Rend. Istit. Mat. Univ. Trieste
Vol. XL, 55–64 (2009)

A Criterion for the Stochasticity of Matrices with Specified Order Relations

Luca Bortolussi and Andrea Sgarro

Abstract. *We tackle the following problem: can one replace a real matrix by a stochastic matrix without altering the order relations between entries? We state a general criterion and a convenient necessary condition. The motivation for this work resides in applications to DNA word design.*

Keywords: Stochastic Matrices, Linear Programming, Separating Hyperplanes
MS Classification 2000: 90C05; 15A51

1. The problem

Studying certain information-theoretic problems [2] which arise in DNA word design,[1] we came upon the following "abstract" problem:

Problem 1.1: Assuming a real $(k \times h)$-matrix $S = \{s_{i,j}\}$ is given, one needs to replace it by a *stochastic* matrix $W = \{w_{i,j}\}$ in such a way that the order relations between entries are all preserved: one has to have $s_{i,j} < s_{r,s}$, $s_{i,j} = s_{r,s}$ or $s_{i,j} > s_{r,s}$ in S iff one has $w_{i,j} < w_{r,s}$, $w_{i,j} = w_{r,s}$ or $w_{i,j} > w_{r,s}$ in W, respectively; $1 \le i, r \le k$, $1 \le j, s \le h$.

In other words, the *start-matrix* S serves only to specify order rela-

[1] DNA word design is a form of error-correction coding which uses DNA strings as codewords, and which is relevant in DNA computing; cf. e.g. [3]. Basically, in [2] one wants to replace "easy-going" *possibilistic* models of channel noise based on "pattern similarities" by means of more committal *probabilistic* models, which are equivalent in the sense that they give rise to the same familty of error-correcting codes; cf. also [7]; cf. e.g. [4] for possibility theory.

tions,[2] but what we actually need is a stochastic matrix, in which the h non-negative components of each row sum exactly to 1. Such a matrix W will be henceforth called *equivalent* to S.

An obvious necessary condition for stochasticity is the following: S should not have *strict domination* between its rows: there cannot be two rows i and s such that $s_{i,j} \geq s_{s,j}$ for all j's unless the two rows coincide entry by entry. In the next Section 2, by slightly deepening this argument, we shall provide a necessary condition for S to have equivalent stochastic matrices. The condition is very simple to check, but counter-example (1) in Section 2 will show that, unfortunately, it is not sufficient.

As the discussion below will soon make clear, ours is basically a problem of linear programming. In this paper we shall provide a criterion for stochasticity at the same level of abstraction as the celebrated Farkas lemma [6], cf. below Criterion 1, Section 4. Before, however, we shall have to re-formulate conveniently our problem, cf. Problem 3.3, Section 3, stating it in terms of Δ-*matrices* as defined in Definition 3.1, Section 3.

2. A Simple Necessary Condition

Our problem of determining whether there exists a stochastic matrix W equivalent to a given start-matrix S may be soon stated in terms of a linear programming problem: we associate a variable to each *distinct* entry of S, say $\mathbf{w} = (w_1, \ldots, w_m)^T$,[3] and we impose the sum of each row of P to be equal to one. Moreover, we impose order constraints of the form $w_i < w_{i+1}$, for each $i = 1, \ldots, m-1$, and the non-negativity constraint $\mathbf{w} \geq \mathbf{0}$ (actually, it is enough to require $w_1 \geq 0$).

---

[2] To no real restriction the entries of S may be assumed to be non-negative real numbers; to avoid trivial specifications, we shall also assume $\max s_{i,j} > 0$ and $k \geq 2$. One may even assume that the maximum entry is 1, and in this case S is called a matrix of *transition possibilities*, cf. [2] and [7]. By the way, ours is *not* a problem of *qualitative probabilities* as in [5], since *compound events* do not play any role whatsoever.

[3] We denote column-vectors in boldface, and interpret equalities and inequalities of vectors componentwise; the superscript T denotes matrix transposition; a real number written in boldface stands for the corresponding constant column-vector.

For example, consider the start-matrix $S = \begin{pmatrix} 1 & 1 & 2 & 3 \\ 3 & 3 & 2 & 1 \end{pmatrix}$ and consider three variables w_1, w_2, w_3 associated respectively to 1,2, and 3. Then, the *associated linear programming problem* is given by the equations

$$\begin{cases} 2w_1 + w_2 + w_3 = 1 \\ w_1 + w_2 + 2w_3 = 1 \end{cases}$$

subject to $0 \leq w_1 < w_2 < w_3$. In this case, as soon checked, the solution set is empty. By the way, one may forget about the non-negativity requirement $w_1 \geq 0$: if one gets a matrix Σ with the same order relations as in S and with each row sum equal to 1, but with negative entries, one just increments all the entries of Σ by the same constant quantity so as to have non-negativity, and then normalises to obtain the desired stochastic matrix W. In the next Section we shall find it convenient to re-cast the linear programming problem in a different way.

The matrix S one starts with might be far from stochasticity, indeed. For example the minimum in row a might be strictly greater than the maximum in row b. Even without going that far, general matrices S may freely have domination between their rows, while stochastic matrices have it only in the limit case when two rows coincide. Actually, in a stochastic matrix row-domination never occurs *however the two rows are permuted*, unless the two rows are a permutation of each other. In practice, one has only to check that there is no domination after ordering rows to be ensured that there never will be however one permutes the row entries, as the following lemma shows:

Lemma 2.1. *Let row a dominate row b, and permute a and b to obtain a^* and b^*, respectively, in such a way as to have non-decreasing order in both a^* and b^*. Then a^* dominates b^*.*

Proof. To no restriction, assume the order of a is already ascending. Take the first column index i such that $b_i > b_{i+1}$ (else $b = b^*$). One has $a_i \geq b_i > b_{i+1}$ and $a_{i+1} \geq a_i \geq b_i$; so, after exchanging $b_i > b_{i+1}$, a is still a dominating row. Now, one may re-arrange b to b^* by successive twiddles (exchanges between adjacent positions; think of the *bubble algorithm* for sorting). □

DEFINITION 2.2. *In two rows of a matrix S there is an inversion when, after re-arranging the rows with respect to the non-decreasing order, there are two positions i and j with $a_i < b_i$, while $a_j > b_j$.*

THEOREM 2.3. *For a matrix S to be equivalent to some stochastic matrix, there must be at least one inversion in each couple of rows, apart from couples of rows which are equal up to a permutation of their entries. However, this condition is not a sufficient one.*

Proof. The necessity has already been argued (the property of having inversions is stable with respect to equivalence). As for a counterexample, take the three-row matrix

$$\begin{pmatrix} a & a & d & d \\ b & c & c & c \\ a & c & c & d \end{pmatrix} \tag{1}$$

with $0 \leq a < b < c < d$; the three rows are already properly arranged in non-decreasing order. In rows 1 and 2 there is an inversion in positions (columns) 1 and 3, in rows 1 and 3 there is an inversion in positions 2 and 3, while in rows 2 and 3 there is an inversion in positions 1 and 4. However, the linear programming problem which one has to solve is

$$a < b < c < d \ , \quad 2a + 2d = 1 \ , \quad b + 3c = 1 \ , \quad a + 2c + d = 1 \ ,$$

whose solution set is empty: actually, the last two equations (after replacing $a + d$ by $1/2$, cf. the first equation) give $b = c = 1/4$, while one should have $b < c$. □

It is no coincidence that the counter-example[4] put forward in the proof is a *three*-row matrix: one can prove that the condition specified in Theorem 2.3 is also *sufficient* for *two*-row matrices (the proof is deferred to Section 4, when the result can be obtained in a much quicker way than we might do now, cf. Corollary 4.3).

When the number of rows is 3 or more, we miss a criterion for equivalence. In the next two sections we shall re-cast our problem,

[4]Applications to DNA word design deal mainly with matrices A such that the row-maxima are all equal: our counterexample can be soon re-cycled to this situation, just assume $e > a, b, c, d$ and add an all-e column.

and show that it is quite general, indeed. We shall provide a criterion for its solution, Criterion 1. However, given the generality of the problem, the criterion we provide is certainly not "simple" in the sense of the necessary condition given in Theorem 2.3, which requires checking rows two at a time. In a way, Criterion 1 still uses "inversions", but they are definitely more complex than those of Definition 2.2. An interesting problem not tackled here would be finding simple *sufficient* conditions, to be set aside the simple necessary condition as in Definition 2.2.

3. A Re-Casting of the Problem in Terms of $\mathbf{\Delta}$-Matrices

As shown in the preceding Section, ours is a linear programming problem: to any start-matrix S we have associated a system of linear equalities and inequalities in the m variables $\mathbf{w} = w_1, \dots, w_m$, which correspond to the distinct entries in the start-matrix S. We move to an alternative and more convenient formulation, which makes it clear that we are dealing with a very general problem, indeed. Along the way, we shall also show how one can get rid of strict inequality constraints, so as to use standard linear-programming algorithms [1].

Consider the consecutive differences among elements of $\mathbf{w}$: $x_i = w_{i+1} - w_i$, for $i = 1, \dots, m-1$; one has $w_k = w_1 + \sum_{i=1}^{k-1} x_k$. Clearly, the system of equalities can be re-written in the form $[\mathbf{m}, A](w_1, \mathbf{x}) = \mathbf{1}$, with $\mathbf{x} = (x_1, \dots, x_{m-1})^T$. Since $w_{i+1} > w_i$, one has $x_i > 0$ ($(w_1, \mathbf{x})$ denotes a column vector whose top component is w_1). Actually, it is enough to deal with the system $A\mathbf{x} = \mathbf{1}$, $\mathbf{x} > 0$: once the shorter system has solution, so has the original system, set e.g. $w_1 = 0$ (use the fact the left-most column $\mathbf{m}$ is constant). So, A is a matrix whose entry $a_{i,j}$ counts the number of occurrences of the unknown increment x_{j-1} in the i^{th} row (equation); a straightforward consequence of the definition is that, for each i, $m \geq a_{i,j} \geq a_{i,j+1}$.

For example, consider again the matrix $S = \begin{pmatrix} 1 & 1 & 2 & 3 \\ 3 & 3 & 2 & 1 \end{pmatrix}$ and rewrite its associated equations as $2w_1+w_2+w_3 \to 4w_1+2x_1+x_2 = 1$ and $w_1 + w_2 + 2w_3 \to 4w_1 + 3x_1 + 2x_2 = 1$. The corresponding Δ-matrix is $A = \begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix}$.

DEFINITION 3.1. *A* Δ-matrix *A is a matrix of non-negative integers,*[5] *whose entries are arranged in non-increasing order in each row.*

DEFINITION 3.2. *The Δ-matrix A associated to a matrix S is obtained from its associated linear programming problem in the unknowns $w_1, w_2, \ldots, w_m$ by setting $x_i = w_{i+1} - w_i$, $1 \le i \le m-1$.*

Clearly, from any Δ-matrix one can go back to a matrix S by just appending on the left a constant column whose repeated integer entry is at least as large as the largest integer in the Δ-matrix. For example, if one starts from the Δ-matrix $A = \begin{pmatrix} 2 & 2 & 2 \\ 4 & 3 & 0 \\ 3 & 3 & 1 \end{pmatrix}$ and appends the constant column $\mathbf{4}$, one re-obtains the counter-example used in the proof of Theorem 2.3.

Thus, Problem 1.1 can be reformulated as follows, after passing from the start-matrix S to its associated Δ-matrix, as in Definition 3.2:

PROBLEM 3.3: *Search for a solution*[6] *of*

$$A\mathbf{x} = \mathbf{1}, \quad \mathbf{x} > \mathbf{0}$$

where A is a Δ-matrix.

We pause for a detour. To get rid of strict inequality constraints, as required by many standard software tools [1], one may use a trick based on the fact that one has at the right of the equality sign a constant column: just add another variable x_m to $\mathbf{x}$ and another column $\mathbf{a_m} = -\mathbf{1}$ to A to get $\overline{A}\mathbf{x} = \mathbf{0}$, with $\overline{A} = [A; -\mathbf{1}]$. With a homogeneous system one soon amends the flaw of having strict inequalities: if the system $\overline{A}\mathbf{x} = \mathbf{0}$ has a positive solution, it has

[5] In the sequel we shall tacitly rule out the case when there are all-zero rows. By so doing, we are ruling out start-matrices S which have a constant row whose entries are equal to the smallest entry. However, such a matrix would be trivial: either it is constant, and then one can provide an equivalent stochastic matrix by just normalising, or it has strict domination between rows, and then there cannot be any equivalent stochastic matrix, as already argued.

[6] Clearly, what matters is only that the column at the right of the equality sign is constant, and so one might replace $\mathbf{1}$ by any constant column-vector $\mathbf{c}$, with $c > 0$.

a positive solution arbitrarily large, and so one can replace $\mathbf{x} > \mathbf{0}$ by $\mathbf{x} \geq \mathbf{1}$.

Now, we proceed to show how general Problem 3.3 is. Think of a problem like $C\mathbf{x} = D,\ \mathbf{x} > \mathbf{0}$, where the coefficients of C and D are strictly positive rationals; the column-vector D need not be constant. Assume that we are able to solve Problem 3.3: then, *a fortiori*, we are able also to solve the new problem, as we now argue. Clearly, one can soon move to a system equivalent to $C\mathbf{x} = D,\ \mathbf{x} > \mathbf{0}$ where all the coefficient are integers. Moreover, if D is not constant, just take a common multiple M of its integer entries d_i to obtain a constant column-vector, and multiply the C-entries in row i by M/d_i to go back to the situation above. If the rows are not properly ordered as in Problem 3.3, one can use a trick. Say there are m unknowns, and so the columns of C are numbered 1 to m. If column $m-1$ (whose entries are strictly positive) does not dominate column m, multiply it by a positive rational coefficient k_{m-1} which is large enough; if column $m-2$ does not dominate the *new* column $m-1$, multiply it by a positive rational coefficient k_{m-2}, and so on. Clearly, the new system fits into the desired mould. In practice, dealing with Problem 3.3 (or with Problem 1.1), is tantamount to dealing with the very general[7] problem of solving linear equations $C\mathbf{X} = D$ in a "strictly positive universe".

4. The Criterion for Stochasticity

In this Section we put forward a "geometric-flavoured" criterion for stochasticity, which uses Δ-matrices. First, however, we need to recall briefly some basic definitions and results in linear programming [6].

A *convex cone* C is a subset of $\mathbb{R}^n$ closed for addition and multiplication by any non-negative scalar, i.e. such that $\mathbf{x_1}, \mathbf{x_2} \in C, \alpha_1, \alpha_2 \geq 0 \Rightarrow \alpha_1\mathbf{x_1} + \alpha_2\mathbf{x_2} \in C$. An important class of convex cones are the *polyhedral cones*, which are defined as the intersection of a finite number of half-spaces: $C = \{\mathbf{y} | A^T\mathbf{y} \leq 0\}$. Equivalently, due to the Minkowsky-Farkas-Weyl theorem [6], they are definable as the set of non-negative linear combinations of a finite number of vectors

[7] As for the rationality constraints, one may get rid of them by using obvious continuity arguments.

$\mathbf{a_1}, \ldots, \mathbf{a_m}$: $C(A) = \{\mathbf{z} | \mathbf{z} = A\mathbf{x}, \mathbf{x} \geq \mathbf{0}\}$, where the matrix A has those m vectors as its columns. A *convex set* K is a subset of $\mathbb{R}^n$ closed for convex combinations, i.e. $x_i \in K$, $\lambda_i \in \mathbb{R}^+$, $\sum \lambda_i = 1$ implies $\sum \lambda_i x_i \in K$. Clearly, every convex cone is a convex set. A fundamental result is the following:

THEOREM 4.1. Separating Hyperplane [6]. *Let $X, Y \subseteq \mathbb{R}^n$ be two convex sets with disjoint interiors. Then there exists a hyperplane W, defined by its* normal vector $\mathbf{w}$, *separating X from Y, i.e. such that $\mathbf{w} \cdot \mathbf{x} \geq 0$ for all $\mathbf{x} \in X$ and $\mathbf{w} \cdot \mathbf{y} \leq 0$ for all $\mathbf{y} \in Y$.*

If one applies the previous theorem to the case in which X is a convex set and Y is a point on the border ∂X of X, one soon obtains the following:

COROLLARY 4.2. Supporting Hyperplane [6]. *Let X be a convex set and $\mathbf{y} \in \partial X$. Then there exists a hyperplane W containing $\mathbf{y}$, with normal vector $\mathbf{w}$, such that $\mathbf{w} \cdot \mathbf{x} \geq 0$ for all $\mathbf{x} \in X$.*

W is called a *supporting hyperplane* for X. Going back to our problem, consider an $n \times m$ Δ-matrix A, $m \geq n$, of rank n. Consider now the cone $C(A)$ spanned by the column vectors $\mathbf{a_1}, \ldots, \mathbf{a_m}$ of A, i.e. $C(A) = \{\mathbf{z} | \mathbf{z} = A\mathbf{x}, \mathbf{x} \geq \mathbf{0}\}$. The existence of a positive solution $\mathbf{x} > 0$ to the equation $A\mathbf{x} = \mathbf{1}$ can be interpreted as the fact that $\mathbf{1}$ *belongs to the interior of* $C(A)$.

Actually, the hypothesis of A having rank n is required for $C(A)$ to have a non-empty interior in $\mathbb{R}^n$ (otherwise it would be contained in a proper subspace of dimension $< n$, which contains no open subset of $\mathbb{R}^n$). If A has rank less than n, we can always remove rows that are linearly dependent.[8]

Applying the separating hyperplane theorem, we can obtain an equivalent geometric condition for the existence of positive solutions to Problem 3.3; we stress once more that the assumption on ranks is not really restrictive:

CRITERION 1. *Let A be a Δ-matrix $n \times m$, $m \geq n$, of rank n.*

---

[8]This operation does not alter the problem, because we can always suppose that $rank(A) = rank([A; \mathbf{1}])$, otherwise no solution to $A\mathbf{x} = \mathbf{1}$ exists. Hence, solutions of a subsystem of maximal rank are also solutions to the original system.

There exists a positive solution to $A\mathbf{x} = \mathbf{1}$ *if and only if for every hyperplane* W *defined by* $\mathbf{w} \cdot \mathbf{x} = 0$*, with* $\mathbf{1} \in W$*, there exist two column vectors* $\mathbf{a_i}$ *and* $\mathbf{a_j}$ *of* A *such that* $\mathbf{w} \cdot \mathbf{a_i} < 0$ *and* $\mathbf{w} \cdot \mathbf{a_j} > 0$*.*

Proof. $Int(C(A)) \neq \emptyset$ in $\mathbb{R}^n$, as $rank(A) = n$, and $\mathbf{1} \in Int(C(A))$ by the hypothesis. Hence, no hyperplane containing $\mathbf{1}$ can be a support for $C(A)$. This means that for each hyperplane W, $\mathbf{1} \in W$, with normal $\mathbf{w}$, there exist $\mathbf{x_1}, \mathbf{x_2} \in C(A)$ such that $\mathbf{w} \cdot \mathbf{x_1} < 0$ and $\mathbf{w} \cdot \mathbf{x_2} > 0$. Indeed, this must be true also for two generators $\mathbf{a_i}, \mathbf{a_j}$ of $C(A)$. In order to prove the other implication, we first show that $\mathbf{1} \in \overline{C(A)}$, the closure of $C(A)$. As $\overline{C(A)}$ is contained in the positive orthant $\mathbb{R}^n_+$, it is sufficient to prove that $\langle \mathbf{1} \rangle \cap \overline{C(A)} \supset \{\mathbf{0}\}$, i.e. that the line $\langle \mathbf{1} \rangle$ spanned by the vector $\mathbf{1}$ non-trivially intersects the closure of $C(A)$ ($\mathbf{0}$ always belongs to the intersection). Suppose not. Both $\overline{C(A)}$ and $\langle \mathbf{1} \rangle$ are convex sets, hence for the hyperplane separation theorem, there exists a hyperplane V, with normal $\mathbf{v}$, such that $\mathbf{v} \cdot \mathbf{z} \geq 0$ for each $z \in \overline{C(A)}$ and $\mathbf{v} \cdot \mathbf{y} \leq 0$ for each $y \in \langle \mathbf{1} \rangle$. But then $\langle \mathbf{1} \rangle \subseteq V$, as $\mathbf{v} \cdot \mathbf{1}$ and $-\mathbf{v} \cdot \mathbf{1}$ must have the same sign (one has $\mathbf{1}$, $-\mathbf{1} \in \langle \mathbf{1} \rangle$). This contradicts the condition specified in the Criterion.

We finally prove that $\mathbf{1} \in Int(C(A))$. The fact that $rank(A) = n$ implies that $Int(C(A)) \neq \emptyset$ in $\mathbb{R}^n$. Now, suppose $\mathbf{1} \in \partial C(A)$. Then, Corollary 4.2 implies that there exists a supporting hyperplane W for $C(A)$ containing $\mathbf{1}$, in contradiction with the condition of the Criterion. □

The "inversions" used in the criterion are certainly not as simple as those in the necessary condition of Section 2. However, for a two-row start-matrix S the criterion reduces to the existence of an inversion as in Definition2.2:

Corollary 4.3. *Let* S *be a start-matrix with two rows having an inversion. Then there exists a stochastic matrix equivalent to it.*

Proof. Let A be the Δ-matrix of S. If S has an inversion, then A has two columns $\mathbf{a_i}$ and $\mathbf{a_j}$ such that $a_{i1} > a_{i2}$ and $a_{j1} < a_{j2}$. In fact, if $s_{1h} > s_{2h}$ and $s_{1k} < s_{2k}$, then $|\{j \mid s_{1j} > s_{2h}\}| > |\{j \mid s_{2j} > s_{2h}\}|$ and $|\{j \mid s_{1j} > s_{1k}\}| < |\{j \mid s_{2j} > s_{1k}\}|$.

In $\mathbb{R}^2$, the only hyperplane containing $(1,1)$ is the line $(1,-1)\cdot(x,y) = x - y = 0$. From $a_{i1} > a_{i2}$ and $a_{j1} < a_{j2}$, we obtain $(1,-1)\cdot\mathbf{a_i} > 0$ and $(1,-1)\cdot\mathbf{a_j} < 0$, hence $A\mathbf{x} = \mathbf{1}$ has a positive solution due to the last Criterion. $\square$

Acknowledgements. We acknowledge support of FIRB LIBI Bioinformatics, INdAM GNCS, and PRIN 2006012773. We thank Gábor Simonyi for his suggestions.

References

[1] Cplex, `http://www.ilog.com/products/cplex/`

[2] L. Bortolussi and A. Sgarro, *Possibilistic channels for dna word design*, in J. Lawry et al., *Soft methods for integrated uncertainty modelling*, Advances in Soft Computing, Springer Verlag, U.S.A. (2006).

[3] A. Condon, R. Corn and A. Marathe, *On combinatorial DNA word design*, J. Comput. Biol. **8 (3)** (2001), 201–220.

[4] D. Dubois and H. Prade, *Fundamentals of fuzzy sets*, Kluwer Academic Press, U.S.A. (2000).

[5] T.L. Fine, *Theories of probability*, Academic Press, U.S.A. (1973).

[6] A. Schrijver, *Theory of linear and integer programming*, Wiley, U.S.A. (1986).

[7] A. Sgarro, *Possibilistic information theory: a coding-theoretic approach*, Fuzzy Sets and Systems **132** (2002), 11–32.

Authors' addresses:

Luca Bortolussi
Dep. of Mathematics and Informatics, University of Trieste, Trieste, Italia
E-mail: `luca@dmi.units.it`

Andrea Sgarro
Dep. of Mathematics and Informatics, University of Trieste, Trieste, Italia
E-mail: `sgarro@units.it`

Received July 9, 2008
Revised September 24, 2008