Covariance-Invariant Mapping of Data Points to Nonlinear Models

Wolfgang M. Grimm

Independent Research and Consulting, Leonberg, 71229 Germany

wolfgang.grimm@daad-alumni.de

ORCID: 0000-0003-3535-3693

MOS subject classification: 11B83, 40E10, 62J05, 65D10

Covariance-Invariant Mapping of Data Points to Nonlinear Models

Abstract. A centroid- and covariance-invariant deterministic mapping of sets of discrete data points to nonlinear models is introduced. Conditions for bijectivity of this mapping are developed. Since the mapping can be accomplished by look-up tables for the special case of equally-spaced data, the resulting mapping algorithm is considered computationally fast. This could be attractive for real-time operations.

Keywords. Data analysis, bijectivity, fast algorithm, approximation, nonlinear regression, cross-central moments.

1. INTRODUCTION.

Mean values such as the centroid of data and their covariance are widely used for mapping of data. In statistics, a mean-value mapping defines the variance functions of Tweedie distributions [1] applied to exponential dispersion models [2]. A covariance mapping is used to generalize the one-dimensional measure of the statistical relation between two random variables in [3]. In computer science, mean-value mappings are defined for digital imaging and computer graphics to accomplish image space rendering and texturing [4]. Mean transformations are defined in matrix analysis [5], or in standardizing statistical distributions [6]. - In contrast, the objective of the mean-value mapping proposed here is the development of an easy-to-apply, reliable and computationally fast mapping of data points to a nonlinear model. For this purpose, a mapping is investigated, which preserves mean values, namely, the centroid and covariance of data points, which are directly related to two model parameters. The approach should be extendable to models with m parameters by taking m-th-order cross-central moments into account. The suggested mapping should have potential applications in various fields ranging from econometrics and technometrics to pattern recognition and computer graphics, which increasingly focuses on real-time construction of models from real-world data [7].

2. MEAN-VALUE MAPPING OF DATA

The following considerations are based on a discrete representation of any straight line by four mean values, which are the arithmetic means of the independent and dependent data $\bar{x} = \sum_{i=1}^{n} x_i \,/n$ and $\bar{y} = \sum_{i=1}^{n} y_i \,/n$, respectively, where $x_i, y_i \in \mathbb{R}$, and the means of products of their deviations from the means, i.e., the covariance $Cov(x, y) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \,/n$ and the variance $Var(x) = Cov(x, x) = \sum_{i=1}^{n}(x_i - \bar{x})^2 \,/n$.

The point $(\bar{x}, \bar{y}\ )$ defines the centroid or "center of mass" [8] of the data points.

Definition 1 (Mean-value form of a straight line). *Any straight line in a two-dimensional Cartesian coordinate system can be represented by the mean-value form*

$$y_i = \frac{Cov(x,y)}{Var(x)}(x_i - \bar{x}) + \bar{y} \tag{1}$$

connecting a set of points (x_i, y_i), $i = 1, \ldots, n$, *with* $x_i, y_i \in \mathbb{R}$ and $i \in \mathbb{N}$. □

The derivation of (1) is straightforward using the normal equations of ordinary least-squares regression such as in [9] implying the following lemma.

Lemma 1. *Any set of points* (x_i, y_i^o), $i = 1, \ldots, n$, $x_i, y_i^o \in \mathbb{R}$, *with the observations* y_i^o *scattered around a straight line segment in a two-dimensional Cartesian coordinate system, fits equation (1) in a least-squares sense by preserving their arithmetic mean and covariance, i.e.,* $\bar{y} = \overline{y^o} = \sum_{i=1}^{n} y_i^o\ /n$ *and* $Cov(x,y) = Cov(x, y^o)$. □

To derive a mapping between a set of data points (x_i, y_i) located on a straight-line segment and a corresponding set of data (x_i, y_i^*) on a nonlinear function, the idea is to define a mapping which preserves the centroid of the data and their covariance, such as the linear regression does according to Lemma 1. Since a set of data points on a nonlinear function (x_i, y_i^*) can be viewed observations and interpreted as a set of data points scattered around a straight-line segment(x_i, y_i), the mapping can be defined in view of Lemma 1 as follows.

Definition 2 (Mean-value mapping). *The mean-value mapping between a set of data points* (x_i, y_i), $i = 1, \ldots, n$, *of a straight-line segment and a set of data points* (x_i, y_i^*), $i = 1, \ldots, n$, *on a nonlinear function* $y_i^* = f(\alpha, \beta, x_i)$ *with the model parameters* α *and* β *is achieved by preserving the centroid* $(\bar{x}, \bar{y}\ )$ *and the covariance of the data, i.e.,*

$$\bar{y} = \bar{y}^* \tag{2}$$

$$Cov(x,y) = Cov(x, y^*). \tag{3}$$

□

Since $Cov(x,y) = \sum_{i=1}^{n} x_i y_i/n - \bar{x}\ \bar{y}$, the mapping by (2) and (3) is equivalent to

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} f(\alpha, \beta, x_i), \tag{4}$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i = \frac{1}{n}\sum_{i=1}^{n} x_i f(\alpha, \beta, x_i). \tag{5}$$

If the parameters α and β of the nonlinear function are known, equations (2) and (3) can be substituted into (1) to obtain a set of data points y_i on a straight-line segment, which can be viewed a mapping from nonlinear to linear.

Remark 1 (Generalization to m--parameter nonlinear models).
Definition 1 is formulated for two parameters $\alpha, \beta \in \mathbb{R}$. However, an extension of the approach to nonlinear models with $j = 1, \dots, m$ parameters is feasible using m-th-order cross-central moments $\mu_j(x, y) = \sum_{i=1}^{n}(x_i - \bar{x})^{j-1}(y_i - \bar{y})\,/n$ with $j = 2, \dots, m$, where μ_2 is the covariance and μ_3 and μ_4 are directly related to the coskewness and cokurtosis [10]. Then, Definition 1 can be replaced by a mean-value form of a higher-order polynomial such as a parabola for $m = 3$, and Definition 2 can be generalized by replacing equation (3) by

$$\mu_j(x, y) = \mu_j(x, y^*)$$

for a nonlinear function $y_i^* = f(p, x_i)$,where p is a vector of $(m - 1)$ real values. □

The reverse mapping from linear to nonlinear means to determine the unknown parameters, such as α and β from (4) and (5) for the two-parameter case. This could be accomplished using root finding methods, where caution is needed since there might be more than a single root. Thus, a bijective mapping is looked for, which exactly pairs each point of one set of the data with one point of the other set and vice versa. In the next section conditions for bijectivity are derived.

3. BIJECTIVITY OF THE MEAN-VALUE MAPPING FOR THE TWO-PARAMETER CASE

For two parameters, dividing equation (3) by (2) leads to

$$\sum_{i=1}^{n}((x_i - \bar{x}) - Cov(x, y)/\bar{y})\, f(\alpha, \beta, x_i) = 0,\ i = 1, \dots, n, \text{ or}$$

$$\sum_{i=1}^{n} A_i\, f(\alpha, \beta, x_i) = 0 \tag{6}$$

using the abbreviation

$$A_i = (x_i - \bar{x}) - \frac{Cov(x,y)}{\bar{y}}. \tag{7}$$

The sum of equation (6) is used together with equation (4) for determining the unknown parameters α and β of the nonlinear model. If the sum would have a single positive real root, the mapping from linear to nonlinear and vice versa would be bijective.

Definition 3 (Bijective-mapping function). Let $x_i \in \mathbb{R}$, $A_i \in \mathbb{R}$. A function $f_B(\alpha, \beta, x_i)$ with two parameters $\alpha, \beta \in \mathbb{R}$, $i = 1, \dots, n$, is a bijective-mapping function if the sum $F(\beta) = \sum_{i=1}^{n} A_i f_B(\alpha, \beta, x_i)$ has exactly one positive real root for a some sequence A_i .□

Applying the rule of signs by Descartes, a bijective-mapping function is obviously the polynomial $f_B(\alpha, \beta, x_i) = \alpha\, \beta^i$, where $x_i = i = 1, \dots, n$ are integer powers. The rule states that the number of positive real roots of the polynomial

$$F_D(\beta) = \sum_{i=1}^{n} A_i\, \beta^i \tag{8}$$

is at most equal to the number of variations in consecutive coefficients A_i, and the difference of both numbers is an even integer [11]. This means that a single sign change in the sequence of the coefficients A_i leads to exactly one positive real root.

Several bijective-mapping functions exists besides the polynomial (8): In [12], Laguerre generalized the rule of signs by Descartes proposing functions of the form

$$F_L(\beta) = \sum_{i=1}^{n} A_i\, f(\beta x_i). \tag{9}$$

In task # 77 in [13], a generalization is the exponential sum

$$F_V(\lambda) = \sum_{i=1}^{n} A_i\, e^{\lambda x_i} , \tag{10}$$

which is ordered such that $0 < x_1 < x_2 < \cdots x_n$. In [14] the generalization of (10) is proved for $x_i \in \mathbb{R}$ and $A_i \in \mathbb{C}$ $(i = 1, \dots, n)$. The representation (10) is a Dirichlet polynomial [15], which can be transformed to the generalized polynomial

$$F_V(\beta) = \sum_{i=1}^{n} A_i\, \beta^{x_i} , \tag{11}$$

by substituting $e^{\lambda} = \beta$ into (10), where β is positive real to ensure that the function is defined in real numbers. Thus, $F_D(\alpha, \beta)$ turns out a special case of $F_V(\alpha, \beta)$ for $x_i = i$. According to [15], the alternative substitution of $e^{x_i} = b_i$ into (10) leads to the alternative representation of the Dirichlet polynomial

$$F_V(b_i, \lambda) = \sum_{i=1}^{n} A_i\, {b_i}^{\lambda}, \tag{12}$$

for $0 < b_1 < b_2 < \cdots b_n$.

Since both transformations do not alter the number of roots, equations (10) to (12) are equivalent. In [15] it is proved that Descartes' rule of signs holds for all generalizations of (10) to (12). In [16] an axiom systems is given for determining function families which generalize the rule of Descartes, and in [17] a summary and a historical view on the generalization efforts is presented.

Theorem 1. Any bijective-mapping function $f_B(\alpha, \beta, x_i)$ with $\{\alpha, \beta, x_i\} \in \mathbb{R}$, $i = 1, \dots, n$, ensures bijectivity of the mean-value mapping if

$$x_1 - \bar{x} < \frac{Cov(x,y)}{\bar{y}} < x_n - \bar{x} \tag{13}$$

is fulfilled, where $x_1 < x_2 < \cdots x_n$, and $\bar{y} > 0$ is fulfilled as a necessary condition. □

Proof. The zero of the straight line $A_i(x_i)$ of equation (7) is $x_i = Cov(x,y)/\bar{y} + \bar{x}$, and it must lie in between the interval $[x_1, x_n]$ to ensure a sign change of the values of the series A_i once the index i is increased from $1\ to\ n$, resulting in (13), which can alternatively be expressed by

$$\sum_{i=1}^{n} \frac{x_i}{x_n} y_i < \sum_{i=1}^{n} y_i < \sum_{i=1}^{n} \frac{x_i}{x_1} y_i,$$

which is identical to the two inequalities

$$\sum_{i=1}^{n} (x_n - x_i) y_i > 0,$$
$$\sum_{i=1}^{n} (x_i - x_1) y_i > 0.$$

Since both must be fulfilled at the same time, their sum must be positive as well, leading to the necessary condition for (13), which becomes $\sum_{i=1}^{n} (x_n - x_1)\, y_i > 0$ or $\bar{y} > 0$. □

In summary, bijectivity of the mean-value mapping is achieved if both is fulfilled: The inequality (13) holds, and $f(\alpha, \beta, x_i) = f_B(\alpha, \beta, x_i)$ is a bijective-mapping function.

Remark 2. The mean-value ratio

$$\phi = \bar{x} + \frac{Cov(x,y)}{\bar{y}} \tag{14}$$

can be defined to abbreviate a scaling of the covariance. Dividing (5) by (4) yields

$$\phi = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i} \tag{15a}$$

$$\phi = \frac{\sum_{i=1}^{n} x_i\, f(\alpha,\beta,x_i)}{\sum_{i=1}^{n} f(\alpha,\beta,x_i)}, \tag{15b}$$

To express the link for the mean-value mapping. Thus, condition (13) becomes

$$x_1 < \phi < x_n. \tag{16}$$

It should be noted that positive y_i always fulfill the range of (16), which is due to (15a) resulting in

$$x_1 \sum_{i=1}^{n} y_i < \sum_{i=1}^{n} x_i y_i < x_n \sum_{i=1}^{n} y_i,\ i = 1, \dots, n$$

holding for any $y_i > 0$. Conditions for negative values of y_i can be obtained as well.

Remark 3. The introduction of the mean-value ratio ϕ leads to the alternative mean-value representation of any straight line segment

$$\frac{y_i}{\bar{y}} = \frac{1}{Var(x)} (\phi - \bar{x})(x_i - \bar{x}) + 1, \tag{17}$$

by substituting (14) into (1), where $\phi > \bar{x}$ means a positive slope of the line segment.

4. SPECIAL CLASS OF TWO-PARAMETER NONLINEAR MODELS LINEAR IN ONE OF THEIR PARAMETERS.

A two-parameter nonlinear model $y_i^* = f(\alpha, \beta, x_i)$ linear in one of its parameters is

$$y_i^* = \alpha\, g(\beta, x_i). \tag{18}$$

Remark 4 (Applicability to nonlinear functions).
There exist several bijective-mapping functions of practical relevance which qualify for the model type (18): The Box Lucas model or a “simple exponential” model such as the Arrhenius model, the nonlinear-resistance model, the time-power model or the economic model for a compounded growth rate. Other functions may be considered as well that are not proven bijective-mapping functions. An application to the Freundlich adsorption isotherm or the Michaelis–Menten model of chemical kinetics is feasible. - Further, an application to periodic functions could be possible for a certain range of bijectivity: An example of rather theoretical nature is the mapping to a sine-function $y_i^* = \alpha \sin(\beta x_i)$, which is bijective for equidistant data $x_i = i$ in a maximum range of $-\pi/2 \leq \beta \leq \pi/2$ that narrows as the number of values n increases. □

Equation (17) can be used to map a set of data (x_i, y_i^*), located on a nonlinear function (18) to a set of data on a line segment by simply substituting $\bar{y} = \alpha \sum_{i=1}^{n} g(\beta, x_i)/n$ and (15b) into (17). The reverse mapping from a set of data of a line segment to that of a nonlinear model can be accomplished by using the following theorem.

Theorem 2. *The mean-value mapping from a set of any scattered data* (x_i, y_i^o), $i = 1, \dots, n$ *to a set of data points* (x_i, y_i^*), $i = 1, \dots, n$, *located on the function of a non-linear model* $y_i^* = \alpha\, g(\beta, x_i)$ *with the unknown parameters* α *and* β,*where* $\{\alpha, \beta, x_i\} \in \mathbb{R}$, *is achieved by substituting the mean-value ratio*

$$\phi = \frac{\sum_{i=1}^{n} x_i y_i^o}{\sum_{i=1}^{n} y_i{}^o},$$

into the equation

$$\sum_{i=1}^{n} (x_i - \phi)\, g(\beta, x_i) = 0, \tag{19}$$

to determine the parameter β, *which is subsequently used for obtaining the parameter*

$$\alpha = \frac{\sum_{i=1}^{n} y_i^o}{\sum_{i=1}^{n} g(\beta, x_i)}. \tag{20}$$

□

Proof. Dividing equation (5) by (4) eliminates the model parameter α leading to (6), i.e., $\sum_{i=1}^{n} A_i\, g(\beta, x_i) = 0$. Since $A_i = x_i - \phi$ from (7) and (14), equation (19) follows. Equation (20) is obtained by substitution of (18) into (4) by replacing y_i by y_i^o. □

Theorem 2 is formulated for any set of scattered observations which, of course, include all straight-line segments, which can be viewed regression lines through the points of these data sets. Thus, all three data sets, namely those of the scattered data, the straight regression line and the nonlinear function have the same centroid and covariance. - Apparently, Theorem 2 is suitable to be applied to nonlinear regression problems such as in real-time image processing [18]. Thus, the approach may have potential to get employed to geophysical systems and automated geographic mapping [19], pattern recognition [20] and computer graphics [21]. Interestingly, fast collision detection uses the covariance to determine the orientation of bounding boxes providing tight fittings for point clouds which represent moving objects [22].

5. APPLICATION TO EQALLY-SPACED DATA SEQUENCES.

If sets of scattered data points (x_i, y_i^o) are equally spaced by $x_i = i = 1, \dots, n$, the mean-value representation of a straight-line segment yields from (17)

$$\frac{y_i}{y_i^o} = \frac{6}{n-1}\left(\frac{2}{n+1}\phi - 1\right)\left(i - \frac{n+1}{2}\right) + 1. \tag{21}$$

A mapping of data on the line segment (21) to a sequence of data points (i, y_i^*) located on a two-parameter nonlinear model $y_i^* = \alpha\, g(\beta, i)$, which is linear in one of its parameters, can be accomplished by Theorem 1 by equations (19) and (20). Considering equally-spaced data, (19) allows to precompute the unknown β for a given range of the mean-value ratio ϕ. Thus, the unknown model parameters can be obtained by using the look-up table $(\phi, n) \rightarrow \beta$, created once and for all in the range of (16), which is valid for any set of data y_i^o. Thus, the mean-value mapping is considered computationally fast due to all mapping information aggregated in the covariance or the mean-value ratio.

5. CONCLUSIONS.

Preserving the centroid and the covariance of data defines not only the linear least squares approach, but as well the proposed mean-value mapping consisting of two consecutive linear and inverse linear mappings of data on a straight line to a nonlinear function.

Applied to sets of scattered data, the main Theorem 2 suggests a novel method for an unbiased fit to a nonlinear model. It preserves the orientation of the data, whereas nonlinear regression does not. Computational advantages of the proposed covariant-invariant mapping are addressed which are relevant for real-time applications: Instead of using root-finding algorithms, which are needed to solve nonlinear least-squares regression problems, a single one-time created look-up table can be used to determine unknown model parameters in the case of equally-spaced data. Further, a conditions for bijectivity of the mapping are stated, yielding an unique one-to-one “linear twin” of a nonlinear model. Further work could focus on the generalization to models with m parameters by taking m-th-order cross-central moments into account such as the coskewness and cokurtosis. Further, the applicability of the covariance-invariant mapping to various fields ranging from econometrics to computer graphics may be worthwhile to explore.

REFERENCES

[1] Tweedie MCK. Some statistical properties of Inverse Gaussian distributions. *The Annals of Math. Stat.* 1956, 28: 362-377.

[2] Jorgensen B. Exponential Dispersion Models. *J. R. Statist.* 1987, Soc. B 49: 127-162. DOI: 10.1111/j.2517-6161.1987.tb01685.x.

[3] Frasinski LJ, Codling, K, Hatherly PA. Covariance Mapping: A Correlation Method Applied to Multiphoton Multiple Ionisation. *Science* 1989; 246:1029–1031. DOI: 10.1126/science.246.4933.1029.

[4] Slater M, Steed A, Chrysanthou Y *Computer Graphics and Virtual Environments – From Realism to Real-Time*. Pearson Education, Harlow, 2002.

[5] Hiai F., Petz D *Introduction to Matrix Analysis and Applications*, Springer Verlag, Heidelberg, 2014. DOI: 10.1007/978-3-319-04150-6.

[6] Adikaram KKLB, Jayantha PA. Zero mean transformation technique that is not effected by missing or removed data. *Proc. 7th Int. Symp. on Multidisciplinary Research for Sustainable Development*, South Eastern University of Sri Lanka, 7th - 8th December, 2017, Sri Lanka, 2017, 881-885.

[7] Hall, PM. Introduction to Machine Learning for Computer Graphics. *ACM SIGGRAPH '14*, Association for Computing Machinery, New York, NY, 2014. DOI: 10.1145/2614028.2615461.

[8] Treeby D. A Moment's Thought - Centers of Mass and Combinatorial Identities. *Math. Mag.* 2017, 90: 19-25. DOI: 10.4169/math.mag.90.1.19.

[9] Bickel PJ, Doksum KA *Mathematical Statistics: Basic Ideas and Selected Topics*, Vol. I, 2nd ed., Prentice Hall, Upper Saddle River, NJ; 2001, 107-111.

[10] Miller M *Mathematics and Statistics for Financial Risk Management*. 2nd ed., Wiley Finance. Wiley, 2014; 53-55. DOI:10.1002/9781118819616.

[11] Descartes R. Combien il peut y auoir de vrayes recines en chaque Equatio, *Des matieres de la géométrie*, 3e série, 1637; 373. Translated by Struik DJ *A Source Book in Mathematics 1200-1800*, Princeton University Press, Princeton, NJ, 1986; 91.

[12] Laguerre EN. Mémoire sur la théorie des équations numériques. *Journal de mathématiques pures et appliquées* 1883, 3e série, 9: 123.

[13] Pólya G, Szegö G *Aufgaben und Lehrsätze aus der Analysis II*, XX of the series: Die Grundlagen der Mathematischen Wissenschaften in Einzeldarstellungen, Springer-Verlag, Berlin, unaltered ed., 1954, 1925; 49.

[14] Voorhoeve M. A Generalization of Descartes' Rule. *J. of the London Math. Society*, s2-20 (1979), 446–456. DOI: 10.1112/jlms/s2-20.3.446.

[15] Jameson GJO. Counting zeros of generalized polynomials: Descartes' rule of signs and Laguerre's extensions. *Math. Gazette* 2006, 90:223-234. DOI:10.2307/40378610.

[16] Haukkanen P, Tossavainen T. A generalization of Descartes' rule of signs and fundamental theorem of algebra. *Appl. Math. and Comp.* 2011, 218: 1203-1207. DOI: 10.1016/j.amc.2011.05.107.

[17] Rahman QI, Schmeisser G *Analytic Theory of Polynomials*, London Mathematical Society Monographs, New Series 26, Oxford University Press, 2002; 353-365.

[18] Liu W, Zhang P, Huang X, Yang J, Shen C, Reid I. Real-time Image Smoothing via Iterative Least Squares. *ACM Trans. Graph.* 2020, 39: 1-24. DOI: 10.1145/3388887.

[19] Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Comm. of the ACM* 1981, 24: 381–395. DOI: 10.1145/358669.358692.

[20] Tobler WR. Comparing figures by regression. *ACM SIGGRAPH Computer Graphics*, 1978, 12: 193-195. DOI: 10.1145/965139.807389.

[21] Fleishman S, Cohen-Or D, Claudio Silva C. Robust moving least-squares fitting with sharp features. *ACM Trans. Graph*. 2005, 24: 544-552. DOI:10.1145/1186822.1073227.

[22] Mukundan R *Advanced Methods in Computer Graphics*, Springer-Verlag, London, 2012. DOI: 10.1007/978-1-4471-2340-8.