Linear Regression with Optimal Rotation

George Livadiotis

Southwest Research Institute, Space Science & Engineering, San Antonio, TX 78238, USA; glivadiotis@swri.edu;
Tel.: +1-210-274-4028

Abstract: The paper shows how the linear regression depends on the selection of the reference frame. The slope of the fitted line and the corresponding Pearson's correlation coefficient are expressed in terms of the rotation angle. The correlation coefficient is found to be maximized for a certain optimal angle, for which the slope attains a special optimal value. The optimal angle, the value of the optimal slope, and the corresponding maximum correlation coefficient were expressed in terms of the covariance matrix, but also in terms of the values of the slope, derived from the fitting at the nonrotated and right-angle-rotated axes. The potential of the new method is to improve the derived values of the fitting parameters by detecting the optimal rotation angle, that is, the one that maximizes the correlation coefficient. The presented analysis was applied to the linear regression of density and temperature measurements characterizing the proton plasma in the inner heliosheath, the outer region of our heliosphere.

1. Introduction

The fitting of a given dataset $\left\{y_i \pm \sigma_{yi}\right\}_{i=1}^{N}$ to the values $\{V_i\}_{i=1}^{N}$ of a statistical model $V(X; \alpha)$ in the domain $X \in D_x \subseteq \mathfrak{R}$ [1–5] involves finding the optimal parameter value $\alpha = \alpha^*$ in $\alpha \in D_\alpha \subseteq \mathfrak{R}$ that minimizes the sum of squared residuals, also known in physical sciences as total square deviations (*TSD*) between model and data,

$$TSD(\alpha)^2 = \sum_{i=1}^{N} \sigma_{yi}^{-2}[y_i - V(x_i; \alpha)]^2, \tag{1}$$

where the inverse of variance of the data measurements $\left\{w_i = \sigma_{yi}^{-2}\right\}_{i=1}^{N}$ is weighting the summation. In the case of the multiparametrical fitting, we consider that the statistical model depends on n independent parameters, $\{\alpha_k\}_{k=1}^{n}$. Then, *TSD* becomes (e.g., [6–9]):

$$TSD(\{\alpha_k\}_{k=1}^{n})^2 = \sum_{i=1}^{N} \sigma_{yi}^{-2}[y_i - V(x_i; \{\alpha_k\}_{k=1}^{n})]^2, \tag{2}$$

or, for a biparametrical (α and β) linear regression, we have:

$$TSD(\alpha, \beta)^2 = \sum_{i=1}^{N} (y_i - \alpha - \beta x_i)^2. \tag{3}$$

Minimization of *TSD* leads to the optimal value of the involved parameters (least-square method), that is,

$$\beta_0 = \frac{\sigma_{xy}^2}{\sigma_{xx}^2}, \ \alpha = \overline{y} - \beta \cdot \overline{x}. \tag{4}$$

The optimal fitting value of the slope β is noted as β_0. The optimal value of the slope β for the inverse linear regression, that is, between (y,x), is:

$$\beta_{90} = \frac{\sigma_{yy}^2}{\sigma_{xy}^2}, \tag{5}$$

where the ratio β_0/β_{90} equals the correlation of these linear regressions, which is maximized for $\beta_0 = \beta_{90}$.

In this paper, we focus on the optimal fitting value of the slope β, derived from the linear regression of data expressed on the rotated Cartesian axes. The slope β and the corresponding Pearson's correlation coefficient are calculated as a function of the rotation angle ϑ, while the optimal fitting value of the slope is derived from the maximization of the correlation coefficient. Finally, we expose an application from the discipline of space physics. In particular, the presented analysis was applied to the linear regression of density and temperature measurements characterizing the proton plasma in the inner heliosheath, the outer region of our heliosphere. Those measurements can be fitted with a line of negative slope, which gives the polytropic index of the plasma. The new statistical method improves the applied linear regression and the estimated value of the polytropic index.

2. Rotated Optimal Slope

The orthogonal Cartesian system is rotated by an angle ϑ, according to

$$\begin{pmatrix} \widetilde{x} \\ \widetilde{y} \end{pmatrix} = \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{6}$$

Then, the total square deviations are given by:

$$TSD(\widetilde{\alpha}, \widetilde{\beta})^2 = \sum_{i=1}^{N} \left(\widetilde{y}_i - \widetilde{\alpha} - \widetilde{\beta}\widetilde{x}_i \right)^2. \tag{7}$$

Next, we express the parameter values, the slope $\widetilde{\beta}$ and intercept $\widetilde{\alpha}$, in terms of their values at the nonrotated system. First, we observe that the slope is transformed as follows:

$$\frac{d\widetilde{y}}{d\widetilde{x}} = \frac{\cos\vartheta + \sin\vartheta \cdot \frac{dy}{dx}}{-\sin\vartheta + \cos\vartheta \cdot \frac{dy}{dx}}, \tag{8}$$

hence,

$$\widetilde{\beta} = \frac{\cos\vartheta + \sin\vartheta \cdot \beta}{-\sin\vartheta + \cos\vartheta \cdot \beta}. \tag{9}$$

In order to find the transform intercept, we consider the yellow-shaded triangle in Figure 1, where we have:

$$\beta = \frac{a - \widetilde{a} \cdot \cos\vartheta}{\widetilde{a} \cdot \sin\vartheta}, \tag{10}$$

thus, we end up with:

$$\widetilde{a} = \frac{a}{\cos\vartheta + \sin\vartheta \cdot \beta}. \tag{11}$$
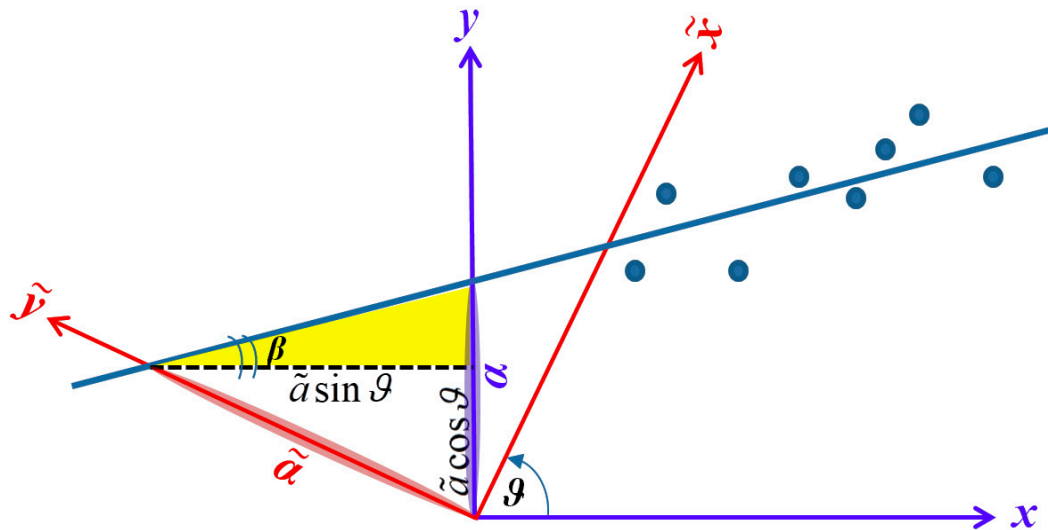
Figure 1. Scheme of the rotated axes and the corresponding symbols. The angle noted by the symbol "))" in the yellow-shaded triangle has its tangent equal to β and is related to Equation (10).

Therefore, the total square deviations in the rotated reference frame is given by:

$$TSD(\widetilde{\alpha},\widetilde{\beta})^2 = \left(\frac{TSD(\alpha,\beta)}{\cos\vartheta + \sin\vartheta \cdot \beta}\right)^2 = \sum_{i=1}^{N}\left(\frac{y_i - \alpha - \beta \cdot x_i}{\cos\vartheta + \sin\vartheta \cdot \beta}\right)^2, \tag{12}$$

(with $\cos\vartheta \neq -\beta$). Hence, the minimization of TSD can be still expressed in terms of the linear parameters α and β of the nonrotated frame; namely, we first derive the normal equation with respect to α:

$$0 = \frac{\partial}{\partial\alpha}\sum_{i=1}^{N}\left(\frac{y_i-\alpha-\beta\cdot x_i}{\cos\vartheta+\sin\vartheta\cdot\beta}\right)^2 = -\frac{2}{(\cos\vartheta+\sin\vartheta\cdot\beta)^2}\cdot\sum_{i=1}^{N}(y_i-\alpha-\beta\cdot x_i)$$
$$= -\frac{2N}{(\cos\vartheta+\sin\vartheta\cdot\beta)^2}\cdot(\overline{y}-\alpha-\beta\cdot\overline{x}) = 0, \tag{13}$$

leading to

$$\overline{y} - \alpha - \beta \cdot \overline{x} = 0. \tag{14}$$

Next, we derive the normal equation with respect to β:

$$0 = \frac{\partial}{\partial\beta}\sum_{i=1}^{N}\left(\frac{y_i-\alpha-\beta\cdot x_i}{\cos\vartheta+\sin\vartheta\cdot\beta}\right)^2 = \frac{2}{(\cos\vartheta+\sin\vartheta\cdot\beta)^3}$$
$$\times\sum_{i=1}^{N}(y_i-\alpha-\beta\cdot x_i)[\cos\vartheta\cdot x_i + \sin\vartheta\cdot(y_i-\alpha)]$$
$$= \frac{2N}{(\cos\vartheta+\sin\vartheta\cdot\beta)^3}$$
$$\times\sum_{i=1}^{N}\left[(\cos\vartheta-\beta\cdot\sin\vartheta)\cdot x_i(y_i-\alpha)-\beta\cdot\cos\vartheta\cdot x_i^2 + \sin\vartheta\cdot(y_i-\alpha)^2\right]$$
$$= \frac{2N}{(\cos\vartheta+\sin\vartheta\cdot\beta)^3}$$
$$\times[(\cos\vartheta-\beta\cdot\sin\vartheta)\cdot\overline{x(y-\alpha)}-\beta\cdot\cos\vartheta\cdot\overline{x^2} + \sin\vartheta\cdot\overline{(y-\alpha)^2}] = 0, \tag{15}$$

or

$$
\begin{aligned}
&(\cos\vartheta - \beta\cdot\sin\vartheta)\cdot\overline{x(y-\alpha)} - \beta\cdot\cos\vartheta\cdot\overline{x^2} + \sin\vartheta\cdot\overline{(y-\alpha)^2} \\
&= (\cos\vartheta - \beta\cdot\sin\vartheta)\cdot\overline{(x-\overline{x})(y-\overline{y})} + (\cos\vartheta - \beta\cdot\sin\vartheta)\cdot\overline{x}(\overline{y}-\alpha) \\
&\quad -\beta\cdot\cos\vartheta\cdot\overline{x^2} + \sin\vartheta\cdot\overline{(y-\overline{y})^2} + \sin\vartheta\cdot(\overline{y}-\alpha)^2 \\
&= (\cos\vartheta - \beta\cdot\sin\vartheta)\cdot\overline{(x-\overline{x})(y-\overline{y})} + (\beta\cos\vartheta - \beta^2\cdot\sin\vartheta)\cdot\overline{x}^2 \\
&\quad -\beta\cdot\cos\vartheta\cdot\overline{x^2} + \sin\vartheta\cdot\overline{(y-\overline{y})^2} + \beta^2\cdot\sin\vartheta\cdot\overline{x}^2 \\
&= (\cos\vartheta - \beta\cdot\sin\vartheta)\cdot\overline{(x-\overline{x})(y-\overline{y})} - \beta\cdot\cos\vartheta\cdot\overline{(x-\overline{x})^2} + \sin\vartheta\cdot\overline{(y-\overline{y})^2} \\
&= (\cos\vartheta - \beta\cdot\sin\vartheta)\cdot\sigma_{xy}^2 - \beta\cdot\cos\vartheta\cdot\sigma_{xx}^2 + \sin\vartheta\cdot\sigma_{yy}^2 = 0.
\end{aligned}
\tag{16}
$$

Hence, we have

$$
\beta = \beta(\vartheta) = \frac{\sigma_{xy}^2 + \sigma_{yy}^2\cdot\tan\vartheta}{\sigma_{xx}^2 + \sigma_{xy}^2\cdot\tan\vartheta},
\tag{17}
$$

or

$$
\beta(\vartheta) = \frac{1 + \frac{\sigma_{yy}^2}{\sigma_{xy}^2}\cdot\tan\vartheta}{\frac{\sigma_{xx}^2}{\sigma_{xy}^2} + \tan\vartheta} = \frac{1 + \beta_{90}\cdot\tan\vartheta}{\beta_0^{-1} + \tan\vartheta}.
\tag{18a}
$$

We observe that $\beta = \beta_0$ for $\vartheta = 0$ and $\beta = \beta_{90}$ for $\vartheta = 90^0$.

Finally, we note that $\widetilde{\beta}(\vartheta)$ is the slope of the fitted curve derived from the fitting performed in the rotated frame, while $\beta(\vartheta)$ is the expression of the slope $\widetilde{\beta}(\vartheta)$ in the nonrotated frame. One may find $\widetilde{\beta}(\vartheta)$, that is, the slope expressed in the rotated frame, simply by substituting Equation (18) in Equation (9):

$$
\widetilde{\beta} = \frac{1 + \tan\vartheta\cdot\beta}{-\tan\vartheta + \beta}.
\tag{18b}
$$

3. Maximized Correlation Coefficient

The Pearson's correlation coefficient is derived as follows. First, we construct the covariance and variances, and then the correlation coefficient at the rotated reference frame. We have:

$$
\overline{\widetilde{x}\,\widetilde{y}} = \sin\vartheta\cos\vartheta\cdot\left(\overline{y^2} - \overline{x^2}\right) + (\cos^2\vartheta - \sin^2\vartheta)\cdot\overline{x\,y},
\tag{19}
$$

$$
\overline{\widetilde{x}}\,\overline{\widetilde{y}} = \sin\vartheta\cos\vartheta\cdot\left(\overline{y}^2 - \overline{x}^2\right) + (\cos^2\vartheta - \sin^2\vartheta)\cdot\overline{x}\,\overline{y},
\tag{20}
$$

thus, we obtain:

$$
\widetilde{\sigma}_{xy}^2 = \sin\vartheta\cos\vartheta\cdot\left(\sigma_{yy}^2 - \sigma_{xx}^2\right) + (\cos^2\vartheta - \sin^2\vartheta)\cdot\sigma_{xy}^2.
\tag{21}
$$

In similar, we have:

$$
\widetilde{\sigma}_{xx}^2 = \cos^2\vartheta\cdot\sigma_{xx}^2 + \sin^2\vartheta\cdot\sigma_{yy}^2 + 2\sin\vartheta\cos\vartheta\cdot\sigma_{xy}^2,
\tag{22}
$$

$$
\widetilde{\sigma}_{yy}^2 = \sin^2\vartheta\cdot\sigma_{xx}^2 + \cos^2\vartheta\cdot\sigma_{yy}^2 - 2\sin\vartheta\cos\vartheta\cdot\sigma_{xy}^2.
\tag{23}
$$

Hence, the Pearson's correlation coefficient for the normal and the rotated frames:

$$
r_0{}^2 = \frac{\sigma_{xy}^4}{\sigma_{xx}^2\sigma_{yy}^2}, \quad r^2(\vartheta) = \frac{\widetilde{\sigma}_{xy}^4}{\widetilde{\sigma}_{xx}^2\widetilde{\sigma}_{yy}^2}.
\tag{24}
$$

Note: While the covariance cov(x,y) or the Pearson's correlation coefficient can be either positive or negative, their algebraic sign has no effect on the method of correlation maximization presented here. Indeed, the examined datasets may be correlated, $r > 0$, or anti-correlated, $r < 0$, but it is important to ask just how big is the correlation or anticorrelation; hence, the meaningful quantity needed to be maximized is the absolute correlation or the square of the correlation, r^2. For the same reason, the

sign of the covariance does not affect the presented analysis, and it is ignored. For this, we denote the covariance with the square symbol σ_{xy}^2, though it may have negative sign.

We observe that

$$
\begin{aligned}
\widetilde{\sigma}_{xx}^2 \widetilde{\sigma}_{yy}^2 &= \left[\cos^2 \vartheta \cdot \sigma_{xx}^2 + \sin^2 \vartheta \cdot \sigma_{yy}^2 + 2\sin \vartheta \cos \vartheta \cdot \sigma_{xy}^2\right] \\
&\quad \times \left[\sin^2 \vartheta \cdot \sigma_{xx}^2 + \cos^2 \vartheta \cdot \sigma_{yy}^2 - 2\sin \vartheta \cos \vartheta \cdot \sigma_{xy}^2\right] \\
&= \sin^2 \vartheta \cos^2 \vartheta \cdot \left(\sigma_{xx}^4 + \sigma_{yy}^4 - 4\sigma_{xy}^4\right) + (1 - 2\sin^2 \vartheta \cos^2 \vartheta) \cdot \sigma_{xx}^2 \sigma_{yy}^2 \\
&\quad + 2\sin \vartheta \cos \vartheta (\cos^2 \vartheta - \sin^2 \vartheta) \cdot \sigma_{xy}^2 (\sigma_{yy}^2 - \sigma_{xx}^2).
\end{aligned}
\tag{25}
$$

Then, we have

$$
\begin{aligned}
\widetilde{\sigma}_{xy}^4 &= \left[\sin \vartheta \cos \vartheta \cdot \left(\sigma_{yy}^2 - \sigma_{xx}^2\right) + (\cos^2 \vartheta - \sin^2 \vartheta) \cdot \sigma_{xy}^2\right]^2 \\
&= (\cos^2 \vartheta - \sin^2 \vartheta)^2 \cdot \sigma_{xy}^4 + \sin^2 \vartheta \cos^2 \vartheta \cdot \left(\sigma_{xx}^4 + \sigma_{yy}^4 - 2\sigma_{xx}^2 \sigma_{yy}^2\right) \\
&\quad + 2\sin \vartheta \cos \vartheta (\cos^2 \vartheta - \sin^2 \vartheta) \cdot \left(\sigma_{yy}^2 - \sigma_{xx}^2\right)\sigma_{xy}^2 \\
&= \sigma_{xy}^4 - \sigma_{xx}^2 \sigma_{yy}^2 + \widetilde{\sigma}_{xx}^2 \widetilde{\sigma}_{yy}^2,
\end{aligned}
\tag{26}
$$

where we used the identity $\sin^4 \vartheta + \cos^4 \vartheta = 1 - 2\sin^2 \vartheta \cos^2 \vartheta$. Hence, we find the invariant:

$$
\widetilde{\sigma}_{xy}^4 - \widetilde{\sigma}_{xx}^2 \widetilde{\sigma}_{yy}^2 = \sigma_{xy}^4 - \sigma_{xx}^2 \sigma_{yy}^2.
\tag{27}
$$

Then, the correlation coefficient at the rotated system becomes:

$$
\begin{aligned}
r^2(\vartheta) - 1 &= \frac{\widetilde{\sigma}_{xy}^4 - \widetilde{\sigma}_{xx}^2 \widetilde{\sigma}_{yy}^2}{\widetilde{\sigma}_{xx}^2 \widetilde{\sigma}_{yy}^2} = \frac{\sigma_{xy}^4 - \sigma_{xx}^2 \sigma_{yy}^2}{\widetilde{\sigma}_{xx}^2 \widetilde{\sigma}_{yy}^2} \\
&= \frac{\sigma_{xy}^4 - \sigma_{xx}^2 \sigma_{yy}^2}{\sigma_{xx}^2 \sigma_{yy}^2} \cdot \frac{\sigma_{xx}^2 \sigma_{yy}^2}{\widetilde{\sigma}_{xx}^2 \widetilde{\sigma}_{yy}^2} = (r_0^2 - 1) \cdot \frac{\sigma_{xx}^2 \sigma_{yy}^2}{\widetilde{\sigma}_{xx}^2 \widetilde{\sigma}_{yy}^2},
\end{aligned}
\tag{28}
$$

or

$$
\frac{1 - r_0^2}{1 - r^2(\vartheta)} = \frac{\widetilde{\sigma}_{xx}^2 \widetilde{\sigma}_{yy}^2}{\sigma_{xx}^2 \sigma_{yy}^2}.
\tag{29}
$$

Hence, we first calculate the right-hand side of Equation (29):

$$
\begin{aligned}
\frac{\widetilde{\sigma}_{xx}^2 \widetilde{\sigma}_{yy}^2}{\sigma_{xx}^2 \sigma_{yy}^2} &= 1 + \sin^2 \vartheta \cos^2 \vartheta \cdot \left(\frac{\sigma_{xx}^2}{\sigma_{yy}^2} + \frac{\sigma_{yy}^2}{\sigma_{xx}^2} - 2 - 4r_0^2\right) \\
&\quad + 2\sin \vartheta \cos \vartheta (\cos^2 \vartheta - \sin^2 \vartheta) \cdot r_0 \cdot \left(\frac{\sigma_{yy}}{\sigma_{xx}} - \frac{\sigma_{xx}}{\sigma_{yy}}\right),
\end{aligned}
\tag{30}
$$

thus, Equation (29) becomes:

$$
\begin{aligned}
\left[\frac{1 - r_0^2}{1 - r^2(\vartheta)} - 1\right] \cdot \frac{1}{r_0^2} &= 4\sin^2 \vartheta \cos^2 \vartheta \cdot \left\{\left[\frac{1}{2r_0}\left(\frac{\sigma_{yy}}{\sigma_{xx}} - \frac{\sigma_{xx}}{\sigma_{yy}}\right)\right]^2 - 1\right\} \\
&\quad + 4\sin \vartheta \cos \vartheta (\cos^2 \vartheta - \sin^2 \vartheta) \cdot \left[\frac{1}{2r_0}\left(\frac{\sigma_{yy}}{\sigma_{xx}} - \frac{\sigma_{xx}}{\sigma_{yy}}\right)\right].
\end{aligned}
\tag{31}
$$

Using trigonometric functions of angle 4ϑ, we easily find:

$$
\begin{aligned}
\left[\frac{1 - r_0^2}{1 - r^2(\vartheta)} - 1\right] \cdot \frac{1}{r_0^2} &= \sin 4\vartheta \cdot \left[\frac{1}{2r_0}\left(\frac{\sigma_{yy}}{\sigma_{xx}} - \frac{\sigma_{xx}}{\sigma_{yy}}\right)\right] \\
&\quad + \frac{1}{2}(1 - \cos 4\vartheta) \cdot \left\{\left[\frac{1}{2r_0}\left(\frac{\sigma_{yy}}{\sigma_{xx}} - \frac{\sigma_{xx}}{\sigma_{yy}}\right)\right]^2 - 1\right\}.
\end{aligned}
\tag{32}
$$

This can be rewritten in a compact way:

$$
\left[\frac{1 - r_0^2}{1 - r^2(\vartheta)} - 1\right] \cdot \frac{1}{r_0^2} = \frac{1}{2}(1 - \cos 4\vartheta) \cdot (X^2 - 1) + \sin 4\vartheta \cdot X,
\tag{33}
$$

where

$$X \equiv \frac{\sigma_{yy}^2 - \sigma_{xx}^2}{2\sigma_{xy}^2} = \frac{1}{2}(\beta_{90} - \beta_0^{-1}) . \tag{34}$$

Given $\sin 4\vartheta = 2\tan 2\vartheta/(1 + \tan^2 2\vartheta)$ and $\cos 4\vartheta = (1 - \tan^2 2\vartheta)/(1 + \tan^2 2\vartheta)$, we rewrite the correlation coefficient as:

$$\left[\frac{1 - r_0^2}{1 - r^2(\vartheta)} - 1 \right] \cdot \frac{1}{r_0^2} = \frac{(\tan 2\vartheta \cdot X + 1)^2}{\tan^2 2\vartheta + 1} - 1 . \tag{35}$$

Note that, as expected, for $\vartheta = 0$ and $\vartheta = 90^0$, the correlation coefficient remains the same, $r_{90}^2 = r_0^2$.

The correlation coefficient is maximized [10] for the optimal angle $\vartheta = \vartheta_{\text{opt}}$:

$$\frac{\partial}{\partial \vartheta} \left[\frac{1 - r_0^2}{1 - r^2(\vartheta)} - 1 \right] \cdot \frac{1}{r_0^2} = 2\sin 4\vartheta \cdot (X^2 - 1) + 4\cos 4\vartheta \cdot X = 0 , \tag{36}$$

or

$$\tan 4\vartheta_{\text{opt}} = \frac{2X}{1 - X^2} . \tag{37}$$

Given the half-tangent formula $\tan 2\vartheta = 2\tan\vartheta/(1 - \tan^2\vartheta)$, we derive the solutions:

$$\text{(i)} \ \tan 2\vartheta_{\text{opt}} = X \ \text{and (ii)} \ \tan 2\vartheta_{\text{opt}} = -\frac{1}{X} \tag{38}$$

or

$$\text{(i)} \ \tan\vartheta_{\text{opt}} = (-1 \pm \sqrt{1 + X^2})/X \ \text{and (ii)} \ \tan\vartheta_{\text{opt}} = X \pm \sqrt{1 + X^2} , \tag{39}$$

where the subscript "opt" denotes the optimal rotation angle, ϑ_{opt}, for which the correlation is maximized (i.e., $r_{\text{opt}}^2 \equiv r^2(\vartheta_{\text{opt}})$). Then, from Equation (35), we derive the respective optimal correlation coefficient for each of the two solutions:

- For the (i) solution, we have a maximum of r^2:

$$\left(\frac{1 - r_0^2}{1 - r^2} - 1 \right)_{\text{opt}} \cdot \frac{1}{r_0^2} = X^2 \ \text{or} \left(\frac{1 - r_0^2}{1 - r^2} \right)_{\text{opt}} = 1 + r_0^2 X^2 , \tag{40}$$

and solving in terms of r_{opt}^2:

$$r_{\text{opt}}^2 = r_0^2 \cdot \frac{1 + X^2}{1 + r_0^2 X^2} . \tag{41}$$

We observe that the maximized correlation coefficient at the rotated system ranges from $r_{\text{opt}}^2 = r_0^2$ for $X = 0$ to $r_{\text{opt}}^2 = 1$ for $X \to \infty$. The case $r_{\text{opt}}^2 = r_0^2$ is consistent with $\vartheta = 0$; hence, we expect for $X = 0$ to have $\vartheta = 0$. The positive root in the (i) solution gives $\vartheta \to 0$ for $X \to 0$, which is consistent with $r_{\text{opt}}^2 = r_0^2$. However, the negative root gives $\vartheta \to 45^0$ for $X \to 0$; thus, the negative sign case must be rejected.

- For the (ii) solution, we have a minimum of r^2:

$$\left(\frac{1 - r_0^2}{1 - r^2} - 1 \right)_{\text{opt}} \cdot \frac{1}{r_0^2} = -1 \ \text{or} \left(\frac{1 - r_0^2}{1 - r^2} \right)_{\text{opt}} = 1 - r_0^2 , \ \text{or} \ r_{\text{opt}}^2 = 0 . \tag{42}$$

We seek the maximum correlation, so we reject the minimum of the (ii) solution and Equation (42). Therefore, the only acceptable solution is (i):

$$\tan\vartheta_{\text{opt}} = (-1 + \sqrt{1 + X^2})/X . \tag{43}$$

Solving Equation (18) in terms of $\tan\vartheta$, we obtain:

$$\beta_{opt} = \beta_0 \cdot \frac{1 + \beta_{90} \cdot \tan\vartheta_{opt}}{1 + \beta_0 \cdot \tan\vartheta_{opt}} \Leftrightarrow \tan\vartheta_{opt} = \beta_0^{-1} \cdot \frac{\beta_0 - \beta_{opt}}{\beta_{opt} - \beta_{90}} . \tag{44}$$

Therefore, the optimal value of the slope can be expressed in terms of the slope for $\vartheta = 0$, β_0, and $\vartheta = 90^0$, β_{90}. Namely,

$$\beta_0^{-1} \cdot \frac{\beta_0 - \beta_{opt}}{\beta_{opt} - \beta_{90}} = \tan\vartheta_{opt} = (-1 \pm \sqrt{1 + X^2})/X , \; X = \frac{1}{2}(\beta_{90} - \beta_0^{-1}) . \tag{45}$$

We obtain $\beta_{opt} = \beta_0$ for $X \to 0$, leading to $r_{opt}^2 = r_0{}^2$, and $\beta_{opt} = \beta_{45} = \beta_0 \cdot (1 + \beta_{90})/(1 + \beta_0)$ for $X \to \infty$, leading to $r_{opt}^2 = 1$; (see: [11]).

In general, the optimal value of the slope is given by:

$$\beta_{opt} = \beta_{90} \cdot \frac{-\frac{1}{2}(1 + \beta_{90}^{-1}\beta_0^{-1}) + \sqrt{1 + \frac{1}{4}(\beta_{90} - \beta_0^{-1})^2}}{\frac{1}{2}\beta_0^{-1}(\beta_{90} - \beta_0^{-1}) - 1 + \sqrt{1 + \frac{1}{4}(\beta_{90} - \beta_0^{-1})^2}} . \tag{46}$$

4. Application to the Inner Heliosheath

Classical collisional particle systems residing in thermal equilibrium have their particle velocity/energy distribution function stabilized into a Maxwell-Boltzmann distribution. On the contrary, space and astrophysical plasmas are collisionless particle systems residing in stationary states characterized by kappa distributions [12–15]. The role of kappa distributions has become increasingly widespread across the physics of space plasma processes, describing particles in the heliosphere, from the solar wind and planetary magnetospheres to the heliosheath and beyond, the interstellar and intergalactic plasmas.

The presented statistical analysis is applied to the measurements of number density n and temperature T of the proton plasma in the inner heliosheath, the outer boundary of our heliosphere. By fitting kappa distributions to the proton energy distributions derived from the IBEX mission [16], Livadiotis et al. [17–19] derived the sky maps of the radially averaged values of several thermodynamic quantities in the inner heliosheath, such as the number density n, the temperature T, and the kappa indices κ_0; these thermal observables parameterize the kappa distributions. The kappa index is the parameter that labels the kappa distribution, and its physical meaning is interwoven with the correlation between particles; in fact, the stronger the correlations, the smaller the kappa index [13,15,20].

The measurements of density and temperature of the inner heliosheath are anticorrelated. In particular, the logarithms of these measurements are linearly related with a negative slope, that is, the polytropic index. The polytropic behavior stands for the powe-law relationship, $T \propto n^{\gamma-1}$, where γ is the polytropic index, along a certain streamline of the plasma flow. A polytrope is a certain thermodynamic process characterized by such a relation. Typically, this is a power-law between two thermodynamic variables,

$$[P(\vec{r})/P(\vec{r}_*)] = [n(\vec{r})/n(\vec{r}_*)]^{\gamma}, \; \text{or} \; [n(\vec{r})/n(\vec{r}_*)] = [T(\vec{r})/T(\vec{r}_*)]^{\nu} , \; \text{with } \nu \equiv 1/(\gamma - 1) , \tag{47}$$

where the exponent a denotes the typical polytropic index; ν is another way of expressing the polytropic index, which corresponds to the effective degrees of freedom. While most of the space plasmas exhibit positive correlations between density and temperature, there are several plasmas with negative correlations, consistent with constant or quasi-constant thermal pressure, that were found in heliosheath [10,17–21] and planetary magnetosheaths (e.g., the low-latitude boundary layer at the terrestrial magnetosheath [22]; in the terrestrial central plasma sheet [23]; and in the Jovian

magnetosheath, [24]). Other plasmas exhibit more complicate relationships that involves also the kappa index [17,25].

The temperature T and density n values of the proton plasma along the equatorial streamline from the heliospheric nose towards the heliotail were shown to be characterized by a negative correlation relationship [10,26,27]. More precisely, it was shown that the variations of T and n follow a near isobaric process, that is a polytropic index $\gamma \approx 0$ (or $\nu \approx -1$).

Here, we examine the linear regression of the data (x,y), where $x = \log(n)$ and $y = \log(1/T)$. The slope β coincides with the quantity $1-\gamma$, thus, the polytropic index γ can be retrieved once we estimate the slope β. The correlation coefficient $r^2(\vartheta)$ and the slope $\beta(\vartheta)$ for a rotated reference frame are plotted in panels (a) and (b) in Figure 2, respectively. We also exclude the data points with small kappa index (as more erroneous). The plots are depicted for different thresholds, κ_{0min}, that is, 0.1 (red), 0.25 (blue), and 0.4 (green). We observe that the larger the kappa index threshold, the more the correlation coefficient shifts to the left, to the negative values of $\tan\vartheta$. This corresponds to smaller values of the slope, thus larger values of the polytropic index. The optimal value of $\tan\vartheta$ is the one that maximizes the correlation coefficient; this is plotted against the kappa index threshold κ_{0min} in panel Figure 2c.
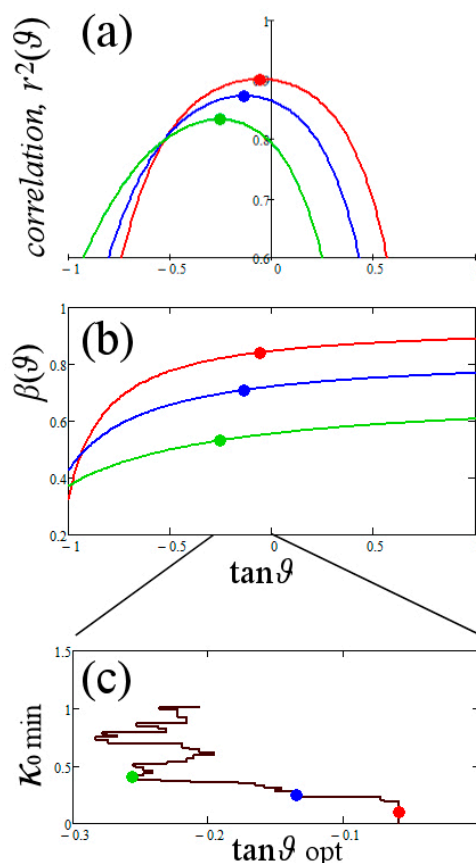

Figure 2. (**a**) The Pearson's correlation coefficient (square) r^2, and (**b**) the slope β, both at the rotation reference frame, are plotted against the tangent of the rotation angle $\tan\vartheta$. We exclude data points with small kappa index, and the plots are depicted for different thresholds (i.e., κ_{0min}: 0.1 (red), 0.25 (blue), and 0.4 (green)). (**c**) The optimal $\tan\vartheta$ is plotted against the kappa index threshold κ_{0min}.

5. Conclusions

The paper presented the dependence of the linear regression on the selection of the reference frame. Both the slope of the fitted line and the corresponding Pearson's correlation coefficient were expressed in terms of the rotation angle. The correlation coefficient was found to be maximized for a certain optimal angle, for which the slope attains a special optimal value. The optimal angle, the value

of the optimal slope, and the corresponding maximum correlation coefficient were expressed in terms of the covariance matrix, but also in terms of the slopes β_0 and β_{90}, namely, those derived from the nonrotated and right-angle-rotated fittings.

The presented analysis can be used as a method for improving the fitting and optimizing the derived values of the involved parameters. The analysis can be further completed for including the following cases:

- the data points are given with errors $\{x_i \pm \sigma_{xi}\}_{i=1}^{N}$, $\left\{y_i \pm \sigma_{yi}\right\}_{i=1}^{N}$.
- the statistical model is nonlinear and multiparametrical, for example, a polynomial of order M, that is, $V(x_i; \{\alpha_k\}_{k=1}^{n}) = \sum_{k=0}^{M} \alpha_k x_i^{k}$.
- the estimation of the errors of the parameters, $\{\delta\alpha_k\}_{k=1}^{n}$.
- minimization of least deviation in L_p norm [10–12,28].

The presented analysis was applied in the density and temperature measurements of the proton plasma in the inner heliosheath, the outer boundary of our heliosphere. Those measurements are known to be fitted with a line with a negative slope, which is related to the polytropic index of the plasma. The application of the new statistical method improved the estimated value of this index.

Finally, Table 1 summarizes the critical equations shown in this paper.

Table 1. Formulae of the developed method for finding the optimal rotation of maximum correlation.

Optimal Quantity	Formula
$\tan\vartheta_{\text{opt}}$	$(-1 + \sqrt{1 + X^2})/X$
β_{opt}	$\beta_0 \cdot \frac{1 + \beta_{90}\cdot\tan\vartheta_{\text{opt}}}{1 + \beta_0\cdot\tan\vartheta_{\text{opt}}}$
$r^2{}_{\text{opt}}$	$r_0{}^2 \cdot \frac{1 + X^2}{1 + r_0^2 X^2}$

Note: We set $\beta_{90} = \sigma_{yy}^2/\sigma_{xy}^2$, $\beta_0 = \sigma_{xy}^2/\sigma_{xx}^2$, and $X = \frac{1}{2}(\beta_{90} - \beta_0^{-1})$.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Kenney, J.F.; Keeping, E.S. Linear Regression and Correlation. In *Mathematics of Statistics*, 3rd ed.; Van Nostrand: Princeton, NJ, USA, 1962; pp. 252–285.
2. McCullagh, P. What is statistical model? *Ann. Stat.* **2002**, *30*, 1225–1310. [CrossRef]
3. Adèr, H.J. Modelling. In *Advising on Research Methods: A Consultant's Companion*; Adèr, H.J., Mellenbergh, G.J., Eds.; Johannes van Kessel Publishing: Huizen, The Netherlands, 2008; pp. 271–304.
4. Melissinos, A.C. *Experiments in Modern Physics*; Academic Press Inc.: London, UK, 1966; pp. 438–464.
5. Burden, R.L.; Faires, J.D. *Numerical Analysis*; PWS Publishing Company: Boston, MA, USA, 1993; pp. 437–438.
6. Livadiotis, G. Approach to general methods for fitting and their sensitivity. *Physica A* **2007**, *375*, 518–536. [CrossRef]
7. Livadiotis, G. Expectation values and Variance based on L^p norms. *Entropy* **2012**, *14*, 2375–2396. [CrossRef]
8. Livadiotis, G. Chi-*p* distribution: Characterization of the goodness of the fitting using L^p norms. *J. Stat Distr. Appl.* **2014**, *1*, 4. [CrossRef]
9. Livadiotis, G.; Moussas, X. The sunspot as an autonomous dynamical system: A model for the growth and decay phases of sunspots. *Physica A* **2007**, *379*, 436–458. [CrossRef]
10. Livadiotis, G.; McComas, D.J. Fitting method based on correlation maximization: Applications in Astrophysics. *J. Geophys. Res.* **2013**, *118*, 2863–2875. [CrossRef]
11. Schmid, J., Jr. The relationship between the coefficient of correlation and the angle included between regression lines. *J. Educ. Res.* **1947**, *41*, 311–313. [CrossRef]

12. Livadiotis, G. *Kappa Distribution: Theory & Applications in Plasmas*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2017.

13. Livadiotis, G.; McComas, D.J. Beyond kappa distributions: Exploiting Tsallis statistical mechanics in space plasmas. *J. Geophys. Res.* **2009**, *114*, A11105. [CrossRef]

14. Livadiotis, G.; McComas, D.J. Understanding kappa distributions: A toolbox for space science and astrophysics. *Space Sci. Rev.* **2013**, *175*, 183–214. [CrossRef]

15. Livadiotis, G. Thermodynamic origin of kappa distributions. *Europhys. Lett.* **2018**, *122*, 50001. [CrossRef]

16. McComas, D.J.; Allegrini, F.; Bochsler, P.; Bzowski, M.; Christian, E.R.; Crew, G.B.; DeMajistre, R.; Fahr, H.; Fichtner, H.; Frisch, P.C.; et al. Global observations of the interstellar interaction from the Interstellar Boundary Explorer (IBEX). *Science* **2009**, *326*, 959. [CrossRef] [PubMed]

17. Livadiotis, G.; McComas, D.J.; Dayeh, M.A.; Funsten, H.O.; Schwadron, N.A. First sky map of the inner heliosheath temperature using IBEX spectra. *Astrophys. J.* **2011**, *734*, 1. [CrossRef]

18. Livadiotis, G.; McComas, D.J.; Randol, B.; Möbius, E.; Dayeh, M.A.; Frisch, P.C.; Funsten, H.O.; Schwadron, N.A.; Zank, G.P. Pick-up ion distributions and their influence on ENA spectral curvature. *Astrophys. J.* **2012**, *751*, 64. [CrossRef]

19. Livadiotis, G.; McComas, D.J.; Schwadron, N.A.; Funsten, H.O.; Fuselier, S.A. Pressure of the proton plasma in the inner heliosheath. *Astrophys. J.* **2013**, *762*, 134. [CrossRef]

20. Livadiotis, G.; McComas, D.J. Invariant kappa distribution in space plasmas out of equilibrium. *Astrophys. J.* **2011**, *741*, 88. [CrossRef]

21. Elliott, H.A.; McComas, D.J.; Zirnstein, E.J.; Randol, B.M.; Delamere, P.A.; Livadiotis, G.; Bagenal, F.; Barnes, N.P.; Stern, S.A.; Young, L.A.; et al. Slowing of the solar wind in the outer heliosphere. *Astrophys. J.* **2019**, in press.

22. Sckopke, N.; Paschmann, G.; Haerendel, G.; Sonnerup, B.U.O.; Bame, S.J.; Forbes, T.G.; Hones, E.W., Jr.; Russell, C.T. Structure of the low-latitude boundary layer. *J. Geophys. Res.* **1981**, *86*, 2099–2110. [CrossRef]

23. Pang, X.X.; Cao, J.B.; Liu, W.; Ma, Y.; Lu, H.; Yang, J.; Li, L.; Liu, X.; Wang, J.; Wang, T.; et al. Polytropic index of central plasma sheet ions based on MHD Bernoulli integral. *J. Geophys. Res.* **2015**, *120*, 4736–4747. [CrossRef]

24. Nicolaou, G.; McComas, D.J.; Bagenal, F.; Elliott, H.A.; Wilson, R.J. Plasma properties in the deep Jovian magnetotail. *Planet. Space Sci.* **2015**, *119*, 222–232. [CrossRef]

25. Ogasawara, K.; Angelopoulos, V.; Dayeh, M.A.; Fuselier, S.A.; Livadiotis, G.; McComas, D.J.; McFadden, J.P. Characterizing the dayside magnetosheath using ENAs: IBEX and THEMIS observations. *J. Geophys. Res.* **2013**, *118*, 3126–3137. [CrossRef]

26. Livadiotis, G.; McComas, D.J. Non-equilibrium thermodynamic processes: Space plasmas and the inner heliosheath. *Astrophys. J.* **2012**, *749*, 11. [CrossRef]

27. Livadiotis, G. Superposition of polytropes in the inner heliosheath. *Astrophys. J. Suppl. Ser.* **2016**, *223*, 13. [CrossRef]

28. Livadiotis, G. Non-Euclidean-normed Statistical Mechanics. *Physica A* **2016**, *445*, 240–255. [CrossRef]