

Note Complémentaire sur La Méthode des Moindres Carrés

Abdelmajid Ben Hadj Salem¹

¹*Résidence Bousten 8, Bloc B, Rue Mosquée Raoudha, 1181 La Soukra Raoudha
Tunisia.*

E-mail: abenhadjsale@gmail.com

ABSTRACT:

In this additional note on least squares, we will come back to a few theorems of the theory of errors and that of least squares which are not well enough known even among geodesists. This is due to the absence of documentation in French on the subject. It is thanks to a work of the Russian school, of the mathematician Yori Vladimirovich LINNIK¹ on the theory of least squares [1], published in French in 1963 which I was inspired to write this note.

Keywords : least squares theory, mathematical statistics, density function of a probability law, weight matrix, normal law, standard deviation.

RÉSUMÉ :

Dans cette note complémentaire sur les moindres carrés, nous reviendrons sur quelques théorèmes de la théorie des erreurs et celle des moindres carrés non assez connus mêmes chez les géodésiens. Ce-ci est dû à l'absence de la documentation en langue française en la matière. C'est grâce à un ouvrage de l'école russe, du mathématicien Yori Vladimirovich LINNIK sur la théorie des moindres carrés [1], édité en français et paru en France en 1963 duquel je suis inspiré pour la rédaction de cette note.

Mots-clefs : théorie des moindres carrés, statistiques mathématiques, fonction de densité d'une loi de probabilités, matrice de poids, loi normale, écart-type.

1. Yori Vladimirovich LINNIK (1915-1972) : Russian mathematician.

Table des matières

1	RAPPELS HISTORIQUES DE LA MÉTHODE DES MOINDRES CARRÉS	3
2	RAPPELS DE PROBABILITÉS ET DE STATISTIQUES	4
2.1	Variable Aléatoire	4
2.1.1	Description d'une variable aléatoire	4
2.1.2	Propriétés	4
2.2	Variable aléatoire normale	7
3	Fonctions linéaires d'un vecteur normal à n dimensions	8
3.1	Propriétés de la transformation linéaire d'un vecteur normal à n dimensions	10

**Note Complémentaire sur La Méthode des Moindres
Carrés
- ABDELMAJID BEN HADJ SALEM -**



FIGURE 1. Professeur Helmut Moritz

A la mémoire de l'éminent géodésien et professeur Helmut Moritz (1936-2022)

1 RAPPELS HISTORIQUES DE LA MÉTHODE DES MOINDRES CARRÉS

Les premiers éléments de cette théorie sont dûs à A. M. Legendre¹ [2] qui les a introduits lors d'études, en 1806, portant sur le calcul des orbites de comètes. C'est à lui, aussi, qu'est dû le nom donné à cette méthode : *méthode des moindres carrés*.

En 1809 K. F. Gauss² [3] donne les premières bases probabilistes à cette méthode [4], puis, en 1810, élabore les méthodes de calcul et introduit des symboles et des notations qui sont encore utilisés. Enfin, en 1821 [5], il découvre toute une série de résultats importants concernant cette méthode.

En 1812 P. S. Laplace³ dans son traité fondamental de calcul des probabilités expose plusieurs résultats qu'il applique à la méthode des moindres carrés.

D'autres résultats non moins importants viennent enrichir la théorie quand, en 1859, P. L. Tchebychev⁴ met au point sa théorie de l'interpolation par la méthode des moindres carrés, à l'aide des polynômes orthogonaux qui portent son nom.

A. A. Markov⁵ en 1898, dans un traité [9] et dans son cours de calcul des probabilités, introduit en statistique mathématique toute une série de notions très importantes explicitant l'essence de la méthode des moindres carrés.

Les applications de la méthode des moindres carrés à l'astronomie et à la géodésie doivent beaucoup aux travaux de F. Helmert⁶[6] à la fin du 19ème siècle. A partir des années vingt du dernier siècle, et grâce aux travaux de A. A. Markov, la méthode des moindres carrés fait partie de la statistique mathématique et est considérée comme un aspect important et naturel de la théorie de l'estimation des paramètres. Dans cette voie, tout un ensemble de résultats intéressants et importants est dû à J. Neyman [7] et F. David [8], Aitken [10] et Rao [11].

En 1946, A. N. Kolmogorov⁷ [12] donne un élégant exposé géométrique de cette

-
1. A. M. LEGENDRE (1752-1833), mathématicien français.
 2. K. F. GAUSS (1777-1855), mathématicien allemand.
 3. P. S. LAPLACE (1749-1827), mathématicien français.
 4. P. L. TCHEBYCHEV (1821-1894), mathématicien russe.
 5. A. A. MARKOV (1856-1922), mathématicien russe.
 6. F. HELMERT (1843-1917), géodésien allemand.
 7. A. N. KOLMOGOROV (1903-1987) : Mathématicien russe.

méthode.

2 RAPPELS DE PROBABILITÉS ET DE STATISTIQUES

2.1 Variable Aléatoire

Définition 2.1. *Variable aléatoire (v.a) : c'est une fonction dont les résultats possibles sont connus mais dont le résultat final ne peut être déterminé, à priori, avant d'effectuer la mesure.*

Exemples :

- fermetures des angles des triangles d'un réseau géodésique,
- épaisseur d'une veine minéralisée,
- concentration d'un polluant dans l'eau souterraine,
- Ph de l'eau de pluie.

2.1.1 Description d'une variable aléatoire

Sans connaître la valeur que prendra le résultat final, on peut parfois connaître la probabilité qu'une variable aléatoire prenne chacun des résultats possibles. C'est la description la plus complète que l'on puisse faire de la variable aléatoire.

La fonction qui décrit ces probabilités est la fonction de densité f pour les variables aléatoires continues et c'est la fonction de masse pour les variables aléatoires discrètes.

2.1.2 Propriétés

On a :

- $f_X(x) \geq 0$, toute probabilité est positive,

- $\sum_{-\infty}^{+\infty} f_X(x) dx = 1$, l'intégrale de la fonction densité donne 1,

- $\sum_a^b f_X(x) dx = P(a \leq X \leq b)$, c'est la probabilité que x prenne une valeur comprise entre a et b .

Certaines quantités résument les caractéristiques principales de la variable aléatoire.

*** - Mesures de tendance centrale :**

- mode : x tel que $f_X(x)$ est maximum,

- médiane : x tel que $P(X < x) = 0.5$,

- moyenne (ou espérance mathématique) : μ_X ou $E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx$,

***- Mesures de dispersion :**

-Variance : $\sigma_X^2 = \int_{-\infty}^{+\infty} (x - E(X))^2 f_X(x) dx$. On note souvent la variance par $D(X)$ ou $\Gamma(X)$.

-Ecart-type : $\sigma_X = \sqrt{\sigma_X^2}$.

-Asymétrie : $E \left[\left(\frac{X - E(X)}{\sigma_X} \right)^3 \right]$.

-Aplatissement : $E \left[\left(\frac{X - E(X)}{\sigma_X} \right)^4 \right]$.

Toutes ces quantités sont généralement, *à priori*, inconnues. On doit donc les estimer à partir d'un ensemble d'observations appelé l'échantillon (par abus de langage, on parlera souvent des échantillons pour désigner ces observations).

A partir de l'échantillon, on peut construire des estimateurs :

- de la moyenne :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.1)$$

- de la variance :

$$\bar{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.2)$$

$$\text{ou } s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.3)$$

- de la fonction de densité : histogramme,

- de la fonction de densité cumulative : courbe des fréquences cumulées ($F_X(x) = P(X \leq x)$) estimée par rang $(x_i)/n$.

Une des caractéristiques importantes d'un estimateur est d'être sans biais, c'est-à-dire d'avoir la même espérance mathématique que la quantité qu'il cherche à évaluer.

Exemple : $E[\bar{X}] = E\left(\frac{\sum_{i=1}^{i=n} X_i}{n}\right) = \frac{1}{n}E\left(\sum_{i=1}^{i=n} X_i\right) = \frac{n}{n}E(X_i) = E(X_i) = \mu_X$
donc \bar{X} est sans biais pour μ_X .

De même, s^2 est sans biais pour σ_X^2 , alors que $\hat{\sigma}^2$ est biaisé.

Passage à plus d'une variable : On peut aussi étudier et décrire le comportement simultané de plus d'une variable aléatoire.

La fonction de densité conjointe : $f_{xy}(x, y)$ donne la probabilité que, simultanément $X = x$ et $Y = y$. On a :

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = 1 \quad (2.4)$$

$$P[x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2] = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{XY}(x, y) dx dy = 1 \quad (2.5)$$

Deux mesures additionnelles permettent de décrire des caractéristiques importantes de la fonction de densité conjointe.

La covariance :

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \quad (2.6)$$

mesure la force du lien linéaire entre les variables X et Y .

Le coefficient de corrélation :

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad (2.7)$$

comme la covariance mais avec des unités "normalisées".

Propriétés de ρ_{XY} , on a :

$$-1 \leq \rho_{XY} \leq 1 \quad (2.8)$$

$$\rho_{XY} = \rho_{aX, bY} \quad (2.9)$$

avec a et b des constantes quelconques. Notons que :

$\rho_{XY} = 0 \implies$ absence de lien linéaire, et $\rho_{XY} \neq 0$ indépendance de X et Y . En effet, on a indépendance si et seulement si $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$. Par contre, l'indépendance de X et $Y \implies \rho_{XY} = 0$.

2.2 Variable aléatoire normale

Définition 2.2. Une variable aléatoire x est normale si la fonction de densité $f(x)$ est de la forme :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right), \quad a = E(x), \quad \sigma = \sigma(x) = \text{l'écart-type de } x \quad (2.10)$$

On note par $\mathcal{N}(a, \sigma)$ l'ensemble des lois normales, par suite la variable aléatoire $x \in \mathcal{N}(a, \sigma)$.

Voyons quelques propriétés numériques de la loi normale (2.10).

Propriété 2.3. Si X est une variable normale $X \in \mathcal{N}(a, \sigma)$, nous aurons :

$$P\{|X - a| \leq 1.96\sigma\} \approx 0.95 \quad (2.11)$$

$$P\{|X - a| \geq 3\sigma\} \approx 0.0027 \quad (2.12)$$

La dernière inégalité est souvent appelée «loi de l'intervalle 3σ ».

Une importante propriété de l'espérance mathématique et de la variance $E(X)$ et $D(X)$ est la suivante :

Propriété 2.4. Si X et Y sont deux variables aléatoires indépendantes :

$$E(XY) = E(X)E(Y) \quad (2.13)$$

à condition que les deux membres aient un sens. Si X_1, X_2, \dots, X_n sont des grandeurs aléatoires indépendantes deux à deux, on a :

$$D(X_1 + \dots + X_n) = D(X_1) + \dots + D(X_n) \quad (2.14)$$

si le second membre a un sens.

Enfin une autre propriété dite l'inégalité de P.L. Tchebychef qui porte sur $D(x)$ et $E(X)$:

Propriété 2.5. Quel que soit $t > 0$:

$$P\{|X - E(X)| \geq t\} \leq \frac{D(X)}{t^2} \quad (2.15)$$

3 Fonctions linéaires d'un vecteur normal à n dimensions

Dans les applications de la méthode des moindres carrés, nous aurons souvent affaire à un vecteur normal à n dimensions, soient :

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

un vecteur aléatoire normal à n dimensions et :

$$Q = Z^T \Lambda Z = \sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} z_i z_j$$

une forme quadratique définie positive de n variables z_1, \dots, z_n correspondante de la matrice Λ :

$$\Lambda = {}_n\Lambda_n = (\lambda_{ij})$$

qui est symétrique et définie positive. Q vérifie : $\forall Z \neq 0, Z^T \Lambda Z > 0$. On note :

$$E(X) = A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_n) \end{pmatrix} \quad (3.1)$$

Comme X est un vecteur normal à n dimensions, sa fonction de densité de probabilité est égale à :

$$f(x_1, \dots, x_n) = C_0 \exp \left[-\frac{1}{2} (X - A)^T \Lambda (X - A) \right] \quad (3.2)$$

où la constante C_0 étant définie par l'égalité :

$$C_0 \cdot \int_{\mathbb{R}^n} \exp \left[-\frac{1}{2} (X - A)^T \Lambda (X - A) \right] dx_1 \dots dx_n = 1 \quad (3.3)$$

On a alors le théorème sans donner la démonstration :

Théorème 3.1. *On a les égalités :*

$$C_0 = \frac{[\det(\Lambda)]^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} \quad (3.4)$$

$$E(X) = A \quad (3.5)$$

$$\Gamma_X = \Lambda^{-1} \quad (3.6)$$

où Γ_X est la matrice des covariances du vecteur X .

La matrice non singulière Γ_X est donc l'inverse de Λ .

La fonction de densité normale $f(X)$ s'écrit :

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} [\det(\Gamma)]^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (X - A)^T \Gamma_X^{-1} (X - A) \right] \quad (3.7)$$

Exemple : Considérons que X est un vecteur normal à deux dimensions :

$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad A = E(X) = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \quad \Gamma_X = \begin{pmatrix} \text{Cov}(x_1 - a_1, x_1 - a_1) & \text{Cov}(x_1 - a_1, x_2 - a_2) \\ \text{Cov}(x_1 - a_1, x_2 - a_2) & \text{Cov}(x_2 - a_2, x_2 - a_2) \end{pmatrix}$$

Introduisons les notations suivantes :

$$\begin{aligned} Cov(x_1 - a_1, x_1 - a_1) &= \sigma_1^2 \\ Cov(x_2 - a_2, x_2 - a_2) &= \sigma_2^2 \\ Cov(x_1 - a_1, x_2 - a_2) &= \rho\sigma_1\sigma_2, \quad |rho| < 1 \end{aligned}$$

La quantité ρ est le coefficient de corrélation entre x_1 et x_2 . Par un calcul facile, nous retrouvons les expressions détaillées des formules (3.6,3.7) données ci-dessus :

$$\left\{ \begin{array}{l} \Gamma_X = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \Gamma_X^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} \\ \\ det(\Gamma_X) = \sigma_1^2\sigma_2^2(1-\rho^2) \\ \\ f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \exp \left[\frac{-2}{1-\rho^2} \left(\frac{(x_1 - a_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - a_1)(x_2 - a_2)}{\sigma_1\sigma_2} + \frac{(x_2 - a_2)^2}{\sigma_2^2} \right) \right] \end{array} \right. \quad (3.8)$$

3.1 Propriétés de la transformation linéaire d'un vecteur normal à n dimensions

Dans les applications de la méthode des moindres carrés, nous aurons souvent affaire au vecteur normal à n dimensions. Soit $X = (x_1, \dots, x_n)^T$ un vecteur normal à n dimensions dont les composantes x_j représenteront des erreurs de mesure aléatoires ou des corrections aux valeurs approchées. De plus, Il nous arrive de former des combinaisons linéaires des observations et de leurs erreurs de la forme :

$$Y = A.X \quad \text{où } A \text{ est une matrice}$$

On a vu quand X est le vecteur de la solution obtenue par moindres carrés et si $Y = AX$ où $A = {}_nA_n$ une matrice carrée $n \times n$ de rang n cest-à-dire inversible, alors la matrice des covariances de Y est donnée par la formule :

$$\Gamma_Y = A.\Gamma_X.A^T \quad (3.9)$$

Alors, maintenant on suppose que la matrice A est une matrice rectangulaire $k \times n$ avec $k < n$ et $rang(A) = k$. Le vecteur $Y = A.X = {}_1Y_k$. On considère que X est un vecteur normal à n dimensions et $E(X) = M$. On donne ci-dessus sans démonstration le théorème suivant :

Théorème 3.2. *Le vecteur $Y = A.X$ est un vecteur aléatoire normal à k dimensions, qui a pour espérance mathématique le vecteur $A.M$ et pour matrice des covariances, la matrice :*

$$\Gamma_Y = A.\Gamma_X.A^T$$

On donne ci-dessous un autre théorème en relation.

Théorème 3.3. (Théorème de R.A. Fisher^a) *Soient X un vecteur normal, à n dimensions ayant un vecteur espérance mathématique nul et dont les composantes sont indépendantes et possèdent toutes une même distribution de probabilité et $F = {}_nF_n$ une transformation orthogonale quelconque ($F^{-1} = F^T$); alors, le vecteur normal à n dimensions $Y = F.X$ aura la même distribution que le vecteur X .*

a. R.A. Fisher (1890 - 1962) : Statisticien anglais.

Démonstration : Nous avons par hypothèse : X est un vecteur normal à n dimensions $X = (x_1, \dots, x_n)^T$ tel que $E(X) = O \implies E(x_i) = 0$. Les x_i sont indépendants et ont une même distribution de probabilité $N(0, \sigma)$, soit $\sigma_{x_1} = \sigma_{x_2} = \dots = \sigma_{x_n} = \sigma$. Par suite la matrice des covariances est $\Gamma_X = \sigma^2.I$ où I est la matrice unité $n \times n$. Soit F une matrice $n \times n$ orthogonale quelconque donc $F^{-1} = F^T$. Soit $Y = F.X$, d'après le théorème 3.2 ci-dessus, $Y = F.X$ est un vecteur normal à n dimensions et :

$$E(Y) = E(F.X) = F.E(X) = F.O = O \quad (3.10)$$

où O est le vecteur nul. On a aussi :

$$\Gamma_Y = F\Gamma_X.F^T = F.(\sigma^2 I).F^T = \sigma^2 F.F^T = \sigma^2 F.F^{-1} = \sigma^2.I \quad (3.11)$$

soit le vecteur Y possède la même distribution que X c'est-à-dire $N(0, \sigma)$.

Soient deux fonctions vectorielles linéaires du vecteur normal à n dimensions X :

$$Y_1 = A_1.X; \quad Y_2 = A_2.X$$

où A_1 est une matrice du type $(m_1 \times n)$ et A_2 est du type $(m_2 \times n)$. Nous pouvons énoncer sans démonstration le théorème suivant :

Théorème 3.4. *La condition nécessaire et suffisante pour que les vecteurs $Y_1 = A_1 \cdot X$ et $Y_2 = A_2 \cdot X$ soient statistiquement indépendants est que :*

$$A_1 \Gamma_X \cdot A_2^T = 0 \quad (3.12)$$

où Γ_X est la matrice des covariances de X .

Références

- [1] LINNIK, Y. V., Méthode des Moindres Carrés - Eléments de la Théorie du Traitement Statistique des Observations,. Traduction de l'ouvrage publié en langue russe sous le titre : METOD NAIMENCHIKH KVADRATOV par les ÉDITIONS DE LITTÉRATURE PHYSICO-MATHÉMATIQUE, Moscou. Traduit par O. ARKHIPOFF. Dunod, Paris, 1963, 341 pages.
- [2] LEGENDRE, A. M., Nouvelles méthodes pour la détermination des orbites des comètes. Paris, 1806, Appendice sur la méthode des moindres carrés.
- [3] GAUSS, C. F., Teoria motus corporum cœlestium, Hamburg, 1809.
- [4] GAUSS, C. F., Disquisitio de elementis ellipticis Palladis. 1810.
- [5] GAUSS, C. F., Theoria combinationis observationum erroribus minimis obnoxiae, 1821.
- [6] HELMERT, F., Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate. Leipzig-Berlin, 1907.
- [7] NEYMAN, J., Statistitcheskaia otsenka kak problema klassitchskoi teorii veroiatnostei, Ouspekhi matem. naouk, byp. 10, 1944, 207-229.
- [8] DAVID, F. N., Probability theory for statistical methods. Cambridge, 1951.
- [9] MARKOV, A. A., Zakon bolchikh tchisel i metod naimenchikh kvadratov, (1898) Izbr. Troudy, Izd. AN SSSR, 1951, pages 233-251.
- [10] EZEKIEL, M., Methods of correlation analysis, 2ème édition, N. Y., 1941.
- [11] RAO, C. R., On the linear combination of observations and the general theory of least squares. Sankhya, vol. 7, p. 3, 1946, 237-256.
- [12] KOLMOGOROV, A. N., Kobosnovaniou metoda naimenchikh kvadratov. Ouspekhi matem. naouk, t. 1, byp. 1, 1946, 57-70.

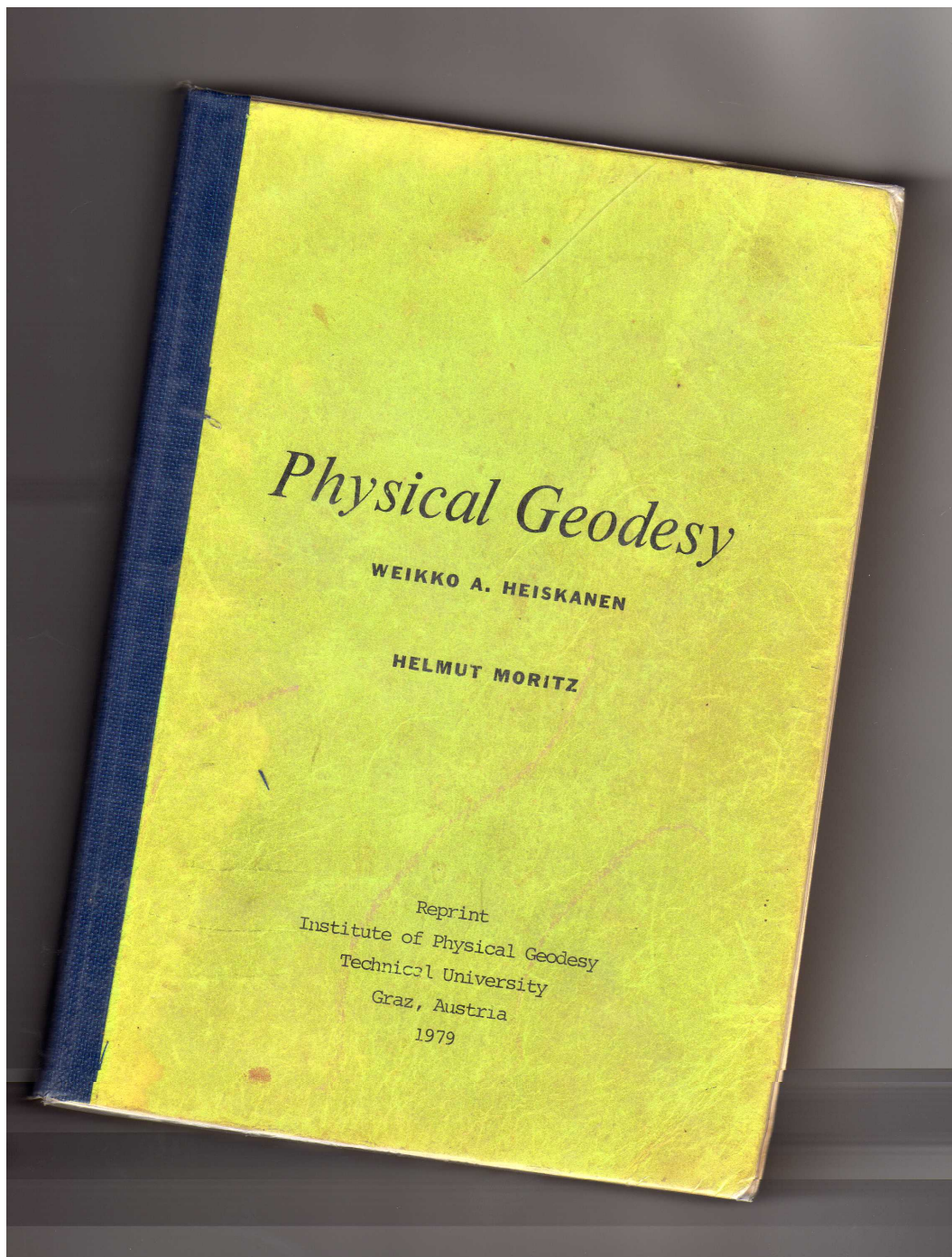


FIGURE 2. La Bible de la "Géodésie-Physique" par Helmut Moritz