PAIRWISE SAMPLING FOR THE NONLINEAR INTERPOLATION OF FUNCTIONS OF VERY MANY VARIABLES (1)

T. TSUDA (2)

ABSTRACT: The difficulty is first pointed out concerning the nonlinear interpolation of functions defined in a space of very many dimensions. There is a method using a sampling technique that works well for \( k \approx 10-20 \) \((k = \text{number of dimensions})\). The sampling errors, however, increase in powers of \( k \), so that for \( k \) greater than the above-quoted values the computation is no more feasible. This is due to the subtraction between two large sums of about the same magnitude, each of which suffers stochastic fluctuations accompanying samplings. To avoid this unfavorable effect, a pairwise sampling technique is considered where one draws two samples at a time when required, one from each of the two sums of terms. By this new avenue of approach, the probabilistic interpretation becomes much more straightforward than hitherto conceived and the reduction of standard errors is also remarkable especially for the cases of very many dimensions.

1. Statement of the Problem.

In a previous paper [1] we established an algorithm of nonlinear interpolation of multivariable functions. The main point of the argument was that when the number of independent variables, \( k \), is greater than 10 (say), the number of terms in the multidimensional Lagrangian interpolation formula turns out to be so enormous, that termwise computation is no more feasible. One way round this difficulty is to identify the assembly of these numerous terms with a parent population, from which some samples are drawn to estimate the parameter of the parent population. If one formulate the problem in such a way that this parameter is none other than the solu-

--- Received il 2-7-73.
(1) This work was done while the author was a Visiting Professor to the Istituto Matematico « Federigo Enriques », Università di Milano, during the summer of 1973.
(2) Department of Electrical Engineering, Hokkaido University, Sapporo, Japan.
tion, then a small number of samples has indeed yielded the solution with a precision of 3 or 4 decimals (see also [2] and [3]).

As an introduction, we now illustrate the difficulty in the nonlinear interpolation where the number of dimensions is as large as, or even larger than, 30. Let \( n \) be the degree of the interpolation polynomial and \( x_i^{(0)}, x_i^{(1)}, \ldots, x_i^{(n)} \) the lattice points on the \( i \)th axis \((i = 1, 2, \ldots, k)\). Using the one-dimensional Lagrange coefficients,

\[
(1) \quad l(s; r) = \prod_{j=0 \atop j \neq s}^{n} \frac{x_r - x_i^{(j)}}{x_r - x_i^{(s)}},
\]

one has

\[
(2) \quad f(x) = \sum_{i} L(i; x) f(i),
\]

where

\[
(3) \quad L(i; x) = \prod_{r=1}^{k} l(i_r; r),
\]

\[
(4) \quad i = (i_1, i_2, \ldots, i_k),
\]

\[
(5) \quad i_r = 0, 1, \ldots, n \quad (r = 1, 2, \ldots, k);
\]

namely, \( L(i; x) \) is the \( k \)-dimensional Lagrangian coefficient and \( f(i) \) stands for the data at a data point \( x = (x_1^{(i_1)}, \ldots, x_k^{(i_k)}) \). By the reason stated in [1] (about half of the coefficients \( L(i; x) \) take on negative values when \( n > 1 \)), one must regroup the terms on the right-hand side of (2), so that there result

\[
(6) \quad f(x) = \sum_{L \geq 0} L'(i, x) f(i) - \sum_{L < 0} L''(i, x) f(i)
\]

where

\[
L'(i; x) = L(i; x) \quad \text{for} \quad L(i; x) \geq 0,
\]

\[
L''(i; x) = -L(i; x) \quad \text{for} \quad L(i; x) < 0.
\]

We then construct a probabilistic model to evaluate the sum of finite series (6). The result is given by

\[
(8) \quad \frac{L'}{N'} \Sigma' f(i) - \frac{L''}{N''} \Sigma'' f(i) \rightarrow \mathcal{L}' E[f; L \geq 0] - \mathcal{L}'' E[f; L < 0] \approx f(x).
\]

Here \( \Sigma' f(i) / N' \) (or equivalently \( E[f; L \geq 0] \)) should give the magnitude of the first sum of series [6] for a sufficiently large \( N' \), the number of samples drawn from the first group of terms in (6) with probability \( \propto |L(i;x)| \).