The Determination of the Location of the Global Maximum of a Function in the Presence of Several Local Extrema

CORNELIS H. SLUMP AND BERNHARD J. HOENDERS

Abstract—**The global maximum of a function can be determined by using information about the number of stationary points in the domain of interest. This information is obtained by evaluating an integral that equals the exact number of stationary points of the function. The integral is based on work by Kronecker and Picard at the end of the nineteenth century. The numerical feasibility of the method is shown by two computed examples, i.e., estimation problems from statistics and optical communication theory. In these examples the global maximum of the likelihood function is obtained by using the total number of stationary points as revealed by the computed integral.**

I. Introduction

In practice, the use of maximum likelihood estimation often leads to a difficult nonlinear numerical problem (see, e.g., Gupta and Mehra [1]). The location of the global maximum of the likelihood function is of special importance for estimation purposes. Only in specific cases (e.g., sufficient statistics or linear likelihood equations) can the global maximum be computed directly. In all other cases, one has to resort to iterative techniques because exhaustive grid search algorithms require a prohibitive amount of computer time. The determination of the global maximum of a function is a notorious numerical problem when there are several local maxima present. The iterative algorithm is in this case very likely to produce a local maximum in the neighborhood of the initial guess. For the applicability of maximum likelihood estimation in practice, good starting values are of great importance, but there is no general method for obtaining them. Usually one repeats the computation and varies the value of the initial guess. The optimum thus obtained is assumed to be the global maximum. However, one cannot be sure not to have missed a maximum; i.e., various computational procedures together with various initial values still might overlook the global optimum, especially when the number of parameters increases. Therefore, information about the total number of local extrema present in a domain would be of great value because it indicates whether or not the repeated iterative procedure has missed a maximum.

This information is provided by an integral derived by Picard [2] from previous work by Kronecker [3] at the end of the nineteenth century. The integrand consists of simple algebraic quantities containing the derivatives up to the third order of the function involved. The integration is over the domain of interest. A discussion of this integral, illustrated, by examples, is given in Hoenders and Slump [4]. This so-called Kronecker–Picard (KP) integral yields the exact number of stationary points of the function to be optimized, i.e., the likelihood function $L(\boldsymbol{x})$. The stationary points of a function $L(\boldsymbol{x})$ are the zeros of the set of equations $\nabla L(\boldsymbol{x}) = 0$. The KP integral is equal to the number of zeros of a set of equations in a domain, provided that all these zeros are simple. We will not consider here the case of multiple zeros (see [5], [6]). The case of multiple zeros will be analyzed in a forthcoming paper, as it is not a trivial extension of the theory presented in this paper. Note that a multiple zero of multiplicity n will contribute the value n to the KP integral. This can be understood heuristically from the remark that a zero of multiplicity n can be considered as the confluence of n simple zeros, which in a certain domain contribute n to the KP integral. We further refer to (19), where $f = x^n$, and R is a rectangle enclosing the point $x = 0$.

The next section is devoted to the underlying theory of the KP integral. The use of this elegant formalism for obtaining the global optimum in maximum-likelihood estimation is illustrated by an example from statistics and an example from optical communication theory.

II. Theory of the Kronecker–Picard Integral

The basic idea of the KP integral is founded in the generalization of the concept of the solid angle into higher dimensional space. The differential solid angle $d\Omega$ with respect to a point P and a surface element $d\sigma$ in $\mathbb{R}^n$ is geometrically defined as the area of the projection of $d\sigma$ on the surface of the unit sphere with origin in P. This is shown in Fig. 1 for $\mathbb{R}^3$ [7, p. 290].

An analytical equation for $d\Omega$ is obtained from the geometrical definition by the following argument. Let $\boldsymbol{r}$

Manuscript received July 1, 1983; revised November 30, 1984. This work was supported by the Netherlands Organization for the Advancement of Pure Research (ZWO). This paper was presented in part at the Fourth Symposium on Information Theory in the Benelux, Louvain, Belgium, May 26–27, 1983.

C. H. Slump was with the Department of Applied Physics, Rijksuniversiteit Groningen, Nijenborgh 18, 9747 AG Groningen, The Netherlands. He is now with the Philips Medical Systems Division, Product Group Imaging, QJ-1 Best, The Netherlands.

B. J. Hoenders is with the Department of Applied Physics, Rijksuniversiteit Groningen, Nijenborgh 18, 9747 AG Groningen, The Netherlands.

0018-9448/85/0700-0490$01.00

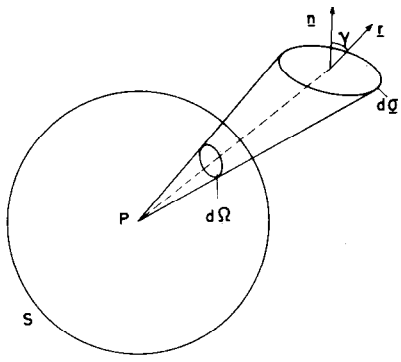

Fig. 1. Differential solid angle $d\Omega$ with respect to P is projection of surface element $d\sigma$ with normal vector n on sphere S with radius 1 about P.

denote the vector connecting the point P and $d\sigma$. The area of the projection of $d\sigma$ on the normal plane at r is da. Then, if γ denotes the angle between r and the normal n of $d\sigma$, i.e., $d\sigma = nd\sigma$, we have

$$d\sigma = \sec(\gamma)\, da = \frac{r}{n \cdot r}\, da, \tag{1}$$

with $r = |r|$ and $d\sigma = |d\sigma|$. Observing that the area da equals r^{n-1} times the projected area of $d\sigma$ on the unit sphere, we have

$$d\Omega = \frac{n \cdot r}{r^n}\, d\sigma. \tag{2}$$

We intuitively expect from the geometrical definition that the total solid angle

$$\Omega = \int_{\partial D_n} \frac{n \cdot r}{r^n}\, d\sigma, \tag{3}$$

where ∂D_n denotes a closed surface in $\mathbb{R}^n$, is independent of the position of the point P as long as P does not cross ∂D_n. The proof of this statement follows from the vanishing of the divergence of the vector field $r^{-n}r$,

$$\nabla \cdot r^{-n} r = 0, \qquad r \neq 0, \tag{4}$$

together with Gauss' theorem,

$$\int_{\partial D_n} r^{-n} n \cdot r\, d\sigma - \int_{\partial D_n'} r^{-n} n \cdot r\, d\sigma' = \int_{D_n} \nabla \cdot r^{-n} r\, d\tau = 0, \tag{5}$$

if $d\tau$ is the volume element of the domain D_n, bounded by the two closed surfaces ∂D_n and $\partial D_n'$ enclosing P.

This result can be easily verified by using geometrical reasoning considering a simple closed curve C in $\mathbb{R}^2$, (viz. Fig. 2).

Let ds denote the length of an infinitesimal line element of the curve C at the point $P' \in C$ at distance r from P. If this curve happens to osculate at the circle with radius r about P, we have $ds = rd\phi$. The projection of ds on this circle equals $rd\phi$. We therefore obtain

$$\cos\gamma\, ds = r\, d\phi, \tag{6}$$

if γ denotes the angle between the two tangent vectors of the circle and the curve C at P'. Recalling that

$$\cos\gamma = \frac{n \cdot r}{r}, \tag{7}$$

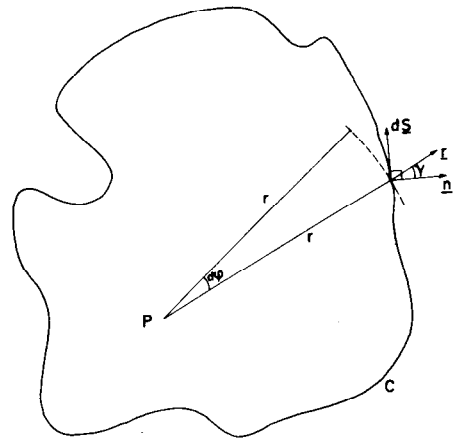

Fig. 2. Simple closed curve C in $\mathbb{R}^2$.

combining (6) and (7) yields

$$\frac{n \cdot r}{r^2}\, ds = d\phi. \tag{8}$$

The total solid angle Ω is obtained by integrating (8) along C. The quantity Ω therefore denotes the total increase of the angle ϕ with respect to P going once along the simple closed curve C. Thus Ω equals 2π if P is enclosed by C; it equals zero if P lies outside C. The latter result can also be derived by drawing a line through p that intersects C in two points, p' and p''. It is obvious that the contributions to Ω of line elements ds' and ds'' at p' and p'' cancel each other, because the sign of $d\phi'$ is opposite to the sign of $d\phi''$, and $d\phi'$ and $d\phi''$ have equal magnitude. (See Fig. 3.)

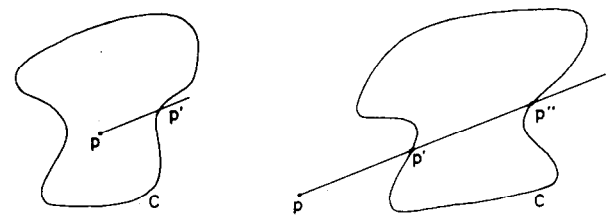

Fig. 3. Dependence of solid angle Ω on location of P inside or outside simple closed curve C in $\mathbb{R}^2$.

The further development of the theory is greatly facilitated by using exterior differential forms. A short heuristic discussion of exterior differential forms, i.e., the exterior product used in this paper, is given in the Appendix. The aim of the exterior calculus is to develop a systematic procedure for the bookkeeping of the partial derivatives occurring when changing the variables in a surface integral or transforming a surface into a volume integral, etc.

The procedure for this calculus is as follows.

1) Replace the products of differentials occurring in the pertinent integral by the exterior products.
2) Apply the appropriate transformation rules, e.g., (11), (A.6).

3) Replace the exterior products by the scalar product in the resulting integral, which yields the expression for the transformed integral.

An example is given in the Appendix, viz. (A.7) and (A.9). We will now proceed with the formulation of the theory of the solid angle in terms of exterior differential forms. Equation (2) and the procedure stated in 1)–3) suggest the following definition of the differential solid angle in terms of an exterior product:

$$d\Omega = \sum_{j=1}^{n} (-1)^{j-1} y_j r^{-n} dy_1 \wedge dy_2 \wedge \cdots dy_{j-1} \wedge d_{j+1} \wedge \cdots \wedge dy_n, \quad (9)$$

where

$$r = \left(y_1^2 + y_2^2 + \cdots y_n^2\right)^{1/2}. \quad (10)$$

We will calculate the total solid angle Ω, integrating $d\Omega$ over a closed surface ∂D_n. A very important property of $d\Omega$ is its closure ([8, Ch. VI, Sec. 8, eq. 54], (A.15) and (A.16)); i.e., a deformation of ∂D_n does not change Ω as long as no singularity of the integrand is crossed by the deformation. From [8, Ch. VI, Sec. 3, eq. 26]

$$dy_1 \wedge dy_2 \wedge \cdots \wedge dy_{j-1} \wedge dy_{j+1} \wedge \cdots \wedge dy_n = \sum_{i_1 < i_2 < \cdots < i_n} \frac{\partial(y_1, y_2, \cdots y_{j-1} \cdots y_{j+1} \cdots y_n)}{\partial(x_{i_1}, x_{i_2}, \cdots x_{i_{n-1}})} \cdot dx_{i_1} \wedge dx_{i_2} \cdots \wedge dx_{i_{n-1}}, \quad (11)$$

where $i_j \in \{1, \cdots n\}$, $j = (1, \cdots n)$, and the summation of the Jacobians occurring at the right side of (11) extends over all possible combinations of $i_1, \cdots i_{n-1}$, which are ordered with respect to their magnitude. For an example of this scheme see (A.8).

Combining (9) and (11) yields

$$\Omega = \int_{\partial D_n} \sum_{j=1}^{n} \sum_{i_1 < i_2 < \cdots < i_n} (-1)^{j-1} r^{-n} y_j \cdot \frac{\partial(y_1, \cdots y_{j-1}, y_{j+1} \cdots y_n)}{\partial(x_{i_1}, x_{i_2}, \cdots x_{i_{n-1}})} dx_{i_1} \wedge \cdots \wedge dx_{i_{n-1}}. \quad (12)$$

When the exterior products are replaced by the ordinary products, this result is equal to the integrals considered by Kronecker [3] and Picard [2],

$$\Omega = \int_{\partial D_n} r^{-n} \sum_i (-1)^{(i-1)n} \cdot \begin{vmatrix} y_1 \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_{i-1}} & \frac{\partial y_1}{\partial x_{i+1}} & \cdots & \frac{\partial y_1}{\partial x_n} \\ y_n \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_{i-1}} & \frac{\partial y_n}{\partial x_{i+1}} & \cdots & \frac{\partial y_n}{\partial x_n} \end{vmatrix} \cdot dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n. \quad (13)$$

The equivalence of (12) and (13) is immediately derived from (13) by using Laplace's expansion rule with respect to the first column of the determinant occurring in (13).

The integral (12) is evaluated easily because, by the closure of the differential form (9), its value is unaffected if we change ∂D_n as long as we do not cross a singularity of the integrand. The singularities of the integrand are the zeros x_l of the set of equations

$$f_j(x) = 0, \qquad j = 1, \cdots, n \quad (14)$$

because r^{-n} is singular for these points. Therefore, we can decompose the integral (12) into a sum of surface integrals ω_l with ∂D_n^l by enclosing a zero x_l of (14). A typical integral ω_l has the form (12) or (13) with boundary ∂D_n^l, and (9) and (10), together with the rules for transformation of integrals, show that

$$\omega_l = \delta_l \int_{\partial D_n'^l} d\Omega, \quad (15)$$

where $\partial D_n'^l$ denotes the image of ∂D_n^l under the transformation $y_j = f_j(x)$, $(j = 1, \cdots, n)$. The number δ_l denotes the degree of the mapping $y_j = f_j(x)$. The absolute value of δ_l is equal to the number of times that ∂D_n^l is mapped onto $\partial D_n'^l$. The number δ_l is positive (resp. negative) if the orientation of $\partial D_n'^l$, i.e., the sign of the Jacobian of the transformation $y_j = f_j(x)$ in a sufficiently small neighborhood of the zero x_l is positive (resp. negative). The number δ_l equals ± 1 if the functions y_j admit a Taylor series expansion around a zero with nonvanishing linear terms [9].

The integral occurring at the right side of (15) is equal to $\Omega_n = 2\pi^{(1/2)n}/\Gamma(n)$, which is the area of the unit sphere in n dimensions [8, Ch. VI, Sec. 8, eq. 53]. This leads to

$$\Omega = \sum_l \omega_l = \Omega_n \sum_l \delta_l. \quad (16)$$

Hence, the total solid angle Ω is equal to the sum of the number of zeros with positive Jacobian, minus the sum of zeros with negative Jacobian (where each zero is counted and weighted with the degree of the mapping), a result obtained by Kronecker [3]. The integral (16) is, therefore, not conclusive if we want to determine the total number of zeros of (14) in the domain D_n. The integral (12) plays an important role in the proofs of many theorems in algebraic topology, e.g., [9], [10], and [11]. Picard [2], however, invented a very simple device making the integral (12) conclusive. He added to the original set of equations (14) the equation $y_{n+1} = zJ(x) = 0$, $|z| < \epsilon$, $x \in D_n$, $z \in \mathbb{R}$, (ϵ arbitrary) and observed that the number of zeros of the set of equations

$$\begin{aligned} f_j(x) &= 0, & zJ(x) &= 0 \\ x &\in D_n & z &\in \mathbb{R} \\ |z| &< \epsilon, & j &= (1, \cdots, n) \end{aligned} \quad (17)$$

is equal to the number of zeros of the original set (14). However, the Jacobian of (17) is equal to $J^2(x)$, i.e., it is nonnegative definite by assumption for each zero x_l.

Equation (8) then shows that the solid angle Ω calculated for the set of equations $y_j = f_j(x)$, $y_{n+1} = zJ(x)$, $j = (1, \cdots, n)$ over D_{n+1} yields the exact number of zeros contained in the domain D_n. Integral (12) is called the Kronecker–Picard integral in honor of its inventors.

We will gain more insight into the properties of the KP integral by considering the zeros of a one-dimensional function. Let $y = f(x)$ denote a suitable function and let its total number of zeros in the interval $a \le x \le b$ be N. The set of functions to be considered is, therefore,

$$y_1 = f(x) \qquad y_2 = zf'(x), \tag{18}$$

and the KP integral taken over the rectangle R, $a \le x \le b$, $-\epsilon \le z \le \epsilon$, is

$$N = (2\pi)^{-1} \cdot \int_R \frac{f(x)\{f'(x)\,dz + zf''(x)\,dx\} - zf'^2(x)\,dx}{f^2 + z^2 f'^2}$$

$$= (2\pi)^{-1} \int_R d\left\{\arctan\left(\frac{zf'}{f}\right)\right\}. \tag{19}$$

The integral (19) is reminiscent of the principle of the argument occurring in the theory of analytic functions.

This integral, viz.

$$\int_C d\{\ln f(\lambda)\} = \int_C d\{\ln|f(\lambda)| + i\phi\},$$

$$f = |f|\exp\{i\phi\}, \tag{20}$$

taken along a contour C enclosing a domain D of analyticity of $f(\lambda)$ in the complex plane, is equal to the number of times that the image of C under the transformation $w = f(\lambda)$ is traversed.

The phase ϕ equals

$$\phi = \arctan\left[\frac{\mathrm{Im}\{f(\lambda)\}}{\mathrm{Re}\{f(\lambda)\}}\right], \tag{21}$$

which is to be compared with the integrand of (19).

Performing the integration over z in (19) leads to

$$N = -\pi^{-1}\epsilon \int_a^b \frac{ff'' - f'^2}{f^2 + \epsilon^2 f'^2}\,dx + \arctan\left\{\epsilon\frac{f'(b)}{f(b)}\right\}$$

$$- \arctan\left\{\epsilon\frac{f'(a)}{f(a)}\right\}, \tag{22}$$

where the multivalued arctan functions have values between $(-1/2)\pi$ and $(1/2)\pi$. In the limit for $\epsilon \to 0$ the contributions of the arctan functions can be neglected. In this case the number of zeros is only determined by the integral. The main contribution to this integral comes from a very narrow region around a zero $x = x_l$ of $f(x)$ because in this region the integrand is $O(\epsilon^{-1})$, whereas it behaves elsewhere as $O(\epsilon)$. This suggests that the integrand behaves as a delta function that is visualized by the following calculation. Let δ denote an arbitrary small positive number. From the Taylor expansion

$$f(x) = f'(x_l)(x - x_l) + O\{(x - x_l)^2\}, \tag{23}$$

we obtain

$$N = -\pi^{-1}\epsilon \sum_l \int_{-\delta}^{\delta} \frac{-f'^2(x_l)}{f'^2(x_l)x^2 + \epsilon^2 f'^2(x_l)}\,dx + O(\delta, \epsilon). \tag{24}$$

A change of variables $x = p\epsilon$ shows that the integral on the right side of (24) equals, apart from a term $O(\epsilon)$, the integral

$$\pi^{-1}\int_{-\infty}^{+\infty} \frac{dp}{1 + p^2} = 1. \tag{25}$$

Equations (22) and (24) show that the behavior of the integrand is like a delta function. Integration over any interval of arbitrary width around a zero of $f(x)$ gives a finite, nonzero contribution to the integral if ϵ tends to zero, whereas the contributions to the integral for closed intervals not containing a zero is negligible, i.e., $O(\epsilon)$.

III. The Determination of the Global Maximum

In this section a method for the determination of the global maximum of the likelihood function will be demonstrated by two examples from different fields of application. The method is based on the Kronecker–Picard integral, which yields the total number of stationary points of the function to be optimized, e.g., the likelihood function, in a domain of interest as has been shown in the previous section. With the properties of the KP integral established, the principle of the method proposed is very simple. The KP integral gives us the total number of stationary points and thus the total number of candidate points for the global optimum. We now subdivide the domain and repeat the calculation of the KP integral for each subdomain separately. A region with no stationary points is further discarded. A region wherein the KP integral reveals the presence of one root is treated as follows. The zero will be located by an iterative zero-locating algorithm, such as [12], taking the center of the pertinent region as starting point. The value of the likelihood function is then calculated at the position of the located zero and stored for further usage. Regions wherein the KP integral detects the presence of two or more zeros of the likelihood equation are subdivided again, and the process is repeated until all zeros are located and all corresponding likelihood values are obtained. Comparison of the stored likelihood function values will reveal the global maximum and also the global minimum in the initial domain of interest. The two estimation problems treated will exemplify the procedure previously outlined.

The first example is taken from statistics and is inspired by the work of Barnett [13]. Barnett evaluates the performance of several iterative procedures for the maximum likelihood estimator when the likelihood equation has multiple roots. Barnett makes various comparisons estimating the location parameter θ of the Cauchy distribution:

$$p(x) = \{\pi(1 + (x - \theta)\}^{-1}. \tag{26}$$

With a sample of n independent observations x_i of a random variable with probability density (26), the likelihood function for the location parameter θ is

$$L(x; \theta) = \prod_{i=1}^{n} \left\{\pi\left(1 + (x_i - \theta)^2\right)\right\}^{-1}. \tag{27}$$

The likelihood equation is obtained from (27) taking the derivative with respect to θ of the logarithm of (27). Estimated values $\hat{\theta}$ for θ should satisfy

$$f(\theta) = \frac{\partial}{\partial\theta} \ln L(x;\theta) = \sum_{i=1}^{n} \frac{2(x_i - \theta)}{1 + (x_i - \theta)^2} = 0. \quad (28)$$

The total number N of zeros of (28) in the interval $[a, b]$ is obtained by calculating the integral given by Picard [2] over the rectangle formed by the lines $\theta = a$, $\theta = b$ and $z = -\epsilon$, $z = +\epsilon$:

$$N = -\pi^{-1}\epsilon \int_a^b \frac{ff'' - f'}{f^2 + \epsilon^2 f'^2}\, d\theta + \pi^{-1} \arctan\left\{\frac{\epsilon f'(b)}{f(b)}\right\} - \pi^{-1} \arctan\left\{\frac{\epsilon f'(a)}{f(a)}\right\}. \quad (29)$$

The values of the arctan function should be taken between $-\pi/2$ and $\pi/2$. From (29) we see that we have to modify the interval if a zero is located at either endpoint of the interval. The method for obtaining the global maximum calculates first the total number of zeros in (a, b). Then the interval is halved and the calculation is repeated for each subinterval. An interval with no zero is discarded. If an interval contains one root, then this zero is located and the corresponding likelihood function value stored. An interval with more than one zero is halved again and so on until all zeros are located. Comparison of the likelihood function values will reveal the absolute maximum. Following [13] we simulated Cauchy distributed random variables with the simple sizes $n = 3, 5, 7, 9, 11, 13, 15, 17$, and 19 for the case $\theta = 0$. In Fig. 4 the function $f(\theta)$ of (28) is drawn for the sample size $n = 5$. In Table I the total number of zeros detected in the interval $(-24, 24)$ are displayed together with the location of the global maximum for the various sample sizes. In the computation ϵ was set equal to one. The whole simulation took about 3 CPU seconds on a CDC Cyber 170/760. The programming was done in Fortran, using routines from the NAG library.

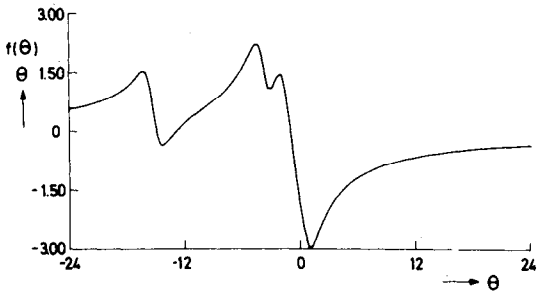

Fig. 4. Function $f(\theta)$ of (28) with sample size n equal to 5.

The second application of the method presented is taken from optical communication theory, where one has to estimate the time of arrival of a signal, e.g., pulse position modulation (PPM) or optical range finding [14], [15]. The various aspects of arrival time estimation has been the subject of many papers (see, e.g., [16], [17], [18]).

TABLE I
GLOBAL MAXIMA $\hat{\theta}$ OF (27) FOR VARIOUS SAMPLE SIZES, TRUE VALUE OF θ EQUALS ZERO

Sample Size n	Number of Zeros in (a, b) $a = -24.0$, $b = 24.0$	Location Global Maximum $\hat{\theta}$
3	5.0	2.422
5	3.0	−0.968
7	1.0	0.086
9	1.0	0.307
11	5.0	0.220
13	3.0	−0.246
15	1.0	0.297
17	1.0	1.093
19	1.0	0.074

In our example, we estimate the position a and width s of a bell-shaped signal from photodetector counts over a fixed time period. The detector readings are realizations of a random point process [15]. The successive counts $\hat{\boldsymbol{n}} = (\hat{n}_1, \cdots, \hat{n}_m)$ are independent Poisson-distributed random variables with parameter

$$\lambda_{t_i}(a, s) = \alpha\left[1 + \beta \exp\left\{-\tfrac{1}{2}s^{-2}(t_i - a)^2\right\}\right] + \lambda_0, \quad (30)$$

where α and β are constants and λ_0 denotes the dark current of the photodetector. Estimated values for a and s correspond to the global maximum of the likelihood function:

$$L(\hat{\boldsymbol{n}}; a, s) = \prod_{i=1}^{m} \exp\left\{-\lambda_{t_i}(a, s)\right\} \lambda_{t_i}(a, s)^{\hat{n}_i} / \hat{n}_i!. \quad (31)$$

The stationary points of the likelihood function (31) are given by the likelihood equations

$$\frac{\partial}{\partial a} \ln L(\hat{\boldsymbol{n}}; a, s) = \sum_{i=1}^{m} \left(-1 + \hat{n}_i \lambda_{t_i}^{-1}\right) \frac{\partial}{\partial a} \lambda_{t_i} = 0$$

$$\frac{\partial}{\partial s} \ln L(\hat{\boldsymbol{n}}; a, s) = \sum_{i=1}^{m} \left(-1 + \hat{n}_i \lambda_{t_i}^{-1}\right) \frac{\partial}{\partial s} \lambda_{t_i} = 0. \quad (32)$$

The set of equations (32) is extended with the equation $zJ(a, s) = 0$, $-\epsilon \le z \le \epsilon$ (Picard's extension), where $J(,)$ denotes the Jacobian of (32). The total number of zeros in the domain of interest is calculated applying integral (13) over D_3 over the extended set of functions. In Fig. 5 the simulated Poisson pulse train is presented. The parameter

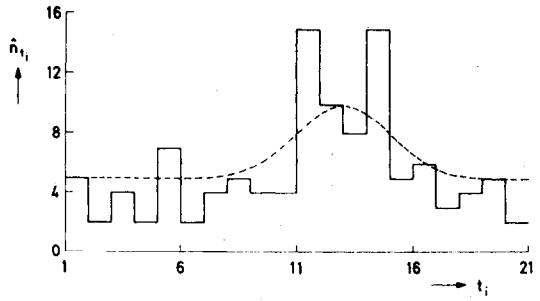

Fig. 5. Simulated Poissonian pulse train (drawn line) and λ_t of (30) (dotted line), with parameter setting $\alpha = 2.0$, $a = 13.0$, $\lambda_0 = 3.0$, $\beta = 2.5$, $s = 2.0$, $m = 21$.

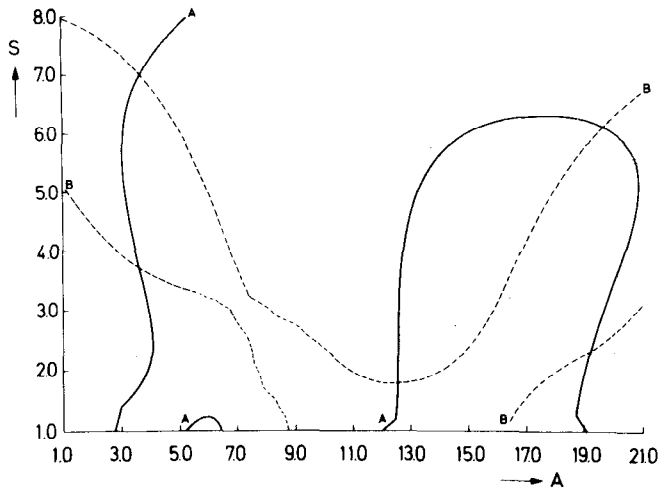

Fig. 6. Curve A, drawn line: $\partial/\partial q \ln L = 0$. Curve B, dotted line: $\partial/\partial s \ln L = 0$.

values are $a = 13.0$, $s = 2.0$, $\alpha = 2.0$, $\beta = 2.5$, $\lambda_0 = 3.0$, and $m = 21$.

Fig. 6, showing the curves $\partial/\partial a \ln L = 0$ and $\partial/\partial s \ln L = 0$, indicates the existence of five stationary points in the domain $1 \leq a \leq 21$, $1 \leq s \leq 8$. The outcome of the numerical integration of (13) for this domain results in a value of 4.82, requiring a low accuracy of 0.5 because we know that the outcome of the KP integral is an integer. The computation takes about 120 CPU with a Fortran program using integration routines from the NAG library.

We now subdivide the domain and repeat the evaluation of the integral (13) for each subdomain until only one or not root of (32) is detected. If the KP integral indicates the existence of one root, then its position can be calculated very quickly with a zero-locating algorithm [12]; we thus obtain the values $\hat{a}$ and $\hat{s}$. Calculation of $L(\hat{\boldsymbol{n}}; \hat{a}, \hat{s})$ in each zero $(\hat{a}, \hat{s})_l$ of (32) reveals the type of the extrema. The results of this scheme are summarized in Table II. The estimated values $\hat{a}$ and $\hat{s}$ corresponding to the absolute maximum are 12.58 and 1.75, respectively. The true values are 13.0 and 2.0.

IV. Discussion

The method by which the global optimum of the likelihood function is obtained has been numerically tested for two different examples. The results show the numerical feasibility, as well as the inherent elegance, of determining the number of zeros of the likelihood equations by means of the KP integral. The most important properties of this device are that it be relatively simple (i.e., the integrand of the integral (13) contains only simple algebraic functions) and, if applied to optimization problems, that it *guarantee* that the optimum is not missed by the algorithm applied, as explained in the Introduction. The method used in this paper for finding the optimum can be modified as follows. Instead of using the nesting procedure, such that the domain of interest is divided into domains only containing at most one stationary point whose position is then determined, we apply the KP integral only once. This gives us the total number of stationary points in the domain N. We then use a zero locating algorithm with different starting values and only stop the search for zeros after N have been determined.

Throughout this paper we only considered the global maximum of the likelihood function in a certain finite domain, discarding the possibility of a global maximum occurring at the boundary of the domain. This situation can also be analyzed using the KP integral. The stationary points of the boundary are determined most conveniently, using the Euler–Lagrangian multipliers. This procedure then leads to the determination of the zeros of an extra set of equations. The KP integral gives us the total number of zeros of this set, so that the optimum of the function at the boundary can be obtained by using the methods explained. This information, together with the knowledge of the optimum of the function in the domain, leads to the desired global maximum.

The explicit form (13) of the Kronecker integral shows that in case of many parameters the pertinent integrals may become prohibitively complicated for numerical evaluation. However, it is still possible to apply the basic idea of this paper using the following procedure. Suppose that we obtain a local maximum, dependent on the starting value chosen. We now would like to test whether the local maximum happens to be the global optimum. To this end, we consider the tangent plane at the alleged global maximum. This tangent plane intersects with the function $z = L(x_1, x_2, \cdots, x_n)$, to be optimized if the determined extremum is not the global maximum. To determine in a simple way whether or not the tangent plane intersects with $L(\cdot)$, we observe that this plane can be generated by a family of lines lying in this plane and parametrically given by $\{x_1(t), x_2(t), \cdots, x_n(t), z(t)\}$. The problem is then to determine whether or not the function $z(t) - L\{x_1(t), \cdots, x_n(t)\}$ has a zero, and this problem can be easily solved with the theory developed in this paper.

TABLE II
Summary of Nesting Procedure (Optical Communication Application)

Nesting Number	Domain of Interest $\leq a \leq$		$\leq s \leq$		Number of Stationary Points by KP Integral	Estimated Values $\hat{a}$	$\hat{s}$	Log Likelihood $\ln L(\boldsymbol{n}; \hat{a}, \hat{s})$	Type of Extremum
0	1.0	21.0	1.0	8.0	4.821				
1	1.0	11.0	1.0	4.5	0.994	3.65	3.70	71.25	global min
1	1.0	11.0	4.5	8.0	0.800	3.71	7.01	72.99	local
1	11.0	21.0	1.0	4.5	1.947				
2	11.0	16.0	1.0	2.75	0.870	12.58	1.75	95.28	global max
2	11.0	16.0	2.75	4.5	-0.881×10^{-3}				
2	16.0	21.0	1.0	2.75	0.816	19.23	2.31	76.91	local
1	11.0	21.0	4.5	8.0	0.945	19.63	6.10	80.87	local

The essential point is that the underlying KP integral, with which we determine the existence of a possible point of intersection of a line and $L(\cdot)$, is a one-dimensional integral, viz. (29). This makes the proposed procedure very attractive from the numerical point of view.

Acknowledgment

We are indebted to Prof. dr. H. A. Ferwerda for a critical reading of the manuscript and many valuable remarks. C. H. Slump gratefully acknowledges the financial support from the Netherlands Organization for the Advancement of Pure Research (ZWO). We thank the referees for valuable comments that improved the presentation of the paper.

Appendix

The aim of this appendix is to give a short heuristic introduction of the theory of exterior (wedge) products systematically developed by Grassmann [19]. The aim of this particular calculus is to develop a systematic procedure for the bookkeeping of the partial derivatives arising from a change of variables in multidimensional integrals or from the change of a surface integral into a volume integral, etc.

Let us consider therefore the law of change of a volume element $dx_1 \cdots dx_n$ under the point transformation

$$y_i = y_i(x_1, \cdots, x_n), \qquad i = 1, \cdots, n, \tag{A.1}$$

$$dy_1 \cdots dy_n = \frac{\partial(y_1, \cdots, y_n)}{\partial(x_1, \cdots, x_n)}\, dx_1 \cdots dx_n, \tag{A.2}$$

if $(\partial(y_1, \cdots, y_n))/(\partial(x_1, \cdots, x_n))$ denotes the Jacobian of the transformation. Grassmann [19] observed that (A.2) cannot be derived by means of ordinary multiplication of the total differentials

$$dy_i = \sum_{j=1}^{n} \frac{\partial y_i}{\partial x_j}\, dx_j, \tag{A.3}$$

and invented a product, called the wedge, or exterior, product leading to (A.2). The wedge product satisfies the following axioms:

$$\begin{aligned} dy_i \wedge (f\, dy_j) &= (f\, dy_i) \wedge dy_j = f(dy_i \wedge dy_j), \\ dy_i \wedge \left(\sum_j g_j\, dy_j\right) &= \sum_j dy_i \wedge (g_j\, dy_j), \\ dy_i \wedge dy_j &= -dy_j \wedge dy_i, \end{aligned} \tag{A.4}$$

where f and g are elements of $\mathbb{R}^n$.

From (A.4) we derive

$$\begin{aligned} dy_i \wedge dy_i &= 0 \\ dy_i \wedge \left(\sum_j g_j\, dy_j\right) &= \sum_j g_j (dy_i \wedge dy_j). \end{aligned} \tag{A.5}$$

The wedge product is therefore reminiscent of the exterior product of two vectors in $\mathbb{R}^n$. We refer to Schwartz [8] and Schreiber [20] where the consistency of these axioms is shown.

From (A.3), (A.4), and (A.5) we observe that

$$\begin{aligned} &dy_1 \wedge dy_2 \cdots \wedge dy_n \\ &= \sum_{i_1, \cdots i_k} \left\{ \epsilon(i_1, \cdots i_k) \frac{\partial y_1}{\partial x_{i_1}} \frac{\partial y_2}{\partial x_{i_2}} \cdots \frac{\partial y_n}{\partial x_{i_n}} \right\} \\ &\quad \cdot dx_1 \wedge dx_2 \cdots \wedge dx_n, \end{aligned} \tag{A.6}$$

where $\epsilon(i_1, \cdots i_k) = \pm 1$, depending on whether the number of interchanges required in permuting $\epsilon(i_1, i_2, \cdots i_k)$ to $(1, 2, \cdots k)$ is even or odd. The summation occurring on the right side of (A.6) is just the definition of the Jacobian of the transformation (A.1) so that we end up with (A.2), if the wedge products are replaced by the ordinary (scalar) product.

Similarly, we obtain the transformation of a surface integral under the change of variables (A.1). Suppose a surface integral in $\mathbb{R}^3$ reads as

$$I = \iint f_1\, dy_1\, dy_2 + \iint f_2\, dy_2\, dy_3 + \iint f_3\, dy_3\, dy_1. \tag{A.7}$$

Replacing $dy_1\, dy_2$ by $dy_1 \wedge dy_2$, $dy_2\, dy_3$ by $dy_2 \wedge dy_3$, and $dy_3\, dy_1$ by $dy_3 \wedge dy_1$, and using (A.3), (A.4), and (A.5), we end up with

$$\begin{aligned} dy_1 \wedge dy_2 = &\left(\frac{\partial y_1}{\partial x_1} \frac{\partial y_2}{\partial x_2} - \frac{\partial y_1}{\partial x_2} \frac{\partial y_2}{\partial x_1} \right) dx_1 \wedge dx_2 \\ &+ \left(\frac{\partial y_1}{\partial x_1} \frac{\partial y_2}{\partial x_3} - \frac{\partial y_1}{\partial x_3} \frac{\partial y_2}{\partial x_1} \right) dx_1 \wedge dx_3 \\ &+ \left(\frac{\partial y_1}{\partial x_2} \frac{\partial y_2}{\partial x_3} - \frac{\partial y_1}{\partial x_3} \frac{\partial y_2}{\partial x_2} \right) dx_2 \wedge dx_3. \end{aligned} \tag{A.8}$$

We obtain similar equations for the wedge products $dy_2 \wedge dy_3$ and $dy_3 \wedge dy_1$. We see that (A.8) is just a special case of (11), which can be derived similarly. Collecting all terms generated by the expansions of the wedge products, and replacing $dx_1 \wedge dx_2$ by $dx_1\, dx_2$, etc., we arrive at

$$\begin{aligned} I = &\iint \begin{vmatrix} f_3 \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ f_2 \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \\ f_1 \frac{\partial y_3}{\partial x_1} & \frac{\partial y_3}{\partial x_2} \end{vmatrix} dx_1\, dx_2 + \iint \begin{vmatrix} f_2 \frac{\partial y_1}{\partial x_2} & \frac{\partial y_1}{\partial x_3} \\ f_1 \frac{\partial y_2}{\partial x_2} & \frac{\partial y_2}{\partial x_3} \\ f_3 \frac{\partial y_3}{\partial x_2} & \frac{\partial y_3}{\partial x_3} \end{vmatrix} dx_2\, dx_3 \\ &+ \iint \begin{vmatrix} f_1 \frac{\partial y_1}{\partial x_3} & \frac{\partial y_1}{\partial x_1} \\ f_3 \frac{\partial y_2}{\partial x_3} & \frac{\partial y_2}{\partial x_1} \\ f_2 \frac{\partial y_3}{\partial x_3} & \frac{\partial y_3}{\partial x_1} \end{vmatrix} dx_3\, dx_1, \end{aligned} \tag{A.9}$$

which is to be compared with (13).

We finally given an heuristic discussion of the concept of closed forms, and to this end we introduce the concept of exterior differentiation. The exterior derivative of a form $a(y)\, dy_1 \wedge dy_2 \cdots \wedge dy_n$ is defined by

$$d(a(y)\, dy_1 \wedge dy_2 \cdots \wedge dy_n) = d(a(y)) \wedge dy_1 \cdots \wedge dy_n. \tag{A.10}$$

Two examples are

$$\begin{aligned} df &= \sum_i \frac{\partial f}{\partial y_i}\, dy_i, \\ d&\left(\sum_{i<j} a_{ij}\, dy_i \wedge dy_j \right) \\ &= \sum_{i<j} \left(\sum_l \frac{\partial a_{ij}}{\partial y_l}\, dy_l \right) \wedge dy_i \wedge dy_j \\ &= \sum_{l<i<j} \left(\frac{\partial a_{ij}}{\partial y_l} - \frac{\partial a_{lj}}{\partial y_i} + \frac{\partial a_{li}}{\partial y_j} \right) dy_l \wedge dy_i \wedge dy_j. \end{aligned} \tag{A.11}$$

The importance of this concept stems from the fact that it allows a very simple generalization of the laws of Gauss and Stokes into higher dimensional space. If

$$\omega = \sum_{i=1}^{n} a_i \, dy_2 \wedge \cdots \wedge dy_{i-1} \wedge dy_{i+1} \cdots \wedge dy_n, \quad \text{(A.12)}$$

and if ∂M denotes the boundary of a domain M in $\mathbb{R}^n$, we have

$$\int_{\partial M} \omega = \int_M d\omega. \quad \text{(A.13)}$$

As an example, we derive Stokes' theorem in two dimensions. If M is a planar region bounded by a closed curve ∂M, we have

$$\int_{\partial M} a_1 \, dy_1 + a_2 \, dy_2 = \iint_M \left(\frac{\partial a_1}{\partial y_2} - \frac{\partial a_2}{\partial y_1} \right) dy_1 \, dy_2. \quad \text{(A.14)}$$

The proof of (A.13) is obtained using (A.10) and integration by parts.

We call a form ω closed in a certain domain if its exterior derivative equals zero, i.e.,

$$d\omega = 0 \quad \text{(A.15)}$$

in that domain. Equation (A.13) indicates that a form is closed if and only if it is possible to change the boundary of an integral without changing its value. This heuristic discussion of the closure concept is to be compared with the paragraph preceding (11). The closure of the form (9) is readily deduced from

$$d\left(\sum_{j=1}^{n} (-1)^{j-1} y_j r^{-n} \, dy_1 \wedge dy_2 \cdots dy_{j-1} \wedge dy_{j+1} \cdots dy_n \right)$$

$$= \sum_{j=1}^{n} \frac{\partial}{\partial y_j} \left(y_j r^{-n} \right) dy_1 \wedge dy_2 \cdots \wedge dy_n = 0. \quad \text{(A.16)}$$

References

[1] N. K. Gupta and R. K. Mehra, "Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 774–783, 1974.

[2] E. Picard, "Sur le nombre des racines communes à plusieurs équations simultanées," *Journ. de Math. Pure et Appl.*, vol. 8, pp. 5–24, 1892.

[3] L. Kronecker, *Werke*, vol. 1. Leipzig: Teubner, 1895, pp. 175–212, 213–226.

[4] B. J. Hoenders and C. H. Slump, "On the calculation of the exact number of zeros of a set of equations," *Computing*, vol. 30, pp. 137–147, 1983.

[5] A. Davidoglou, "Sur le nombre de racines communes à plusieurs équations," *Compt. Rend. Hebd.*, vol. 133, pp. 784–786, 860–863, 1901.

[6] G. Tzitzéica, "Sur le nombre des racines communes à plusieurs équations," *Compt. Rend. Hebd.*, vol. 133, pp. 918–920, 1901.

[7] W. Kaplan, *Advanced Calculus*. Reading, MA: Addison-Wesley, 1952.

[8] L. Schwartz, *Cours d'Analyse*, vol. 2. Paris: Hermann, 1967.

[9] E. Picard, *Traité d'Analyse*, vol. 1. Paris: Gauthier-Villars, 1922, pp. 150–153.

[10] V. Hutson and J. S. Pym, *Applications of Functional Analysis and Operator Theory*. London: Academic, 1980, Ch. 14.

[11] D. H. Sattinger, *Topics in Stability and Bifurcation Theory (Lecture Notes in Mathematics)*, vol. 309. Berlin: Springer, 1973.

[12] M. J. D. Powell, "A hybrid method for nonlinear algebraic equations," in *Numerical Methods for Nonlinear Algebraic Equations*, P. Rabinowitz, Ed. London: Gordon and Breach, 1970.

[13] V. D. Barnett, "Evaluation of the maximum-likelihood estimator where the likelihood equation has multiple roots," *Biometrika*, vol. 53, pp. 151–165, 1966.

[14] R. M. Gagliardi and S. Karp, *Optical Communications*. New York: Wiley, 1976.

[15] D. L. Snyder, *Random Point Processes*. New York: Wiley, 1975.

[16] I. Bar-David, "Minimum-mean-square-error estimation of photon pulse delay," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 326–330, 1975.

[17] G. Lee and G. Schroeder, "Optical pulse timing resolution," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 114–118, 1976.

[18] I. Bar-David and M. Levy, "Advent of nonregularity in photon-pulse delay estimation," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 122–124, 1978.

[19] H. Grassmann, *Lineare Ausdehnunglehre*. Leipzig: Teubner, 1844; reprinted in H. Grassmann, *Gesammelte Werke*. Leipzig: Teubner, 1894.

[20] M. Schreiber, *Differential Forms, A Heuristic Introduction*. Berlin: Springer, 1977.