A NEW MODEL FOR BIOLOGY

This chapter describes a mathematical model of epigenesis, and starts by translating the traditional but vague definitions of that concept into expressions that are increasingly more precise. The first step consists in defining epigenesis as *a convergent increase of complexity*. The second states that this process is equivalent to *a reconstruction from incomplete information*, and in the third step this becomes *a reconstruction from incomplete projections*. In this way, we can model epigenesis as a special case of the problem of reconstructing structures from projections, a problem that arises in many fields (for example in computerised tomography) and whose mathematics is well known. What is less well known is that a reconstruction can be achieved even when the starting information is incomplete, provided that appropriate memories and codes are employed. This is illustrated with a few practical examples, and the logic of that unusual kind of reconstruction is described first in words and then in formulae. At that stage we can go back to biological epigenesis and conclude that a convergent increase of complexity in organic life necessarily requires organic codes and organic memories. And this gives us the two critical concepts that will be used in the rest of the book for an entirely new approach to the problem of biological complexity.

The logic of embryonic development

The discovery of genes that control embryonic development has started a true revolution in biology, both from an experimental and from a theoretical point of view. On the experimental side, it has opened fields of research that previously seemed unapproachable.

From a theoretical point of view, it has inspired the conclusion that embryonic development is the execution of a genetic program, in the sense that all processes of ontogenesis depend, more or less indirectly, on the transcription of genes. Unfortunately, many have also concluded that the central problem of development – the problem of form – has been, *in principle*, resolved. Many details are still to be worked out, it is said, but the "logic" is now clear because the form of an organism depends on its genes.

In his book *The Problems of Biology* (1986), John Maynard Smith has lucidly sounded a note of caution against this attitude:
"It is popular nowadays to say that morphogenesis (that is the development of form) is programmed by the genes. I think that this statement, although in a sense true, is unhelpful. Unless we understand how the program works, the statement gives us a false impression that we understand something when we do not … One reason why we find it so hard to understand the development of form may be that we do not make machines that develop: often we understand biological phenomena only when we have invented machines with similar properties … and we do not make 'embryo' machines."

Maynard Smith's point can also be expressed in another way: embryonic development is a process that increases the complexity of a living system, but we do not know how to build machines that increase their own complexity, and we cannot therefore understand the logic of development. We can also leave aside the physical construction of machines and concern ourselves only with their planning. If we could prove, with a mathematical model, that it is possible to increase the complexity of a system, we already would have taken a major step forward. The search for the logic of development begins therefore with the search of a mathematical model for systems which are capable of increasing their own complexity.

At this point, however, a formal distinction between two very different cases is called for. An increase in complexity took place even during the history of life, but in this case new structures arose by chance mutations, and the increase was a *divergent* process. In embryonic development, on the contrary, new structures are never formed by chance, and we are dealing with a *convergent* increase of

complexity. This is the great difference between evolution and ontogenesis, and such a dichotomy does require two very different types of mathematical models.

In the case of evolution, we already have algorithms that simulate the effects of natural selection, and we do therefore understand how a divergent increase of complexity can take place. But we do not have algorithms that describe a convergent increase, and it is for this reason that the logic of the embryos still eludes us. The real key to embryonic development is the logic of systems which are capable of increasing their complexity *in a convergent way*, and in order to understand this we need, if not a machine, at least a model that is functioning according to that logic.

Reconstruction from incomplete projections

The starting-point for a new model of embryonic development is the reconstruction of structures from their projections, a problem which arises in many fields such as computerised tomography and electron microscopy. The image produced by X-rays on a radiographic plate, for example, is a projection of a three-dimensional body on a two-dimensional surface, and this process is inevitably accompanied by a loss of information. The result, to quote Hounsfield (1972), is *"like having a whole book projected on a single sheet of paper, so that the information of any one page cannot be extracted from the superimposed information of all the other pages"*. In order to reconstruct the original structure, therefore, it is necessary to collect a plurality of projections at different angles, as shown in Figure 3.1. The minimum number of projections that must be collected is known from basic theorems, and has an clear intuitive meaning. The projections taken at different angles carry different information, and their totality must contain (in a compressed form) all the information that was present in the original structure.

As for the reconstruction algorithms, we can divide them in two major groups: iterative and non-iterative techniques. A non-iterative method produces the final result with a formula which is applied only

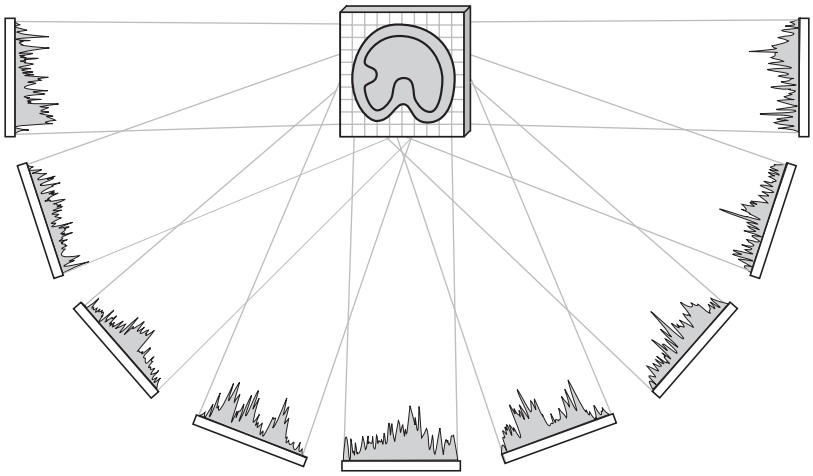

Figure 3.1 During the projection of a three-dimensional structure onto a two-dimensional plane, information is lost, and it is therefore necessary to collect a plurality of projections at different angles in order to reconstruct the original structure.

once to the experimental data. In this case the reconstruction is precise, because the formula provides a rigorous solution, but the procedure is cumbersome because all data are processed together. The iterative algorithms have been introduced precisely in order to simplify the reconstruction procedure and still obtain satisfactory results. In these cases, a reconstruction produces only an approximation of the original structure, and it is therefore necessary to repeat the operations many times in order to get progressively closer to the original structure.

Iterative algorithms are clearly less precise than single-application techniques, but their great advantage is that they introduce the *time dimension* in the computation, and this makes them particularly suitable to simulate biological processes. Even more important is the fact that the temporal dimension allows us to reconsider the problem of the minimum number of projections that are required for a complete reconstruction. During an iterative procedure, we could discover properties of the original structures that were not recorded in the projections, and in this case we could obtain a reconstruction even if

the number of projections is appreciably lower than the theoretical minimum. This brings us face to face with an entirely new problem: *the problem of reconstructing structures from incomplete projections*, where projections are said to be *incomplete* when their number is at least one order of magnitude less than the theoretical minimum. The problem, in other words, is to make a complete reconstruction with an amount of information which is much lower than that of the original structure.

The interesting point is that this is a mathematical version of the problem that we face in embryonic development. The fertilised egg contains far less information than the adult organism (whatever criterion is used to measure information in biological systems), and embryonic development can be described therefore as a process that is reconstructing a structure from *incomplete information*. This is another way of saying that embryonic development is a process that increases the complexity of a living system. The reconstruction of structures from incomplete information, in short, is a model that could help us understand how it is possible for a system to obtain a convergent increase of complexity.

A memory-building approach

An iterative reconstruction algorithm produces a series of pictures which are increasingly more accurate approximations of the original structure. Any reconstructed picture is affected by errors, and in general there is no way of knowing where the errors are falling, but there are two outstanding exceptions to this rule. The values which are below the minimum or above the maximum are clearly "illegal", and the algorithm gives us the precise coordinates of the points where they appear. This makes it possible to correct those errors by setting to the minimum or to the maximum all values which are respectively below or above the legal limits. This operation (which is called a *reconstruction constraint*) does improve the results and so it is normally applied at regular intervals. Every time that we apply this constraint, however, we lose some information, because the coordinates of the

illegal values are lost. Keeping information about errors may not seem important, but let us assume that it could be, and let us see if we can save it. This can be done by using a "memory" picture where we store the coordinates and the values of the illegal points before applying the constraint. In this case we need a more complex algorithm, because we have to perform in parallel two different reconstruction: one for the structure and one for the memory, as illustrated in Figure 3.2. But what is the point of keeping a memory of the reconstruction errors?

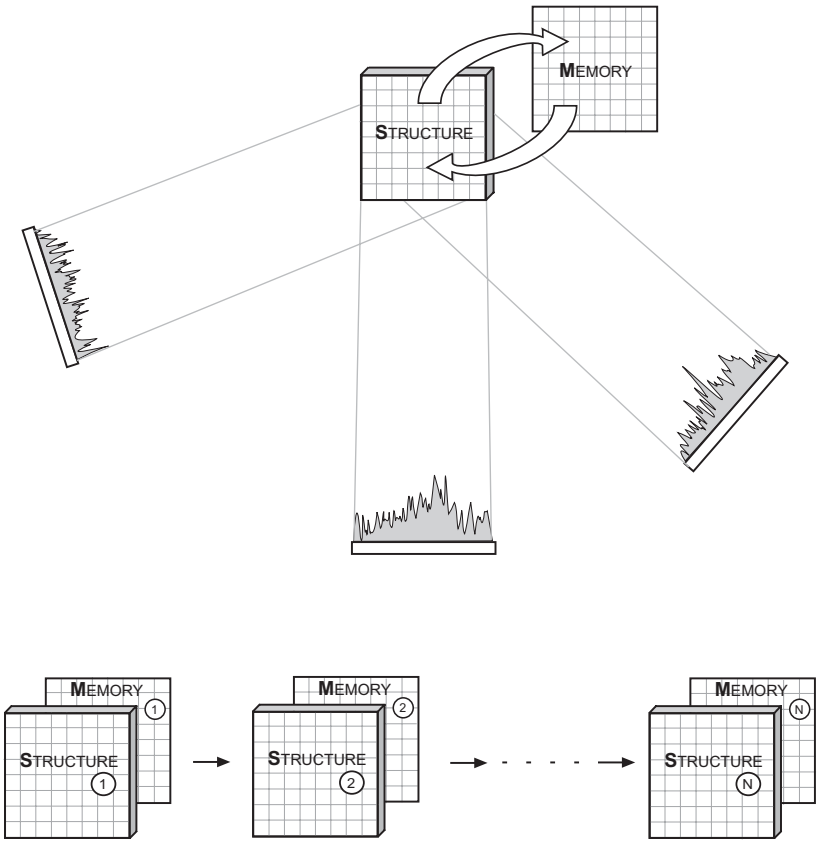

Figure 3.2 A reconstruction from incomplete projections is a method where structure matrices and memory matrices are reconstructed in parallel.

The point is that we can study their pattern, and that turns out to have unexpected features. Since the errors are random events, we would expect a statistical distribution, but this is not what happens in all cases. It is true that in many points the errors are totally random, but there invariably are other points where this does not happen. In those points the illegal values keep reappearing each time, and always with the same sign, which explains why such points have been called *vortices*. Figure 3.3 is a schematic illustration of what happens. The patterns of the illegal values look totally random when they are examined one by one (Figure 3.3A), but when they are memorised together (Figure 3.3B) the statistical fluctuations disappear and only the vortices stand up. Now we have a new type of information before us. When an illegal value has consistently reappeared in the same point for a number of times (we can choose 5, 10 or any other convenient number), we can reasonably conclude that the value of that point is either a minimum or a maximum. We can therefore "fix" the value of that point, and this means that the total number of the unknowns is reduced by one. By repeating the operation, the number of the unknowns becomes progressively smaller, and when it reaches the number of the equations a complete reconstruction is possible. That is the result we were looking for. With appropriate "tools" we can indeed obtain a complete reconstruction of the original structure from incomplete information.

Let us now take a closer look at those "tools". One is the memory picture that must be reconstructed in parallel with the structure, but that is not all. The new information of the vortices appear in the memory space, but we use that information in the structure space, because it is here that we reduce the number of the unknowns. We are in fact transferring information from the memory space to the structure space with a *conventional rule* of the type *"If a vortex appears in the memory space, fix the corresponding point in the structure space to a minimum or a maximum."* A reconstruction from incomplete information, in short, does not require only a memory. It requires memory and codes. The reconstruction memory is where new information appears. The reconstruction codes are the tools that transfer information from the memory to the structure.

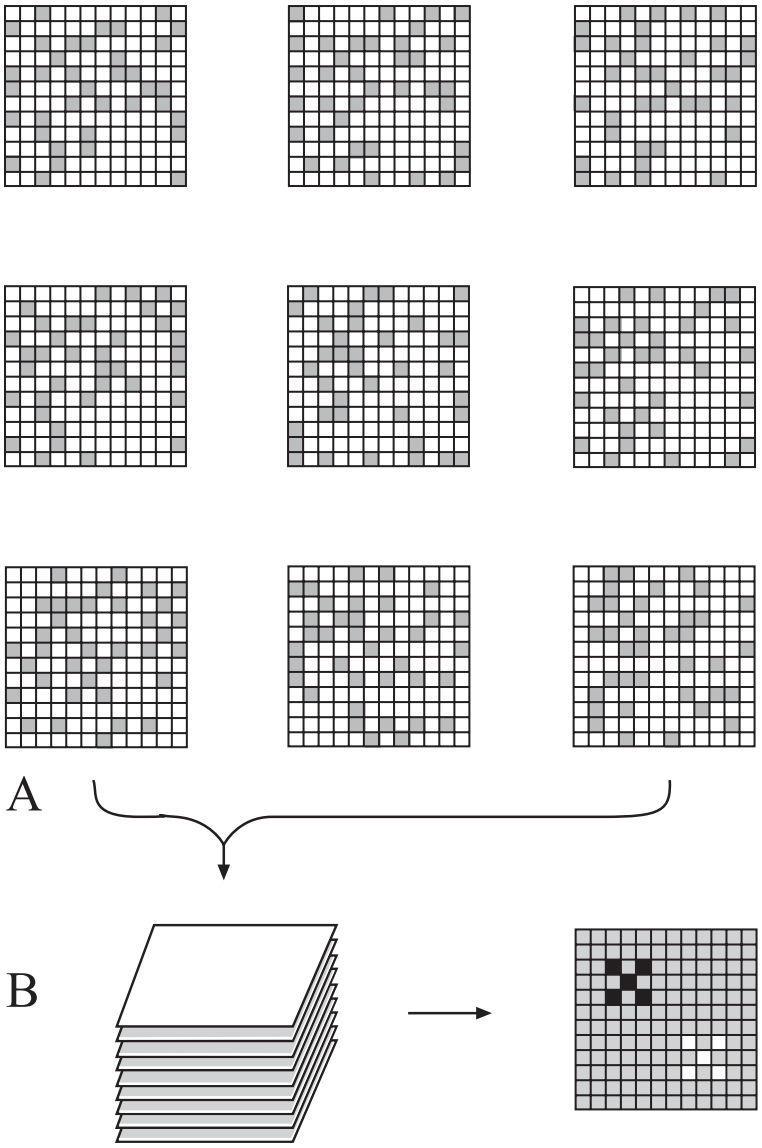

Figure 3.3 The errors produced by an iterative reconstruction algorithm have patterns which appear, at each iteration, completely random (A), but if successive patterns are memorised together, it is possible to observe regular structures appearing in the memory matrix (B).

The biological implications of the above model are straightforward. Embryonic development is also a reconstruction of structures from incomplete information, and so it must employ organic memories and organic codes. Before looking for the presence of these biological tools in nature, however, we must examine the mathematics of the new reconstruction method. Today there is a widespread belief that a convergent increase of information is impossible (despite the evidence from the embryos) and only mathematics can give us the proof of the contrary.

The algebraic method

The simplest case is the reconstruction of two-dimensional structures from one-dimensional projections. A digitised two-dimensional structure, for example a television picture, can be described as an $n \cdot n$ matrix $[f_{ij}]$ of side D and cells (i,j) of side $d = D/n$ (Figure 3.4). A projection of the picture at an angle ϑ is a set of parallel rays (ϑ,k) which totally cover the picture at the angle ϑ, and any projection ray can be represented by an $n \cdot n$ matrix (Figure 3.5) where each element $a_{ij}^{\vartheta k}$ is the fraction of the cell (i,j) which is contained within the ray (ϑ,k). The picture matrix and the ray-matrices are easily transformed into vectors (Figure 3.6). More precisely, the picture matrix $[f_{ij}]$ is replaced by a column-vector $[f_z]$, and the ray matrices $[a_{ij}^{\vartheta k}]$ are described by row-vectors $[a_z^{\vartheta k}]$ with the transformations:

$$f_{ij} \longrightarrow f_z \quad \text{and} \quad a_{ij}^{\vartheta k} \longrightarrow a_z^{\vartheta k} \quad \text{with} \quad z = 1, \ldots\ldots\ldots, n^2 = t$$

In this way, the projection values $g_{\vartheta k}$ of any ray (ϑ,k) are described by the scalar product of the vectors $[a_z^{\vartheta k}]$ and $[f_z]$:

$$g_{\vartheta k} = a_1^{\vartheta k} f_1 + a_2^{\vartheta k} f_2 + \ldots\ldots\ldots + a_t^{\vartheta k} f_t$$

which is a linear equation with $t = n^2$ unknowns.

If we have p projections of a picture and each projection contains r rays, we have a system of $p \cdot r$ equations in n^2 unknowns, and a solution

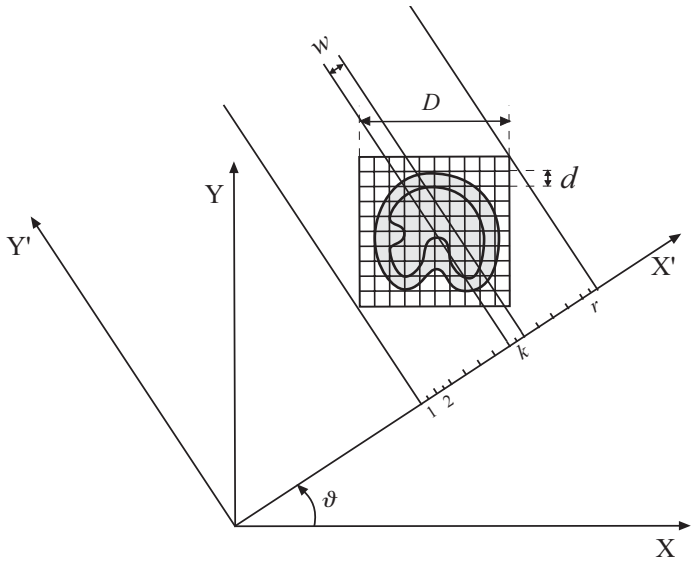

Figure 3.4 A digitized two-dimensional structure can be represented by a matrix of side D which is made of $n \cdot n$ cells, or pixels, of side d. The projection of a picture can be represented by a set of adjacent parallel rays, of equal width w, which totally covers the matrix of the picture.

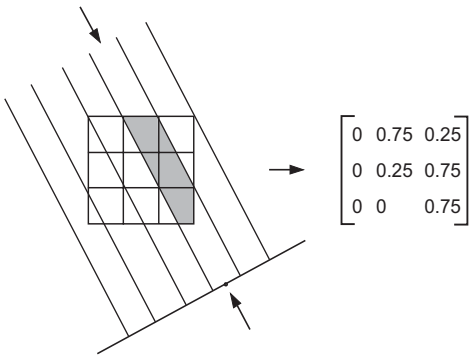

Figure 3.5 A projection ray that crosses an $n \cdot n$ picture matrix can also be represented by an $n \cdot n$ matrix, where each cell (i,j) of the ray matrix contains a number that represents the fraction of the cell (i,j) of the picture matrix which is contained within the projection ray.

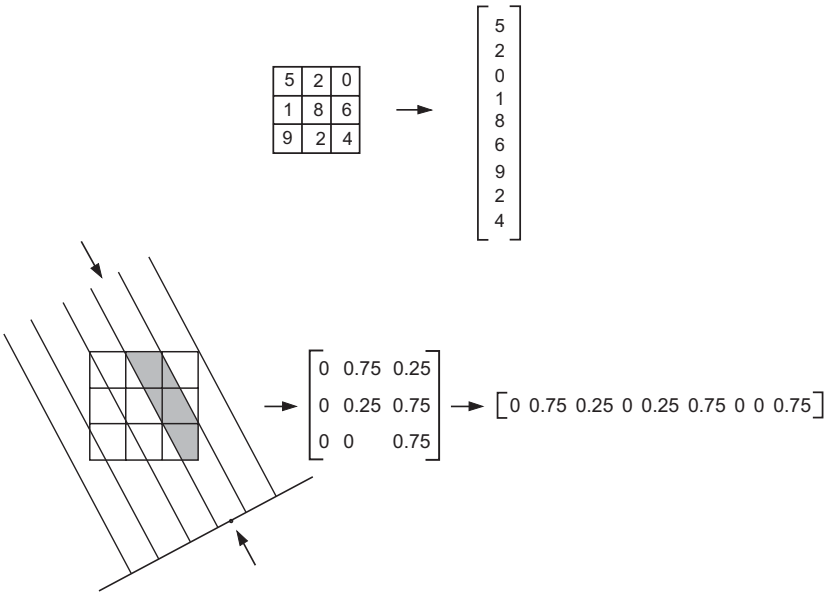

Figure 3.6 A picture matrix and a ray matrix can both be represented by vectors. More precisely, by a column vector for the picture matrix, and by a row vector for the ray matrix.

exists if the number of linearly independent equations is equal to the number of unknowns, i.e. if

$$p \cdot r = n^2$$

In order to have all equations in a compact form, the double index (ϑ,k) of each ray is replaced by the single index (h) with the transformations

$$g_{\vartheta k} \longrightarrow g_h \quad \text{and} \quad a_z^{\vartheta k} \longrightarrow a_{hz} \quad \text{with} \quad h = 1, \ldots\ldots, p \cdot r = t$$

In this way, all the projections of a picture are represented by a single column vector $[g_h]$, and the geometrical parameters form a matrix $[a_{hz}]$, known as the *weighting factors matrix*, which has $p \cdot r = t$ rows

and $n^2=t$ columns, i.e. a matrix of $t^2=n^4$ cells. All projection equations of a picture can be represented therefore by a single matrix equation:

$$\begin{bmatrix} a_{11} \ldots\ldots\ldots\ldots a_{1t} \\ \\ \\ a_{t1} \ldots\ldots\ldots\ldots a_{tt} \end{bmatrix} \begin{bmatrix} f_1 \\ \vdots \\ f_t \end{bmatrix} = \begin{bmatrix} g_1 \\ \vdots \\ g_t \end{bmatrix} \qquad (3.1)$$

A reconstruction is a procedure which reverses the projection process, and the reconstruction equations can therefore be obtained from equation 3.1 with a matrix $[b_{hz}]$ which represents the *inverse weighting factors matrix*:

$$\begin{bmatrix} f_1 \\ \vdots \\ f_t \end{bmatrix} = \begin{bmatrix} g_1 \\ \vdots \\ g_t \end{bmatrix} \begin{bmatrix} b_{11} \ldots\ldots\ldots\ldots b_{1t} \\ \\ \\ b_{t1} \ldots\ldots\ldots\ldots b_{tt} \end{bmatrix} \qquad (3.2)$$

The values f_z of the reconstructed picture are obtained therefore by the following equations:

$$\begin{aligned} f_1 &= g_1 b_{11} + g_2 b_{12} + \ldots\ldots + g_t b_{1t} \\ &\vdots \\ f_t &= g_1 b_{t1} + g_2 b_{t2} + \ldots\ldots + g_t b_{tt} \end{aligned} \qquad (3.3)$$

Once the weighting factors are calculated, the reconstruction values are obtained by equations 3.3 with simple additions and multiplications. This classic algebraic method, known as *matrix inversion*, is rigorous and straightforward, but in practice it is employed

only with small pictures because the weighting factors matrix contains n^4 cells, and its dimensions become quickly prohibitive with increasing values of n (for a picture with $100 \cdot 100$ cells we would need a weighting factors matrix with $100^4 = 10^8$ cells).

The theoretical limit

Matrix inversion is not widely used in practice, but from a theoretical point of view is extremely useful, because it allows us to calculate the minimum number of projections that are required for a complete reconstruction. If we have p projections of a structure, and each projection contains r rays, a reconstruction procedure amounts to solving a system of $p \cdot r$ equations in n^2 unknowns, and algebra tells us that a solution exists only if the number of *linearly independent* equations is equal to the number of the unknowns.

The condition that equations are linearly independent is easily understandable, because it amounts to saying that projections obtained at different angles must transport different information (if they didn't, the total information of the projections would be inferior to that of the original picture and the reconstruction would be impossible). In practice, the linear independence condition implies that (1) the angle between any two projections must be greater than a critical minimum (which means that the projections must be equally distributed in the 180° angular range), and (2) the ray width (w) and the cell width (d) must have the same order of magnitude, i.e.

$$w \approx d \tag{3.4}$$

Since $d = \frac{D}{n}$ and $w = \frac{D(\vartheta)}{r} \approx \frac{D}{r}$

equation 3.4 is equivalent to $\frac{D}{n} \approx \frac{D}{r}$ and therefore $n \approx r$

This means that the requirement $p \cdot r = n^2$ becomes $p \cdot n \approx n^2$, which amounts to

$$p \approx n \tag{3.5}$$

The result is that *the minimum number of projections that are required for reconstructing a structure of* n^2 *unknowns is comparable to the square root of the number of the unknowns.*

It is important to notice that, in real-life applications, the actual number of projections must always be *greater* (often much greater) than the theoretical minimum, because of the need to compensate the inevitable loss of information which is produced by various types of noise. It is also important to notice that the theoretical minimum obtained with *non-algebraic* methods (Crowther *et al.*, 1970) is never inferior to the algebraic minimum. Equation 3.5, in other words, is the *lowest possible* estimate of the minimum number of projections that are required for a complete reconstruction of any given structure.

ART: an iterative algebraic method

The first algebraic reconstruction method was described by Hounsfield in 1969 in a patent application for computerised tomography, and an equivalent version was published independently by Gordon, Bender and Herman in 1970 with the name of *ART (Algebraic Reconstruction Technique).* Instead of resorting to the matrix inversion approach (which requires matrices of n^4 cells), the reconstructions of this iterative method are performed with matrices of n^2 cells, and are therefore much simpler to handle. The algorithm starts with a uniform matrix [f_{ij}^{0} = constant], and performs an iterative sequence of corrections which tend to bring the reconstructed matrix increasingly closer to the original structure. The corrections consist in calculating the differences between the projection values of the original structure ($g_{\vartheta k}$) and those of the matrix reconstructed at iteration q, ($g_{\vartheta k}^{q}$), and then in redistributing these differences among

the cells of the reconstruction matrix. The reconstruction values at iteration $q+1$ are obtained therefore from those of iteration q with the algorithm

$$f_{ij}^{q+1} = f_{ij}^{q} + \frac{g_{\vartheta k} - g_{\vartheta k}^{q}}{N_{\vartheta k}} \tag{3.6}$$

where $N_{\vartheta k}$ is the number of cells whose central points (i,j) are inside the ray (ϑ,k).

Gordon and Herman (1974) gave to equation 3.6 the name of *"unconstrained ART"*, and called *"partially constrained ART"* the same algorithm subjected to the constraint that negative values are set to zero at the end of each iteration. In addition to this, they called *"totally constrained ART"* the version where negative values are set to zero and values which exceed the maximum M are set to M. At first, it may appear that the totally constrained algorithm requires an a priori knowledge of the maximum M, but in practice it is always possible to obtain a satisfactory estimate of M even without that information. This can be achieved with some preliminary runs of *unconstrained ART* and *partially constrained ART*, because it can be shown that the maxima M^u and M^p obtained with these algorithms satisfy the relationship

$$M^u \leq M \leq M^p$$

and an average of M^u and M^p gives an estimate of M which becomes increasingly accurate as the number of iterations increases.

Gordon and Herman have also proposed a variety of formulae which allow one to compute the distance between the original picture and the reconstructed matrix, and therefore to evaluate the efficiency of a reconstruction algorithm. The ART method, in conclusion, is simple, fast and versatile, which explains why it has become an ideal starting-point for research on a new class of reconstruction algorithms.

The memory matrix

In reconstructions performed with iterative algorithms we usually find, at each iteration, values that are below the minimum and above the maximum, but we have already seen that it is always possible to bring these "illegal" values within the legitimate range. Let us assume, however, that we want to discover something else about those irregular values, apart from the fact that they do exist. It could be interesting, for example, to find out whether their distribution in space is totally random or is following some kind of regularity.

In order to answer this kind of questions, we can perform reconstructions by using not only the structure matrix $[f_{ij}]$ but also an additional matrix $[m_{ij}]$, of the same size, where we "memorise" the illegal values which appear at each iteration. This allows us to conserve a "memory" of them even when they have been erased from the structure matrix, and for this reason their matrix has been called the *memory matrix.*

The construction of the memory matrix is performed by taking as a starting point a totally "blank" matrix $[m_{ij}^{0} = 0]$, and by applying the following operations:

$$\begin{array}{lll} \text{If } f_{ij} \leq 0 & \longrightarrow & m_{ij} = m_{ij} - \gamma \\ \text{If } f_{ij} \geq M & \longrightarrow & m_{ij} = m_{ij} + \gamma \\ \text{otherwise} & \longrightarrow & m_{ij} = m_{ij} \end{array} \qquad (3.7)$$

where γ is a parameter which is chosen to represent the presence of an "illegality" in any convenient way.

The combination of a totally constrained algorithm with equations 3.7 of the memory matrix allows us to build, at each iteration, two very different matrices: the structure matrix where the reconstruction appears, and the memory matrix where the parameters of the illegal values are gradually accumulated.

If the distribution of these values were totally random, the memory matrix would tend to remain uniform, but in reality its behaviour is much more complex than that. At many points the illegal values do

have a random behaviour, in the sense that they appear and disappear in a statistical way, but at other points the illegalities keep reappearing with absolute regularity at each iteration, and always with the same sign. These points clearly behave as "attractors" of density, and for this reason have been called *vortices*. More precisely, the names *negative vortices* and *positive vortices* have been given to the points (or cells) where values appear which are respectively smaller than the minimum and greater than the maximum for T consecutive iterations (where T is a parameter which is chosen by the operator).

By indicating with V_0 the negative vortices and with V_M the positive ones, the recognition of the vortices is performed, every T iterations, with the following criteria:

$$\begin{array}{lll} \text{If } m_{ij} = -T\gamma & \longrightarrow & m_{ij} = V_0 \\ \text{If } m_{ij} = +T\gamma & \longrightarrow & m_{ij} = V_M \\ \text{otherwise} & \longrightarrow & m_{ij} = m_{ij} \end{array} \qquad (3.8)$$

Another important result is obtained by applying this method to pictures of many different kinds, because it has been noticed that the space distribution of the vortices is *picture-dependent*. The vortices' pattern does not depend therefore on general characteristics of the algorithm, but on specific properties of the examined picture. It is as if a picture had a specific image in the memory space exactly as it has one in the real space. This brings us immediately to the following question: *Is it possible to use the information that appears in the memory matrix to improve the reconstruction in the structure matrix?*

The question is absolutely natural because the vortices appear to have a precise, and often even obvious, meaning. If a negative (or a positive) density value keeps reappearing in the same point for T consecutive times, it is clear that in the original structure that point must be a minimum (or a maximum). But if this is true, it is clearly useless to keep treating that point as an *unknown*, and we can therefore erase it from the list of the unknowns. The advantage of this operation is obvious: *while the number of equations* ($p \cdot r$) *remains constant, the number of the unknowns* (n^2) *is decreasing.*

If this is confirmed, the problem of reconstructing structures from

incomplete projections could be solved. The key obstacle, in this problem, is precisely the fact that the number of equations is smaller than the number of unknowns, but if the unknowns are continuously reduced, eventually they would reach the same number as the equations, and at that point an exact reconstruction would be guaranteed. As we can see, the production of "illegal" density values – which was looking like a structural defect of the algorithm – opens the way to unexpected developments.

Density modulation

The first algorithm to use memory matrices was presented at Brookhaven's first international workshop on reconstruction techniques with the name of *"density modulation"* (Barbieri, 1974). This method recognizes the vortices with equations 3.7 and 3.8, and then subtracts them from the list of the unknowns. By indicating with $N^0_{\vartheta k}$ and $N^M_{\vartheta k}$ the number of negative and positive vortices that fall in the ray (ϑ,k), the values of the reconstructed matrix at iteration $q+1$ are obtained with the following instructions:

$$\text{If} \quad m_{ij} = V_0 \text{ or } V_M \longrightarrow f_{ij}^{q+1} = f_{ij}^{q}$$

$$\text{otherwise} \quad f_{ij}^{q+1} = f_{ij}^{q} + \frac{g_{\vartheta k} - g_{\vartheta k}^{q}}{N_{\vartheta k} - N^0_{\vartheta k} - N^M_{\vartheta k}} \tag{3.9}$$

The results obtained with density modulation depend, as we have seen, upon the choice of a parameter T that represents how many times an illegal value must appear in a cell in order to be considered a vortex. If $T = 10$, for example, it is reasonable to conclude that the point in question is a true vortex, but in this case the procedure is lengthy and the number of unknowns decreases very slowly. The choice of $T = 5$, on the other hand, increases the speed of the algorithm but also increases the probability of making mistakes in vortex recognition.

The first reconstructions performed with density modulation were made with the choice T = 5, and the results (Figure 3.7D) clearly showed that some points had been erroneously classified as vortices. Despite these mistakes, however, the reconstructions obtained with density modulation were greatly superior to those of the other algorithms (Figure 3.7B and 3.7C), and the memory method therefore is effective even when the choice of its parameters is not ideal. The most important result, however, is another one. The original pictures (Figure 3.7A)

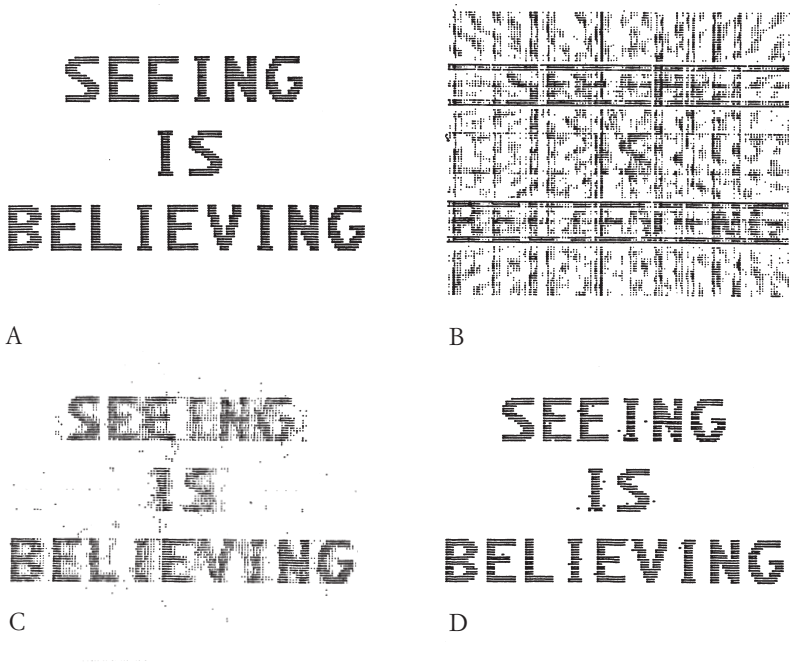

Figure 3.7 A black-and-white picture (A) reconstructed from 12 projections with Convolution (B), ART (C) and density modulation (D). The original picture was a $120 \cdot 120$ matrix, and in order to perform a complete reconstruction it would have been necessary to work with 120 projections in a full 180^{o} angular range.

were matrices with $120 \cdot 120$ cells, and we know, from equation 3.5, that a complete reconstruction requires a minimum of 120 projections. The reconstructions of Figure 3.7 were made instead with only 12 projections, i.e. only 10% of the minimum information was actually used. This is clearly an example of *reconstruction from incomplete projections.* The results obtained with $T = 5$ would surely have been better with $T = 10$, but this is not the point. What really matters is that the main goal has been achieved even with $T = 5$. That goal was the proof that *the memory matrix does allow us to decrease the number of unknowns*, and the results tell us that this is precisely what happens.

The hypotheses that were made about density modulation, therefore, are valid: a memory matrix does allow us to obtain new information about the structure that we are reconstructing, and we can progressively move towards the point where a complete reconstruction becomes possible.

MRM: the family of memory algorithms

One of the interesting features of density modulation is that the reconstructions of black-and-white pictures (Figure 3.7) contain fewer errors than those obtained with grey (or *chiaroscuro*) pictures, i.e. with pictures which have intermediate degrees of density (Figure 3.8). This is understandable, because in black-and-white images all points are either minima or maxima, and the number of vortices is potentially very high. In grey pictures, instead, minima and maxima are far less numerous, and therefore the number of points that can be taken away from the list of the unknowns is much smaller.

This result is interesting because it focuses our attention on the *individual* features of the memory matrix. If only vortices are memorized, it is obvious that the algorithm performs better with pictures that have a high potential number of vortices, but if *other* features could be memorized, it would become possible to reduce substantially the unknowns even with grey pictures. We have therefore the problem of discovering if other features exist which allow us to

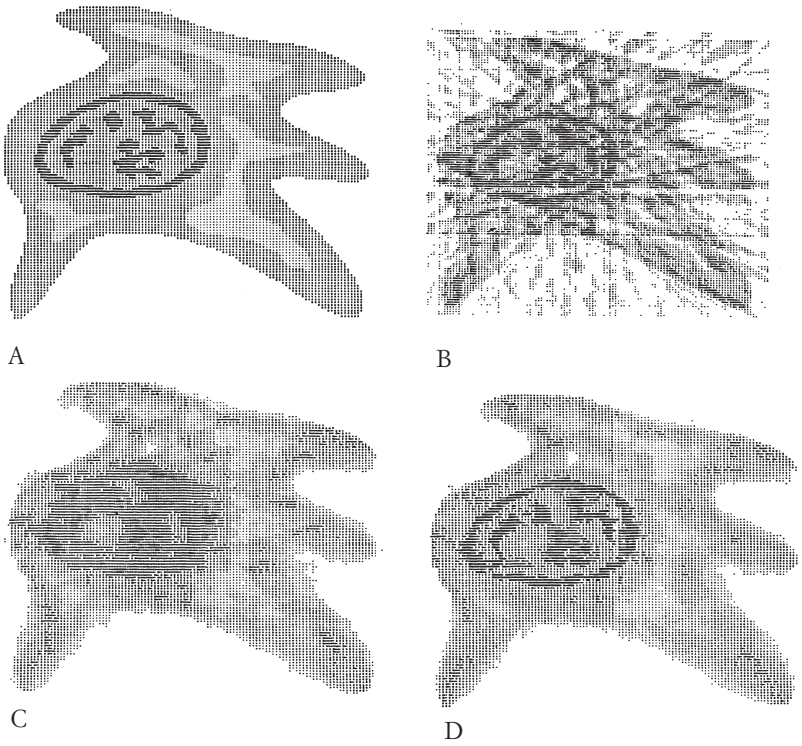

Figure 3.8 A grey, or *chiaroscuro*, picture (A) reconstructed from 12 projections with Convolution (B), ART (C) and density modulation (D). As in the previous case, a complete reconstruction would have required 120 projections equally spaced in the 180° range.

reduce the number of the unknowns, i.e. if there are other types of memory matrices.

A first hint came from the discovery that, in some cells, the reconstructed values can remain virtually unchanged for many consecutive iterations. In order to find these cells – which are called *stationary points* – it is necessary to keep a record of the values obtained in any two consecutive iterations, and to store their differences in a

memory matrix with the instruction

$$m_{ij} = m_{ij} + | f_{ij}^{q+1} - f_{ij}^{q} | \tag{3.10}$$

After a predetermined number T of iterations, the sum of these increments is evaluated, and one can see whether it has exceeded a threshold $T\delta$, where δ is so small a quantity that any density change inferior to it can be regarded as practically insignificant. The cells where that sum is not greater than $T\delta$ are regarded as stationary points, and their formal recognition (with a label S) is performed every T iterations with the following criteria:

$$\begin{aligned} \text{If} \quad & m_{ij} > 0 \text{ and} \leq T\delta \quad \longrightarrow \quad m_{ij} = S \\ & \text{otherwise} \quad \longrightarrow \quad m_{ij} = m_{ij} \end{aligned} \tag{3.11}$$

When stationary points have been identified, it is no longer necessary to treat them like the other points, and we can take them away from the list of the unknowns with the same procedure that was adopted for the vortices.

In practice one can use two different memory matrices – one for vortices and another for stationary points – but it is also possible to use the same matrix for both points. In this case, by indicating with $N^{0}_{\vartheta k}$, $N^{M}_{\vartheta k}$ and $N^{S}_{\vartheta k}$ respectively the negative vortices, the positive vortices and the stationary points that fall within the ray (ϑ,k), the values of the reconstruction matrix at iteration $q+1$ are calculated with the following algorithm:

$$\begin{aligned} \text{If} \quad & m_{ij} = V_0 \,,\, V_M \text{ or } S \quad \longrightarrow \quad f_{ij}^{q+1} = f_{ij}^{q} \\ \text{otherwise} \quad & f_{ij}^{q+1} = f_{ij}^{q} + \frac{g_{\vartheta k} - g_{\vartheta k}^{q}}{N_{\vartheta k} - N^{0}_{\vartheta k} - N^{M}_{\vartheta k} - N^{S}_{\vartheta k}} \end{aligned} \tag{3.12}$$

There are, in conclusion, at least two different types of memory

matrices, and one can use them either separately or together. This makes it important to distinguish between the memory matrix method and the particular algorithms which are based on it, and in order to underline such a distinction it is convenient to adopt a new terminology. The family of all algorithms which use memory matrices is referred to as MRM (*Memory Reconstruction Method*), whereas any individual member of this family is indicated with the label MRM followed by a number. More precisely, MRM-1 is the algorithm which employs only the vortex memory (density modulation), MRM-2 uses only the stationary points memory, and MRM-3 is the algorithm of equation 3.12 which exploits both memories.

At this point we are left with the problem of discovering yet more memory matrices, and here we have plenty of suggestions. It is plausible, for example, that a memory of *boundaries*, or more generally a memory of *discontinuities*, could be built, but we can leave these developments to the future. We have seen that the memory matrix method can indeed perform reconstructions from incomplete information, and therefore we already have what we were looking for: a model that may help us understand the *logic* of embryonic development.

The two general principles of MRM

When we speak of mathematical models for biology, we usually refer to formulae (such as the Hardy–Weinberg theorem, or the Lotka–Volterra equations) that effectively describe some features of living systems. In our case, embryonic development is not described by integrals and deconvolutions, and the formulae of the reconstruction algorithms cannot be a direct description of what happens in embryos. There is however another type of mathematical model. The formulae of energy, entropy and information, for example, apply to all natural processes, irrespective of their mechanisms, and at this more general level there could indeed be a link between reconstruction methods and embryonic development. For our purposes, in fact, what really matters are not the formulae *per se*, but

the general conclusions that they allow us to reach, and among these there are at least two which are indeed worthy of attention.

In the MRM model, the initial memory matrix is a *tabula rasa*, a white page that is gradually filled during the reconstruction process, while the reconstructed picture starts with a uniform image, and becomes progressively differentiated in the course of time. A reconstruction with the MRM model, in other words, is a set of *two* distinct reconstructions that are performed in parallel. The point is that this *double* reconstruction is necessary for reasons that are absolutely general.

A picture and its projections are both structures of the real space, and, when projections are incomplete, there is no possibility of perfoming exact reconstructions if information comes only from structures of the real space (or from equivalent structures of the Fourier space). Only in a *related but autonomous* space we can find genuinely new information, and the memory space is precisely that type of independent world. It is in fact the only space where a system can get the extra information that allows it to increase its own complexity. The MRM model, in other words, leads to a universal concept: to the principle that *there cannot be a convergent increase of complexity without memory.*

The second fundamental characteristic of the MRM model is that information can be transferred from memory space to real space only by suitable *conventions.* In order to decrease the number of the unknowns in real space, it is necessary to give a *meaning* to the structures that appear in memory space, and this too is a conclusion whose validity is absolutely general. Real space and memory space must be autonomous worlds, because if they were equivalent (like real space and Fourier space, for example) they would convey the same information and no increase in complexity would be possible. But between two independent worlds there is no necessary link, and no information can be transferred automatically from one to the other. The only bridge that can establish a link between such worlds is an *ad hoc* process, i.e. a convention or a code. This amounts to a second universal principle: *there cannot be a convergent increase of complexity without codes.*

The Memory Reconstruction Method, in conclusion, gives us two general principles that must be valid for *all* systems which increase their own complexity, and embryos *are* such systems. The MRM model predicts therefore the existence of biological structures which are equivalent to reconstruction codes and to memory matrices. More precisely, the model leads to the conclusion that in embryos there must be codes and memories which are made of organic molecules, i.e. *organic codes* and *organic memories.* At this point, therefore, we can go back to biology and look for the existence of such structures in real life.