NOSC

ADA087566

Technical Report 552

RECURSIVE LEAST SQUARES LATTICE ALGORITHMS — A GEOMETRICAL APPROACH

M. J. Shensa

15 December 1979

Final Report: July — August 1979

DTIC
ELECTE
AUG 6 1980

A

NAVAL OCEAN SYSTEMS CENTER
SAN DIEGO, CALIFORNIA 92152

80 8 3 038

NOSC

NAVAL OCEAN SYSTEMS CENTER, SAN DIEGO, CA 92152

AN ACTIVITY OF THE NAVAL MATERIAL COMMAND

SL GUILLE, CAPT, USN
Commander

HL BLOOD
Technical Director

ADMINISTRATIVE INFORMATION

ACKNOWLEDGMENT

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NOSC Technical Report 552	2. GOVT ACCESSION NO. AD-A087 566	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) RECURSIVE LEAST SQUARES LATTICE ALGORITHMS – A GEOMETRICAL APPROACH		5. TYPE OF REPORT & PERIOD COVERED Final Report, July - Aug 79
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) M. J. Shensa		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Ocean Systems Center San Diego, CA 92152		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Ocean Systems Center San Diego, CA 92152		12. REPORT DATE 15 December 1979
		13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

time recursives
least squares algorithms
lattice algorithms

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Several time-recursive least squares algorithms have been developed in recent years. In this report a geometrical formalism is defined which utilizes a nested family of metric spaces indexed by the data time interval. This approach leads to a simplified derivation of the so-called recursive least squares lattice algorithms (recursive in time and order). In particular, it is found that the resulting structure provides a single framework which encompasses an entire family of fairly complex algorithms as well as providing geometrical insight into their behavior.

DD FORM 1473 1 JAN 73   EDITION OF 1 NOV 68 IS OBSOLETE
S/N 0102-LF-014-6601

CONTENTS

i

INTRODUCTION

In 1973 Markel and Gray [1] presented a geometrical formulation of the lattice approach to regression analysis (also used by Itakura and Saito [2]). Their goal was the solution of a stochastic problem (least mean squares or so-called autocorrelation or normal equations) and consequently, the relations derived were functions of expected values of variables. In particular, the results rely on the true autocorrelation matrix. More recently, Morf et al. ([3]–[8]) have developed time-recursive algorithms for the solution of the least squares problem; i.e., the minimization of a finite time average of errors rather than the expectation of the error. The least squares problem also involves a set of normal equations, but in this case the relevant matrix is a finite-data approximation to the autocorrelation matrix.

In this paper we utilize an inner-product formalism similar to that of [1] to solve the least squares problem. The structure so obtained is of considerable interest inasmuch as it forms the basis for a whole class of recursive least squares lattice algorithms ([3]–[9]). This apporach also enjoys several other advantages. First, it provides a simplified derivation of the above mentioned algorithms. (It is especially noteworthy that the inversion lemma for a partitioned matrix is not needed.) Secondly, the geometrical nature of the variables defined makes their role in the algorithms readily apparent and lends some physical intuition as well. Finally, many of the relations proved in [1] for expected values are shown to hold true for the actual variables. These relations, such as the expression of the lattice gain coefficients in terms of inner products, can provide important insights into the algorithms' behavior; i.e., regarding numerical stability, convergence (effect of $T \to \infty$), implementation, etc. ([1],[9]).

DEFINITIONS

We begin with some vector notation. Let $x(t)$ be a discrete time series, and define $\bar{x}$ to be the vector whose t^{th} component is given by

$$[\bar{x}]_t = x(t) \quad t = 0, 1, \ldots . \tag{1}$$

Although this vector lies in an infinite dimensional space ($\bar{x} \in R^\infty$), all our operations shall be on finite dimensional subspaces thus alleviating the need for sophisticated mathematical techniques. However, if so desired, the reader may assume an upper bound T_0 such that $0 \leq t \leq T_0$; i.e., $\bar{x} \in R^{T_0}$. By shifting $\bar{x}$, we obtain a family of vectors $\bar{x}^i$

$$[\bar{x}^i]_t = \begin{cases} x(t-i) & t \geq i \\ 0 & 0 \leq t < i \end{cases} . \tag{2}$$

This family generates a nested set of subspaces S^p defined by

$$S^p = \left\{ \bar{x}^0, \bar{x}^1, \ldots, \bar{x}^p \right\} \quad p = 0, \ldots, p_0 \tag{3}$$

1

where the curly brackets indicate the linear space spanned by the vectors it encloses. It is assumed that the $\bar{x}^i$'s are linearly independent for $0 \leqslant i \leqslant p_0$; thus, S^p is p + 1 dimensional for $p \leqslant p_0$. Note also that $S^p \subset S^r$ for $p \leqslant r$.

We next introduce a family of pseudo-metrics on R^∞ defined as follows:

$$\langle \bar{x}, \bar{y} \rangle_T = \sum_{t=0}^{T} w^{T-t} [\bar{x}]_t [\bar{y}]_t \tag{4}$$

$$|| \bar{x} ||^2_T = \langle \bar{x}, \bar{x} \rangle_T$$

where $0 < w \leqslant 1$ is constant factor which "exponentially windows" the data ([8]–[9]). Although $< >_T$ is singular on R^∞, it is a true metric on S^p for sufficiently large T. More precisely, a necessary and sufficient condition for $< >_T$ to be a metric on S^p for all $p \leqslant p_0$ is that the truncated vectors $\bar{x}^{i'}$ defined by

$$[\bar{x}^{i'}]_t = \begin{cases} [\bar{x}^i]_t & t \leqslant T \\ 0 & t > T \end{cases} \tag{5}$$

be linearly independent for $0 \leqslant i \leqslant p_0$ (see Appendix). We shall always assume this to be the case. Note that a necessary condition is $T > p_0$.

It is also convenient to define the linear shift operator z^i by

$$[z^i \bar{x}]_t = \begin{cases} [x]_{t+i} & t+i \geqslant 0 \\ 0 & t+i < 0 \end{cases} \tag{6}$$

where i may be any integer. It follows that

$$z^{-i} \bar{x} = \bar{x}^i . \tag{7}$$

Note that $zz^{-1} = I$ but $z^{-1}z \neq I$ where I is the identity. A short calculation utilizing definitions (6) and (4) yields the relation

$$\langle z^{-1} \bar{x}, \bar{y} \rangle_T = \langle \bar{x}, z\bar{y} \rangle_{T-1} . \tag{8}$$

Equation (8) will prove extremely important inasmuch as it relates time shift properties to order (i.e., to the index i of the family $\bar{x}^i$). It differs from the stochastic case of Markel and Gray [1] inasmuch as z and z^{-1} are not quite adjoint operators (T is replaced by T-1

2

in (8)). However, the shift properties are sufficiently powerful to allow $< >$ to replace expectation in the formalism.

The least squares problem whose solution we are ultimately seeking is the following: Given sequences $x(t)$ and $y(t)$, find g_i^p such that

$$|| \bar{y} - \sum_{i=0}^{p} g_i^p \bar{x}^i ||_T^2 = \min_q || \bar{y} - \sum_{i=0}^{p} q_i \bar{x}^i ||_T^2 . \qquad (9)$$

Actually this is a family of problems where the order p may range over $0 \leqslant p \leqslant p_0$ and the time T over $p_0 \leqslant T \leqslant T_0$. Note that the weights w appearing in the definition of the metric, (4), may be thought of as simply error weights, or as a sliding exponential window $(x(t) \rightarrow (\sqrt{w})^{T-t} x(t))$ of the data [9].

LATTICE ORDER RECURSION RELATIONS

We shall repeatedly have need of the phrase "orthogonal projection with respect to the metric $< >_T$," and hence shall use the abbreviation "proj_T." Let $\sum_{i=0}^{p-1} b_i^p(T) \bar{x}^i$ be the proj_T of $\bar{x}^p$ onto S^{p-1} (where $S^{-1} = $ empty set). Then we define $\bar{\beta}^p(T)$ by

$$\bar{\beta}^p(T) = \bar{x}^p - \sum_{i=0}^{p-1} b_i^p(T) \bar{x}^i . \qquad (10)$$

These vectors have several important properties.

We first note that $\bar{\beta}^p(T)$ is the proj_T of $\bar{x}^p$ onto the orthogonal complement of S^{p-1} in S^p; i.e., $\langle \bar{\beta}^p(T), \bar{x}^p \rangle_T = \langle (S^{p-1})^\perp, \bar{x}^p \rangle_T$. Also $S^p = S^{p-1} \oplus \bar{\beta}^p$ so that $\bar{\beta}^i$ for $0 \leqslant i \leqslant p$ form an orthogonal basis of S^p. In physical terms we may consider $\bar{\beta}^p(T)$ as the error residual of a least squares backward predictor (estimator of $\bar{x}^p$ by $\bar{x}^i$'s for $i < p$) since

$$|| \bar{x}^p - \sum_{i=0}^{p-1} b_i^p(T) \bar{x}^i ||^2 = \min_q || \bar{x}^p - \sum_{i=0}^{p-1} q_i \bar{x}^i ||^2 . \qquad (11)$$

For future convenience, we define the magnitude-squared of $\bar{\beta}^p(T)$ by

$$\epsilon_\beta^p(T) = || \bar{\beta}^p(T) ||_T^2 . \qquad (12)$$

It follows from the definition of $\bar{\beta}^p$ (T) that it is uniquely determined by the three conditions

$$\bar{\beta}^p (T) \ \epsilon \ S^p \tag{13a}$$

$$\bar{\beta}^p (T) - \bar{x}^p \ \epsilon \ S^{p-1} \tag{13b}$$

$$\langle \bar{\beta}^p(T), \bar{x}^i \rangle_T = 0 \text{ for } 0 \leqslant i \leqslant p-1 \ . \tag{13c}$$

Equations (13) merely state that the set of $\bar{\beta}^i$ (T) is obtained by applying the Gram-Schmidt orthogonalization procedure to $\bar{x}^i$, $0 \leqslant i \leqslant p$ with respect to the metric $\langle \ \rangle_T$.

As in [1] we introduce a set of auxiliary vectors $\bar{\alpha}^p$ (T) to aid in the order recursion. Let Q^p be the subspace

$$Q^p = \left\{ \bar{x}^i, \ldots, \bar{x}^p \right\}, \tag{14}$$

and let $\displaystyle\sum_{i=1}^{p} a_i^p$ (T) $\bar{x}^i$ be the proj$_T$ of $\bar{x}^0$ onto Q^p. Then we define $\bar{\alpha}^p$(T) by

$$\bar{\alpha}^p (T) = \bar{x}^0 - \sum_{i=1}^{p} a_i^p \ \bar{x}^i \tag{15}$$

with magnitude squared

$$\epsilon_\alpha^p (T) = || \bar{\alpha}^p (T) ||_T^2 \ . \tag{16}$$

These $\bar{\alpha}^p$(T) represent forward predictor residuals and are uniquely determined by the conditions

$$\bar{\alpha}^p(T) \ \epsilon \ S^p \tag{17a}$$

$$\bar{\alpha}^p(T) - \bar{x}^0 \ \epsilon \ Q^p \bullet \tag{17b}$$

$$\langle \bar{\alpha}^p(T), \bar{x}^i \rangle_T = 0 \text{ for } 1 \leqslant i \leqslant p \ . \tag{17c}$$

We now show that $\bar{\alpha}^p$(T) and $\bar{\beta}^p$(T-1) may be computed recursively via the relations

$$\bar{\alpha}^{p+1}(T) = \bar{\alpha}^p(T) + K_\alpha^p(T) \ z^{-1} \ \bar{\beta}^p (T-1) \tag{18}$$

4

$$\bar{\beta}^{p+1}(T) = z^{-1}\,\bar{\beta}^p(T-1) + K_\beta^p(T)\,\bar{\alpha}^p(T) \tag{19}$$

where $K_\alpha^p(T)$ and $K_\beta^p(T)$ are constants which we shall subsequently determine (see (26) and (29)).

To prove (18) we note that $z^{-1}\bar{\beta}^p(T-1) \in Q^{p+1} \subset S^{p+1}$, and from (8)

$$\langle z^{-1}\bar{\beta}^p(T-1), \bar{x}^i \rangle_T = \langle \bar{\beta}^p(T-1), z\bar{x}^i \rangle_{T-1} \quad \text{for } i > 0$$

$$= \langle \bar{\beta}^p(T-1), \bar{x}^{i-1} \rangle_{T-1}$$

$$= 0 \quad \text{for } 1 \leqslant i \leqslant p . \tag{20}$$

Thus for any constant K, $\bar{\phi} \triangleq \bar{\alpha}^p(T) + K\,z^{-1}\bar{\beta}^p(T-1)$ satisfies (17a) and (17b) with p replaced by p + 1, and also satisfies

$$\langle \bar{\phi}, \bar{x}^i \rangle_T = 0 \quad \text{for } 1 \leqslant i \leqslant p . \tag{21}$$

In consideration of (21), condition (17c) will be satisfied for p + 1 provided K is chosen such that

$$0 = \langle \bar{\phi}, \bar{x}^{p+1} \rangle_T = \langle \bar{\alpha}^p(T) + Kz^{-1}\bar{\beta}^p(T-1), \bar{x}^{p+1} \rangle_T \quad ;$$

i.e., if we choose $K = K_\alpha^p(T)$ where

$$K_\alpha^p(T) = -\frac{\langle \bar{\alpha}^p(T), \bar{x}^{p+1}(T) \rangle_T}{\langle z^{-1}\bar{\beta}^p(T-1), \bar{x}^{p+1}(T) \rangle_T} . \tag{22}$$

A similar analysis verifies (19) where $K_\beta^p(T)$ is chosen to satisfy

$$K_\beta^p(T) = -\frac{\langle z^{-1}\bar{\beta}^p(T-1), \bar{x}^0 \rangle_T}{\langle \bar{\alpha}^p(T), \bar{x}^0 \rangle_T} . \tag{23}$$

The appearance of T-1 in the above equations should be noted. The z^{-1} was necessary in order to obtain a vector whose "$\bar{x}^{p+1}$ coefficient" was non-zero (see eq. (22)) which in turn necessitated T-1 in order that eq. (20) be valid (c.f. remarks after eq. (8)).

5

SIMPLIFICATION OF K_α^p AND K_β^p

Equations (22) and (23) may be rewritten in a form which provides a simple geometric interpretation. We first note that since $\bar{\beta}^p(T-1)$ is a proj_{T-1} of $\bar{x}^p$, we have $\langle \bar{\beta}^p(T-1), \bar{x}^p \rangle_{T-1} = \langle \bar{\beta}^p(T-1), \bar{\beta}^p(T-1) \rangle_{T-1} = \epsilon_\beta^p(T-1)$. Thus

$$\langle z^{-1}\bar{\beta}^p(T-1), \bar{x}^{p+1} \rangle_T = \langle \bar{\beta}^p(T-1), \bar{x}^p \rangle_{T-1} = \epsilon_\beta^p(T-1) . \tag{24}$$

Similarly

$$\langle \bar{\alpha}^p(T), \bar{x}^0 \rangle_T = \epsilon_\alpha^p(T) . \tag{25}$$

We now show that the numerators of (22) and (23) are both given by

$$k^p(T) \triangleq \langle \bar{\alpha}^p(T), z^{-1}\bar{\beta}^p(T-1) \rangle_T . \tag{26}$$

We have, using (20),

$$\langle \bar{\alpha}^p(T), z^{-1}\bar{\beta}^p(T-1) \rangle_T = \langle \bar{x}^0 - \sum_{i=1}^{p} a_i^p(T)\bar{x}^i, z^{-1}\bar{\beta}^p(T-1) \rangle_T$$

$$= \langle \bar{x}^0, z^{-1}\bar{\beta}^p(T-1) \rangle_T , \tag{27}$$

and from (17c)

$$\langle \bar{\alpha}^p(T), z^{-1}\bar{\beta}^p(T-1) \rangle_T = \langle \bar{\alpha}^p(T), z^{-1}(\bar{x}^p - \sum_{i=0}^{p-1} b_i^p(T-1)\bar{x}^i) \rangle_T$$

$$= \langle \bar{\alpha}^p(T), z^{-1}\bar{x}^p \rangle_T$$

$$= \langle \bar{\alpha}^p(T), \bar{x}^{p+1} \rangle_T . \tag{28}$$

It thus follows from (22), (23) and (26)–(28) that

$$K_\alpha^p(T) = -\frac{k^p(T)}{\epsilon_\beta^p(T-1)}$$

$$\tag{29}$$

$$K_\beta^p(T) = -\frac{k^p(T)}{\epsilon_\alpha^p(T)} .$$

For ease of notation let $\bar{\alpha} = \bar{\alpha}^p(T)$ and $\bar{\beta} = z^{-1} \bar{\beta}^p(T-1)$. Then

$$K_\alpha = \frac{-\langle \bar{\alpha}, \bar{\beta} \rangle_T}{|| \bar{\beta} ||_T^2} \quad \text{and} \quad K_\beta = \frac{-\langle \bar{\alpha}, \bar{\beta} \rangle_T}{|| \bar{\alpha} ||_T^2} \ . \tag{30}$$

It follows from the Cauchy-Schwartz inequality that their product satisfies

$$0 < K_\alpha K_\beta < 1 \ . \tag{31}$$

This condition is weaker than that of [1] where the gains are individually less than one. However, relations (30) are exact for arbitrary input x(t) whereas [1] assumes a knowledge of the true correlation matrix, a condition rarely met in practice. For w = 1 and under reasonable conditions on x, $\langle \bar{x}^i, \bar{x}^j \rangle_T \triangleq R_{ij}$ approaches the true correlation matrix as $T \to \infty$ so that $\lim_{T \to \infty} | K_\alpha(T) | < 1$ and $\lim_{T \to \infty} | K_\beta(T) | < 1$. This is not true for w < 1, and either $| K_\alpha(T) |$ or $| K_\beta(T) |$ (but not both) may be greater than one. This may have serious stability consequences ([1] – [2], [9]).

TIME UPDATES

We introduce a set of auxiliary vectors $\bar{\xi}^p$ to aid in the time update defined as follows: Let $\bar{e}^T$ be the vector representing the Tth coordinate axis; i.e.,

$$[\bar{e}^T]_t = \begin{array}{ll} 1 & \{t = T \\ 0 & \{t \neq T \end{array} \ . \tag{32}$$

Then $\bar{\xi}^p(T)$ is defined as the proj_T of $\bar{e}^T$ onto S^p

$$\bar{\xi}^p(T) \ \epsilon \ S^p$$

$$\langle \bar{\xi}^p(T), \bar{y} \rangle_T = \langle \bar{e}^T, \bar{y} \rangle_T \tag{33a}$$

$$= [\bar{y}]_T \ \text{for} \ \bar{y} \ \epsilon \ S^p \ . \tag{33b}$$

(We could characterize $\bar{\xi}^p(T) = \sum_{i=0}^{p} c_i^p(T) \bar{x}^i$ as in equations (10) – (11), but this is not needed in the development.) Thus, $\bar{\xi}^p(T)$ picks out the Tth coordinate of all vectors $\bar{y}$ in S^p. It is worth observing that $\bar{\xi}^p$ is a measure of the influence of the most recent data point (coordinate axis $\bar{e}^T$) on the pth order model (subspace S^p).

7

We define the squared magnitude of $\bar{\xi}^p(T)$ to be $\sigma^p(T)$

$$\sigma^p(T) = \|\bar{\xi}^p(T)\|_T^2 . \tag{34}$$

An important property of $\sigma^p(T)$ is that

$$0 \leqslant \sigma^p(T) \leqslant \sigma^q(T) \leqslant 1 \quad \text{for } p \leqslant q . \tag{35}$$

Relation (35) follows from definition (34) and the nesting of the S^p's: $S^p \subset S^q$ so that the magnitude of the projections $\bar{\xi}^p$ must be an increasing function of p. Also
$[\bar{\xi}^p(T)]_T = \sigma^p(T) = \|\bar{\xi}^p(T)\|_T^2 \leqslant [\bar{\xi}^p(T)]_T^2$ which implies $[\bar{\xi}^p(T)]_T \leqslant 1$.

Since $S^{p+1} = S^p \oplus \bar{\beta}^{p+1}(T)$ and the set of vectors $\bar{\beta}^i(T) / \sqrt{\epsilon_\beta^i(T)}$ for

$0 \leqslant i \leqslant p+1$ form an orthonormal$_T$ basis of S^{p+1}, the proj$_T$ of $\bar{e}^T$ onto S^{p+1} is equal to its projection onto S^p plus the vector

$$\langle \bar{e}^T, \bar{\beta}^{p+1}(T) / \sqrt{\epsilon_\beta^{p+1}(T)} \rangle_T \quad \bar{\beta}^{p+1}(T) / \sqrt{\epsilon_\beta^{p+1}(T)} \quad . \text{ Thus,}$$

$$\bar{\xi}^{p+1}(T) = \bar{\xi}^p(T) + \frac{\langle \bar{e}^T, \bar{\beta}^{p+1}(T) \rangle_T}{\epsilon_\beta^{p+1}(T)} \bar{\beta}^{p+1}(T)$$

$$= \bar{\xi}^p(T) + \frac{[\bar{\beta}^{p+1}(T)]_T}{\epsilon_\beta^{p+1}(T)} \bar{\beta}^{p+1}(T) . \tag{36}$$

This gives an order recursion for $\bar{\xi}^p(T)$ and $\sigma^p(T)$.

We now derive the time update recursion for $\bar{\beta}^p$. A useful relationship which follows immediately from definition (4) is

$$\langle \bar{x}, \bar{y} \rangle_T = w \langle \bar{x}, \bar{y} \rangle_{T-1} + [\bar{x}]_T [\bar{y}]_T . \tag{37}$$

8

Thus,

$$\langle \bar{\beta}^p (T-1), \bar{x}^i \rangle_T = w \langle \bar{\beta}^p (T-1, \bar{x}^i \rangle_{T-1} + [\bar{\beta}^p (T-1)]_T [\bar{x}^i]_T$$

$$= [\bar{\beta}^p (T-1)]_T [\bar{x}^i]_T \quad 0 \leq i \leq p-1$$

$$= [\bar{\beta}^p (T-1)]_T \langle \bar{\xi}^{p-1} (T) \bar{x}^i \rangle_T . \tag{38}$$

Consider $\bar{\phi} \triangleq \bar{\beta}^p (T-1) - [\bar{\beta}^p (T-1)]_T \bar{\xi}^{p-1} (T)$. Then, from (38), $\langle \bar{\phi}, \bar{x}^i \rangle_T = 0$ for $0 \leq i \leq p-1$. Also $\bar{\phi} - \bar{x}^p \in S^{p-1}$ and it follows that $\bar{\phi}$ satisfies (13a) – (13c); i.e., $\bar{\phi} = \bar{\beta}^p (T)$ and

$$\bar{\beta}^p (T) = \bar{\beta}^p (T-1) - [\bar{\beta}^p (T-1)]_T \bar{\xi}^{p-1} (T) . \tag{39}$$

A similar calculation (one must use (8) because of z^{-1}) yields the time update for $\bar{\alpha}^p$

$$\bar{\alpha}^p (T) = \bar{\alpha}^p (T-1) - [\bar{\alpha}^p (T-1)]_T z^{-1} \bar{\xi}^{p-1} (T-1) . \tag{40}$$

The time update for k^p easily follows from equations (26), (28) and (40) by utilizing the identities (8) and (37).

$$- k^p (T) = - \langle \bar{\alpha}^p (T), \bar{x}^{p+1} \rangle_T$$

$$= - \langle \bar{\alpha}^p (T-1), \bar{x}^{p+1} \rangle_T + [\bar{\alpha}^p (T-1)]_T \langle z^{-1} \bar{\xi}^{p-1} (T-1), \bar{x}^{p+1} \rangle_T$$

$$= - wk^p (T-1) - [\bar{\alpha}^p (T-1)]_T [\bar{x}^{p+1}]_T$$

$$+ [\bar{\alpha}^p (T-1)]_T \langle \bar{\xi}^{p-1} (T-1), \bar{x}^p \rangle_{T-1} . \tag{41}$$

It follows from (36) (or simple geometric considerations) that

$$\langle \bar{\xi}^{p-1} (T-1), \bar{x}^p \rangle_{T-1} = \langle \bar{\xi}^p (T-1), \bar{x}^p \rangle_{T-1} - \frac{[\bar{\beta}^p (T-1)]_{T-1}}{\epsilon_{\beta}^p (T-1)} \langle \bar{\beta}^p (T-1), \bar{x}^p \rangle_{T-1}$$

$$= [\bar{x}^p]_{T-1} - [\bar{\beta}^p (T-1)]_{T-1} .$$

This combined with (41) yields

$$k^p(T) = wk^p(T-1) + [\bar{\alpha}^p(T-1)]_T \, [\bar{\beta}^p(T-1)]_{T-1} . \qquad (42)$$

Two useful relations which may be used in deriving various forms of the least squares algorithm ([9]) are obtained by taking the scalar product$_T$ of (39) and (40) with $\bar{\xi}^p(T)$

$$[\bar{\beta}^p(T)]_T = [\bar{\beta}^p(T-1)]_T \, (1 - \sigma^{p-1}(T)) \qquad (43a)$$

$$[\bar{\alpha}^p(T)]_T = [\bar{\alpha}^p(T-1)]_T \, (1 - \sigma^{p-1}(T-1)) \qquad (43b)$$

where we have used (36) in deriving (43a).

The time updates for $\bar{\xi}^p$, ϵ_α^p, and ϵ_β^p rely on the following lemmas:

$$\langle z^{-1} \, \bar{\xi}^{p-1}(T-1), \bar{x}^0 \rangle_T = \langle z^{-1} \, \bar{\xi}^{p-1}(T-1), \bar{x}^0 - \bar{\alpha}^p(T) \rangle_T$$

$$= \langle \bar{\xi}^{p-1}(T-1), z \, (\bar{x}^0 - \bar{\alpha}^p(T)) \rangle_{T-1}$$

$$= [\bar{x}^0 - \bar{\alpha}^p(T)]_T \quad \text{(from (36))}$$

$$= [\bar{x}^0]_T - [\bar{\alpha}^p(T)]_T \qquad (44a)$$

and similarly

$$\langle \bar{\xi}^{p-1}(T), \bar{x}^p \rangle_T = [\bar{x}^p]_T - [\bar{\beta}^p(T)]_T . \qquad (44b)$$

We can now calculate $\bar{\xi}^p(T)$ from $\bar{\xi}^p(T-1)$. Let $\bar{\phi} = z^{-1} \, \bar{\xi}^{p-1}(T-1) + K \, \bar{\alpha}^p(T)$. Then $\langle \bar{\phi}, \bar{x}^i \rangle_T = [x^i]_T$ for $p \geq i > 0$. For $i = 0$ we determine K by

$$[x^0]_T = \langle \bar{\phi}, \bar{x}^0 \rangle_T$$

$$= \langle z^{-1} \, \bar{\xi}^{p-1}(T-1), \bar{x}^0 \rangle_T + K \langle \alpha^p(T), \bar{x}^0 \rangle_T .$$

Using (44a) we have

$$K = \frac{[\bar{\alpha}^p(T)]_T}{\epsilon_\alpha^p(T)} .$$

Thus,

$$\bar{\xi}^p(T) = z^{-1} \bar{\xi}^{p-1}(T-1) + \frac{[\bar{\alpha}^p(T)]_T}{\epsilon_\alpha^p(T)} \bar{\alpha}^p(T)$$

$$= z^{-1} \bar{\xi}^p(T-1) - \frac{[\bar{\beta}^p(T-1)]_{T-1}}{\epsilon_\beta^p(T-1)} \bar{\beta}^p(T-1) + \frac{[\bar{\alpha}^p(T)]_T}{\epsilon_\alpha^p(T)} \bar{\alpha}^p(T) \quad (45)$$

where we have used (36).

Also, combining (44a) with (40), we have

$$\epsilon_\alpha^p(T) = \langle \bar{\alpha}^p(T), \bar{x}^0 \rangle_T$$

$$= \langle \bar{\alpha}^p(T-1), \bar{x}^0 \rangle_T - [\bar{\alpha}^p(T-1)]_T \langle z^{-1} \bar{\xi}^{p-1}(T-1), \bar{x}^0 \rangle_T$$

$$= w \langle \bar{\alpha}^p(T-1), \bar{x}^0 \rangle_{T-1} + [\bar{\alpha}^p(T-1)]_T \left([\bar{x}^0]_T - \langle z^{-1} \bar{\xi}^{p-1}(T-1), \bar{x}^0 \rangle_T \right)$$

$$= w \epsilon_\alpha^p(T-1) + [\bar{\alpha}^p(T-1)]_T [\bar{\alpha}^p(T)]_T . \quad (46)$$

Finally, it follows from (44b) that

$$\epsilon_\beta^p(T) = w \epsilon_\beta^p(T-1) + [\bar{\beta}^p(T-1)]_T [\bar{\beta}^p(T)]_T . \quad (47)$$

SOLUTION OF LEAST SQUARES PROBLEM

The least squares problem, (9), may now easily be solved by noting that

$$\sum_{i=0}^p g_i^p(T) \bar{x}^i \text{ is simply the } proj_T \text{ of } \bar{y} \text{ onto } S^p. \text{ We define}$$

$$\bar{\gamma}^p(T) = \bar{y} - \sum_{i=0}^p g_i^p(T) \bar{x}^i . \quad (48)$$

The by now familiar procedure yields the order update for $\bar{\gamma}^p$ in the form

$$\bar{\gamma}^{p+1}(T) = \bar{\gamma}^p(T) + K_\gamma^{p+1}(T) \bar{\beta}^{p+1}(T) \quad (49)$$

11

where

$$K_\gamma^{p+1}(T) = -\frac{k_\gamma^{p+1}(T)}{\epsilon_\beta^{p+1}(T)}$$ (50)

and

$$k_\gamma^{p+1}(T) = \langle \bar{\gamma}^p(T), \bar{x}^{p+1} \rangle_T$$

$$= \langle \bar{\gamma}^p(T), z^{-1} \bar{\beta}^p(T-1) \rangle_T .$$ (51)

Similarly, the time updates for $\bar{\gamma}^p$ and $\bar{k}_\gamma^p$ are found to be

$$\bar{\gamma}^p(T) = \bar{\gamma}^p(T-1) - [\bar{\gamma}^p(T-1)]_T \, \xi^p(T)$$ (52)

$$k_\gamma^p(T) = wk_\gamma^p(T-1) + [\bar{\gamma}^{p-1}(T-1)] \, [\beta^p(T)]_T$$ (53)

with

$$[\bar{\gamma}^p(T)]_T = [\bar{\gamma}^p(T-1)]_T \, (1 - \sigma^p(T)) .$$ (54)

The output variables we seek are usually $[\bar{\gamma}^p(T-1)]_T$ as in noise cancelling, or $z^p(T) \stackrel{\Delta}{=} y(T) - [\bar{\gamma}^p(T-1)]_T$ as in equalization [9]. (We address the determination of $g_i^p(T)$ in the next section.) To achieve this, our algorithm need consider only a single component of the vectors $\bar{\alpha}^p, \bar{\beta}^p$, and $\bar{\gamma}^p$. More specifically, we choose

$$\alpha^p(T) = [\bar{\alpha}^p(T)]_T$$ (55a)

$$\beta^p(T) = [\bar{\beta}^p(T)]_T$$ (55b)

$$\gamma^p(T) = [\bar{\gamma}^p(T-1)]_T .$$ (55c)

where removal of the bar indicates a particular vector component. We remark that relations (43) and (54) allow us to alternate between $[\cdot (T-1)]_T$ and $[\cdot (T)]_T$, but for the applications mentioned (55c) is the most prudent choice [9].

The algorithm is now obtained by taking the Tth component of (18) and (19) and the $T + 1^{st}$ of (49). Order updates for the ϵ's are obtained by taking the scalar product of (18),

(19) and (49) with $\bar{\alpha}^p (T)$, $z^{-1} \bar{\beta}^p (T)$, and $\tilde{\gamma}^p (T)$, respectively. These, combined with the time updates of the k's ((42) and (53)) and relations (29) and (37), yield the algorithm of reference [9]

$$\alpha^0 (T) = \beta^0 (T) = x (T) \tag{56a}$$

$$\epsilon_\alpha^0 (T) = \epsilon_\beta^0 (T) = w\epsilon_\alpha^0 (T-1) + x^2 (T) \tag{56b}$$

$$\sigma^{-1} (T) = 0 \tag{56c}$$

$$\gamma^{-1} (T) = y (T) . \tag{56d}$$

For $p = 0, \ldots, p_0$

$$k^p (T) = wk^p (T-1) + \frac{\alpha^p (T) \beta^p (T-1)}{1 - \sigma^{p-1} (T-1)} \tag{57a}$$

$$\alpha^{p+1} (T) = \alpha^p (T) - \frac{k^p (T)}{\epsilon_\beta^p (T-1)} \beta^p (T-1) \tag{57b}$$

$$\beta^{p+1} (T) = \beta^p (T-1) - \frac{k^p (T)}{\epsilon_\alpha^p (T)} \alpha^p (T) \tag{57c}$$

$$\epsilon_\alpha^{p+1} (T) = \epsilon_\alpha^p (T) - (k^p (T))^2 / \epsilon_\beta^p (T-1) \tag{57d}$$

$$\epsilon_\beta^{p+1} (T) = \epsilon_\beta^p (T-1) - (k^p (T))^2 / \epsilon_\alpha^p (T) \tag{57e}$$

$$\sigma^p (T) = \sigma^{p-1} (T) + (\beta^p (T))^2 / \epsilon_\beta^p (T) \tag{57f}$$

$$\gamma^p (T) = \gamma^{p-1} (T) - \frac{k_\gamma^p (T-1)}{\epsilon_\beta^p (T-1)} \frac{\beta^p (T)}{1 - \sigma^{p-1} (T)} \tag{57g}$$

$$k_\gamma^p (T) = wk_\gamma^p (T-1) + \gamma^{p-1} (T) \beta^p (T) . \tag{57h}$$

Note that (56b) follows from (37). Aspects of the implementation of the above algorithm such as initialization may be found in [9].

13

THE NATURAL ISOMORPHISM

It is sometimes desirable to obtain the filter coefficients g_i^p (T) (as in the fast Kalman algorithm, [4]-[5]). This may be achieved via a natural isomorphism which is the analogue of that implicit in [1]. The only difference is that here the inner product is defined with respect to the estimated correlation matrix rather than the true correlation matrix.

More precisely, consider the set of polynomials in z^{-1} of the form

$$A(z^{-1}) = \sum_{i=0}^{\infty} a_i z^{-i} \tag{58}$$

where it is assumed that A has finite degree; i.e., $a_i = 0$ for $i > p$ for some $p < \infty$. Define a family of inner products $(\cdot, \cdot)_T$ on these polynomials by

$$(A(z^{-1}), B(z^{-1}))_T = \sum_{i,j} a_i R_{ij} (T) b_j \tag{59}$$

where

$$R_{ij} (T) = \sum_{t=0}^{T} w^{T-t} x(t-i) \ x(t-j) \ . \tag{60}$$

Note that the conditions of the Appendix insure that $(\cdot, \cdot)_T$ is non-singular on all polynomials of degree less than p_0.

A(z) may be thought of as an operator on the vector space of equation (1) and the mapping $A(z^{-1}) \to A(z^{-1})\bar{x}$ maps the polynomials of degree p onto S^p. It is easy to see that this mapping is a metric space isomorphism (for $p \leqslant p_0$).

$$(A(z^{-1}), B(z^{-1}))_T = \langle A(z^{-1})\bar{x}, B(z^{-1}\bar{x}) \rangle_T \ . \tag{61}$$

It is a simple matter to translate the recursions of the previous sections into recursions of polynomial coefficients via the inverse isometry. For example, let

$\bar{\xi}^p$ (T) $= \displaystyle\sum_{i=0}^{p} c_i^p$ (T)$\bar{x}^i$. Since the $\bar{x}^i$ form a basis for S^p, by matching the coefficients of

$\bar{x}^i$ on both sides of equation (39), we obtain

$$b_i^p \ (T) = b_i^p \ (T-1) - [\bar{\beta}^{\ p} (T-1)]_T \ c_i^{p-1} \ (T-1) \ \ 0 \leqslant i \leqslant p - 1 \ . \tag{62}$$

14

Similarly, since $\{\bar{y}, \bar{x}^i; i = 1, \ldots, p\}$ form a basis for $S^p \oplus y$, equation (52) yields the filter coefficients

$$g_i^p(T) = g_i^p(T-1) - [\gamma^{p-1}(T-1)]_T \, c_i^p(T). \tag{63}$$

This approach, the development of the algorithm through the polynomial coefficients $(a_i^p, b_i^p, c_i^p, g_i^p)$ is most practical if the order is fixed at some value $p = p_0$, and the time updates are used (c.f. [5]). In this case the computational burden is $0(p_0)$ per time update. If g_i^p for $0 \leqslant p \leqslant p_0$, and $0 \leqslant i \leqslant p$ is required there are $p_0^2/2$ variables to evaluate, and the number of computations becomes $0(p_0^2)$. Note, also, that for fixed $p = p_0$ the expressions such as $[\bar{\beta}^{p_0}(T-1)]_T$ which appear throughout (as in (62)) may be computed directly from their definitions in $0(p_0)$ operations. For example,

$$[\bar{\beta}^{p_0}(T-1)]_T = x(T-p_0) - \sum_{i=0}^{p_0-1} b_i^{p_0}(T-1) \, x(T-i).$$

CONCLUDING REMARKS

In this paper we defined a class of metric spaces which resulted in a geometrical derivation of least squares lattice algorithms. This inner-product formalism is closely related to that found in [1], but involves the actual input data rather than its second order statistical properties. This is reflected in the use of a set of metrics indexed by time to replace the expected value. Note that such an approach lends itself naturally to a time-recursive formulation.

An advantage of the above structure is that it provides a single framework which encompasses an entire family of fairly complex algorithms ([3]–[9]). Its geometrical nature also provides a guide for the intuition which should be of use in implementation and future investigations. Finally, we wish to point out that there remain many questions which were beyond the scope of this paper. Among these are stability properties, the role of w for $w < 1$, the significance of the scalar gains K and σ, and the spectral properties of the lattice decomposition. Some of the present techniques may also generalize to the case treated in [10].

APPENDIX

INDEPENDENCE ASSUMPTION

The metric $< \quad >_T$ is singular on $S^p \leftrightarrow \exists \ r_i$ not all zero and

$$\bar{y} = \sum_{i=0}^{p} r_i \ \bar{x}^i \text{ such that } ||\bar{y}||_T^2 = 0 \leftrightarrow \sum_{t=0}^{T} w^{T-t} [\bar{y}]_t^2 = 0 \leftrightarrow [\bar{y}]_t = 0 \text{ for } t \leqslant T \leftrightarrow$$

$$\sum_{i=0}^{p} r_i \ \bar{x}^{i'} = 0 \leftrightarrow \text{ the } \bar{x}^{i'} \text{ are linearly dependent. Note that } \bar{x}^{i'} \text{ are defined in (5).}$$

Note also $< \quad >_T$ is singular $\leftrightarrow \exists \ r_i$ not all zero such that

$$\langle \sum_{i=0}^{p} r_i \ \bar{x}^i, x^j \rangle_T = 0 \text{ for } 0 \leqslant j \leqslant p \leftrightarrow \text{ The matrix } R^p (T) \text{ defined by (60) with i,}$$

$j \leqslant p$ satisfies $\sum_{i=0}^{p} R_{ij}^p (T) \ r_i = 0 \leftrightarrow R^p (T)$ is singular.

The independence condition on the vectors $\bar{x}^{i'}$ of (5) is thus a mixing condition. It says that by the time T all the modes of the estimated correlation matrix R^{p0} must be excited; i.e., $R^{p0} (T)$ is nonsingular.

REFERENCES

1. J. Markel, A. H. Gray Jr., "On Autocorrelation Equations as Applied to Speech Analysis," *IEEE Trans. on Audio and Electroacoustics,* Vol. AU-21, No. 2, April 1973, pp. 69-79.

2. F. Itakura, S. Saito, "Digital Filtering Techniques for Speech Analysis and Systems," *Proc. 7th Int. Cong.,* Budapest, Paper 25-C-1, pp. 261-264, 1971.

3. M. Morf, "Fast Algorithms for Multivariable Systems," Ph.D. Thesis, Stanford, CA, 1974.

4. L. Ljung, M. Morf, D. Falconer, "Fast Calculation of Gain Matrices for Recursive Estimation Schemes," *Int. Journal of Control,* Vol. 27, No. 1, 1978, pp. 1-19.

5. D. D. Falconer, L. Ljung, "Application of Fast Kalman Estimation to Adaptive Equalization," *IEEE Trans. Comm.,* Vol. COM-26, No. 10, Oct. 1978, pp. 1439-1446.

6. M. Morf, D. Lee, "Fast Algorithms for Speech Modeling," Final Report to Defense Communication Agency, Contract DCA 100-77-C-0005, Information Systems Laboratory, Stanford University, Stanford, CA, Nov. 1978.

7. M. Morf, D. Lee, "Recursive Least Squares Ladder Forms for Fast Parameter Tracking," *Proc. of the 1978 IEEE Conf. on Decision and Control,* San Diego, CA, Jan. 1979, pp. 1362-1367.

8. E. H. Satorius, J. D. Pack, "Application of Least Squares Lattice Algorithms to Adaptive Equalization," submitted to IEEE Trans. on Comm.

9. E. H. Satorius, M. J. Shensa, "On the Application of Recursive Least Squares Methods to Adaptive Processing," International Workshop on Applications of Adaptive Control, Yale University, New Haven, Conn., Aug. 1979.

10. B. Friedlander, M. Morf, T. Kailath, L. Ljung, "New Inversion Formulas for Matrices Classified in Terms of Their Distance from Teoplitz Matrices," to appear in *Linear Algebra and Its Applications,* 1979.