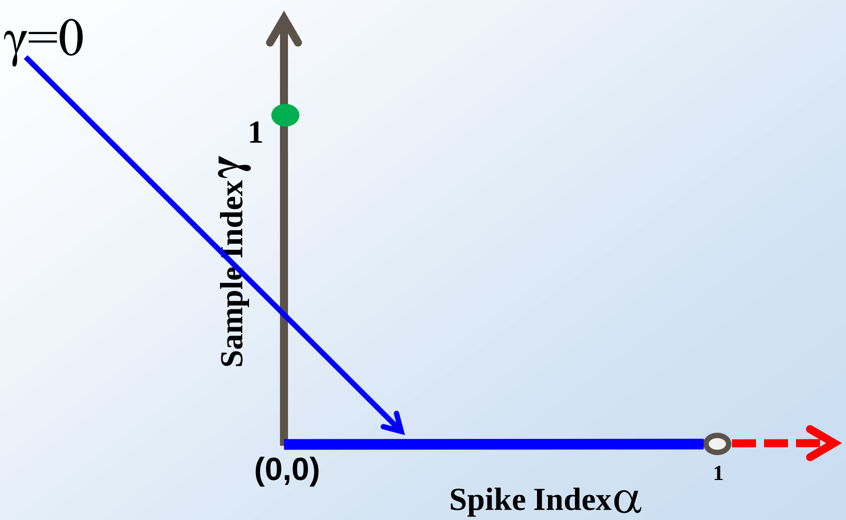
- Paul (2007) : boundary case α=0, γ=1

HDLSS asymptotics

- Jung and Marron (2009): subspace consistent when α >1, γ=0

$\lambda_j = c_j d^{\alpha}$
$n \sim d^{\gamma}$

- Jung and Marron (2009): str. incon. when α <1, γ=0

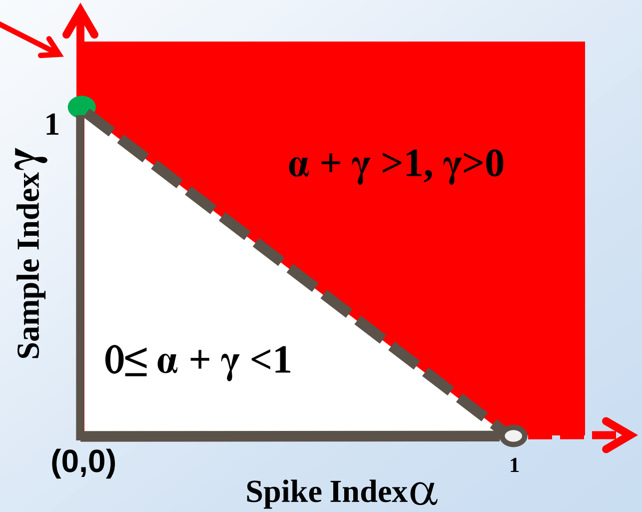


Multi-Spike General Framework

Our result : bridge between settings

- consistent when $\alpha + \gamma > 1, \gamma > 0$
- strongly inconsistent when $\alpha + \gamma < 1$

$\lambda_j = c_j d^{\alpha}$
$n \sim d^{\gamma}$



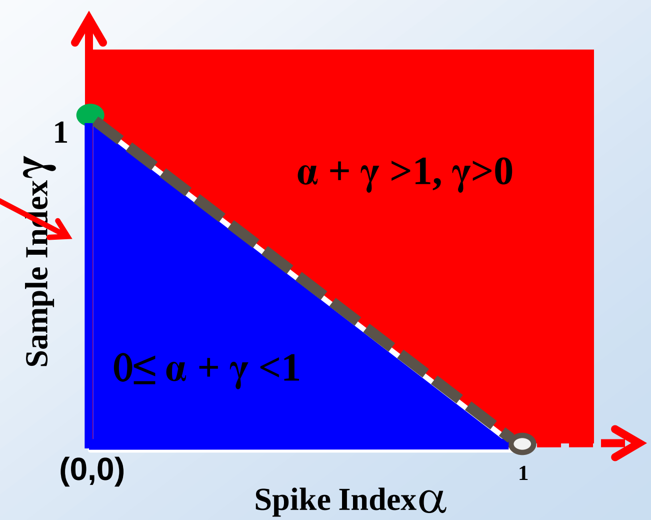
Multi-Spike General Framework

UNC, Stat & OR

Our result : bridge between settings

- consistent when $\alpha + \gamma > 1$, $\gamma > 0$
- strongly inconsistent when $\alpha + \gamma < 1$

$\lambda_j = c_j d^{\alpha}$
$n \sim d^{\gamma}$

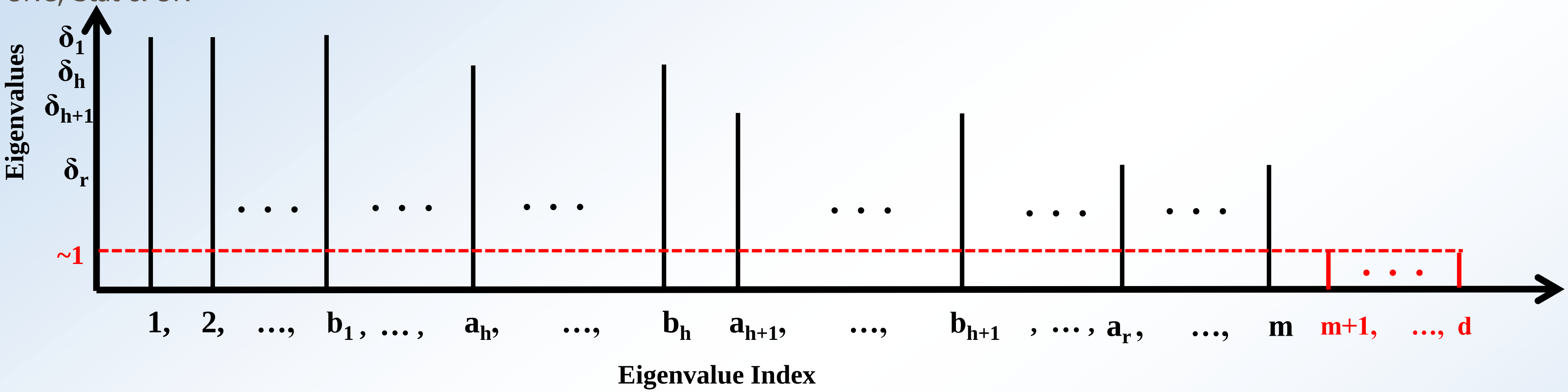


Outline

- Motivation & Background
- PCA Asymptotics
- Spike Covariance Models
- **Theoretical Results of PCA**
- Sparse PCA
- Summary
- Analysis of Tree Data

Assumptions on Spikes

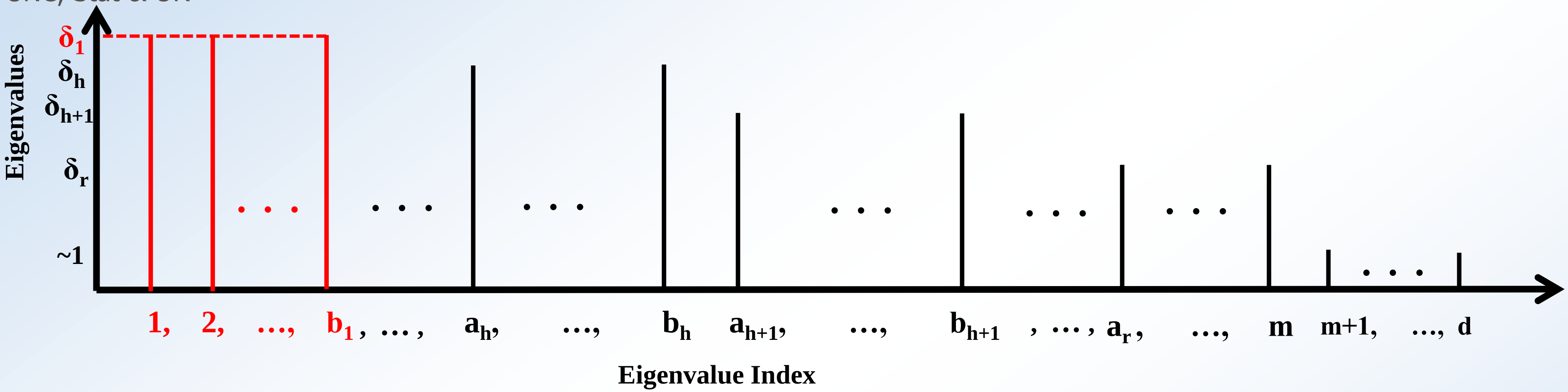
UNC, Stat & OR



Assumption: as n → ∞

- $\lambda_{m+1}, \ldots, \lambda_d \sim 1$

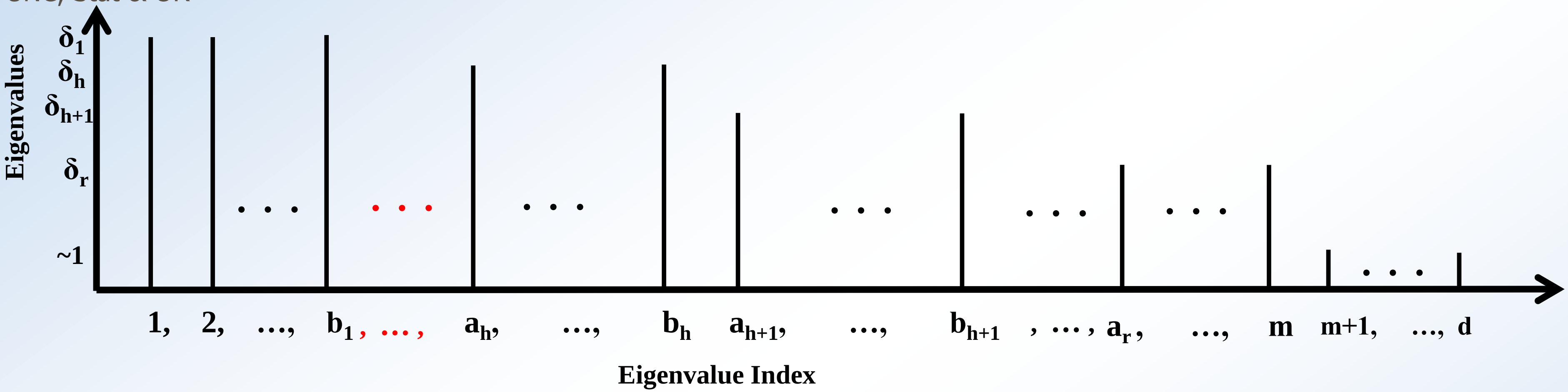
Assumptions on Spikes



Assumption: as n → ∞

- $\lambda_{m+1}, \ldots, \lambda_d \sim 1$
- $\lambda_1, \ldots, \lambda_{b_1} \rightarrow \delta_1$

Assumptions on Spikes

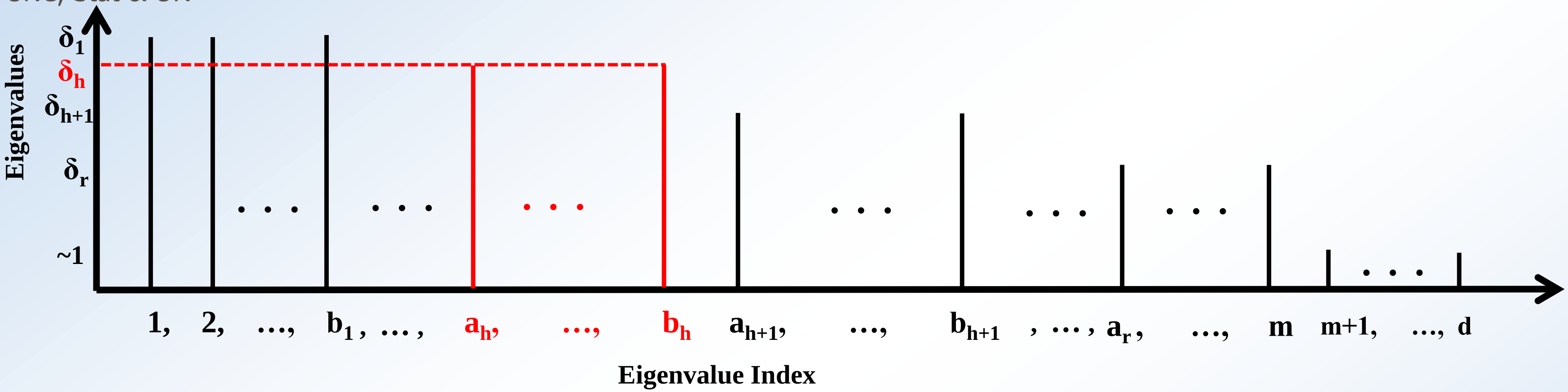


Assumption: as n → ∞

- $\lambda_{m+1}, \ldots, \lambda_d \sim 1$
- $\lambda_1, \ldots, \lambda_{b_1} \rightarrow \delta_1$

⋮

Assumptions on Spikes



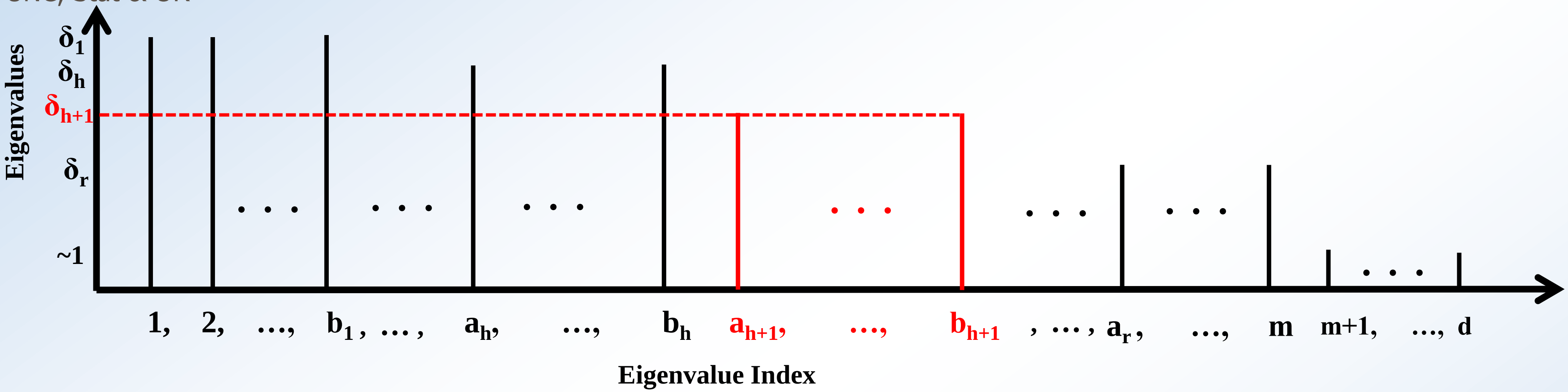
Assumption: as $n \rightarrow \infty$

- $\lambda_{m+1}, \ldots, \lambda_d \sim 1$
- $\lambda_1, \ldots, \lambda_{b_1} \rightarrow \delta_1$

⋮

- $\lambda_{a_h}, \ldots, \lambda_{b_h} \rightarrow \delta_h$

Assumptions on Spikes



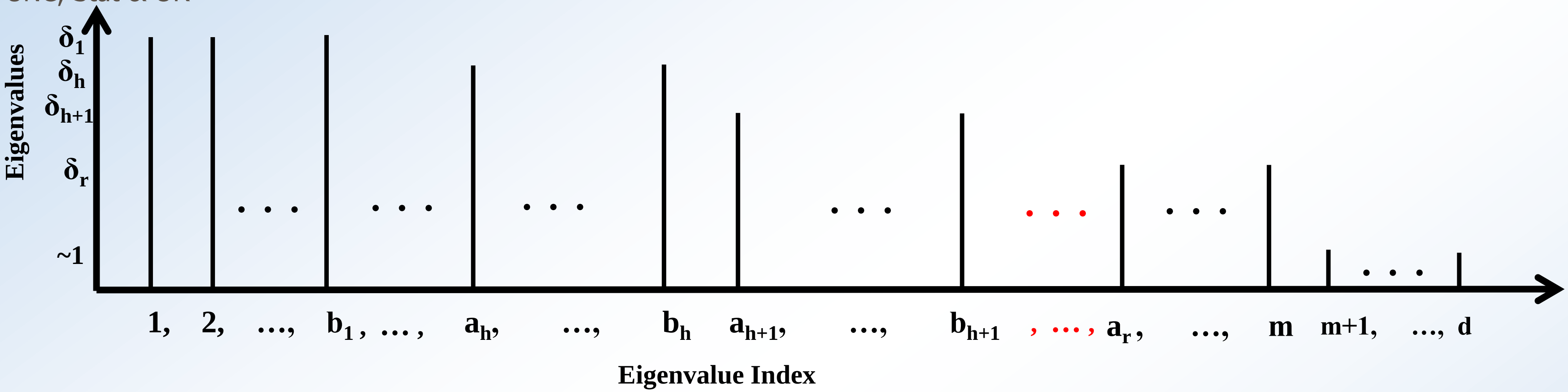
Assumption: as $n \to \infty$

- $\lambda_{m+1}, \ldots, \lambda_d \sim 1$
- $\lambda_1, \ldots, \lambda_{b_1} \to \delta_1$

  ⋮
- $\lambda_{a_h}, \ldots, \lambda_{b_h} \to \delta_h$
- $\lambda_{a_{h+1}}, \ldots, \lambda_{b_{h+1}} \to \delta_{h+1}$

Assumptions on Spikes

UNC, Stat & OR



Assumption: as $n \rightarrow \infty$

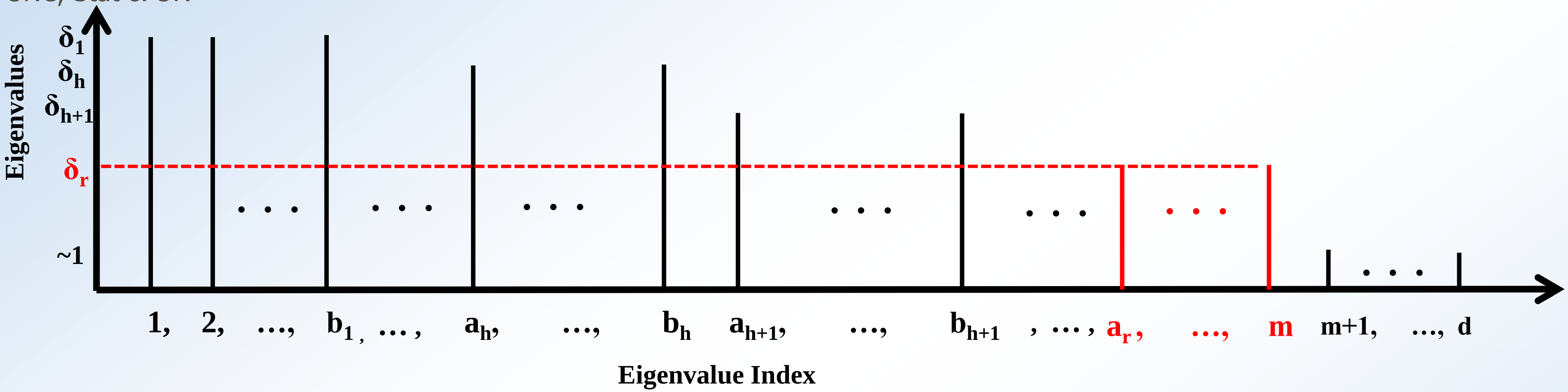
- $\lambda_{m+1}, \ldots, \lambda_d \sim 1$
- $\lambda_1, \ldots, \lambda_{b_1} \rightarrow \delta_1$

⋮

- $\lambda_{a_h}, \ldots, \lambda_{b_h} \rightarrow \delta_h$
- $\lambda_{a_{h+1}}, \ldots, \lambda_{b_{h+1}} \rightarrow \delta_{h+1}$

⋮

Assumptions on Spikes



Assumption: as n → ∞

- $\lambda_{m+1}, \ldots, \lambda_d \sim 1$
- $\lambda_1, \ldots, \lambda_{b_1} \rightarrow \delta_1$

⋮

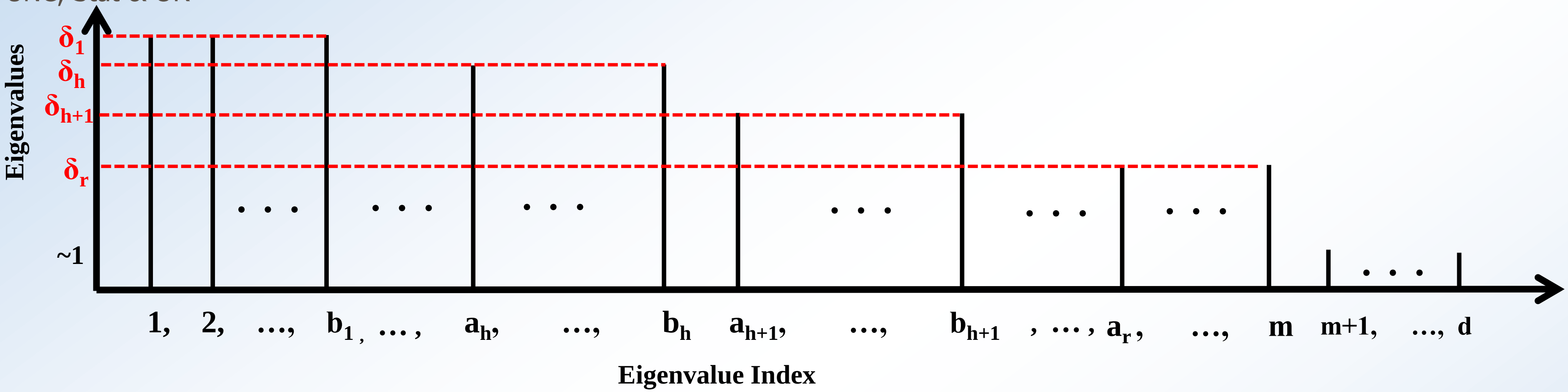
- $\lambda_{a_h}, \ldots, \lambda_{b_h} \rightarrow \delta_h$
- $\lambda_{a_{h+1}}, \ldots, \lambda_{b_{h+1}} \rightarrow \delta_{h+1}$

⋮

- $\lambda_{a_r}, \ldots, \lambda_m \rightarrow \delta_r$

Assumptions on Spikes

UNC, Stat & OR



Assumption: as $n \to \infty$

- $\lambda_{m+1}, \ldots, \lambda_d \sim 1$
- $\lambda_1, \ldots, \lambda_{b_1} \to \delta_1$

⋮

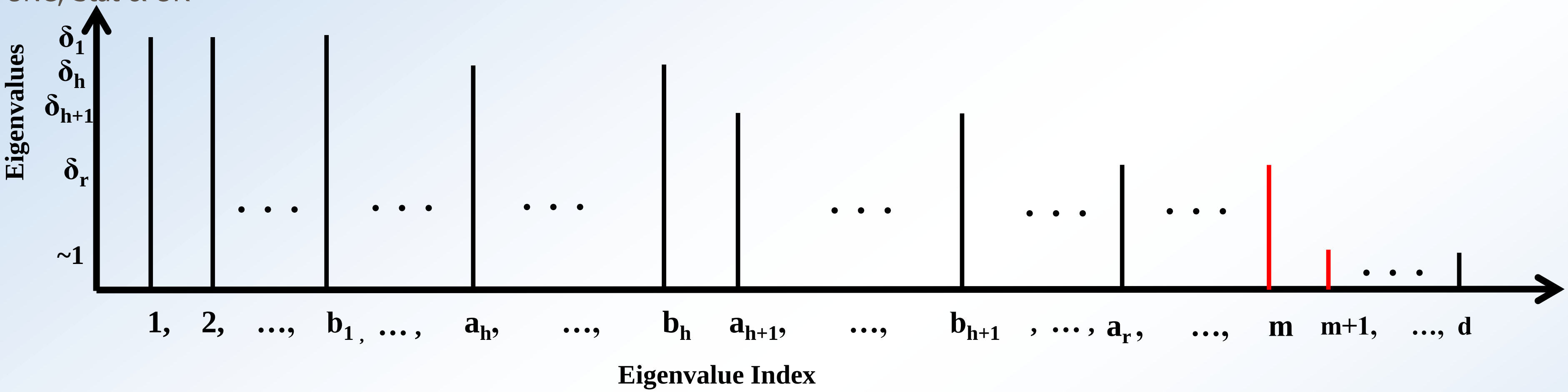
- $\lambda_{a_h}, \ldots, \lambda_{b_h} \to \delta_h$
- $\lambda_{a_{h+1}}, \ldots, \lambda_{b_{h+1}} \to \delta_{h+1}$

⋮

- $\lambda_{a_r}, \ldots, \lambda_m \to \delta_r$

- $\overline{\lim}\ \delta_{i+1}/\delta_i < 1$

Assumptions on Spikes



Assumption: as $n \to \infty$

- $\lambda_{m+1}, \ldots, \lambda_d \sim 1$
- $\lambda_1, \ldots, \lambda_{b_1} \to \delta_1$

⋮

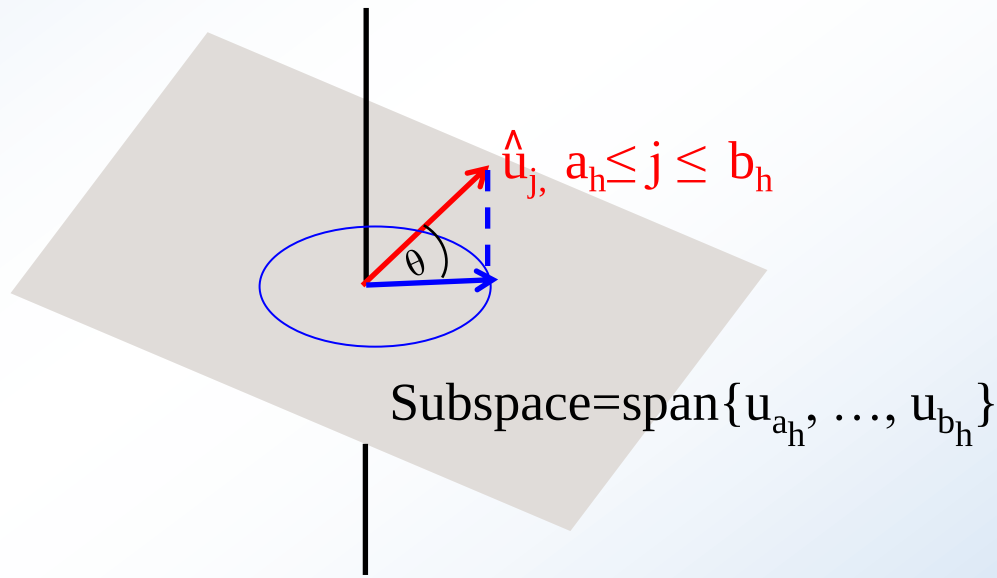
- $\lambda_{a_h}, \ldots, \lambda_{b_h} \to \delta_h$
- $\lambda_{a_{h+1}}, \ldots, \lambda_{b_{h+1}} \to \delta_{h+1}$

⋮

- $\lambda_{a_r}, \ldots, \lambda_m \to \delta_r$

- $\overline{lim}\ \delta_{i+1}/\delta_i < 1$
- $\lambda_{m+1}/\lambda_m \to 0$

Subspace & Eigenvalue Consistency

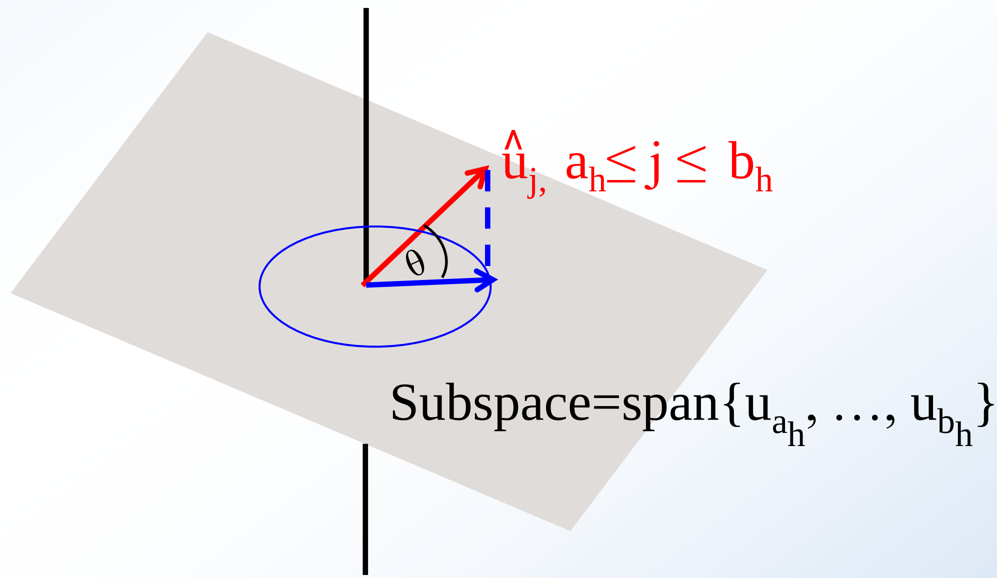


As $n \to \infty$, $d \to \infty$, or $d\&n \to \infty$

- Subspace consistency: $\theta \to 0$

UNC, Stat & OR

Subspace & Eigenvalue Consistency

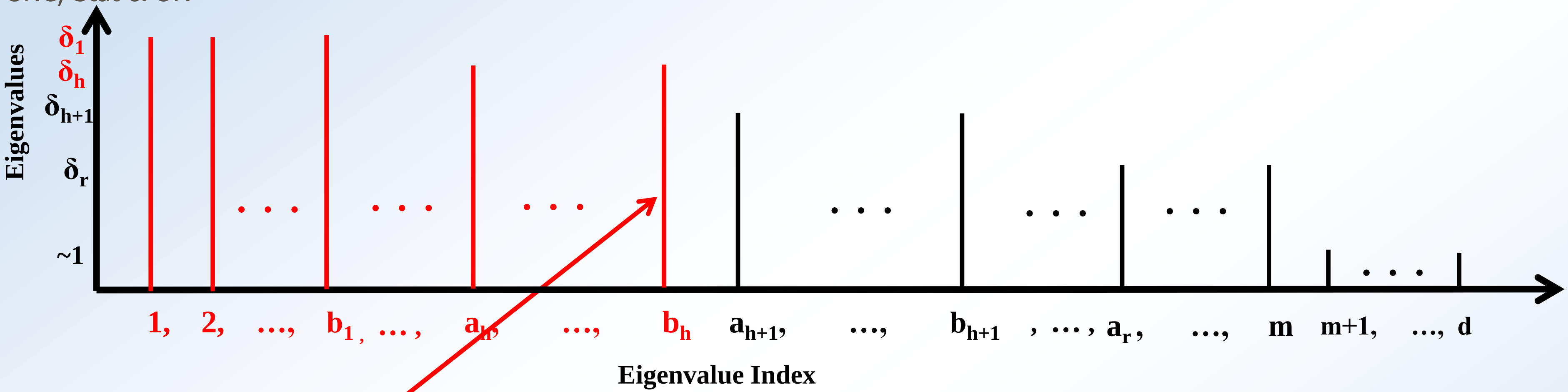


As $n \to \infty$, $d \to \infty$, or $d \& n \to \infty$

- Subspace consistency: $\theta \to 0$
- Eigenvalue consistency: $\hat{\lambda}_j / \lambda_j \to 1$

Main Theorem 1

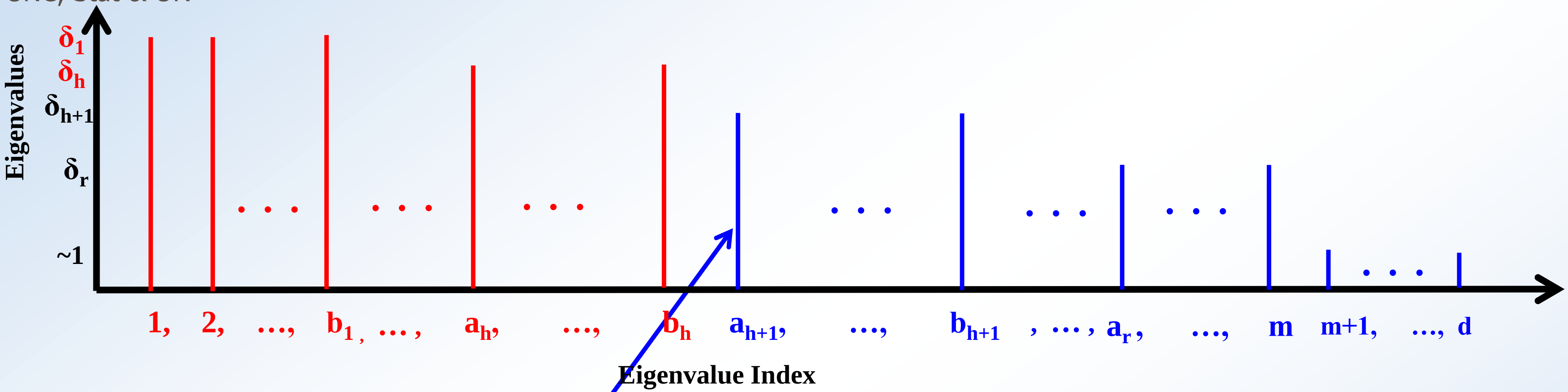
UNC, Stat & OR



Assumption: as $n \to \infty$

- If $d/(n\delta_h) \to 0$, then $\hat{\lambda}_j$ is consistent and $\hat{u}_j$ is subspace consistency, $j \le b_h$

Main Theorem 1



Assumption: as $n \to \infty$

- If $d/(n\delta_h) \to 0$, then $\hat{\lambda}_j$ is consistent and $\hat{u}_j$ is subspace consistency, $j \leq b_h$
- In addition $d/(n\delta_{h+1}) \to \infty$, then $\hat{\lambda}_j \overset{a.s.}{\approx} d/n$ and $\hat{u}_j$ is strongly inconsistent, $j > b_h$

Remark

- If h=0, all $\hat{u}_j$ are strongly inconsistent

Remark

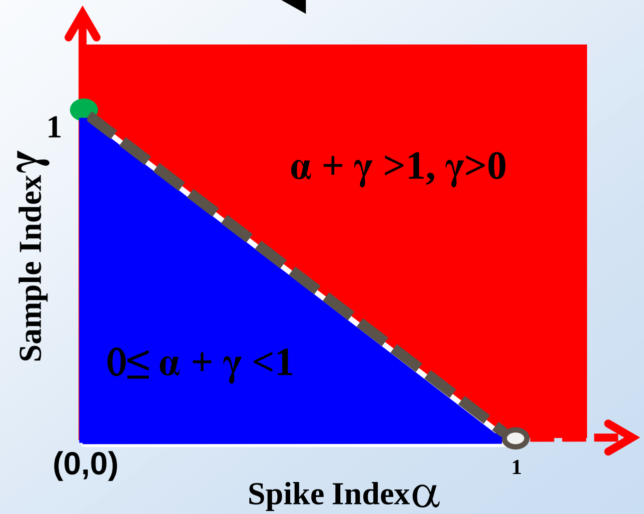
- If h=0, all $\hat{u}_j$ are strongly inconsistent

- If h=r, all $\hat{u}_j$ are subspace consistent

Remark

UNC, Stat & OR

- If h=0, all $\hat{u}_j$ are strongly inconsistent
- If h=r, all $\hat{u}_j$ are subspace consistent
- If a_h=b_h, subspace consistency becomes consistency (Example 2)



Remark

- If h=0, all $\hat{u}_j$ are strongly inconsistent
- If h=r, all $\hat{u}_j$ are subspace consistent
- If $a_h=b_h$, subspace consistency becomes consistency (Example 2)
- For fixed n and d$\to\infty$, condition $\overline{\lim}\ \delta_{i+1}/\delta_i < 1$ should be strengthened to $\lim\ \delta_{i+1}/\delta_i = 0$

Main Theorem 2

Boundary case for single spike model

Assumption: as $n \rightarrow \infty$

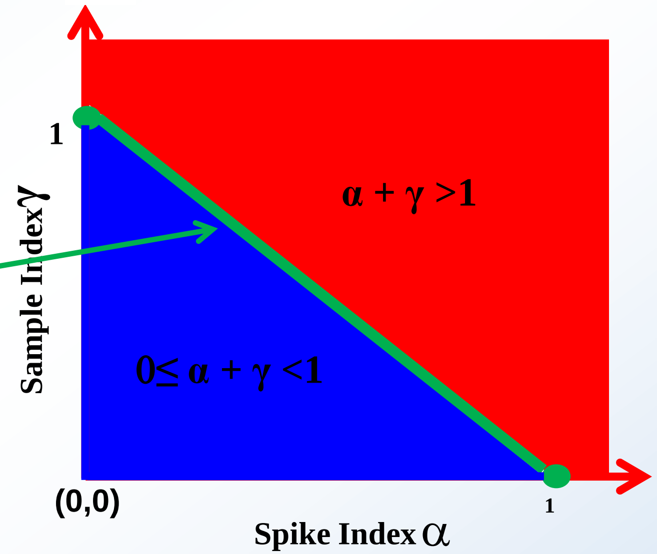
- $\lambda_1 >> \lambda_2 = \ldots = \lambda_d = 1$

Main Theorem 2

Boundary case for single spike model

Assumption: as $n \to \infty$

- $\lambda_1 >> \lambda_2 = \ldots = \lambda_d = 1$
- $d/(n\lambda_1) \to c$

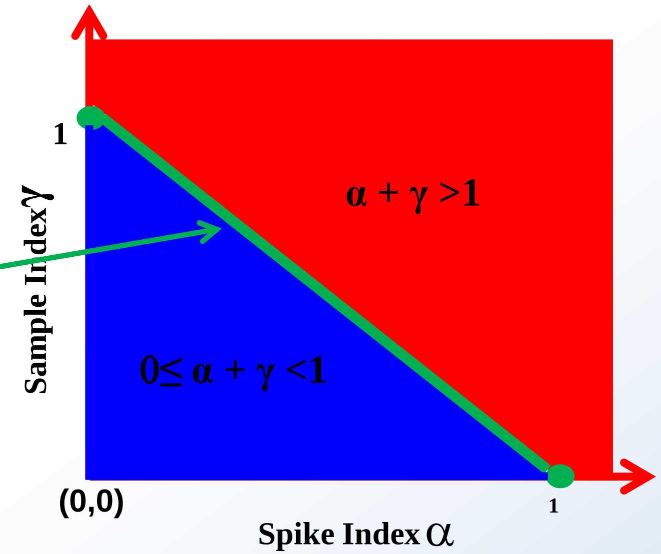


Main Theorem 2

Boundary case for single spike model

Assumption: as $n \to \infty$

- $\lambda_1 >> \lambda_2 = \ldots = \lambda_d = 1$
- $d/(n\lambda_1) \to c$



Result

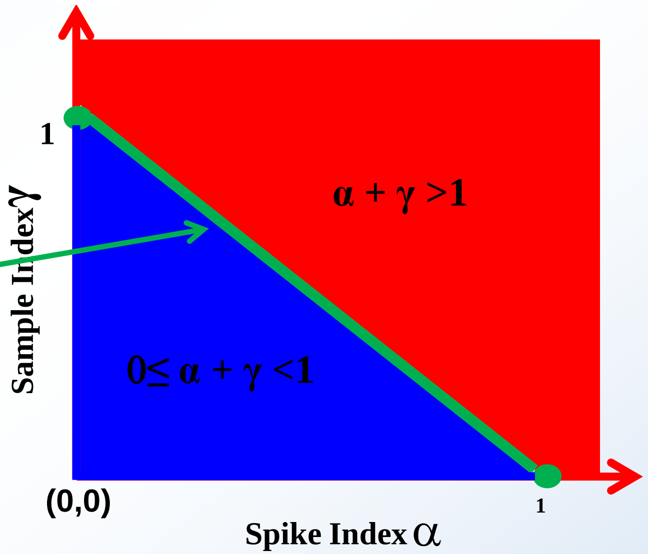
- $\hat{\lambda}_1 / \lambda_1 \xrightarrow{a.s.} 1+c$, and $n\hat{\lambda}_j/d \xrightarrow{a.s.} 1$, $j>1$,

Main Theorem 2

Boundary case for single spike model

Assumption: as $n \to \infty$

- $\lambda_1 >> \lambda_2 = \ldots = \lambda_d = 1$
- $d/(n\lambda_1) \to c$



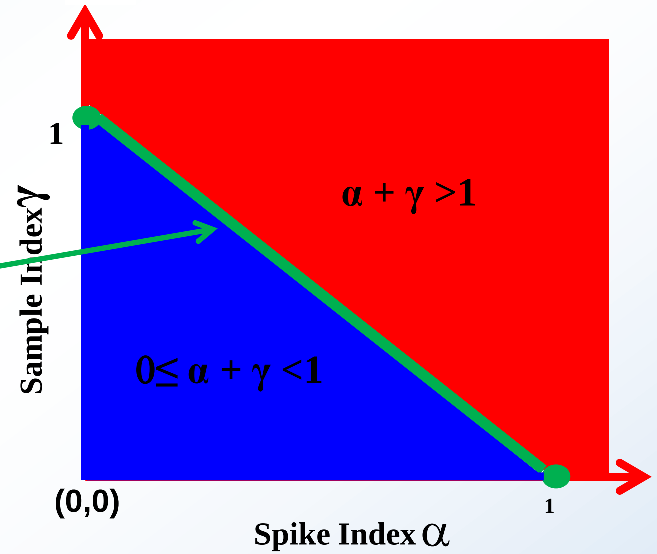
Result

- $\hat{\lambda}_1 / \lambda_1 \xrightarrow{a.s.} 1+c$, and $n\hat{\lambda}_j/d \xrightarrow{a.s.} 1$, $j>1$,
- $|<\hat{u}_1, u_1>| \xrightarrow{a.s.} 1/(1+c)$

Main Theorem 2

Boundary case for single spike model

Assumption: as $n \to \infty$



- $\lambda_1 >> \lambda_2 = \ldots = \lambda_d = 1$
- $d/(n\lambda_1) \to c$

Result

- $\hat{\lambda}_1 / \lambda_1 \xrightarrow{a.s.} 1+c$, and $n\hat{\lambda}_j/d \xrightarrow{a.s.} 1$, $j>1$,
- $|<\hat{u}_1, u_1>| \xrightarrow{a.s.} 1/(1+c)$
- $\hat{u}_j$, $j>1$, are strongly inconsistent with convergence rate $(n/d)^{1/2}$

Outline

- Motivation & Background
- PCA Asymptotics
- Spike Covariance Models
- Theoretical Results of PCA
- **Sparse PCA**
- Summary
- Analysis of Tree Data

Single-Component Spike Model

Recall Example 1:

- $\lambda_1 \sim d^\alpha, \ \lambda_2 = \ldots = \lambda_d = 1$
- Spike index: α

Sparse PCA

Johnstone and Lu (2009)

- PCA strongly inconsistent if and only if $d/n \to \infty$
- But sparse PCA is consistent

Jung and Marron (2009)

- HDLSS: n fixed and $d \to \infty$
- PCA consistent when $\alpha > 1$
- PCA strongly inconsistent when $\alpha < 1$
- Performance of PCA under the sparsity assumption???

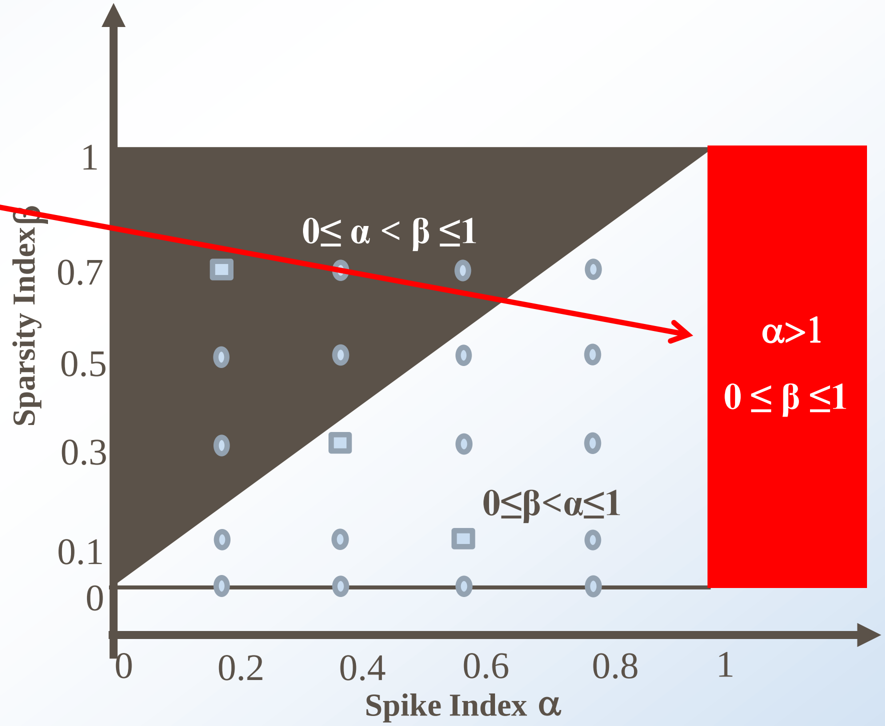
Sparsity Assumption

- $u_1 \sim (\overbrace{1, \ldots, 1}^{[d^\beta]}, 0, \ldots, 0)$
- $[d^\beta]$: the integer part of d^β
- $0 \le \beta \le 1$: sparsity index

Sparse PCA in HDLSS Settings

Conventional PCA

- Consistent when α >1, 0≤β≤1

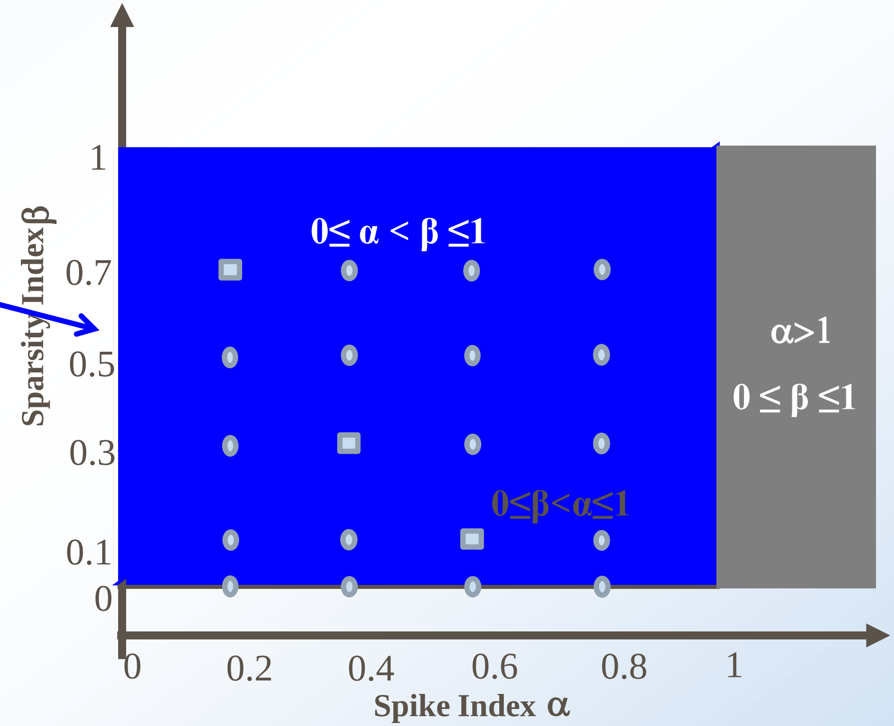


Sparse PCA in HDLSS Settings

Conventional PCA

- Consistent when α >1, 0≤β≤1

- Strongly inconsistent when α <1, 0≤β≤1



Sparse PCA in HDLSS

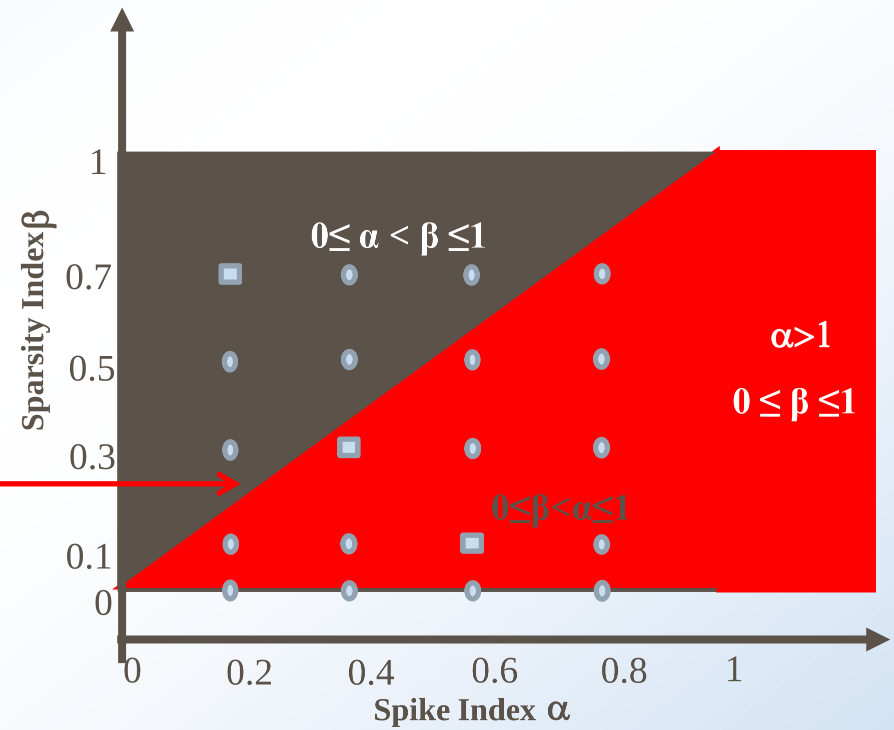
Conventional PCA

- Consistent when α >1, 0≤β≤1

- Strongly inconsistent when α <1, 0≤β≤1

Sparse PCA

- Consistent when 0≤β<α≤1 and α >1, 0≤β≤1



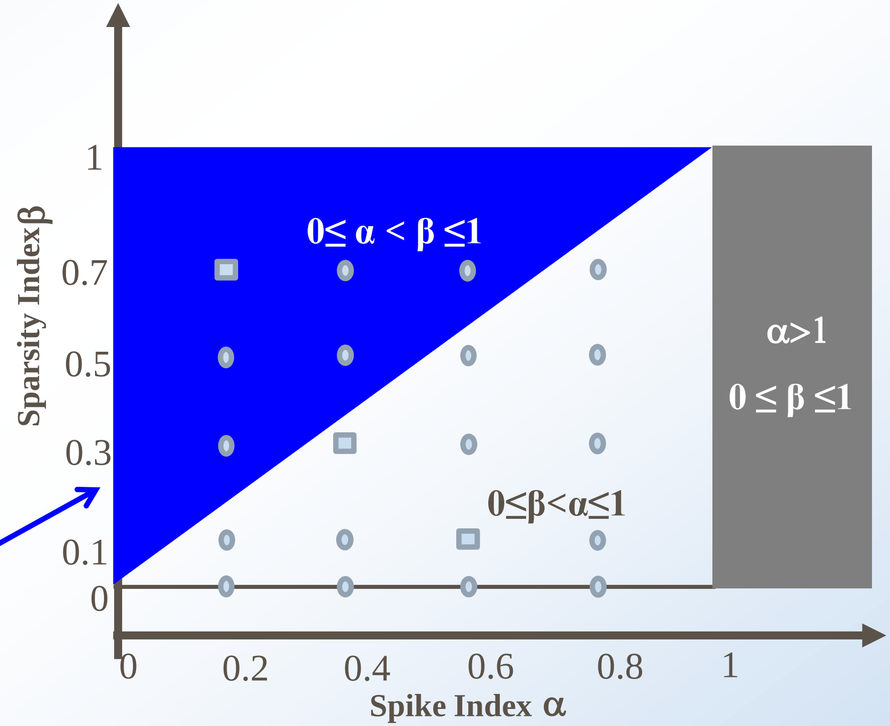
Sparse PCA in HDLSS Settings

Conventional PCA

- Consistent when $\alpha > 1$, $0 \le \beta \le 1$
- Strongly inconsistent when $\alpha < 1$, $0 \le \beta \le 1$

Sparse PCA

- Consistent when $0 \le \beta < \alpha \le 1$ and $\alpha > 1$, $0 \le \beta \le 1$
- Strongly inconsistent when $0 \le \alpha < \beta \le 1$



Sparse PCA in HDLSS Settings

Conventional PCA

- Consistent when α >1, 0≤β≤1

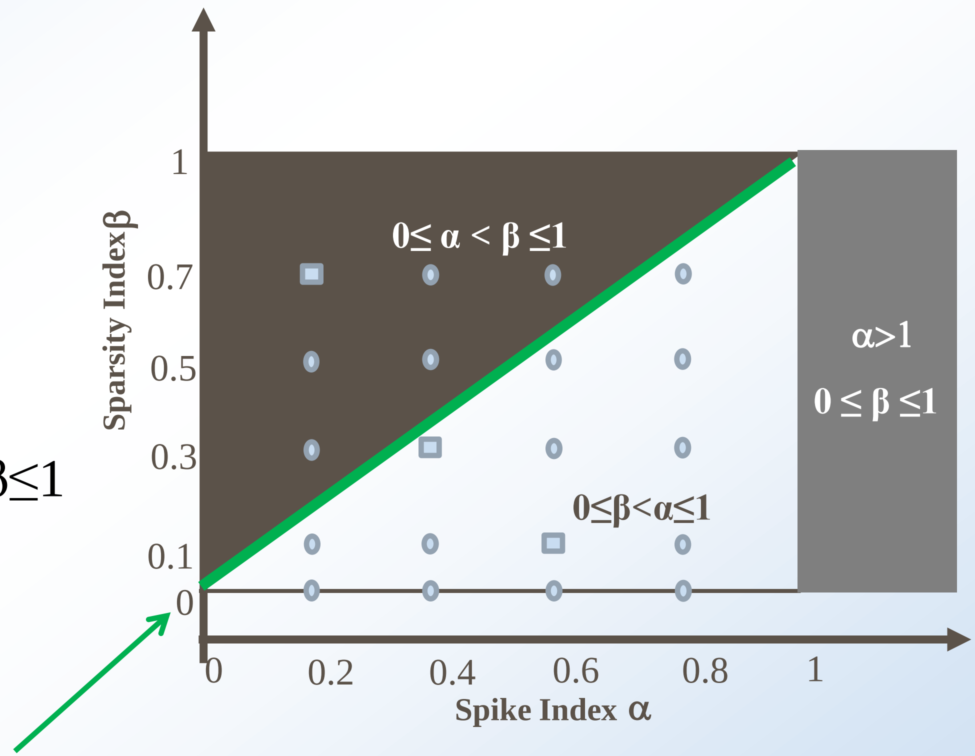
- Strongly inconsistent when α <1, 0≤β≤1

Sparse PCA

- Consistent when 0≤β<α≤1 and α >1, 0≤β≤1

- Strongly inconsistent when 0≤ α < β ≤1

- Marginal inconsistent when 0≤ α = β ≤1



Simulation Studies

- $n=25,\ d=10{,}000$
- $\alpha=0.2, 0.4, 0.6, 0.8;\quad \beta=0, 0.1, 0.3, 0.5, 0.7$
- $\lambda_1=d^\alpha,\quad \lambda_2=\ldots=\lambda_d=1$
- $$u_1 \sim (\overbrace{1, \ldots, 1}^{[d^\beta]}, 0, \ldots, 0)$$
- $2 \le i \le [d^\beta]$, $$u_i \sim (\overbrace{1, \ldots, 1}^{i-1}, -i+1\ 0, \ldots, 0)$$
- $i>[d^\beta]$, $$u_i \sim (\overbrace{0, \ldots, 0}^{i-1}, 1\ 0, \ldots, 0)$$
- Data matrix
- $$X=U_1 d^{\alpha/2} Z_1^T + \sum_{i=2}^{d} U_i Z_i^T, \quad \text{with} \quad Z_i \sim N(0, I_n)$$

Outline

- Motivation & Background
- PCA Asymptotics
- Spike Covariance Models
- Theoretical Results of PCA
- Sparse PCA
- **Summary**
- Analysis of Tree Data

PCA & Sparse PCA

- Build a general framework to study PCA asymptotics
  Shen et al. (2011) (under review)

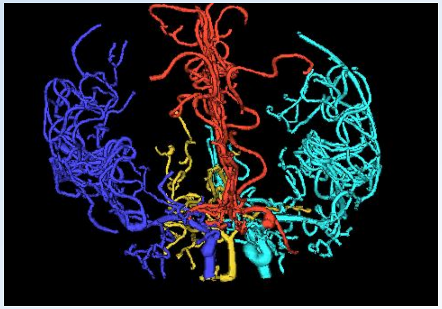
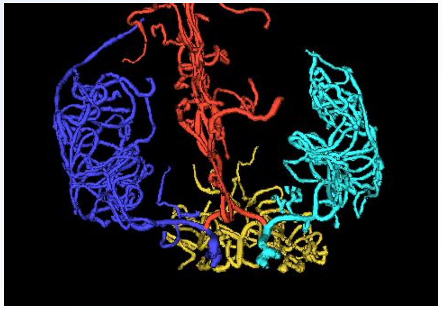
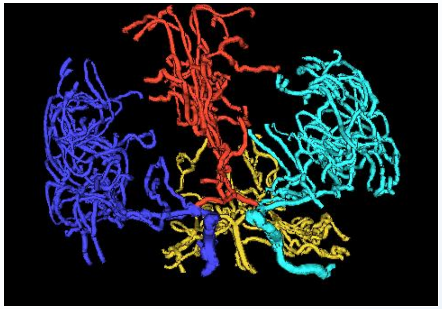
- Introduce sparse PCA asymtptotics in HDLSS
  Shen et al. (2011) (resumbitted)

- Build a general framework to study sparse PCA asymptotics

Outline

- Motivation & Background
- PCA Asymptotics
- Spike Covariance Models
- Theoretical Results of PCA
- Sparse PCA
- Summary
- **Analysis of Tree Data**

Population of Blood Vessel Trees

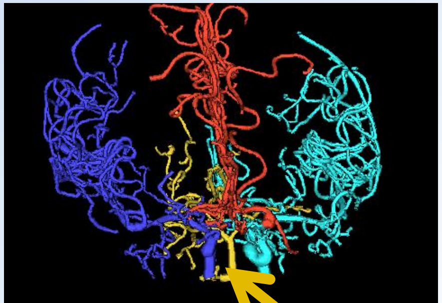
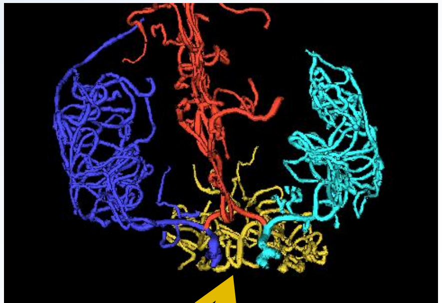
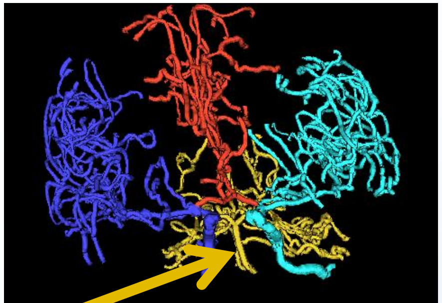
 ,  , … , 

- n=98
- Statistical goals:

  1. Population variation
  2. Age difference
  3. Gender difference
  4. Build model

Population of Blood Vessel Trees

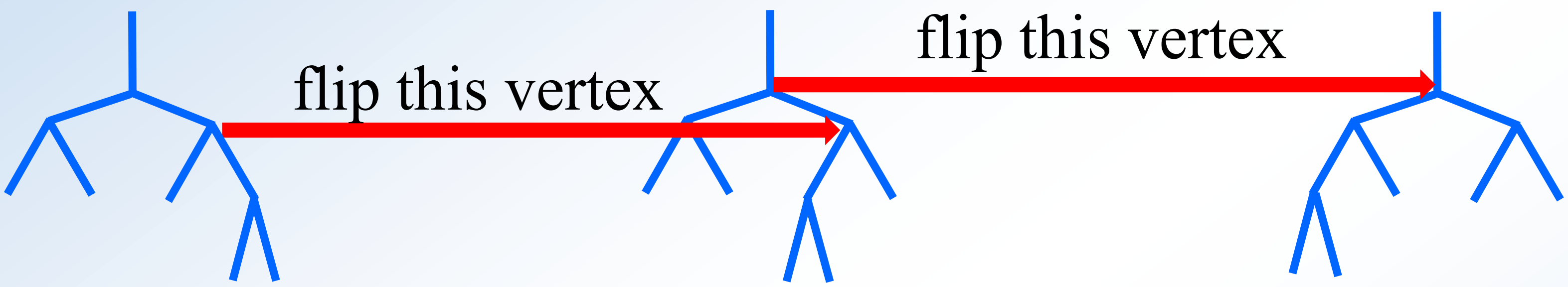
UNC, Stat & OR

 ,  , … , 

- n=98
- Statistical goals:

  1. **Population Variation**
  2. Age difference
  3. Gender difference
  4. Build model

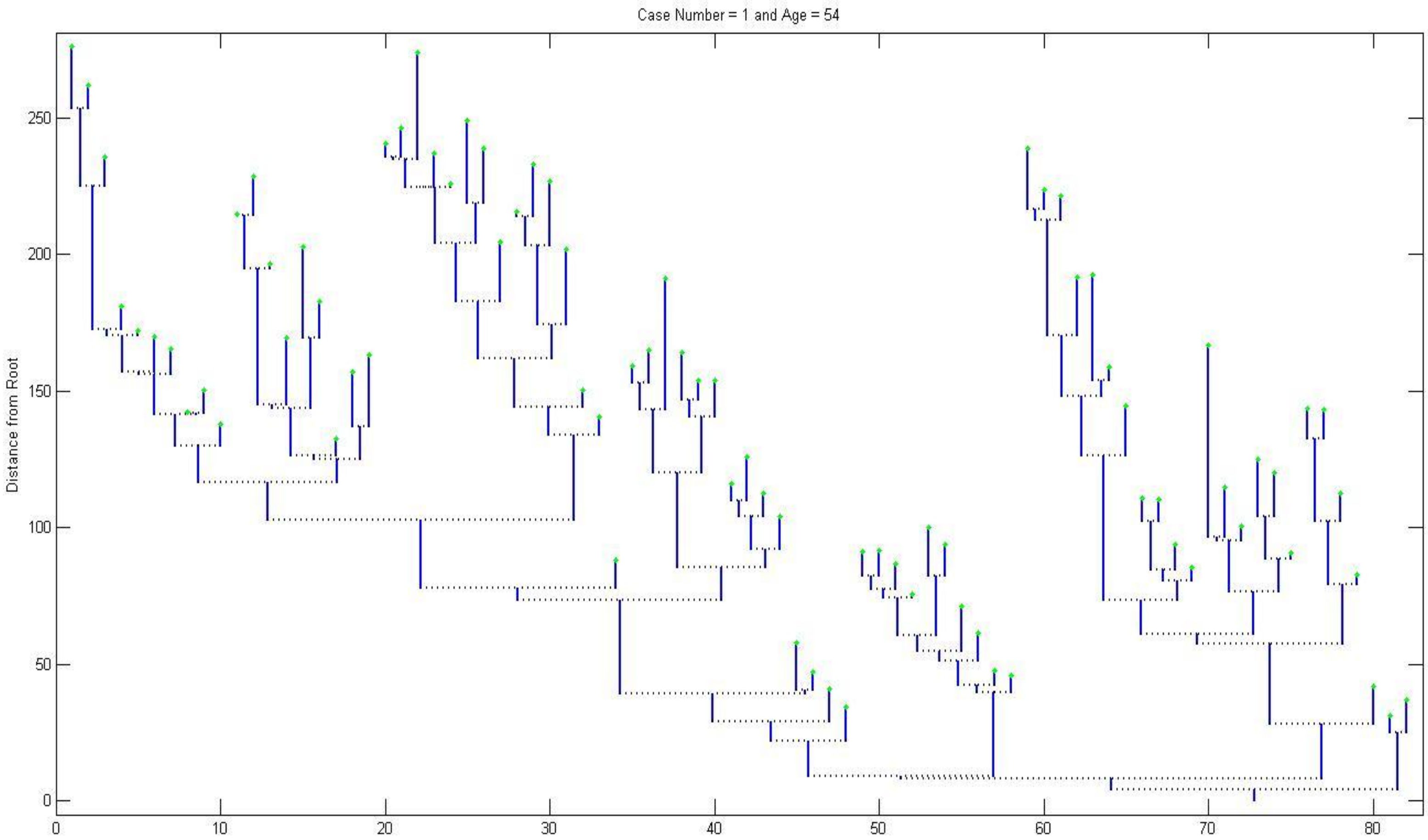
Descendant Correspondence



- Embed 3-d tree in 2-d
- More descendants to the left

Individual Back Tree

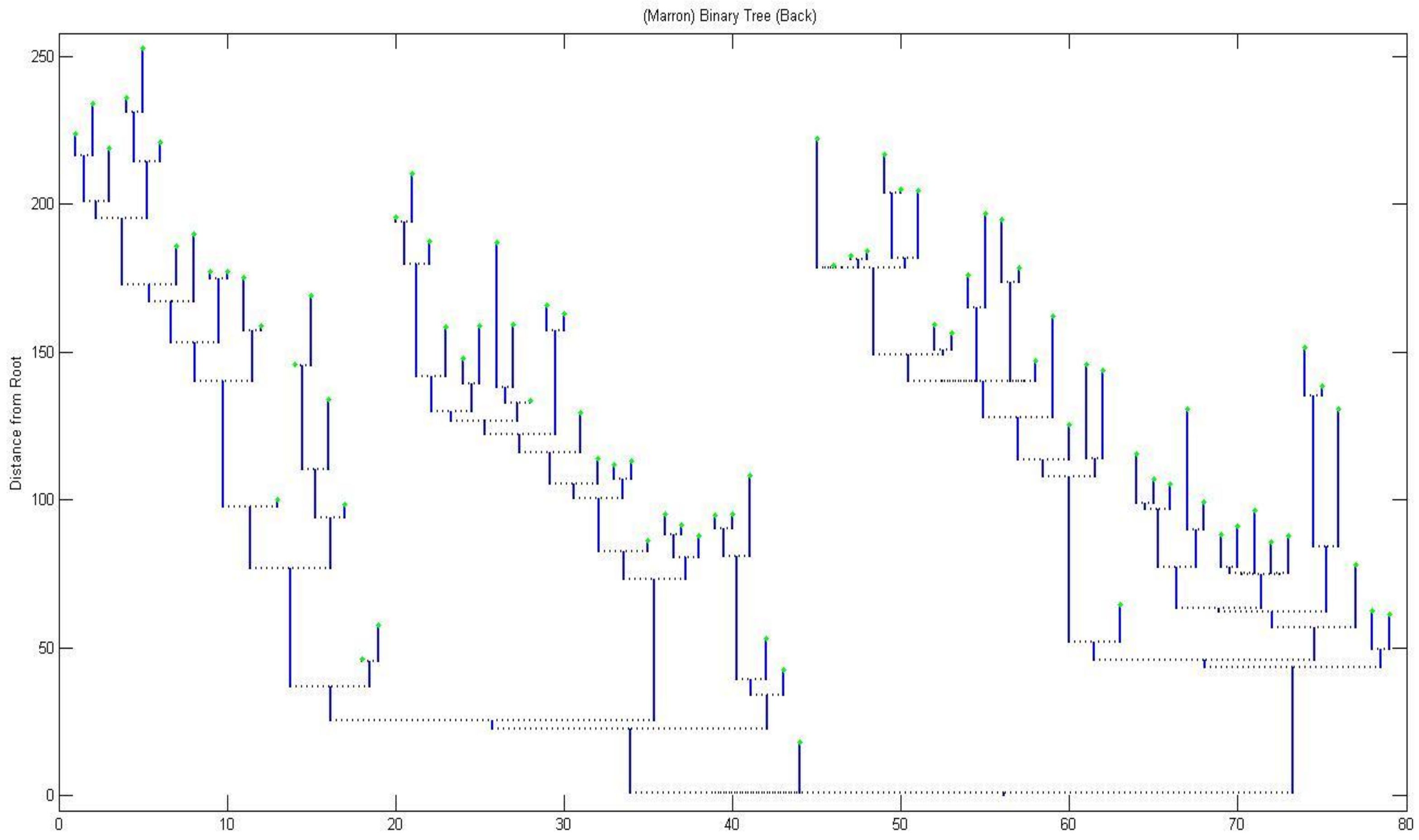
Descendant Correspondence with Branch Length



UNC, Stat & OR

Marron's Back Tree

Descendant Correspondence with Branch Length

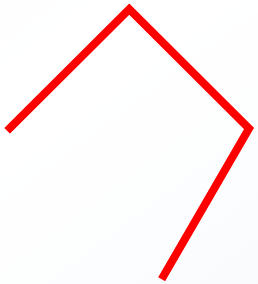


Dyck Path Representation

Example 1, Assume that we have three following trees

Tree 1

Tree 2

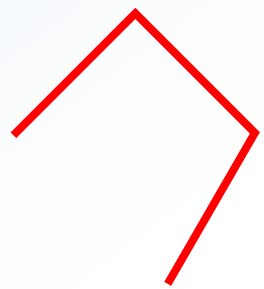


Tree 3

Support Tree: union of trees

Tree 1

Tree 2



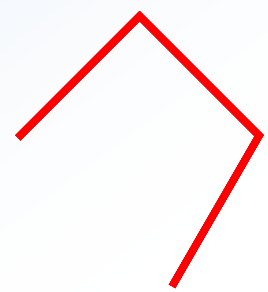
Tree 3

Tree 1

Support Tree: union of trees

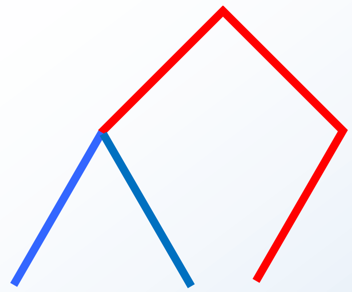
Tree 1

Tree 2



Tree 3

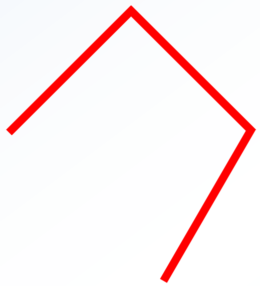
Tree 1,2



Support Tree: union of trees

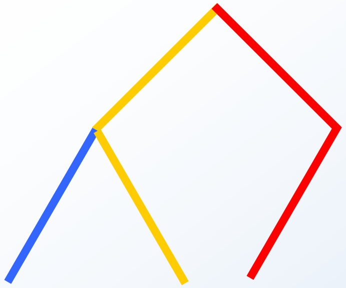
Tree 1

Tree 2



Tree 3

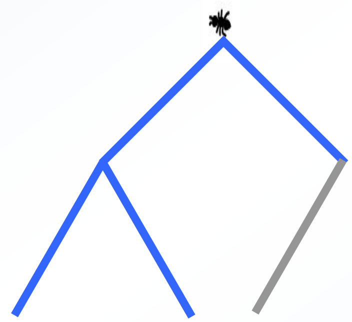
Tree 1,2,3



Dyck Path Representation

Now, we show how to transform the first tree as a curve.

Tree 1/ Support Tree

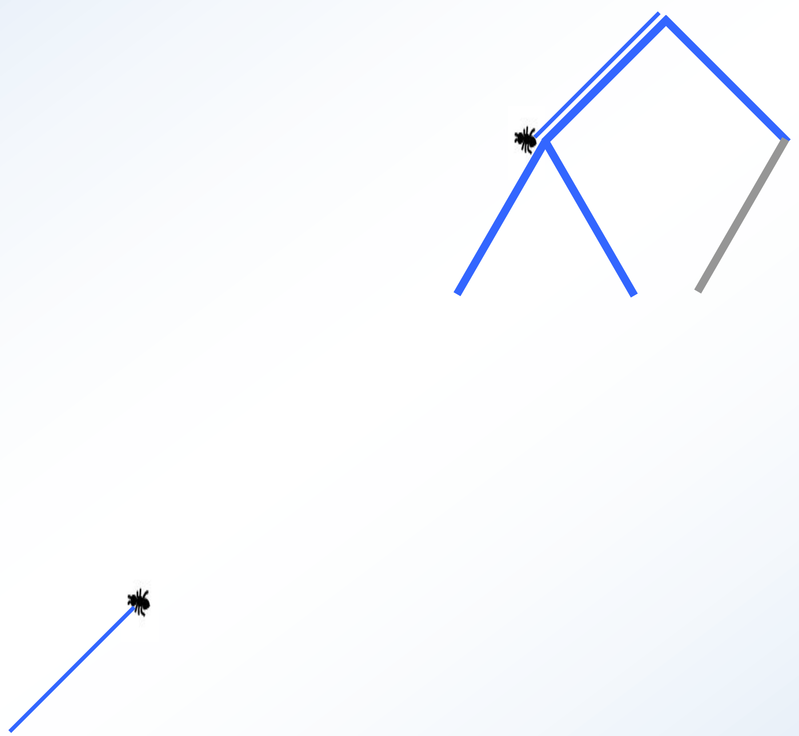


Dyck Path Representation

UNC, Stat & OR

Now, we show how to transform the first tree as a curve.

Tree 1/ Support Tree

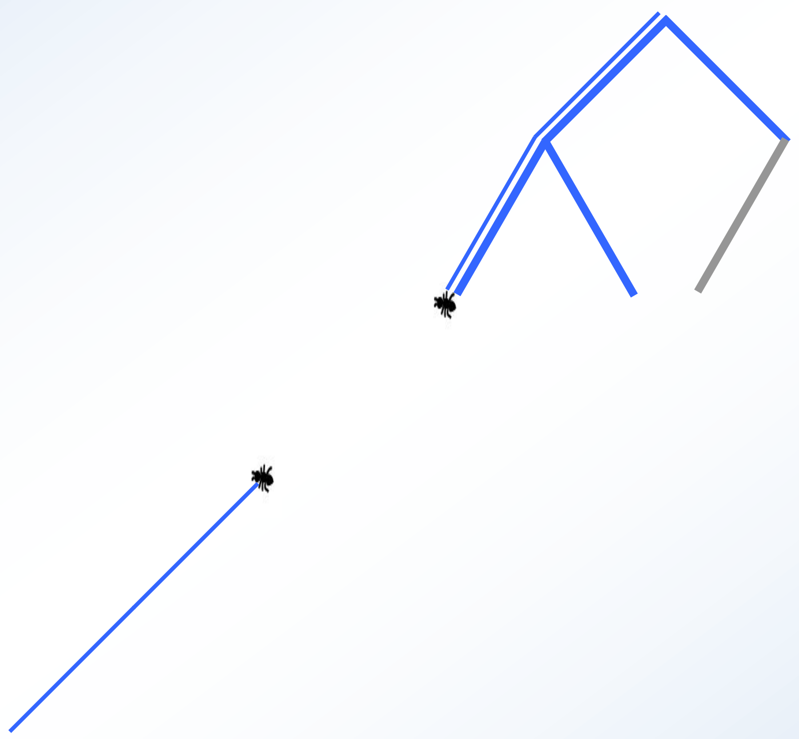


UNC, Stat & OR

Dyck Path Representation

Now, we show how to transform the first tree as a curve.

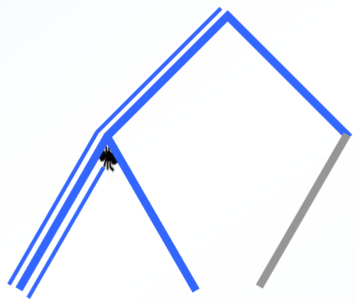
Tree 1/ Support Tree

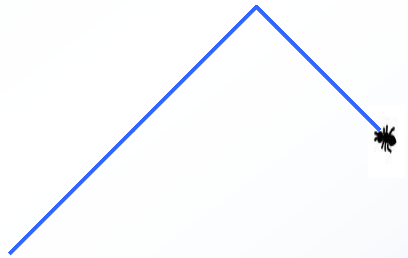


Dyck Path Representation

Now, we show how to transform the first tree as a curve.

Tree 1/ Support Tree

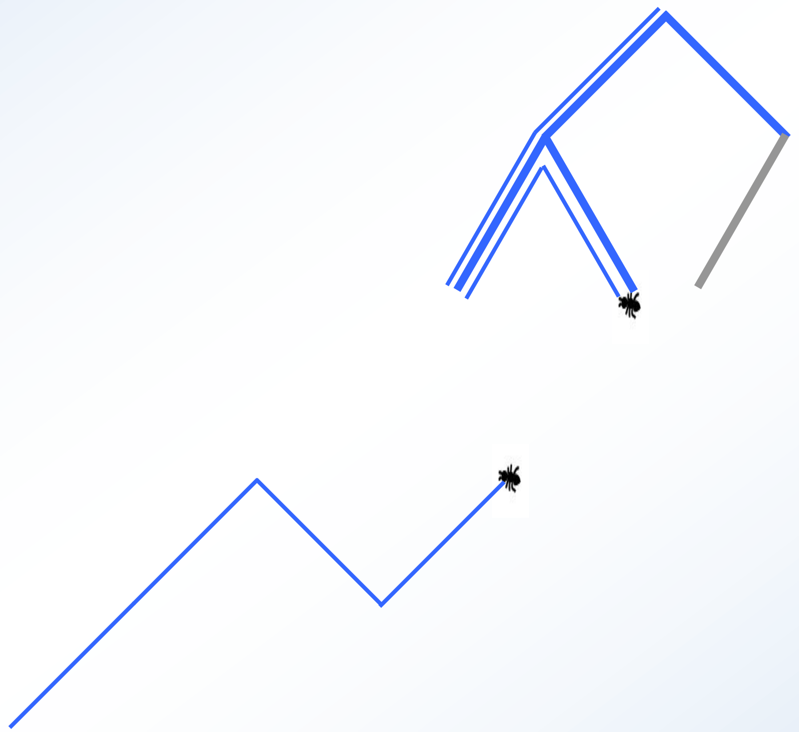




Dyck Path Representation

Now, we show how to transform the first tree as a curve.

Tree 1/ Support Tree

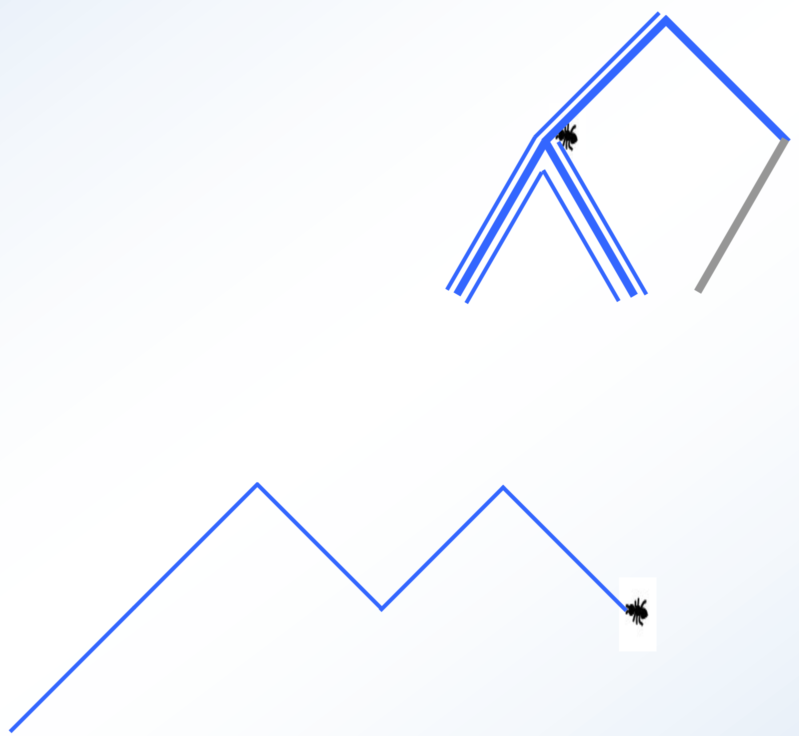


Dyck Path Representation

UNC, Stat & OR

Now, we show how to transform the first tree as a curve.

Tree 1/ Support Tree



Dyck Path Representation

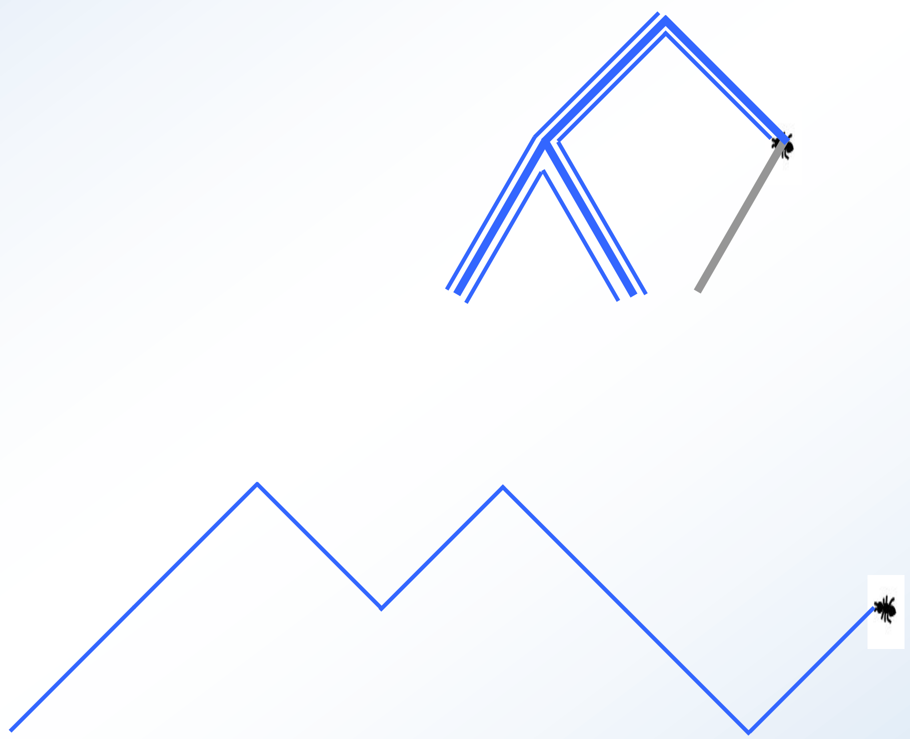
Now, we show how to transform the first tree as a curve.

Tree 1/ Support Tree

Dyck Path Representation

Now, we show how to transform the first tree as a curve.

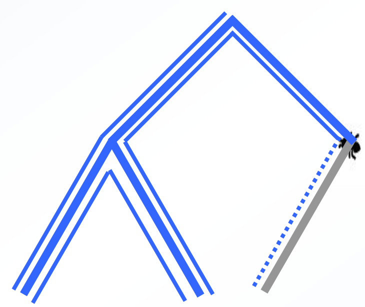
Tree 1/ Support Tree



Dyck Path Representation

Now, we show how to transform the first tree as a curve.

Tree 1/ Support Tree



Dyck Path Representation

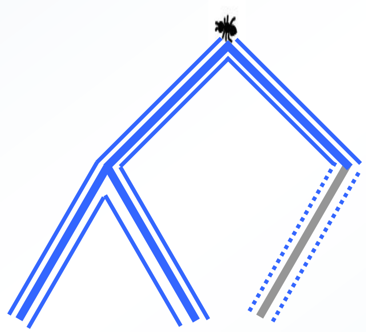
Now, we show how to transform the first tree as a curve.

Tree 1/ Support Tree

Dyck Path Representation

Now, we show how to transform the first tree as a curve.

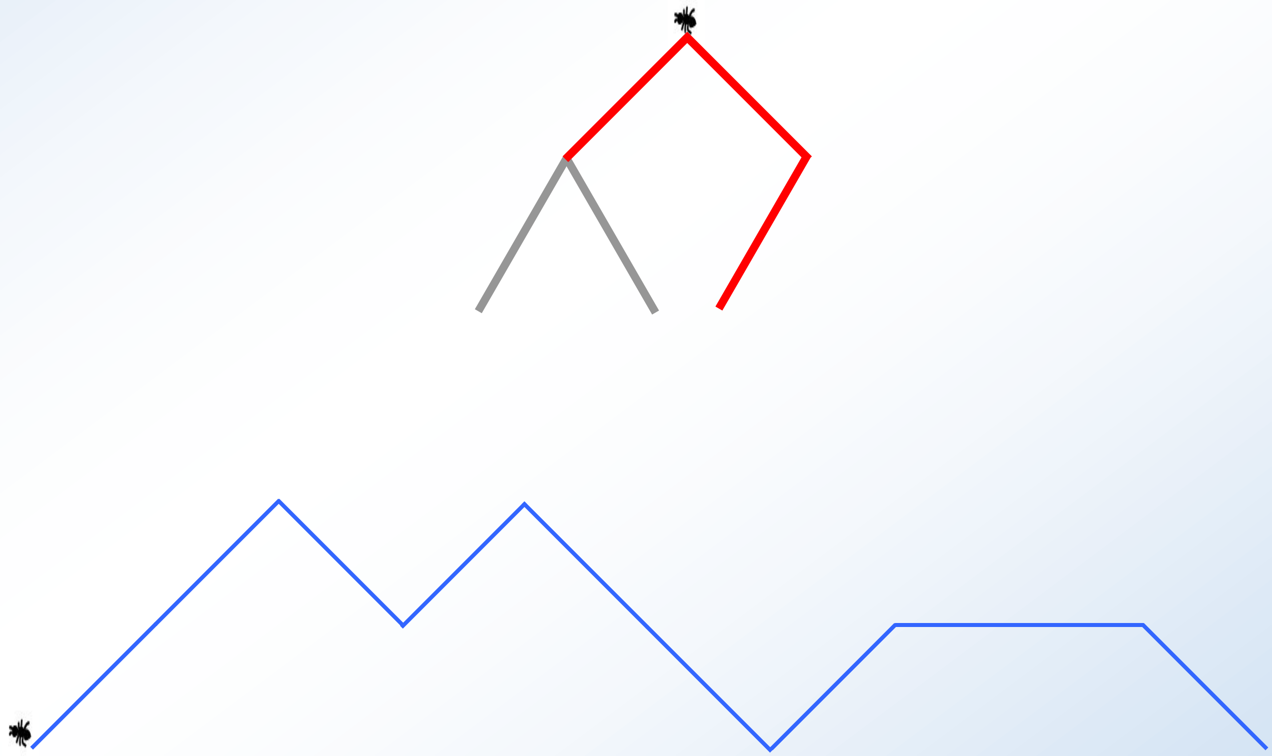
Tree 1/ Support Tree



Dyck Path Representation

Now, we show how to transform the second tree as a curve.

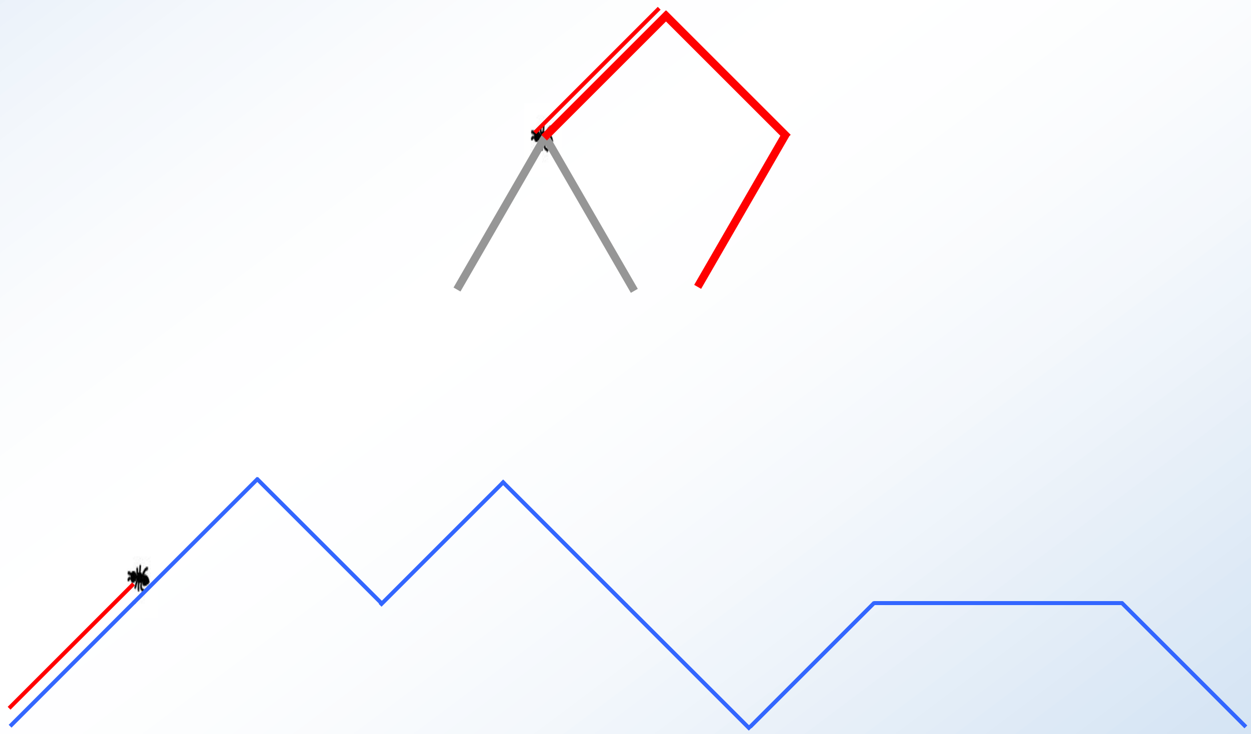
Tree 2/ Support Tree



Dyck Path Representation

Now, we show how to transform the second tree as a curve.

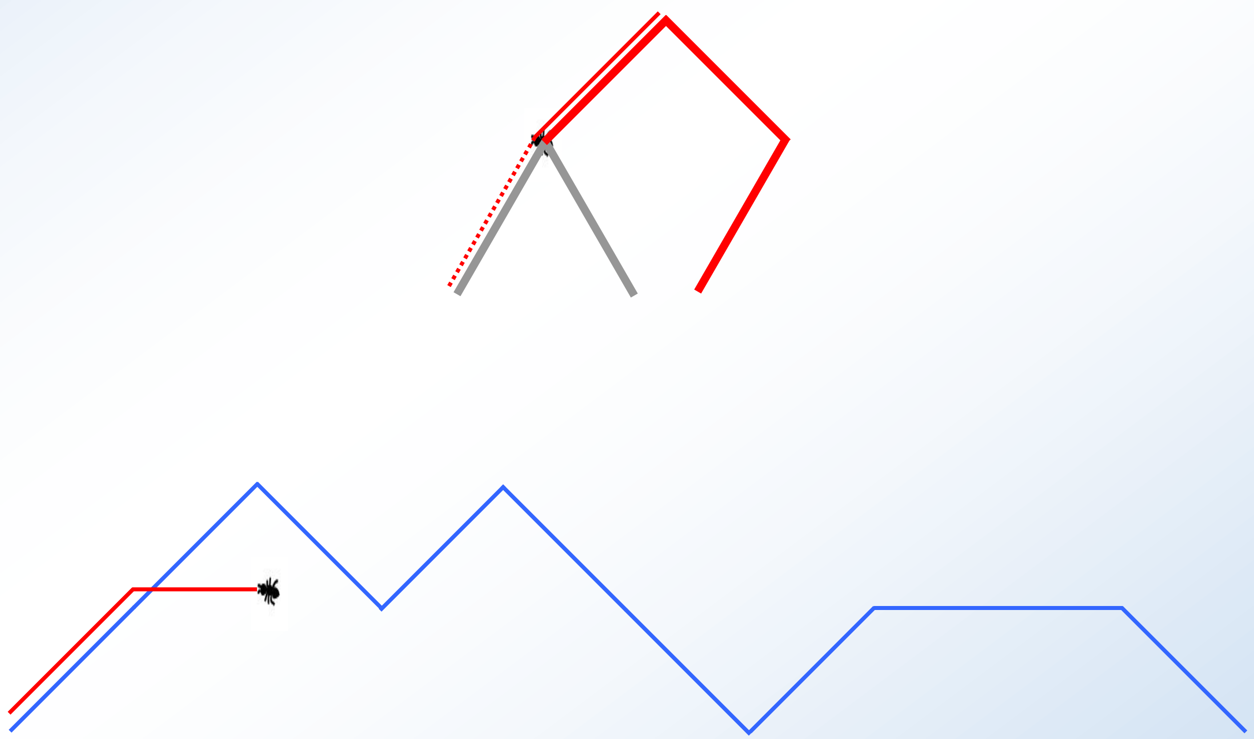
Tree 2/ Support Tree



Dyck Path Representation

Now, we show how to transform the second tree as a curve.

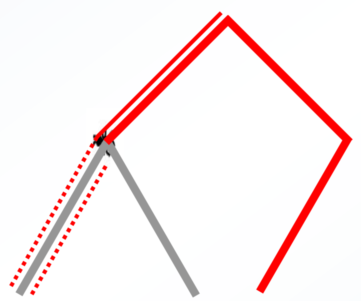
Tree 2/ Support Tree



Dyck Path Representation

Now, we show how to transform the second tree as a curve.

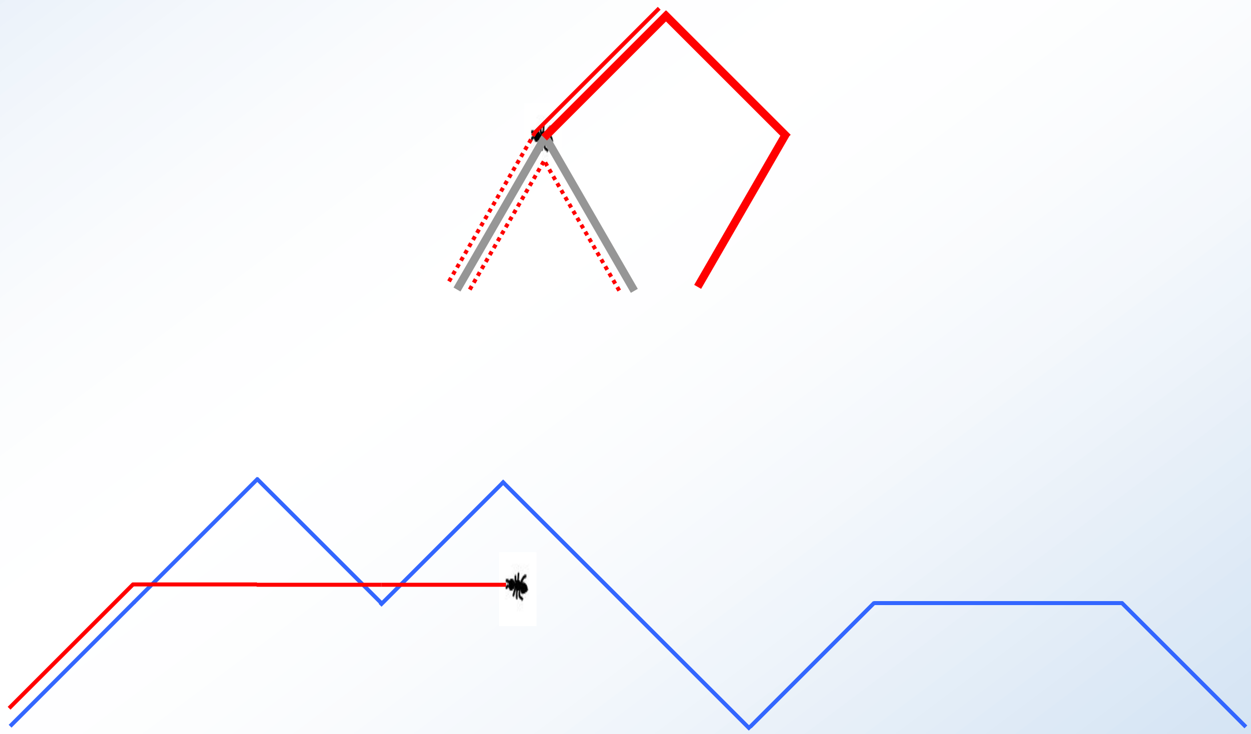
Tree 2/ Support Tree



Dyck Path Representation

Now, we show how to transform the second tree as a curve.

Tree 2/ Support Tree

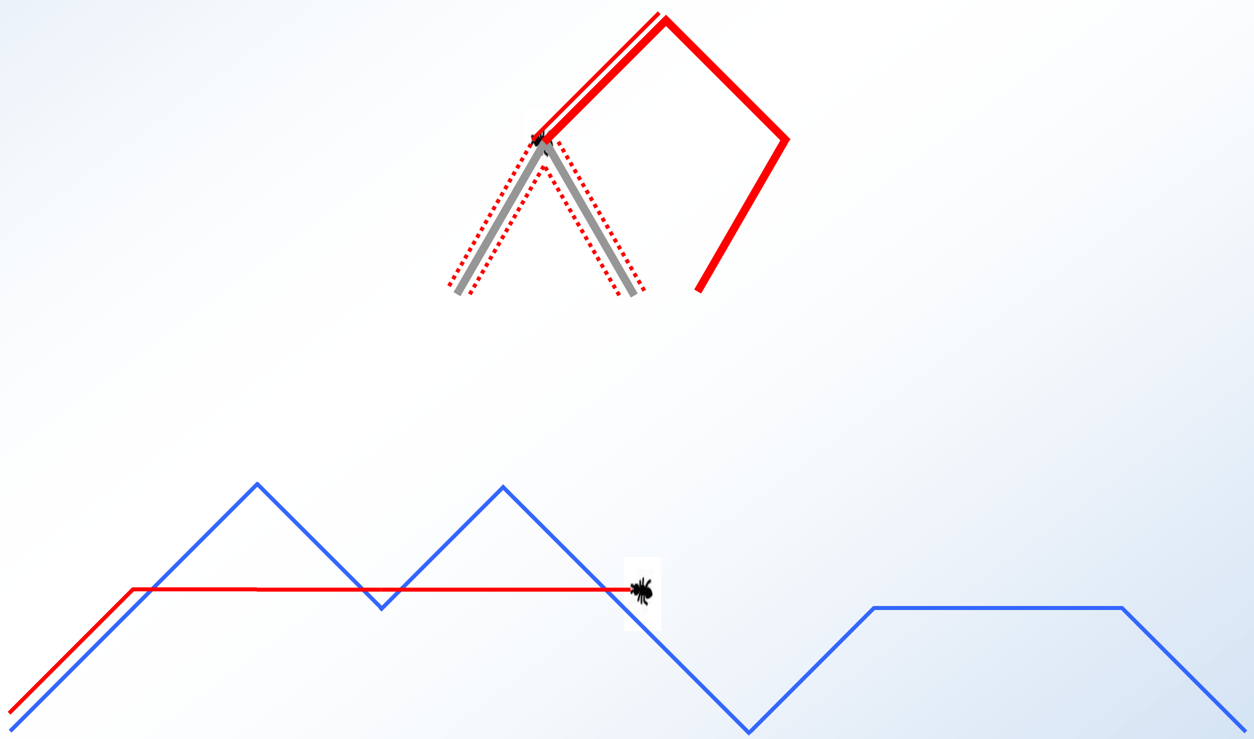


Dyck Path Representation

UNC, Stat & OR

Now, we show how to transform the second tree as a curve.

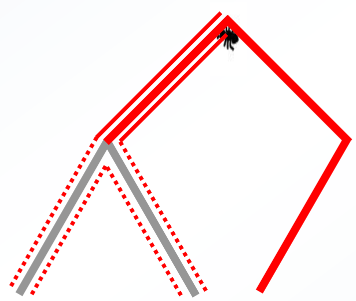
Tree 2/ Support Tree

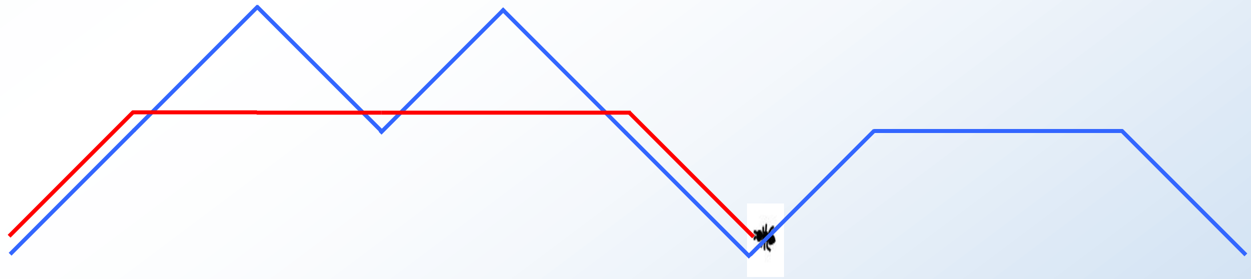


Dyck Path Representation

Now, we show how to transform the second tree as a curve.

Tree 2/ Support Tree

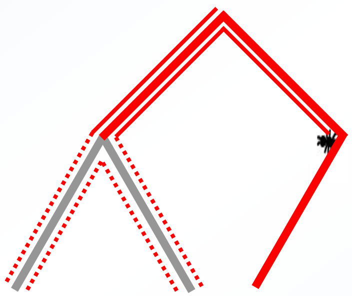




Dyck Path Representation

Now, we show how to transform the second tree as a curve.

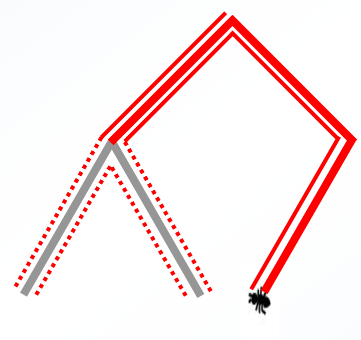
Tree 2/ Support Tree

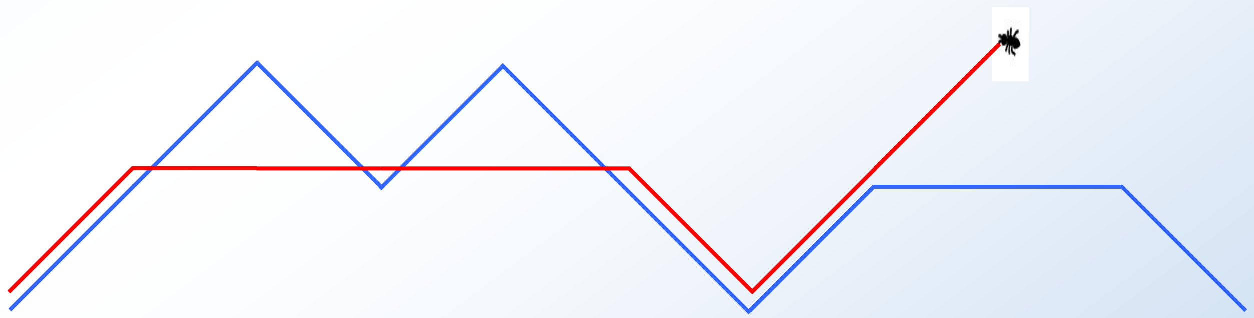


Dyck Path Representation

Now, we show how to transform the second tree as a curve.

Tree 2/ Support Tree

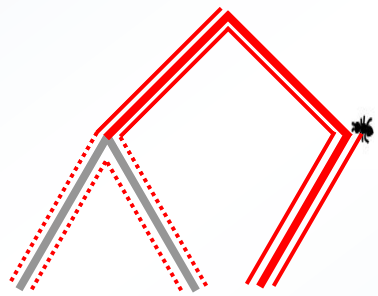


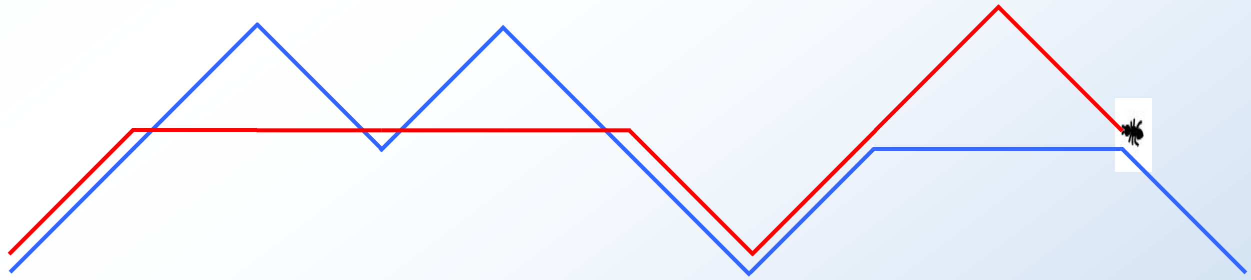


Dyck Path Representation

Now, we show how to transform the second tree as a curve.

Tree 2/ Support Tree

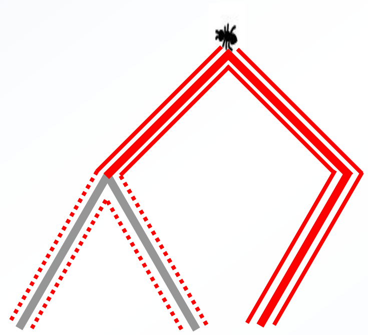




Dyck Path Representation

Now, we show how to transform the second tree as a curve.

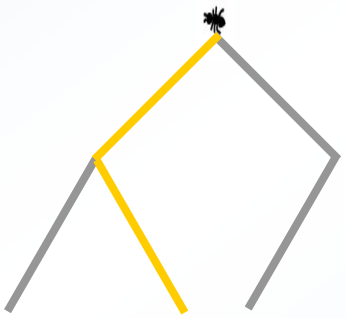
Tree 2/ Support Tree



Dyck Path Representation

Now, we show how to transform the third tree as a curve.

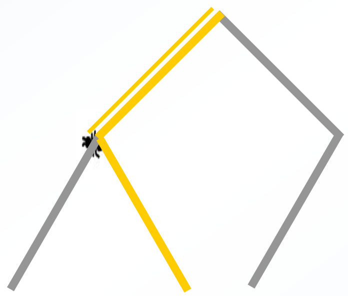
Tree 3/ Support Tree

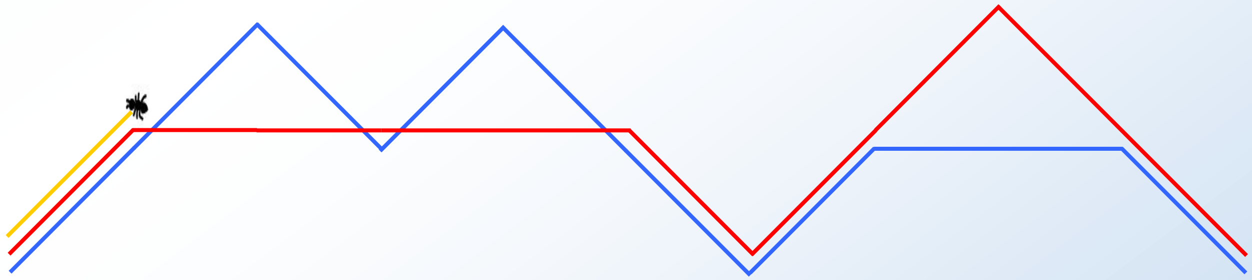


Dyck Path Representation

Now, we show how to transform the third tree as a curve.

Tree 3/ Support Tree

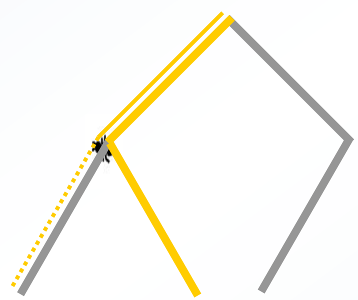


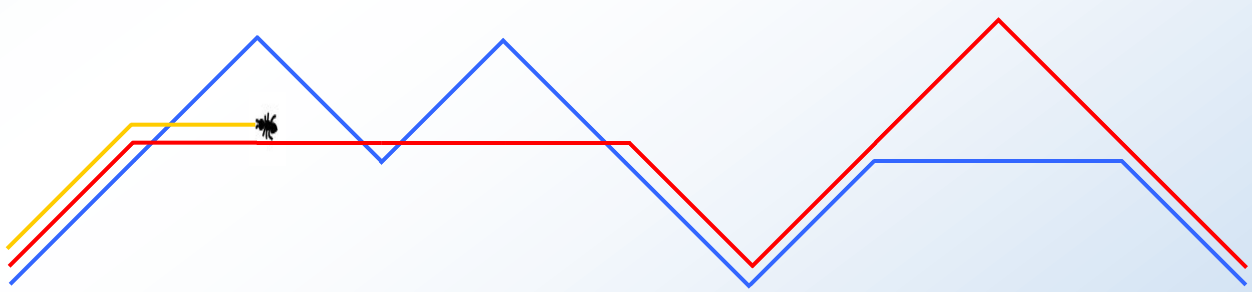


Dyck Path Representation

Now, we show how to transform the third tree as a curve.

Tree 3/ Support Tree

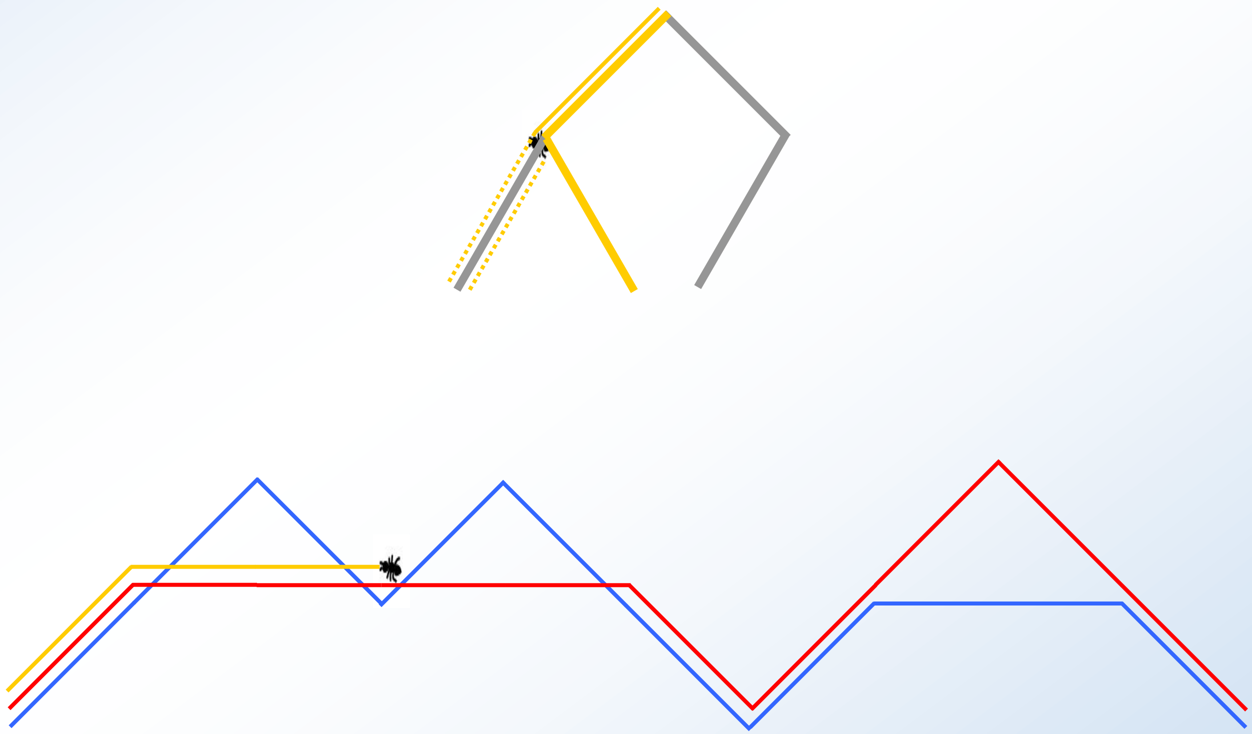




Dyck Path Representation

Now, we show how to transform the third tree as a curve.

Tree 3/ Support Tree

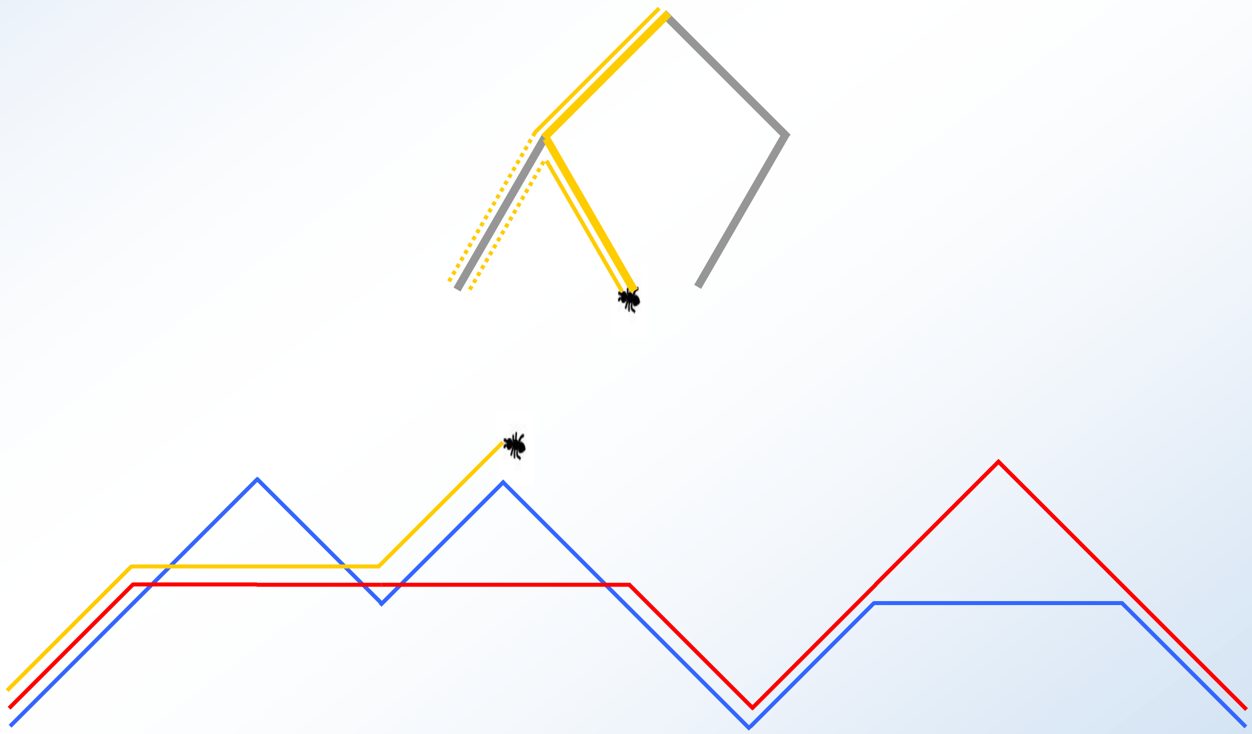


Dyck Path Representation

UNC, Stat & OR

Now, we show how to transform the third tree as a curve.

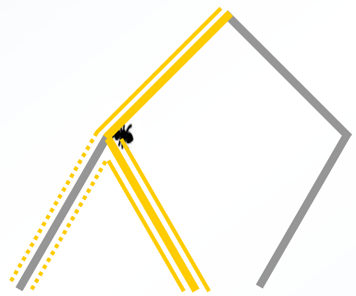
Tree 3/ Support Tree

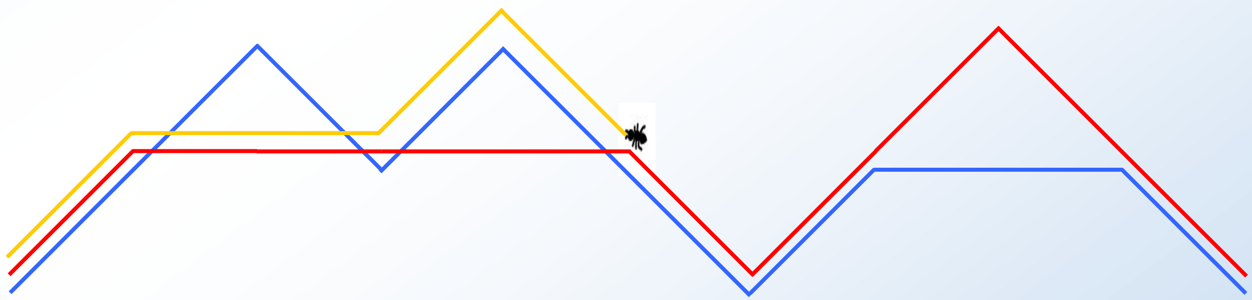


Dyck Path Representation

Now, we show how to transform the third tree as a curve.

Tree 3/ Support Tree

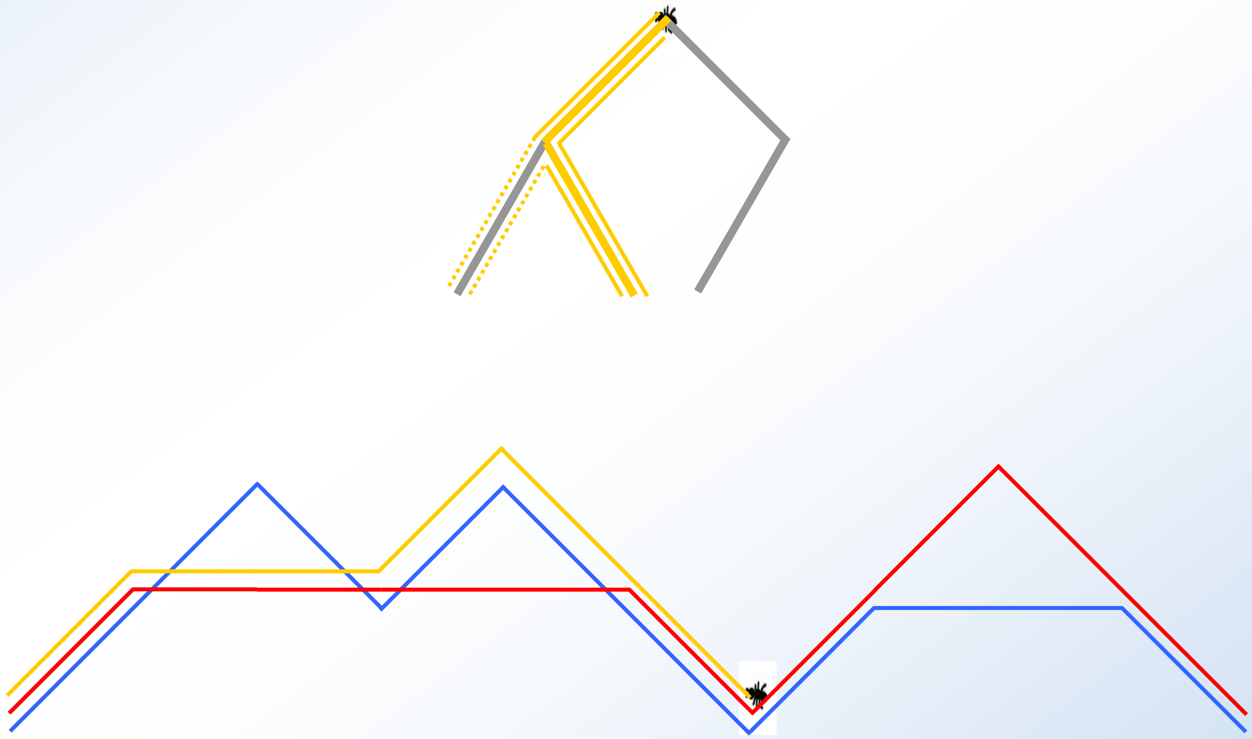




Dyck Path Representation

Now, we show how to transform the third tree as a curve.

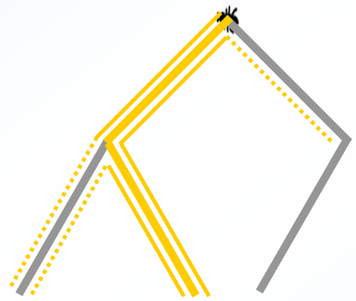
Tree 3/ Support Tree

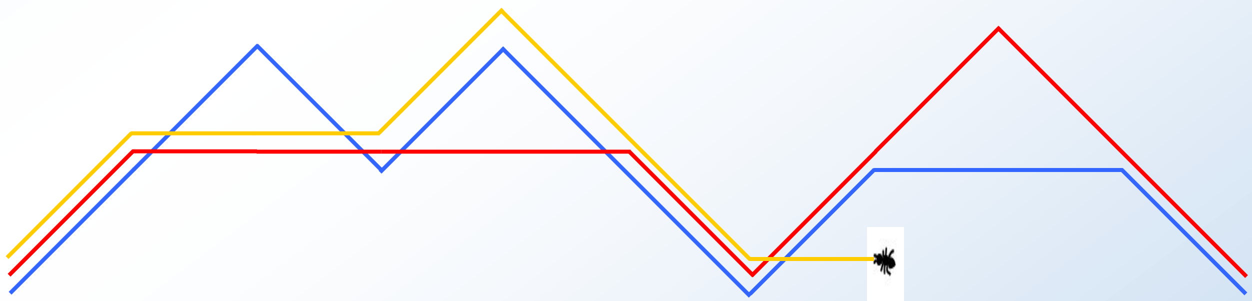


Dyck Path Representation

Now, we show how to transform the third tree as a curve.

Tree 3/ Support Tree

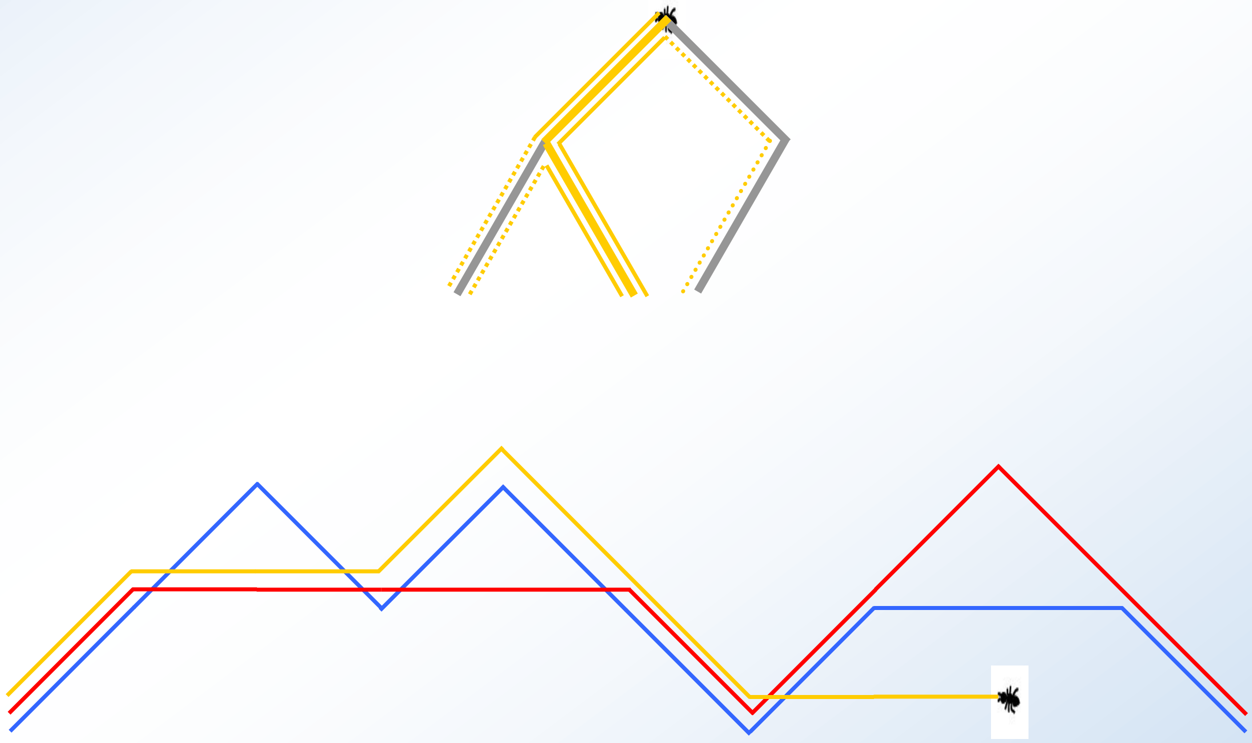




Dyck Path Representation

Now, we show how to transform the third tree as a curve.

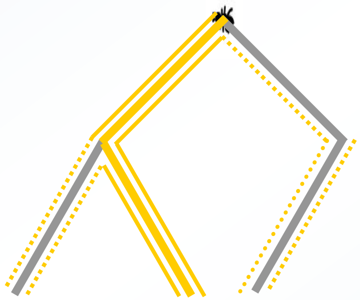
Tree 3/ Support Tree



Dyck Path Representation

Now, we show how to transform the third tree as a curve.

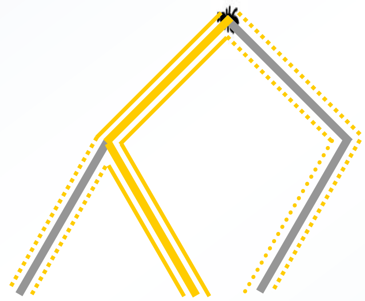
Tree 3/ Support Tree

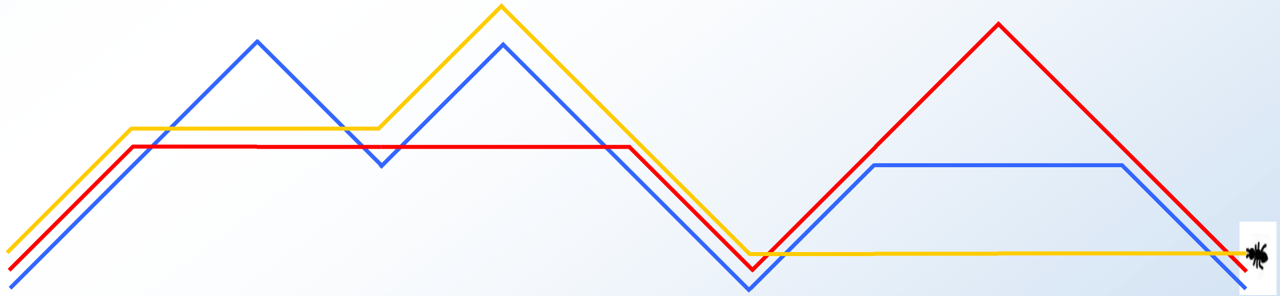


Dyck Path Representation

Now, we show how to transform the third tree as a curve.

Tree 3/ Support Tree





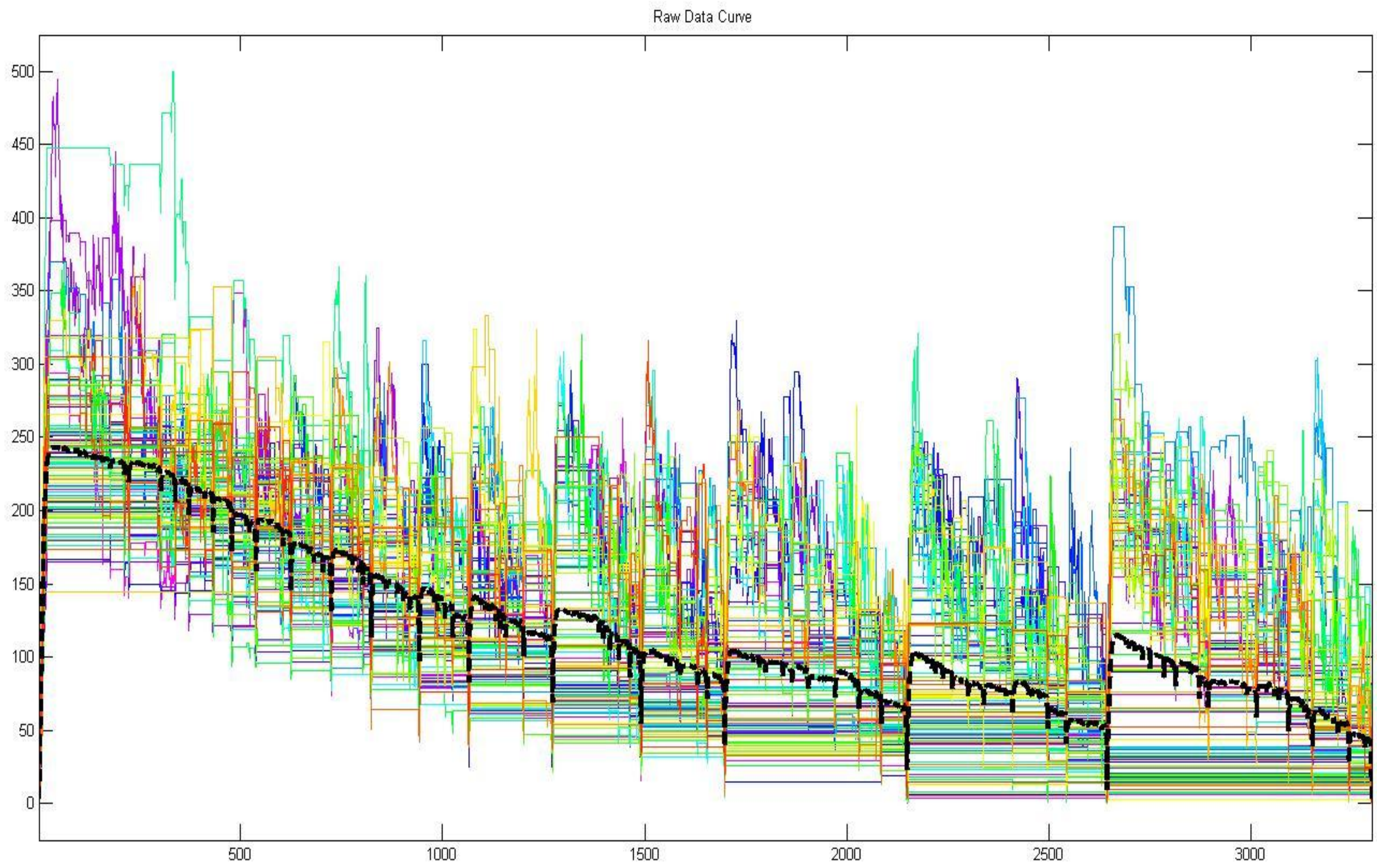
Dyck Path Representation

The Dyck Path:

- The curve connecting the coordinate points (x, y)
- X-value: the number of steps that the ant passed
- Y-value: the corresponding branch height

Dyck Path Curves (Back Tree)

UNC, Stat & OR



Dyck Path Curves

Properties:

- Flat curve segments correspond to missing branches

- Rainbow color corresponds to age
  ranging from magenta (for young) to red (for old)

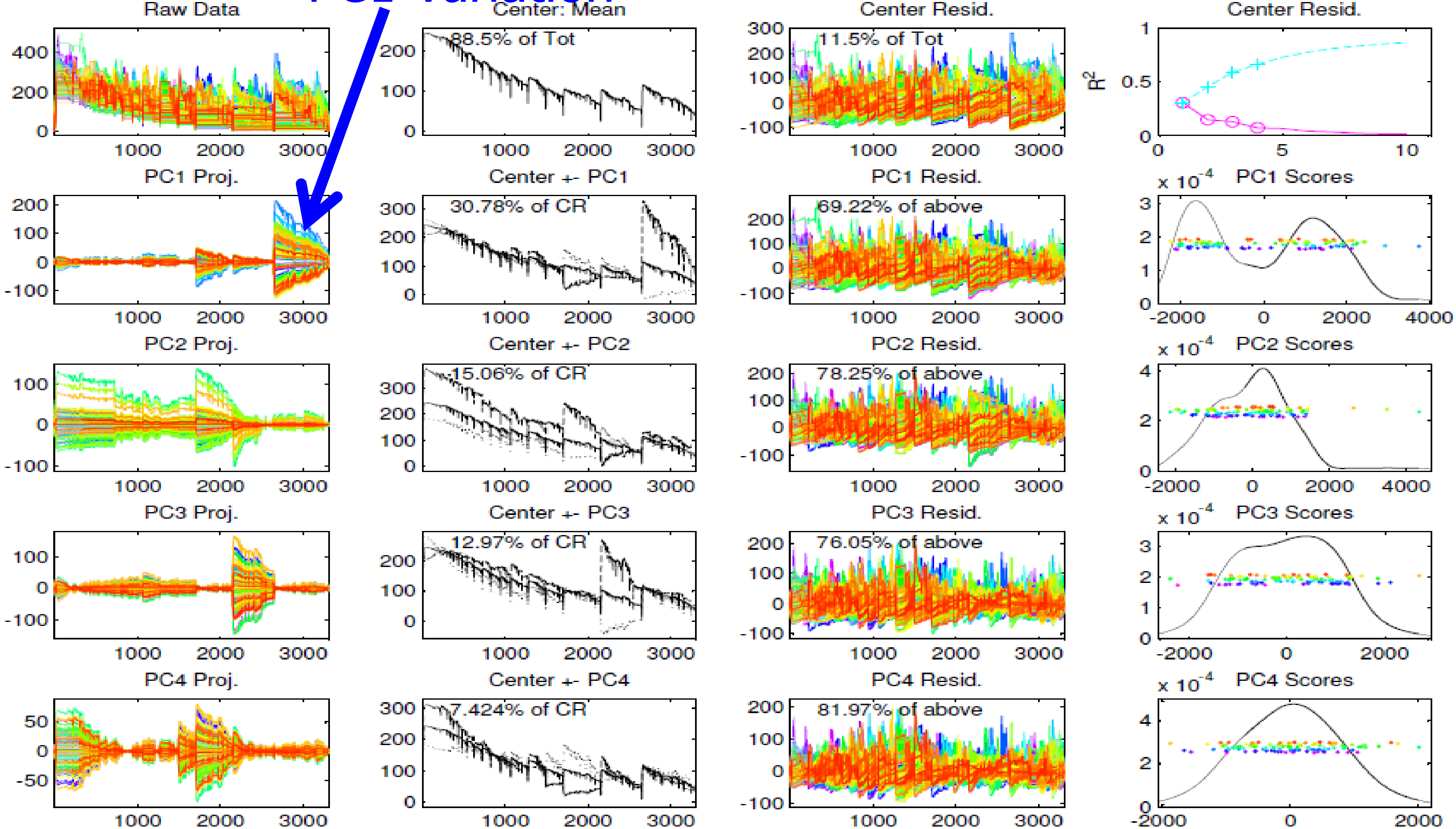
- The left part is taller than the right part
  the descendant correspondence

- The range of x-value is twice of the branch number
  every branch is passed twice - Dyck Path

PCA of the Dyck Path Curves (Back Tree)

UNC, Stat & OR

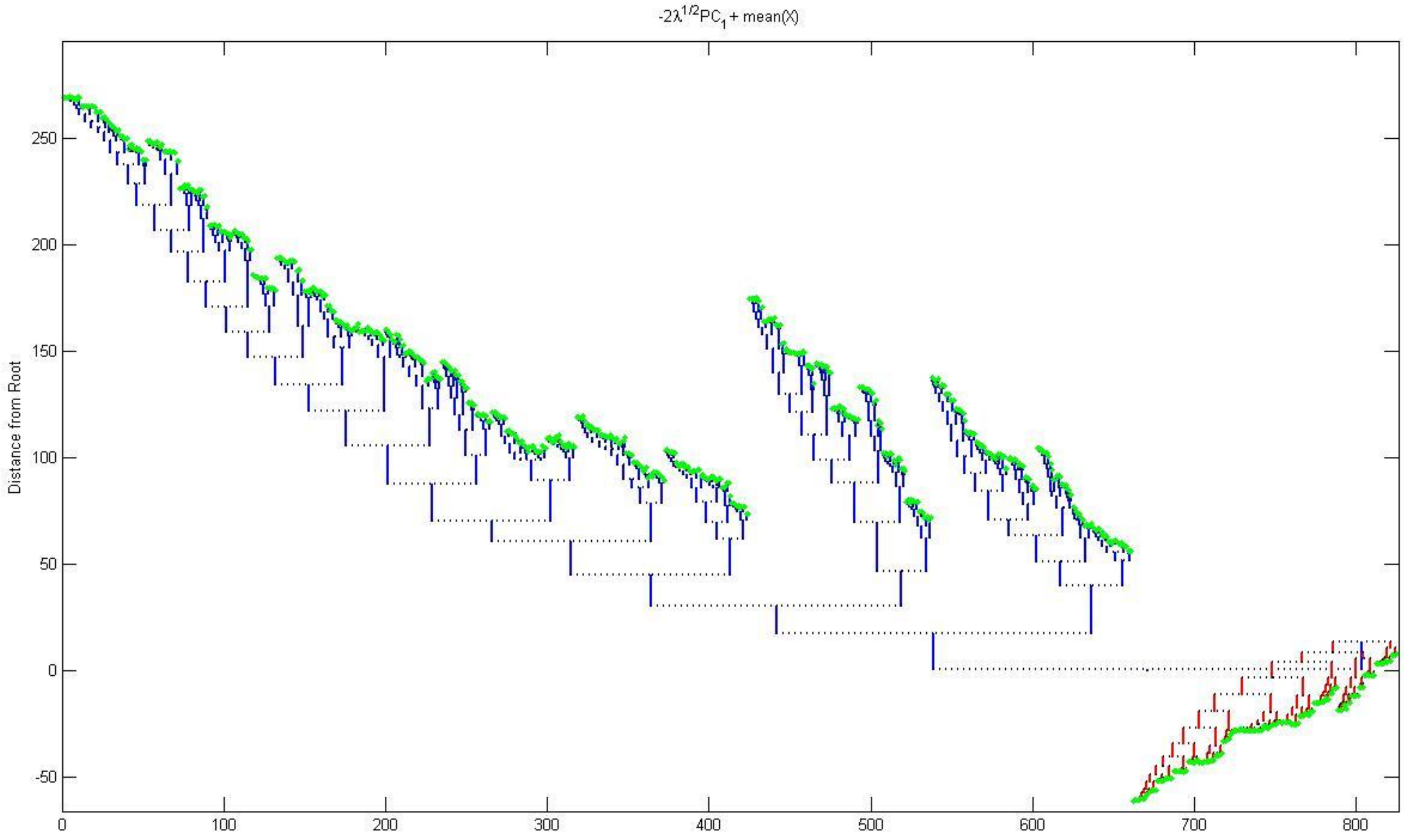
PC1 Variation



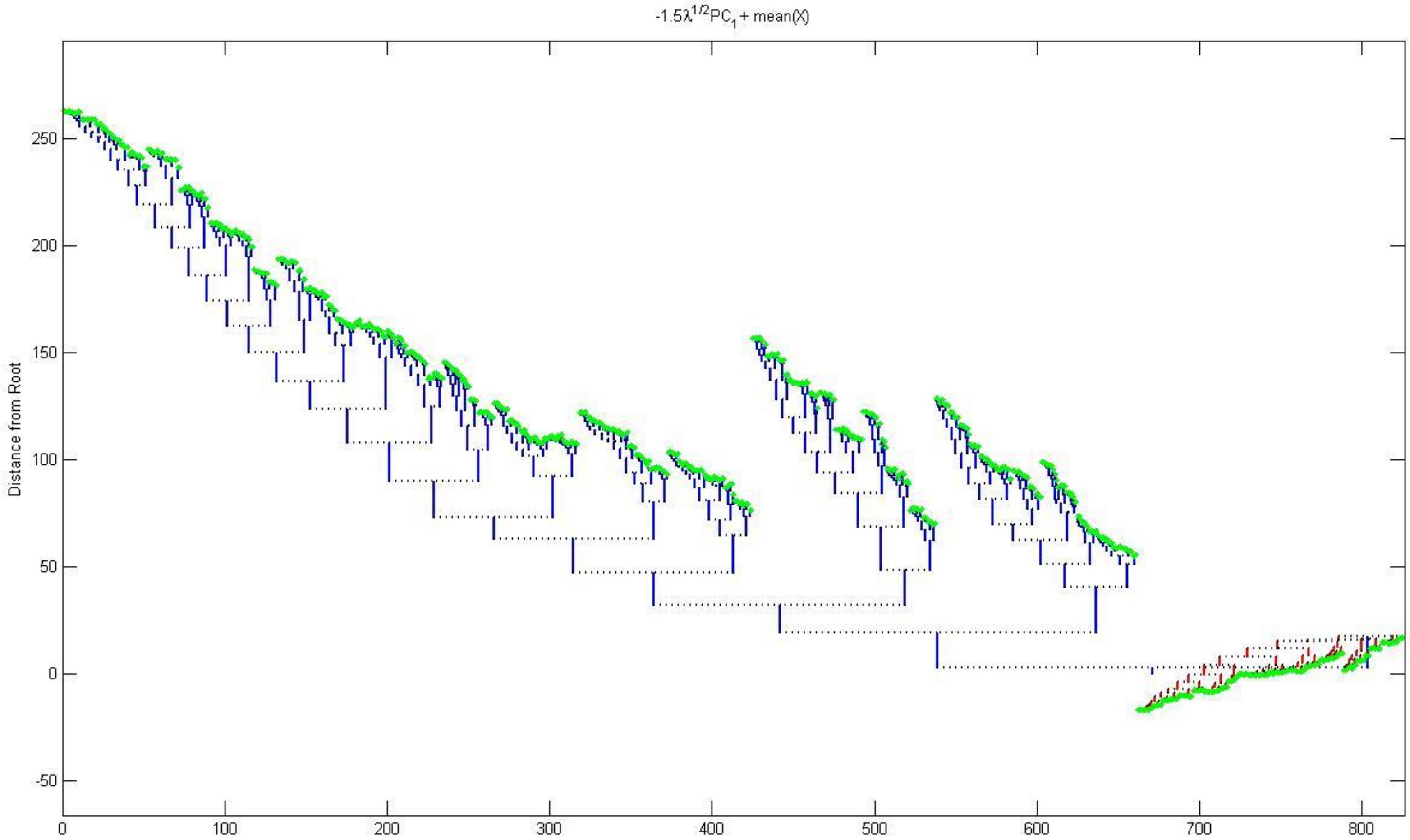
Tree interpretation of the PC direction

PC1 Direction (Back Tree)

UNC, Stat & OR

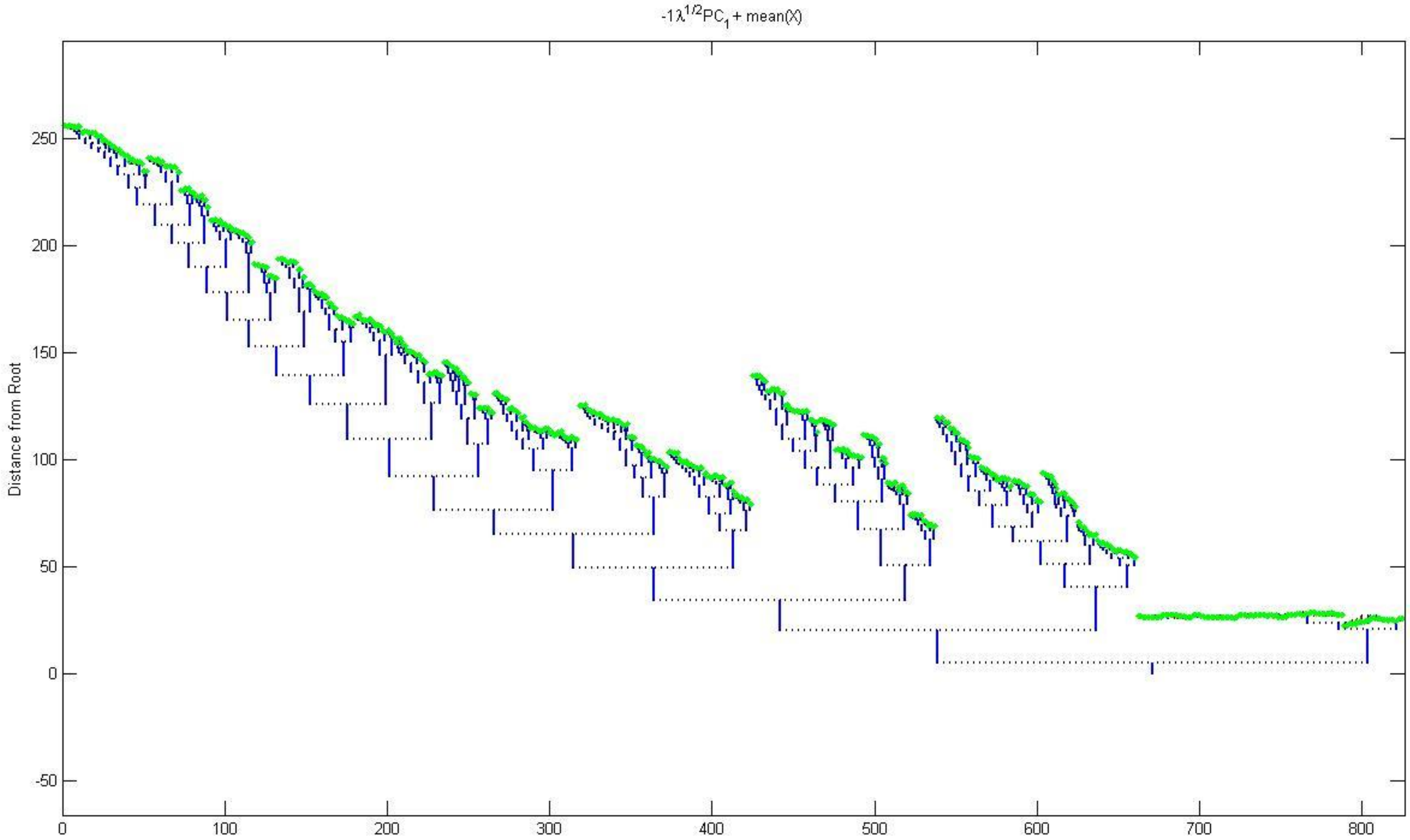


PC1 Direction (Back Tree)



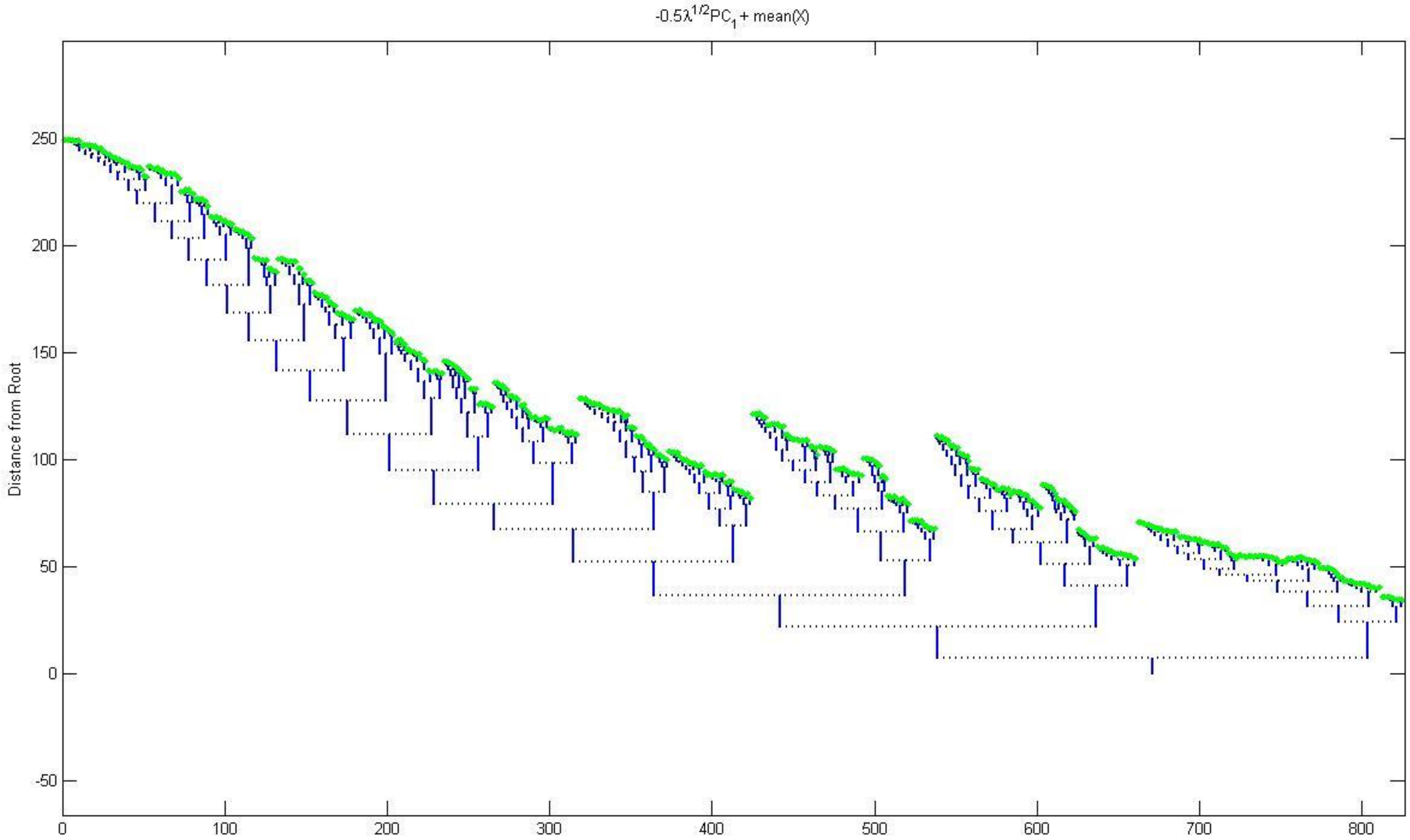
PC1 Direction (Back Tree)

UNC, Stat & OR



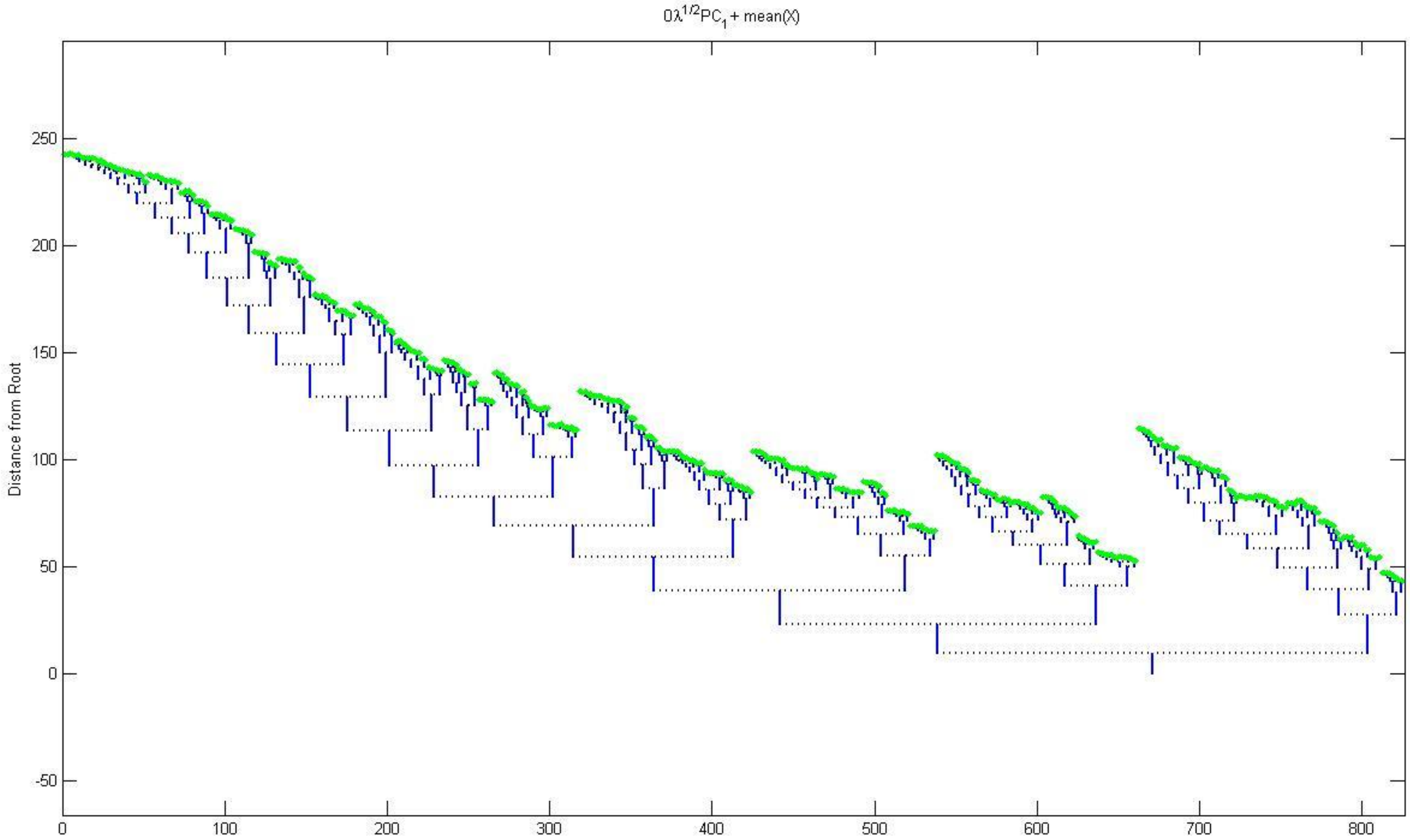
PC1 Direction (Back Tree)

UNC, Stat & OR

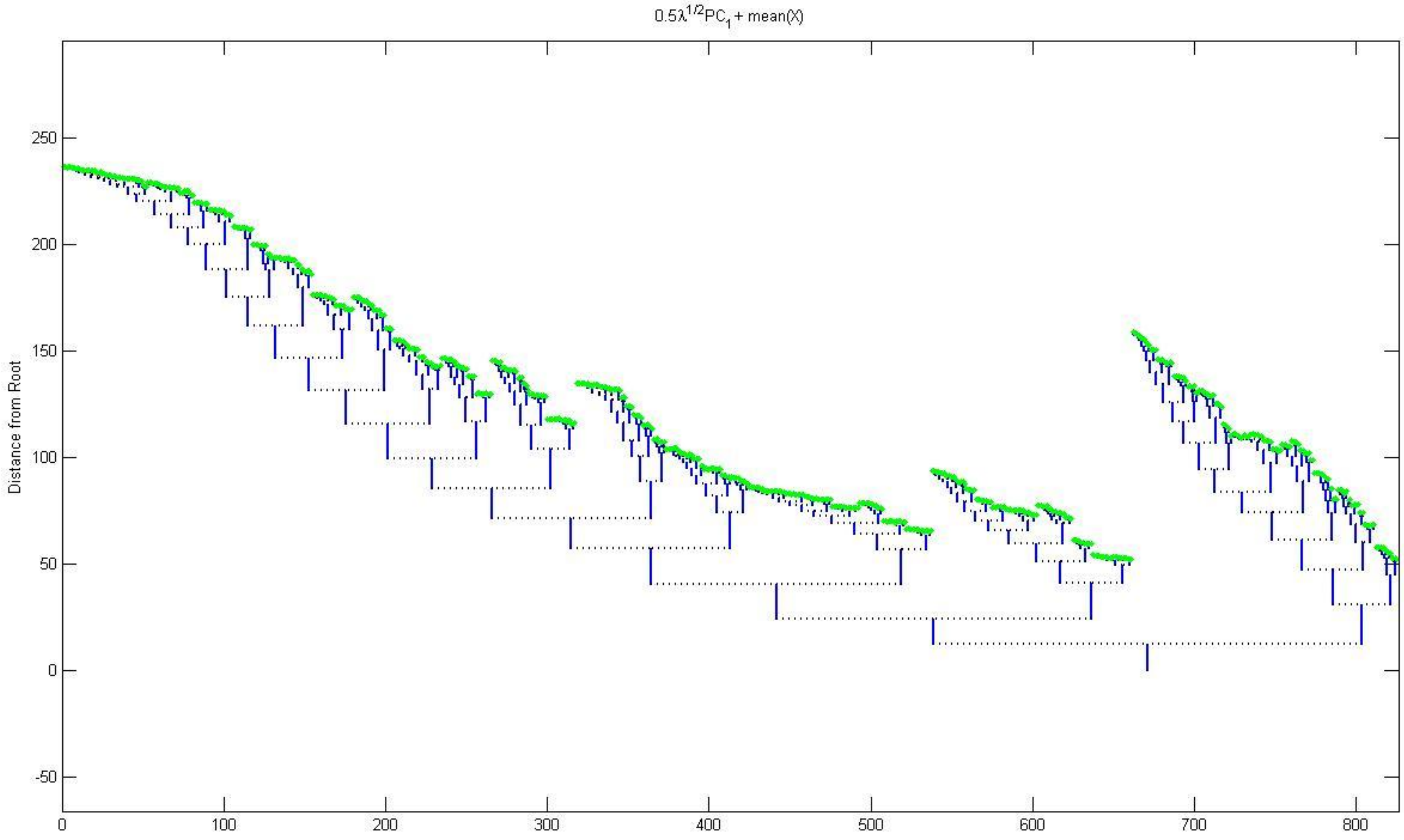


PC1 Direction (Back Tree)

UNC, Stat & OR

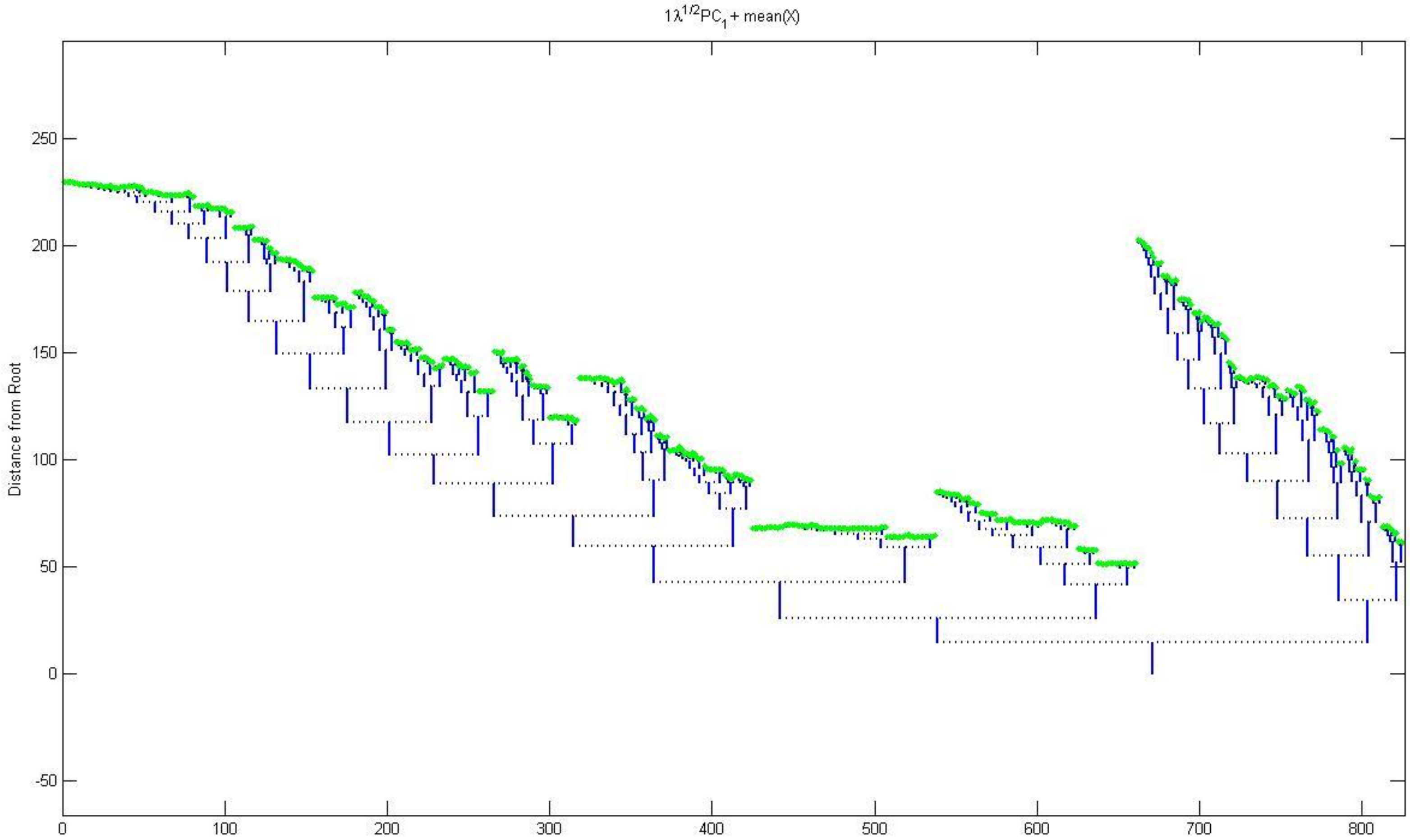


PC1 Direction (Back Tree)

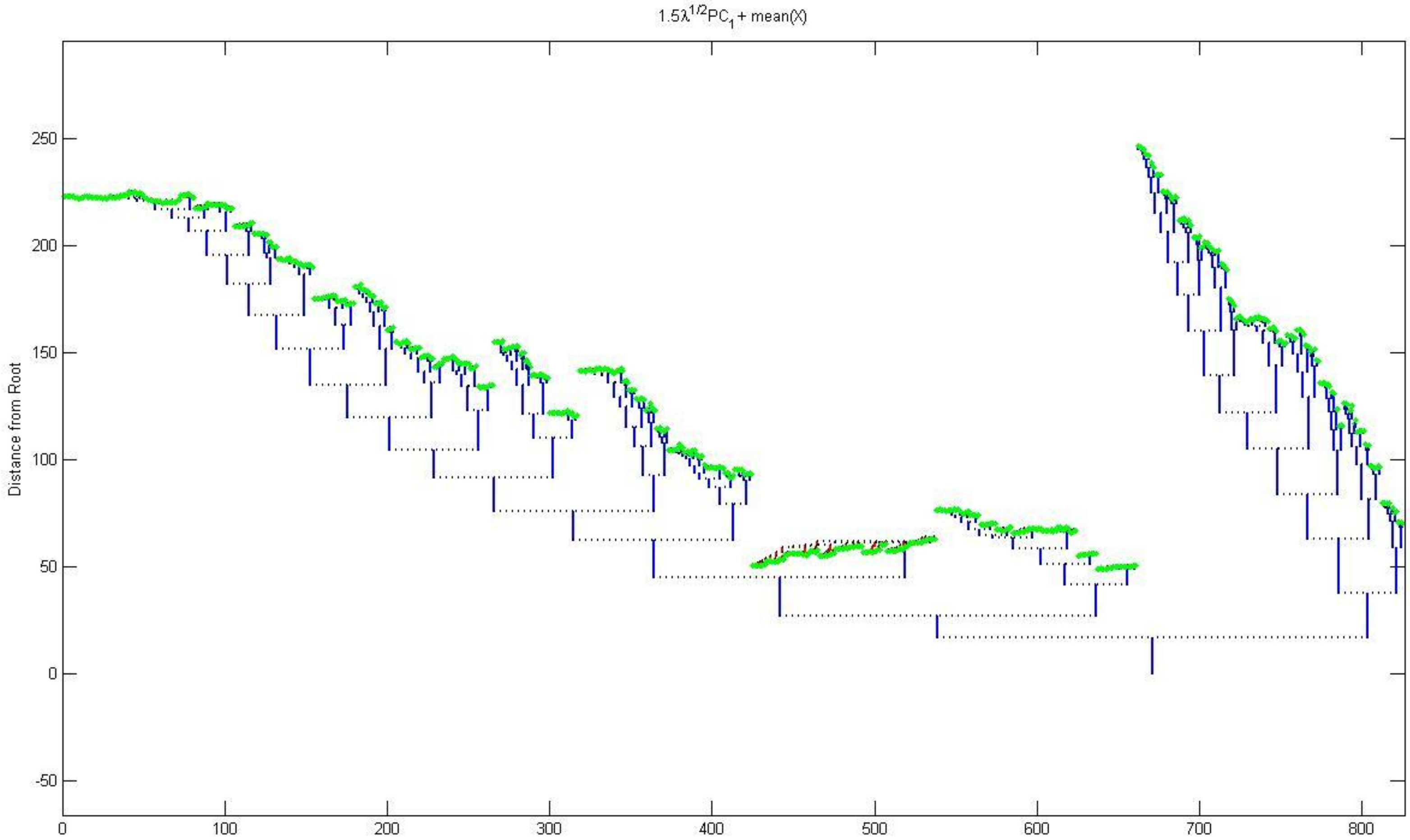


PC1 Direction (Back Tree)

UNC, Stat & OR

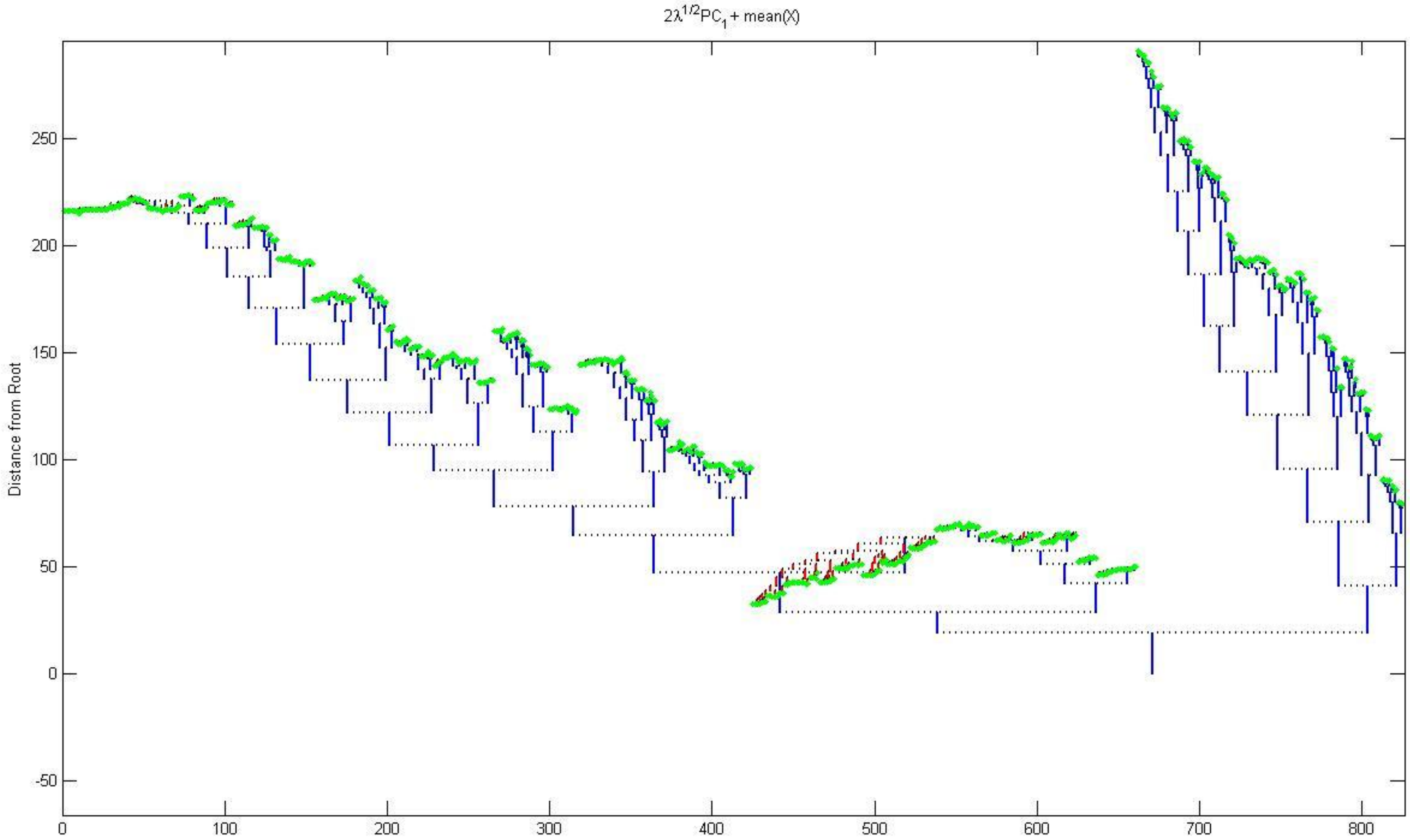


PC1 Direction (Back Tree)



PC1 Direction (Back Tree)

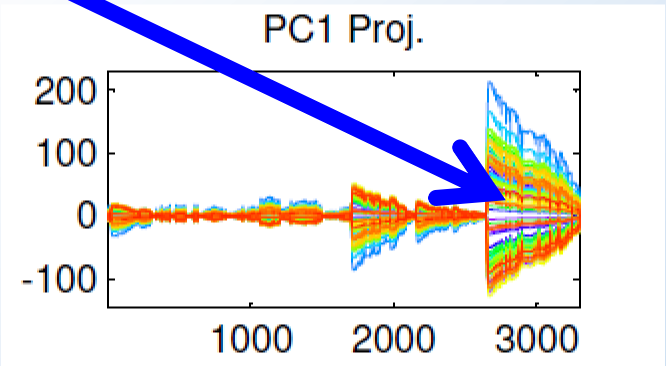
UNC, Stat & OR



Tree Interpretation of the PC direction (Back Tree)

Summary :

- Main variation: banches in the right part of the binary trees
- Reflects the result from the PCA of the Dyck path curves



Thank you !