An approximation approach to the problem of the acquisition of phonotactics in Optimality Theory

Giorgio Magri
Laboratoire de Linguistique Formelle, CNRS and University of Paris 7
magrigrg@gmail.com

Abstract

The *problem of the acquisition of phonotactics* in Optimality Theory is intractable. This paper offers a way to cope with this hardness result: the problem is reformulated as a well known integer program (the *Assignment problem* with *linear side constraints*) paving the way for the application to phonotactics of *approximation algorithms* recently developed for integer programming.

Knowledge of the *phonotactics* of a language is knowledge of its distinction between licit and illicit forms. The acquisition of phonotactics represents a distinguished and important stage of language acquisition. In fact, in carefully controlled experimental conditions, nine-month-old infants already react differently to licit and illicit sound combinations (Jusczyk et al., 1993). They thus display knowledge of phonotactics already at an early stage of language development.

Usually, the problem of the acquisition of the phonotactics of a language given a finite set of linguistic data is formalized as the problem of finding a smallest language in the typology that is consistent with the data (Berwick, 1985; Manzini and Wexler, 1987; Prince and Tesar, 2004; Hayes, 2004; Fodor and Sakas, 2005). Section 1 formulates the problem of the acquisition of phonotactics along these lines within the mainstream phonological framework of *Optimality Theory* (Prince and Smolensky, 2004; Kager, 1999).

Unfortunately, (such a formulation of) the problem of the acquisition of phonotactics in OT turns out to be intractable (NP-complete): for any attempted efficient solution algorithm, there are some instances of the problem where the algorithm fails (Magri, 2010; Magri, 2012b). This hardness result holds for the *universal* formulation of the problem, in the sense of Heinz et al. (2009): there are no restrictions on the constraint set that defines the OT typology and indeed the OT typology itself figures as an input to the problem.

There are two strategies to cope with this hardness result. One approach weakens the formulation of the problem through proper restrictions on the constraint set: certain constraint sets are implausible from a phonological perspective, and should therefore be ignored in the proper formulation of the problem (Magri, 2011; Magri, 2012c). This approach raises interesting challenges, as it requires a through investigation of the algorithmic implications of various generalizations developed by phonologists on what counts as a "plausible" OT constraint set. Another approach is to bypass this difficulty, and weaken the formulation of the problem by lowering the standard for success: we settle on an *approximate* solution, namely a "small" language rather than a smallest language. This paper paves the way for the latter approach.

I focus on the specific formulation of the problem of the acquisition of OT phonotactics developed in Prince and Tesar (2004). In Sections 2 and 3, I show that this formulation of the problem can be restated as a classical integer program, namely the *Assignment problem* with *liner side constraints* (AssignLSCsPbm). The theory of *approximation algorithms* for integer programing is a blooming field of Computer Science (Bertsimas and Weismantel, 2005). In particular, powerful approximation algorithms have been recently developed for the AssignLSCsPbm. A state-of-the-art algorithm is due to Arora et al. (2002). The integer programming formulation developed in this paper thus paves the way for a new approximation approach to the problem of modeling the acquisition of phonotactics within OT. In Magri (2012a), I report simulation results with Arora's et. al. (2002) algorithm on various instances of the problem of the acquisition of phonotactics.

Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology (SIGMORPHON2012), pages 52–61,
Montréal, Canada, June 7, 2012.

1 Formulation of the problem

1.1 Basic formulation

A *typology* in Optimality Theory (OT) is defined through a 4-tuple $\tau = (\mathcal{X}, \mathcal{Y}, Gen, \mathfrak{C})$, where $\mathcal{X}$ is the set of *underlying forms*; $\mathcal{Y}$ is the set of *candidate surface forms*; Gen is the *generating function* that pairs an underlying form $x \in \mathcal{X}$ with a set $Gen(x) \subseteq \mathcal{Y}$ of surface forms called the *candidates* for x; and $\mathfrak{C}$ is the set of n *constraints* $C_1, \ldots, C_n$. Each constraint C_i is a function that maps a pair (x, y) of an underlying form $x \in \mathcal{X}$ and a candidate $y \in Gen(x)$ into a number $C_i(x, y)$, called the corresponding *number of violations*. The constraint set is split into the subset $\mathcal{M}$ of markedness constraints and the subset $\mathcal{F}$ of *faithfulness constraints*. As the constraint set is finite and can therefore only distinguish among a finite number of forms, I can assume that the set of underlying forms $\mathcal{X}$ is finite, as well as the candidate set $Gen(x)$ for any underlying form $x \in \mathcal{X}$.

Let π be a *ranking*, namely a total order over the constraint set. I denote by $\mathsf{OT}_\pi \colon \mathcal{X} \to \mathcal{Y}$ the *OT grammar* corresponding to the ranking π, as defined in Prince and Smolensky (2004). And I denote by $\mathcal{L}(\pi)$ the language corresponding to the ranking π, namely the range of the corresponding grammar OT_π (or, more explicitly, the set of all and only those surface forms $\hat{y} \in \mathcal{Y}$ such that there exists an underlying form $x \in \mathcal{X}$ such that $\mathsf{OT}_\pi(x) = \hat{y}$). Throughout the paper, I use x for an underlying form, $\hat{y}$ for a surface form which is an intended winner, and y for a surface form which is an intended loser.

The *Problem of the acquisition of phonotactics* in OT can be stated as in (1) in its universal formulation (Berwick, 1985; Manzini and Wexler, 1987; Prince and Tesar, 2004; Hayes, 2004). We are given an OT typology as well as a finite set $P \subseteq \mathcal{X} \times \mathcal{Y}$ of linguistic data. These data consist of pairs $(x, \hat{y})$ of an underlying form $x \in \mathcal{X}$ and a corresponding intended winner form $\hat{y} \in Gen(x)$. I assume that P is *consistent*, namely that there exists at least a ranking π such that $\mathsf{OT}_\pi(x) = \hat{y}$ for every pair $(x, \hat{y}) \in P$. We are asked to return a ranking π which has two properties. First, π is *consistent*: the corresponding OT grammar maps x into $\hat{y}$ for every pair $(x, \hat{y}) \in P$. Second, π is *restrictive*: there exists no other ranking π' consistent with P too such that the language $\mathcal{L}(\pi')$ corresponding to π' is a proper subset of the language $\mathcal{L}(\pi)$ corresponding to π. A solution algorithm needs to run in time polynomial in the number of constraints $|\mathfrak{C}|$ and the numbers of forms $|\mathcal{X}|, |\mathcal{Y}|$ (recall that $\mathcal{X}$ and $\mathcal{Y}$ are finite).

(1) *given*: an OT typology $\tau = (\mathcal{X}, \mathcal{Y}, Gen, \mathfrak{C})$ and a finite set $P \subseteq \mathcal{X} \times \mathcal{Y}$ of data;
find: a ranking π s.t. $P \subseteq \mathsf{OT}_\pi$ and there is no π' s.t. $P \subseteq \mathsf{OT}_{\pi'}$ and $\mathcal{L}(\pi') \subset \mathcal{L}(\pi)$;
time: $\max\{|\mathfrak{C}|, |\mathcal{X}|, |\mathcal{Y}|\}$.

Problem (1) is NP-complete: there exists no efficient algorithm that is able to solve any instance of the problem (Magri, 2010; Magri, 2012b).

An interesting variant of the problem (1) assumes that we are given only the surface forms but not the corresponding underlying forms. Prince and Tesar (2004) and Hayes (2004) suggest that we can circumvent this difficulty as follows. Assume that the set of underlying forms and the set of surface forms coincide, namely $\mathcal{X} = \mathcal{Y}$. Assume furthermore that the typology is *output driven* (Tesar, 2008): a surface form $\hat{y}$ belongs to the language $\mathcal{L}(\pi)$ corresponding to a ranking π iff the corresponding grammar OT_π maps that form $\hat{y}$ (construed as an underlying form) into itself (construed as a surface form), as stated in (2)

(2) $$\hat{y} \in \mathcal{L}(\pi) \iff \mathsf{OT}_\pi(\hat{y}) = \hat{y}.$$

In this case, a way to cope with the lack of the underlying forms is to assume that the underlying form corresponding to a given surface form $\hat{y}$ is the completely faithful underlying form $\hat{y}$ itself. For this reason, I stick with the formulation (1) of the problem, whereby we are provided with both surface and underlying forms.

1.2 ERC notation

Consider an underlying form $x \in \mathcal{X}$ and two different candidate forms $y, \hat{y} \in Gen(x)$, with the convention that $\hat{y}$ is the intended winner for x while y is a loser. Following Prince (2002), all the relevant information concerning the underlying/winner/loser form triplet $(x, \hat{y}, y)$ can be summarized into the corresponding *elementary ranking condition* (ERC), namely the n-tuple $\mathbf{e}$ with entries $e_1, \ldots, e_n \in \{\text{L}, e, \text{W}\}$ defined as in (3).

(3) $$(x, \hat{y}, y) \implies \mathbf{e} = \boxed{e_1 \mid \ldots \mid e_i \mid \ldots \mid e_n}$$
$$e_i \doteq \begin{cases} \text{W} & \text{if } C_i(x, \hat{y}) < C_i(x, y) \\ \text{L} & \text{if } C_i(x, \hat{y}) > C_i(x, y) \\ e & \text{if } C_i(x, \hat{y}) = C_i(x, y) \end{cases}$$

In words, The ith entry e_i is e_i = W iff constraint C_i assigns more violations to (x, y) than to $(x, \hat{y})$ and thus favors the intended winner $\hat{y}$ over the loser y; e_i = L iff the opposite holds; finally, $e_i = e$ iff the constraint C_i assigns the same number of violations to the two pairs (x, y) and $(x, \hat{y})$.

A ranking π can be represented as a permutation over $\{1, \ldots, n\}$, with the understanding that $\pi(i) = j$ means that the ranking π assigns constraint C_i to the jth stratum of the ranking, with the convention that the stratum corresponding to $j = n$ (to $j = 1$) is the top (bottom) of the ranking. For every such permutation π, let $\mathbf{e}_\pi$ be the n-tuple $\mathbf{e}$ with the components reordered according to π in decreasing order, as in (4).

(4) $\mathbf{e}_\pi \doteq (e_{\pi(n)}, \ldots, e_{\pi(1)})$

The ERC $\mathbf{e}$ is *OT-consistent* with π provided the left-most component of $\mathbf{e}_\pi$ different from e is a W.

For each of the pairs $(x, \hat{y})$ in the set P given with an instance of the problem (1), consider each loser candidate $y \in Gen(x)$ different from $\hat{y}$, construct the ERC corresponding to the underlying/winner/loser form triplet $(x, \hat{y}, y)$ as in (3) and organize all these ERCs one underneath the other into an *ERC matrix* with n columns and many rows (the order of the ERCs does not matter). I denote a generic ERC matrix by $\mathbf{E}$ and I say that a ranking π is *OT-consistent* with $\mathbf{E}$ provided it is consistent with each of its ERCs. The problem of the acquisition of phonotactics in (1) can thus be equivalently restated in ERC notation as in (5).

(5) *given*: an OT typology $\tau = (\mathcal{X}, \mathcal{Y}, Gen, \mathfrak{C})$ and an ERC matrix $\mathbf{E}$;
find: a ranking π s.t. π is OT-consistent with $\mathbf{E}$ and there is no π' consistent with $\mathbf{E}$ too s.t. $\mathcal{L}(\pi') \subset \mathcal{L}(\pi)$;
time: $\max\{|\mathfrak{C}|, |\mathcal{X}|, |\mathcal{Y}|\}$.

The latter formulation of the problem is only partially stated in terms of ERC notation, as the condition $\mathcal{L}(\pi') \subset \mathcal{L}(\pi)$ still requires knowledge of the entire OT typology. This difficulty is tackled in the next Subsection.

1.3 Restrictiveness measures

Let a *restrictiveness measure* be a function μ which takes a ranking π and returns a number $\mu(\pi) \in \mathbb{N}$ that provides a relative measure of the size of the language $\mathcal{L}(\pi)$ corresponding to π, in the sense that the (strict) monotonicity property in (6) holds for any two rankings π, π'.

(6) If $\mathcal{L}(\pi') \subset \mathcal{L}(\pi)$, then $\mu(\pi') < \mu(\pi)$.

Any solution of the optimization problem (7) is a solution of the corresponding instance (5) of the problem of the acquisition of phonotactics. In fact, if π solves (7) then there cannot exist any other ranking π' consistent with the ERC matrix that corresponds to a smaller language $\mathcal{L}(\pi') \subset \mathcal{L}(\pi)$, since (6) would imply that $\mu(\pi') < \mu(\pi)$, contradicting the hypothesis that π is a solution of (7).

(7) *minimize*: $\mu(\pi)$;
subject to: π is OT-consistent with the given ERC matrix $\mathbf{E}$;
time: number of columns and rows of $\mathbf{E}$.

As problem (7) is stated completely in terms of the ERC matrix $\mathbf{E}$, the time required by a solution algorithm needs to scale just with the size of $\mathbf{E}$.

From now on, I will focus on the new formulation (7). Thus, I need a restrictiveness measure (6). Of course, not just any restrictiveness measure will do. For instance, the function (8), which pairs a ranking π with the cardinality of its language $\mathcal{L}(\pi)$, trivially satisfies (6).

(8) $\mu(\pi) \doteq |\mathcal{L}(\pi)|$.

Yet, this is not a *good* restrictiveness measure, because there seems to be no way to compute $\mu(\pi)$ without actually computing the language $\mathcal{L}(\pi)$, which requires knowledge of the entire typology.

Prince and Tesar (2004) suggest a better candidate, which is defined for any ranking π as in (9). Recall that the constraint set $\mathfrak{C} = \mathcal{F} \cup \mathcal{M}$ is split up into the subset $\mathcal{F}$ of faithfulness constraints and the subset $\mathcal{M}$ of markedness constraints. For each faithfulness constraint $F \in \mathcal{F}$, determine the number $\mu(F)$ of markedness constraints $M \in \mathcal{M}$ ranked by π *below* that faithfulness constraint, i.e. $\pi(F) > \pi(M)$. Finally, add up all these numbers $\mu(F)$ together to determine the value $\mu(\pi)$.

(9) $$\mu(\pi) \doteq \sum_{F \in \mathcal{F}} \underbrace{\Big|\{M \in \mathcal{M} \mid \pi(F) > \pi(M)\}\Big|}_{\mu(F)}$$

Is the function μ defined in (9) is a restrictiveness measure? namely, does it satisfy condition (6)? Prince and Tesarconjecture that it is, based on the following intuition. Markedness (faithfulness) constraints work against (towards) the preservation of the underlying contrasts. Thus, a small (large) language should arise by ranking the markedness (faithfulness) constraints as high as possible. And a ranking that ranks the markedness (faithfulness) constraints as high (low) as possible is a ranking that minimizes Prince and Tesar's function (9).

I endorse Prince and Tesar's conjecture that (9) is a restrictiveness measure, at least for the cases of interest.[1] In Magri (2012a), I backup this claim by looking at a case study, namely the typology corresponding to the large constraint set considered in Pater and Barlow (2003). In the rest of this paper, I thus focus on the reformulation (7) of the problem of the acquisition of phonotactics, with μ defined as in (9). The latter formulation of the problem of the acquisition of phonotactics is NP-complete too (Magri, 2010; Magri, 2012b). In the rest of this paper, I thus develop an integer programming formulation of the latter problem, that allows approximation algorithms for integer programming to be used in order to tackle the problem of the acquisition of phonotactics. The reasoning is split up into two steps. In Section, 2, I develop an integer programming formulation of the objective function, namely the alleged restrictiveness measure in (9). And in Section 3, I turn to an integer programming formulation of the OT-consistency condition.

2 An integer programming restatement of the restrictiveness measure

A *square matrix of order* n is a collection of n^2 real numbers displayed into n columns and n rows. I denote a square matrix of order n as $\mathbf{X} = [x_{i,j}]_{i,j=1}^{n}$, with the understanding that $x_{i,j}$ is the element of the matrix $\mathbf{X}$ which sits in the ith row and the jth column. I denote by $\mathbb{R}^{n \times n}$ the vector space of all square matrices of order n.

A square matrix $\mathbf{X} = [x_{i,j}]_{i,j=1}^{n}$ is called a *permutation matrix* iff its elements $x_{i,j}$ satisfy the following three conditions: (i) they are all 0 or 1; (ii) each column contains a unique 1; (iii) each row contains a unique 1. I denote by $\mathcal{P}^n$ the set of all $n!$ permutation matrices of order n. To illustrate, I list $\mathcal{P}^n$ with $n = 3$ in (10).

(10) $$\begin{bmatrix} 1&0&0\\0&1&0\\0&0&1 \end{bmatrix} \begin{bmatrix} 1&0&0\\0&0&1\\0&1&0 \end{bmatrix} \begin{bmatrix} 0&1&0\\1&0&0\\0&0&1 \end{bmatrix}$$
$$\begin{bmatrix} 0&0&1\\1&0&0\\0&1&0 \end{bmatrix} \begin{bmatrix} 0&1&0\\0&0&1\\1&0&0 \end{bmatrix} \begin{bmatrix} 0&0&1\\0&1&0\\1&0&0 \end{bmatrix}$$

Permutation matrices play a special role in convex geometry (Webster, 1984, par. 5.8).

There is a natural correspondence between permutation matrices of order n and rankings over n constraints $C_1, \ldots, C_n$. Recall that a ranking π is a permutation over $\{1, 2, \ldots, n\}$, with the understanding that $\pi(i) = j$ means that the ranking π assigns the constraint C_i to the jth stratum, with the convention that the stratum corresponding to $j = n$ is the top stratum. I use i as the index ranging over constraints and j as the index ranging over strata. Thus, a ranking π can be identified with that (unique) permutation matrix $\mathbf{X} = [x_{i,j}]_{i,j=1}^{n} \in \mathcal{P}^n$ such that $x_{i,j} = 1$ iff the ranking π assigns the constraint C_i to the jth stratum, namely $\pi(i) = j$. To illustrate, I list in (11) the rankings over $\{C_1, C_2, C_3\}$ corresponding to the six permutation matrices in (10), respectively.

(11) $$C_3 \gg C_2 \gg C_1,\ \ C_2 \gg C_3 \gg C_1,\ \ C_3 \gg C_1 \gg C_2,$$
$$C_1 \gg C_3 \gg C_2,\ \ C_2 \gg C_1 \gg C_3,\ \ C_1 \gg C_2 \gg C_3$$

I denote by $\pi_{\mathbf{X}}$ the ranking corresponding to a permutation matrix $\mathbf{X} \in \mathcal{P}^n$ and by $\mathbf{X}_\pi \in \mathcal{P}^n$ the permutation matrix corresponding to a ranking π. Prince and Tesar's restrictiveness measure (9) of a ranking π can be straightforwardly read off the corresponding permutation matrix $\mathbf{X}_\pi$, as follows.

Define the *scalar product* $\langle \mathbf{X}, \mathbf{Y} \rangle \in \mathbb{R}$ between two arbitrary square matrices $\mathbf{X} = [x_{i,j}]_{i,j=1}^{n}, \mathbf{Y} = [y_{i,j}]_{i,j=1}^{n} \in \mathbb{R}^{n \times n}$ as in (12) (namely as the Euclidean scalar product of $\mathbb{R}^{n^2}$).

[1] Prince and Tesar's conjecture that (9) is a restrictiveness measure runs into a straightforward problem when the constraint set $\mathfrak{C}$ contains both positional and faithfulness constraints. Yet, there are various ways to circumvent this difficulty posed by positional constraints. One way could be to weigh differently the two types of faithfulness constraints in the determination of restrictiveness. Thus, we could switch from the definition in (9) to the variant in (i), where $\mathcal{F}_{pos}$ is the set of positional faithfulness constraints, $\mathcal{F}_{gen}$ is the set of general faithfulness constraints and α is a positive coefficient.

(i) $$\mu_\alpha(\pi) \doteq \sum_{F \in \mathcal{F}_{pos}} \left| \left\{ M \in \mathcal{M} \,\middle|\, \pi(F) > \pi(M) \right\} \right| + \\ + \alpha \sum_{F \in \mathcal{F}_{gen}} \left| \left\{ M \in \mathcal{M} \,\middle|\, \pi(F) > \pi(M) \right\} \right|$$

Another way to deal with positional faithfulness constraints could be to ignore altogether rankings where a positional faithfulness constraint is ranked below the corresponding general faithfulness constraint. This is trivial to obtain, by adding a proper ERC to the ERC matrix given with an instance of the problem (7).

(12) $$\langle \mathbf{X}, \mathbf{Y} \rangle \doteq \sum_{i,j=1}^{n} x_{i,j} y_{i,j}.$$

A function $f : \mathbb{R}^{n \times n} \to \mathbb{R}$ is called *linear* iff there exists a square matrix $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ such that (13) holds for any square matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$.

(13) $$f(\mathbf{X}) = \langle \mathbf{\Sigma}, \mathbf{X} \rangle.$$

Linear functions are the "simplest" possible convex functions, namely the ones that yield the easiest optimization problems.

Let me assume that the first m constraints in $\mathfrak{C}$ are the faithfulness constraints while the remaining $n - m$ constraints are the markedness constraints, namely that $\mathcal{F} = \{C_1, \ldots, C_m\}$ and $\mathcal{M} = \{C_{m+1}, \ldots, C_n\}$. Consider the matrix $\mathbf{\Sigma}_{n,m} \in \mathbb{R}^{n \times n}$ defined as follows: its first m rows each have the form $[0, 1, \ldots, n-2, n-1]$; the remaining $n - m$ rows are all null. To illustrate, I give in (14) the matrix $\mathbf{\Sigma}_{n,m}$ with $n = 7, m = 4$.

(14) $$\mathbf{\Sigma}_{7,4} \doteq \left[\begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The following Claim 1 explains how to compute the restrictiveness $\mu(\pi)$ of a ranking π according to (9) out of the corresponding permutation matrix $\mathbf{X}_\pi$; see Appendix A.1. This Claim shows an important property of Prince and Tesar's restrictiveness measure: it can be described as a linear function over the set of permutation matrices.

Claim 1 *The restrictiveness $\mu(\pi)$ of a ranking π according to (9) can be computed as follows:*

(15) $$\mu(\pi) = \langle \mathbf{\Sigma}_{n,m}, \mathbf{X}_\pi \rangle - \frac{1}{2} m(m-1)$$

namely as the scalar product $\langle \mathbf{\Sigma}_{n,m}, \mathbf{X} \rangle$ between the matrix $\mathbf{\Sigma}_{n,m}$ and the corresponding permutation matrix $\mathbf{X}_\pi$, minus the constant $\frac{1}{2} m(m-1)$ which does not depend on the ranking.[2] ■

[2] I have noted in footnote 1 that the conjecture that the function μ in (9) is a restrictiveness measure runs into problems for constraint sets that contain both general and positional faithfulness constraints. And I have suggested that a possible way out is to to switch from the definition (9) to the variant in (i). Let me now point out that the latter variant too can be described as a linear function over permutation matrices. In fact, let $\mathbf{\Sigma}_{n,m,\alpha}$ be as the matrix $\mathbf{\Sigma}_{n,m}$ defined above, but with the rows corresponding to general faithfulness constraints multiplied by α. Then, $\mu_\alpha(\mathbf{X})$ coincides with $\langle \mathbf{\Sigma}_{n,m,\alpha}, \mathbf{X} \rangle$, but for a constant.

The problem of the acquisition of phonotactics (7) with Prince and Tesar's alleged restrictiveness measure (9) can thus be restated as the optimization problem (16).

(16) *minimize*: $\langle \mathbf{\Sigma}_{n,m}, \mathbf{X} \rangle$;
subject to: $\mathbf{X} \in \mathcal{P}^n$ and $\pi_{\mathbf{X}}$ is consistent with the given ERC matrix $\mathbf{E}$.

Here, I have dropped the constant $\frac{1}{2} m(m-1)$ which appears in (15), as it does not affect the optimization problem.

3 An integer programming formulation of the OT-consistency condition

The reformulation in (16) makes use of the notion of OT-consistency with a given ERC matrix and this notion is currently stated in terms of rankings rather than in terms of the corresponding permutation matrices. We need to restate the latter condition directly in terms of permutation matrices. In this Section, I point out two strategies for doing that. The first approach hinges on a classical observation by Prince and Smolensky (2004) that OT consistency can be restated as linear consistency in the case of exponentially spaced weights. The second approach requires a larger number of linear conditions, but is shown to provide a better reformulation (i.e. a tighter relaxation).

3.1 An initial formulation of OT-consistency

Given an ERC $\mathbf{e} = [e_1, \ldots, e_n]$, consider the corresponding square matrix $\mathbf{A}_\mathbf{e} = [a_{i,j}]_{i,j=1}^n \in \mathbb{R}^{n \times n}$ defined in (17). Here, t_i is the *sign* of the ERC's entry e_i, namely t_i is equal to -1, 0 or $+1$ depending on whether e_i is equal to L, e or W. Thus, the entry $a_{i,j}$ in the ith row and the jth column of the matrix (17) consists of the sign t_i multiplied by 2^j.

(17) $$\mathbf{A}_\mathbf{e} = \begin{bmatrix} 2^1 t_1 & 2^2 t_1 & \ldots & 2^j t_1 & \ldots & 2^n t_1 \\ & & & \vdots & & \\ 2^1 t_i & 2^2 t_i & \ldots & 2^j t_i & \ldots & 2^n t_i \\ & & & \vdots & & \\ 2^1 t_n & 2^2 t_n & \ldots & 2^j t_n & \ldots & 2^n t_n \end{bmatrix}$$

Intuitively, this entry $a_{i,j} = 2^j t_i$ is the weight of the sign t_i under the assumption that the constraint

C_i is assigned to the jth stratum.

The following claim offers a restatement of OT-consistency between an ERC and a ranking in terms of the permutation matrix corresponding to that ranking. This claim is just a restatement in matrix form of the observation by Prince and Smolensky (2004) that OT consistency is equivalent to a linear condition with exponentially spaced weights; see Subsection A.2.

Claim 2 *A ranking π is OT-consistent with an ERC $\mathbf{e}$ iff $\langle \mathbf{A}_{\mathbf{e}}, \mathbf{X}_\pi \rangle \geq 0$, where $\langle \mathbf{A}_{\mathbf{e}}, \mathbf{X}_\pi \rangle$ is the scalar product (12) between the matrix $\mathbf{A}_{\mathbf{e}}$ corresponding to the ERC $\mathbf{e}$ and the permutation matrix $\mathbf{X}_\pi$ corresponding to the ranking π.* ■

The current formulation (16) of the problem of the acquisition of phonotactics can thus be restated as the optimization problem in (18).

(18) FIRST INTEGER REFORMULATION:
minimize: $\langle \mathbf{\Sigma}_{n,m}, \mathbf{X} \rangle$;
subject to: $\mathbf{X} \in \mathcal{P}^n$ s.t. $\langle \mathbf{A}_{\mathbf{e}}, \mathbf{X} \rangle \geq 0$ for every ERC $\mathbf{e}$ of the ERC matrix $\mathbf{E}$.

Problem (18) is an optimization problem over permutation matrices $\mathbf{X} \in \mathcal{P}^n$. The *objective function* is the linear function $\langle \mathbf{\Sigma}_{n,m}, \mathbf{X} \rangle$. And the feasible set is defined in terms of linear *side conditions* $\langle \mathbf{A}_{\mathbf{e}}, \mathbf{X} \rangle \geq 0$. Problem (18) is thus an *integer program*. In particular, it is an *Assignment problem* with *linear side constraints* (AssignLSCsPbm) (Bertsimas and Weismantel, 2005).

3.2 Another formulation of OT-consistency

Let $\ell(\mathbf{e})$ be the number of entries equal to L in an ERC $\mathbf{e} = [e_1, \ldots, e_n]$. Assume without loss of generality that $\ell(\mathbf{e}) > 0$, as ERCs with no L's can be ignored. For every stratum $\bar{\jmath} \in \{1, \ldots, n\}$, consider the square matrix $\mathbf{A}^{\bar{\jmath}}_{\mathbf{e}} = [a_{i,j}]^n_{i,j=1}$ with n rows and n columns whose generic element $a_{i,j}$ is defined as in (19).

(19) $$a_{i,j} \doteq \begin{cases} 1 & \text{if } e_i = \text{L},\ j \geq \bar{\jmath} \\ -1 & \text{if } e_i = \text{W},\ j \geq \bar{\jmath} + \ell \\ 0 & \text{otherwise} \end{cases}$$

The following claim offers another restatement of OT-consistency between an ERC and a ranking in terms of the permutation matrix corresponding to that ranking; see Subsection A.3.

Claim 3 *A ranking π is OT-consistent with an ERC $\mathbf{e}$ iff $\langle \mathbf{A}^{\bar{\jmath}}_{\mathbf{e}}, \mathbf{X}_\pi \rangle \leq 0$ for every $\bar{\jmath} \in \{1, \ldots, n\}$, where $\langle \mathbf{A}^{\bar{\jmath}}_{\mathbf{e}}, \mathbf{X}_\pi \rangle$ is the scalar product (12) between the matrix $\mathbf{A}^{\bar{\jmath}}_{\mathbf{e}}$ corresponding to the ERC $\mathbf{e}$ and the stratum $\bar{\jmath}$ and the permutation matrix $\mathbf{X}_\pi$ corresponding to the ranking π.* ■

The current formulation (16) of the problem of the acquisition of phonotactics can thus be alternatively restated as the optimization problem (20).

(20) SECOND INTEGER REFORMULATION:
minimize: $\langle \mathbf{\Sigma}_{n,m}, \mathbf{X} \rangle$;
subject to: $\mathbf{X} \in \mathcal{P}^n$ s.t. $\langle \mathbf{A}^{\bar{\jmath}}_{\mathbf{e}}, \mathbf{X} \rangle \leq 0$ for every ERC $\mathbf{e}$ of the ERC matrix $\mathbf{E}$ and every $\bar{\jmath} \in \{1, \ldots, n\}$.

Again, (20) is another instance of the AssignLSCsPbm. The feasible set in the latter formulation (20) involves n times more inequalities than the formulation (18).

3.3 Comparing the two formulations

Problems (18) and (20) are two different formulations of the original problem (16) of the acquisition of phonotactics. They are thus equivalent, in the sense that a solution to any of the two problems is also a solution to the other and furthermore to the original problem. This Subsection explains why, nonetheless, the latter formulation (20) is better than the former formulation (18).

Both (18) and (20) are optimization problems over permutation matrices $\mathbf{X} \in \mathcal{P}^n$. The latter condition on the matrix $\mathbf{X} = [x_{i,j}]^n_{i,j=1}$ means that conditions (21) hold for any $i, j = 1, \ldots, n$.

(21) $$x_{i,j} \in \{0, 1\} \qquad \sum_{i=1}^{n} x_{i,j} = 1, \quad \sum_{j=1}^{n} x_{i,j} = 1$$

Problems (18) and (20) are *integer* optimization problems because of the condition $x_{i,j} \in \{0, 1\}$ in (21). This condition can be *relaxed*, requiring the entires $x_{i,j}$ to be not necessarily 0 or 1 but instead any number in between 0 and 1. Thus, let $\mathcal{P}^n_{\text{rel}}$ be the set of matrices that satisfy the relaxed conditions (22), known as the *Birkhoff polytope*.

(22) $$x_{i,j} \in [0, 1] \qquad \sum_{i=1}^{n} x_{i,j} = 1, \quad \sum_{j=1}^{n} x_{i,j} = 1$$

Relaxing the integer constraint $\mathbf{X} \in \mathcal{P}^n$ into the continuous constraint $\mathbf{X} \in \mathcal{P}^n_{\text{rel}}$, yields the two corresponding problems (23) and (24).

(23) FIRST RELAXATION:

minimize: $\langle \mathbf{\Sigma}_{n,m}, \mathbf{X} \rangle$;

subject to: $\mathbf{X} \in \mathcal{P}^n_{\text{rel}}$ s.t. $\langle \mathbf{A}_{\mathbf{e}}, \mathbf{X} \rangle \leq 0$ for any ERC $\mathbf{e}$ of the ERC matrix.

(24) SECOND RELAXATION:

minimize: $\langle \mathbf{\Sigma}_{n,m}, \mathbf{X} \rangle$;

subject to: $\mathbf{X} \in \mathcal{P}^n_{\text{rel}}$ s.t. $\langle \mathbf{A}^{\bar{\jmath}}_{\mathbf{e}}, \mathbf{X} \rangle \geq 0$ for any ERC $\mathbf{e}$ of the ERC matrix and any stratum $\bar{\jmath} \in \{1, \ldots, n\}$.

These linear programs (23) and (24) are the *relaxations* of the two integer programs (18) and (20).

The relaxation of an integer program provides a lower bound on the solution of that integer program. This lower bound is used by solution algorithms for the integer program. Of course, linear relaxations that provide tight bounds yield improved solution algorithms for the original integer problem (Bertsimas and Weismantel, 2005). Despite the fact that the two original integer programs (18) and (20) are equivalent, the two corresponding relaxations (23) and (24) are not. Claim 4 ensures that the feasible set of the relaxation (24) is a subset of that of the relaxation (23), so that the lower bound provided by a solution of the former will be at least as tight as the lower bound provided by a solution of the latter.

Claim 4 *If a matrix* $\mathbf{X}$ *belongs to the feasible set of problem (24), then it also belongs to the feasible set of problem (23).* ■

The following counterexample shows that the lower bound provided by the relaxation (24) is not just as tight as but actually tighter than the bound provided by the relaxation (23). Given the ERC matrix (25), the solution to the corresponding problem (7) is the ranking $F_2 \gg M \gg F_1$: the faithfulness constraint F_1 is redundant and should therefore be ranked at the bottom.

(25)
$$\mathbf{E} = \begin{array}{c} \begin{matrix} F_1 & F_2 & M \end{matrix} \\ \left[\begin{matrix} \text{W} & \text{W} & \text{L} \\ e & \text{W} & \text{L} \end{matrix}\right] \end{array}$$

The solutions of the two corresponding relaxations (23) and (24) are provided in (26).[3]

(26)
$$\mathbf{X}_{(23)} = \begin{array}{c} \\ F_1 \\ F_2 \\ M \end{array} \begin{array}{c} \begin{matrix} st1 & st2 & st3 \end{matrix} \\ \left[\begin{matrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{matrix}\right] \end{array} \quad \mathbf{X}_{(24)} = \begin{array}{c} \\ F_1 \\ F_2 \\ M \end{array} \begin{array}{c} \begin{matrix} st1 & st2 & st3 \end{matrix} \\ \left[\begin{matrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{matrix}\right] \end{array}$$

The relaxation (23) has a non-integral solution; the relaxation (24) is thus stronger because its solution is integral. The latter solution indeed represents the desired ranking, as it assigns F_2 to the top 3rd stratum (because of the 1 in the second column and third row) and F_1 to the bottom 1st stratum (because of the 1 in the first row and first column).

4 Conclusion

In this paper, I have focused on Prince and Tesar's (2004) formulation (7) of the problem of the acquisition of phonotactics, in terms of the alleged restrictiveness measure (9). This problem is NP-complete. To cope with this hardness result, in this paper I have looked for an integer programming formulation of the latter problem. The formulation in (20) has emerged as the best formulation among those considered, namely the one that yields the tightest relaxation. This problem (20) is an instance of a classical integer program, namely the *Assignment problem* with *linear side constraints* (AssignLSCsPbm). The result obtained in this paper thus paves the way for the efficient application of approximation algorithms for the AssignLSCsPbm to the problem of the acquisition of phonotactics in OT. In Magri (2012a), I report simulation results with Arora's et. al. (2002) algorithm, a state-of-the-art approximation algorithm for the AssignLSCsPbm.

Acknowledgments

I wish to thank A. Albright for endless discussion on the problem of the acquisition of phonotactics. This work was supported in part by a 'Euryi' grant from the European Science Foundation to P. Schlenker, by a grant from the Fyssen Research Foundation, and by the LABEX-EFL grant.

Appendix: proof of the main results

A.1 Proof of claim 1

Consider the example of the permutation matrix $\mathbf{X}$ in (27). There are seven constraints (hence $n = 7$), four of which are faithfulness constraints (hence $m = 4$). I have fringed each row of $\mathbf{X}$ with the name of the constraint it corresponds to and I have

[3]These solutions have been computed with the Matlab codes `RelaxedSubPbmFirstFormulation.m` and `RelaxedSubPbmSecondFormulation.m`, that solve the two relaxations (23) and (24), respectively. These codes are available on the author's website. The two codes use the two subroutines `MatrixToVectorConverter.m` and `VectorToMatrixConverter.m`, that are available on the author's website too.

fringed each column of $\mathbf{X}$ with the stratum it corresponds to.

$$
(27)\quad \mathbf{X} = \begin{array}{c} \\ F_1 \\ F_2 \\ F_3 \\ F_4 \\ M_5 \\ M_6 \\ M_7 \end{array} \begin{array}{c} \begin{array}{ccccccc} st1 & st2 & st3 & st4 & st5 & st6 & st7 \end{array} \\ \left[\begin{array}{ccccccc} 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 \end{array}\right] \end{array}
$$

As prescribed by our conventions, the first four rows correspond to the four faithfulness constraints, the bottom three rows correspond to the markedness constraints; the leftmost column corresponds to the bottom stratum and the rightmost column corresponds to the top stratum.

The ranking $\pi_{\mathbf{X}}$ that corresponds to $\mathbf{X}$ can be obtained as follows: the 1 in the first column of $\mathbf{X}$ says that the markedness constraint M_6 is assigned by $\pi_{\mathbf{X}}$ to the bottom stratum $j = 1$; the 1 in the second column of $\mathbf{X}$ says that the faithfulness constraint F_1 is assigned to the next stratum $j = 2$; and so on. Thus, $\pi_{\mathbf{X}}$ is the ranking (28).

$$
(28)\quad F_4 \gg F_2 \gg M_7 \gg M_5 \gg F_3 \gg F_1 \gg M_6
$$

According to (9), the restrictiveness $\mu(\pi_{\mathbf{X}})$ of this ranking $\pi_{\mathbf{X}}$ is $8 = 3{+}3{+}1{+}1$: 3 markedness constraints underneath F_4, another 3 underneath F_2, 1 underneath and F_3 as all as underneath F_1. Here is a way to quickly compute this number directly from the permutation matrix $\mathbf{X}$.

Consider the matrix (29) obtained from the matrix (27) through the following two steps. *First*, all 1's which appear in the bottom three rows of $\mathbf{X}$ (and thus correspond to markedness constraints) are replaced with 0's.

$$
(29)\quad \begin{array}{c} \\ F_1 \\ F_2 \\ F_3 \\ F_4 \\ M_1 \\ M_2 \\ M_3 \end{array} \begin{array}{c} \begin{array}{ccccccc} st1 & st2 & st3 & st4 & st5 & st6 & st7 \end{array} \\ \left[\begin{array}{ccccccc} 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{5} & 0 \\ 0 & 0 & \mathbf{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}
$$

Second, each 1 which appears in one of the top four rows of $\mathbf{X}$ (and thus corresponds to a faithfulness constraint) is replaced with the number which identifes the corresponding column, diminished by 1. Thus for example, the 1 in the second row in the matrix $\mathbf{X}$ in (27) is replaced by a 5 in (29), since it occurs in the sixth column.

Next, let's scan the columns of the matrix (29) from left to right, assigning to each column which is not all zeros a progressive index k starting from $k = 0$, as made explicit in (30).

$$
(30)\quad \begin{array}{c} \\ F_1 \\ F_2 \\ F_3 \\ F_4 \\ M_1 \\ M_2 \\ M_3 \end{array} \begin{array}{c} \begin{array}{ccccccc} & k_1{=}0 & k_3{=}1 & & & k_2{=}2 & k_4{=}3 \end{array} \\ \left[\begin{array}{ccccccc} 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{5} & 0 \\ 0 & 0 & \mathbf{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}
$$

Now we can straightforwardly read out of (30) the number of markedness constraints ranked by $\pi_{\mathbf{X}}$ below each faithfulness constraint: F_1 has only one markedness constraint ranked below it, which is precisely the number $i_1 = 1$ which appears in the row corresponding to F_1 diminished by the value $k_1 = 0$ which corresponds to the column where that number appears; F_2 has three markedness constraints ranked below it, which is precisely the number $i_2 = 5$ which appears in the row corresponding to F_2 diminished by the value $k_2 = 2$ which corresponds to the column where that number appears; and so on.

Since $\mu(\pi_{\mathbf{X}})$ is defined in (9) as the sum over each faithfulness constraint of the number of markedness constraints ranked below that faithfulness constraint, we get the right result as in (31).

$$
\begin{aligned}
(31)\quad & \mu(\pi_{\mathbf{X}}) = \\
& = \mu(F_1) + \mu(F_2) + \mu(F_3) + \mu(F_4) \\
& = (i_1 - k_1) + (i_2 - k_2) + (i_3 - k_3) + (i_4 - k_4) \\
& = (1 - 0) + (5 - 2) + (2 - 1) + (6 - 3) \\
& = 8
\end{aligned}
$$

Note that the sum in the second line of (31) can be rearranged as follows:

$$
\begin{aligned}
(32)\quad & \mu(\pi_{\mathbf{X}}) = \\
& = (i_1 + i_2 + i_3 + i_4) - (k_1 + k_2 + k_3 + k_4) \\
& = (i_1 + i_2 + i_3 + i_4) - (0 + 1 + 2 + 3)
\end{aligned}
$$

It is trivial to check directly from the definition (12) of scalar product that the first term $i_1 + i_2 + i_3 + i_4$ in the second line of (32) is the scalar product $\langle \mathbf{\Sigma}_{7,4}, \mathbf{X} \rangle$ between the permutation matrix $\mathbf{X}$ in (27) and the matrix $\mathbf{\Sigma}_{7,4}$ in (14). Thus, the first term in the second line of (32) corresponds to the first term in (15). It is also trivial to check that the second term $0+1+2+3$ in the second line of (32) is equal to $\frac{1}{2}m(m-1)$ for $m=4$. Thus, the second term in the second line of (32) corresponds to the second term in (15).

A.2 Proof of claim 2

Consider a ranking π, namely a permutation over $\{1,\dots,n\}$. Let π^{-1} be its inverse. Recall that $\pi(i) = j$ means that constraint C_i is assigned by the ranking π to the jth stratum, with the top stratum being the one corresponding to $j = n$. Thus, $\pi^{-1}(j)$ is the constraint assigned by π to the jth stratum. Given an ERC $\mathbf{e} = [e_1,\dots,e_n]$, let $k = k(\mathbf{e}) \in \{1,\dots,n\}$ be univocally defined by conditions (33): they say that the constraints assigned by π to the top strata $k+1,\dots,n$ all have an e in the ERC $\mathbf{e}$ so that the constraint assigned by π to the kth stratum is the highest one that does not have an e in the ERC.

(33) a. $e_{\pi^{-1}(k+1)} = \ldots = e_{\pi^{-1}(n)} = e.$
b. $e_{\pi^{-1}(k)} \neq e.$

Thus, π is OT-consistent with the ERC $\mathbf{e}$ iff $e_{\pi^{-1}(k)} = \text{W}$. To prove Claim 2, I thus prove the equivalence (34), where $\mathbf{X}_\pi = [x_{i,j}]_{i,j=1}^n$ is the permutation matrix corresponding to π and $\mathbf{A}_\mathbf{e} = [a_{i,j}]_{i.j=1}^n$ is the matrix defined in (17).

(34) $$\langle \mathbf{A}_\mathbf{e}, \mathbf{X}_\pi \rangle > 0 \iff e_{\pi^{-1}(k)} = \text{W}.$$

Assume that $e_{\pi^{-1}(k)} = \text{W}$; then I can reason as follows, following Prince and Smolensky (2004):

(35) $$\begin{aligned}\langle \mathbf{A}_\mathbf{e}, \mathbf{X}_\pi \rangle &= \sum_{i,j=1}^n x_{i,j} a_{i,j} &&= \sum_{i,j=1}^n x_{i,j} 2^j t_i \\ &= \sum_{i=1}^n t_i \sum_{j=1}^n x_{i,j} 2^j &&= \sum_{i=1}^n t_i 2^{\pi(i)} \\ &= \sum_{j=1}^n t_{\pi^{-1}(j)} 2^j &&> 2^k - \sum_{j=1}^{k-1} 2^j > 0\end{aligned}$$

The proof of the reverse implication is analogous.

A.3 Proof of claim 3

To illustrate why claim 3 holds, consider the concrete case of the ERC $\mathbf{e}$ in (36).

(36) $$\mathbf{t} = \begin{bmatrix} \overset{C_1}{\text{W}} & \overset{C_2}{\text{W}} & \overset{C_3}{e} & \overset{C_4}{e} & \overset{C_5}{\text{L}} \end{bmatrix}$$

A ranking π is OT-consistent with this ERC $\mathbf{e}$ provided it ranks either C_1 or C_2 above C_5. This condition is equivalent to the set of implications (37). For example, the the third implication says that if, π assigns C_5 to either stratum 3, or 4 or 5 (the latter being the top stratum), then π must assign either C_1 or C_2 to either stratum 4 or 5.

(37) $$\begin{aligned} C_5 \in \{1,2,3,4,5\} &\Longrightarrow C_1 \in \{2,3,4,5\} \vee C_2 \in \{2,3,4,5\} \\ C_5 \in \{2,3,4,5\} &\Longrightarrow C_1 \in \{3,4,5\} \vee C_2 \in \{3,4,5\} \\ C_5 \in \{3,4,5\} &\Longrightarrow C_1 \in \{4,5\} \vee C_2 \in \{4,5\} \\ C_5 \in \{4,5\} &\Longrightarrow C_1 \in \{5\} \vee C_2 \in \{5\} \\ C_5 \in \{5\} &\Longrightarrow C_1 \in \emptyset \vee C_2 \in \emptyset \end{aligned}$$

Consider the permutation matrix $\mathbf{X} = [x_{i,j}]_{i,j=1}^{n=5}$. Recall that $x_{i,j} = 1$ iff the corresponding ranking π satisfies the condition $\pi(i) = j$ namely it assigns constraint C_i to the jth stratum. Thus, the implications in (37) can be restated in terms of permutation matrices rather than rankings as in (38), in the sense that a ranking π satisfies (37) iff the corresponding permutation matrix $\mathbf{X}_\pi$ satisfies (38). The five inequalities (38) can be written in

(38) $$\begin{aligned} x_{5,1} + x_{5,2} + x_{5,3} + x_{5,4} + x_{5,5} &\leq x_{1,2} + x_{1,3} + x_{1,4} + x_{1,5} + x_{2,2} + x_{2,3} + x_{2,4} + x_{2,5} \\ x_{5,2} + x_{5,3} + x_{5,4} + x_{5,5} &\leq x_{1,3} + x_{1,4} + x_{1,5} + x_{2,3} + x_{2,4} + x_{2,5} \\ x_{5,3} + x_{5,4} + x_{5,5} &\leq x_{1,4} + x_{1,5} + x_{2,4} + x_{2,5} \\ x_{5,4} + x_{5,5} &\leq x_{1,5} + x_{2,5} \\ x_{5,5} &\leq 0 \end{aligned}$$

(39) $$2x_{5,1}+4x_{5,2}+8x_{5,3}+16x_{5,4} \leq 2x_{1,1}+4x_{1,2}+8x_{1,3}+16x_{1,4}+32x_{1,5}+2x_{2,1}+4x_{2,2}+8x_{2,3}+16x_{2,4}+32x_{2,5}$$

(40) $$\begin{aligned} 4x_{5,1} + 4x_{5,2} + 4x_{5,3} + 4x_{5,4} &\leq 4x_{1,2} + 4x_{1,3} + 4x_{1,4} + 4x_{1,5} + 4x_{2,2} + 4x_{2,3} + 4x_{2,4} + 4x_{2,5} \\ 4x_{5,2} + 4x_{5,3} + 4x_{5,4} &\leq 4x_{1,3} + 4x_{1,4} + 4x_{1,5} + 4x_{2,3} + 4x_{2,4} + 4x_{2,5} \\ 8x_{5,3} + 8x_{5,4} &\leq 8x_{1,4} + 8x_{1,5} + 8x_{2,4} + 8x_{2,5} \\ 16x_{5,4} &\leq 16x_{1,5} + 16x_{2,5} \end{aligned}$$

(41) $$4x_{5,1} + 8x_{5,2} + 16x_{5,3} + 32x_{5,4} \leq 4x_{1,2} + 8x_{1,3} + 16x_{1,4} + 32x_{1,5} + 4x_{2,2} + 8x_{2,3} + 16x_{2,4} + 32x_{2,5}$$

matrix notation as $\langle \mathbf{A}_{\mathbf{e}}^{\bar{\jmath}}, \mathbf{X} \rangle \leq 0$ for $\bar{\jmath} = 1, \ldots, 5$.

A.4 Proof of claim 4

To illustrate why claim 4 holds, consider again the concrete case of the ERC (36). As just noted, the conditions $\langle \mathbf{A}_{\mathbf{e}}^{\bar{\jmath}}, \mathbf{X} \rangle \leq 0$ for $\bar{\jmath} = 1, \ldots, 5$ enforced by the relaxation (24) boil down to the inequalities (38). The condition $\langle \mathbf{A}_{\mathbf{e}}, \mathbf{X} \rangle \geq 0$ enforced by the relaxation (23) boils down to the inequality (39). In order to prove claim 4 in this specific case, I thus need to show that, if $\mathbf{X} \in \mathcal{P}_{\text{rel}}^{n}$ satisfies inequalities (38), then it also satisfies inequalities (39). Indeed, the last inequality in (38) says that $x_{5,5}$ is null, and can thus be dropped from the other four inequalities (38). Multiplying the first inequality in (38) by 4, the second by 4, the third by 8 and the fourth by 16, I get (40). Summing the inequalities (40) together, I get the inequality (41). As $x_{i,j} \geq 0$, I can weaken the inequality (41) by dividing the left hand side by 2 and by adding $2x_{1,1}$ and $2x_{2,1}$ to the right hand side, thus obtaining the desired inequality (39).

References

Sanjeev Arora, Alan Frieze, and Haim Kaplan. 2002. A New Rounding Procedure for the Assignment Problem with Applications to Dense Graph Arrangement Problems. *Mathematical Programming*, 92.1:1–36.

Dimitris Bertsimas and Robert Weismantel. 2005. *Optimization over Integers*. Dynamic Ideas, Belmont, Massachusetts.

Robert Berwick. 1985. *The acquisition of syntactic knowledge*. MIT Press, Cambridge, MA.

Janet Dean Fodor and William Gregory Sakas. 2005. The subset principle in syntax: costs of compliance. *Linguistics*, 41:513–569.

Bruce Hayes. 2004. Phonological Acquisition in Optimality Theory: The Early Stages. In R. Kager, J. Pater, and W. Zonneveld, editors, *Constraints in Phonological Acquisition*, pages 158–203. Cambridge University Press.

Jeffrey Heinz, Gregory M. Kobele, and Jason Riggle. 2009. Evaluating the Complexity of Optimality Theory. *Linguistic Inquiry*, 40:277–288.

P. W. Jusczyk, A. D. Friederici, J. M. I. Wessels, V. Y. Svenkerud, and A. Jusczyk. 1993. Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32:402–420.

René Kager. 1999. *Optimality Theory*. Cambridge University Press.

Giorgio Magri. 2010. Complexity of the Acquisition of Phonotactics in Optimality Theory. In Jeffrey Heinz, Lynne Cahill, and Richard Wicentowski, editors, *Proceedings of SIGMORPHON 11: the 11th biannual meeting of the ACL Special Interest Group on Computational Morphology and Phonology*, pages 19–27, Uppsala, Sweden. Association for Computational Linguistics.

Giorgio Magri. 2011. An online model of the acquisition of phonotactics within Optimality Theory. In L. Carlson, C. Hölscher, and T. Shipley, editors, *Proceedings of CogSci 33: the 33rd annual conference of the Cognitive Science Society*, Austin, TX:. Cognitive Science Society.

Giorgio Magri. 2012a. An approximation approach to the problem of the acquisition of phonotactics in optimality theory. manuscript available on the author's website; this is a longer version of the present paper.

Giorgio Magri. 2012b. Complexity of the acquisition of Phonotactics in Optimality Theory. Accepted at *Linguistic Inquiry*.

Giorgio Magri. 2012c. Restrictiveness of error-driven ranking algorithms: an initial assessment. Manuscript in progress.

M. Rita Manzini and Ken Wexler. 1987. Parameters, Binding Theory, and Learnability. *Linguistic Inquiry*, 18.3:413–444.

Joe Pater and Jessica A. Barlow. 2003. Constraint conflict in cluster reduction. *Journal of Child Language*, 30:487–526.

Alan Prince and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell. As Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder, and Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ, April 1993. Rutgers Optimality Archive 537 version, 2002.

Alan Prince and Bruce Tesar. 2004. Learning Phonotactic Distributions. In R. Kager, J. Pater, and W. Zonneveld, editors, *Constraints in Phonological Acquisition*, pages 245–291. Cambridge University Press.

Alan Prince. 2002. Entailed Ranking Arguments. ROA 500.

Bruce Tesar. 2008. Output-Driven Maps. ms., Rutgers University; ROA-956.

Roger Webster. 1984. *Convexity*. Oxford University Press.