Chapter 21
The Method of Least-Squares

D.D. Kosambi, Poona

The English version of this paper appeared two years after the Chinese "original." During the 1950s and early 1960s, DDK visited China several times on exchange programs. This paper was probably written when he visited the Academia Sinica on an exchange program between India and China as an expert in statistics from TIFR [DDK-JK]. *This was a visit of several months, ample time for DDK to write his paper and have it translated into Chinese.*

This note begins with a discussion of possible metrics in probability spaces associated with independent random variables; the Euclidean metric (in suitable coordinates) turns out to be the only one admissible. The method of least squares is known to be derived from such a concept of distance. In the second section, a unique least-squares solution is derived for general linear systems of equations in abstract spaces even when there may be no proper solution in the usual sense, the two coinciding when the ordinary solution exists. This is of considerable importance for diffusion theory and the integral equations for atomic energy piles. The final section gives a sketch of the extension to general nonlinear systems of equations.

1. We start with a system of measurable sets called "simple events" such that the adjunction of the "compound events" obtained by set addition and set multiplication gives an aggregate of measurable Borel sets constituting a Boolean set algebra. The union $A \cup B$ of two sets is the compound event "A or B"; the intersection $A \cap B$ is the compound event "A and B (simultaneously)"; the operational laws for the dual operations "cap" $= \cap$ and "cup" $= \cup$ being as usual in Boolean algebra, which

Published in the Journal of the Indian Society of Agricultural Statistics **11**, 49–57 (1959). The Chinese version appeared in Advancement in Mathematics **3**, 485–491 (1957). Reprinted with permission.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_21

contains the null set O and the universe I. The probability measure is regulated by the postulates [1]:

$$P(I) = 1. \tag{21.1a}$$

$$P(A) \geq P(B) \text{ if } A \cup B = A, \ i.e., \text{ if } A \supset B. \tag{21.1b}$$

$$\text{If } A \cap B = O, \ P(A \cup B) = P(A) + P(B). \tag{21.1c}$$

Taking $A = I$, $B = O$ in Eq.(21.1c), it follows that $P(O) = O$. With (21.1a) and (21.1b), this gives $O \leq P(A) \leq 1$ for all sets of the ensemble. Finally, seeing that $A \cup B$ is the union of three mutually non-intersecting sets $(A - A \cap B)$, $A \cap B$, $(B - A \cap B)$, we obtain the general result:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)\,. \tag{21.2}$$

This could have been substituted for the third postulate in (21.1a, 21.1b and 21.1c) with the addition of $P(O) = O$. The events are to be regarded as reduced modulo, the ideal of all sets with measure zero. The restriction to Borel sets, though not always necessary, permits infinite repetition of the two operations $\cup, \cap$.

Definition *Two events A, B such that $A \cap B = O$ are called mutually exclusive. Non-null events $A_1, A_2, \ldots A_n \ldots$ such that $P(A_i \cap A_j \cap A_k \ldots) = P(A_i)P(A_j)$ $P(A_k) \ldots$ for any finite section $i, j, k \ldots$ are called mutually independent events.*

It follows that two mutually exclusive events cannot be mutually independent, nor can two events one of which wholly includes the other; these are the extreme case of zero and unit conditional probability, always omitting from the classification the trivial extremes, O, I. Starting with any simple event of the algebra, we can build an ordered maximal chain of such simple events, with O and I at the two ends, each event of the chain including all preceding members and being included in all that follow, while no other simple event of the algebra outside the chain has this property, with respect to all sets of the particular chain. We consider hereafter *only such Boolean probability algebras whose simple events can be split up into a finite number of maximal chains, every event of each chain being independent of every event in any other chain*.

In the first place, each such chain can be mapped upon the real line segment $(0, 1)$ by the correspondence $A \to [O, P(A)]$. But we need also a map on the whole real axis $-\infty \leq x \leq +\infty$, which is connected with the $(0, 1)$ measure map by a distribution function $F(x)$, which is monotonically non-decreasing, with $F(-\infty) = 0$, $F(+\infty) = 1$. Any set A of the chain can be mapped upon the interval $(-\infty, \alpha)$ on the line such that $F(\alpha) < P(A)$ if $x < \alpha$, while $F(a) = P(A)$. Using one dimension for each such ordered chain, we map the Boolean algebra upon an n-dimensional continuum $(x_1, x_2 \ldots x_n)$, where the image of a simple event is a section from $-\infty$ to $+\infty$ in all dimensions except one, where the section extends only from $-\infty$ to α. The measure image on the unit hypercube is the rectangular parallelepiped of side unity

in all except one dimension, where the side is the interval $[O, P(A)]$. Compound events are derived from these by set union and set intersection.

Theorem 1 *If an n-dimensional probability space be associated with a Boolean algebra of events such that each dimension represents a chain of events independent of all the others, and if the space is endowed with a Riemann metric plus a measure function which give a true map upon the unit hypercube, then the metric can only be Euclidean.*

Proof For the Riemann metric, $ds^2 = \Sigma g_{ij} dx_i dx_j$. The measure of any k-dimensional event in the x-space is given, for $1 \leq k \leq n$ by an integral of the form $\int f_k(x_{i_1}, \ldots, x_{i_k}) \sqrt{|g_{ij}|}\, dx_{i_1} \ldots dx_{i_k}$. But if the region be the compound event $A_1 \cap A_2 \cap \cdots \cap A_k$, it follows that the integral must break up into a product of k separate integrals for all $k \leq n$. Therefore, any principal minor as well as the whole determinant $|g_{ij}|$ must reduce to a product of diagonal terms: $g_{11}(x_1)$ $g_{22}(x_2) \cdots g_{nn}(x_n)$, and correspondingly for each of its principal minors. The measure function f, essentially the derivative of the distribution, assumed to exist and be continuous, will similarly break up into a product of factors, but that is of lesser interest here. It is clear that the cross terms of the tensor g_{ij} all vanish, with $ds^2 = g_{11}(x_1)dx_1^2 + g_{22}(x_2)dx_2^2 + \cdots + g_{nn}(x_n)dx_n^2$. The g_{ii} are positive from the hypothesis of positive measure for any chain (we need not invoke the positive definition form of the metric here), permitting a transformation of coordinate variables defined by $dx'_r = \sqrt{g_{rr}}\, dx_r$. These are the Euclidean coordinates of the space. □

We have two simple corollaries:—

Corollary 1 *If the space of n random variables be endowed with a Riemann metric and a measure (distribution density) function which permit the original random variables to be replaced by n independent random functions thereof, then the curvature tensor of the original space must vanish, the space being Euclidean.*

The new variables amount simple to a non-singular transformation of coordinates. But there, the space will have the Euclidean coordinates of the preceding theorem; hence its curvature tensor will vanish in both coordinate systems. For the second corollary, we need a topological result, [2] that a Riemann metric exists when the space may be covered by neighborhoods such that each pair of points may be joined by one and only one arc (lying wholly within the neighborhood) of a previously defined class which we may call paths. Then, for any compact portion of the space that can be so covered, a Riemann metric can be assigned whereof the given paths are actually the geodesics. In the present case, we have one compact space, namely the unit hypercube, to which the result may be applied, working back to the original space if the f function is continuous, giving us:

Corollary 2 *If the space of random variables is only endowed with a continuous measure density function, and a set of continuous paths with the property that any two points sufficiently close have a unique path join, then the space also possesses a Riemann metric, hence is Euclidean if the concept of independent random variables is applicable by suitable transformation.*

With Euclidean space, if the compound probability density function of several independent random variables depends only upon the distance, it follows immediately that the distribution of each variate must be normal (Gaussian) [3]. From this to the usual motivation of least squares is only one step, for the best approximation to the population mean from a sample is that which minimizes the sampling variance, which is a sum of squares (the distance, in fact, to a hyperplane), hence the arithmetic mean.

2. We deal throughout with real variables, though the extension to complex or other number systems causes little difficulty. The system of m linear equations in $n < m$ real variables

$$\sum_{j=1}^{n} A_{ij}x_j - y_i = 0\,; \quad i = 1, 2, \ldots, m > n \tag{21.3}$$

has no solution in general. But it has always a least-squares solution minimizing

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j - y_i \right)^2 , \tag{21.4}$$

thereby, specifying the values of x as solutions of the n equations:

$$\sum_{r=1}^{n} C_{kr}x_r - z_k = 0\,, \quad C_{kr} = \sum_{q=1}^{m} A_{qk}A_{qr}\,, \quad z_k = \sum_{q=1}^{m} A_{qk}y_q\,. \tag{21.5}$$

Here, every free index runs through the values $1, 2, \ldots, n$. The system (21.5) has a unique solution in general, coinciding with the exact solution of (21.3) should those equations be compatible. Clearly, we can take formal passage to the limit to an integral equation of the first kind:

$$\int A(s,t)x(t)dt = y(s)\,, \tag{21.6}$$

and to other general linear systems. This is the work of the present section, regardless of probability considerations.

We begin with a vector space V over the field C of all real numbers, the elements $x, y, \cdots$ being in V and constants $a, b \cdots$ in C give $ax + by + \cdots$ also in V. We further require a symmetric bilinear scalar product $x \cdot y$ as a mapping of $V \times V$ into C, with the properties: $x \cdot y = y \cdot x$, and $x \cdot (ay + bz) = a(x \cdot y) + b(x \cdot z)$. This leads to a quadratic *norm* $x \cdot x$ of which we demand that $x \cdot x = 0$ if and only if $x = 0$, which amounts to reduction of V with respect to elements of zero norm. We shall assume that V is complete with respect to convergence in the norm. The usual condition that the norm be positive is easily imposed, for it must always be of the

same sign. If there were two distinct elements x, y with $x \cdot x > 0$, $y \cdot y < 0$, the quadratic in $\lambda : (x + \lambda y) \cdot (x + \lambda y) = 0$ would have real roots, giving an element with vanishing norm, of the form $x + \lambda y$. But this cannot be zero identically, for then $x \cdot x = \lambda^2 (y \cdot y)$, which is impossible because the two norms had initially opposite signs. Hence, the norm must always have the same sign, and there is no loss of generality in taking it always positive.

We avoid the trivial cases where V contains only the element 0, or only multiples of a single element ϕ. Two nonzero elements ϕ, ψ are defined as orthogonal if their scalar product vanishes: $\phi \cdot \psi = 0$, while an element with unit norm (always to be had by multiplication with a suitable constant) is called normal. The assumption is that V has an orthonormal basis $\phi_1, \phi_2, \ldots \phi_n, \ldots$ not necessarily finite, but (by the Hilbert theorem) at most denumerable, and that the Riesz–Fisher theorem applies so that with any convergent Σa_r^2, there always exists a function in V represented by $\Sigma a_n \phi_n$; this is necessary for the completeness of the space, which we have assumed.

To correspond to the matrices in (21.3), we need two-sided linear associative operators $S, T, \ldots$ defined over V, i.e., Tx and $xT \subset V$ for all $x \subset V$; with $(ax + by)T = a(xT) + b(yT)$, $T(ax + by) = a(Tx) + b(Ty)$. For xT, we shall also write $T^* x$, the *adjoint* of T. This adjoint is governed by the operational rule: $(T^*)^* = T$. If we define the operator product ST by $(ST)x = S(Tx)$, with $x(ST) = (xS)T$, it follows that $STx = S(xT^*) = (xT^*)S^*$, whence $(ST)^* = T^* S^*$, the star operation for the adjoint of these linear operators thus satisfying four of the basic postulates for a C^* algebra in the sense of Gelfand and Neimark. We may write SxT for $S(xT) = ST^* x = xTS^* = T^* xS^*$, according to convenience, without confusion. The scalar product $x \cdot (Ty)$ is similarly abbreviated $xTy = yT^* x$, at will.

Using the orthonormal basis for V, it is seen that the T operation amounts to a linear matrix transformation for the coordinates (Fourier coefficients) of an element. All operations may be visualized and theorems proved by use of the matrix representation. For Hilbert spaces (vector spaces with infinite basis), the argument has to be restricted in general to such operators as may be separated into two additive portions of which one is finite dimensional, the other with arbitrarily small norm. That is, the operators must be *bounded*: $(Tx) \cdot (Tx) \leq M(x \cdot x)$ for all $x \subset V$, M depending only upon T. We shall deal only with non-singular bounded operators, and remark that a symmetric operator such that $T = T^*$ has always a real spectrum. To each T, there correspond always the two symmetric operators TT^* and $T^* T$, of which the latter is assumed to have a discrete spectrum for our main result.

The entire least-squares procedure rests upon the following [4]

Lemma *The orthonormal portion of V which does not lie in TV is mapped into zero by T^* that is, $T^*(V - TV) = 0$.*

Proof If the transformed space TV is the whole of V, the result is trivially true. If not call $\bar{V}$, the orthonormal component of V not in TV. The generic scalar product of a function in $\bar{V}$ and another in V is $\bar{V}TV = VT^* \bar{V}$. By hypothesis, this scalar product is zero, which means that every element in the whole of V is orthogonal to every element in $T^* \bar{V}$, which is impossible, except when $T^* \bar{V} = 0$, proving the lemma. □

The main least-squares equation now takes the form

$$Tx - y = 0 \tag{21.7}$$

with T, y given, x to be found.

This need not have a solution at all, as Tx lies necessarily in TV, while the given y may have a component outside TV. The norm of the left-hand side is

$$(Tx - y) \cdot (Tx - y) \equiv xT^*Tx - 2(xT^*y) + (y \cdot y)\,. \tag{21.8}$$

To minimize this, give a variation to x, replacing x by $x + \delta x$. Subtracting the original value in (21.8) from the varied value gives

$$2\delta x \cdot (T^*Tx - T^*y) + (T\delta x) \cdot (T\delta x)\,. \tag{21.9}$$

The coefficient of δx is equated to zero for a minimum as usual, for the remainder is positive, while we take the norm of δx as tending to zero. This gives us the following:

Theorem 2 *The least-squares solution of* $Tx - y = 0$ *is given by*

$$T^*Tx - T^*y = 0\,. \tag{21.10}$$

Our lemma $T^*(V - TV) = 0$ makes the solution possible. Naturally, there are some simple restrictions upon the operator in question. In terms of eigenfunctions and eigenvalues, these give Picard's solution [5] of integral equations of the first kind and the corresponding least-squares solution, which may be subsumed in the following:

Theorem 3 *The least-squares solution of* $Tx - y = 0$ *exists if and only if* $\Sigma(\phi_n T^*y)^2/\lambda_n^4$ *converges, where* ϕ_n *and* λ_n^2 *are the eigenfunctions and eigenvalues respectively of* $T^*T\phi - \lambda^2\phi = 0$. *In particular, if the orthonormal set* $\{\phi_n\}$ *furnish a basis for* V, *we have the exact solution (for that portion of* V *in which* y *lies).*

The proof is as follows: The non-singular operator T^*T leaves the origin invariant in V, hence by continuity maps some portion of V on to some neighborhood of O, in the map space T^*V. We (assume the operator T^*T to have a discrete spectrum, and) expand T^*y in terms of the eigenfunctions. The lemma above says that T^*y cannot be orthogonal to all these eigenfunctions without vanishing identically, while the condition of the theorem merely requires T^*y to lie in the transformed neighborhood of the origin.

The result is independent of the norm. That is, our norm was best taken with respect to the identity, the symmetric operators $xI = Ix = x$ for all $x \subset V$. Any other symmetric operators may be used for the least-squares norming provided $SV = V$, and $xSx = 0$ if and only if $x = 0$. The result is of great use in the solution of integral equations when nothing is known about the closure of the eigenfunctions of the particular kernel.

3. The square sum (21.4) of the linear equation (21.3) amounts to the sum of weighted squares of distances from a generic point (x) to the various hyperplanes. The same idea can be extended, therefore, to nonlinear hypersurfaces. We look for the point or points from which the sum of squares of distances to a given set of (weighted) hypersurfaces is minimum, which is included in the set of points where the distance sum is stationary and which is all we shall investigate without insisting upon a true minimum. The geometric picture tells us that the point sought is common to all the surfaces if they have a common interesection, or that point which lies on the intersection of normals to each of the surfaces. In mathematical notation, let the surfaces be:

$$f_1(x_1, x_2, \ldots, x_n) = c_1,\ f_2(x) = c_2, \ldots f_m(x) = c_m\,;\quad m > n\,. \tag{21.11}$$

The point sought is the solution of the equations:

$$\frac{\partial F}{\partial x} = 0\,,\quad \frac{\partial F}{\partial u} = 0\,,\quad \frac{\partial f}{\partial v} = 0\,, \tag{21.12}$$

where

$$F \equiv \sum_i (x_i - u_i)^2 + (x_i - v_i)^2 + \cdots + \lambda_1 f_1(u) + \lambda_2 f_2(v) + \cdots$$

and the unindexed letters $x, u, v \ldots$ each represent the set of n variables, the index being understood, even in the partial differentiation. This is Lagrange's method of multipliers leading to two sets of equations:

$$x_i = \frac{u_i + v_i + \cdots}{m} \tag{21.13a}$$

$$2(x_i - u_i) - \frac{\lambda_1 \partial f_1}{\partial u_i} = 0\,,\ 2(x_i - v_i) - \frac{\lambda_2 \partial f_2}{\partial v_i} = 0\,, \ldots \tag{21.13b}$$

These lead to compatibility conditions:

$$\frac{\lambda_1 \partial f_1}{\partial u_i} + \frac{\lambda_2 \partial f_2}{\partial v_i} + \cdots = 0\,;\quad i = 1, 2, \ldots, n\,, \tag{21.14}$$

which merely reflects our previous lemma $T^*(V - TV) = 0$ in the total extended space. For linear equations, the process is as before, and for the general case, the extension is fairly clear.

We begin with an abstract vector space V such that $x \subset V$. This V is extended over a variety which was formerly a finite Abelian group and may for the present be taken as an indexed variety. The extension is then indicated by V_α, with variables $u_\alpha \subset V_\alpha$. The α-space has to be compact, with an abstract integral which we shall denote by Σ_α and which has the properties of a Lebesgue–Stieltjes integral, while $\Sigma_\alpha 1 = M$. The scalar product is defined over the extended space as $\Sigma_\alpha (x - u_\alpha) \cdot (x - u_\alpha)$. Finally,

$f(x)$ is a general operator, mapping x into the real field, $f_\alpha(u)$ being the (suitably but completely defined) extended operation, understood as $f_\alpha(u_\alpha)$.

We need further the generalized partial differentiation, which is defined as the infinitesimal operator of the (Abelian) Lie group in the space of $f(x)$ when the base space of x undergoes a translation $x \to x + h$; the Lie group is generated by the usual exponential representation, which leads to a Lie-Taylor series expansion which is the formal representation, and in the analytic case converges to give an exact representation. Our nonlinear functional operators f need not be analytic nor even arbitrarily differentiable, for they may be approximated by such at need; but the f-operators must at least be continuous in the first derivative for the analogue of (21.12) to be valid. By introducing an orthonormal basis and coordinates for V, the partial derivative becomes just the ordinary partial derivative in the coordinates; generically, we represent this by f'. The results are then summed up as follows:

The least-squares solutions of the simultaneous projections $f_\alpha(x) = 0$ *are given by* $x = (1/M)\Sigma_\alpha u_\alpha$, *provided the extended variables* u_α *satisfy*

$$\lambda_\alpha f'_\alpha(u_\alpha) = \left(\frac{1}{M}\right)\Sigma_\alpha u_\alpha - u_\alpha \,. \tag{21.15}$$

The λ_α *and* u_α *being so chosen as to further satisfy* $f_\alpha(u_\alpha) = 0$.

References

1. A. Kolmogorov, Grundbegriffe der Wahrscheinlichkeitsrechnung, Ergebnisse d. Mathematik, **2**(3), (1933). (Berlin, the opening sections).
2. D.D. Kosambi, The metric in path-space. Tensor **3**, 67–74 (1954).
3. D.D. Kosambi, The geometric method in mathematical statistics. Am. Math. Mon. **51**, 382–389 (1944).
4. L.H. Loomis, in *Introduction to Abstract Harmonic Analysis* (New York, 1953), p. 27. (for the classical result).
5. R. Courant, D. Hilbert, in *Methoden der mathematischen Physik I*, (Berlin, 1931), pp. 134–36. (E. Picard, in *Rendiconti del Circolo Mathematico di Palermo* **29**, 79–97 (1910)).
6. D.D. Kosambi, An extension of the least-squares method for statistical estimation. Ann. Eugen. **13**, 257–261 (1947).