Vladimir DRAGOVIĆ[*], Borislav GAJIĆ[**]

BRIDGING STATISTICS, GEOMETRY, AND MECHANICS

Abstract: We emphasis the importance of bridges between statistics, mechanics, and geometry. In par ticular, we developed and employed links between pencils of quadrics, moments of inertia, and linear and orthogonal regressions. For a given system of points in $\mathbb{R}^k$ representing a sample of a full rank, we recently constructed a pencil of confocal quadrics which provided a useful geometric tool to study the data.

Key words and phrases: *ellipsoid of concentration, confocal pencil of quadrics, planar moments of inertia, restricted regression, regularization and shrinkage, restricted PCA*

1. Introduction

We will talk here about bridges between statistics, mechanics, and geometry. In particular, we talk about the links we developed and employed between pencils of quadrics, moments of inertia, and linear and orthogonal regressions. This cross-fertilization between the areas appears to be beneficial for each of them. Individual quadrics have been in use in statistics from late XIX century. Recently, we constructed a new object, a confocal pencil of quadrics, associated to a given data set, see [**15, 16**]. We demonstrated in [**15, 16**] that this confocal pencil of quadrics is a natural and very useful instrument to understand the data.

1.1. Confocal pencils of quadrics.

We age going to recall about confocal pencils of conics in the plane and its generalizations, the confocal pencils of quadrics in $\mathbb{R}^k$ for for any k. We will also recall the definition of the associated Jacobi elliptic coordinates.

The family of confocal conics in the plane can be given analytically:

$$\mathcal{C}_\lambda \ : \ \frac{x^2}{\alpha - \lambda} + \frac{y^2}{\beta - \lambda} = 1. \tag{1.1}$$

[*] Vladimir Dragović, Department of Mathematical Sciences, The University of Texas at Dallas, Richardson, TX, USA, Mathematical Institute, of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

[**] Borislav Gajić, Mathematical Institute, of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

We notice that the family (1.1) contains two types of smooth conics: ellipses, when $\lambda < \beta$, and hyperbolas, when $\lambda \in (\beta, \alpha)$. Gometrically, they all share a common pair of foci. We also notice that there are two degenerated conics in the confocal family: the x-axis, for $\lambda = \beta$; and the y-axis, for $\lambda = \alpha$.

Each point in the plane, which is not a focus of the confocal family, lies on exactly two conics $\mathcal{C}_{\lambda_1}$ and $\mathcal{C}_{\lambda_2}$ from (1.1) – one ellipse and one hyperbola, which are orthogonal to each other at the intersection point.

Let an ellipsoid be given in $\mathbb{R}^k$ by:

$$\mathcal{E} \ : \ \frac{x_1^2}{\alpha_1} + \cdots + \frac{x_k^2}{\alpha_k} = 1, \quad \alpha_1 > \alpha_2 > \cdots > \alpha_k > 0. \tag{1.2}$$

The family of quadrics confocal with $\mathcal{E}$ is:

$$\mathcal{Q}_\lambda(\mathbf{x}) = \frac{x_1^2}{\alpha_1 - \lambda} + \cdots + \frac{x_k^2}{\alpha_k - \lambda} = 1, \tag{1.3}$$

where λ is a real parameter. We will say that confocal quadrics $\mathcal{Q}_\lambda, \mathcal{Q}_\mu$ are *of the same type* if there exists $j, j = 1, \ldots, k-1$ such that $\alpha_j > \lambda, \mu > \alpha_{j+1}$ or $\alpha_k > \lambda, \mu$. For a point given by its Cartesian coordinates $\mathbf{x} = (x_1, \ldots, x_k)$, the Jacobi elliptic coordinates $(\lambda_1, \ldots, \lambda_k)$ are defined as the solutions of the equation in λ: $\mathcal{Q}_\lambda(\mathbf{x}) = 1$. The quadrics $\mathcal{Q}_{\lambda_1}, \mathcal{Q}_{\lambda_2}, \ldots, \mathcal{Q}_{\lambda_k}$ which contain a given point $\mathbf{x}$ are of different types [**4**]. Jacobi introduced the Jacobi elliptic coordinates in [**31**] in 1838 when he used them to integrate the equations of geodesics on ellipsoids.

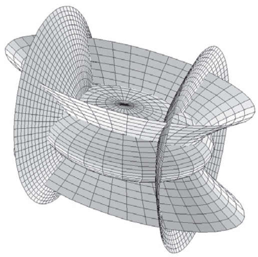

Figure 1. From [**17**] and [**18**]: Confocal quadrics in $\mathbb{R}^3$: one ellipsoid, one 1-sheeted hyperboloid, and one 2-sheeted hyperboloid, intersecting orthogonally at eight points.

1.2. Application to billiard systems within ellipses. Similar in spirit to the use of the Jacobi elliptic coordinates in solving the equations of geodesics on ellipsoids, is the application of these coordinates to billiards within ellipsoids.

The theory of mathematical billiards is one of theoretical models for motion of a ball inside a billiard table, and its reflection off the boundary of the table. Let us suppose that a planar domain is given. *Mathematical billiard* in this domain is a dynamical system where a material point of unit mass moves under inertia without constraints and friction within the domain, and obeys the billiard reflection law off the boundary [**32**]. The billiard reflection law coincides with the law reflection of light in geometric optics. This theory provides an idealized model of the physical reality in many aspects. For example, a usual billiard ball is replaced by a material point, and the friction and spin are neglected. Nevertheless, this model has many important and natural applications, for example in geometric optics. Thus, the billiard dynamics has two different regimes: the first one is inside the billiard domain, when we assume that the material point moves under the inertia, i.e. uniformly along straight lines. The second regime concerns with the impacts off the boundary. We assume here that the impacts are *absolutely elastic*. In other words, the geometric billiard law is satisfied: the impact and reflection angles are congruent to each other and the speed remains unchanged after the impacts. Here, the trajectories of the mathematical billiards are polygonal lines with vertices at the boundary.

We focus here only on *elliptical billiards* – defined by an ellipse in the Euclidean plane as the boundary of a billiard table:

$$\mathcal{E}\ :\ \frac{x^2}{\alpha}+\frac{y^2}{\beta}=1,\quad \alpha>\beta>0.$$

The key property of such billiards is the existence of caustics. Each trajectory of the elliptic billiard has *a caustic*: a curve such that each segment of the trajectory lies on a line tangent to the caustic. The existence of a caustic is a geometric manifestation of the integrability of billiard systems within ellipses.

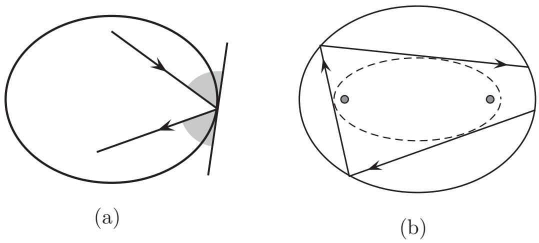

FIGURE 2. 2(a): Billiard reflection law; 2(b): The caustic of a billiard trajectory.

1.3. Quadrics in Statistics. Ellipses, as two-dimensional quadrics, got incorporated in statistics in 1886 in Galton's paper [**26**]. That seminal paper introduced the law of regression.

To study the hereditary transmissions, Galton collected the data about the height of 930 adult children and 205 of their respective parentages. He introduced a "mid-parent" height, as a weighted average of the heights of the parents, and assigned it to each pair of parents. He established the average regression from mid-parent to offsprings and from offsprings to mid-parent. Using this data Galton formulated the law of regression toward mediocrity:

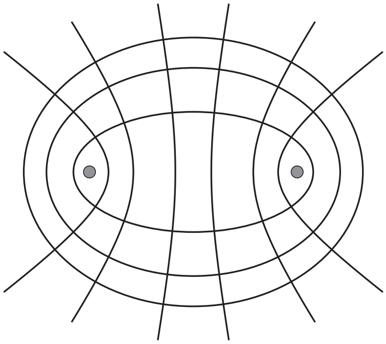

Figure 3. Confocal family of conics in plane

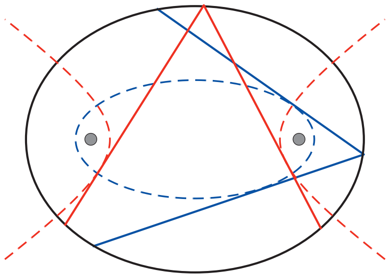

Figure 4. A billiard trajectory with an ellipse as the caustic and a billiard trajectory with a hyperbola as a caustic.

When Mid-Parents are taller than mediocrity, their Children tend to be shorter than they. When Mid-Parents are shorter than mediocrity, their Children tend to be taller than they.

Thus, the notion of *regression* got into statistics, thanks to Galton. From mathematical perspective, the background method of least squares, appeared in early 1800's, due to Gauss and Legendre.

Galton discovered an important use of ellipses in statistical analysis. Here we quote Galton:

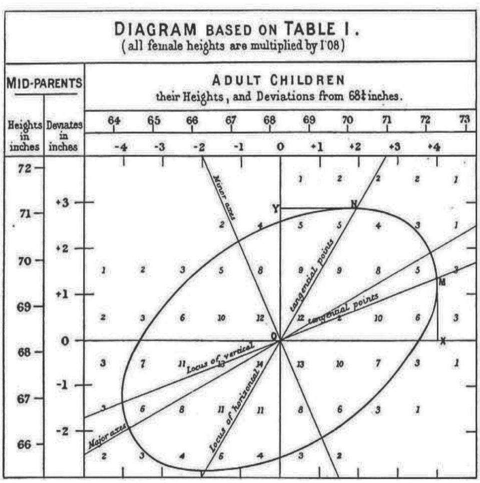

Figure 5. From [**26**].

"*...I found it hard at first to catch the full significance of the entries in the table...They came out distinctly when I 'smoothed' the entries by writing at each intersection of a horizontal column with a vertical one, the sum of entries of four adjacent squares... I then noticed that lines drawn through entries of the same value formed a* *series of concentric and similar ellipses. Their common center ... corresponded to* $68\frac{1}{4}$ *inches. Their axes are similarly inclined. The points where each ellipse in succession was touched by a horizontal tangent, lay in a straight line inclined to the vertical in the ratio of* $\frac{2}{3}$*; those where they were touched by a vertical tangent lay in a straight line inclined to the horizontal in the ratio of* $\frac{1}{3}$*. These ratios confirm the values of average regression already obtained by a different method, of* $\frac{2}{3}$ *from mid-parent to offspring, and of* $\frac{1}{3}$ *from offspring to mid-parent...These and other relations were evidently a subject for mathematical analysis and verification...I noted these values and phrased the problem in abstract terms such as a competent mathematician could deal with, disentangled from all references to heradity, and in that shape submitted it to Mr. Hamilton Dixson, of St. Peter's College, Cambridge...*

I may be permitted to say that I never felt such a glow of loyalty and respect towards the sovereignty and magnificent sway of mathematical analysis as when his answer reached me, confirming, by purely mathematical reasoning, my various and laborious statistical conclusions with far more minuteness than I had dared to hope... His calculations corrected my observed value of mid-parental regression from $\frac{1}{3}$ *to* $\frac{6}{17.6}$*, the relation between the major and minor axis of the ellipse was changed 3 per cent. (it should be as* $\frac{\sqrt{7}}{\sqrt{2}}$*)..."*

The notions of moments in statistics came from mechanics, where they were originally introduced in the three-dimensional space. The next Section 1.4 reviews these notions from a mechanics perspective.

1.4. Axial and planar moments of inertia in $\mathbb{R}^3$. We review the notions of the axial moment of inertia and the operator of inertia with respect to a point, as they are introduced in mechanics. Suppose that N points $M_1, ..., M_N$ with masses $m_1, m_2, ..., m_N$ are given in the space $\mathbb{R}^3$. Sum of the masses $m = \sum_{j=1}^{N} m_j$ is the total mass of the system of the points, while *the center of masses* of the N given points is the point C that satisfy

$$\sum_{j=1}^{N} m_j \overrightarrow{CM_j} = 0.$$

For a given line $u \subset \mathbb{R}^3$, *the axial moment of inertia* I_u is defined as

$$I_u = \sum_{j=1}^{N} m_j d_j^2,$$

where d_j is the distance from the point M_j to the line u, $j = 1, \dots, N$.

Suppose that the line u contains point O and that it is defined with the unit vector $\mathbf{u}_0$.

Then the axial moment of inertia I_u for the axis u can be rewritten in the form:

$$I_u = \sum_{i=1}^{N} m_i d_i^2 = \sum_{i=1}^{N} m_i \langle \mathbf{u}_0 \times \overrightarrow{OM_i}, \mathbf{u}_0 \times \overrightarrow{OM_i} \rangle = \sum_{i=1}^{N} m_i \langle \overrightarrow{OM_i} \times (\mathbf{u}_0 \times \overrightarrow{OM_i}), \mathbf{u}_0 \rangle = \langle I_O \mathbf{u}_0, \mathbf{u}_0 \rangle.$$

If one considers u as an arbitrary line that contains fixed point O, the relation

$$I_u = \langle I_O \mathbf{u}_0, \mathbf{u}_0 \rangle$$

defines *the operator of inertia at the point* O and it is denoted by I_O. In the Cartesian coordinates $Oxyz$ the matrix of the operator I_O is symmetric and positive-definite [**4**].

Its diagonal elements I_{11}, I_{22}, I_{33} are the moments of inertia for the coordinate axes Ox, Oy, Oz respectively. For example:

$$I_{11} = \sum_{j=1}^{N} m_j (y_j^2 + z_j^2).$$

The non-diagonal elements of this matrix are called *the centrifugal moments of inertia.* For example:

$$I_{12} = -\sum_{p=1}^{N} m_p x_p y_p,$$

and similarly, one can define all other I_{ij}.

As for any symmetric matrix, one can choose an orthogonal basis in which the matrix of the operator I_O has a diagonal form:

$$I_O = \mathrm{diag}(I_1, I_2, I_3).$$

The scalars I_1, I_2, I_3 are called *the principal axial moments of inertia* while the corresponding coordinates are called *principal coordinates.*

The inertia operator I_O defines *the axial ellipsoid of inertia at the point* O:

$$\langle I_O u, u \rangle = 1.$$

In the principal coordinates, the equation of the axial ellipsoid of inertia becomes

$$I_1 x^2 + I_2 y^2 + I_3 z^2 = 1.$$

An important formula that makes a connection between the axial moments of inertia for two parallel axes, where one of them contains the center of masses, is the essence of the Huygens-Steiner Theorem (see e.g. [**4**]).

Theorem 1.1 (The Huygens-Steiner Theorem, [**4**]). *Let the axis u contain the center of masses C and let u_1 be a line parallel to u. Denote by I_{u_1} and I_u the corresponding axial moments of inertia of a given system of points with the total mass m. Then*

$$I_{u_1} = I_u + md^2, \tag{1.4}$$

where d is the distance between the parallel lines u and u_1.

As an immediate consequence, one gets an important property of the center of masses:

Corollary 1.1. *Given a system of points and one direction. Among all the lines parallel to the given direction, the least moment of inertia is attained for the line which contains the center of masses of the set of points.*

In a similar manner as for an axial moments of inertia, for a given set of points and for a given plane $\pi \subset \mathbb{R}^3$ one can introduce the planar moment of inertia. It is defined by formula:

$$J_\pi = \sum_{j=1}^{N} m_j D_j^2,$$

where D_j is the distance between the point M_j and the plane π, for $j = 1, \ldots, N$. For vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ define the operator

$$\langle J_O \mathbf{n_1}, \mathbf{n_2} \rangle = \sum_{j=1}^{N} m_j \langle \overrightarrow{OM_j}, \mathbf{n_1} \rangle \langle \overrightarrow{OM_j}, \mathbf{n_2} \rangle.$$

If the plane π is determined by its point O and the normal unit vector $\mathbf{n}$, then the planar moment of inertia J_π can be rewritten in the form

$$J_\pi = \sum_{j=1}^{N} m_j \langle \overrightarrow{OM}_j, \mathbf{n} \rangle^2 = \langle J_O \mathbf{n}, \mathbf{n} \rangle. \tag{1.5}$$

This is the reason to call J_O *the planar inertia operator at the point* O.

In the Cartesian coordinates $Oxyz$, the diagonal elements of the matrix of the planar inertia operator are the planar moments of inertia for the coordinate planes. For example:

$$J_{11} = \sum_{j=1}^{N} m_j x_j^2$$

is the moment of inertia for the coordinate plane Oyz. The nondiagonal elements of the planar inertia operator are also called *the centrifugal moments of inertia*. For example:

$$J_{12} = \sum_{p=1}^{N} m_p x_p y_p,$$

and similarly for other J_{ij}. The planar inertia operator is symmetric and positive-definite. Thus, one can choose a basis in which its matrix has a diagonal form:

$$J_O = \mathrm{diag}(J_1, J_2, J_3).$$

The extremal property of the center of masses is fulfilled in the planar case also. It is consequence of an analog of the Huygens-Steiner Theorem for the case of the planar moments of inertia.

Proposition 1.1. [**16**] *Given a system of points in* $\mathbb{R}^3$ *with the total mass* m *and the center of masses* C*. If the planes* π *and* π_1 *are parallel and* π *contains the center of masses* C*, then*

$$J_{\pi_1} = J_\pi + mD^2,$$

where D *is the distance between the parallel planes* π *and* π_1*.*

An immediate and important conclusion follows from the coordinate expressions for the planar and axial operators of inertia J_O and I_O of the same system of points and with respect to the same point O: both operators have a diagonal form in the same orthogonal basis, called the principal basis. Using the Pythagorean theorem, one can see that the axial moment of inertia for the axis u is the sum of the planar moments of inertia for two orthogonal planes having the line u as their intersection. For example, for the principal axes, we have:

$$I_i = J_j + J_p, \tag{1.6}$$

where (i, j, p) is a cyclic permutation of $(1, 2, 3)$.

2. The classical results of Pearson and their generalizations

Karl Pearson was one of the founding fathers of modern statistics. He investigated the question of the hyperplane which minimizes the mean square distance from a given set of points in $\mathbb{R}^k$, for any $k \geq 3$. In his own words, Pearson formulated the problem [**34**]: *"In the case we are about to deal with, we suppose the observed variables–all subject to error–to be plotted in plane, three-dimensioned or higher space, and we endeavour to take a line (or plane) which will be the 'best fit' to such a system of points. Of course the term 'best fit' is really arbitrary; but a good fit will clearly be obtained if we make the sum of the squares of the perpendiculars from the system of points upon the line or plane a minimum."*

Pearson noticed that the notion of the best fit is not uniquely determined. In the measurement error models, also known as regression with errors in variables (EIV), [**9**], a natural choice is the choice of the squares of the perpendiculars. This was indicated in Pearson's above note: "we suppose the observed variables–all subject to error". Such models are in use when the both predictors and responses are known with some error. In the usual linear regressions, the situation is different: for the predictors the exact values are known. Only responses are assumed to be known with some error. Thus, the squares of distances along one of the axes is in use in such regression models. We called that *directional regression* in a given direction. We talked more about its geometric aspects in the last section of [**16**].

Here we are going to explain in more details the classical simple linear regression model and the regression with error in variables (EIV) models, [**9**]. For a classical simple regression model, it is assumed that the values $(x^{(i)})_{i=1}^N$ are known, fixed values, as for example values set up in advance in the experiment. The values $(y^{(i)})_{i=1}^N$ are observed values of uncorrelated random variables Y_i, $i = 1, \dots, N$ with the same variance σ^2. It is assumed that the predictors $x^{(i)}$ and responses $(y^{(i)})_{i=1}^N$ are related with a linear

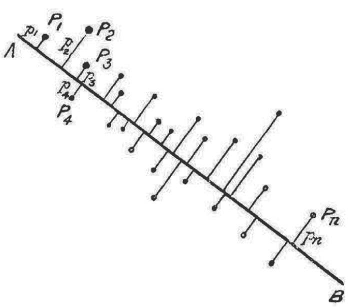

FIGURE 6. From [**34**].

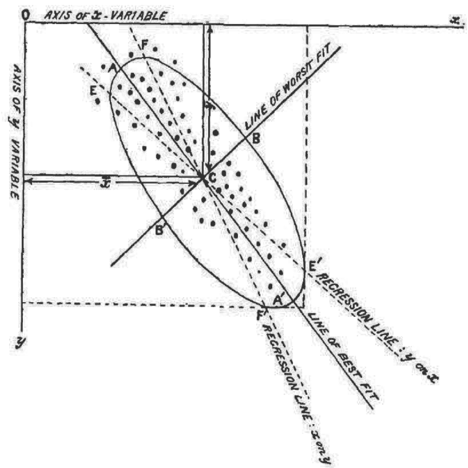

FIGURE 7. From [**34**].

relationship:

$$EY_i = \alpha + \beta x^{(i)}, \quad i = 1, \dots, N.$$

This can be restated as

$$Y_i = \alpha + \beta x^{(i)} + \epsilon_i, \quad i = 1, \dots, N,$$

where ϵ_i are called *the random errors* and they are uncorrelated random variables with zero expectation and the same variance σ^2. In such models the regression is of Y on x, i.e. in the vertical direction. This model was in the background of the Galton study [**26**], mentioned above.

Important situations appear when predictors are known only up to some error. They are described by measurement error models. There the observed pairs $(x^{(i)}, y^{(i)})_{i=1}^N$ are sampled from random variables (X_i, Y_i) with means satisfying the linear relationship

$$EY_i = \alpha + \beta(EX_i), \quad i = 1, \dots, N.$$

Denoting $EX_i = \xi_i$, the errors in variables model can be defined as

$$Y_i = \alpha + \beta\xi_i + \epsilon_i, \quad X_i = \xi_i + \delta_i, \quad i = 1, \dots, N,$$

where both X_i and Y_i have error terms which belong to mean zero normal distributions, such that all ϵ_i, $i = 1, \dots, N$ have the same variance σ_ϵ^2 and all δ_i, $i = 1, \dots, N$ have the same variance σ_δ^2. Since x_i and y_i are both known with an error, the orthogonal least square method is more natural to apply here, that is to say to apply the orthogonal regression. This we are going to introduce next under the assumption that $\eta = \sigma_\epsilon^2/\sigma_\delta^2 = 1$ and in a general dimension k. One should also mention that applications of the orthogonal least square method in the models with measurement errors have limitations, based on the fact that value of η could be unknown, see e.g. [**8**], [**7**]. Here we deals with the cases when *η is known*.

The case $\eta = 1$ historically originated from [**1, 2**]. Pearson established *orthogonal regression*, using the squares of the perpendiculars that corresponds to the case $\eta = 1$. Our aim is to study geometric aspects of the orthogonal regression. We also adopted Pearson's generality full rank assumption, that the given system of points is in a general position, which means that the points do not belong to a hyper-plane.

Let N points $(x_1^{(i)}, x_2^{(i)}, \dots, x_k^{(i)})_{i=1}^N$ be given. The mean values of the coordinates define *the centroid*:

$$\bar{x}_j = \frac{1}{N}\sum_{i=1}^N x_j^{(i)}, \quad j = 1, \dots, k.$$

The variances are

$$\sigma_{x_j}^2 = \frac{1}{N}\sum_{i=1}^N (x_j^{(i)} - \bar{x}_j)^2, \quad j = 1, \dots, k.$$

Due to the generality full rank assumption, all $\sigma_{x_j}^2$, for $j = 1, \dots, k$ are non-zero. Then, *the correlations* are

$$r_{jl} = \frac{p_{jl}}{\sigma_{x_j}\sigma_{x_l}}, \quad j, l = 1, \dots, k, l \neq j,$$

where

$$p_{jl} = \frac{1}{N}\sum_{i=1}^N (x_j^{(i)} - \bar{x}_j)(x_l^{(i)} - \bar{x}_l), \quad j, l = 1, \dots, k, l \neq j,$$

are *the covariances*. *The covariance matrix* K is a $(k \times k)$ matrix with the diagonal elements

$$K_{jj} = \sigma_{x_j}^2, \quad j = 1, \dots, k,$$

and the off-diagonal elements

$$K_{jl} = p_{jl}, \quad j, l = 1, \ldots, k, l \neq j.$$

The covariance matrix is always symmetric positive semidefinite. Here, due to the generality full rank assumption, K is a positive-definite matrix. In particular, it has the inverse K^{-1} and all its eigenvalues are positive. Pearson defined *the ellipsoid of residuals* by the equation

$$\sum_{j,l=1}^{k} K_{jl} x_j x_l = const.$$

Denote the eigenvalues of K as $\mu_1 \geq \cdots \geq \mu_k > 0$.

THEOREM 2.1. *[Pearson, [**34**]] The minimal mean square distance from a hyperplane to the given set of N points is equal to the minimal eigenvalue of the covariance matrix K. The best-fitting hyperplane contains the centroid and it is orthogonal to the corresponding eigenvector of K. Thus, it is the principal coordinate hyperplane of the ellipsoid of residuals which is normal to the major axis.*

Then Pearson studied the lines which best fit to the given set of points and proved

THEOREM 2.2. *[Pearson, [**34**]] The line which fits best the given system of N points contains the centroid and coincides with the minor axis of the ellipsoid of residuals.*

Pearson integrated the visualization of the linear regression with the orthogonal regression in the planar case in [**34**] in Fig. 7. The ellipse in Fig. 7 is dual to the ellipse of residuals and it coincides with the object studied by Galton.

The main results of [**16**] were to generalize the above classical results of Pearson in the following directions.

The first result: *For a given system of N points in $\mathbb{R}^k$, for any $k \geq 2$, under the full rank assumption, we consider all hyperplanes which equally fit the given system of points. In other words, for any fixed value s not less than the smallest eigenvalue μ_k of the covariance matrix K, we consider all hyperplanes for which the mean sum of square distances to the given set of points is equal to s. Starting from the ellipsoid of residuals, we are going to effectively construct a pencil of confocal quadrics with the following property: For each $s \geq \mu_k$ there exists a quadric from the confocal pencil which is the envelope of all the hyperplanes which s-fit the given system of points.*

We stress that the ellipsoid of residuals does not belong to the confocal family of quadrics. The construction of this confocal pencil of quadrics is fully effective, though quite involved. The obtained pencil of confocal quadrics have the same center as the ellipsoid of residuals and moreover, the same principal axes.

EXAMPLE 2.1. Let us recall that μ_k denotes the smallest eigenvalue of the covariance matrix K. In the case $s = \mu_k$ there is only one hyperplane which s fits the given set of N points. This is the best-fitting hyperplane described in Theorem 2.1. The envelope of this single hyperplane is this hyperplane itself. This hyperplane is going to be a degenerate quadric from our confocal pencil of quadrics.

The second result: *For a given system of N points in $\mathbb{R}^k$, for any $k \geq 2$, under the full rank assumption, find the best fitting hyperplane under the condition that it contains a selected point in $\mathbb{R}^k$. We also provide an answer to the questions of the best fitting line and more general the best fitting affine subspace of dimension ℓ, $1 \leq \ell \leq k-1$ under the condition that they contain a given point.*

A careful look at the Galton's figure (see Fig. 5) discloses an intriguing geometric fact that the line of linear regression of y on x intersects the ellipse at the points of vertical tangency, while the line of linear regression of x on y intersects the ellipse at the points of horizontal tangency. Further analysis of this phenomenon leads us to our third result.

The third result is to formulate linear regression in $\mathbb{R}^k$ in a coordinate free, i.e. in an invariant form. We answered the following question: *for a given direction and for a given system of N points, under the generality full rank assumption, what is the best fitting hyperplane in the given direction, under the condition that it contains a selected point in $\mathbb{R}^k$.*

Apparently, the second and the third result are obtained using the same confocal pencil of quadrics constructed in relation with the first result.

3. Conclusion

For a given system of points in $\mathbb{R}^k$, under the full rank assumption, we constructed in [**16**] an explicit pencil of confocal quadrics with the following properties:

(i) All the hyperplanes for which the hyperplanar moments of inertia for the given system of points are equal, are tangent to one of the quadrics from the pencil of quadrics. As an application, we developed regularization procedures for the orthogonal least square methods, analogues of lasso and ridge methods from linear regression. Another motivation for this study was the gradient descent methods in machine learning. An optimization algorithm may not be guaranteed to arrive at the minimum in a reasonable amount of time. As pointed out in e.g. [**28**] it often reaches some quite low value of the cost function equal to some value s_0 quickly enough to be useful. Here we deal with the hyperplanar moment as the cost function, in application to the orthogonal least square. From [**16**] we know that the hyperplanes which generate the hyperplanar moment equal to s_0 are all tangent to the given quadric from the confocal pencil of quadrics, where the pencil parameter is determined through the value s_0.

(ii) For any given point P among all the hyperplanes that contain it, the best fit is the tangent hyperplane to the quadric from the confocal pencil corresponding to the maximal Jacobi elliptic coordinate of the point P. The worst fit among the hyperplanes containing P is the tangent hyperplane to the ellipsoid from the confocal pencil that contains P. We also determined the best and worst fit among ℓ-dimensional planes containing P, for any $\ell : 1 \leq \ell \leq k-1$.

Both results (i) and (ii) are generalizations of the famous theorem of Pearson on orthogonal regression [**34**], or in other words on the orthogonal least square method (see e.g. [**9**]). For the original Pearson's statement, see Theorem 2.1 above. It is well known that the Pearson result also initiated the Principal Component Analysis (PCA), see e.g. [**3**]. Similarly, our results have a natural interpretation in terms of PCA. The confocal pencil of quadrics provides a universal tool to solve the Restricted Principal Component Analysis restricted at any given point which we formulated and solved in [**16**]. Our generalizations of the Pearson Theorem have natural and important applications in the statistics of the measurement error models, for which the orthogonal regression is known to provide a natural framework, see [**9**], [**24**], [**8**].

Acknowledgements. We are grateful to Maxim Arnold for indicating the reference [**27**] and for interesting discussions. We are also grateful to Pankaj Choudhary, Sam Efromovich, and Frank Konietschke for interesting discussions. This research has been partially supported by Mathematical Institute of the Serbian Academy of Sciences and Arts, the Science Fund of Serbia grant Integrability and Extremal Problems in Mechanics, Geometry and Combinatorics, MEGIC, Grant No. 7744592 and the Ministry for Education, Science, and Technological Development of Serbia and the Simons Foundation grant no. 854861.

References

[1] Adcock, R. J., *Note on the method of least squares.* Analyst 4, 183–184, (1877).

[2] Adcock. R. J., *A problem in least squares.* Analyst 5, 53–54, (1878).

[3] Andreson, T. W., *An Introduction to Multivariate Statistical Analysis* John Wiley and Sons, 1958

[4] Arnold, V. I., *Mathematical methods of classical mechanics*, Graduate Texts in Mathematics, **60**, 2nd edition, Springer-Verlag, New York, 1989, pp. xvi+508

[5] Audin, M., *Courbes algébriques et systèmes intégrables: géodesiques des quadriques*, Exposition. Math. **12**, (1994) pp. 193–226.

[6] Bialy, M., Tabachnikov, S., *Dan Reznik's identities and more*, arXiv:2001.08469, European Journal of Mathematics, , 10.1007/s40879-020-00428-7, (2020).

[7] Carroll, R. J., Ruppert, D. *The use and misuse of orthogonal regression estimation in linear errors-in-variables models.* The American Statistician, 50, 1–6, (1996).

[8] Carroll, R., Ruppert, D., Stefanski, L., Crainiceanu, C., *Measurement Error in Nonlinear Models, A Modern Perspective* Second Edition, Monographs on Statistics and Applied Probability 105, Taylor and Francis (2006).

[9] Casella, G., Berger, R. L., *Statistical Inference*, Second Edition, Duxbury Advanced Series, Duxbury

[10] Cohen, J. E., D'Eustachio, P., and Edelman, G. M., *The specific antigenbinding cell populations of individual fetal mouse spleens: Repertoire composition, size and genetic control.* J. Exp. Med. 146. 394–411, (1977)

[11] Cohen, J. E. and D'Eustachio, P., *An affine linear model for the relation between two sets of frequency counts. Response to a query.* Biometrics 34,514–516, (1978).

[12] Cramér, H., *Mathematical Methods of Statistics,* Princeton: Princeton University Press, (1946).

[13] DeGroot, M., Schervish, M., *Probability and Statistics*, Fourth Edition, Adison-Wesley, 2012.

[14] Dempster, A., *Elements of continuous multivariate analysis*, Addison-Wesley, 1969.

[15] Dragović, V. Gajić, B. (2023) Points with rotational ellipsoids of inertia, envelopes of hyperplanes which equally fit the system of points in R^k, and ellipsoidal billiards, Physica D: Nonlinear Phenomena, Volume 451, 133776

[16] Dragović, V., Gajić, *Orthogonal and Linear Regressions and Pencils of Confocal Quadrics*, preprint, arXiv: 2209.01679, 2022.

[17] Dragović, V., Radnović, M., *Hyperelliptic Jacobians as billiard algebra of pencils of quadrics: Beyond Poncelet porisms*, Advances in Mathematics, Vol. 219, 2008, pp. 1577–1607.

[18] Dragović, V., Radnović, M., *Poncelet Porisms and Beyond: Integrable Billiards, Hyperelliptic Jacobians and Pencils of Quadrics*, Frontiers in Mathematics, Basel: Springer, (2011).

[19] Fedorov, Y. N., Kozlov, V. V., *Various aspects of n-dimensional rigid body dynamics*, Translations of the American Mathematical Society-Series 2, 168, 141–172.

[20] T. B. Fomby, R. C. Hill, S. R. Johnson, *Advanced Econometric Methods*, (1984) Springer Science + Business Media.

[21] Forbes, J. D., *Further experiments and remarks on the measurement of heights by the boiling point of water*, Transactions of the Royal Society of Edinburgh, 21: 135–143, (1857).

[22] Friendly, M., Monette, G., Fox, J,. *Elliptic insights: understanding statistical methods through elliptical geometry.* Statist. Sci. 28 (2013), no. 1, 1–39.

[23] Fuller, W. A., *An affine linear model for the relation between two sets of frequency counts: Response to query.* Biometrics 34, 517–521, (1978).

[24] Fuller, W. A., *Measurement Error Models*, John Willey and Sons, 1987.

[25] Gajić, B., *The rigid body dynamics: classical and algebro-geometric integration*, Zb. Rad.(Beogr.), 16(24), (2013), 5–44.

[26] Galton, F. *Regression towards mediocrity in hereditary stature*, Journal of the Anthropological Institute, 15, (1886) pp. 246–263.

[27] Gardiner, M., *On "Confocal Quadrics of Moments of Inertia" pertaining to all Planes in Space, and Loci and Envelopes of Straight Lines whose "Moments of Inertia" are Constants*, Proceedings of the Royal Society of Victoria, 5, (1893), 200–208.

[28] Goodfellow, I., Bengio, Y., Courville, A., *Deep learning* The MIT Press, 2016.

[29] Hastie, T., Tibshirani, R., Friedman, J, *The elements of statistical learning* Second Edition, Springer

[30] Hilbert, D., Cohn-Vossen, S., *Anschauliche Geometri*, Springer, 1932. English translation: *Geometry and the imagination*. Translated by P. Neményi. Chelsea Publishing Co., New York, N. Y., 1952. ix+357 pp.

[31] Jacobi, C., *Jacobi's Lectures on Dynamics*, Springer, TRIM 51, 2009.

[32] Kozlov, V. V., Treshchev, D. V., *Billiards: A Genetic Introduction to the Dynamics of Systems with Impacts*, Translations of Mathematical Monographs, 89. Providence, RI: Amer. Math. Soc., (1991).

[33] Newton, I., *Philosophia Naturalis Principia Mathematica*, (1st edition 1687, 2nd edition 1713, 3rd edition 1726).

[34] Pearson, K,. *On Lines and Planes of Closest Fit to Systems of Points in Space*, Phil. Mag. 2 (1901), 559–572.

[35] Prasolov, V. V., Tikhomirov, V. M., *Geometry*, Translations of Mathematical Monographs, Vol. 200, AMS, (2001).

[36] Rogers, A. J., *Concentration Ellipsoids, Their Planes of Support, and the Linear Regression Model*, Econometric Reviews, 32:2, 220-243, DOI: 10.1080/07474938.2011.608055, (2013).

[37] Seber, G., Lee, A., *Linear Regression Analysis* Second Edition, John Wiley and Sons, 2003.

[38] Stachel, H., *Recalling Thread Construction of Quadrics*, J. Geom. Graph., 24, No 2, 143–157.

[39] Suslov, G.K., *Fundamentals of Anaytical Mechanics, Vol.1*, Kiev, 1900, [in Russian].

[40] Vinberg, E. B. *A Course in Algebra* Graduate Studies in Mathematics, Vol. 16, AMS.

[41] Weisberg, S., *Applied Linear Regression*, Second Edition, John Willey and Sons, 1985.