# تحلیل ژئودزیک اصلی داده‌های آماری در فضای غیراُقلیدسی شکل

**حمیدرضا فتوحی - موسی گل علی زاده**

گروه آمار، دانشکدهٔ علوم ریاضی، دانشگاه تربیت مدرس

**چکیده:** یکی از هدفهای تحلیل آماری شکل، علاوه بر دستیابی به برآوردی از میانگین شکل، برآورد واریانس شکل است. این هدف از طریق روش تحلیل مولفهٔ اصلی قابل حصول می‌باشد. بدلیل محدودیت استفاده از روش تحلیل مولفهٔ اصلی برای مجموعه داده‌هایی از فضای اُقلیدسی، این روش برای داده‌های برآمده از آمار شکل که ماهیتاً داده‌های غیراُقلیدسی هستند، قابـــل کاربرد نیست. دراین حالت می‌توان از تحلیل ژئودزیک اصلی و یا تقریب خطی آن به عنوان تعمیمی از تحلیل مولفهٔ اصلی به فضای غیراُقلیدسی استفاده نمود. در مقالهٔ حاضر پس از ارائه جنبه‌های تئوری روش تحلیل ژئودزیک اصلی، عملکرد آن در یک مطالعه شبیه‌سازی و یک مثال واقعی، مورد ارزیابی قرار خواهد گرفت.

**واژه‌های کلیدی:** تحلیل مولفهٔ اصلی، واریانس شکل، تحلیل ژئودزیک اصلی، فضای غیراُقلیدسی شکل، صفحهٔ مماس

## ۱ مقدمه

امروزه پیشرفت در تکنولوژی سبب گردیده است که مطالعه اشیأ از روی تصاویر هندسی آنها به آسانی امکان‌پذیر باشد. این موضوع علم آمار را با مسائل متفاوت و جدیدی مواجه کرده است که شکل شی در اینگونه مسائل یک منبع اطلاعاتی مناسب بشمار می‌رود. از اینرو اگر قبلاً کاربرد آمار در سایر علوم تا حدودی ملموس نبوده است اما در عصر کنونی کاربرد آمار در بسیاری از علوم مثل زیست شناسی، کامپیوتر، کشاورزی، پزشکی، صنعت و ... غیر قابل انکار است.

در اوایل دهه ۸۰ میلادی بوکشین (۱۹۸۶) و کندال (۱۹۸۴) از جمله کسانی بودند که به طور همزمان به مطالعه آماری اشیاء با در نظر گرفتن ساختار هندسی آنها می‌پرداختند. حاصل زحمات آنها و محققینِ بعد از آنها، منجر به پیدایش شاخه جدیدی در آمار به نام آمار شکل گردیده است. یکی از جنبه‌های مهم تحلیل آماری شکل، برآورد ساختارتغییرات مربوط به شکل، ازطریق نمونه تصادفی اخذشده ازاشیا می‌باشد.

جهت محاسبه واریانس شکل کوتز و همکاران (۱۹۹۲) و کنت (۱۹۹۴) برای اولین بار استفاده از تحلیل مولفه اصلی (PCA) را در فضاهای اُقلیدسی به عنوان تقریبی از فضای شکل پیشنهاد کردند. دلیل تقریب فضای شکل با یک فضای اُقلیدسی به این خاطر است که فضای شکل یک فضای غیراُقلیدسی می‌باشد و لذا استفاده مستقیم از تحلیل مولفه‌های اصلی در آن امکان پذیر نمی‌باشد. اما بهتر است واریانس شکل بدون هیچ تقریب و مستقیماً در همان فضای غیراُقلیدسی شکل محاسبه گردد. بعبارتی دیگر تعمیمی از PCA به فضای غیراُقلیدسی یکی از راه‌حل‌های ممکن است. تحلیل ژئودزیک اصلی[۱] (PGA) تعمیمی از PCA به فضای غیراُقلیدسی می‌باشد. چنین تعمیمی اولین بار توسط فلتچر (۲۰۰۴) برای تحلیل تصاویر کامپیوتری پیشنهاد شد. قابل ذکر است که در مقالهٔ حاضر فضای مماسی متناظر با منیفلد M با مبداء مختصات $p \in M$ را به صورت $T_pM$ نمایش خواهیم داد.

در بخش دوم این مقاله ابتدا نحوه به کارگیری PCA در فضای شکل به اختصار معرفی می‌گردد. در بخش سوم تعمیم PGA و روش تقریب آن در یک فضای اُقلیدسی ارائه می‌گردد. همچنین در بخش چهارم عملکرد آنها در یک مطالعه شبیه‌سازی و یک مثال واقعی به منظور برآورد واریانس حرکت یک مولکول DNA، مورد ارزیابی قرار خواهد گرفت. سپس نتیجه‌گیری و پیشنهاداتی برای تحقیقات آتی در بخش پنجم ارائه خواهد شد.

## ۲ استفاده از PCA در آمار شکل

پس از محاسبه میانگین شکل، برآورد واریانس یکی از مسائل مورد علاقه در تحلیل آماری شکل می‌باشد. از آنجائیکه فضای شکل یک فضای غیراُقلیدسی است امکان محاسبه واریانس به صورتی که در فضای اُقلیدسی مرسوم است، وجود نخواهد داشت (درایدن و ماردیا، ۱۹۹۸). جهت رفع این مشکل کنت (۱۹۹۴) استفاده از PCA در فضای مماسی به عنوان تقریبی از فضای غیراُقلیدسی شکل را که به طور خلاصه در زیر آمده است، پیشنهاد نمود. بدلیل محدودیت در فضای مقاله از بیان مقدمات آمار شکل خودداری می‌گردد. خواننده علاقه‌مند جهت مطالعه دربارهٔ آنها می‌تواند به درایدن و ماردیا (۱۹۹۸) مراجعه کند.

فرض کنید $n$ ماتریس $m \times (k+1)$ بُعدی $X_1, X_2, ..., X_n$، که معرف پیکره‌بندی $n$ شی مختلف می‌باشد، در اختیار است. با پیش‌ضرب این ماتریس‌ها توسط زیر ماتریس

[۱] Principal Geodesic Analysis

هلمرت $H$ با بُعد $k \times (k+1)$ و سپس تقسیم آنها به نرم اُقلیدسی اشیاء مورد نظر به ماتریس‌های جدیدی می‌رسیم که نسبت به تبدیلات انتقال و مقیاس پایا می‌باشند. آنها را بعنوان اطلاعات قبلِ شکل با نمایش ماتریسی $Z_1, Z_2, ...Z_n$ در نظر بگیرید. در این صورت مختصات متغیرهای شکل در فضای مماسی با مرکزیت میانگین شکل را می‌توان توسط رابطهٔ

$$v_j = (I_{km-m} - vec(\hat{\mu})vec(\hat{\mu})^T)(vec(Z_i\hat{\Gamma}_i) \quad j = 3, ..., k$$

تقریب زد که $vec(.)$ بیانگر عملگر برداری کننده، $(\hat{\mu})$ برآورد میانگین شکل به روش پروکراستس تام و $\hat{\Gamma}$ برآورد ماتریس دوران براساس انطباق بهینهٔ قبلِ شکل‌ها به روش پروکراستس تام می‌باشد (کنت، ۱۹۹۴).

اکنون ماتریس واریانس- کواریانس مختصات‌های شکل در فضای مماسی را به صورت $S_v = \sum_{i=1}^{n}(v_i - \bar{v})(v_i - \bar{v})^T$ در نظر بگیرید. واضح است که به کمک $S_v$ اعمال روش PCA در فضای مماسی امکان‌پذیر خواهد بود. بویژه، اثر $j$ امین مولفه بر روی مشاهده $i$ ام به صورت زیر بدست خواهد آمد.

$$x_{ij} = \gamma_j^T(v_i - \bar{v}) \quad i = 1, 2, ..., n; \quad j = 1, 2, 3, ..., p \leq n$$

که در آن $\gamma_j, \ j = 1, 2, ..., p$ بردارهای ویژه حاصل از تجزیه طیفی ماتریس واریانس- کواریانس $S_v$ و $\lambda_j$ مقدارهای ویژه متناظر با $\gamma_j$ می‌باشد. همچنین می‌توان با تقسیم $j$ امین مولفه بر مقدار ویژه متناظرش، به اثر استاندارد شده $j$ امین مولفه بر مشاهده $i$ ام از طریق رابطه زیر دست یافت.

$$c_{ij} = x_{ij}/\lambda_j^{1/2}, \quad i = 1, 2, ..., n; \quad j = 1, 2, 3, ..., p$$

حال با توجه به اینکه هر مولفه اصلی بیانگر درصدی از تغییرات کل می‌باشد، درصدی از عمدهٔ تغییرات شکل، پیرامون میانگین جامعه تحت هر مولفه قابل محاسبه است. بویژه با محاسبه میانگین شکل، معمولاً تغییرات اصلی شکل تحت هر مولفه از طریق کمیت $v = \bar{v} + c_{ij}\lambda^{1/2}\gamma_j$ به ازای $c_{ij} = \pm 3$ توصیف می‌شود.

## ۳ استفاده از PGA در آمار شکل

جهت رفع مشکلِ استفاده از روش PCA در فضای غیراُقلیدسی می‌توان دو رویکرد ذیل را مورد نظر قرار داد:

۱) تقریب فضای غیرأقلیدسی با یک فضای أقلیدسی هم رفتار با آن و به کار بردن PCA در آنجا (همانطور که در بخش ۲ معرفی شد).
۲) به کار بردن روش PGA به عنوان تعمیمی از تحلیل مولفه‌های اصلی به فضای غیرأقلیدسی.
روش PGA مشابه روش PCA، علاوه بر کاهش بعد داده‌های مورد مطالعه روشی برای برآورد واریانس جامعه می‌باشد. باید توجه داشت که بر خلاف روش PCA که استفاده از آن محدود به فضای اقلیدسی است روش PGA را می‌توان برای زمانی که فضای مورد مطالعه، یک فضای غیرأقلیدسی است نیز بکار برد. در تحلیل ژئودزیک اصلی با فرض غیرأقلیدسی بودن فضای مورد مطالعه، هدف تصویر داده‌ها به داخل زیر فضایی با بعد کمتر می‌باشد به طوری که داده‌های تصویر شده دارای بیشترین پراکندگی در اطراف میانگین باشند. لازمه رسیدن به این مهم تعمیم برخی مفاهیم اساسیِ PCA به فضای غیرأقلیدسی می‌باشد، که در ادامه به توضیح در مورد آنها خواهیم پرداخت.

### ۱.۳ فاصله ریمانی و واریانس آماری

واریانس نمونه‌ای مشاهدات متعلق به یک فضای غیرأقلیدسی را می‌توان بعنوان مجموع مربعات فاصله ریمانیِ مشاهدات از میانگین به صورت زیر تعریف کرد (پنس، ۱۹۹۹)

$$\sigma^{۲} = \frac{۱}{N}\sum_{i=۱}^{N} d(\mu, x_i)^{۲} = \frac{۱}{N}\sum_{i=۱}^{N} \| Log_{\mu}(x_i) \|^{۲},$$

که در آن $d$ متر تعریف شده در فضای غیرأقلیدسی و $\| . \|$ بیانگر نرم‌أقلیدسی می‌باشد. همچنین $Log_{\mu}(x_i)$ بیانگر نگاشتی به منظور انتقال مشاهدهٔ $x_i, \quad i = ۱, ۲, ..., n$ از فضای غیرأقلیدسی به فضای مماسی با مرکزیت $\mu$ (به عنوان تقریبی از فضای غیر أقلیدسی) می‌باشد و به نگاشت لگاریتمی معروف است (فلتچر و همکاران، ۲۰۰۴).
از طرف دیگر با توجه به اینکه در روش PCA به دنبال یک زیرفضای أقلیدسی از فضای أقلیدسیِ اولیه خواهیم بود به طوری که تصویر داده‌ها در آن دارای بیشترین واریانس باشد. بنابراین در مقایسه با روش PCA زمانی که فضای مورد مطالعه غیر أقلیدسی است، با تعمیم مفهوم زیر فضای أقلیدسی به زیر فضای غیرأقلیدسی(زیرمنیفلد ژئودزیک از منیفلد اصلی)، به دنبال زیر منیفلدی می‌باشیم که تصویر داده‌ها در آن دارای بیشترین واریانس باشد.

یکی از نکات مهم دیگر که در تعمیم PCA به فضای غیرأقلیدسی مدنظر قرار داد مفهوم تصویرِ متعامد نقطهٔ $x \in M$ به داخل یک زیر منیفلد $H \subseteq M$ می‌باشد. این

موضوع را می‌توان به عنوان نقطه‌ایی متعلق به زیرمنیفلد $H$ که دارای نزدیکترین فاصله ریمانی از $x \in M$ می‌باشد به صورت زیر در نظر گرفت:

$$\pi_H(x) = \arg\min_{y \in H} d(x,y)^2$$

فرض کنید $T_\mu M$ یک فضای مماسی است که مبداء آن میانگین مشاهدات ($\mu$) است. می‌دانیم رفتار $T_\mu M$ حول ($\mu$) مشابه رفتار منیفلد $M$ حول میانگین خودش است. حال می‌توان گفت تحلیل ژئودزیک اصلی معادل دستیابی به بردارهای مستقل خطی $\{\boldsymbol{\nu}_1 ... \boldsymbol{\nu}_k\}$ می‌باشد که تشکیل یک پایه متعامد برای $T_\mu M$ می‌دهند. به عبارتی دیگر داریم:

$$T_\mu M = span\{\boldsymbol{\nu}_1 ... \boldsymbol{\nu}_k\},$$

همچنین در قیاس با PCA، هر یک از $\boldsymbol{\nu}_i$ ها به ازای $i = 1, 2, ...k$، جهت i امین ژئودزیک اصلی بوده که از مجموعهٔ روابط زیر بدست می‌آیند (فلتچر، ۲۰۰۴):

$$\begin{cases} \boldsymbol{\nu}_1 = argmax_{\|\nu\|=1} \sum_{i=1}^{N} \| Log_\mu(\pi_H(x_i)) \|^2 \\ H = Exp_\mu(span\{\boldsymbol{\nu}\}) \end{cases}$$
$$\vdots \qquad (1)$$
$$\begin{cases} \boldsymbol{\nu}_k = argmax_{\|\nu\|=1} \sum_{i=1}^{N} \| Log_\mu(\pi_H(x_i)) \|^2 \\ H = Exp_\mu(span\{\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_{k-1}, \boldsymbol{\nu}\}) \end{cases}$$

که در آن $Exp_\mu(x_i)$ بیانگر نگاشتی به منظور انتقال مشاهدهٔ $x_i, \quad i = 1, 2, ..., n$ از فضای مماسی با مرکزیت $\mu$ (به عنوان تقریبی از فضای غیراقلیدسی) به فضای غیراُقلیدسی می‌باشد. واضح است که $Exp_\mu(x_i)$ معکوس نگاشت لگاریتمی است و از آن به عنوان نگاشت نمایی یاد می‌شود.

حال با در نظر گرفتن $V_k = span\{\boldsymbol{\nu}_1 ... \boldsymbol{\nu}_k\}$، می‌توان $H_k = Exp_\mu(V_k)$ را به عنوان $k$ امین زیر منیفلد ژئودزیک از منیفلد اولیه در نظر گرفت که نه تنها دارای بعد کوچکتری نسبت به منیفلد اولیه است، بلکه تصویر مشاهدات در آن دارای بیشترین واریانس پیرامون میانگین جامعه خواهد بود.

### ۲.۳ تقریب خطی PGA

در بسیاری از مواقع محاسبه تصویر یک نقطه به داخل یک زیر منیفلد ژئودزیک مشکل و حتی در صورت امکان از نظر محاسباتی بسیار زمان‌بر است (سن، ۲۰۰۸). در عوض می‌توان تصویر را در یک فضای مماسی با مرکزیّت میانگین که هم رفتار با فضای منیفلد M می‌باشد تقریب زد.

فرض کنید که $H \subseteq M$ یک زیر منیفلد از منیفلد $M$ و گذرنده از نقطهٔ $p \in M$ باشد. آنگاه می‌توان تصویر $x \in M$ به روی زیر منیفلد ژئودزیک H را توسط رابطهٔ زیر تقریب زد:

$$\pi_H(x) = argmin_{y \in H} \parallel Log_x(y) \parallel^2 \simeq argmin_{y \in H} \parallel Log_p(x) - Log_p(y) \parallel^2,$$

همچنین فاصله ژئودزیک تصویر $\pi_H(x)$، از نقطه $p \in M$ را می‌توان به صورت تقریبی از رابطه زیر بدست آورد(فلتچر و همکاران، ۲۰۰۴):

$$Log_p(\pi_H(x)) \simeq argmin_{\nu \in T_pH} \parallel \ Log_p(x) - \nu \parallel^2, \qquad (۲)$$

اگر $T_PH$ بیانگر فضای مماسی با مبداء مختصات $p$ و متناظر با زیر منیفلد $H$ باشد، و با فرض اینکه $\{\boldsymbol{\nu}_1 ... \boldsymbol{\nu}_k\}$ یک پایه متعامد برای $T_PH$ است، آنگاه رابطه تقریبی ذکر شده در رابطهٔ (۲) را می‌توان به صورت زیر بازنویسی نمود:

$$Log_p(\pi_H(x)) \simeq \sum_{i=1}^{k} < \nu_i, Log_p(x) >, \qquad (۳)$$

که نماد $< ., . >$ نشانگر حاصلضرب درونی است.

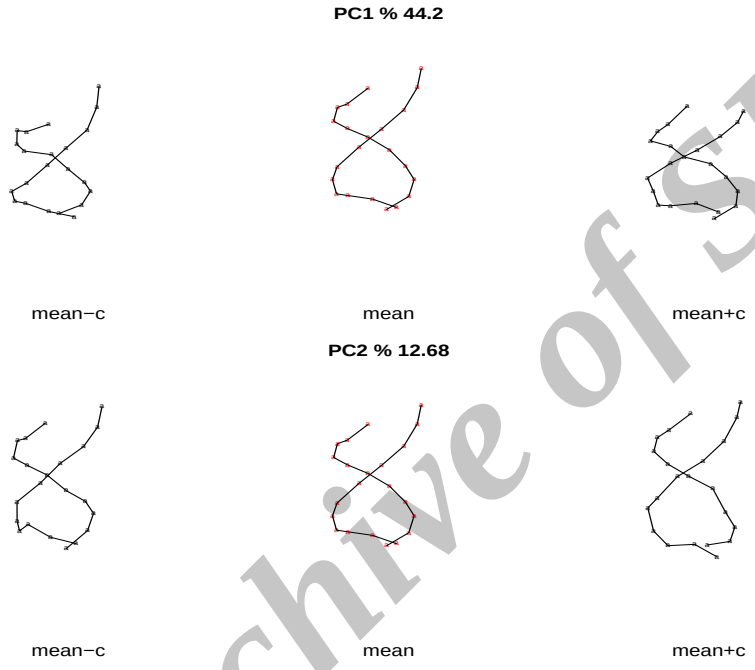
بنابراین با جایگزینی تقریب ارائه شده در رابطه (۳) در روابط (۱)، می‌توان مقدار جهت‌های ژئودزیک اصلی را از حل روابط زیر بدست آورد:

$$\begin{cases} \nu_1 \simeq argmax_{\|\nu\|=1} \sum_{i=1}^{N} < \nu, Log_\mu(x_i) >^2 \\ \vdots \\ \nu_k \simeq argmax_{\|\nu\|=1} \sum_{i=1}^{N} \sum_{j=1}^{k-1} < Log_\mu(x_i), \nu_j >^2 + < \nu, Log_\mu(x_i) >^2 \end{cases} \qquad (۴)$$

لازم به ذکر است که جهت‌های تقریبی بدست آمده از روابط (۴) را می‌توان به عنوان جهت‌های مربوط به تحلیل مولفه‌های اصلی در فضای اُقلیدسی در نظر گرفت. لذا مشابه روش PCA به کمک بردارهای جدید $\{U_1, U_2, ..., U_N\}$، می‌توان از تجزیه طیفی ماتریس واریانس- کواریانس حاصل از آنها یعنی $S = \frac{1}{N} \sum_{i=1}^{N} U_i U_i^T$ که در آن $U_i = Log_\mu(x_i)$ ، به ازای $i = 1, 2, ..., N$ می‌باشد، استفاده نموده و جهت‌های ژئودزیک تقریبی را بدست آورد. در ادامه با ارائه یک مثال واقعی نحوه به کارگیری روش‌های ارائه‌شده درمقاله ر بررسی نموده و به‌تفسیر خروجی‌های مربوطه می‌پردازیم.

## ۴ مثال کاربردی

داده‌های مورد استفاده در این مقاله مربوط به حرکت یک مولکول DNA در ۳۰ زمان متفاوت می‌باشد. که اطلاعات مربوط به هر زمان، در قالب ماتریسی با بُعد ۳ × ۲۲ ذخیره شده است. این داده‌ها در بسته آمار شکل (*shapes*) در نرم افزار آماری *R*، موجود و از طریق پایگاه http://cran.r-project.org/web/packages/shape قابل دسترسی می‌باشد. در این مثال جهت برآورد ساختار تغییرات مربوط به حرکت مولکول DNA، ابتدا فضای شکل را با یک فضای اُقلیدسی هم رفتار با آن تقریب زده و دو مولفه اصلی اول را در نمودار (۱) رسم کرده‌ایم.

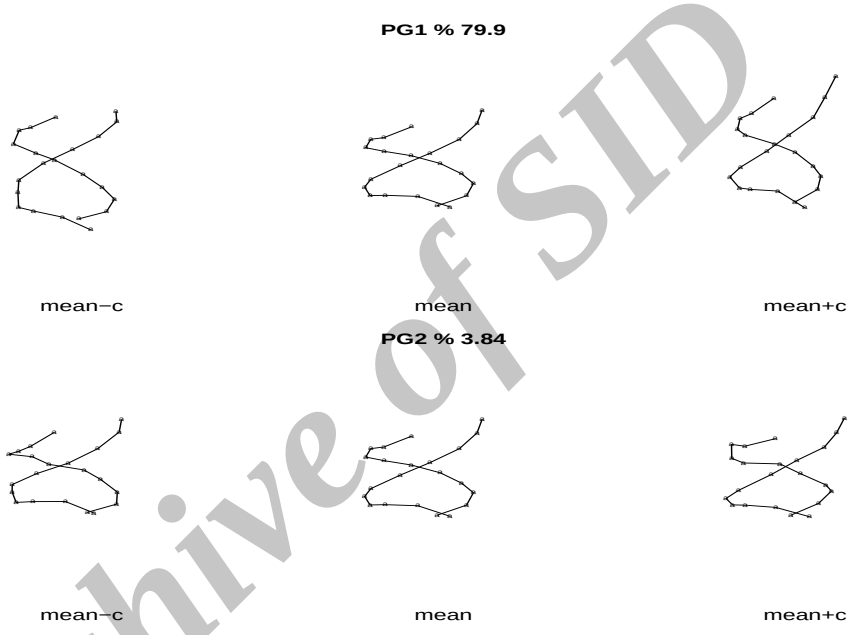



شکل ۱.: دو مولفه اصلی اول برای حرکت یک مولکول DNA. هر ردیف افقی بیانگر عمده تغییرات پیرامون میانگین در داخل یک فضای مماسی اُقلیدسی و هم‌رفتار با فضای غیراُقلیدسی به ازای امتیاز استاندارد ۳ =c می‌باشد.

بنا به این نمودار، اولین ردیف بیانگر عمدهٔ پراکندگی پیرامون میانگین شکل تحت اولین مولفه اصلی می‌باشد. در واقع از روی این ردیف و با توجه به مقدار ویژه متناظر با اولین جهت اصلی می‌توان نتیجه‌گرفت که درحدود ۴۴/۲ درصد از پراکندگی

کل مربوط به نقاط شاخصی است که قسمت میانی بین دو شاخک مولکول DNA و پیرامون آن را تشکیل می‌دهند. همچنین با توجه به ردیف دوم می‌توان دریافت که ۱۲/۶۸ درصد از تغییرات کل مربوط به قسمت‌های انتهایی شاخک‌های تشکیل دهندهٔ مولکول DNA می‌باشد.

به منظور توصیف دقیقتر واریانس حرکت DNA در فضای واقعی غیرأقلیدسی شکل، و بدون وجود محدودیت ذکر شده ناشی از بکارگیری روش PCA، با انجام شبیه‌سازی و اعمال تحلیل ژئودزیک اصلی تقریبی، دو ژئودزیک تقریبی اول را در نمودار (۲) رسم کرده‌ایم.




شکل ۲.: دو ژئودزیک تقریبی اول برای حرکت یک مولکول DNA. هر ردیف افقی بیانگر عمده تغییرات پیرامون میانگین در داخل یک فضای أقلیدسی و هم‌رفتار با فضای غیرأقلیدسی به ازای امتیاز استاندارد ۳ =c می‌باشد.

بنا به نمودار اولین ردیف بیانگر عمدهٔ پراکندگی پیرامون میانگین شکل تحت اولین ژئودزیک اصلی می‌باشد. در حقیقت از روی این ردیف و با توجه به مقدار ویژه متناظر با اولین جهت ژئودزیک اصلی می‌توان نتیجه گرفت که در حدود ۷۹ درصد از

پراکندگی کل مربوط به میزان کشیدگی مولکول DNA می‌باشد. همچنین ردیف دوم این مطلب را بیان می‌کند که ۳/۸۴ درصد از تغییرات کل مربوط به نقاط شاخص میانی و نقاط انتهایی است. قابل ذکر است که یکی از دلایل محاسبه جهت‌های ژئودزیک تقریبی، کوتاه بودن زمان لازم در تعیین آنها است. واضح است که تحقیق بیشتر در مورد حرکت مولکول نیازمند لحاظ نمودن نظر متخصصان زیست‌شناسی و در عین حال مدل‌بندیِ جامع‌تر است. این موضوع هدف تحقیقات آتی ما خواهد بود.

## بحث و نتیجه‌گیری

ارزیابی تغییرات دقیق داده‌های آماری شکل، که ذاتاً داده‌های غیراُقلیدسی هستند، یکی از موضوعات مهم و مورد علاقهٔ محققین علوم کاربردی است. بنا به روش‌های سنتی، PCA تا حدودی به این موضوع پاسخ می‌دهد. اما ضعف عمده آن این است که داده‌های اولیه را در یک فضای خطی مماسی تصویر می‌نماید. تحلیل ژئودزیک اصلی که در فضای واقعی شکل مورد استفاده قرار می‌گیرد از این مشکل مبرّا می‌باشد. در این مقاله با استفاده از یک مثال کاربردی مربوط به حرکت مولکول DNA، دو روش PCA و PGA تقریبی مورد مقایسه قرار گرفت.

## مراجع


Bookstein, F.L. (1986), Size and Shape Spaces for Landmark Data in Two Dimensions (with discussion), *Statistical Science,* **1**, 181-242.

Cootes, T.F., Taylor, C.J., Cooper, D.H., and Graham.J. (1992), Training Models of Shape from Sets of Examples, In Hogg, D.C. and Boyle, R.D.(eds), *British Machine Vision Conference,* 9-18, Springer-Verlag, Berlin.

Fletcher, P. (2004). *Statistical Variability in Nonlinear Spaces: Application to Shape Analysis and DT-MRI*, Ph.D Thesis, University of North Carolina at Chapel Hill.

Fletcher, P., Lu, C., Pizer, S., Joshi, S. (2004), Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape, *Medical Imaging, IEEE Transactions,* **23**, 995-100

Dryden, I. and Mardia, K. (1998), *Statistical Shape Analysis.* New York, Wiley

Kendall, D.G. (1984), Shape Manifolds, Procrustean Metrics and Complex Projective Spaces, *Bulletin of the London Mathematical Society*, **16**, 81-121

Kent, J.T. (1994), The Complex Bingham Distribution and Shape Analysis, *Journal of the Royal Statistical Society. Series B,* **56**, 285-299

Sommer, S., Lauze, F., Hauberg, S. and Nielson, M. (2010), Manifold Valued Statistics, Exact Principal Geodesic Analysis and the Effect of Linear Approximations, *Lecture Notes in Compute Science*, **6316**, 43-56

Sen, S. K. (2008), *Classification on Manifolds*, Ph.D Thesis, University of North Carolina at Chapel Hill.

Kendall, W. S. (1990), Probability, Convexity, and Harmonic Maps with Small Image-I: Uniqueness and fine existence, *Proc. Lond. Math. Soc*, **3**, 371-406

Pennec, X. (1999), Probabilities and Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements, *IEEE Workshop on Nonlinear Signal and Image Processing, Antalaya, Turkey.*