Appl. Math. and Comp. Sci., 1994, vol.4, No.1, 53–78

ON STATISTICAL RECONSTRUCTION OF VECTOR RANDOM FIELDS

MACIEJ NIEDŹWIECKI*

The problem of statistical reconstruction of missing samples arbitrarily located within the rectangular segment of a Gaussian vector random field is considered. It is shown that the reconstruction is made up of the deterministic (predictable and hence unique) component and the stochastic (unpredictable and hence nonunique) component. First, the reconstruction procedure is briefly sketched for general homogenous random fields. Next, it is shown that a major computational simplifications can be achieved if the unilateral conditional Markov (CM) model of the field is adopted. For CM fields the possibility of decomposing the restoration procedure into a number of computationally less demanding sub–problems is also discussed.

1. Introduction

The paper deals with the problem of statistical reconstruction of missing samples arbitrarily located within the $N_1 \times N_2$ rectangular segment of a homogenous vector random field (see Fig. 1(a))

$$\mathcal{Y} = \Big\{ y(i,j),\ 1 \le i \le N_1,\ 1 \le j \le N_2 \Big\}$$

Since one of the possible applications of the presented theory as well as the main motivation behind its development comes from the area of image processing (texture analysis and synthesis, masking techniques, etc.), we will further regard $\mathcal{Y}$ as a vectorially scaned (from left to right and from top to bottom) image and the n–dimensional vector $y(i,j)$ — as the intensity of the corresponding vector of pixels.

By *statistical reconstruction* we will mean the process of generating the missing samples in such a way that the reconstructed ("patched") image remains statistically "indistinguishable" from the original one, i.e. it is not possible to locate missing parts of the image by looking at its reconstructed version.

Samples may be scheduled for reconstruction for several different reasons such as:

- degradation of the recording medium due to aging or mishandling (old photographs, scratched optical disks),
- uncorrectable transmission errors in modern image coding systems,
- intentional deletion of some fragments of the image performed as a part of the editing process (as an example consider the task of "closing up" the background after removing a specific object).

* Department of Automatic Control, Faculty of Electronics, Technical University of Gdańsk, ul. Narutowicza 11/12, Gdańsk, Poland

Depending on the case the patterns of missing samples may range from small, isolated spots to large clusters consisting of tens or even hundreds of pixels.

We note that the formulation of the reconstruction problem stated above is considerably more general than the classic attempt to minimize a certain distance between the original signal and its reconstructed version. There is a number of applications where reconstruction based on distance minimization may be not satisfactory as it fails, in general, to satisfy the indistinguishibility test mentioned earlier. For example, when parts of the image missing due to the physical damage or human intervention are filled up it is usually demanded that such cosmetic changes should be unnoticeable.

We will show that the optimal reconstruction is made up of two components. The first (predictable) component is unique and can be interpreted as the orthogonal projection of known samples on the space spanned by unknown samples. The second component removes the "covariance defect" which results if the unknown samples are replaced simply by their predictions (or, more precisely, extrapolations). Since this component can be obtained as any realization of a random variable with pre–specified covariance matrix, it is nonunique (many such corrections can be generated infinitely). The nonunique component can play an important role if the reconstruction procedure is applied to weakly correlated images (fields) or images with large clusters of unknown samples. The paper extends results derived for multivariate time series (Niedźwiecki, 1993).

2. Results for General Random Fields

2.1. Notation

Let $\mathcal{Y}_0 = \{y(i,j), (i,j) \in \Omega_0\}$ where $\Omega_0 \subset \Omega = \{(i,j) : 1 \le i \le N_1, 1 \le j \le N_2\}$ denote the set of d_O observed (known) samples and let $\mathcal{Y}_M = \{y(i,j), (i,j) \in \Omega_M\}$, $\Omega_M = \Omega - \Omega_0$, be the set of d_M missing samples

$$\mathcal{Y}_0 \cup \mathcal{Y}_M = \mathcal{Y}$$

$$d_O + d_M = d$$

where $d = N_1 \times N_2$ and all dimensions are specified in terms of $n \times 1$ blocks.

Denote by $i \oplus j = N_2(i-1) + j$ the lexicographical coordinate of any cell $(i,j) \in \Omega$ and let

$$\mathcal{K} = \Big\{ i \oplus j,\ (i,j) \in \Omega \Big\} = \Big\{ 1, \ldots, d \Big\}$$

$$\mathcal{K}_0 = \Big\{ i \oplus j,\ (i,j) \in \Omega_0 \Big\}$$

$$\mathcal{K}_M = \Big\{ i \oplus j,\ (i,j) \in \Omega_M \Big\}$$

The vector of lexicographically ordered array of samples $\mathcal{Y}$ will be written as

$$x = \Big[x^T(1), \ldots, x^T(d) \Big]^T$$

where $x(i \oplus j) = y(i,j)$, $(i,j) \in \Omega$ (see Fig. 1(b)).

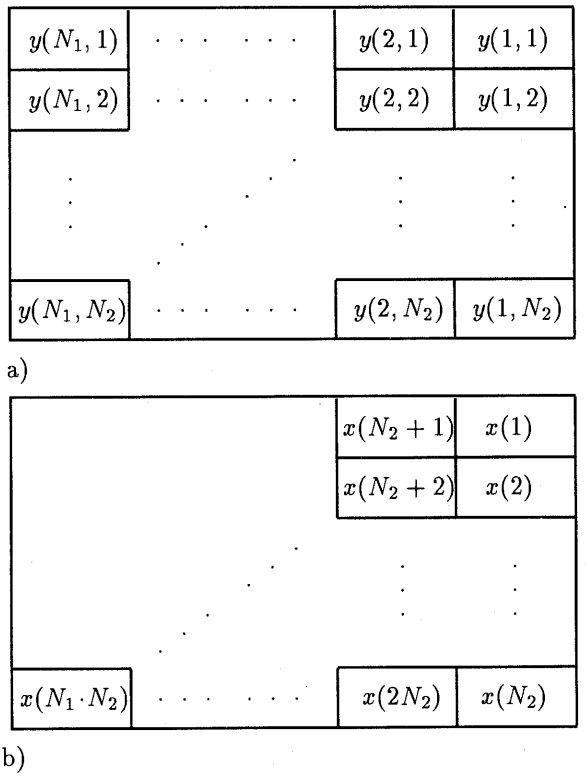

Fig. 1. A rectangular segment of a homogenous vector random field (a) and its lexicographically ordered counterpart (b).

Consider matrices $A[d \times d_1]$, $B[d_1 \times d]$ and $C[d \times d]$, made up of identical blocks of arbitrary dimensions and let $\mathcal{I} = \{r_1, \ldots, r_{d_2}\} \subset \mathcal{K}$, $\mathcal{J} = \{r_1, \ldots, r_{d_3}\} \subset \mathcal{K}$. We will denote by $A_{\langle \mathcal{I}|}$ the $[(d - d_2) \times d_1]$ block matrix obtained after removing from A d_2 block rows indicated by the set $\mathcal{I}$. Similarly, by $B_{|\mathcal{I}\rangle}$ we will denote the $[d_1 \times (d - d_2)]$ block matrix obtained from B after removing its d_2 columns. Finally, by

$$C_{\langle \mathcal{I}|\mathcal{J}\rangle} \triangleq (C_{\langle \mathcal{I}|})_{|\mathcal{J}\rangle} = (C_{|\mathcal{J}\rangle})_{\langle \mathcal{I}|}$$

we will denote the $[(d - d_2) \times (d - d_3)]$ block matrix obtained from C after removing simultaneously the corresponding d_2 rows and d_3 columns.

The following notation will be also used

$$A^T_{|\mathcal{I}\rangle} \triangleq (A^T)_{\langle \mathcal{I}|} = (A_{\langle \mathcal{I}|})^T$$

$$B^T_{\langle \mathcal{I}|} \triangleq (B^T)_{|\mathcal{I}\rangle} = (B_{|\mathcal{I}\rangle})^T$$

$$C^T_{\langle \mathcal{J}|\mathcal{I}\rangle} \triangleq (C^T)_{\langle \mathcal{J}|\mathcal{I}\rangle} = (C_{\langle \mathcal{I}|\mathcal{J}\rangle})^T$$

Note that the operations of matrix transpose and block (row or column) deletion do not commute (i.e., generally, $(A^T)_{|\mathcal{I}} \neq (A_{|\mathcal{I}})^T$ etc.) which means that some caution is needed when using the above shorthands.

Using the notation introduced above the vectors of observed samples x_0 and missing samples x_M can be written down as

$$x_0 = x_{(\mathcal{K}_M|}, \qquad x_M = x_{(\mathcal{K}_0|}$$

2.2. The Case of a Known Covariance Structure of the Field

Consider the problem of reconstruction of missing samples coming from a homogenous, Gaussian random field $\{y(i,j)\}$ with known covariance structure. More precisely, we will assume that $\mathcal{Y}$ is a $d \times d$ dimensional array of n–dimensional zero–mean random vectors $y(i,j)$ with known covariance matrices

$$E\left[y(i,j)y^T(k,l)\right] \triangleq R_{i-k,j-l}$$

forming the positive–definite symmetric $d \times d$ block Toeplitz matrix

$$\mathcal{R} = E[xx^T] = \begin{bmatrix} \mathcal{R}_0 & \mathcal{R}_1 & \cdots & \mathcal{R}_{N_1-1} \\ \mathcal{R}_1^T & \mathcal{R}_0 & \cdots & \mathcal{R}_{N_1-2} \\ \vdots & & \ddots & \vdots \\ \mathcal{R}_{N_1-1}^T & \mathcal{R}_{N_1-2}^T & \cdots & \mathcal{R}_0 \end{bmatrix} \tag{1}$$

with Toeplitz (but in general nonsymmetric) $N_2 \times N_2$ blocks

$$\mathcal{R}_k = \begin{bmatrix} R_{k,0} & R_{k,1} & \cdots & R_{k,N_2-1} \\ R_{k,-1} & R_{k,0} & \cdots & R_{k,N_2-2} \\ \vdots & & \ddots & \vdots \\ R_{k,1-N_2} & R_{k,2-N_2} & \cdots & R_{k,0} \end{bmatrix} \tag{2}$$

It is well known in statistics (see e.g. Schweppe, 1973) that the solution to the quadratic optimization problem

$$E\left[\| \hat{x}_M - x_M \|^2\right] \mapsto \min$$

can be written down in the form

$$\begin{aligned} \hat{x}_M &= E\left[x_M|x_0\right] = E[x_M x_0^T]\left[E[x_0 x_0^T]\right]^{-1} x_0 \\ &= \mathcal{R}_{(\mathcal{K}_0|\mathcal{K}_M)}\left[\mathcal{R}_{(\mathcal{K}_M|\mathcal{K}_M)}\right]^{-1} x_0 = \left[(\mathcal{R}^{-1})_{(\mathcal{K}_0|\mathcal{K}_0)}\right]^{-1}(\mathcal{R}^{-1})_{(\mathcal{K}_0|\mathcal{K}_M)} x_0 \end{aligned} \tag{3}$$

and can be interpreted as the orthogonal projection of the vector of known samples x_0 on the space of unknown samples x_M. We claim, however, that replacement of the missing samples by their predictions (extrapolations) does not complete the task

of their statistical reconstruction. Actually, even though the projection preserves the cross–correlation between the observed and missing samples

$$E\left[\widehat{x}_M x_0^T\right] = E\left[E[x_M|x_0]x_0^T\right] = E\left[x_M x_0^T\right] \tag{4}$$

it distorts the covariance structure of the reconstructed data, namely it produces samples suffering from the "covariance defect"

$$\mathrm{cov}[\widehat{x}_M] < \mathrm{cov}[x_M] \tag{5}$$

In order to see (5) note that

$$x_M = \widehat{x}_M + x_M^{\perp} \tag{6}$$

and, due to orthogonality of $\widehat{x}_M$ and $x_M^{\perp}$

$$\mathrm{cov}[x_M] = \mathrm{cov}[\widehat{x}_M] + \mathrm{cov}[x_M^{\perp}]$$

Because of the covariance defect the reconstruction–by–prediction scheme fails to satisfy the "indistinguishibility from the original" test mentioned in the introduction. The effect is quite clear for weakly correlated random fields. Consider, for example, the limiting case of a white noise field $\{y(i,j)\}$. Because of the lack of spatial correlation between the samples we have

$$\widehat{x}_M = 0$$

However, the incomplete data segment "patched up" with zeros can be considered a rather poor reconstruction of the original as the locations of the missing samples can be easily told from the results. Obviously, for the white noise the covariance defect turns simply into the variance defect — the reconstruction–by–prediction scheme does not preserve the variance of the recovered signal at the points of interest.

Finally, we note that the same effect may be strongly emphasized for large clusters of missing samples (since, vaguely speaking, the energy of the reconstructed signal $\widehat{y}(i,j)$ dies away to zero as we move towards the center of such large clusters).
An obvious way of removing the covariance defect, which we postulate hereby, is by adding to $\widehat{x}_M$ *any* realization of the random variable (independent of $\widehat{x}_M$)

$$\Delta x_M \sim \mathcal{N}(0, \mathcal{R}_{\perp})$$

where

$$\mathcal{R}_{\perp} = \mathrm{cov}[x_M] \; - \; \mathrm{cov}[\widehat{x}_M] = \mathcal{R}_{(\mathcal{K}_0|\mathcal{K}_0)}$$

$$-\mathcal{R}_{(\mathcal{K}_0|\mathcal{K}_M)}\left(\mathcal{R}_{(\mathcal{K}_M|\mathcal{K}_M)}\right)^{-1}\mathcal{R}_{(\mathcal{K}_M|\mathcal{K}_0)} = \left((\mathcal{R}^{-1})_{(\mathcal{K}_0|\mathcal{K}_0)}\right)^{-1} \tag{7}$$

Observe that while $\widehat{x}_M$, the predictable component of x_M, is uniquely determined by the vector of known samples x_0, its complement Δx_M is not. Consequently, due to the nonuniqueness of Δx_M, infinitely many reconstructions of an incomplete data segment can be generated, all statistically indistinguishable from the original.

Finally, it is important to realize that even though the covariance–retouched image $\widehat{x}_M + \Delta x_M$ is not closest, in the sense of the Euclidean norm, to the original image x_M, it can be regarded as an ideal "forgery" of x_M.

Remark 1. Even though formally inclusion of the covariance retouche Δx_M in the reconstruction can be seen as simply "adding some noise" to the predictable component of the signal, such an interpretation may be very deceptive. Since the matrix $\mathcal{R}_\perp$ depends on the covariance properties of the reconstructed field it is clear that the signal Δx_M will somehow inherit these properties. In particular, if the reconstructed field is *smooth* (i.e., in statistical terms, strongly correlated) the corresponding covariance retouche will also be a smooth signal.

We note that the degree of spatial correlation of the field is one of the factors which decides upon the visibility of the covariance defect. In case of weakly correlated fields the effect will be apparent even for small blocks of missing samples. Conversley, for highly correlated sources covariance retouching will be essential only if missing samples form relatively large clusters.

Remark 2. Since covariance retouching amounts to "making up" some missing pieces of information — in order to hide away the fact that the data were incomplete — it may be inadvisable should any decisions be made based on the results of reconstruction.

2.3. Adaptive Reconstruction

The reconstruction procedure presented in section 2.2 was derived under the assumption that all covariance matrices $R_{i,\pm j}$ $(0 \leq i < N_1, \quad 0 \leq j < N_2)$ were known. If not, one might consider the possibility of replacing true covariance matrices by their sample estimates. Unfortunately, the solution based on direct estimation of process covariance matrices can be hardly recommended in practice. Analyzing carefully the structure of the matrix $\mathcal{R}$ (1) one finds out that it is made up of $N_1 N_2 + (N_1 - 1)(N_2 - 1) > d$ different matrices $R_{i,j}$, i.e. that the number of matrices one should estimate is greater than the number of available samples. Hence, from the statistical point of view the problem is ill posed.

The possible way out of this difficulty could be by adopting a parsimonyous parametric model of a random field, that is, the model which allows for the covariance structure of the field to be characterized in terms of a small number of matrix coefficients. The corresponding reconstruction procedure, analogous to the one suggested for 1–D reconstruction (Niedźwiecki, 1993) would consist of three steps:

1. Estimate parameters of the adopted spatial model of the random field.
2. Obtain the estimates of the covariance matrices in (2) exploiting the relationship between the (theoretical) covariance structure of the field and parameters of its spatial model.
3. Apply the non–adaptive reconstruction procedure after replacing unknown covariance matrices with their estimates.

Unfortunately, in the case of 2–D reconstruction the second step above may cause serious problems as there is no unique relationship between the sequence of autocorrelation matrices and the sequence of spatial matrix coefficients — the one–to–one

mapping often described as the autocorrelation matching property of 1–D signals is lost for 2–D processes (Marple, 1988; Marzetta, 1988).

We will show that for a special class of spatial models, known as conditional Markov models, it is possible to skip the second step of the algorithm summarized above, i.e. it is possible to base the reconstruction procedure upon the estimates of model coefficients without the need of estimating covariance matrices of the process. The approach works for almost all patterns of missing samples (it fails only in the case where some of the missing samples are located in a small "restricted" area) and offers the means for decomposition of the reconstruction problem into a number of computationally less demanding sub–problems.

3. Reconstruction Based on Conditional Markov Models of Random Fields

3.1. Notation

Denote by $\prec$ the symbol of ordering among pairs of indexes: $(i,j) \prec (k,l)$ iff $i < k$ or $(i = k \quad \text{and} \quad j < l)$. Using this ordering the entire plane can be divided at any point (i,j) (after exclusion of (i,j) itself) into two parts: the nonsymmetric upper–half plane and nonsymmetric lower half–plane, i.e. the "past" and "future", respectively.

We will consider a special class of spatial interaction models known as causal (supported on the nonsymmetric upper half–plane) conditional Markov (CM) models. Denote by

$$P = \Big\{(i_k, j_k) : k = 1, \ldots, p;\ (0,0) \prec (i_1, j_1) \prec \ldots \prec (i_p, j_p)\Big\}$$

the set of indexes determining p neighbor elements for $y(i,j)$.

The causal (unilateral) CM model of a vector random field $\{y(i,j)\}$ is defined as (Chellappa and Kashyap, 1982; Kashyap, 1983)

$$y(i,j) = \sum_{k=1}^{p} A_k y(i - i_k, j - j_k) + n(i,j) \tag{8}$$

where $A_1, \ldots, A_p$ are the coefficient matrices of the model and $\{n(i,j)\}$ is a zero mean white vector process with covariance matrix $\operatorname{cov}[n(i,j)] = \rho$.

Suppose that $P \subset \Omega$. We will partition the finite lattice Ω into mutually exclusive and totally inclusive subsets Ω_B, the boundary set, and Ω_I, the interior set:

$$\Omega_B = \Big\{(i,j) \in \Omega \text{ such that } (i-k, j-l) \notin \Omega \text{ for at least one } (k,l) \in P\Big\}$$

$$\Omega_I = \Omega - \Omega_B$$

Likewise, denote by

$$\mathcal{Y}_B = \Big\{ y(i,j), (i,j) \in \Omega_B \Big\}$$

$$\mathcal{Y}_I = \Big\{ y(i,j), (i,j) \in \Omega_I \Big\}$$

a similar decomposition of the sample space $\mathcal{Y}$

$$\mathcal{Y}_B \cup \mathcal{Y}_I = \mathcal{Y}$$

$$d_B + d_I = d$$

where d_B and d_I stand for numbers of samples in $\mathcal{Y}_B$ and $\mathcal{Y}_I$, respectively.

Finally, denoting by $\mathcal{K}_B$ and $\mathcal{K}_I$ the sets of lexicographical coordinates of Ω_B and Ω_I, we can define the vectors of boundary and interior samples as

$$x_B = x_{(\mathcal{K}_I|}, \qquad x_I = x_{(\mathcal{K}_B|}$$

3.2. Basic Scheme

Suppose that we know ρ and $A_1, \ldots, A_p$, i.e., we know all coefficient matrices in the conditional Markov model (8). Under Gaussian assumptions

$$\widehat{x}_M = E[x_M|x_0] = \arg\max_{x_M} p(x_M|x_0) = \arg\max_{x_M} p(x) \tag{9}$$

For a homogenous random field truncated to a finite lattice Ω the likelihood function $p(x)$ can be expressed in the form

$$p(x) = p(\mathcal{Y}) = p(\mathcal{Y}_I, \mathcal{Y}_B) = p(\mathcal{Y}_I|\mathcal{Y}_B)\pi(\mathcal{Y}_B) \tag{10}$$

Denote by $w = \big[w^T(1), \ldots, w^T(d)\big]^T$ the vector of lexicographically ordered array of noise samples

$$w(i \oplus j) = n(i,j)$$

and let $e = \big[e^T(1), \ldots, e^T(d)\big]^T$ where

$$e(r) = \rho^{-\frac{1}{2}} w(r)$$

The conditional likelihood can be written down as

$$p(\mathcal{Y}_I|\mathcal{Y}_B) = p(x_I|x_B) = \prod_{r \in \mathcal{K}_I} p\Big(x(r) \mid \text{ all } x(s) \text{ such that } r > s \in \mathcal{K} \Big)$$

$$\propto \exp\Big\{ -\frac{1}{2} \sum_{r \in \mathcal{K}_I} \| w(r) \|^2_{\rho^{-1}} \Big\} = \exp\Big\{ -\frac{1}{2} \sum_{r \in \mathcal{K}_I} \| e(r) \|^2 \Big\} \tag{11}$$

$$= \exp\Big\{ -\frac{1}{2} \| e_{(\mathcal{K}_B|} \|^2 \Big\}$$

Let $\mathcal{K}_P = \{i \oplus j, (i,j) \in P\} = \{r_1, \ldots, r_p\}$, $r_k = i_k \oplus j_k$.

Introduce the lower band $d \times d$ block Toeplitz matrix $\mathcal{C}$:

$$\mathcal{C} = \begin{bmatrix} C_0 & 0 & \cdots & & & \cdots & 0 & 0 \\ \vdots & C_0 & & & & & & 0 \\ C_1 & \vdots & \ddots & & & & & \vdots \\ \vdots & C_1 & & & & & & \\ C_p & \vdots & \ddots & & & & & \\ 0 & C_p & & & & & & \\ 0 & 0 & \ddots & & & & & \\ \vdots & \vdots & & C_p & \cdots & C_1 & \cdots & C_0 \quad 0 \\ 0 & 0 & & 0 & C_p & \cdots & C_1 & \cdots \quad C_0 \end{bmatrix} \tag{12}$$

where

$$\mathcal{C}[r,s] = \begin{cases} C_0 = \rho^{-1/2} & \text{if } r = s \\ C_k = -\rho^{-1/2} A_k & \text{if } r - s = r_k \in \mathcal{K}_P \\ 0 & \text{elsewhere} \end{cases} \tag{13}$$

Then it is straightforward to show that

$$\mathcal{C}_{(\mathcal{K}_B|} x = e_{(\mathcal{K}_B|} \tag{14}$$

Combining (14) with (11) we get

$$p(\mathcal{Y}_I|\mathcal{Y}_B) \propto \exp\left\{-\frac{1}{2} x^T \mathcal{C}_{|\mathcal{K}_B)}^T \mathcal{C}_{(\mathcal{K}_B|} x\right\} \tag{15}$$

For a homogenous Gaussian field the second component of the likelihood function, the prior density of boundary samples $\pi(\mathcal{Y}_B)$, can be expressed in the form

$$\pi(\mathcal{Y}_B) = \pi(x_B) \propto \exp\left\{-\frac{1}{2} x_B^T R_B^{-1} x_B\right\} = \exp\left\{-\frac{1}{2} x^T \mathcal{B} x\right\} \tag{16}$$

where

$$R_B = E[x_B x_B^T], \qquad \mathcal{B} = I_{|\mathcal{K}_I)} R_B^{-1} I_{(\mathcal{K}_I|} \tag{17}$$

and I denotes the $d \times d$ (block) identity matrix.

Finally, combining (15) and (16) we get

$$p(x) \propto \exp\{-J(x)\} \tag{18}$$

where

$$J(x) = \frac{1}{2} x^T \Pi x \tag{19}$$

and

$$\Pi = \mathcal{C}_{|\mathcal{K}_B)}^T \mathcal{C}_{(\mathcal{K}_B|} + \mathcal{B} \tag{20}$$

Maximization of the likelihood function with respect to x_M is equivalent to minimization of the quadratic form $J(x)$. Note that

$$\frac{\partial J(x)}{\partial x_M} = \Pi_{(\mathcal{K}_0|}x = \Pi_{(\mathcal{K}_0|\mathcal{K}_0)}x_M + \Pi_{(\mathcal{K}_0|\mathcal{K}_M)}x_0 \tag{21}$$

Hence, after setting the partial derivative (21) to zero at $x_M = \widehat{x}_M$, we get

$$\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}\widehat{x}_M = -\Pi_{(\mathcal{K}_0|\mathcal{K}_M)}x_0 \tag{22}$$

i.e. (see the remark below)

$$\widehat{x}_M = -\left[\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}\right]^{-1}\left[\Pi_{(\mathcal{K}_0|\mathcal{K}_M)}\right]x_0 \tag{23}$$

Remark 3. Note that $p(x) = \mathcal{N}(0, \mathcal{R})$ and, according to (18), $p(x) = \mathcal{N}(0, \Pi^{-1})$. Consequently

$$\Pi = \mathcal{R}^{-1} \tag{24}$$

and the matrix $\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}$ appearing in (23) is always invertible (as a principle submatrix of $\mathcal{R}^{-1}$ it is positive–definite).

With the predictable part of the reconstruction given by (23) it only remains to determine the covariance matrix $\mathcal{R}_\perp$ of its unpredictable part $x_M^\perp$. We will show that

Lemma 1.

$$\mathcal{R}_\perp = \mathrm{cov}\left[x_M^\perp\right] = \mathrm{cov}\left[\Delta x_M\right] = \left(\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}\right)^{-1} \tag{25}$$

Proof. Observe that

$$\mathcal{B}_{(\mathcal{K}_0|}x = \mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_0)}x_M + \mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_M)}x_0 = \mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_B)}x_I + \mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_I)}x_B$$

and

$$\mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_B)} = 0$$

Hence

$$\mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_0)}x_M = -\mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_M)}x_0 + \mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_I)}x_B \tag{26}$$

Similarly, after multiplying both sides of (14) by the matrix $\mathcal{C}_{(\mathcal{K}_0|\mathcal{K}_B)}^T$ we get

$$\mathcal{C}_{(\mathcal{K}_0|\mathcal{K}_B)}^T \mathcal{C}_{(\mathcal{K}_B|}x = \mathcal{C}_{(\mathcal{K}_0|\mathcal{K}_B)}^T \mathcal{C}_{(\mathcal{K}_B|\mathcal{K}_0)}x_M + \mathcal{C}_{(\mathcal{K}_0|\mathcal{K}_B)}^T \mathcal{C}_{(\mathcal{K}_B|\mathcal{K}_M)}x_0$$

$$= \mathcal{C}_{(\mathcal{K}_0|\mathcal{K}_B)}^T e_{(\mathcal{K}_B|}$$

resulting in

$$\mathcal{C}^T_{(\mathcal{K}_0|\mathcal{K}_B)}\mathcal{C}_{(\mathcal{K}_B|\mathcal{K}_0)}x_M = -\mathcal{C}^T_{(\mathcal{K}_0|\mathcal{K}_B)}\mathcal{C}_{(\mathcal{K}_B|\mathcal{K}_M)}x_0 + \mathcal{C}^T_{(\mathcal{K}_0|\mathcal{K}_B)}e_I \tag{27}$$

where $e_I = e_{(\mathcal{K}_B|}$.

Adding equations (26) and (27) sidewise we arrive at

$$\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}x_M = -\Pi_{(\mathcal{K}_0|\mathcal{K}_M)}x_0 + \mathcal{C}^T_{(\mathcal{K}_0|\mathcal{K}_B)}e_I + \mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_I)}x_B \tag{28}$$

leading to (c.f. (22))

$$\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}[x_M - \widehat{x}_M] = \mathcal{C}^T_{(\mathcal{K}_0|\mathcal{K}_B)}e_I + \mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_I)}x_B = \epsilon_M \tag{29}$$

Since for the unilateral CM model (8) the noise vector e_I is independent of x_B and

$$\mathrm{cov}[e_I] = I_{(\mathcal{K}_B|\mathcal{K}_B)}$$

$$\mathrm{cov}[x_B] = R_B$$

it holds

$$\begin{aligned}\mathrm{cov}[\epsilon_M] &= \mathcal{C}^T_{(\mathcal{K}_0|\mathcal{K}_B)}\mathrm{cov}[e_I]\mathcal{C}_{(\mathcal{K}_B|\mathcal{K}_0)} + \mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_I)}\mathrm{cov}[x_B]\mathcal{B}^T_{(\mathcal{K}_I|\mathcal{K}_0)} \\ &= \mathcal{C}^T_{(\mathcal{K}_0|\mathcal{K}_B)}\mathcal{C}_{(\mathcal{K}_B|\mathcal{K}_0)} + \mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_0)} = \Pi_{(\mathcal{K}_0|\mathcal{K}_0)}\end{aligned} \tag{30}$$

leading finally to

$$\mathrm{cov}[\widehat{x}_M] = \left[\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}\right]^{-1}\mathrm{cov}[\epsilon_M]\left[\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}\right]^{-1} = \left[\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}\right]^{-1}$$

Since the matrix $\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}$ is positive definite, a Cholesky decomposition technique can be used for the purpose of solving (22). We note that $\mathcal{R}_\perp^{1/2}$, the right–hand Cholesky factor of the matrix $[\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}]^{-1}$ (related to the left–hand factor of $\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}$), can be used for the purpose of generating the vector of corrections Δx_M via

$$\Delta x_M = \mathcal{R}_\perp^{1/2}\eta$$

where η is any realization of a d_M–dimensional Gaussian variable with zero mean and unit covariance matrix.

3.3. Scheme Based on "Inverse" CM Models

It is a well–known fact that every process described by the causal CM model (8) admits also the following "inverse" (anti–causal, i.e. supported on nonsymmetric lower half plane) CM model equation

$$y(i,j) = \sum_{k=1}^{p} A_k^\star y(i+i_k, j+j_k) + n^\star(i,j) \tag{31}$$

where $A_1^\star, \ldots, A_p^\star$ is the set of "inverse" CM model matrices and $\{n^\star(i,j)\}$ denotes another white noise sequence, different from $\{n(i,j)\}$, characterized by a covariance matrix $\rho^\star$.

In the scalar case it is possible to show that (Marple, 1988)

$$a_k = a_k^\star, \quad k = 1, \ldots, p \qquad \rho = \rho^\star$$

where $a_1, \ldots, a_p, \rho$ and $a_1^\star, \ldots, a_p^\star, \rho^\star$ denote scalar counterparts of the corresponding matrices. However, there is no trivial relationship between the matrices $A_1, \ldots, A_p, \rho$ and $A_1^\star, \ldots, A_p^\star, \rho^\star$ in a general vector case.

The duality between the "normal" CM description (8) and "inverse" CM description (31) is analogous to the duality between forward–time and backward–time models in 1–D signal processing. It will play a crucial role in our further developments. Denote by $\mathcal{K}_B^\star$ and $\mathcal{K}_I^\star$ the sets of lexicographical coordinates of $\Omega_B^\star$ and $\Omega_I^\star$, the boundary set and the interior set for the "inverse" CM model (31).

Let

$$e^\star = \left[e^\star(1), \ldots, e^\star(d)\right]^T$$

(where $e^\star(r) = (\rho^\star)^{-1/2} w^\star(r)$ and $w^\star(i \oplus j) = n^\star(i,j)$) be the corresponding vector of normalized noise samples.

Finally, let $\mathcal{C}^\star$ denote the upper band $d \times d$ block Toeplitz matrix

$$\mathcal{C}^\star = \begin{bmatrix} C_0^\star & \cdots & C_1^\star & \cdots & C_p^\star & 0 & 0 & \cdots & 0 \\ 0 & C_0^\star & \cdots & C_1^\star & \cdots & C_p^\star & 0 & \cdots & 0 \\ \vdots & & \ddots & & \ddots & & \ddots & & \\ & & & & & & & C_p^\star & 0 \\ & & & & & & & \vdots & C_p^\star \\ & & & & & & & C_1^\star & \vdots \\ \vdots & & & & & & & \vdots & C_1^\star \\ 0 & & & & & & & C_0^\star & \vdots \\ 0 & 0 & \ddots & & & & & 0 & C_0^\star \end{bmatrix} \tag{32}$$

where

$$\mathcal{C}^\star[r,s] = \begin{cases} C_0^\star = (\rho^\star)^{-1/2} & \text{if } r = s \\ C_k^\star = -(\rho^\star)^{-1/2} A_k^\star & \text{if } s - r = r_k \in \mathcal{K}_P \\ 0 & \text{elsewhere} \end{cases} \tag{33}$$

Similarly as it took place before one can show that

$$p(x) \propto \exp\left\{-J^\star(x)\right\} = \exp\left\{-\frac{1}{2} x^T \Pi^\star x\right\}$$

where

$$\Pi^\star = (\mathcal{C}^\star)^T_{|\mathcal{K}_B^\star)}\mathcal{C}^\star_{(\mathcal{K}_B^\star|} + \mathcal{B}^\star \tag{34}$$

and the boundary covariance matrix $\mathcal{B}^\star$ is defined analogously to $\mathcal{B}$ in (17).

Since models (8) and (31) describe the same vector field characterized by the covariance matrix $\mathcal{R}$, it must hold

$$J(x) = J^\star(x)$$

$$\Pi^\star = \Pi = \mathcal{R}^{-1}$$

which means that (20) and (34) form two equivalent descriptions of the same positive–definite matrix.

3.4. Evaluation of the Projection Matrix

Using the relationships derived in the preceding section we will show that almost all elements of the $d \times d$ matrix Π can be expressed directly either in terms of the CM matrices $C_0, \ldots, C_p$ or in terms of the "inverse" CM matrices $C_0^\star, \ldots, C_p^\star$ (or in both ways).

First, we note that

$$\mathcal{B}_{(\mathcal{K}_B|} = \mathcal{B}^\star_{(\mathcal{K}_B^\star|} = 0$$

$$\mathcal{B}_{|\mathcal{K}_B)} = \mathcal{B}^\star_{|\mathcal{K}_B^\star)} = 0$$

Consequently it holds

$$\Pi_{(\mathcal{K}_B|} = \mathcal{C}^T_{(\mathcal{K}_B|\mathcal{K}_B)}\mathcal{C}_{(\mathcal{K}_B|}, \qquad \Pi_{(\mathcal{K}_B^\star|} = (\mathcal{C}^\star)^T_{(\mathcal{K}_B^\star|\mathcal{K}_B^\star)}\mathcal{C}^\star_{(\mathcal{K}_B^\star|} \tag{35}$$

and

$$\Pi_{|\mathcal{K}_B)} = \mathcal{C}^T_{|\mathcal{K}_B)}\mathcal{C}_{(\mathcal{K}_B|\mathcal{K}_B)}, \qquad \Pi_{|\mathcal{K}_B^\star)} = (\mathcal{C}^\star)^T_{|\mathcal{K}_B^\star)}\mathcal{C}^\star_{(\mathcal{K}_B^\star|\mathcal{K}_B^\star)} \tag{36}$$

A careful analysis of the above relationships leads to the conclusion that the only elements of Π which cannot be directly expressed in terms of CM matrices (neither "normal" nor "inverse") have the form $\Pi[i,j]$ $(i,j \in \mathcal{K}_C)$ where $\mathcal{K}_C = \mathcal{K}_B \cap \mathcal{K}_B^\star$ is the set which will be further referred to as the *critical boundary zone* (see Fig. 2).

Consequently, if all missing samples are located outside the critical boundary zone, i.e. if $\mathcal{K}_C \subset \mathcal{K}_0$, the projection matrix in (23) can be determined without the need to compute R_B (which is a condition of explicit solvability of the reconstruction problem).

Two important special cases are worth mentioning.

First, if $\mathcal{K}_B \subset \mathcal{K}_0$ that is, if all samples belonging to the boundary set Ω_B are known, we have $\mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_0)} = 0$, $\mathcal{B}_{(\mathcal{K}_0|\mathcal{K}_M)} = 0$ and hence

$$\Pi_{(\mathcal{K}_0|\mathcal{K}_0)} = \mathcal{C}^T_{(\mathcal{K}_0|\mathcal{K}_B)}\mathcal{C}_{(\mathcal{K}_B|\mathcal{K}_0)}$$

$$\Pi_{(\mathcal{K}_0|\mathcal{K}_M)} = \mathcal{C}^T_{(\mathcal{K}_0|\mathcal{K}_B)}\mathcal{C}_{(\mathcal{K}_B|\mathcal{K}_M)}$$

i.e. the projection matrix can be expressed directly in terms of parameters of the causal CM model (8).

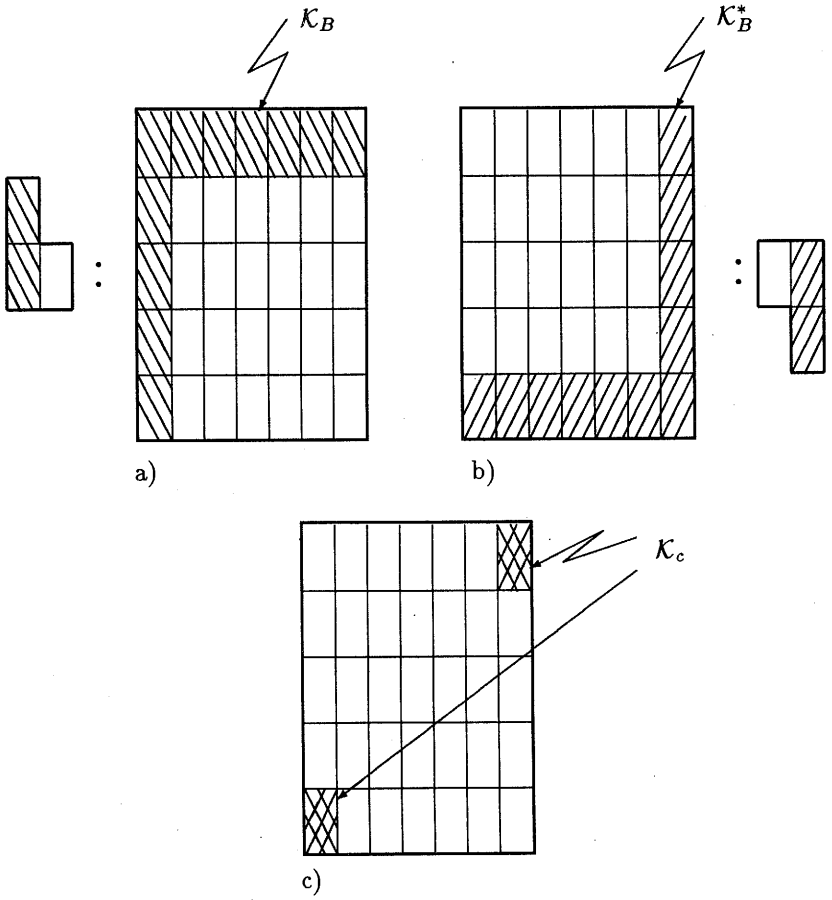

Fig. 2. The boundary set (a), "inverse" boundary set (b) and critical boundary zone (c) for the image obeying a second–order unilateral CM model (supporting pixels for $(0,0)$ are $(0,-1)$ and $(-1,-1)$).

Second, in the case where $\mathcal{K}_B^* \subset \mathcal{K}_0$, it holds

$$\Pi_{(\mathcal{K}_0|\mathcal{K}_0)} = (\mathcal{C}^\star)^T_{(\mathcal{K}_0|\mathcal{K}_B^*)} \mathcal{C}^\star_{(\mathcal{K}_B^*|\mathcal{K}_0)}$$

$$\Pi_{(\mathcal{K}_0|\mathcal{K}_M)} = (\mathcal{C}^\star)^T_{(\mathcal{K}_0|\mathcal{K}_B^*)} \mathcal{C}^\star_{(\mathcal{K}_B^*|\mathcal{K}_M)}$$

i.e. the projection matrix can be written down in terms of parameters of the anti–causal model (31).

If $\mathcal{K}_M \cap \mathcal{K}_C \neq \emptyset$, the procedure becomes quite complicated as it requires solving a difficult inverse problem — determination of the covariance structure of the field given parameters of its spatial model. Since the autocorrelation matching property does not apply to random fields (except for special cases) only approximate methods can be used to recover the 2–D autocorrelation sequence $\{R_{i,j}\}$ from the CM matrices $C_0, \ldots, C_p$ or $C_0^\star, \ldots, C_p^\star$ (Marple, 1988).

Remark 4. We note that in the scalar case

$$\mathcal{C}^\star = \mathcal{C}^T$$

and hence

$$(\mathcal{C}^\star)^T_{|\mathcal{K}^*_B|}\mathcal{C}^\star_{|\mathcal{K}^*_B|} = \mathcal{C}_{|\mathcal{K}^*_B|}\mathcal{C}^T_{|\mathcal{K}^*_B|} \tag{37}$$

For scalar random fields the matrix Π is centrosymmetric

$$\Pi[i,j] = \Pi[d-i+1, d-j+1]$$

which is a consequence of the fact that it is an inverse of a centrosymmetric matrix $\mathcal{R}$ (Marple, 1988). Unfortunately this property does not extend to the vector case where $\mathcal{R}$ can be shown to be (only) block–centrosymmetric (inverse of a centrosymmetric matrix is centrosymmetric but inverse of a block–centrosymmetric matrix is not block–centrosymmetric).

Remark 5. Not surprisingly, existence of the critical boundary zone is closely related to the correlation matching problem mentioned before.
Consider, for example, the following simple autoregressive (AR) process with quarter plane region of support

$$y(i,j) = A_1 y(i,j-1) + A_2 y(i-1,j) + A_3 y(i-1,j-1) + n(i,j) \tag{38}$$

for which the set of Yule–Walker equations takes the form

$$\begin{bmatrix} R_{0,0} & R_{0,1} & R_{1,0} & R_{1,1} \\ R^T_{0,1} & R_{0,0} & R_{1,-1} & R_{1,0} \\ R^T_{1,0} & R^T_{1,-1} & R_{0,0} & R_{0,1} \\ R^T_{1,1} & R^T_{1,0} & R^T_{0,1} & R_{0,0} \end{bmatrix} \begin{bmatrix} I \\ -A_1^T \\ -A_2^T \\ -A_3^T \end{bmatrix} = \begin{bmatrix} \rho \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{39}$$

According to (39), $k_R = 5$ unique autocorrelation matrices $R_{0,0}$, $R_{0,1}$, $R_{1,0}$, $R_{1,1}$, $R_{1,-1}$ are required to solve for $k_A = 4$ autoregressive (plus noise covariance) matrices A_1, A_2, A_3, ρ. There are, therefore, insufficient AR coefficients to uniquely determine autocorrelation coefficients. The critical boundary zone consists in this case of $k_C = 2$ elements, $y(1,N_1)$ and $y(N_2,1)$, situated in the lower left and upper right corners, respectively.

One can show that for a general 2–D AR process the size of the critical boundary area is proportional to the "degree of ill–conditioning" of (39), namely

$$k_C = 2(k_R - k_A)$$

Decomposition of the Reconstruction Problem

Consider any pixel $(i,j) \in \Omega$, $i \oplus j = r \in \mathcal{K}$. Denote by

$$N_r = \left\{ s : \; s \doteq r + r_i - r_j, \quad i,j = 0,\ldots,p \right\} \cap \mathcal{K}$$

where $r_0 = 0$, the set of all pixels in $\mathcal{K}$ directly related to r via the CM equation (8) (or (31)), including the pixel r itself — see Figure 3(a).

Suppose, for the time being, that there is only one missing sample $x(r)$ and that it is located in the interior set $\mathcal{Y}_I$, i.e. $r \in \mathcal{K}_I$.

Denote by ν, $0 \leq \nu \leq p$, the largest index such that $r + r_\nu \leq d$. A careful examination of (14)–(20) leads to the conclusion that

$$\frac{\partial J(x)}{\partial x(r)} = \frac{1}{2} \sum_{j=0}^{\nu} \frac{\partial}{\partial x(r)} \| x(r + r_j) - \sum_{k=1}^{p} A_k x(r + r_j - r_k) \|^2_{\rho^{-1}}$$

i.e. that the gradient of the quadratic form $J(x)$ evaluated with respect to the unknown vector $x(r)$ depends *only* on the data vectors located in N_r. Consequently $\widehat{x}(r)$, the predictable part of $x(r)$ obtained by solving $\partial J(x)/\partial \widehat{x}(r) = 0$, will be a function of samples located in the neighbor set $N_r - \{r\}$ and will not depend on the remaining samples.

By reiterating the above argument for the whole set of pixels $\mathcal{F} \subset \mathcal{K}_I$ one can show that the corresponding set of neighbor elements takes the form

$$\bigcup_{r \in \mathcal{F}} N_r - \mathcal{F}$$

Since the process can be alternatively modelled using the inverse CM equation (31), the neighborhood description given above remains valid for all sets $\mathcal{F} \subset \mathcal{K}_I^\star$. Finally, combining arguments for the "normal" and "inverse" models one can show that it holds for every set $\mathcal{F} \subset (\mathcal{K}_I \cup \mathcal{K}_I^\star)$.

Consider now the case where there is a single missing sample $x(r)$ located in the critical boundary zone, i.e. $r \in \mathcal{K}_C = \mathcal{K} - (\mathcal{K}_I \cup \mathcal{K}_I^\star)$. Based on model (8) one finds out that the gradient $\partial J(x)/\partial x(r)$ is a function of samples located in $N_r \cup \mathcal{K}_B$. Similarly, using the "inverse" model description, one can show that $\partial J(x)/\partial x(r)$ depends on samples located in the set $N_r \cup \mathcal{K}_B^\star$.

Combining both arguments (which must hold simultaneously) one comes to the conclusion that the set of neighbor elements for $x(r)$ is

$$(N_r \cup \mathcal{K}_B) \cap (N_r \cup \mathcal{K}_B^\star) - \{r\} = (N_r \cup \mathcal{K}_C) - \{r\}$$

and, more generally, that the neighborhood of any set $\mathcal{F}$, $\mathcal{F} \cap \mathcal{K}_C \neq \emptyset$ can be written down as

$$\left(\bigcup_{r \in \mathcal{F}} N_r \cup \mathcal{K}_C \right) - \mathcal{F}$$

Based on the discussion carried out above, the Markovian neighborhood $N(\mathcal{F})$ of any set $\mathcal{F} \subset \mathcal{K}$ can be defined in the form (see Fig. 3(b))

$$N(\mathcal{F}) = \begin{cases} \bigcup_{r \in \mathcal{F}} N_r - \mathcal{F} & \text{if } \mathcal{F} \cap \mathcal{K}_C = \emptyset \\ \left(\bigcup_{r \in \mathcal{F}} N_r \cup \mathcal{K}_C \right) - \mathcal{F} & \text{if } \mathcal{F} \cap \mathcal{K}_C \neq \emptyset \end{cases} \tag{40}$$

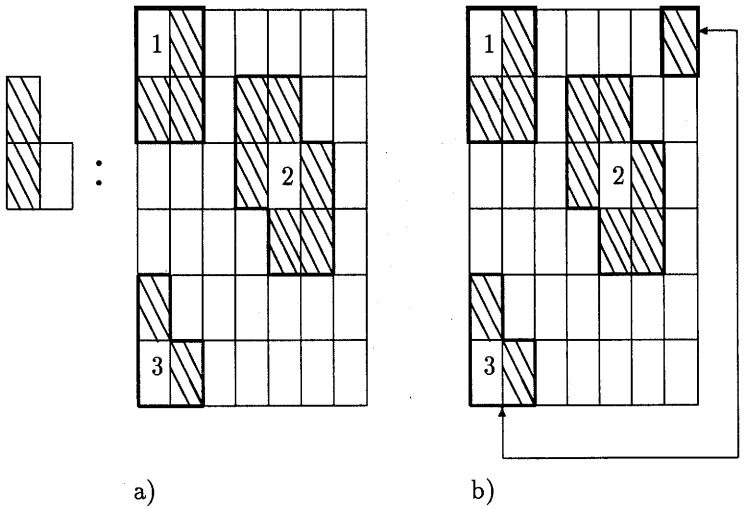

Fig. 3. a) The neighbor set N_r for three pixels differently located within the image obeying a second-order unilateral CM model,
b) Markovian neigborhoods for three different sets of pixels for the same image.

Denote by $\mathcal{Y}_{\mathcal{F}}$ and $\mathcal{Y}_{N(\mathcal{F})}$ the sets of all samples located in $\mathcal{F}$ and $N(\mathcal{F})$, respectively. Let $\overline{\mathcal{Y}}_{\mathcal{F}} = \mathcal{Y} - \mathcal{Y}_{\mathcal{F}}$. The term *Markovian neighborhood* stems from the fact that

$$p(\mathcal{Y}_{\mathcal{F}}|\overline{\mathcal{Y}}_{\mathcal{F}}) = p(\mathcal{Y}_{\mathcal{F}}|\mathcal{Y}_{N(\mathcal{F})}) \tag{41}$$

i.e. $\mathcal{Y}_{N(\mathcal{F})}$ serves as a sufficient statistics for the purpose of estimation of $\mathcal{Y}_{\mathcal{F}}$.

According to (41) it holds

$$p(\mathcal{Y}_M|\mathcal{Y}_0) = p(\mathcal{Y}_M|\mathcal{Y}_{N(\mathcal{K}_M)}) \tag{42}$$

where $N(\mathcal{K}_M)$, the Markovian neighborhood of $\mathcal{K}_M$, is a subset of $\mathcal{K}_0$. There are two important consequences of (42).

First, it is evident from (42) that the Markovian reconstruction of x_M does not depend on the samples located in

$$\mathcal{K}_D = \mathcal{K}_O - N(\mathcal{K}_M)$$

– such samples will be further referred to as dummy. Not surprisingly, those columns of the projection matrix $[\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}]^{-1}\Pi_{(\mathcal{K}_0|\mathcal{K}_M)}$ which correspond to dummy variables in x_0 can be shown to be null. Removing such null vectors (along with the corresponding elements of x_0) results in the following reduced-order version of (23)

$$\widehat{x}_M = -\left[\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}\right]^{-1}\left[\Pi_{<\mathcal{K}_0|\mathcal{K}_S>}\right] x_S \tag{43}$$

where

$$\mathcal{K}_S = \mathcal{K}_M \cup \mathcal{K}_D, \qquad x_S = x_{<\mathcal{K}_S|}$$

Note that while the matrix $\Pi_{(\mathcal{K}_0|\mathcal{K}_M)}$ in (23) is $d_M \times d_O$–dimensional, the dimension of $\Pi_{<\mathcal{K}_0|\mathcal{K}_S>}$ in (43) is $d_M \times d_S$ where $d_S = d_O - d_D$ and d_D denotes the number of dummy samples. Since usually (especially for low-order CM models) it holds that $d_S \ll d_O$, the modified scheme (43) may offer huge computational savings over (23).

The second simplification mentioned above takes place in the case where the indicator set $\mathcal{K}_M$ can be partitioned into q mutually exclusive and totally inclusive subsets $\mathcal{F}_1, \ldots, \mathcal{F}_q$

$$\mathcal{F}_1 \cup \ldots \cup \mathcal{F}_q = \mathcal{K}_M, \qquad \mathcal{F}_i \cap \mathcal{F}_j = \emptyset \text{ for } i \neq j \tag{44}$$

such that

$$N(\mathcal{F}_i) \subset \mathcal{K}_0, \qquad i = 1, \ldots, q \tag{45}$$

According to (42) it is sufficient to know samples in $N(\mathcal{F}_i)$ in order to reconstruct $\mathcal{Y}_{\mathcal{F}_i}$. Consequently, the statistical reconstruction problem for one large array of samples can be decomposed into q lower–dimensional (i.e. computationally less demanding) sub–problems.

The decomposition procedure (see Fig. 4) can be summarized as follows:

1. Find regions $\mathcal{F}_1, \ldots, \mathcal{F}_q$ satisfying conditions (44) and (45) ; note that even though the sets $\mathcal{F}_i$, $i = 1, \ldots, q$ must be exclusive, the sets $N(\mathcal{F}_i)$, $i = 1, \ldots, q$ may partially overlap.
2. Determine minimal rectangular arrays $\Omega_1, \ldots, \Omega_q \subset \Omega$ which contain Markovian neighborhoods $N(\mathcal{F}_1), \ldots, N(\mathcal{F}_q)$.
3. Solve the reconstruction problem for each of the rectangular sub–arrays of Ω; since after the decomposition some of the dummy samples in sub–arrays may turn out to be unknown (cf. Fig. 4) it is essential to use the reduced–order formula (43) instead of (23).

Separability of the reconstruction problem can also be easily explained in terms of some structural properties of the projection matrix. Since Π is a band matrix, it is possible to show that the simultaneous deletion of its $d_{\mathcal{F}_i}$ rows and columns indicated by any neighborhood set $\mathcal{F}_i$ results in a matrix which is two–block diagonal. Repeating this argument q times we find out that the matrix $\Pi_{(\mathcal{K}_0|\mathcal{K}_0)}$ (and hence also its inverse appearing in (23) and (43)) is q–block diagonal which leads in a natural way to q lower–dimensional reconstruction problems.

3.6. Differences with Respect to the 1–D Case

As we have already stressed, there are some important differences between the 1–D and 2–D reconstruction (even though solutions to both problems look similarly)

1. Since 2–D signals do not have the autocorrelation matching property the "classic" (i.e. correlation–based) solution to the reconstruction problem (7) has a very limited practical value. The same remark applies to any approach (such as reconstruction based on ARMA models) which requires "translation" of spatial coefficients to autocorrelation coefficients.

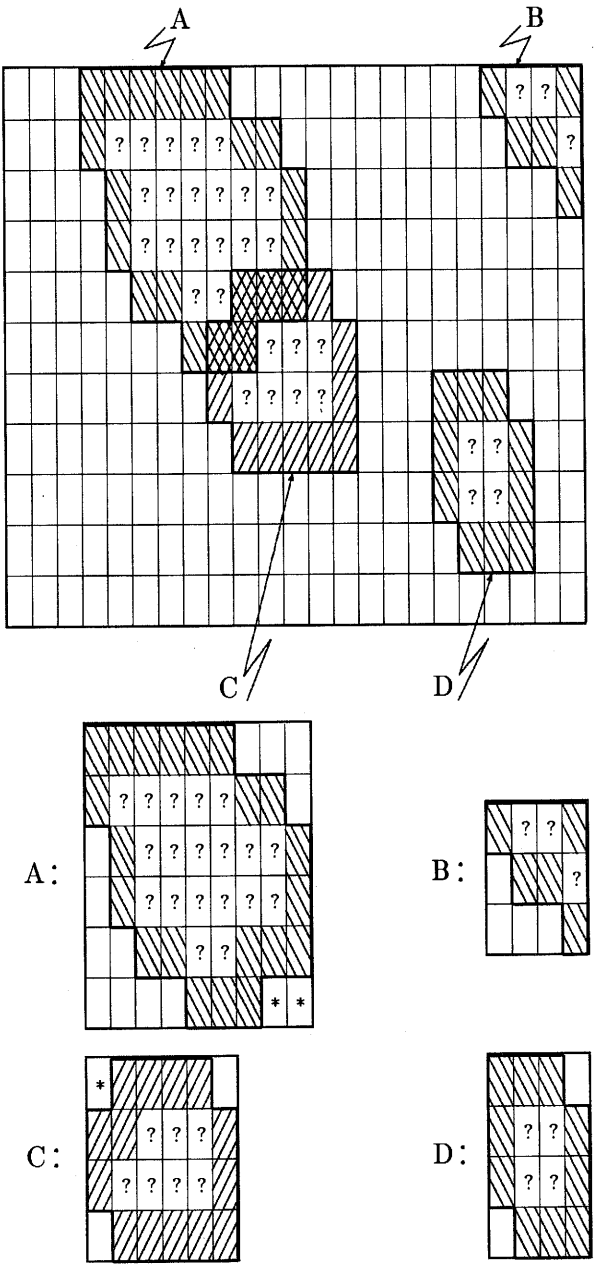

Fig. 4. An example of decomposition of the reconstruction problem into four lower–dimensional sub–problems A, B, C and D. Shaded areas surrounding blocks of unknown samples denote the corresponding Markovian neighborhoods. Empty cells denote dummy samples. Those dummy samples located in sub–arrays which turn out to be unknown are marked with asterisks.

2. Boundary problems are much more emphasized in the 2–D case. In particular, for a chosen spatial model solvability of the reconstruction problem depends on the pattern of missing samples — if any of the missing samples is located in the (model–dependent) critical boundary zone, then no closed–form solution to the problem can be derived.
3. In the vector case considered in this paper the projection matrix is specified in terms of both "normal" and "inverse" CM matrices. The fact that in the general case both representations have to be used simultaneously is a unique feature of 2–D reconstruction.

3.7. Adaptive Reconstruction

The formulae derived in section 3 are based on the assumption that all coefficient matrices of the CM model (8) and/or of the "inverse" CM model (31) are known. If not, the following adaptive version of the reconstruction procedure can be used

1. Estimate coefficient matrices $A_1, \ldots, A_p, \rho$ and/or $A_1^\star, \ldots, A_p^\star, \rho^\star$ from the set of available data $\mathcal{Y}_0$.
2. Apply the non–adaptive reconstruction procedure after replacing all unknown matrices with their sample estimates.

The complete analysis of the first step of the algorithm summarized above is a non-trivial problem which falls beyond the scope of this paper. Basically, a number of different identification techniques can be adopted for the purpose of estimation of coefficient matrices in (8) (along with some related structural parameters such as model orders and neighbor locations) — see e.g. (Azimi–Sadjadi, 1991; Chellappa and Kashyap, 1982; Kashyap, 1983).

Perhaps the simplest solution is to use the estimates

$$\widehat{\mathcal{A}} = \left\{\widehat{A}_1, \ldots, \widehat{A}_p\right\} = \arg\min_{\mathcal{A}} \sum_{r \in \mathcal{K}_E} \| \varepsilon_{\mathcal{A}}(r) \|^2$$

$$\widehat{\mathcal{A}}^\star = \left\{\widehat{A}_1^\star, \ldots, \widehat{A}_p^\star\right\} = \arg\min_{\mathcal{A}^\star} \sum_{r \in \mathcal{K}_E^\star} \| \varepsilon_{\mathcal{A}^\star}(r) \|^2 \tag{46}$$

$$\widehat{\rho} = \frac{1}{d_E} \sum_{r \in \mathcal{K}_E} \varepsilon_{\widehat{\mathcal{A}}}(r) \varepsilon_{\widehat{\mathcal{A}}}^T(r)$$

$$\widehat{\rho}^\star = \frac{1}{d_E^\star} \sum_{r \in \mathcal{K}_E^\star} \varepsilon_{\widehat{\mathcal{A}}^\star}(r) \varepsilon_{\widehat{\mathcal{A}}^\star}^T(r) \tag{47}$$

where $\mathcal{K}_E \subset \mathcal{K}_0$ and $\mathcal{K}_E^\star \subset \mathcal{K}_0$ denote the sets of all points at which the residuals

$$\varepsilon_{\mathcal{A}}(r) = x(r) - \sum_{i=1}^{p} A_i x(r - r_i)$$

$$\varepsilon_{\mathcal{A}^\star}(r) = x(r) - \sum_{i=1}^{p} A_i^\star x(r + r_i)$$

can be computed and d_E, $d_E^\star$ denote numbers of such points (after vectorization (46) becomes a standard least squares problem).

The approach above can be expected to work quite well for low–order models or in the case where the missing samples can form a few clusters. In all other cases d_E $(d_E^\star)$, the number of samples available for identification based on (46), may turn out to be too small. In particular, one can show such maliciously chosen patterns of missing samples for which the set $\mathcal{K}_E$ $(\mathcal{K}_E^\star)$ is *empty* even if $d_M \ll d_O$ (it is possible to remove $[N_1N_2/p]$ elements from the $N_1 \times N_2$ rectangular array in such a way — the construction depends on the topology of the set of supporting pixels P — that $N_r \not\subset \mathcal{K}_O$ for every possible location of $r \in \mathcal{K}$!).
In all such cases, including the pathological ones indicated above, some special estimation techniques developed for the purpose of identification of non–uniformly sampled random fields should be applied (see Jones (1980) for the solution of the related problem for unevenly sampled time series). The more detailed discussion of this issue is left for further studies.

Remark 6. In the scalar case where the "normal" and "inverse" model coefficients coincide one can use the following bidirectional estimates

$$\widehat{\mathcal{A}} = \left\{\widehat{a}_1, \ldots, \widehat{a}_p\right\} = \arg\min_{\mathcal{A}} \Big(\sum_{r \in \mathcal{K}_E} \varepsilon^2_{\mathcal{A}}(r) + \sum_{r \in \mathcal{K}_E^\star} \varepsilon^2_{\mathcal{A}^\star}(r) \Big)$$

$$\widehat{\rho} = \frac{1}{d_E + d_E^\star} \Big(\sum_{r \in \mathcal{K}_E} \varepsilon^2_{\widehat{\mathcal{A}}}(r) + \sum_{r \in \mathcal{K}_E^\star} \varepsilon^2_{\widehat{\mathcal{A}}^\star}(r) \Big)$$

4. Example

Consider the task of reconstruction of the incomplete image (array) shown in Figure 5(a). In order to keep the dimensionality of the problem low it is assumed that there are only four missing samples: three located in the upper right corner of the image and one in the central area. For the same reason a simple three–point AR model (38) is adopted.

First of all we note that using the rules described in section 3.5, the problem can be decomposed into two sub–problems each of which can be embedded (after element renumbering) in the same 3×3 subarray shown in Figure 5(b). Furthermore, in the case under consideration we have (for both sub–problems)

$$\mathcal{K} = \{1,2,3,4,5,6,7,8,9\}$$

$$\mathcal{K}_B = \{1,2,3,4,7\}$$

$$\mathcal{K}_B^\star = \{3,6,7,8,9\}$$

$$\mathcal{K}_C = \{3,7\}$$

and

$$\mathcal{C} = \begin{bmatrix} C_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ C_1 & C_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_1 & C_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ C_2 & 0 & C_1 & C_0 & 0 & 0 & 0 & 0 & 0 \\ C_3 & C_2 & 0 & C_1 & C_0 & 0 & 0 & 0 & 0 \\ 0 & C_3 & C_2 & 0 & C_1 & C_0 & 0 & 0 & 0 \\ 0 & 0 & C_3 & C_2 & 0 & C_1 & C_0 & 0 & 0 \\ 0 & 0 & 0 & C_3 & C_2 & 0 & C_1 & C_0 & 0 \\ 0 & 0 & 0 & 0 & C_3 & C_2 & 0 & C_1 & C_0 \end{bmatrix}$$

$$\mathcal{C}^\star = \begin{bmatrix} C_0^\star & C_1^\star & 0 & C_2^\star & C_3^\star & 0 & 0 & 0 & 0 \\ 0 & C_0^\star & C_1^\star & 0 & C_2^\star & C_3^\star & 0 & 0 & 0 \\ 0 & 0 & C_0^\star & C_1^\star & 0 & C_2^\star & C_3^\star & 0 & 0 \\ 0 & 0 & 0 & C_0^\star & C_1^\star & 0 & C_2^\star & C_3^\star & 0 \\ 0 & 0 & 0 & 0 & C_0^\star & C_1^\star & 0 & C_2^\star & C_3^\star \\ 0 & 0 & 0 & 0 & 0 & C_0^\star & C_1^\star & 0 & C_2^\star \\ 0 & 0 & 0 & 0 & 0 & 0 & C_0^\star & C_1^\star & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & C_0^\star & C_1^\star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & C_0^\star \end{bmatrix}$$

Consequently (cf. (35))

$$\Pi_{\langle \mathcal{K}_B \rangle} = \begin{bmatrix} \Pi_{51} & \Pi_{52} & \Pi_{53} & \Pi_{54} & \Pi_{55} & \Pi_{56} & \Pi_{57} & \Pi_{58} & \Pi_{59} \\ \Pi_{61} & \Pi_{62} & \Pi_{63} & \Pi_{64} & \Pi_{65} & \Pi_{66} & \Pi_{67} & \Pi_{68} & \Pi_{69} \\ \Pi_{81} & \Pi_{82} & \Pi_{83} & \Pi_{84} & \Pi_{85} & \Pi_{86} & \Pi_{87} & \Pi_{88} & \Pi_{89} \\ \Pi_{91} & \Pi_{92} & \Pi_{93} & \Pi_{94} & \Pi_{95} & \Pi_{96} & \Pi_{97} & \Pi_{98} & \Pi_{99} \end{bmatrix}$$

$$= \begin{bmatrix} C_0^T & C_1^T & C_2^T & C_3^T \\ 0 & C_0^T & 0 & C_2^T \\ 0 & 0 & C_0^T & C_1^T \\ 0 & 0 & 0 & C_0^T \end{bmatrix} \begin{bmatrix} C_3 & C_2 & 0 & C_1 & C_0 & 0 & 0 & 0 & 0 \\ 0 & C_3 & C_2 & 0 & C_1 & C_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & C_3 & C_2 & 0 & C_1 & C_0 & 0 \\ 0 & 0 & 0 & 0 & C_3 & C_2 & 0 & C_1 & C_0 \end{bmatrix} \tag{48}$$

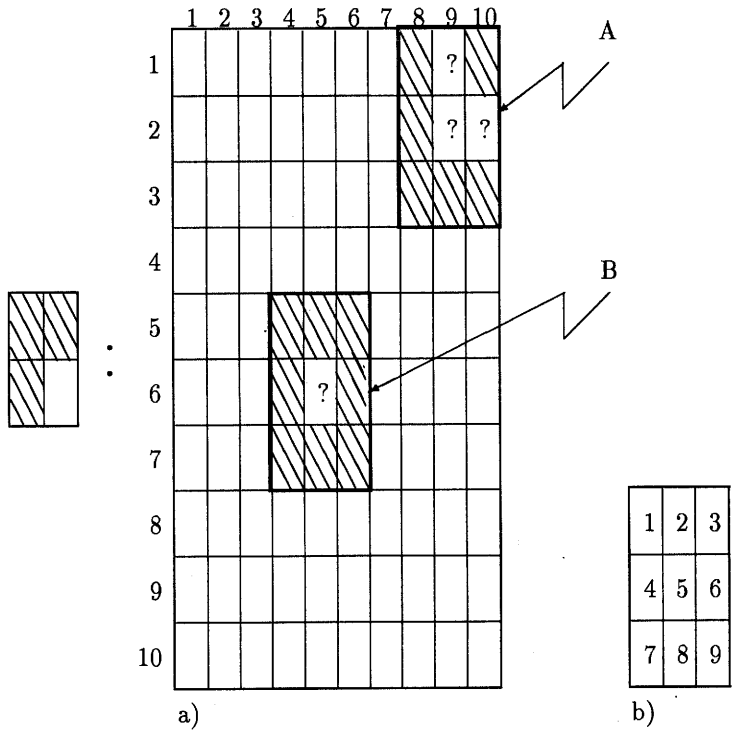

Fig. 5. A simple reconstruction problem (a) which can be decomposed into two lower–dimensional sub–problems A and B; each of the sub–problems can be embedded in the same 3×3 array (b).

$$\Pi_{(\mathcal{K}_B^\star|} = \begin{bmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} & \Pi_{14} & \Pi_{15} & \Pi_{16} & \Pi_{17} & \Pi_{18} & \Pi_{19} \\ \Pi_{21} & \Pi_{22} & \Pi_{23} & \Pi_{24} & \Pi_{25} & \Pi_{26} & \Pi_{27} & \Pi_{28} & \Pi_{29} \\ \Pi_{41} & \Pi_{42} & \Pi_{43} & \Pi_{44} & \Pi_{45} & \Pi_{46} & \Pi_{47} & \Pi_{48} & \Pi_{49} \\ \Pi_{51} & \Pi_{52} & \Pi_{53} & \Pi_{54} & \Pi_{55} & \Pi_{56} & \Pi_{57} & \Pi_{58} & \Pi_{59} \end{bmatrix}$$

$$= \begin{bmatrix} (C_0^\star)^T & 0 & 0 & 0 \\ (C_1^\star)^T & (C_0^\star)^T & 0 & 0 \\ (C_2^\star)^T & 0 & (C_0^\star)^T & 0 \\ (C_3^\star)^T & (C_2^\star)^T & (C_1^\star)^T & (C_0^\star)^T \end{bmatrix} \begin{bmatrix} C_0^\star & C_1^\star & 0 & C_2^\star & C_3^\star & 0 & 0 & 0 & 0 \\ 0 & C_0^\star & C_1^\star & 0 & C_2^\star & C_3^\star & 0 & 0 & 0 \\ 0 & 0 & 0 & C_0^\star & C_1^\star & 0 & C_2^\star & C_3^\star & 0 \\ 0 & 0 & 0 & 0 & C_0^\star & C_1^\star & 0 & C_2^\star & C_3^\star \end{bmatrix} \tag{49}$$

Based on (35) and (36) the symmetric 9×9 matrix Π can be determined up to four elements associated with the critical boundary zone (Π_{33}, Π_{37}, Π_{73}, Π_{77}) — see Figure 6.

	1	2	3	4	5	6	7	8	9
1	−	−	−	−	±	±	−	±	±
2	−	−	−	−	±	±	−	±	±
3	−	−	?	−	±	+	?	+	+
4	−	−	−	−	±	±	−	±	±
5	±	±	±	±	±	±	±	±	±
6	±	±	+	±	±	+	+	+	+
7	−	−	?	−	±	+	?	+	+
8	±	±	+	±	±	+	+	+	+
9	±	±	+	±	±	+	+	+	+

Fig. 6. Structure of the matrix Π: the '+' sign denotes elements which can be expressed in terms of CM matrices $C_0, \ldots, C_3$; the '−' sign denotes elements which can be expressed in terms of "inverse" CM matrices $C_0^\star, \ldots, C_3^\star$ (those elements which can be characterized in both ways are marked with '±'); elements associated with the critical boundary zone are identified with questionmarks.

We are ready now to solve the two sub–problems described earlier

Sub–problem A (reconstruction of corner elements). We note that

$$\mathcal{K}_M = \{2,5,6\}$$

$$\mathcal{K}_O = \{1,3,4,7,8,9\}$$

Consequently, after changing the local numbering (within the subarray) to the global one (corresponding to the original image), one gets

$$\begin{bmatrix} \widehat{y}(1,9) \\ \widehat{y}(2,9) \\ \widehat{y}(2,10) \end{bmatrix} = - \begin{bmatrix} \Pi_{22} & \Pi_{25} & \Pi_{26} \\ \Pi_{52} & \Pi_{55} & \Pi_{56} \\ \Pi_{62} & \Pi_{65} & \Pi_{66} \end{bmatrix}^{-1} \begin{bmatrix} \Pi_{21} & \Pi_{23} & \Pi_{24} & \Pi_{27} & \Pi_{28} & \Pi_{29} \\ \Pi_{51} & \Pi_{53} & \Pi_{54} & \Pi_{57} & \Pi_{58} & \Pi_{59} \\ \Pi_{61} & \Pi_{63} & \Pi_{64} & \Pi_{67} & \Pi_{68} & \Pi_{69} \end{bmatrix} \begin{bmatrix} y(1,8) \\ y(1,10) \\ y(2,8) \\ y(3,8) \\ y(3,9) \\ y(3,10) \end{bmatrix}$$

where (cf. (48)–(49)):

$$\Pi_{21} = (C_1^\star)^T C_0^\star$$

$$\Pi_{22} = (C_1^\star)^T C_1^\star + (C_0^\star)^T C_0^\star$$

$$\Pi_{23} = (C_0^\star)^T C_1^\star$$

$$\Pi_{24} = (C_1^\star)^T C_2^\star$$

$$\Pi_{25} = (C_1^\star)^T C_3^\star + (C_0^\star)^T C_2^\star = C_3^T C_1 + C_2^T C_0 = \Pi_{52}^T$$

$$\Pi_{26} = (C_0^\star)^T C_3^\star = C_3^T C_0 = \Pi_{62}^T$$

$$\Pi_{27} = \Pi_{28} = \Pi_{29} = 0$$

$$\Pi_{51} = (C_3^\star)^T C_0^\star = C_0^T C_3 = \Pi_{59}^T$$

$$\Pi_{53} = (C_2^\star)^T C_1^\star = C_1^T C_2 = \Pi_{57}^T$$

$$\Pi_{54} = (C_3^\star)^T C_2^\star + (C_1^\star)^T C_0^\star = C_2^T C_3 + C_0^T C_1$$

$$\begin{aligned}\Pi_{55} &= (C_3^\star)^T C_3^\star + (C_2^\star)^T C_2^\star + (C_1^\star)^T C_1^\star + (C_0^\star)^T C_0^\star \\ &= C_3^T C_3 + C_2^T C_2 + C_1^T C_1 + C_0^T C_0\end{aligned}$$

$$\Pi_{56} = (C_2^\star)^T C_3^\star + (C_0^\star)^T C_1^\star = C_1^T C_0 + C_3^T C_2 = \Pi_{65}^T$$

$$\Pi_{58} = \Pi_{25}$$

$$\Pi_{61} = \Pi_{64} = \Pi_{67} = 0$$

$$\Pi_{63} = C_0^T C_2$$

$$\Pi_{66} = C_0^T C_0 + C_2^T C_2$$

$$\Pi_{68} = C_2^T C_1$$

$$\Pi_{69} = C_2^T C_0$$

Sub–problem B (reconstruction of the central element). We have

$$\mathcal{K}_M = \{5\}$$

$$\mathcal{K}_O = \{1,2,3,4,6,7,8,9\}$$

leading to

$$\widehat{y}(6,5) = -\Pi_{55}^{-1} \begin{bmatrix} \Pi_{51} & \Pi_{52} & \Pi_{53} & \Pi_{54} & \Pi_{56} & \Pi_{57} & \Pi_{58} & \Pi_{59} \end{bmatrix} \begin{bmatrix} y(5,4) \\ y(5,5) \\ y(5,6) \\ y(6,4) \\ y(6,6) \\ y(7,4) \\ y(7,5) \\ y(7,6) \end{bmatrix}$$

where the matrices Π_{55} and $\Pi_{51} = \Pi_{59}^T$, $\Pi_{52} = \Pi_{58}^T$, $\Pi_{53} = \Pi_{57}^T$, $\Pi_{54} = \Pi_{56}^T$ were already specified in part A.

5. Conclusions

We have shown that the optimal reconstruction of an incomplete random field consists of two components: the unique one which can be obtained as a result of orthogonal projection of known samples on the space spanned by unknown samples, and the nonunique (artificially generated) one added to remove the "covariance defect" introduced by the first component.

The closed–form solution to the problem was derived for fields described by conditional Markov models. Finally, the possibility of decomposing the restoration procedure into a number of computationally less demanding sub–problems was discussed.

References

Azimi–Sadjadi M.R. (1991): *Estimation and identification for 2–D block Kalman filtering.* — IEEE Trans. Signal Processing, v.39, No.8, pp.1885–1889.

Chellappa R. and R.L. Kashyap (1982): *Digital image restoration using spatial interaction models.* — IEEE Trans. Acoustics Speech and Signal Processing, v.30, No.3, pp.461–472.

Jones R.H. (1980): *Maximum likelihood fitting of ARMA models to time series with missing observations.* — Technometrics, v.22, No.4, pp.389–395.

Kashyap R.L. and Chellappa R. (1983): *Estimation and choice of neighbors in spatial–interaction models of images.* — IEEE Trans. Information Theory, v.29, No.1, pp.60–72.

Marple Jr.S.L. (1988): *Digital Spectral Analysis.* — Englewood Cliffs, New Jersey: Prentice–Hall.

Marzetta T.L. (1988): *Two–dimensional linear prediction: autocorrelation arrays, minimum–phase prediction error filters and reflection coefficient arrays.* — IEEE Trans. Acoustics Speech and Signal Processing, v.28, No.6, pp.725–733.

Niedźwiecki M. (1993): *Statistical reconstruction of multivariate time series.* — IEEE Trans. Signal Processing, v.41, No.1, pp.451–457.

Schweppe F.C. (1973): *Uncertain Dynamic Systems.* — Englewood Cliffs, New Jersey: Prentice–Hall.

Received December 3, 1993