RuCCS TR−1

June, 1993

In Defense of the Number *i*

Anatomy of a Linear Dynamical Model of Linguistic Generalizations

Alan Prince

Technical Reports
of the
Rutgers University
Center for Cognitive Science

RuCCS-TR

ISSN 1070-5996

Publication Program.

Technical Reports of the Rutgers University Center for Cognitive Science (RuCCS-TR) is a series of occasional papers reporting research in the cognitive sciences conducted by members of the Rutgers University Center for Cognitive Science and their collaborators.

Subscription Prices.

Individual issues are priced separately.

Inquiries.

All inquiries should be addressed to the editor: S. Bergelson, Rutgers University Center for Cognitive Science, Rutgers University, PO Box 1179, Piscataway, NJ 08855. Email inquiries may be addressed to admin@ruccs.rutgers.edu.

Photocopying.

Rutgers University Center for Cognitive Science
Rutgers, The State University
PO Box 1179
Piscataway, NJ 08855

RuCCS TR−1

In Defense of the Number *i*

Anatomy of a Linear Dynamic Model of Linguistic Generalizations

Alan Prince
Center for Cognitive Science
Department of Linguistics

Technical Report #1, June 1993
Rutgers University Center for Cognitive Science
Rutgers University
PO Box 1179
Piscataway, NJ 08855

Table of Contents

Preface

This report presents the results of an analytic investigation of a significant new approach to prosodic structure, the *Dynamic Linear Model* (Goldsmith 1991abc, 1992, in press; Goldsmith & Larson 1990; Larson 1992). The name displays the chief formal properties of the model: it is *dynamic*, because it involves a recurrent network which evolves in time, and *linear*, because the updating function is nothing more than a weighted sum of activations and biases. Linearity is crucial to the present enterprise, because it allows the model to be solved exactly. With an exact solution in hand, considerable progress can be made in determining the fundamental properties of the model. Most connectionist models have crucial nonlinearities, which are often directly responsible for their interesting behavior; but nonlinearity almost always entails the impossibility of exact solution, and the would-be analyst must use coarser methods to obtain a picture, often highly incomplete, of how the model behaves in general. The methods involve statistical approaches, and (far more commonly) extensive experimental probing. In this, there is a parallel to the methods typically used to explore linguistic theories: because of their intrinsic complexity, or merely because of disciplinary tendencies, theories are not infrequently explored through application to data problems, and analytic investigation of their structure and consequences is subordinated, postponed, or entertained principally in the context of encounters with factual material. With the *Dynamic Linear Model*, we are able to go beyond the usual limitations and achieve a surprisingly precise understanding of how the model parses reality.

The components of the argument have been arranged so as to maximizing accessibility. Part I, §§0-1, lays out the properties of the model in an essentially qualitative way; the aim is to characterize the behavior of the model and to measure it against what is known about the basic prosodic patterns of human language. The formal analysis supporting this discussion is presented in Parts II and III. Further formal analysis is found Part I, §3, and extension of the model from the discrete to the continuous occupies Part I, §2.

The Parts of the report were originally drafted and circulated in 1991 (Parts II & III) and 1992 (Part I). They have been lightly re-edited here. Additional references have been added to relevant work that has appeared in the interim.

I would like thank Paul Smolensky for valuable discussion of this and related material; his views on the analysis of connectionist networks have influenced the course of this enterprise. Thanks also to András Kornai for discussion and encouragement. Neither of these individuals should be charged with responsibility for any errors that may have crept into the text or the argument. I learned much about the Dynamic Linear Model and its promise from lucid presentations by John Goldsmith and by Gary Larson at the 1990 CLS meeting, at the University of New Hampshire conference on *Connectionism & Language* in May, 1990, and at the 1991 University of Illinois *Organization of Phonology* © Conference. The Mazer Fund of Brandeis University provided useful hardware. This research was supported by NSF Grant BNS-90 16806.

RuCCS TR-1
Part I

Remarks
on the Goldsmith-Larson Dynamic Linear Model as a Theory of Stress
with
Extension to the Continuous Linear Theory
And Additional Analysis

Alan Prince
Department of Linguistics, Rutgers,
Rutgers University Center for Cognitive Science

(Originally circulated January, 1992)

Abstract

Part I of this report characterizes and assesses the Goldsmith-Larson Dynamic Linear Model (DLM) as a theory of linguistic stress systems, building on the analytic results of Parts II and III. The discussion is qualitative, eschewing formal details, and oriented to evaluating the linguistic import of the DLM. A variety of significant properties are reviewed, but it is shown that the fundamental computational assumption of the model (linearity) leads to a many nonlinguistic behaviors in the models — for example, dependence on the absolute length of strings in determining the placement of stresses; and a completely gradual transition between LR→ and ←RL iterative systems. The second section shows that the DLM is a discrete approximation to a forced, more-than-lightly damped harmonic oscillator; in the Canonical Models, the damping is critical. The fundamental equation of the Critical Continuous Linear Theory of stress is stated. In the third section, formal analyis is presented in support of the new assertions in section one. Closed-form solution for the DLM's treatment of the vector $\Sigma = \Sigma \mathbf{e}_k$ is obtained in the Canonical Models and the solution space is classified. This vector is particularly significant in the economy of the model, in that it plausibly represents a string of a syllables undifferentiated as to weight, the syllabic substrate of the simplest class of stress patterns.

Remarks
on the Goldsmith-Larson Dynamic Linear Model as a Theory of Stress with Extension to the Continuous Linear Theory and Additional Analysis

Alan Prince

0. Introduction

0.0 Setting

Prosodic theory deals inevitably in notions of prominence, relative and absolute. The SPE theory of stress calculates with an integer-valued stress feature, generalized from structuralist analysis. Though often thought of as a absolute quantitative measure, the n-ary stress 'feature' really works with an *ordinal* ranking — [*n* stress] meaning in essence 'nth most prominent' stress in a domain.[1] Current stress theory offers two ways looking at at the role of prominence, with its notions of phonological constituency and the metrical grid (Liberman, 1975 et seq.); each of these contains both ordinal and absolute structures.

Within phonological constituents, the infinite, implicitly ordinal scale of SPE is shrunk to a binary ordinal contrast — strongest ('head') vs. anything else ('nonhead') — and concomitantly generalized to pertain to all units in the hierarchy of units. The ordinal scale thereby generated takes on some absolute qualities when the units of structure are given categorial status, like 'foot' and 'prosodic word', which have certain fixed properties. The Grid presents an explicit hierarchical layering of prominence ranks, delimiting the access of linguistic predicates (most notably 'adjacent') to a gradient structure. Here too absolute interpretation may be imposed when levels of the Grid are associated with categories of analysis. Principles of well-formedness may thus appeal to constituency-based notions ('head', 'sister', 'foot,' etc.) or to grid-based measures ('clash', 'lapse', etc.). The attempt to divine the interdependencies and empirical extension of these notions continues to inspire vigorous research to the present day. Goldsmith

[1] This re-interpretation accords with the way the scale runs; 1 is bigger than 2, because 1 means *first*, not '1 dollop' of some substance, of which 2 dollops could only be more. Furthermore, it makes sense of the SPE Stress Subordination Convention, which holds that assignment of [1 stress] automatically decrements all other stress-features in the domain by 1. From the stress-as-substance point of view this is puzzling indeed, mere algorithm grinding. Ordinally speaking, however, the 'assignment' of [1 stress] means declaring that a certain element is most prominent in the domain; every thing else takes a subordinate position in the lengthened queue. Whence are rationalized the 'numerical anfractuosities' that so perplexed Ladefoged and the magniloquent van der Slice. Of course, this interpretation is not that of the authors of SPE, who for example hold that a domain can contain [1 stress] and [3 stress] without [2 stress].

and Larson have recently put forth a model of prominence computation that differs considerably from familiar prosodic theory (Goldsmith and Larson 1990; Goldsmith 1991, 1992, in press; Larson 1992). The model involves iterative computation of real-numbered prominence values in a spreading-activation network. Because the model is dynamic, and because the calculation procedure involves linear equations, we will refer to it as the Dynamic Linear Model (DLM). In its current stage of development, the DLM does not aim to offer an account of the full hierarchy of prosodic structure in a single network. It is used to locate *peaks* of prominence in a sequence of units, making the contrast between nucleus and non-nucleus in the syllable when its units are taken to be segments, between stress and non-stress when its units are taken to be syllables. (If the units were regarded as stresses, the model would distinguish primary from secondary.) The conceptual affinity is therefore with the metrical grid (as indeed Goldsmith has frequently observed), though without the extended hierarchy; one might say that the DLM offers a fresh perspective on matters handled by two adjacent rows of the grid, the most basic structure of relative prominence. In particular, the theory holds out the promise of obtaining a smooth and principled transition from intrinsic prominence at one level (e.g. syllabic) to derived prominence at the next (e.g. stress) through its uniform, numerical treatment of prominence at all levels.

In Part I, we will explore the properties of the DLM as a theory of stress, drawing on and adding to the analytic results of Parts II and III. We will first present a qualitative assessment of the model's properties, avoiding formal details, so that readers can come to an understanding of the model's linguistic import without having to master its algebra. We then turn to formal analysis. We show that DLM is a discrete approximation to a critically or heavily damped harmonic oscillator, exhibiting the relevant differential equations, which shed considerable light on what the network actually accomplishes. We conclude with proof of the new claims made in the qualitative discussion.

First, some background.

0.1 How the DLM Computes

In the Goldsmith-Larson spreading-activation model of stress and syllable structure (DLM), the linguistic string is represented as a network of nodes with mutual interconnection between neighbors. The network for a string of length *n* can be pictured like this, using arrows to represent paths of influence:

(1) $N_1 \leftrightarrows N_2 \leftrightarrows \ldots \leftrightarrows N_{n-1} \leftrightarrows N_n$

Each node is endowed with an unchanging *bias*, which represents sonority or weight, conceived of as the intrinsic disposition of a segment or syllable to occupy a position of high prominence. Each node has an *activation* level, which (because of the way it is derived) takes into account not only the node's own bias but also the *activation* of the node's neighbors. The computation of activation is iterative: every node passes activation to adjacent nodes in each cycle of computation, and the cycles repeat forever. The effects of node-activation are modulated by *weights* on the links between nodes. Goldsmith and Larson postulate that all leftward links have the same weight (notated α); similarly all rightward links have the same weight (notated β). Weights may be positive, negative, or even 0. The activation of a given node is updated by weighting the activation on the adjacent nodes and summing the result with the node's own intrinsic bias. We can redraw the connection-diagram of the network to show the role of the link weights and biases:

(2) *Network Architecture*

$$\begin{array}{ccccccc} & \alpha & & \alpha & & \alpha & \\ \mathbf{N}_1 & \leftrightarrows & \mathbf{N}_2 & \leftrightarrows & \ldots & \leftrightarrows & \mathbf{N}_n \\ & \beta & & \beta & & \beta & \\ \uparrow & & \uparrow & & & & \uparrow \\ \boldsymbol{B}_1 & & \boldsymbol{B}_2 & & \ldots & & \boldsymbol{B}_n \end{array}$$

The updating procedure can be written out like this, using a_k for the activation level of node *k*, b_k for the intrinsic bias, α for the weight associated with leftward-moving activation, β for the rightward weight:

(3) $$a_k \leftarrow \alpha \cdot a_{k+1} + \beta \cdot a_{k-1} + b_k$$

This is a recipe for generating new activation levels for each node (left-hand-side), given the current levels (right-hand-side). Notice that the activation of node *k* itself (a_k) *does not enter at all* into the calculation of the new level for node *k*. One can think of *activation* as measure of a node's influence on its neighbors, and the fixed bias as a measure of a node's influence on itself.[2]

The calculation starts with all activations a_k set to 0. The update scheme (3) on the first cycle of iteration gives each node an activation equal to its bias; serious computation then begins. (Usually serious, that is: if both α and β are zero, there is of course no neighborly interaction at all; and if all biases are zero, all activations remain perpetually zero.) All activations are recalculated in each cycle of computation. But in favorable circumstances, it will happen that the set of activations will change less and less with each succeeding cycle, settling (in the limit, typically) on stable values that will repeat themselves without change. These stable activation values are the *output* of the network.

Since linguistic strings have various lengths and a network has but one, it is necessary to define a notion of *model* more abstract than network. Let a model $M_{\alpha\beta} = \langle \alpha, \beta, \mathbb{N} \rangle$, where $\mathbb{N}$ is the set of string-networks of all finite lengths. We are interested in what the model makes of every possible sequence of node biases. Let $\mathbf{b} = (b_1, \ldots, b_n)$, a vector (i.e. string) of biases, representing the assignment of some numerically-measured property to the syllables or segments in the string. Let **b*** be the string of activations ultimately attained by the computational procedure — the result of setting the model to work on **b**. We can write

(4) $$\mathbf{b}^* = M_{\alpha\beta}(\mathbf{b})$$

[2] To emphasize the fact that a node's current activation-level has *no* influence on the immediate update of its own activation, we might write: $a_k \leftarrow \alpha \cdot a_{k+1} + \mathbf{0} \cdot a_k + \beta \cdot a_{k-1} + b_k$.

The function $M_{\alpha\beta}$ produces its output by computing according the iterative scheme in eq. (3). For a model $M_{\alpha\beta}$ to make anything at all of a string of biases **b**, the iterative scheme must settle on a stable, finite value for the activation of each node. At this point of convergence or equilibrium, each new cycle of computation will produce an output exactly equal to its input. The activation of each node is stable and in a stable relation with the activations of its neighbors. The era of *becoming* has come to an end; history is over; a node's activation *is* a weighted sum of adjacent activations and self-bias. For such a 'fixed point' **b*** = $(a_1,...,a_n)$ we have

(5) $$a_k = \alpha \cdot a_{k+1} + \beta \cdot a_{k-1} + b_k$$

At convergence, the '←' is replaced by '='. The stable activations of the *n* nodes are described in *n* equations like (5), one for each node. The equations portray the final activations, the a_k's toward which a convergent networks tends, as a function of the network parameters α,β and the values of the input b_k's, fleshing out the import of eq. (4). The function associated with a model $M_{\alpha\beta}$ is a set of *n* linear equations in *n* unknowns, which can be solved explicitly, allowing us to use analytical methods to investigate the structure of the DLM (Parts II and III below). One useful basic result of such analysis is that we can determine exactly when parameters α and β will produce convergent networks. Any $M_{\alpha\beta}$ will converge for all inputs if, and only if, $|\alpha\beta| \leq 1/4$ (Part II below). Outside this region, models fail to settle, exploding to infinity, or under special circumstances, entering oscillatory regimes.

0.2 Models and Theories

The output activation sequence **b*** of a convergent model $M_{\alpha\beta}$ counts as a linguistic description when its numerical structures are interpreted with reference to linguistic constructs such as 'sonority','syllable', 'stress' and so on. For Goldsmith and Larson, it is the position of local *maxima* in the string, rather than absolute activation values, that determines the interpretation.[3] When the nodes are taken to represent segment positions, with the biases representing sonority values, then the local maxima in the output are interpreted as syllable peaks. If the network nodes are taken for syllables, with the biases representing weight or intrinsic (lexical) stresses or accents, then the local maxima in the output are stresses. A given model, plus a crucial interpretive component that finds maxima, maps a string of segments to a string of syllables, or a string of syllables to a stress pattern — or indeed any string of linguistic units with a numerical prominence structure defined on it — to a modified version of itself.

[3] It is necessary to refine the notion of *maximum* at play here, since networks can easily produce equality between adjacent nodes (v. Part III). While it might be sensible to regard a sequence like [1,2,1,1] as stressed on the second node, it is not plausible that [1,2,2,1] should be viewed as completely unstressed, on a par with [1,1,1,1]. We therefore introduce the notion of *quasi-maximum*: a node is a quasi-maximum if it's greater than at least one its neighbors but neither neighbor is greater than it. Every maximum is a quasi-maximum, and in [1,2,2,1] both of the 2's are quasi-maxima.

Each $M_{\alpha\beta}$ is a specific grammar of (an aspect of) prosody. The set of all such grammars then comprises a linguistic theory of that aspect of prosody. As with other such theories, there are two general claims to explore about the success of the theory:

(A) *Descriptive Inclusiveness*. For any actual linguistic prosodic system (stress, syllable structure), there is some α,β and some set of biases such that the patterns of the system are generated by $M_{\alpha\beta}$.

(B) *Predictive Validity*. Every parameter setting of α,β, and biases describes an authentic linguistic system.

Neither claim will hold up, of course, under scrutiny, but this is no reason to abandon the investigation. Exactly as with most other known theories, we work with relativized versions of the general claims.[4] The theory may not be descriptively inclusive (A), but it does offer new descriptions of complex phenomena (v. Goldsmith, Larson refs.). The theory may not be predictively valid (B), but it does offer interesting, unexpected entailments (Part III below). This will establish the significance of the approach in the minds of most serious researchers.[5] The DLM marks the first attempt to derive the characteristics of a rich, well-understood linguistic domain from the behavior of a dynamical system, and deserves investigation not only because of such low-level empirical successes as can be obtained, but because understanding it will point the way toward deeper models as yet unimagined.

1. Qualitative Characterization of Stress in the DLM

1.0 How Patterns are Built Up

Despite the existence of descriptive overlap (an empirical necessity), which may stir dire visions of 'notational variant' or 'mere implementation' in the minds of some thinkers, the leading ideas of the DLM are quite distinct from those of symbolic theories. The constraints of metrical theory are imposed by what amounts to Boolean logic. 'Do not place a new entry in the grid *if* it would be level-adjacent to another entry.' The DLM, by contrast, works through *addition*. A lone stress causes activation to spread throughout the string it sits in. If there are several stresses — positive biases — in the string, the global result is exactly the *sum* of the very activation-patterns caused by each independently, in the absence of the others. A stress does not see other stresses, does not influence them in the way they send out their activation.[6]

---

[4] According to reliable authorities, Quantum Electrodynamics does not require this dispensation. The downside is that no one seems to know what the theory is *about* — a small price to pay in the circumstances.

[5] This point of view, which emphasizes the value of ideas, is of course familiar from Chomsky's remarks over the years. Dissent from it is common in practice, under the constraints of advocacy and trepidation.

[6] This is consequence of the linearity of the equations defining the computation performed in the DLM. Such behavior is of course characteristic of many familiar wave phenomena, which are controlled by linear equations; see §2 below for analysis of the DLM as a wave generator.

The metrical grid can be thought of as a Boolean network; a given node does not work from a weighted *sum* of the state of its neighbors and its own intrinsic bias, but evaluates this sort of information according to a scheme built up from Boolean connectives. This is not commensurable with the additive method. To see this, consider the effect of a lexically-prespecified stress in a string. Assume that an iambic (minimum-first) pattern is to be developed from left to right (LR→); assume that the the fixed stress is in an odd position. Let us write χ for the lexical stress and its consequences. The following grid fragment would result

	x		x	χ		χ...	
x	x	x	x	χ	x	x	x...

The LR→ unfolding of the pattern results from a local interaction between adjacent grid positions. What's important is that the fixed lexical stress puts an absolute end to the influence of the stress that precedes it; a new calculation begins.

In the DLM, there is no such curtailing of influence. Each stress sends its activation out into the unlimited distance, bouncing and echoing off the ends of the string forever, and that wave of activation rolls through anything in its way. A lexical stress will make itself felt because its own wave *adds* to the others, not because it inhibits their propagation.

The DLM is thus capable of very significant long-distance effects. If for example, the leftward weight α is 1 while the rightward weight β is 0, the bias of any node is simply copied onto every node to its left. If the leftward weight α is -1, then positive and negative copies of a node's bias spread leftward in an alternating pattern, the immediate neighbor receiving a negative copy, the next one over receiving a positive copy (since $-1 \times -1 = +1$), and so on. A high-activation node will strongly suppress the ultimate activation achieved by alternate nodes to its left.[7] To see this at work, consider the following example, in which $\beta = 0$ (no rightward transmission at all) and $\alpha = -1$ (giving alternating waves going leftward).

(6) *Wave Cancellation*

$\alpha = -1$, $\beta = 0$	N_1	N_2	N_3	N_4	N_5	N_6
Biases	0	0	1	0	0	1
Wave from N_3	1	−1	1	0	0	0
Wave from N_6	−1	1	−1	1	−1	1
SUM of Waves	0	0	0	1	−1	1

The alternating wave pumped by the bias on Node 6 cancels the wave associated with Node 3.

[7] This raises an interesting question: is there a setting of parameters such that a node can annihilate itself? The answer is no. We need only consider the case where the network has just one non-zero bias. Recall that the initial condition of the net is all activations 0. If in the process of iteration we arrive at a state where all activations are 0, we are back at the beginning and are doomed to repeat the cycle endlessly. But for $|\alpha\beta| \leq ¼$, the network converges on a fixed output and has no oscillatory regimes.

Phenomena of this character have not been noted in the linguistic domain, of course. The DLM (and numerical approaches generally) offer exponential decay of amplitude with distance as a way of controlling such influences. As just seen, and as will be seen in more detail in §2 below, the rate of decay that can be managed by the DLM is not always dazzlingly fast (indeed it is not simply exponential, but somewhat faster or slower that an exponential with the same decay factor would be). Further, the DLM is equally capable of expressing exponential *explosion* of influence with distance. This fact merely highlights the generality of the explanatory problem that is evident in table (6): not every region of the parameter space is one that stress patterns live in.

The fundamental conceptual issue is whether stress patterns ever add — not only in the dramatic sense of total cancellation, but in any sense at all. This marks an important dividing line between the symbolic paradigm and the particular quantitative approach embodied in the DLM. We will argue that the evidence, from relatively subtle details of the DLM, indicates that they do not. However, in pursuit of this question we will uncover a variety of unexpected properties, some of which offer new perspectives on classical descriptive problems.

1.1 Culmination and the Barrier Models[8]

Since complicated input is processed by summing its simple components, it is instructive to examine the very simplest building blocks out of which complex structures can be constructed. These are the bias-strings that are zero everywhere except for a single 1. Let us use the notation ${}_n\mathbf{e}_k$ to represent a string of length *n*, with 1 as bias on the k^{th} node, zero bias elsewhere.[9] It should be clear that from the full set of such basic strings, any bias sequence whatever can be built up by addition and by multiplication by a numerical scaling factor.[10] For example, the activation string (2, -3) is just 2×(1,0) + -3×(0,1). Once we understand how the DLM treats the $\mathbf{e}_k$'s, we are well-positioned to understand its general behavior.

These basic constructional units can be thought of as syllable strings with a single lexical accent. An input string with two accents such as σ́σσσσ́σ is just the sum σ́σσσσσ + σσσσσ́σ.

Numerically, this is (1,0,0,0,0,0,0) + (0,0,0,0,1,0). The result of applying any model to σ́σσσσ́σ is exactly the same as applying the model to σ́σσσσσ and to σσσσσ́σ separately and then summing the individual results. Furthermore, the influence of an accent can be magnified or diminished or inverted by multiplying the basic string by some constant factor, say 1.2×(σ́σσσσ) or −3×(σ́σσσσ). We can mix multiplication and addition to get objects like σ́σσσσ + 1.2×(σσσσ́σ). In this way a string of syllables with any conceivable numerically-representable structure can be analyzed as the weighted sum of the basic one-accent strings, the $\mathbf{e}_k$'s, and the processing in the models respects this analyis completely.

---

[8] The formal analysis supporting the assertions made here is found in Part III below.

[9] We will be able suppress the pre-subscript, fortunately.

[10] To add strings, add the elements in corresponding ordinal positions, first with first, second with second, and so on, just as in table (6). To multiply a constant times a string, multiply each position in the string by the constant.

Let's use the notation $\mathbf{e}_k$* to represent the result of processing $\mathbf{e}_k$ in a some model $M_{\alpha\beta}$. The model $M_{\alpha\beta}$ itself can be thought of as a kind of stress rule; the string $\mathbf{e}_k$* is the output of the rule, given an underlying form $\mathbf{e}_k$. (One caveat: the absolute activation values have no meaning, only the location of quasi-maxima among them; the actual output of the DLM should be a string of, say, 0's and 1's, demarcating the quasi-maxima.) Output based on complex input is analyzable as the weighted sum of output based on simple inputs: if $(b_1, b_2, ...)$ is a string of biases, then we can write this basic observation down as follows:

(7) *Analysis of Complex Input*

$$(b_1, b_2, \ldots) = b_1 \times (1, 0, \ldots) + b_2 \times (0, 1, \ldots) + \ldots = b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + \ldots$$

$$M_{\alpha\beta}(b_1, b_2, \ldots) = b_1 \times M_{\alpha\beta}(\mathbf{e}_1) + b_2 \times M_{\alpha\beta}(\mathbf{e}_2) + \ldots = b_1\mathbf{e}_1{}^* + b_2\mathbf{e}_2{}^* + \ldots$$

Knowledge of the characteristics of the $\mathbf{e}_k$* will therefore open the doors to understanding the nature of the DLM.

The fundamental properties of the $\mathbf{e}_k$*, established in Part III, are these:

1. *Alternation of Sign*. The node *k* has positive activation. If α is negative, then alternate nodes preceding *k* have negative activation. If β is negative, then alternate nodes *following* *k* have negative activation. In this way, binary alternation of stress emerges.

2. *Culmination*. If α and β are both positive, then all activation is positive, and there is one and only one maximum value (which may indeed be shared between two adjacent nodes). Such models assign one stress to a string (or at most two on adjacent syllables, if each quasi-maximum counts as a stress).

3. *The Barrier Property*. Exactly where the culminative maximum falls when α and β are positive is matter of some interest. With simple input like the $\mathbf{e}_k$, the result is particularly striking: For a given positive α and β, the maximum on $\mathbf{e}_k$* falls no further than a certain fixed distance from one edge.

We can think of the node *p* beyond which there is no surface accent as a kind of barrier to the transmission of influence. If the lexical accent lies at or *inside* the barrier — that is, on node *p* or between node *p* and the relevant edge — then the lexical accent itself is realized. (The unit bias on node *k* of $\mathbf{e}_k$ causes a a maximum to surface on node *k* of $\mathbf{e}_k$*.) If, however, the lexical accent lies beyond the barrier, the output maximum ends up on the barrier node itself and not on node *k*. Here, the bias on node *k* of $\mathbf{e}_k$ leads to a maximum on node *p* of $\mathbf{e}_k$*. As an example, consider the input-output map of a model with a barrier on node 3:

(8) *Barrier on Node 3*

Input		*Output*
σ́σσσσ	→	σ́σσ̲σσ
σσ́σσσ	→	σσ́σ̲σσ
σσσ́σσ	→	σσσ̲́σσ
σσσσ́σ	→	σσσ̲́σσ
σσσσσ́	→	σσσ̲́σσ

The Barrier Property is remarkable in a couple of respects. First of all, the location of the barrier depends only on α and β — the basic parameters of the model — and not at all on the length of the input string.[11] The barrier's position is measured in absolute terms from an edge, not by some function relativized to length. This is of course highly desirable, since dependence on the absolute length of a string is not observed in language. Second, the Barrier Property reflects a kind of subtle behavior that is common among stress and accent systems: accent recedes as far back as it can from an edge (typically the end) within some window (often something like 3 syllables), but within that window a lexical marking or special rule supersedes the recessionary trend. Stress or accent will never be found *outside* the window, beyond the barrier. Examples of this general type — blurring details— would include Greek, Latin, Pirahã, English, and Spanish (for recent discussion, see Kager 1992).

The Barrier Property emerges unexpected and uncoaxed from the basic design of the DLM, providing a kind of explanation-from-first-principles for a much-discussed phenomenon. Results of this character, without real parallel in competing systems, intensify the interest of the whole project.[12]

Further investigation of the Barrier Property indicates that the result is incomplete in various ways, however. First of all, a barrier can be placed on *any* node (measured from an edge), not just on the 2nd or 3rd unit from the edge, the commonly-encountered positions. To encourage a feel for this, let's look at some actual parameter settings and their effects. The important factor in barrier-placement turns out to be the (positive) square root of the *ratio* of α and β, $(\alpha/\beta)^{1/2}$, which we will call r. If we set $\alpha\beta = 1/4$, we get very simple solutions to the DLM, which allow for explicit statement of the conditions on barriers. Let's call all $M_{\alpha\beta}$ for which $\alpha\beta = 1/4$ the 'Canonical Models'. (Note that fixing the product of the two parameters leaves their ratio completely free to vary, so the full range of behavior is still exemplified; no generality is lost in focussing attention on the Canonical Models.) The model with a barrier at p we will call a 'p-Model'. The following table shows how things come out:

[11] Whence the celebratory title *Theorema Egregium* applied to Thm. 13 of Part III below.

[12] The Barrier Property entails that the input-output function is not invertible: given the output array marking the location of extrema, one cannot in general point to a single input that would underly it, even when one knows α and β. For example, surface accent on the 3rd syllable in a 3-Model could come from any $\mathbf{e}_k$ with $k \geq 3$. In natural language systems of this sort, the position of underlying accents in morphemes and strings of morphemes is arrived at by paradigmatic (not syntagmatic) arguments, often complex.

(9) *Barrier Location in the Canonical Models, measured from the beginning*

Model #	Range of *r*	Range Length
1-Model	∞ *to* 2	∞
2-Model	2 *to* 1.5	1/2
3-Model	1.5 *to* 1.33^{+}	1/6
4-Model	1.33^{+} *to* 1.25	1/12
j-Model	j/(j−1) *to* (j+1)/j	1/ j(j−1)
∞-Model	1	0

A entirely parallel situation obtains for cases where the barriers are reckoned from the end rather than the beginning of the string. Let's call these the (−*p*)-Models.

(10) *Barrier Location in the Canonical Models, measured from the end*

Model #	Range of *r*	Range Length
−1-Model	0 *to* .5	1/2
−2-Model	.5 *to* $.66^{+}$	1/6
−3-Model	$.66^{+}$ *to* .75	1/12
−4-Model	.75 *to* .8	1/20
−j-Model	(j−1)/j *to* j/(j+1)	1/ j(j+1)
−∞-Model	1	0

The endpoints of the ranges for the finite barriers are not to be included in the range of *r*. (At the endpoint between the k-Model and the (k+1)-Model, nodes *k* and *k+1* share the maximal value.) In the ∞-Model, every $\mathbf{e}_k$* has its maximum at *k*, since every *k* is less than ∞, so that the surface form exactly reflects the lexical specification. The −∞-Model behaves identically.

The range of *r* runs from 0 to ∞ (excluding the endpoints). But all the action in the Canonical Models takes places in the range ½ — 2. For *r* < ½, every $\mathbf{e}_k$* has final stress; for every *r* > 2, stress is initial on every $\mathbf{e}_k$*.

Notice that the length of the range *decreases* as the barrier recedes from the edge (here, from the beginning of the string). It might therefore be possible to address the problem of unlimited barrier-distances — to establish the primacy of the 1-,2-, and 3-Models — by coarsening the DLM's ability to set parameter values, introducing some quantization into the model, as it were. Also worth exploring would be the possibility of arranging things so that the primacy of the lower distances was a mere statistical artifact.

A second, deeper problem is that the Barrier Property holds only of the processing of the $\mathbf{e}_k$ and not generally over all input. With prosody in mind, one would hope that in the case of multiple accents, the effect would be to maximize the leftmost or rightmost inside the barrier. But nothing of the sort emerges. Instead, the presence of even one other stress can introduce significant effects of string-*length* into the calculation. Consider the set of input strings beginning σ́σσ́.... Setting $\alpha\beta = ¼$, and $(\alpha/\beta)^{½} = 1.4$, to produce a 3-Model, we find the following input-output map:

σ́σσ́	→	σ́σσ
σ́σσ́σ	→	σ́σσσ
σ́σσ́σσ	→	σσ́σσσ
σ́σσ́σσσ	→	σσ́σσσσ

Surface accent hits the leftmost underlying accent for 3- and 4- syllable strings, but settles on an unfortunate compromise between the underlying accents in longer strings: syllable 2, underlyingly unaccented but sitting right between the two basic accents.

A sense of how this happens can be garnered from a direct comparison of the inner workings of the 4- and 5-syllable cases, presented in the following two tables (all values rounded).

(11) *Four Syllables*: $\alpha\beta$ = ¼, r = 1.4. Input: /σ́σσ́σ/

Biases:	1	0	1	0
Wave from N_3	1.57	2.24	**2.40**	.86
Wave from N_1	**1.60**	.86	.41	.15
***SUM* of Waves**	**3.17**	3.10	2.81	1.00

(12) *Five Syllables*: $\alpha\beta$ = ¼, r = 1.4. Input: /σ́σσ́σσ/

Biases:	1	0	1	0	0
Wave from N_3	1.96	2.80	**3.00**	1.43	.51
Wave from N_1	**1.67**	.95	.51	.24	.09
***SUM* of Waves**	3.63	**3.75**	3.51	1.67	.60

This length-dependent contrast is a direct consequence of the additive mechanism that powers the DLM. Its failure to match reality provides us with compelling evidence that stress patterns do not add together (as indeed Boolean symbolic models predict). We are left with the conclusion that the Barrier Property is a remarkable result which points in the direction of new modes of explanation, although richer dynamical assumptions are evidently required to actually arrive at a sound alternative to current theory.

1.2 Quantity-Insensitivity

Analyzing the behavior of the mono-accentual $\mathbf{e}_k$'s lays the foundation for understanding complex patterns. Among these, one is of obvious interest: the string in which all biases are equal. This provides the natural representation for quantity-insensitive prosody, in which the input string is analyzed as a sequence of undifferentiated syllables. Since there is no reason to consider any other activation level besides 1, let us focus on a string we will call Σ, in which every unit has 1 as bias. The string Σ is the sum of all $\mathbf{e}_k$'s of a given length. We will base our assertions on formal analysis of the Canonical Models, where $\alpha\beta = ¼$. (Details are found in §3 below.) Recall that it is the *ratio* not the *product* of α and β that is crucial to determining the effect of the model on its input. Imposing other conditions on the product $\alpha\beta$ adds nothing to the range of patterning of maxima, so long as $0 < \alpha\beta < ¼$.) Here again will write $r = (\alpha/\beta)^{½}$.

We will examine the two fundamental patterns that can be imposed on Σ^* — culmination in a single maximum value (positive α and β); and alternation of maxima (α and β negative). We will find a number of effects that are quite interesting in themselves, but which indicate that the extremal patterns of the DLM are rather different from those of linguistic prosody.

1.2.1 Culmination in Σ^*

For the $\mathbf{e}_k$* — the surface forms of the $\mathbf{e}_k$ — the location of the maximum is determined by r alone; hence the desirable independence from the string length n. In Σ^*, however, the culminative position is *always* a function of both r and n. As length increases, the effects of string length n diminish and indeed disappear in the limit: there the culmination principle becomes strictly a function of r and is identical to that relevant to the $\mathbf{e}_k$*. But for short strings (like those witnessed in languages), strong length effects are unavoidable.

The overall pattern works like this. For r less than ½, the maximum falls on the last unit of the string, just as[13] for the $\mathbf{e}_k$*. For $r = 1$, the theoretical maximum falls exactly at the *midpoint* of the string: $(n+1)/2$. This is, of course, an extreme case of length-dependence. If the string is of odd length, then there is a node sitting at this point which receives the maximum. If the string is of even length, then the abstract midpoint is flanked by two actual nodes, which share equally in the highest activation in the string. This behavior is highly nonlinguistic. No language has a stress rule putting stress right in the middle of the string; the notion 'exact

[13] Since with $r < ½$, we have end-stress on every $\mathbf{e}_k$*, with a steady rise to that maximum over the entire string, it is clear that adding up all the $\mathbf{e}_k$*'s of a given length will put a maximum on the last node of the sum.

middle' is surely not available to grammar. And no language has a rule putting *two* adjacent stresses across the midpoint of even-length strings.

If we fix r somewhere between ½ and 1, the maximum will lie between the end of the string and the midpoint. As string-length n increases, the maximum drifts toward its r-dependent limit position. Let us look at some examples to clarify how this works.

Suppose $r = .6$. The maximum is final for strings of length 2 and 3, and penultimate for longer strings.

Suppose $r =.7$. The maximum is final for length 2, penultimate for length 3-8, and antepenultimate for longer strings.

Suppose $r = .8$. The maximum is final for strings of length 2, penultimate for length 3-5, antepenultimate for length 6-10, pre-antepenultimate for length 11-82, and finally fifth-from-the-end for length 83 and above.

As r runs between 1 and ∞, the same pattern repeats in mirror image, reckoning maximal position from the *beginning* of the string to the midpoint.[14]

These findings show that Σ is essentially unmanageable as a model of a syllable-string that is to receive one accent by phonological rule encoded in α, β values. The accent cannot be made to sit still as length varies, except at the very edges.

1.2.2 Alternating Patterns in Σ^*

When α and β are both negative, each $\mathbf{e}_k{}^*$ shows a binary alternating pattern of positive and negative activations, anchored at node k, which is positive. Adding up all the $\mathbf{e}_k{}^*$ to get Σ^* gives rise to alternating patterns of maxima and minima (with exceptions to be discussed below), providing a natural model of familiar patterns of alternating stress. Here considerable regions of length-independence are to be found.

As in standard prosody, there are crucial differences in behavior that depend on the parity (oddness/evenness) of the string, rather than on its absolute length. These will allow us to construct an account of the four basic alternating types: iambic and trochaic, right-to-left and left-to-right.

The basic facts are these:

- (i) Odd-length Σ^* are stable throughout the entire range of r: they have maxima on all odd-numbered nodes.
- (ii) Even-length Σ^* show three classes of behavior:
  - (a) For r greater than approximately 1.211, maxima fall on even-numbered nodes.
  - (b) For r less than approximately .826, maxima fall on odd-numbered nodes.
  - (c) There is a transitional region between .826 and 1.211 (roughly) in which there are length-dependent failures of strict alternation.

[14] Here again, most of the road to ∞ is barren. The maximum falls on the first node for $r > 2$, regardless of string length. As with the $\mathbf{e}_k{}^*$'s, all the crucial action in the Canonical Models is compressed into the range ½ — 2.

Let's put aside the transitional region for the moment. In the secure area $r < .826$, all maxima in all strings fall on odd-numbered nodes. This is the correlate of trochaic, LR→, with mono-syllabic feet stressed. A more exact parallel is available in the pure grid theory of Prince 1983, which like the DLM does not recognize constituents: build the grid *Peak-First, LR→.*

In the secure area $r > 1.211$, maxima fall on even nodes in even-length strings, odd nodes in odd-length strings. Greater perspicuity is attained when we count node-numbers from the end rather than the beginning: then maxima fall always on odd-numbered nodes. (This is exactly the mirror image of the pattern for $r < .826$.) The approximate correlate in foot theory is: iambic, ←RL. Again the more exact parallel is the grid-theoretic rubric *Peak-First, ←RL.*

The processing of $\sum$ yields two alternating patterns, both peak-first, which would be assigned in opposite directions in a directional theory. Two other possible patterns remain: trough-first in either direction, in grid terminology; or iambic, LR→ and trochaic, ←RL (the nearest correlates in foot theory). These must be attained by applying the models to $-\sum$, i.e. by multiplying the just-described patterns by -1, which will exchange maxima and minima, as Goldsmith has suggested (Goldsmith in press).

The results of this survey can be tabulated as follows:

(13) *Classification of Binary-Alternating Systems*

	$\sum$*		$-\sum$*	
$r < .826$	Peak-First ≈ trochaic	LR→	Trough-First ≈ iambic	LR→
$r > 1.211$	Peak-First ≈ iambic	←RL	Trough-First ≈ trochaic	←RL

This system is of course heir to the complaints registered against the original grid-theory, which induced the same classification. At issue is whether the rows and the columns of the table define natural groupings (Hayes 1985). The vertical category *Peak-First* (or, equivalently, $\sum$*-based) mixes iambic and trochaic, as does the category *Trough-First* (or $-\sum$*-based). Similarly, the horizontal category LR→ (small *r*) mixes the two rhythmic types, as does the category ←RL (big *r*). Modern theory has insisted on the primacy of the iambic/trochaic distinction, which is lost in grid-type classifications that recognize only the whole string and not the foot as the essential prosodic domain. Regardless of such deeper empirical problems, it is notable that the DLM is able to generate a fair facsimile of the range of alternating systems, and in a stable, length-independent way.[15]

More interesting, perhaps, and more disconcerting is the existence of the transitional region between the two stable zones. The greater interest springs from the fact that the

[15] Multiplying bias by -1 cannot be recommended as a generally allowable procedure, however, inasmuch as it would turn high-bias items like heavy syllables into rejectors rather than attractors of stress.

transitional region is a distinctive property of the DLM not shared by known symbolic theories, and furthermore the kind of property that is intrinsic to the mechanics of the DLM. Because the DLM is linear, small changes in parameters will result in small changes in behavior; there can be no catastrophes, singularities, or sudden reversals. Consequently, two opposite forms of behavior — for example, what is describable as LR→ vs. ←RL stressing of string — must always be linked by a gradual transition, as one set of additive wave components grows in strength relative to its competitors.

In the transitional region, there is a one-by-one flipping of maxima and minima as *r* increases, going through a stage of complete equality between each adjacent pair of nodes. (Recall that odd-length strings are stable, so only even length strings are affected.) This process is portrayed iconically in the following display, which follows the transition of ∑* from LR→ Peak-First to ←RL Peak-First.

Transition in 6-unit ∑*						LR→	←RL	*r*
X	*x*	*X*	*x*	*X*	*x*	6	0	Small *r*
X	*x*	*X*	*x*	*X*	**X**	5	1	↓
X	*x*	*X*	*x*	**x**	X	4	2	↓
X	*x*	*X*	**X**	x	X	3	3	*r* = 1
X	*x*	**x**	X	x	X	2	4	↓
X	**X**	x	X	x	X	1	5	↓
x	X	x	X	x	X	0	6	Big *r*

The string passes through every mixture of left-to-rightness and right-to-leftness. In the process, various patterns with adjacent quasi-extrema are generated. A particular interesting case occurs when *r* = 1; the equal quasi-extrema straddle the exact midpoint of the string.

A further important characteristic of the transitional region is *length-dependence*: a given setting of *r* can produce rather different patterns in strings of different length. For example, take *r* = 1.2. The general pattern is Peak First, ←RL (≈iambic). But in strings of length 6 and 8 the first 'foot' is inverted — that is, these lengths show a mixture of (what are usually thought of as) directionalities: the first two units are LR→, the remaining units ←RL. The set of patterns looks like this:

(14) *Length Dependence in the Transitional Region of the Parameter Space*

Length	Stress Pattern
2	σσ́
4	σσ́σσ́
6	σ́σσσ́σσ́
8	σ́σσσ́σσ́σσ

10 σσ́σσ́σσ́σσ́σσ́

Empirically, of course, the kind of behavior seen in the transitional region is unattested, both in the patterns themselves, with their adjacent quasi-extrema (particularly quasi-maxima), and in the dependence on absolute segmental length. Indeed, it is even-length strings that are stable in the natural world, with odd-lengths falling under special constraints that deal with unpaired syllables. The existence of the transitional region is a direct consequence of the model's design: its linearity. We are led to the conclusion that stress patterns do not, in fact, combine additively; and that the descriptive success of the basically Boolean-based symbolic theories entails that dynamical models require more complex design if they are to achieve a better match with reality.

1.3 Summary of Discussion

The DLM achieves significant success in modeling basic features of stress patterns in natural language. The models in which α and β are both positive give an account of culminative patterns, those with just one local maximum. For mono-accentual input, such models show the Barrier Property, which limits the location of surface accent to an edge-most window. Underlying accents at or beyond the barrier give rise to surface accent at the barrier, while an accent inside the barrier gets realized in its underlying position. Location of the barrier is counted from an edge, and is independent of string length. Unfortunately, the Barrier Property holds only of mono-accentual input; the presence of even one more underlying accent can lead to length-dependent pathologies.

The quirks of multi-accented input in the culminative regime are clearly visible in strings with uniform activation throughout. Examining Σ, the sum of all mono-accentual input strings, with activation 1 everywhere, we found that placement of the culminative maximum in the output Σ^* was thoroughly length-dependent. Memorably, for $r = 1$, the culminative accent falls right in the middle of the string and is not measured from an edge at all. For even-length strings this entails a shared quasi-extremum over the middle two nodes. The location of the maximum in Σ^* only becomes stable for long string-lengths; the exact length at which stability is achieved depends on r. With increase of length in shorter strings, there is a drift of the maximum toward its asymptotic position, the location of which is a function of r. Because of the phenomenon of length-dependent stress-placement, the input string Σ turns out to provide a poor model of quantity-insensitive culmination.

When α and β are both negative, the entire Prince 1983 classification of alternating patterns is generable in a length-independent fashion, if uniform negative activation is admitted. Between two regions of parametric stability, there is a transitional zone in which unattested double-extrema patterns — adjacent quasi-maxima (stresses) and and adjacent quasi-minima (unstresses) — are produced, in a length-dependent fashion.

The treatment of alternation raises a fundamental question. Since there are only 4 useful cells in the table — indeed, only 2 cells are differentiated by r-values, the other 2 being generated by the device of multiplying the input by -1 — do we really want or need continuously many parametric values to describe them? One clear justification for scalar parameters is the modeling of scalar reality: but the basic interpretive assumption — that only extrema are significant — quantizes the output of the model, binarily. There is, nevertheless, a more subtle justification, of a type that has been argued by Paul Smolensky in a broader context (Smolensky 1988, for example). Some types of behavior emerge from the very way that the model computes; their existence is diagnostic of a certain computational modes. Such behaviors can often be forced from competing models, but if they are inevitable, we praise the models that have them for achieving explanation-from-principle. We praise them, that is, when the behaviors mirror reality; in the contrary situation, we are likely to be more circumspect. The computational assumptions of the DLM lead to desirable results, like the Barrier Property and the ability to model alternation, but they leave undescribed many basic patterns (e.g. edgemost accent wins) and lead just as directly to length-dependence, to unattested gradual transitions between sharply defined categories, and to additive compromises that are not characteristic of real systems. Although we are forced to the conclusion that stress patterns do *not* add (as current theory predicts), we must recognize that the DLM marks a real advance in the direction of finding new principles of linguistic form, and therefore deserves careful analysis and vigorous extension.

2. The Continuous Linear Theory of Stress.

2.0 Background[16]

The DLM computes according to the following iterative scheme, repeated from (3):

(15) $$a_k \leftarrow \alpha \cdot a_{k+1} + \beta \cdot a_{k-1} + b_k$$

This can be more concisely formulated in vector notation:

(16) $$\mathbf{a} \leftarrow \mathbf{W}_n\, \mathbf{a} + \mathbf{b}$$

Here **a** is the vector of node activations (initially **0**), **b** is the vector of fixed biases (the input to the scheme), and $\mathbf{W}_n$ is an n×n tridiagonal matrix with 0 on the main diagonal, α on the supradiagonal and β on the subdiagonal. Suppose that $|\alpha\beta| \leq 1/4$, so that the iteration converges for any **b**. Let **b*** represent the ultimate activation vector (the fixed point) that the iteration settles on, given a particular **b**. In the limit, we have:

(17) $$\mathbf{b}^* = \mathbf{W}_n\mathbf{b}^* + \mathbf{b}$$

This can be re-written like this:

(18) $$(\mathbf{I} - \mathbf{W}_n)\mathbf{b}^* = \mathbf{b}$$

Or, more usefully,

(19) $$\mathbf{b}^* = (\mathbf{I} - \mathbf{W}_n)^{-1}\, \mathbf{b}$$

Inverting the matrix $(\mathbf{I} - \mathbf{W}_n)$ will immediately give the modeling function $M_{\alpha\beta}$ that associates inputs **b** with outputs **b***.

Since $M_{\alpha\beta}$ is linear it is sensible to inquire about its effects on the canonical basis vectors $\mathbf{e}_k$, which have 1 in the k^{th} coordinate and 0 elsewhere. It turns out that two formulas are required for $\mathbf{e}_{jk}{}^*$, the j^{th} coordinate of $M_{\alpha\beta}(\mathbf{e}_k)$: one for those coordinates *preceding* the k^{th}, and another for those *following* the k^{th}. (The formulas agree on $\mathbf{e}_{kk^*}$.)

Let us call the formula for the coordinates preceding *k*, the ‘Initial Wave’, since it can be thought of as presenting samples from a wave standing over the initial segment of the string, from node 1 to node *k*. Similarly, let us call the formula for the coordinates following *k*, the ‘Final Wave’.

The general solution involves a certain amount of complication, which we will sweep aside forthwith, but it is worthwhile to exhibit the formulas.

[16] Proofs of assertions made here about the DLM and further details are found in Part III below.

We introduce a parameter u, sensitive to the product $\alpha\beta$, defined as follows:

(20) $$4\alpha\beta \cosh^2 u = 1, \qquad u \in \mathbb{C}$$

Recall that we have a parameter r, sensitive to the ratio α/β.

$$r = \sqrt{\frac{\alpha}{\beta}}\ , \qquad \alpha\beta \neq 0$$

The parameter r is either positive real or i times a positive real. To state the solutions concisely, let us define the following functions:[17]

(21) $$U_k(u) = \sinh(ku) / \sinh(u)$$

We have then for $\mathbf{e}_{jk}*$,

(22) *Initial Wave.*

$$\vec{e}^{\,*}_{jk} = 2(\cosh u)\frac{U_{n-k+1}}{U_{n+1}}\ (\mathrm{sgn}\,\alpha)^{k-j}\ U_j\ r^{k-j}\ , \quad j \le k$$

(23) *Final Wave*

$$\vec{e}^{\,*}_{jk} = 2(\cosh u)\frac{U_k}{U_{n+1}}\ (\mathrm{sgn}\,\beta)^{k-j}\ U_{n-j+1}\ r^{k-j}\ , \quad j \ge k$$

Our interest will center on the conditions that prevail when α and β *agree* in sign; indeed, we will focus on the models obtained when the weight parameters are subject to the constraint $\alpha\beta = ¼$, which puts u at 0. These we refer to as the 'Canonical Models'. The relevant formulas for $\mathbf{e}_{jk}*$ simplify greatly: observe that as $u \to 0$, $U_k \to k$, eliminating all reference to sinh. Since the extremal behavior of the Canonical Models is identical in the relevant respects to that of the related models for which α and β are either both positive or both negative, and since only the extremal behavior has empirical interpretation, we lose nothing.

[17] These are just the $(k\text{-}1)^{st}$ Chebyshev polynomials of the second kind applied to the argument cos iu. The structure of the DLM is tied up in various ways with that of the Chebyshev polynomials — for example, the Chebyshev polynomial of the first kind T_{n+2} has its extrema at $\lambda_k/(4\alpha\beta)^{½}$, where the λ_k are the eigenvalues of $\mathbf{W}_n$. The figure n+2 shows up because the networks of DLM implicitly run between node 0 and node n+1, which always have 0 activation. Thanks to András Kornai for directing me to the Chebyshev polynomials.

In the Canonical Models, α and β are interdefinable, so there is really only one weight parameter, r.

(24) *The Weight Parameter in the Canonical Models*

$$r = \sqrt{\frac{\alpha}{\beta}} = 2\,|\alpha| = \frac{1}{2|\beta|}$$

The leftward weight is $\pm r$, the rightward weight is $\pm 1/r$, where $r > 0$, and the symbols $\pm$ are interpreted uniformly so as to ensure $\alpha\beta = +¼$. (We do not admit $r = 0$, i.e. $\alpha = 0$.) The basic iterative scheme for the Canonical Models comes out like this:

(25) $$a_k \leftarrow ½\,(\pm\mathbf{r})\cdot a_{k+1} + ½(\pm\mathbf{r})^{-1}\cdot a_{k-1} + b_k$$

The solutions for the models take these forms:

(26) *Initial Wave in the Canonical Models*

$$\vec{e}^{\,*}_{jk} = \frac{2\,(n-k+1)}{n+1}\, j\,(\pm r)^{k-j} \quad , \quad j \le k$$

(27) *Final Wave in the Canonical Models*

$$\vec{e}^{\,*}_{jk} = \frac{2\,k}{n+1}\,(n-j+1)\,(\pm r)^{k-j}, \quad j \ge k$$

2.1 The Continous Linear Theory

If all time is eternally present
All time is unredeemable.
—*Burnt Norton*

To move to the continous theory, we need to replace $j \in Z^+$ with $x \in \mathbb{R}$. We will also want to deal with $\rho =_{def} \log(\pm r)$ rather $\pm r$ itself. For conciseness, we write N for n+1. Separating out the bits that depend on x from those that do not, we have

(28) *Initial Wave*

$$\mathrm{I}(x, k) = \frac{2}{N}\,(N-k)\,e^{\rho k}\;\; x\, e^{-\rho x}$$

(29) *Final Wave*

$$\mathrm{F}(x, k) = \frac{2}{N}\,k e^{\rho k}\;\;(N-x)\,e^{-\rho x}$$

Notice that these both take the form of the product of a linear term with an exponential term.

It is worthwhile to examine the weight parameters. The network parameter $\pm r$ may be positive or negative. As is clear from equations (26) and (27), with $-r$ there is alternation in sign in $\mathbf{e}_k{}^*$ spreading out in both directions from node k (which is positive). The *amplitude* of the activation is the same, of course, as for $+r$. The $+/-$ split on r divides the world into single-maximum, culminative patterns ($+r$) and alternating patterns ($-r$).
When the weight parameter is $-r$, ρ is complex:

$$\rho \quad = \quad \log(-r) \quad = \quad \log r + i\pi$$

since $\log(-1) = i\pi$ (picking a handy branch of the logarithm function). The effect on the functions I and F comes from the exponential term, which expands as follows:

$$e^{-(\log r + i\pi)x} \quad = \quad e^{-(\log r)x}(\cos \pi x \; - \; i \sin \pi x)$$

For integer x, this boils down to $\cos n\pi = (-1)^n$, creating alternation of sign, as desired.

The parameter r also induces an important further split in behavior. For $r > 1$, the amplitude of the Final Wave falls steeply from k to n. This is in part because $r^{\,k-j}$ decreases as j grows. (Equivalently, the exponential $e^{-\rho x}$ in I and F decreases as x increases. Note that $\log r = \mathrm{Re}(\rho) > 0$.) Supporting the decrease in the exponential term is the fact that the linear term $(N-x)$ in the Final Wave is both positive and decreasing between k and n; hence the fall.

The Initial Wave, on the other hand, will show more interesting behavior: the exponential term still decreases, but the linear term — merely x — *increases*. The Initial Wave may therefore include a *maximum* in amplitude somewhere inside the span running from 1 to k.

If $r < 1$, this pattern of effects occurs in mirror image. This is evident from the fact the rightward weight factor $1/r$ is the reciprocal of the leftward weight factor r. As the leftward factor ranges from 1 to ∞, the rightward factor ranges from 1 to 0, and vice versa. In the log domain, the contrast is between positive $\mathrm{Re}(\rho)$, with $r > 1$, and negative $\mathrm{Re}(\rho)$, with $r < 1$. This perfect symmetry means that we only need to focus on one interval or the other. Results obtained in one interval transfer immediately to other, reversing the sense of the string (i.e. treating the end as the beginning and counting node-numbers from right to left.)

For simplicity of discussion, but without loss of generality, we impose the restriction $r \geq 1$. In the log domain, this means $\mathrm{Re}(\rho) \geq 0$. In short, we will be looking at the following two cases:[18] $\rho = \log r$ and $\rho = \log r + i\pi$, with $\log r \geq 0$.

---

[18] Recall from §1 that all the maximum-shifting as r varies is confined to the interval [½,2]. Outside that interval the maximum falls on the edge-most node. Taking a generous view of things, we need allow no greater range for ρ than [-1,1] to get all significant behavior. Eliminating mirror-image redundancy, we have $0 \leq \rho \leq 1$.

2.2 Towards The Continuous Theory

The functions I(x, k) and F(x, k) are immediately recognizable as solutions to the following differential equation:

(30) $$(D + \rho)^2 \psi = 0$$

The general solution is as follows:

$$\psi(x) = (c_1 + c_2 x)e^{-\rho x}$$

Arriving at I(x, k) and F(x, k) requires setting appropriate boundary conditions that determine the free constants c_1 and c_2. There are two important conditions. First, the Initial and Final Waves vanish at their outside boundaries, points $x = 0$ and $x = N$ respectively. Second, the two waves must agree at point k in the string.

(31) *Essential Boundary conditions*.
a. I(0,k) = F(N,k) = 0
b. I(k,k) = F(k,k) > 0

These conditions determine I and F up to a multiplicative constant; to get an exact result, we need to pick a value for I(k,k) = F(k,k). The choice here is of no great significance, since the *extremal* behavior of I and F is unaffected by multiplying them by a positive constant. If we want the solutions to be identical to the values computed by the discrete network, we must pick

$$F(k, k) = I(k, k) = \frac{2}{N}\, k\,(N-k)\, e^{\rho k}$$

— a choice that might not otherwise recommend itself.

Equation (30) describes the *critically-damped harmonic oscillator*. The prototypical physical model of a harmonic oscillator goes likes this: imagine a mass attached to an anchored spring; pull the mass some distance in the direction away from the rest point; at t = 0 release the mass. There is a function f(t) that describes the displacement of the mass after its release, a solution to a 2^{nd} order differential equation derived from Newtonian considerations. With damping, the mass does oscillate freely, but is itself in contact with some damping mechanism that applies a resistance ρ to the motion. In the case of *critical* damping, the differential equation that governs the oscillation simplifies to (30). The mass does not oscillate at all; displacement decreases steadily with increasing time and asymptotes out at the 0 or equilibrium position ($e^{-\rho t} \to 0$ as $t \to \infty$ for positive ρ). See fig. (1) at the end of Part I for a graph of this course of events.

The stress-theoretic application is quite different in character and requires a rather broader perspective on the critical damping function. The physical application assumes an initial state of affairs — the mass displaced and unmoving before release, i.e. f(0) = A, f′(0) = 0, and the equation describes what happens as time moves forward. By contrast, the stress application assumes that conditions are known at both edges of the domain under scrutiny (these being positions 0 and k for the Initial Wave, positions k and N for the Final Wave), and uses the equation *to calculate the location and amplitude of a displacement that would lead to these edge conditions*, given the value of ρ.

The effects of this 'displacement' must be conceived of as running both forwards and backwards in time. Suppose, for example, that we are looking at $\mathbf{e}_5$*. The parameter ρ can be chosen so as to put a maximum on the third node; $\rho = ⅓$ will do the job exactly. This maximum is the analog of the initial displacement that sets off the spring apparatus. But we track its effects not only forward in time to node k, but also backward in time through nodes 2 and 1, and indeed to node 0, where displacement vanishes.

Fig. (2) at the end of Part I shows the generic shape of an Initial Wave under the condition ρ real (and positive, as agreed on). It has three notable characteristics.

(a) I(x, k) has its maximum at $x = 1/\rho$.
(b) I(x, k) $\to -\infty$ as $x \to -\infty$.
(c) I(x, k) is strictly positive for all $x > 0$, approaching 0 as an asymptote as $x \to \infty$.

Property (b) is interesting: it shows that the 'damped' linear oscillator is damped in only one temporal direction. Running time backwards past the maximal displacement, the wave's amplitude decreases without bound as x ('time') heads toward $-\infty$. Observe that for complex ρ, with an alternating pattern, Fig. (2) traces the envelope of the relative maxima of the real part of the function.

Fig. (3) shows the generic shape of the *Final Wave* under the same conditions on ρ. The notable characteristics are these:

(a) F(x, k) has its *minimum* at $x = N + 1/\rho$.
(b) F(x, k) $\to \infty$ as $x \to -\infty$.
(c) F(x, k) is positive for all $x < N$. At $x = N$ it goes negative, and heads back toward 0 after its minimum, so that F(x, k) $\to 0^-$ as $x \to \infty$.

Note that the Final Wave has the same basic *shape* as the Initial Wave. To get the Final Wave's shape, turn the Initial Wave upside down (reflect through x-axis) and shift it rightward so that the 0-crossing is at N. (It is also multiplicatively scaled.)

We are now in a position to characterize the behavior of the DLM in terms of the the critical-damping equation (30). What the DLM computes, given $\mathbf{e}_k$, is implicitly the location and magnitude of a certain displacement and the evolution (both backwards and forwards in time) of a critically-damped linear oscillator, with amplitude-decay-factor ρ, to which that displacement is administered.[19]

The point at which this action-initiating displacement occurs is the extremum of each of the waves. For the Initial Wave, the crucial displacement hits at $x = 1/\rho$, and is positive in amplitude. For the Final Wave, the displacement occurs at $x = N + 1/\rho$, and is negative;

[19] Our one mild departure from the physical model is allowing the amplitude decay factor to be complex.

everything we see of the Final Wave runs backward from that displacement. (This description is based on our assumption that ρ is positive; for ρ < 0, invert the sense of the string.) The DLM computes the values of the waves for integral x. The DLM provides a discrete approximation to the underlying continuous functions, in the sense that the DLM obtains, at integral points, a numerical solution to the critical damping equation, to any desired degree of accuracy.

The shape of the waves is determined by ρ and N, and by the requirement that the value at k be positive. The point k of input $\mathbf{e}_k$ marks the point where we switch from the Initial Wave to the Final Wave; crucially, that is, where the linear term switches from positive slope to negative slope (given ρ ≥ 0). The DLM fits *two* harmonic oscillators to the boundary conditions imposed by $\mathbf{e}_k$. We cannot simply patch the two together at point k and claim a single solution to the equation, since the patched function is nondifferentiable at k. (Note the discontinuity in slope at k.) We therefore need to take one more step forward to arrive at a full understanding of the DLM from the differentiable point of view.

2.3 The Continuous Theory Made Smooth

‘‘And the rough places plain.’’

It is instructive to consider the vector $\Sigma = \Sigma\mathbf{e}_k$. For the j^th^ coordinate $\Sigma_j{}^*$ of Σ^* we have

$$\vec{\Sigma}_j = \sum_{k=1}^{n} \vec{e}_{jk}^{\,*}$$

To arrive at $\Sigma_j{}^*$ we add the j^{th} coordinates of all $\mathbf{e}_k{}^*$’s. As we sum over k, the process splits into two parts. In the first, k precedes (or equals) j, and j is in the Final Wave that comes after k; in the second part, k is beyond j and the node j is in the Initial Wave that comes before k.

$$\vec{\Sigma}_j = \sum_{k=1}^{j} \mathrm{F}(j,k) + \sum_{k=j+1}^{n} \mathrm{I}(j,k)$$

The transition to the continuous theory is obtained by considering all points x, $0 \le x \le n+1 = N$, instead of just the integer arguments dealt with by the DLM. (Notice that 0, N could just as well have been used as limits on the sums above, because the activation functions are 0 at these points.) This allows us to define a function $S(x|1)$ that computes the stress at x on the assumption that *every* point in the interval [0, N] has bias 1.

(32) *Stress Function with Uniform Bias 1*

$$\mathrm{S}(x|1) \quad = \quad \int_0^x \mathrm{F}(x,t)\,dt \;+\; \int_x^N \mathrm{I}(x,t)\;dt$$

Since we are integrating over *t*, we can pull out constants and terms dependent on *x* to arrive at the following:

$$\mathrm{S}(x) \quad = \quad \frac{2e^{\rho x}}{N}\left[\;(N-x)\int_0^x t\,e^{\rho t}dt \;+\; x\int_x^N (N-t)\,e^{\rho t}dt\;\right]$$

Applying the differential operator associated with critical damping we obtain:

$$(\boldsymbol{D} + \rho)^2\,S(x) \quad = \quad -2$$

This shows that the constant bias serves as *driving force* applied to the oscillator. More generally, if s(x) is forcing function defined for every point in the domain of S, we will have the following extension of eq. (32):

$$\mathrm{S}(x\,|\,\mathrm{s}) \quad = \quad \int_0^x \mathrm{F}(x,t)\,\mathrm{s}(t)\,dt \;+\; \int_x^N \mathrm{I}(x,t)\,\mathrm{s}(t)\;dt$$

Once again, this function is remarkably well-behaved under the critical damping operator:

$$(\boldsymbol{D}+\rho)^2\,S(x\,|\,\mathrm{s}) \quad = \quad -2\,\mathrm{s}(x)$$

This result establishes that network *bias* is the analog of a physical *force*; and that a bias vector **b** is a discretized version of a function s(*x*) which describes a time-varying driving force applied to the oscillator. The DLM, in full, is thus a discrete approximation to a forced, damped oscillator.

In moving to the continuous theory, we might as well drop out multiplicative factors that are irrelevant artifacts of the network calculation. The fundamental equation of the Critical Continuous Linear Theory of Stress then becomes the following:

(33) *Fundamental Equation — CCL* Θ

$$\boxed{(\boldsymbol{D}+\rho)^2\,\mathrm{S}(x) \quad = \quad -\,\mathrm{s}(x)}$$

Equation (33) can be rewritten so as to display the solution directly in terms of S. Recall that

$$e^{\rho x}\,(\boldsymbol{D}+\rho)^k\,\mathrm{f}(x) \quad = \quad \boldsymbol{D}^k\,e^{\rho x}\,\mathrm{f}(x)$$

This leads to the following reformulation, which some may find more perspicuous:

$$S(x) \quad = \quad -e^{-\rho x} D^{-2} e^{\rho x} \, \mathrm{s}(x)$$

Let us now turn to the matter of obtaining the desired solution directly from the fundamental equation. Let us consider only the case of constant bias, say $\mathrm{s}(x) \equiv \mathrm{k}$. By inspection, the general solution of the equation is this:

$$S(x) \quad = \quad (c_1 + c_2 x)\, e^{-\rho x} - \frac{k}{\rho^2}$$

The constants c_1 and c_2 are determined by the boundary conditions:

$$\mathrm{S}(0) \quad = \quad 0$$

$$\mathrm{S}(N) \quad = \quad 0$$

We have

$$c_1 \quad = \quad \frac{k}{\rho^2}$$

$$c_2 \quad = \quad \frac{k}{N\rho^2}\,(e^{\rho N} - 1)$$

This leads to the following:

(34) *Explicit form of S for Constant Driving Force*

$$\mathrm{S}(x) \quad = \quad \frac{k}{\rho^2 N}\left[\,(N-x)\,e^{-\rho x} + x\,e^{\rho(N-x)} - N\,\right]$$

It's worth noting that the heart of $\mathrm{S}(x)$ is function we can call $\psi(x)$, which has the same extremal behavior as $\mathrm{S}(x)$.

$$\psi(x) \quad = \quad (N-x)\,e^{-\rho x} + x\,e^{\rho(N-x)}$$

The function ψ solves the homogeneous critical damping equation, i.e. equation (33) with $\mathrm{s}(x) \equiv 0$, under the boundary conditions $\psi(0) = N$, $\psi(N) = N$. More generally, if y_s is a particular solution of the fundamental equation for some choice of $\mathrm{s}(x)$ and it happens to be the case that $y_s(0) = y_s(N)$, then we will have

A glance at the fundamental equation (33) raises an obvious question: why is the driving

$$S(x) \quad = \quad -\frac{y_s(0)}{N}\,\psi \;+\; y_s$$

force *negated*? Some insight may be obtained by examining equation (34), which gives the explicit formula for constant forcing. Observe that as x $\rightarrow \infty$, S(x) asymptotes out at $-k/\rho^2$, not at 0. Recall the general shape of the critical damping curve as seen in figs. (2, 3) for the Initial and Final Waves: the curve crosses 0 once and never crosses it again, though it approaches it asymptotically. (More generally, the critical damping curve crosses its asymptote line once and then returns to it in the limit.) But the boundary conditions on S(x) require that it cross 0 twice, once at x = 0, once at x = N. This requires forcing, i.e. pushing the basic curve, which is described by ψ, *down*. A positively forced damped oscillator will asymptote out at some positive value and never reach zero at all; a negatively-forced critically-damped oscillator will asymptote out at some negative value, crossing zero to reach it. Notice that the negative driving force acts, nonetheless, as a positive multiplier on the value of S(x). This is because of the task that is being performed by the equation: it calculates the location of a maximal displacement, and the evolution (forwards and backwards in time) of the system that has that particular displacement. With negative forcing, the maximal displacement must be all the greater so that the amplitude reaches 0 only at the boundaries 0, N and not before.

Critical damping in the physical world precludes oscillation; damping there must be ‘light’ to allow it. Under light damping, the amplitude of a sinusoidal wave is subject to exponential decay. This cannot happen in the stress theories we are examining. The nonsinusoidal factor is itself not simply exponential but is rather the product of a linear function and an exponential decay function. Alternation of stress — actual oscillation — occurs here under the control of the critical damping equation when the amplitude decay factor ρ is complex. Because log ($-r$) is a periodic function, oscillation emerges in the critical damping scenario.

We conclude with the observation that the *noncritical* DLM, for α and β agreeing in sign but not necessarily bound by the constraint $\alpha\beta$ = ¼, satisfies the equation

(35) *General Equation for* $\alpha\beta > 0$

$$\left[\boldsymbol{D}^2 \;+\; 2\rho\boldsymbol{D} \;+\; (\rho^2 - u^2)\right] S(x) \quad = \quad \frac{-u}{\sinh u}\, s(x)$$

where

$$\rho \quad = \quad \log\sqrt{\frac{\alpha}{\beta}}$$

$$u \quad = \quad \cosh^{-1}\frac{1}{2\sqrt{\alpha\,\beta}} \quad = \quad \log\frac{1+\sqrt{1-4\alpha\beta}}{2\sqrt{\alpha\beta}}$$

Note that the general equation becomes identical to the one examined above when $u = 0$, i.e. when $\alpha\beta = ¼$. The differential operator in the generalized equation is factored as $(D+\rho+u)(D+\rho-u)$. Consequently, the generic solution is

$$c_1 e^{-(\rho+u)x} + c_2 e^{-(\rho-u)x} = e^{-\rho x}\left[c_1 e^{-ux} + c_2 e^{ux}\right]$$

Appropriate boundary conditions will turn the bracketed expression into the desired sinh terms in eqs. (22) and (23).

The general equation (35) describes a *heavily damped* harmonic oscillator. Like the critically-damped special case, this device does not actually oscillate, but sinks from its maximum displacement toward its asymptote, usually 0 in the prototypical spring-mass-damper model. The point of interest is that the critically damped version makes the most rapid descent toward asymptote for a given value of the term (ρ^2-u^2). A Canonical Model with weight factor ρ_1 is thus being compared with another model with factor ρ_2, such that $\rho_1^2 = \rho_2^2-u^2$. The Canonical Models are optimal in the sense that they are doing the best that can be done, within the constraint of linearity, to avoid the consequences of a lingering decay of amplitude, and hence of additive interaction between emanations from different stresses in the same string. This best is apparently not good enough, as shown above, and linearity is clearly the culprit.

2.4 Summary

By examining the solutions to the DLM, we have determined that for $\alpha\beta > 0$ the DLM is a discrete approximation to a severely damped, forced harmonic oscillator. For $\alpha\beta = ¼$, the damping is critical, entailing the interesting extremal property of most rapid decay of wave amplitude. The parameter $\rho = \log r$ is the amplitude decay characteristic of the device, which may be complex, allowing for alternation of stress. The bias vector that serves as input to the DLM is revealed as the discrete version of a time-varying driving force that acts on the oscillator. This suggests moving from the DLM to the Critical Continuous Linear Theory of stress and syllable structure, whose fundamental equation is

(36) *Critical Continuous Linear Theory*

$$\left(\frac{d}{dx} + \rho\right)^2 \mathrm{S}(x) + \mathrm{s}(x) = 0$$

3. Formal Analysis of $\sum^*$ in the Canonical Models

Here we establish the results which lie behind the discussion of $\sum^*$ in §1 above. For convenience of reference, we repeat the explicit solutions for the coordinates of $\mathbf{e}_k^*$.

(37) *Initial Wave in the Canonical Models*

$$\vec{e}^{\,*}_{jk} = \frac{2\,(n-k+1)}{n+1}\; j\, r^{k-j} \quad , \qquad j \le k$$

(38) *Final Wave in the Canonical Models*

$$\vec{e}^{\,*}_{jk} = \frac{2\,k}{n+1}\;(n-j+1)\, r^{k-j}, \qquad j \ge k$$

To deal with $\sum^*$, we need first of all the familiar formula for summing a geometric series:

(39) *Geometric Series*

$$\sum_{k=0}^{n} r^k = \frac{r^{n+1}-1}{r-1}$$

We also need the perhaps less familiar formula:

(40) *Geometric-like Series with Special Coefficients*

$$\sum_{k=0}^{n-1} (n-k)\, r^k = \frac{r^{n+1}-(n+1)r+n}{(r-1)^2}$$

Eq. (40) can be readily derived from eq. (39), given the following observation:

$$\sum_{k=0}^{n-1} (n-k)\, r^k = -r^{n+1}\frac{d}{dr}\sum_{k=0}^{n-1} r^{k-n} = -r^{n+1}\frac{d}{dr}\,\frac{1}{r^n}\sum_{k=0}^{n-1} r^k$$

Invoking equation (39) and performing the indicated computation yields eq.(40) .

We want to find Σ_j*, the jth entry in Σ*. We have

$$\vec{\Sigma}^* = \sum_{k=1}^{n} \vec{e}_k^{\,*}$$

To find Σ_j* we sum over the jth coordinates of all $\mathbf{e}_k$*. Since the relevant formulas differ depending on whether *j* falls to the left or right of *k*, we are dealing with two summations. One running from 1 to *j* ($k \le j$) covers the cases where *j* is in a Final Wave. The other running from *j*+1 to *n* covers the cases where *j* is in an Initial Wave.

$$\vec{\Sigma}_j^* = \sum_{k=1}^{n} \vec{e}_{jk}^{\,*} = \sum_{k=1}^{j} \vec{e}_{jk}^{\,*} + \sum_{k=j+1}^{n} \vec{e}_{jk}^{\,*}$$

The full formula looks like this:

(41) Σ_j* *as Sum of Sums*

$$\vec{\Sigma}_j^* = \frac{2}{n+1}\left[(n-j+1)\sum_{k=1}^{j} k \left(\frac{1}{r}\right)^{j-k} + j \sum_{k=j+1}^{n} (n-k+1)\, r^{k-j} \right]$$

To apply eq. (40), we need to adjust the indices of summation. For the first sum, let the new index $k_{new} = j - k_{old}$.

When $k_{old} = j$, $k_{new} = 0$.
When $k_{old} = 1$, $k_{new} = j - 1$.

The first sum-containing term in eq. (41) now computes out as follows:

$$\begin{aligned}
(n-j+1)\sum_{k=1}^{j} k \left(\frac{1}{r}\right)^{j-k} &= (n-j+1)\sum_{k=0}^{j-1} (j-k) \left(\frac{1}{r}\right)^{k} \\
&= (n-j+1)\frac{\left(\frac{1}{r}\right)^{j+1} - (j+1)\left(\frac{1}{r}\right) + j}{\left(\left(\frac{1}{r}\right) - 1\right)^2} \\
&= (n-j+1)\frac{r}{(r-1)^2}\left[r^{-j} + jr - (j+1) \right]
\end{aligned}$$

For the second summation-term in eq.(41), let the new index $k_{new} = k_{old} - j - 1$.

When $k_{old} = j + 1$, $k_{new} = 0$.
When k_{old} = n, $k_{new} = n - j - 1$.

The second sum-containing factor in eq. (41) now computes out as follows:

$$j \sum_{k=j+1}^{n} (n-k+1)\, r^{k-j} \;=\; j \sum_{k=0}^{n-j-1} (n-j-k)\, r^{k+1} \;=\; jr \sum_{k=0}^{n-j-1} (n-j-k)\, r^{k}$$

$$= \; j \frac{r}{(r-1)^2} \left[\, r^{n-j+1} - (n-j+1)r + (n-j) \,\right]$$

Combining these two results, we get

$$(n-j+1)\frac{r}{(r-1)^2}\left[\, r^{-j} + jr - (j+1) \,\right] + j\frac{r}{(r-1)^2}\left[\, r^{n-j+1} - (n-j+1)\,r + (n-j) \,\right]$$

$$= \; \frac{r}{(r-1)^2} \left[\, j r^{n-j+1} + (n-j+1)\, r^{-j} - (n+1) \,\right]$$

This gives us the bracketted terms in eq.(41). The final expression for $\sum_j^*$ is therefore:

(42) *Explicit Form of* $\sum_j^*$

$$\vec{\Sigma}_j^* \;=\; \frac{2r}{(n+1)(r-1)^2} \left[\, j r^{n-j+1} + (n-j+1)\, r^{-j} - (n+1) \,\right]$$

Since we are only interested in the extremal behavior of $\sum_j^*$ as a function of *j*, it is desirable to re-write eq. (42) so that all the terms dependent on *j* are on one side.

$$\frac{(n+1)(r-1)^2}{2r}\, \vec{\Sigma}_j^* + (n+1) \;=\; j r^{n-j+1} + (n-j+1)\, r^{-j}$$

It is worthwhile to name the expression that depends on *j*.

(43) *Definition of* Ψ

$$\Psi(r,j) \;=\; j r^{n-j+1} + (n-j+1)\, r^{-j}$$

The function Ψ(r,j) has, for fixed *r* and *n*, the same extrema as $\sum_j^*$. Note the relationship between Ψ and the formulas (37) and (38) for the Initial and Final Waves.

For the case where $\alpha,\beta < 0$, need merely observe that the terms $\mathrm{sgn}(\alpha)^{k-j}$ and $\mathrm{sgn}(\beta)^{k-j}$ in the formulas for $\mathbf{e}_j$* are equal to $(-1)^{k-j}$, and can be amalgamated with r^{k-j} as $(-r)^{k-j}$. The formulas for $\sum_j$* and Ψ can then be applied directly, with $-r$ for r. To reduce notational clutter, we introduce the symbol $\sigma_k = (-1)^k = +1$ for even k, -1 for odd k.

$$\vec{\Sigma}_j^* = \frac{-2r}{(n+1)(r+1)^2}\left[j\,\sigma_{n-j+1}\,r^{n-j+1} + (n-j+1)\,\sigma_j\,r^{-j} - (n+1)\right]$$

$$= \frac{2r}{(n+1)(r+1)^2}\left[j\,\sigma_{n-j}\,r^{n-j+1} + (n-j+1)\,\sigma_{j-1}\,r^{-j} + (n+1)\right]$$

The function $\Psi(-r,j)$ comes out like this:

$$\Psi(-r,j) = \frac{(n+1)(r+1)^2}{-2r}\vec{\Sigma}_j^* + (n+1) = j\,\sigma_{n-j+1}\,r^{n-j+1} + (n-j+1)\,\sigma_j\,r^{-j}$$

Note that $\Psi(-r,j)$ has maxima where $\sum_j$* has minima, and vice versa, due to the inversion of sign. Computation of maxima should for perspicuity be based on $-\Psi(-r,j)$, which for conciseness we will call Ψ*. This can be written conveniently as follows, making a slight adjustment in the σ-terms:

$$\boxed{\Psi^*(r,j) = -\Psi(-r,j) = j\,\sigma_{n-j}\,r^{n-j+1} + (n-j+1)\,\sigma_{j-1}\,r^{-j}}$$

Let us now consider the location of maxima for $\alpha,\beta > 0$. First, notice that $\Psi(r,j)$ has the following form, the product of an exponential in j with a linear expression in j:

$$\Psi(r,j) = r^{-j}\left[j(r^{n+1}-1) + (n+1)\right]$$

We need to examine the behavior of $\Psi(r,j)$ for fixed r, varying j.

$$\frac{\partial\Psi}{\partial j} = r^{-j}(r^{n+1}-1) - r^{-j}(\log r)\left[j(r^{n+1}-1) + (n+1)\right]$$

Imposing the condition $\partial\Psi/\partial j = 0$ yields the following:

(44) *Maximal point of* Ψ

$$\boxed{\frac{1}{\log r} - \frac{n+1}{r^{n+1}-1} = j_{\max}}$$

At $r = 0$, eq. (44) is not defined, but determination of the limits is straightforward.

As $r \to 0$, $j_{\max} \to n+1$. (The log term vanishes and the power term goes to -(n+1).)

As $r \to \infty$, $j_{\max} \to 0$.

Thus, the position of the maximum in $\sum$* goes from the end, node n, to the beginning, node 1, as r goes from 0 to ∞.

The disconcerting thing about eq. (44) is its dependence on *n*. The location of the maximum in Σ^* depends on the exact length of Σ^*, as well as on the value of *r*.

Consider the case of $r = 1$. Eq. (44) is not defined for $r = 1$, but two applications of L'Hôpital's Rule establish that as $r \to 1$, $j \to (n+1)/2$.

$$\lim_{r \to 1} \left[\frac{1}{\log r} - \frac{n+1}{r^{n+1}-1} \right] = \lim_{r \to 1} \frac{r^{n+1} - 1 - (n+1)\log r}{(r^{n+1}-1)\log r}$$

$$= \lim_{r \to 1} \frac{(n+1)\, r^{n+1} - (n+1)}{r^{n+1} - 1 + (n+1)(\log r)\, r^{n+1}} \quad \text{, by L'Hopital's rule}$$

$$= \lim_{r \to 1} \frac{(n+1)^2\; r^{n}}{(n+1)r^{n} + (n+1)r^{n} + (n+1)^2 r^{n}(\log r)} \quad \text{, by L'Hopital's rule}$$

$$= \frac{n+1}{2}$$

Thus, as shown in Part III from considerations of symmetry, when $r = 1$ the maximum in Σ^* falls right in the middle of the string. (When the string is of even length, the maximum value is shared by the two nodes flanking the center point of the string.) This is an extreme case of length dependence, but the behavior for other values of *r* is no less striking.

For $r > 1$, the term $(n+1)/(r^{n+1}-1)$ in eq. (44) vanishes as $n \to \infty$, and we have $j_{max} \to 1/\log r$. This is exactly the formula for the location of a barrier at *j*, in the $\mathbf{e}_k^*$'s, imposed by *r* (established in Part III). Recall that this means that for any input $\mathbf{e}_k$, the maximum on $\mathbf{e}_k^*$ will fall on node *j*, if $j \geq k$, otherwise on *k* itself. (In this latter case, think of *k* as being inside the barrier at $j = 1/\log r$.)

For $r < 1$, we need to consider the limit reached by the expression $(n+1) - j_{max}$. As $n \to \infty$, $r^{n+1} \to 0$ and the term $-(n+1)/(r^{n+1}-1)$ in eq. (44) goes to $(n+1)$. The expression $(n+1) - j_{max}$ asymptotes to $-1/\log r$. This is again exactly the formula for locating a barrier with respect to the *end* of the string, on node $(n - j_{max} + 1)$. We expect this symmetry between the models for which $\infty > r > 1$ and those for which $0 < r < 1$, given the reciprocal relation between the weight terms *r* and $1/r$.

Observe now that the location of the extremal point varies in a regular way with increase of *n*. We can consider eq. (44) to give *j* as a function J of *r* and *n*. Let us focus on the term involving n; rewriting it for convenience with $x = n + 1$, we have

$$J(r,x) = \frac{x}{r^{x} - 1}$$

We will show that this function is strictly decreasing on the interval of interest. Differentiating with respect to x, we find

$$\frac{\partial J}{\partial x} = \frac{r^x - 1 - x r^x \log r}{(r^x - 1)^2}$$

Since the denominator is positive, we need to assure ourselves that the numerator is negative. We need

$$1 - x \log r < r^{-x}$$

Recall that the Taylor series expansion for r^{-x} around 0 is

$$r^{-x} = 1 - x \log r + \tfrac{1}{2}(x \log r)^2 - \tfrac{1}{6}(x \log r)^3 + \ldots$$

Truncating this series after the 2nd term leaves a positive remainder (as may be checked by examining any remainder formula for Taylor series), and we have the result.

This means that j_{max}, the value of j for which $\sum_j$* is the local maximum, will start out near an edge and move steadily inward with increase of n until it reaches its asymptotic value. It will commonly be the case, then, that the value of j_{max} will change notably as n increases.

For example, suppose $r = .6$. The maximum is final for strings of length 2 and 3, and penultimate (2nd-to-last) for longer strings.

Suppose $r = .7$. Then the maximum is final for length 2, penultimate for length 3-8, and antepenultimate (3rd-to-last) for longer strings.

Suppose $r = .8$. The maximum is final for strings of length 2, penultimate for length 3-5, antepenultimate for length 6-10, pre-antepenultimate (4th from last) for length 11-82, and finally fifth-from-the-end for length 83 and above.

As noted in §1, there is no obvious application in the linguistic realm for this sort of behavior, and indeed it renders Σ useless as a model of linguistic structure.

*****

Let us turn now to the alternating patterns, those where $\alpha, \beta < 0$. We need to examine the pattern of maxima in $\Psi^*(r, j)$, for fixed r, repeated here for convenience:

$$\Psi^*(r,j) \quad = \quad j\,\sigma_{n-j}\, r^{\,n-j+1} + (n-j+1)\,\sigma_{j-1}\, r^{-j}$$

This breaks down cleanly into two cases, depending on whether n is odd or even.

If n is odd, then $\sigma_{n-j} = \sigma_{j-1}$. We can re-write Ψ^* like this:

$$\begin{aligned}\Psi^*_{odd}(r,j) \quad &= \quad \sigma_{j-1}\, r^{-j}\left[\, j\, r^{\,n-j+1} + (n-j+1)\,\right] \\ &= \quad \sigma_{j-1}\, r^{-j}\, \varphi_{odd}(j)\end{aligned}$$

Since $\varphi_{odd}(j)$ is always positive for the cases we are interested in ($0 \le j \le n$), the sign of Ψ_{odd}^* is exactly σ_{j-1}, positive for odd j, negative for even j. Therefore, for odd-length strings, Ψ_{odd}^* has the pattern [+−+−+−...+], with maxima on odd nodes, minima on even nodes, no matter what r is. The vector Σ^* will show exactly the same pattern of maxima (not necessarily *signs*, though). Odd-length strings are therefore entirely stable in their pattern of maxima, and they show no effects of string-length nor indeed any effect of the value of r.

The situation is richer in even-length strings. With even n, we have $\sigma_{n-j} = \sigma_j = -\sigma_{j-1}$, and the maximum-determining function $\Psi^*(r, j) = -\Psi(-r, j)$ comes out like this:

$$\begin{aligned}\Psi^*_{even}(r,j) \quad &= \quad \sigma_j\left(j\, r^{\,n-j+1} - (n-j+1)\, r^{-j}\right) \\ &= \quad \sigma_j\, r^{-j}\left[\, j\, r^{\,n+1} - (n-j+1)\,\right] \\ &= \quad \sigma_j\, r^{-j}\ \varphi_{even}(j)\end{aligned}$$

The sign pattern of Ψ_{even}^* is sensitive to the sign of $\varphi_{even}(j)$, which will indeed always switch from negative to positive for some real $j > 0$. To see this, observe that $\varphi_{even}(0) = -(n+1)$ and that φ_{even} is strictly increasing ($d\varphi_{even}/dj = r^{n+1}+1$, always positive). Because it's linear in j, φ_{even} must indeed become positive.

This change of sign has major impact on the shape of Σ^*. It is therefore useful to determine the value of j (call it j_0) for which φ_{even} is zero.

$$j_0\, r^{\,n+1} - (n+1-j_0) \quad = \quad 0$$

$$j_0 \quad = \quad \frac{n+1}{r^{\,n+1}+1}$$

For $j > j_0$, φ_{even} is positive. Maxima will occur on on even-numbered nodes counting from the beginning (where $\sigma_j = +1$); equivalently, on odd-numbered nodes counting right-to-left from the end (where $\sigma_{n-j} = +1$). Thus, beyond the point j_0, Σ^* has the pattern ...−+−+].

When φ_{even} is negative, the opposite pattern occurs. For $j < j_0$, guaranteeing $\varphi_{even} < 0$, the maxima occur on odd-numbered nodes counting from the beginning (where $\sigma_j = -1$); equivalently, on even-numbered nodes counting from the end (where $\sigma_{n-j} = -1$). Thus, in the span preceding j_0, Σ^* has the pattern [+−+−... .

Uniform alternation of maxima throughout the string — the image of familiar alternating stress patterns — can be obtained by careful placement of the zero-crossing point of φ_{even}. Putting it before the string of nodes begins or after it ends will ensure uniform behavior throughout. For small enough j_0, the function φ_{even} will be positive for all values of j that count as node-numbers. Maxima occur on even-numbered nodes. If j_0 is large enough, then φ_{even} is negative for strings less than n in length, and maxima occur on odd nodes.

This gives a characterization of alternating patterns, but the specter of length-dependence is not far to seek. Since j_0 is function of n as well as r, the setting of r that gives uniform alternation for one class of string-lengths will not always work for strings of different length. Fortunately, the parameter space exhibits two safe zones (even-max or iambic, odd-max or trochaic) in which there is stability for all lengths; separating these is a transitional zone where behavior is more varied.

The transition is itself of great interest. To see exactly how it evolves, observe that increase of r causes and increase in the slope of φ_{even} and moves j_0, the zero point of $\Psi_{even}^*(r, j)$, to the *left*. Recall that beyond j_0, the linear function φ_{even} is positive; before it, negative. Beyond j_0, therefore, the maxima fall agree in parity with the node-numbers; beyond it, they disagree. At j_0, then, two oppositely-oriented trains of alternation meet. This yields a structure we can represent iconico-alphabetically like this: [**O** e **O** e...<j_0>...o **E** o **E**].

With increase of r and leftward shift of j_0 there is a gradual transition from all-even maxima to all-odd maxima. This is of course exactly the sort of behavior that we expect from a linear system: the transition between any two states must be gradual.

Let us now determine the detailed structure of the safe zones and the transition.

When j_0 falls between two integers k and $k+1$, the corresponding functions $\Psi_{even}^*(r, k)$ and $\Psi_{even}^*(r, k+1)$ will be of the *same sign*. This does not of itself tell us which of Σ_k^* and Σ_{k+1}^* is the greater. However, we can easily find the point at which the two nodes are exactly *equal* by solving the equation $\Sigma_k^* - \Sigma_{k+1}^* = 0$. (This turns out to be the same as solving $\Psi_{even}^*(r, k) - \Psi_{even}^*(r, k+1) = 0$. This gives us the equation

(45) *Equality between Nodes k and k+1*

$$k\,r^{n+2} + (k+1)\,r^{n+1} - (n-k+1)\,r - (n-k) \quad = \quad 0$$

Call the polynomial $P_n(r, k)$; call the relevant zero ${}_n r_k$. By the Harriot-Descartes rule of sign-changes, $P_n(r, k)$ has exactly one positive zero. Therefore, there is always a unique point ${}_n r_k$ at which node k and node $k+1$ assume identical values in a string of length n.

With eq. (45) in hand, we can determine the boundaries of the safe zones where alternation is binary and independent of length. For the odd-max system (trochaic) we want

$$\vec{\Sigma}_1^* - \vec{\Sigma}_2^* \quad > \quad 0$$

The values of r that yield *equality* between Σ_1* and Σ_2* are given by the solution of $P_n(r,1) = 0$ for each fixed n.

$$r^{n+2} - 2\,r^{n+1} - n\,r - (n-1) \quad = \quad 0$$

We need to know the *largest* r that solves the equation, for any n. Any r larger than that will ensure the primacy of Σ_1*. We assert (but refrain from proving) that, as n increases, the values of r increase to a maximum and then decrease back toward 1. Application of Newton's method reveals that for even n (the only cases we're interested in), the largest r — call it R — occurs at $n = 6$ and has the value $R = 1.21102+$. Any r larger than R will ensure maxima on every even node (iambic regime).

The other boundary to the safe zone is found by examining the condition

$$\vec{\Sigma}_{n-1}^* - \vec{\Sigma}_n^* \quad > \quad 0$$

The resulting equation is

$$(n-1)\,r^{n+2} + n\,r^{n+1} - 2\,r - 1 \quad = \quad 0$$

Considerations of symmetry entail that the lower bound falls at $1/R$, and indeed the substitution $r \to 1/r$ transforms one equation into the other. The value of $1/R$ is .82575−. Any r less than $1/R$ will guarantee maxima on odd nodes (trochaic regime) for all lengths.

Within the transitional zone, where $1/R < r < R$, there is a gradualistic change from all even maxima to all odd maxima. To see how this takes place, we need to look at the behavior of the polynomial $P_n(r, k)$ near its positive zero ${}_n r_k$. We repeat the formula for $P_n(r, k)$ for convenience of reference.

$$k\,r^{n+2} + (k+1)\,r^{n+1} - (n-k+1)\,r - (n-k) \quad = \quad 0$$

The key observation is that as r increases so that $r > {}_n r_k$, the polynomial $P_n(r, k)$ becomes *positive*. To see this, observe that it starts out negative (since $P_n(0, k) = -(n-k) < 0$), crosses the r-axis at its one positive zero ${}_n r_k$, and stays positive thereafter, heading off to ∞.

From this it follows that increase of r will cause an even node flanking the zero of $P_n(r, k)$ to increase in activation and an odd node flanking the zero to decrease. To establish this, note the following:

$$\Delta\Sigma^* = \Sigma_k^* - \Sigma_{k+1}^* = \frac{2r}{(n+1)(r+1)^2}\left[\Psi^*(r,k) - \Psi^*(r,k+1)\right]$$

$$= c\ \Delta\Psi^*, \quad c > 0$$

This is just a positive constant times the difference in the Ψ^*'s. But $\Delta\Psi^*$ comes out like this:

$$\Delta\Psi^* = \Psi^*(r,k) - \Psi^*(r,k+1) = \sigma_k r^{-(k+1)} \boldsymbol{P}_n(r,k)$$

The sign of $\Delta\Sigma_k^*$ is thus determined by the sign of $\Delta\Psi^*$, which is in turn just the sign of $\sigma_k P_n(r, k)$. Since we are considering the case where $P_n(r, k) > 0$, the sign of $\Delta\Sigma_k^*$ is just σ_k.

From this it follows that in the node sequence *Odd Even*, for $r > {}_nr_k$, we must have *Odd* < *Even*, since the difference in activation [*Odd* − *Even*] is *negative*. Similarly, in *Even Odd*, we have *Even* > *Odd*.

We can now put together a complete picture of the course of events as r increases. Focus first on the node pair *E*(ven) *O*(dd). First, for small enough r, we are in the odd-max (trochaic) regime. The sign pattern of the Ψ^*'s is [− +], so that $E < O$. As r increases, j_0 in its leftward shift will on fall on node O, yielding [− 0] in the Ψ^*'s. With further increase, j_0 will sit between E and O, giving the sign pattern [− −]. With yet further increase, r will reach ${}_nr_E$, and we will have $E = O$, both Ψ^*'s negative. Beyond this, we have $E > O$, with the Ψ^*'s still negative. As r increases, j_0 falls on E and then continues leftward, rendering E permanently positive and O permanently negative in their Ψ^*'s. Thus, with increase of r the *E O* relation smoothly transits from from $E < O$ (trochaic) to $E > O$ (iambic), with an intermediate point of equality.

An *O E* sequence behaves in an entirely comparable way, except that it starts off, for small r, with a left-side maximum, $O > E$ and a sign pattern in the Ψ^*'s of [+ −]. As r increases, the E node's Ψ^* goes through 0 to positive status; then $O = E$, and finally we have $O < E$, both positive, and continuing in this vein, ultimately the Ψ^* sign pattern [− +], a full reversal.

We have defined a quasi-maximum as a node *n* such that the activation of *n* is greater than *or equal to* the activation of its adjacent nodes. In the following table, modified from §1 above, we portray the transition in maximum-structure of a string of a length 6 stretch of string, using capitalization to express quasi-maximum-hood:

Transition in 6-unit $\sum$*						r
X	*x*	*X*	*x*	*X*	*x*	Small r
X	*x*	*X*	*x*	*X*	**X**	↓
X	*x*	*X*	*x*	**x**	X	↓
X	*x*	*X*	**X**	x	X	r = 1
X	*x*	**x**	X	x	X	↓
X	**X**	x	X	x	X	↓
x	X	x	X	x	X	Big r

Each of the stages in this chart corresponds to some value-range in r. It is notable that the effects of a given r vary with string length. As an indicator of these effects, consider the location of j_0, the point where Ψ* = 0. Let us take r = 1.2, chosen to be within the transitional range (roughly .826 — 1.211).

Length	j_0
2	1.1
4	1.4
6	1.5
8	1.5
10	1.3

This seemingly mild variation gives rise to the following rather startling variations in quasi-maximal structure, represented here in an obviously iconic way:

Length	*QMS*
2	[. *X*]
4	[. *X* . *X*]
6	[*X* . . *X* . *X*]
8	[*X* . . *X* . *X* . *X*]
10	[. *X* . *X* . *X* . *X* . *X*]

Finally, we examine the remarkable instance of length-dependence occurs when $r = 1$. A glance at $P_n(1,k)$ will tell which two nodes k and k+1 are rendered equal by this choice of r.

$$P_n(1,k) \quad = \quad k + (k+1) - (n-k+1) - (n-k) \quad = \quad 4k - 2n$$

From $P_n(1,k) = 0$ we get

$$k \quad = \quad \frac{n}{2}$$

Thus the two nodes $n/2$ and $n/2 + 1$, straddling the mid-point, agree in activation and share quasi-extremum status: if $n/2$ is odd, both are quasi-maxima; both are quasi-minima if $n/2$ is even.

Fig.1. Critically Damped Harmonic Oscillator
with extension to negative time

$$y = c\,(1 + \rho\, x)\, e^{-\rho x}$$

Fig. 1

Fig. 2. Initial Wave for $\rho > 0$

$$y = c\, x\, e^{-\rho x}$$

$1/\rho$

Fig. 2

Fig. 3. Final Wave for $\rho > 0$

$$y = c\,(N - x)\,e^{-\rho x}$$

$N + 1/\rho$

Fig. 3

RuCCS TR-1
Part II

CONVERGENCE
of the Goldsmith-Larson Dynamic Linear Model
of Sonority and Stress Structure

Alan Prince
Department of Linguistics
Center for Cognitive Science
Rutgers, the State University of New Jersey

(Originally circulated May, 1991)

Abstract

The affine iteration $\mathbf{a} \leftarrow \mathbf{W}_n\mathbf{a} + \mathbf{b}$, where $\mathbf{W}_n$ is an n×n tridiagonal matrix with 0 on the main diagonal, $\alpha \in \mathbb{R}$ on the superdiagonal and $\beta \in \mathbb{R}$ on the subdiagonal, which has been used by Goldsmith and Larson in models of syllable structure and stress patterns, is shown to converge iff $|\alpha\beta| < 1 / [4 \cos^2 (\pi/(n+1))]$.

Convergence of the Goldsmith-Larson Dynamic Linear Model of Sonority and Stress Structure

Alan Prince

(May, 1991)

Consider the following iterative system, a dynamic linear model of syllabic and stress structure (DLM) put forth in the work of John Goldsmith and Gary Larson [refs.]:

(1) $$a_k \leftarrow \alpha \cdot a_{k+1} + \beta \cdot a_{k-1} + b_k \,, \qquad 1 \le k \le n$$

This describes the dynamics of a network consisting of a string of *n* nodes, in which activation is passed to and is received from adjacent neighbors. All leftward links bear weight α and all rightward links weight β. The activation of node k is a_k; node k also has an inherent unchanging bias b_k. For convenience, assume that $a_0 = a_{n+1} = 0$.

We rewrite the DLM iteration scheme in matrix form as follows:

(2) $$\vec{a} \leftarrow \boldsymbol{W}_n \vec{a} + \vec{b}$$

The vector **a** is initially **0**. $\mathbf{W}_n$ is the n×n tridiagonal matrix schematized in (3):

(3) $$\boldsymbol{W}_n = \begin{pmatrix} 0 & \alpha & 0 & \dots & & \\ \beta & 0 & \alpha & 0 & \dots & \\ 0 & \beta & 0 & \alpha & 0 & \dots \\ \vdots & \vdots & & & & \\ 0 & 0 & \dots & \beta & 0 & \alpha \\ 0 & 0 & \dots & \dots & \beta & 0 \end{pmatrix}$$

The DLM iteration will converge if and only if $\mathbf{W}_n^k \to 0$ as $k \to \infty$. In the trivial case where either $\alpha = 0$ or $\beta = 0$, it happens that $\mathbf{W}_n^n = 0$, so convergence is guaranteed. To investigate the more interesting case where $\alpha\beta \neq 0$, we calculate the eigenvalues of $\mathbf{W}_n$ in order to determine the asymptotic behavior of $\mathbf{W}_n^k$. We must solve the following n^{th} degree equation in λ:

(4) $$|W_n - \lambda I| = 0$$

The determinant D_n we need to compute looks like this:

$$(5) \qquad D_n = |W_n - \lambda I| = \begin{vmatrix} -\lambda & \alpha & 0 & \dots & & \\ \beta & -\lambda & \alpha & 0 & \dots & \\ 0 & \beta & -\lambda & \alpha & 0 & \dots \\ \vdots & \vdots & & & & \\ 0 & 0 & \dots & \beta & -\lambda & \alpha \\ 0 & 0 & \dots & \dots & \beta & -\lambda \end{vmatrix}$$

The determinant D_n has symmetry properties which allow us to write out a recurrence formula for it. Performing first three steps of the cofactor expansion yields the following:

$$(6) \qquad D_n = -\lambda D_{n-1} - \alpha\beta D_{n-2}$$

Since $D_2 = \lambda^2 - \alpha\beta$, the initial conditions must be as in (7)

$$(7) \qquad \begin{aligned} D_0 &= 1 \\ D_1 &= -\lambda \end{aligned}$$

Recurrence relations like that in equation (6) are commonly solved in elementary number theory; observe that eq. (6) defines the Fibonacci sequence when $\lambda = -1$ and $\alpha\beta = -1$. Paul Smolensky has provided the following perspicuous approach to deriving the solution in radicals. Observe that (6) is a homogeneous 2nd order difference equation, which can be re-written as (8):

$$(8) \qquad D_n + \lambda D_{n-1} + \alpha\beta D_{n-2} = 0$$

To solve this, we start with the Ansatz (9):

$$(9) \qquad D_n = c r^n$$

Substituting this into (8) yields

$$(10) \qquad \begin{aligned} 0 &= c r^n + c\lambda r^{n-1} + c\alpha\beta r^{n-2} \\ &= c r^{n-2} (r^2 + \lambda r + \alpha\beta) \end{aligned}$$

For $r \neq 0$, this has two solutions, given by the quadratic formula:

(11)
$$r_1 = \tfrac{1}{2}\left(-\lambda + \sqrt{\lambda^2 - 4\alpha\beta}\right)$$
$$r_2 = \tfrac{1}{2}\left(-\lambda - \sqrt{\lambda^2 - 4\alpha\beta}\right)$$

The general solution of eq. (8) is then

(12)
$$D_n = c_1 r_1^n + c_2 r_2^n$$

Plugging this into the initial conditions (7) gives the following:

(13)
$$D_0 = c_1 r^0 + c_2 r^0 = c_1 + c_2 = 1 \;, \qquad \textit{by the initial conditions } (7)$$

(14)
$$\begin{aligned} D_1 &= c_1 r_1 + c_2 r_2 = \tfrac{1}{2}\left[c_1\left(-\lambda + \sqrt{\lambda^2 - 4\alpha\beta}\right) + c_2\left(-\lambda - \sqrt{\lambda^2 - 4\alpha\beta}\right)\right] \\ &= \tfrac{1}{2}\left[-\lambda(c_1 + c_2) + (c_1 - c_2)\sqrt{\lambda^2 - 4\alpha\beta}\right] \\ &= -\frac{\lambda}{2} + \frac{(c_1 - c_2)}{2}\sqrt{\lambda^2 - 4\alpha\beta} \qquad \textit{since } c_1 + c_2 = 1 \\ &= -\lambda \qquad \textit{by the initial conditions } (7). \end{aligned}$$

For $\lambda^2 \neq 4\alpha\beta$, we may re-arrange the last two lines of (14) to get this:

(15)
$$c_1 - c_2 = \frac{-\lambda}{\sqrt{\lambda^2 - 4\alpha\beta}}$$

Adding to (15) the equation $c_1 + c_2 = 1$, from (13), yields this:

(16)
$$2c_1 = 1 + \frac{-\lambda}{\sqrt{\lambda^2 - 4\alpha\beta}} \;, \qquad \textit{so}$$
$$c_1 = \frac{-\lambda + \sqrt{\lambda^2 - 4\alpha\beta}}{2\sqrt{\lambda^2 - 4\alpha\beta}} = \frac{r_1}{\sqrt{\lambda^2 - 4\alpha\beta}}$$

Since $c_1 + c_2 = 1$, we must have

(17)
$$c_2 = \frac{\lambda + \sqrt{\lambda^2 - 4\alpha\beta}}{2\sqrt{\lambda^2 - 4\alpha\beta}}$$
$$= \frac{-r_2}{\sqrt{\lambda^2 - 4\alpha\beta}}$$

Whence, plugging into (12),

$$D_n = r_1^n \frac{r_1}{\sqrt{\lambda^2 - 4\alpha\beta}} - r_2^n \frac{r_2}{\sqrt{\lambda^2 - 4\alpha\beta}} = \frac{r_1^{n+1} - r_2^{n+1}}{\sqrt{\lambda^2 - 4\alpha\beta}} \tag{18}$$

To solve $D_n = 0$, let us consider those cases where λ is real. The radical formula can be rendered more manageable by the following substitution for λ:

$$\lambda = 2\sqrt{\alpha\beta}\ \cos\theta\ , \quad \mathit{for}\ \ 0 < \theta < \pi \tag{19}$$

The condition on Θ reflects the fact that $\lambda^2 \neq 4\alpha\beta$ if (18) is to make sense. The substitution (19) allows us to explore a range of values for λ, both real (for $\alpha\beta > 0$) and imaginary (for $\alpha\beta < 0$), which will turn out to cover all the cases. The substitution yields the following:

(20)

$$\begin{aligned} \sqrt{\lambda^2 - 4\alpha\beta} &= \sqrt{4\alpha\beta\cos^2\theta - 4\alpha\beta} \\ &= 2\sqrt{|\alpha\beta|}\sqrt{\cos^2\theta - 1} \\ &= 2i\sqrt{\alpha\beta}\sin\theta \end{aligned}$$

(21)

$$\begin{aligned} r_1 &= \tfrac{1}{2}(-2\sqrt{\alpha\beta}\cos\theta + 2i\sqrt{\alpha\beta}\sin\theta) \\ &= -\sqrt{\alpha\beta}(\cos\theta - i\sin\theta) \\ &= -\sqrt{\alpha\beta}\, e^{-i\theta} \end{aligned}$$

(22)

$$\begin{aligned} r_2 &= \tfrac{1}{2}(-2\sqrt{\alpha\beta}\cos\theta - 2i\sqrt{\alpha\beta}\sin\theta) \\ &= -\sqrt{\alpha\beta}(\cos\theta + i\sin\theta) \\ &= -\sqrt{\alpha\beta}\, e^{i\theta} \end{aligned}$$

Plugging these into (18), we find

$$\begin{aligned} D_n &= (-\sqrt{\alpha\beta})^{n+1} \frac{e^{-i(n+1)} - e^{i(n+1)}}{2i\sqrt{\alpha\beta}\sin\theta} \\ &= (-\sqrt{\alpha\beta})^n \frac{\sin(n+1)\theta}{\sin\theta} \end{aligned} \tag{23}$$

Note that $\sin\Theta \neq 0$. We must also assume that $\alpha\beta \neq 0$, as we have throughout.

To solve $D_n = 0$, we divide out all the irrelevant terms, obtaining

(24) $$\sin(n+1)\theta = 0$$

This immediately gives us:

(25) $$\theta = \frac{k\pi}{n+1} \quad , \quad 0 < k \leq n$$

Note that k>0 because sin $\Theta \neq 0$. Similarly, $k \neq n+1$. Thus, we have computed n eigenvalues λ_k, using eq. (19):

(26) $$\lambda_k = 2\sqrt{\alpha\beta}\ \cos k\pi/(n+1) \quad , \qquad 1 \leq k \leq n$$

Since there are n distinct eigenvalues given by this formula, and since the equation we wish to solve is of the n^{th} degree, we conclude that all eigenvalues have been enumerated.

Since the n×n matrix $\mathbf{W}_n$ has n distinct eigenvalues, it is diagonalizable. Thus there is a matrix $\mathbf{Q}$ such that $\mathbf{W}_n = \mathbf{QV}_n\mathbf{Q}^{-1}$, with $\mathbf{V}_n$ diagonal. Now, $\mathbf{W}_n{}^m = (\mathbf{QV}_n\mathbf{Q}^{-1})^m = \mathbf{QV}_n{}^m\mathbf{Q}^{-1}$. Thus, as $m \to \infty$, $\mathbf{W}_n{}^m \to 0$ iff $\mathbf{V}_n{}^m \to 0$ iff $\lambda_k{}^m \to 0$. Since all λ_k are of the simple form r or ir, $r \in \mathbb{R}$, to get $\lambda_k{}^m \to 0$ as $m \to \infty$, we must have $|\lambda_k| < 1$. To guarantee convergence of the DLM iteration, then, we need

(27)

$$|\lambda_k| = |2\sqrt{\alpha\beta}\ \cos k\pi/(n+1)| < 1 \quad , \qquad i.e.$$

$$\sqrt{|\alpha\beta|} < \frac{1}{2\ |\cos k\pi/(n+1)|} \quad , \qquad i.e.$$

$$|\alpha\beta| < \frac{1}{4\cos^2 k\pi/(n+1)}$$

Since the cosine term is largest for k = 1 (equivalently k = n), the fraction on the r.h.s of (27) is smallest, and we can conclude that the exact condition for convergence is therefore that given here:

(28) $$|\alpha\beta| < \frac{1}{4\cos^2 \pi/(n+1)}$$

Since the cosine term on the r.h.s. of (28), (4) is always less than 1, the entire r.h.s is always great than ¼, for all n. We therefore arrive at the following general condition:

(29) **DLM Convergence Theorem**. Any Dynamic Linear Model with $|\alpha\beta| \leq$ ¼ will converge, for all n, n the number of nodes in the network.

○○○○○

RuCCS TR-1
Part III

Closed-form Solution of the Goldsmith-Larson Dynamic Linear Model of Syllable and Stress Structure and Some Properties Thereof

Alan Prince
Department of Linguistics
Center for Cognitive Science
Rutgers, the State University of New Jersey

(Originally circulated June, 1991)

Abstract

Closed-form solution is obtained for the convergent cases of the affine iteration $\mathbf{a} \leftarrow \mathbf{W}_n\mathbf{a} + \mathbf{b}$, where $\mathbf{W}_n$ is an n×n tridiagonal matrix with 0 on the main diagonal, $\alpha \in \mathbb{R}$ on the supradiagonal and $\beta \in \mathbb{R}$ on the subdiagonal. This has been proposed by Goldsmith and Larson as a key element in models of syllable structure and stress patterns. We aim to find an expression for $\mathbf{b}^* = (\mathbf{I} - \mathbf{W}_n)^{-1}\mathbf{b}$, which gives the fixed point $\mathbf{b}^*$ achieved by starting from any $\mathbf{b}$ and initial condition $\mathbf{a} = 0$.

Since the function relating input $\mathbf{b}$ to output $\mathbf{b}^*$ is linear, we examine the effect of the iteration on the canonical basis vectors $\mathbf{e}_k$, and develop expressions for $\mathbf{e}_k^*$. Particularly simple solutions exist for the case $\alpha\beta = ¼$. Let $r = (\alpha/\beta)^{½}$. Then we have:

$$\vec{e}^{\,*}_{jk} = \operatorname{sgn}(\alpha^{k-j})\ \frac{2(n-k+1)}{n+1}\ j\, r^{k-j}\ , \qquad j \le k$$

$$\vec{e}^{\,*}_{jk} = \operatorname{sgn}(\beta^{k-j})\ \frac{2k}{n+1}\ (n-j+1)\, r^{k-j}\ , \qquad k \le j$$

A variety of basic properties of the $\mathbf{e}_k^*$ are demonstrated, under various constraints on the signs of α and β, including these:

(1) The signs of the entries in $\mathbf{e}_k^*$ ($\mathbf{e}_{jk}^*$) to the left of $\mathbf{e}_{kk}^*$ are determined entirely by the sign of α, alternating from $\mathbf{e}_{kk}^*$ when $\alpha < 0$. Similarly the signs of the $\mathbf{e}_{jk}^*$ to the right of $\mathbf{e}_{kk}^*$ are determined entirely by the sign of β.

(2) When the signs of αβ agree, $\alpha\beta > 0$, the absolute values of the $\mathbf{e}_{jk}^*$ have one and only maximum, to which the values rise and from which the values fall monotonically. This maximum may be shared between 2 adjacent nodes. The maximum can be placed on any entry $\mathbf{e}_{jk}^*$ in a given $\mathbf{e}_k^*$ by suitable choice of parameters α and β.

(3) When the signs of αβ disagree, $\alpha\beta < 0$, there are 3 types of behavior with respect to absolute values of the $\mathbf{e}_{jk}^*$; (i) a simple rise to $\mathbf{e}_{kk}^*$ and fall from it; (ii) a condition in which low amplitude ripples develop, moving in from the edge to $\mathbf{e}_{kk}^*$ over a very flat region, with a rapid fall-off beyond $\mathbf{e}_{kk}^*$; (iii) a uniform fall (or uniform rise) over the whole vector; i.e. maximum value at one edge or the other.

(4) The placement of maxima of absolute value, for given α and β, does not depend on the length of the vector (n of $\mathbf{W}_n$), but is measured from an edge.

Since maximum-behavior is shown to be uniform in each of the sign-conditions, it is suggested that αβ can be fixed, reducing the permitted range of parametric variation. For $\alpha\beta > 0$, attention may be limited to the models $\alpha\beta = ¼$, which are easy to solve, and therefore deserve to be called the ‘Canonical Models’.

Part III concludes with discussion of the formal and linguistic ramifications of the results, along with a tentative classifications of the models. It is shown, for example, that models with $\alpha\beta > 0$ can be linguistically interpreted as allowing an accent to fall no further from a specified edge than syllable j, with lexical accents surfacing only when they fall inside this limit. Indeed, the Canonical Models can easily be parametrized directly in terms the limit. Limited attention is paid to behavior of inputs more complex than $\mathbf{e}_k$, that is, weighted sums of $\mathbf{e}_k$; but the vector $(1, 1, 1, ...) = \Sigma_k\, \mathbf{e}_k$ is examined in the Canonical Models, with $\alpha=\beta=\pm½$. This vector can be interpreted as a string of syllables undifferentiated as to quantity. Its image under the model is shown to have a complete mirror-image symmetry, and, therefore, in the condition $\alpha=\beta=-½$, a kind of edge-in pattern of alternating maxima which, in the even-length case, induces clashes or lapses mid-string. Various other properties are noted, and it is suggested that limiting the accuracy with which parameters may be set, along with a requirement that models be behaviorally stable over the range of accuracy, might lead to a resolution of some of the nonlinguistic properties of the model.

Closed-form Solution of the Goldsmith-Larson Dynamic Linear Model of Syllable and Stress Structure and Some Properties Thereof

Alan Prince

(June, 1991)

We examine the following iterative system, a dynamic linear model (DLM) which has been proposed by John Goldsmith and Gary Larson as a key element in the theory of syllable structure and stress patterns:

(1) $$a_k \leftarrow \alpha \cdot a_{k+1} + \beta \cdot a_{k-1} + b_k \,, \qquad 1 \le k \le n$$

This describes the dynamics of a network consisting of a string of *n* nodes, in which activation is passed to and is received from adjacent neighbors. All leftward links bear weight α and all rightward links weight β. The activation of node k is a_k; node k also has an inherent unchanging bias b_k. For convenience, assume that $a_0 = a_{n+1} = 0$.

We rewrite the DLM iteration scheme in matrix form as follows:

(2) $$\vec{a} \leftarrow \boldsymbol{W}_n \vec{a} + \vec{b}$$

$\mathbf{W}_n$ is the n×n tridiagonal matrix schematized in (3):

(3) $$\boldsymbol{W}_n = \begin{pmatrix} 0 & \alpha & 0 & \ldots & & \\ \beta & 0 & \alpha & 0 & \ldots & \\ 0 & \beta & 0 & \alpha & 0 & \ldots \\ \vdots & \vdots & & & & \\ 0 & 0 & \ldots & \beta & 0 & \alpha \\ 0 & 0 & \ldots & \ldots & \beta & 0 \end{pmatrix}$$

The DLM iteration will converge iff condition (28), (4) is met, as shown above in Part II.

(4) $$|\alpha\beta| < \frac{1}{4\cos^2 \pi/(n+1)}$$

Suppose α and β are fixed so that DLM converges. Let the **b*** be the result when the DLM iteration converges given a bias vector **b**. The vector **b*** is a fixed point of the iteration, so we have

(5)
$$\vec{b}* = W_n \vec{b}* + \vec{b}$$

That is,

(6)
$$(I - W_n)\vec{b}* = \vec{b}$$

This yields a formula for **b*** in terms of **b**.

(7)
$$\vec{b}* = (I - W_n)^{-1}\vec{b}$$

The form of the entries in $(\mathbf{I} - \mathbf{W}_n)^{-1}$ can be determined straightforwardly from a calculation on $(\mathbf{I} - \mathbf{W}_n)$, which looks like this:

(8)
$$I - W_n = \begin{pmatrix} 1 & -\alpha & 0 & \ldots & & \\ -\beta & 1 & -\alpha & 1 & \ldots & \\ 0 & -\beta & 1 & -\alpha & 0 & \ldots \\ \vdots & \vdots & & & & \\ 0 & 0 & \ldots & -\beta & 1 & -\alpha \\ 0 & 0 & \ldots & \ldots & -\beta & 1 \end{pmatrix}$$

Let $\mathbf{E}_m$ be the determinant of $(\mathbf{I} - \mathbf{W}_m)$. Let c_{jk} be the generic entry in $(\mathbf{I} - \mathbf{W}_n)^{-1}$. We have the following result:

(9)
$$c_{jk} = \alpha^{k-j}\,\frac{E_{n-k}\,E_{j-1}}{E_n}\,, \qquad j \le k$$
$$c_{jk} = \beta^{j-k}\,\frac{E_{n-j}\,E_{k-1}}{E_n}\,, \qquad j \ge k$$

Sketch of proof. These formulas are obtained by the standard inversion procedure, whereby $(A^{-1})_{jk} = (-1)^{j+k}\,|\mathrm{cof}(A)_{kj}|\,/|A|$. The cofactor of a_{jk} is the matrix obtained from A by omitting row *j* and column *k*. Two observations are required to apply this procedure to the matrix $(\mathbf{I} - \mathbf{W}_n)$ and get the result in eq. (9).

First, there's a generalization of the notion of a triangular matrix to what we might call 'block triangular', in which the elements along the diagonal are blocks rather than single entries. Such a matrix will look, e.g., like this:

$$(10) \qquad \begin{pmatrix} a & b & c & & & \\ d & e & f & & 0 & \\ g & h & i & & & \\ j & k & l & m & & \\ n & 0 & p & q & r & s \\ t & u & v & w & x & y \end{pmatrix}$$

This matrix might be called 'lower block triangular'; reflecting it through the diagonal, so that the all-0 part lies in the lower triangle, also produces a block-triangular matrix. Just as the determinant of an upper or lower triangular matrix is equal to the product of its diagonal elements, so the determinant of a block-triangular matrix is equal to the product of the *determinants* of its diagonal *blocks*.

Second, the cofactor of any entry a_{jk} in $(\mathbf{I} - \mathbf{W}_n)$ turns out to be block-triangular. In particular, for $j \geq k$ $cof(a_{jk})$ has one block of size $(j-1)\times(j-1)$ in the upper left-hand corner, another block of size $(n-k)\times(n-k)$ in the lower right-hand corner. For $j \leq k$ $cof(a_{jk})$ has one block of size $(k-1)\times(k-1)$ in the upper left-hand corner, another block of size $(n-j)\times(n-j)$ in the lower right-hand corner. If $j > k$, so that a_{jk} is in the upper triangle, the all-zero part of the cofactor matrix lies in the lower triangle, and the two blocks are connected by a string of $-\beta$'s on the diagonal, $(j-k)$ of them. If $j < k$, so that a_{jk} is in the lower triangle, its cofactor is lower block-triangular, as in ex. (10), and the two blocks are connected by a diagonal of $-\alpha$'s, $(k-j)$ in number.

The formulas in eq. (9) follow immediately. The fate of the $(-1)^{j+k}$ factor in the generic formula is worth tracking: it has the same sign as $(-\alpha)^{k-j}$ and $(-\beta)^{j-k}$, so it simply disappears from the calculation. ☐

Let us evaluate the determinants in eq. (9). The expression $|\mathbf{I} - \mathbf{W}_n| = \mathbf{E}_n$ has the general form of the determinant evaluated previously in the process of finding the eigenvalues of $\mathbf{W}_n$. From Part II, eqs. (11) & (18), we derive, with $-\lambda = 1$, this formula:

$$(11) \qquad E_n = \frac{(1 + \sqrt{1 - 4\alpha\beta}\,)^{n+1} - (1 - \sqrt{1 - 4\alpha\beta}\,)^{n+1}}{2^{n+1}\sqrt{1 - 4\alpha\beta}}$$

To render this more manageable, we introduce a parameter u with this definition:

$$(12) \qquad 4\alpha\beta\cosh^2 u = 1 \;, \qquad u \in \mathbb{C}$$

Let u be complex and restricted to $u = U + i\Theta$, where $U > 0$ and $\Theta = 0$ or $\Theta = \pi/2$; or $U = 0$ and $0 < \Theta < \pi/(n+1)$. Then a unique u can be found for any values of α and β such that $4\alpha\beta \cosh^2 u = 1$.

★ For $0 < \alpha\beta < ¼$, i.e. $4\alpha\beta < 1$, we must have cosh u > 1, and $u \in \mathbb{R}^+$.

★ For $¼ < \alpha\beta < 1/[4 \cos^2 \pi/(n+1)]$, i.e. $4\alpha\beta > 1$, we have cosh u < 1 and so $u = i\Theta$, for $0 < \Theta < \pi/(n+1)$, so that $\cosh u = \cosh i\Theta = \cos \Theta$.

★ For $\alpha\beta < 0$, we need $\cosh u = i/(2|\alpha\beta|^{½})$. Therefore $u = U + i\pi/2$ and $\cosh u = i \sinh U$, $U \in \mathbb{R}^+$.

★ Note that $u \neq 0$, for this implies $\alpha\beta = ¼$, generating 0 denominators. This case will be handled separately.

Using eq. (12) as a substitution into the radical term in eq.(11), we arrive at this:

(13)
$$\begin{aligned}\sqrt{1-4\,\alpha\beta} &= \sqrt{4\,\alpha\beta\cosh^2 u \,-\, 4\,\alpha\beta} \\ &= 2\sqrt{\alpha\beta}\,\sqrt{\cosh^2 u \,-\, 1} \\ &= 2\,\sqrt{\alpha\beta}\ \sinh u\end{aligned}$$

From this we have the following reduction of $\mathbf{E}_n$:
(14)

$$E_n = \frac{(2\sqrt{\alpha\beta}\cosh u \,+\, 2\sqrt{\alpha\beta}\sinh u)^{n+1} \,-\, (2\sqrt{\alpha\beta}\cosh u \,-\, 2\sqrt{\alpha\beta}\sinh u)^{n+1}}{2^{n+1}\,(2\sqrt{\alpha\beta}\ \sinh u)}$$

$$= 2^{n+1}\sqrt{\alpha\beta}^{\,n+1}\ \frac{(\cosh u \,+\, \sinh u)^{n+1} \,-\, (\cosh u \,-\, \sinh u)^{n+1}}{2^{n+1}\sqrt{\alpha\beta}\ (2\sinh u)}$$

$$= \sqrt{\alpha\beta}^{\,n}\ \frac{e^{(n+1)u} \,-\, e^{-(n+1)u}}{2\sinh u}\ , \qquad \textit{so that}$$

$$E_n = \sqrt{\alpha\beta}^{\,n}\ \frac{\sinh(n+1)u}{\sinh u}$$

NOTES ON THE SOLUTION:

1. In the case where $\alpha\beta > ¼$, $u = \cosh^{-1}(4\alpha\beta)^{-½} = i\Theta$, $\Theta \in \mathbb{R}$, and we have *cos* Θ replacing *cosh u* in (12); the outcome is the same as eq. (14), except that *sinh (...u)* is replaced by *sin (...*Θ*)*.

2. To see more clearly what's happening in the case where $\alpha\beta < 0$, i.e. where α and β have different signs, we can work entirely with the reals. Let $c = |\alpha\beta|$. The following substitution will prove useful:

(15)
$$1 = 2\sqrt{c}\,\sinh u$$

The radical term and the expressions involving it come out as follows:
(16)

$$\sqrt{1 - 4\alpha\beta} = \sqrt{1 + 4c}$$

$$= \sqrt{4c\sinh^2 u + 4c}$$

$$= 2\sqrt{c}\cosh u$$

(17)

$$1 + \sqrt{1 + 4c} = 2\sqrt{c}\sinh u + 2\sqrt{c}\cosh u$$

$$= 2\sqrt{c}(\sinh u + \cosh u)$$

$$= 2\sqrt{c}\,e^{u}$$

(18)

$$1 - \sqrt{1 + 4c} = 2\sqrt{c}\sinh u - 2\sqrt{c}\cosh u$$

$$= 2\sqrt{c}(\sinh u - \cosh u)$$

$$= -2\sqrt{c}\,e^{-u}$$

Putting these into eq. (11) yields this:

(19)

$$E_n = 2^{n+1}\sqrt{c}^{n+1}\frac{e^{(n+1)u} - (-1)^{n+1}e^{-(n+1)u}}{2^{n+1}\sqrt{c}\,(2\cosh u)}$$

$$= \sqrt{c}^{n}\frac{\sinh(n+1)u}{\cosh u}, \qquad (n+1)\ \textit{even},$$

$$= \sqrt{c}^{n}\frac{\cosh(n+1)u}{\cosh u}, \qquad (n+1)\ \textit{odd}.$$

END OF NOTES.

Since the map $G:\mathbf{b} \mapsto \mathbf{b}^*$ given by $(\mathbf{I} - \mathbf{W}_n)^{-1}$ is linear, its properties are fully determined by its effect on a basis for $\mathbb{R}^n$. It is instructive, therefore, to calculate the map $\mathbf{e}_k \mapsto \mathbf{e}_k^*$, where $\mathbf{e}_k$ is the canonical basis vector with 1 in the k^{th} coordinate and 0 elsewhere. Since $\mathbf{b} = \Sigma b_k \mathbf{e}_k$, the behavior of $\mathbf{b} \mapsto \mathbf{b}^*$ in general is just the additive superimposition of individual terms of the form $G(b_k\mathbf{e}_k) = b_k G(\mathbf{e}_k) = b_k \mathbf{e}_k^*$.

In terms of the underlying network, the j^{th} coordinate of $\mathbf{e}_k^*$, $\pi_j(\mathbf{e}_k^*)$, which we will write as $\mathbf{e}_{jk}^*$, represents the isolated effect on the j^{th} node of the unit bias at the k^{th} node, since all nodes except k have bias 0. Because of linearity, effects emanating from different node-biases *do not interact* but merely superimpose, and once we have calculated the effect of the isolated unit biases, in principle we have the wherewithal to learn everything about how the network works.

What we need to calculate, then, is $\mathbf{e}_k^* = (\mathbf{I} - \mathbf{W}_n)^{-1}\, \mathbf{e}_k$. This is simply the k^{th} column of $(\mathbf{I} - \mathbf{W}_n)^{-1}$, whose elements are described by the formulas in eq. (9). The E_m's can be expressed in the terms given by eq. (14). Starting with $j \le k$, that is, node j to the left of node k, we arrive at the following:

(20)
$$\begin{aligned}\vec{e}_{jk}^{\,*} &= \alpha^{k-j}\, \frac{\sqrt{\alpha\beta}^{\,n-k}\ \sqrt{\alpha\beta}^{\,j-1}\ \sinh(n-k+1)u\ \sinh ju}{\sqrt{\alpha\beta}^{\,n}\ \sinh u\ \sinh(n+1)u} \\ &= (\operatorname{sgn}\alpha)^{k-j} \sqrt{\frac{\alpha}{\beta}}^{\,k-j}\ \frac{\sinh(n-k+1)u\ \sinh ju}{\sqrt{\alpha\beta}\ \sinh u\ \sinh(n+1)u}\ , \qquad j \le k\end{aligned}$$

We use the notation (sgn x) = 1 if $x \ge 0$, -1 if $x < 0$. Observe that the final denominator term in no way depends on *j* or *k*; it is strictly a function of α,β, and n, and therefore reflects a property of the network as a whole. We assign it a special symbol:

(21)
$$K_{\alpha\beta n} = \frac{1}{\sqrt{\alpha\beta}\ \sinh u\ \sinh(n+1)u}$$

We arrive at the following expression for $\mathbf{e}_{jk}^*$:

(22)
$$\vec{e}_{jk}^{\,*} = K_{\alpha\beta n}\ (\operatorname{sgn}\alpha)^{k-j} \sqrt{\frac{\alpha}{\beta}}^{\,k-j}\ \sinh(n-k+1)u\ \sinh ju\ , \quad j \le k$$

To determine the rightward effects of a unit bias at node k, we turn our attention to $\mathbf{e}_{jk}^*$ for $j \ge k$. Using the second equation in (9), we find

(23)
$$\begin{aligned}\vec{e}_{jk}^{\,*} &= K_{\alpha\beta n}\ (\operatorname{sgn}\beta)^{j-k} \sqrt{\frac{\beta}{\alpha}}^{\,j-k}\ \sinh(n-j+1)u\ \sinh ku \\ &= K_{\alpha\beta n}\ (\operatorname{sgn}\beta)^{k-j} \sqrt{\frac{\alpha}{\beta}}^{\,k-j}\ \sinh(n-j+1)u\ \sinh ku\ , \quad j \ge k\end{aligned}$$

These formulas for $\mathbf{e}_{jk}$* depend crucially on two measures of distance: from the node k, and from the edges of the network. They evince a division of the node string into 3 segments: the central stretch from j to k (or k to j), inclusive, and its flanks. The exponential term pertains to the central segment; the sinh terms to the flanks.

The exponential term $(\alpha/\beta)^{(k-j)/2}$ depends on the *signed* distance from node k: leftward is positive, rightward is negative (these directions could be switched if we used β/α instead of α/β). At node k this term is 1; node k is then like the origin for a simple exponential e^{-ax}, where a = $\log(\alpha/\beta)^{½}$. (The minus sign pops up because rightward is positive in the standard coordinate system.) This term, when $\alpha\beta > 0$ so that $(\alpha/\beta)^{½}$ is real, describes an exponential decay or inflation, depending on whether $\alpha/\beta > 1$ or $\alpha/\beta < 1$, running smoothly from one end of $\mathbf{e}_k$* to the other. If we think of k as fixed and j varying, this can be made even more clear by placing $(\alpha/\beta)^{(k/2}$ among the constant terms at the beginning of the expression, leaving $(\alpha/\beta)^{-j/2}$ as the exponential term.

The sinh terms directly reflect the size of the flanks, where by ‘flank’ we mean a contiguous stretch running from an edge-node to node j (node k resp.) which does not cross node-k (node j resp.). When j < k, the j-flank runs from $\mathbf{e}_{1k}$* to $\mathbf{e}_{jk}$* and the relevant term is sinh ju, as in eq. (22), which counts the number of nodes from node 1 to node j. The k-flank runs from $\mathbf{e}_{nk}$* to $\mathbf{e}_{kk}$* and its term is sinh (n−k+1)u, which counts the number of nodes from $\mathbf{e}_{kk}$* to $\mathbf{e}_{nk}$*. When j > k, mirror-image-wise, the j-flank runs from $\mathbf{e}_{nk}$* to $\mathbf{e}_{jk}$*, giving rise to the term sinh (n−j+1)u; the k-flank runs from $\mathbf{e}_{1k}$* to $\mathbf{e}_{kk}$*, yielding sinh ku.

If we think of k as fixed, and vary j, the heart of the activation function for j ≤ k and the function for j ≥ k can be resolved into the product of an exponential term and a sinh term dependent on j: sinh ju or sinh (n−j+1)u. (The sinh term dependent on k is constant within each of these two conditions.) Using the notation $d_{\#}(m)$ to indicate the number of nodes from the designated boundary # to node m, we can write

(24)
$$f_k^*(j) = \sqrt{\frac{\alpha}{\beta}}^{\,-j} \sinh d_{\#}(j)\,u$$

But while the exponential term runs smoothly and montonically throughout $\mathbf{e}_k$*, the sinh term splits into two parts at the node k. For j ≤ k, we have # taken to be the beginning of the node string, so that $d_{\#}(j)$ = j and we are dealing with sinh ju, which runs up from sinh u to sinh k. For k ≥ j, # is the end and $d_{\#}(j)$ = n−j+1, so that we have sinh $d_{\#}(j)$ = sinh (n−j+1)u, which runs down from sinh(n−k+1)u to sinh u. Thus we have a near mirror-image pattern in the sinh term.

As noted above, the formulas we have derived do not work for $\alpha\beta = ¼$. In that case, u = 0 and denominators vanishes. However, a limit argument easily establishes the result for these values, which turns out to be of particular interest. In the case of leftward spreading from the node with unit bias, where $j \le k$, we need to ascertain

(25) $$\lim_{\alpha\beta \to \frac{1}{4}} \left[(\operatorname{sgn}\alpha)^{k-j} \sqrt{\frac{\alpha}{\beta}}^{\,k-j} \frac{\sinh(n-k+1)u \ \ \sinh ju}{\sqrt{\alpha\beta}\ \sinh u \ \ \sinh(n+1)u} \right]$$

Since lim (xy) = (lim x)(lim y) if the limits exist, to evaluate expression (24), we break it into factors that can be easily analyzed.

(26)

$$\frac{1}{\sqrt{\alpha\beta}} \sqrt{\frac{\alpha}{\beta}}^{\,k-j} = 2 \sqrt{\frac{\alpha}{\beta}}^{\,k-j} \ , \ \textit{when} \ \ \sqrt{\alpha\beta} = \tfrac{1}{2} \ \ ; \qquad \textit{whence also}$$

$$= 2(2|\alpha|)^{k-j} = 2^{k-j+1}|\alpha|^{k-j} \ , \qquad \textit{since} \ \ \beta = 1/4\alpha \ \ , \qquad \textit{and}$$

$$= 2(2|\beta|)^{j-k} = 2^{j-k+1}|\beta|^{j-k} \ , \qquad \textit{since} \ \ \alpha = 1/4\beta .$$

(27)

$$\lim_{u \to 0} \left[\frac{\sinh(n-k+1)u}{\sinh u} \right] = \lim_{u \to 0} \left[\frac{(n-k+1) \ \cosh(n-k+1)u}{\cosh u} \right] , \quad \textit{by} \ \ \textit{L'hôpital's} \ \ \textit{rule}$$

$$= n-k+1$$

(28)

$$\lim_{u \to 0} \left[\frac{\sinh ju}{\sinh(n+1)u} \right] = \lim_{u \to 0} \left[\frac{j \ \cosh ju}{(n+1) \ \cosh(n+1)u} \right] , \quad \textit{by} \ \ \textit{L'hôpital's} \ \ \textit{rule}$$

$$= \frac{j}{n+1}$$

Putting these together, we find

(29) $$\vec{e}_{jk}^{\ *} = 2 \ (\operatorname{sgn}\alpha)^{k-j} \ \frac{j \ (n-k+1)}{n+1} \sqrt{\frac{\alpha}{\beta}}^{\,k-j} \ , \qquad j \le k \, , \ \alpha\beta = \tfrac{1}{4}$$

Similarly,

(30)
$$\vec{e}_{jk}{}^{*} = 2\ (\mathrm{sgn}\,\beta)^{k-j}\ \frac{k\ (n-j+1)}{n+1}\ \sqrt{\frac{\alpha}{\beta}}^{\,k-j}\ ,\qquad j \geq k\ ,\ \alpha\beta = \frac{1}{4}$$

For $\alpha\beta = ¼$, then, the equations simplify remarkably. The exponential term remains as before, but the sinh terms become linear. Thinking of k as fixed and j as variable, the sinh term rises as a line with slope $2(n-k+1)/(n+1)$ from node 1 to node k, and falls as line with slope $-2k/(n+1)$ from node k to node n. It is intuitively clear why this should be so: for $0 < \alpha\beta < ¼$ we have the function sinh $d_{\#}(j)u$ term which, graphed, is concave upward; for $\alpha\beta > ¼$, this becomes sin $d_{\#}(j)u$, whose graph is concave downward; the straight line marks the limiting case attained as each approaches the other.

****

From formulas (9), (22), and (23) we can deduce the basic properties of the behavior of the DLM as α and β vary. From the general formulas (9) for entries in $(\mathbf{I} - \mathbf{W}_n)^{-1}$ we can derive the most basic structure of the $\mathbf{e}_k$*.

We repeat those formulas here for convenience:

(31)

$$\vec{e}_{jk}{}^{*} = \alpha^{k-j}\ \frac{E_{n-k}\ E_{j-1}}{E_n}\ ,\qquad j \leq k$$

$$\vec{e}_{jk}{}^{*} = \beta^{j-k}\ \frac{E_{n-j}\ E_{k-1}}{E_n}\ ,\qquad j \geq k$$

It turns out that E_m are *positive*. This can be seen from the following argument. The eigenvalues of $(\mathbf{I} - \mathbf{W}_n)^{-1}$ are $1/(1-\lambda_k)$, λ_k an eigenvalue of $\mathbf{W}_n$. The determinant of $(\mathbf{I} - \mathbf{W}_n)^{-1}$ is equal to the product of its eigenvalues. From the eigenvalue calculation in Part II above, eq. (26), it follows that the λ_k come in positive, negative pairs. (For odd n, $\mathbf{W}_n$ also has an eigenvalue 0.) Thus the determinant of $(\mathbf{I} - \mathbf{W}_n)^{-1}$ is the product of terms of the form $1/(1-\lambda_k{}^2)$. Either λ_k is real and $|\lambda_k| < 1$, or $\lambda_k = ip$ for real p so that $1-\lambda_k{}^2 = 1+p^2$. It follows that every such term is positive, and their product is positive.

The sign of $\mathbf{e}_{jk}$* therefore depends entirely on the sign of α^{k-j} for j < k and β^{j-k} for j > k. We arrive at the following classification of behaviors:

Theorem 1. Classification of Sign Patterns in $\mathbf{e}_k$*

a. $\mathbf{e}_{kk}$* > 0. Node k's activation ($\mathbf{e}_{kk}$*) is positive.

b. If $\alpha > 0$, then $\mathbf{e}_{jk}$* is positive for $1 \le j \le k$, i.e. to the left of $\mathbf{e}_{kk}$*.

c. If $\beta > 0$, then $\mathbf{e}_{jk}$* is positive for $k \le j \le n$, ie. to the right of $\mathbf{e}_{kk}$*.

d. If $\alpha < 0$, then the sign of $\mathbf{e}_{jk}$* alternates leftward from $\mathbf{e}_{kk}$*, j < k.

e. If $\beta < 0$, then the sign of $\mathbf{e}_{jk}$* alternates rightward from $\mathbf{e}_{kk}$*, j > k.

Proof. Just given.

We now turn to a more fine-grained analysis of the results of varying α and β. Our goal is to characterize the number and position of activation maxima in $\mathbf{e}_k$*, which Goldsmith (1991) has proposed as the linguistically significant property of the underlying network. To that end, we will consider only the absolute value of the activation; that is we will ignore the terms (sgn $\alpha)^{k-j}$ and the (sgn $\beta)^{j-k}$ in eqs. ?,(22), and (23), which have an effect that is quite independent of activation value. Since we shall often have occasion to refer to specific $|\mathbf{e}_{jk}*|$ from the vector *abs*($\mathbf{e}_k$*) consisting of the absolute values of the entries in e_k*, let us introduce the more perspicuous notations [j] $=|\mathbf{e}_{jk}*|$ and [m,...,p] for the sequence of entries beginning with $|\mathbf{e}_{mk}*|$ = [m] and ending with $|\mathbf{e}_{pk}*|$ = [p]. In what follows, we will also use $\mathbf{e}_k$* for abs($\mathbf{e}_k$*) = [1,...,n].

We are interested in how the the activation values of [j] change with changing *j*. Much can be learned about the distribution of maxima by treating the basic activation formulas as functions of a real variable *x* rather than integer-valued *j*. We can think of the network as sampling the continuous function at discrete points. Sluicing away everything that does not depend on *j*, i.e. various positive constants , and moving to the continuous case, we have two functions to examine, continuous versions of eq.(24):

(32)

$$\mathrm{f}(x) = \sqrt{\frac{\alpha}{\beta}}^{\,-x} \sinh ux \ , \quad x \le k$$

$$\mathrm{g}(x) = \sqrt{\frac{\alpha}{\beta}}^{\,-x} \sinh(n - x + 1)u \ , \quad x \ge k$$

We need derivatives of these functions. For convenience,let a = log $(\alpha\beta)^{1/2}$. First, f(x):

(33)
$$f'(x) = -a\, e^{-ax} \sinh ux + u\, e^{-ax} \cosh ux$$

An extremum occurs only if

(34)
$$u e^{-ax} \cosh ux - a e^{-ax} \sinh ux = 0 \ , \quad i.e.$$

$$u \cosh ux - a \sinh ux = 0 \qquad , \ i.e. \qquad \frac{\cosh ux}{\sinh ux} = \frac{a}{u}$$

As we know from the above discussion, the function g(x) is a lightly disguised version of f(x). Let us put it into the same form as f(x), so that conclusions drawn from analysis of the one transfer transparently to the other. Let x = n−y+1; then we have

(35)
$$g(x) = g_1(y) = \sqrt{\frac{\alpha}{\beta}}^{\,-(n-y+1)} \sinh uy$$

$$= \sqrt{\frac{\beta}{\alpha}}^{\,n-y+1} \sinh uy$$

We can discard the constant part of the exponential term, as it does not affect the location of extrema; we have, then,

(36)
$$h(y) = \sqrt{\frac{\beta}{\alpha}}^{\,-y} \sinh uy$$

$$= e^{-by} \sinh uy \ , \qquad for \quad b = \log\sqrt{\beta/\alpha}$$

The character of the solutions to these equations depends on the status of *u*, which in turn depends on $\alpha\beta$. There are two basic cases to consider,

I. $\alpha\beta > 0$ and

II. $\alpha\beta < 0$

Condition I further splits into two subcases:

Ia. $0 < \alpha\beta < 1/4$ and

Ib. $1/4 < \alpha\beta < [4 \cos^2 \pi/(n+1)]^{-1}$.

Let us pursue each of these in turn.

I. $\alpha\beta > 0$.

For the case ¼ > $\alpha\beta$ > 0, general considerations based on the above discussion suggest the following: there ought to be a single maximum in $\mathbf{e}_k$*. The exponential term is monotonic; the sinh term rises and then falls. Suppose, for concreteness, that $\alpha > \beta$, so that the exponential term increases. The exponential term agrees in direction up to node k with the sinh term; the maximum on this segment lies at node k. On the second segment, from node k to node n, the two terms disagree in direction. Because of monotonicity, the interaction is rather simple, allowing for a maximum anywhere between node k and node n, depending on the parameters − but only for one such maximum. As it happens, the case $\alpha\beta$ > ¼ also allows only one maximum − the non-monotonicity of sin x doesn't get enough room to play. Let us proceed to the proof of both these assertions.

Theorem 2. Suppose 0 < $\alpha\beta$ < ¼. There is single maximum of the absolute value of activation in $\mathbf{e}_k$*, for all k, $1 \le k \le n$. Activation rises monotonically to the maximum, and falls monotonically from it.

Proof. Since 0 < $\alpha\beta$ < ¼, from $4\alpha\beta \cosh^2 u = 1$, we have $0 < u < \infty$. Therefore, ux > 0, both cosh and sinh are strictly increasing, and since cosh y > sinh y, for all y, it follows that eq. (34) has a solution iff a > u . This relation between $a=\log(\alpha/\beta)^{1/2}$ and $u=\cosh^{-1}(\tfrac{1}{2}(\alpha\beta)^{-1/2})$ can always be arranged, with proper choice of α and β, as can $a \le u$. To evaluate the character of the extremum, we examine f″(x):

(37) $$f''(x) = -a f'(x) + u^2 e^{-ax} \sinh ux - a u e^{-ax} \cosh ux$$

But at f′(x) = 0 we have cosh ux = (a/u) sinh ux, so this comes out as

(38) $$f''(x) = e^{-ax}(u^2 - a^2)\sinh ux \ , \quad f'(x) = 0$$

Since a>u when f′(x) = 0, it follows that f″(x) < 0 at such points. Therefore, all points where f′(x) =0 are maxima. Since f′(x) = 0 has exactly one solution when a > u, and none otherwise, f(x) has one and only maximum and that occurs iff a > u. Furthermore, since f′(x) > 0 before the maximum and f′(x) < 0 after it, it follows that f is monotonic up to the maximum and monotonic down from the maximum. For $a \le u$, f is monotonic increasing.

Similar remarks apply to g(x). From the direct parallel between h(x) and f(x) we have, immediately, that there is a maximum iff b > u and none iff $b \le u$. If $b \le u$, then h(y) is monotonic increasing, but g(x) monotonically decreasing, because x = n−y+1. Since b = −a, we have the following relations between f(x) and g(x):

i) If a> u > 0 > b , then *f* has a maximum; *g* decreases.
ii) If $u \ge a$ and $u \ge b$, then *f* increases; *g* decreases.
iii) If b > u > 0 > a, then *g* has a maximum; *f* increases.

These results apply to f(x) and g(x) without regard for limitations on the domain of x. Application to the circumscribed domains $1 \le x \le k$ and $k \le x \le n$ is, however, straightforward.

i) Suppose a>u>0>b. The function f has a maximum. If it occurs for x < 0, then [1] bears the maximum in $\mathbf{e}_k$*. If it occurs for x > k, then [k] bears the maximum. Otherwise the maximum occurs on [j] = $\mathbf{e}_{jk}$*, for $1 \le j \le k$. Note that activation falls from [k] to [n], since g is decreasing. Note also that since [k] belongs to both spans, we can conclude that everything in [k,...,n] to the right of k is less than anything in [1,...,k].

ii) Suppose u≥a and u≥b. Then f increases and g decreases. Therefore [k] bears the maximum.

iii) Suppose b>u>0>a. This is just the mirror image of case (i). The maximum falls on [j], k≤j ≤n. □

Let us turn to the case $\frac{1}{4} < \alpha\beta < [4 \cos^2 \pi/(n+1)]^{-1}$. We have u = iΘ, so that cosh u = cos Θ, and sinh u = i sin Θ, where 0 < Θ < π/(n+1). (The i divides out, ensuring a real answer for [j].) The term sin Θj in the activation function therefore ranges over the interval 0 < Θj < nπ/(n + 1). The appearance of the sin Θj, which is not monotonic on the interval under consideration, suggests that new behaviors might arise, but no such untoward event happens. Potential maxima in [1,...,k] are located by solving the equation below, paralleling eq. (34):

$$\theta \cos \theta x - a \sin \theta x = 0 \tag{39}$$

This has a unique solution for any legitimate choice of $a = \log(\alpha\beta)^{1/2}$ and $\Theta = \cos^{-1}(4\alpha\beta)^{-1/2}$, namely $x = \Theta^{-1} \cot^{-1}(a/\Theta)$, where $0 < \cot^{-1}\phi < \pi$. Furthermore, it is clear that the expression on the l.h.s starts out positive, turns 0, then remains negative: (Θ cos Θx) starts out at (arbitrarily near) Θ>0, and runs down to (arbitrarily near) −Θ, while (a sin Θx) starts out at (arbitrarily near) 0 and ends up back at (arbitrarily near) 0. These lines must cross. Thus the zero of the first derivative is at a *maximum*, to which the activation function rises monotonically and from which it falls monotonically.

Correspondingly, a maximum for the function g related to the case $j \geq k$ is guaranteed by the parallel equation with $b = \log(\beta/\alpha)^{1/2}$ in place of a, and $y = n - x + 1$ in place of x. Here we have $y = \Theta^{-1} \cot^{-1}(b/\Theta)$. Since $\cot^{-1}(-\phi) = \pi - \cot^{-1}(\phi)$, and since $b = -a$, we can derive the following relation between the solution for $j \leq k$
(call it x_f) and the solution for $j \geq k$ (call it x_g):
(40)

$$x_f = \frac{1}{\theta} \cot^{-1} \frac{a}{\theta} , \qquad \textit{and}$$

$$y_g = \frac{1}{\theta} \cot^{-1} \frac{-a}{\theta} = \frac{\pi}{\theta} - \frac{1}{\theta} \cot^{-1} \frac{a}{\theta}$$

$$= \frac{\pi}{\theta} - x_f , \qquad \textit{but}$$

$$y_g = n + 1 - x_g , \qquad \textit{so}$$

$$x_g - x_f = n + 1 - \frac{\pi}{\theta}$$

Noting that $\pi/(n+1) > \Theta$, i.e. that $\pi/\Theta > n+1$, we conclude that $x_g < x_f$. This means that the maximum for g(x) occurs *to the left* of the maximum for f(x). This will lead us immediately to the following strengthened version of Thm. 2:

Theorem 3. Let α and β be chosen so that the DLM converges. Suppose $\alpha\beta > 0$. There is single maximum of the absolute value of activation in $\mathbf{e}_k$*, for all k, $1 \leq k \leq n$. Activation rises monotonically to the maximum, and falls monotonically from it.

Proof. Theorem 2 establishes the result for $\alpha\beta < ¼$. Two cases remain: $\alpha\beta > ¼$ and $\alpha\beta = ¼$. Let us begin with $\alpha\beta > ¼$.

We know that each activation function has a maximum. We need to show that, nevertheless, $\mathbf{e}_k$* = [1,...,n] has only one local maximum. Suppose that the left hemi-network [1,...,k] has a maximum on [j], $1 \leq j < k$. Now we know from the above remarks that the maximum of the continuous activation function *g* for the right hemi-network [k,...,n] occurs *before* x = j. . Since *g* decreases after its maximum, it follows that activation decreases from [k] to [n].

Suppose now that the maximum for the left hemi-network [1,...,k] falls on [k]. In terms of the corresponding continuous activation function f(x) this is ambiguous between two cases: (i) f has a maximum at some x near k, which would entail that [k] is a left maximum; and (ii) the maximum for f(x) is somewhere off to the right, so that f(x) is rising toward it when it is cut off, as it were by node k. In case (i), g(x) has its maximum on or before x = k, [k] is the maximal node in the whole network. In case (ii), g(x) can have its maximum anywhere from x=k to x=n, depending on how far to the right the maximum of f(x) falls, and the maximum of $\mathbf{e}_k$* will fall in [k,...,n]. This covers all cases for $\alpha\beta > 1/4$.

Now, to conclude, we must consider what happens when $\alpha\beta = 1/4$. The relevant continous activation function and its derivatives are as follows:

(41)

$$\mathrm{f}(x) = x\, e^{-ax} \quad , \qquad a = \log\sqrt{\alpha/\beta}$$

$$\mathrm{f}'(x) = -axe^{-ax} + e^{-ax}$$

$$\mathrm{f}''(x) = -a\,\mathrm{f}'(x) - a\,e^{-ax}$$

When $\mathrm{f}'(x) = 0$, we have

(42)

$$x = 1/a \ , \qquad for\ \mathrm{f}'(x) = 0,\ a \neq 0.$$

$$\mathrm{f}''(x) = -a\,e^{-ax}$$

Notice that the sign of $\mathrm{f}''(x)$ depends on the sign of $a = \log(\alpha/\beta)^{1/2}$. For this case, then, f has both maxima and minima, depending on the α,β parameter settings, so long as $a \neq 0$: for $\alpha>\beta$, there is a maximum at x = 1/a; for $\alpha<\beta$, there is a minimum at $x = 1/a = -|1/a|$ and no other extrema. Of course, in the discrete system we are modeling, there are only positive-numbered nodes. Thus, for $\alpha/\beta > 1$, the left hemi-network [1,...,k] has a maximum [j], $1 \leq j \leq k$. For $\alpha/\beta < 1$, activation increases monotonically, and [k] bears the maximum. Parallel considerations apply in the usual mirror-image form to the right hemi-network [k,...,n]. Thus, when $\alpha/\beta > 1$, so that $\beta/\alpha < 1$, there is falling activation from node k to node n; the whole network's maximum then falls in [1,...,k]. When $\alpha/\beta < 1$, so that $\beta/\alpha > 1$, the network's maximum falls in the right hemi-network [k,...,n]. When $\alpha = \beta$, $a = \log(\alpha/\beta)^{1/2} = 0$, and f(x) rises to infinity on a straight line, g(x) falls the same way, and [k] bears the maximum, which is $2k(n-k+1)/(n+1)$, by direct computation. □

Theorem 4. Increasing α/β moves the maximum to left. Increasing β/α moves the maximum to right. Setting $\alpha=\beta$ puts maximum on [k] if $\alpha\beta \le 1/4$, and beyond the midpoint of the network for $\alpha\beta > 1/4$.

Proof. Consider the case $\alpha\beta < 1/4$. Focus on the left hemi-network [1,...,k]. If $\log(\alpha/\beta)^{1/2} = a \le u$, then there is no maximum in the associated continuous activation function f(x), which increases monotonically; in which case the maximum falls on [k]. Increasing α/β will lead to a > u, which puts a maximum in f(x), located by this formula:

(43)

$$u \cosh ux = a \sinh ux \ , \qquad i.e.,$$

$$x = \frac{1}{u} \tanh^{-1} \frac{u}{a}$$

Since $\tanh^{-1}$ is strictly increasing, the decrease in (u/a) caused by increasing *a* shrinks *x*, moving the maximum leftward. The same remark holds for $\tan^{-1}$, covering the $\alpha\beta > 1/4$ case. Argument for right hemi−network is entirely parallel, in the usual mirror-image fashion. If $\alpha = \beta$, for $\alpha\beta < 1/4$ then both f and g are increasing, putting the maximum on [k]. For $\alpha\beta < 1/4$, $\alpha = \beta$ yields $\Theta \cos \Theta x = 0$, i.e. $\Theta x = \pi/2$, $x = \pi/2\Theta$. Since $\Theta < \pi/(n+1)$, we have $\pi/2\Theta > (n+1)/2$, and the maximum falls beyond the midpoint of the network, counting from the relevant edge. □

Remark. We have found that all convergent networks for which $\alpha\beta > 0$ show the same range of behaviors with respect to presence of maxima. All have one maximum activation value among the [j]; values rise monotonically to it, and fall monotonically from it. This suggests that admitting a 2-dimensional α,β-parameter space may be unnecessary. The parameters can be restricted to the hyperbola $\alpha\beta = 1/4$, on which the network has a remarkably simple solution, with [j] turning on the product of an exponential and a piecewise linear function.

II. $\alpha\beta < 0$

Let us turn now to the case $\alpha\beta < 0$. The seemingly trivial change in sign structure (from *agree* to *disagree*) entails an entirely different behavioral repertory in the network. This can be intuited from the change induced in the activation function $f(t) = e^{-at} \sinh ut$. The parameter $u = U + i\pi/2$ is now crucially complex, as indeed is $a = \log\,(\alpha/\beta)^{1/2} = \log|\alpha/\beta|^{1/2} + i\pi/2$. The function f(t) describes a curve wandering about the complex plane. The exponential term spirals around the origin; the sinh term describes another kind of spiral, winding in the opposite direction. The point-wise product of the two is a complicated curve in the half-plane *real* > 0 which oscillates across the real axis, sometimes turning back on itself. Rather than sampling a real function $f(x):\mathbb{R} \to \mathbb{R}$ when x is an integer, the network's nodes take on the values of a squiggling curve $f(t):\mathbb{R} \to \mathbb{C}$ when it crosses the real axis, which happens with integral *t*. Comparing the relative values of adjacent [j−1], [j], [j+1] is no longer reducible to determining local maxima in a continuous function that threads through them in a simple fashion. Rather, between each pair of node values the associated continuous function loops off into the complex plane and comes back. Consequently, the localistic methods used above are not applicable. Instead, we must directly compare [j] and its neighbors. It turns out to be informative (ultimately indeed, more precisely informative than the continuous methods can be) to examine the *ratio* between [j] and its immediate neighbors. When [j]/[j−1] > 1 and [j]/[j+1] > 1, we have a maximum at [j].

As a preliminary to the analysis, it is useful to cash in the product of the two complex terms for their real value. For the exponential term, we have

(44) $$\sqrt{\frac{\alpha}{\beta}}^{\,-j} = i^{-j}\sqrt{\left|\frac{\alpha}{\beta}\right|}^{\,-j} = (-i)^{j}\sqrt{\left|\frac{\alpha}{\beta}\right|}^{\,-j}$$

The sinh term becomes

(45)

$$\sinh(U + i\,\pi/2)\,j = \sinh Uj\,\cosh i\frac{\pi j}{2} + \cosh Uj\,\sinh i\frac{\pi j}{2}$$

$$= \sinh Uj\,\cos\frac{\pi j}{2} + i\cosh Uj\,\sin\frac{\pi j}{2}$$

The trigonometric terms take on the values {0,1,−1} for integral *j*, while the term $(-i)^j$ cycles among the values {1, −i, −1, i}. Their product ends up real, and alternates between sinh Uj, for j even, and cosh Uj, j odd, as can be seen from the following table:

j	$(-i)^j$	sinh uj	$(-i)^j$ sinh uj
4m	1	sinh Uj	sinh Uj
4m+1	*−i*	*i cosh Uj*	*cosh Uj*
4m+2	−1	−sinh Uj	sinh Uj
4m+3	*i*	*−i cosh Uj*	*cosh Uj*

We must therefore consider different functions, depending on whether j is even or odd. Write f_{jk}* for [j] = $\mathbf{e}_{jk}$* with all j-free factors divided out, and let r = $|\alpha/\beta|^{½}$:

(46) $$f_{jk}^{\,*} = r^{-j}\sinh Uj\,, \quad j \textit{ even}$$
$$f_{jk}^{\,*} = r^{-j}\cosh Uj\,, \quad j \textit{ odd}$$

We want to establish the following characterization of the structure of [1,...,k] as the parameters vary:

1. For $|\alpha|<1$, [j] rises monotonically to [k], $1\leq j\leq k$.

2. For $|\alpha|> 1/(1-|\beta|)$, activation falls monotonically from [1] to [k].

3. For $1 < |\alpha| < (1-|\beta|)^{-1}$, as α increases, a low-amplitude alternating ripple with maxima on [j], j even, and minima on [j],j odd, j < k, spreads from [1] to [k].

The usual mirror-imaging remarks apply to [k,...,n].

Let us now establish a series of assertions that will lead to this characterization. In what follows, we assume $\alpha\beta < 0$. Recall that we are abstracting away from sign alternations and by the symbol [j] we always mean the *absolute value* of the activation of node j. When discussing the network segment [1,...,k] we mean by 'maximum on [j]' that [j] > [j−1] and [j] > [j+1], for 1<j<k; when j=1, we mean only [1] > [2], and when j=k, we mean only [k] > [k−1]. Similar remarks hold for the segment [k,...,n]. It will be useful to introduce the term 'quasi-maximum on [j]', by which we mean [j] ≥ [j−1] and [j] ≥ [j+1]. All discussion will be centered on the left hemi-network [1,...,k], since conclusions about it can be immediately converted into conclusions about [k,...,n].

Lemma 1. No quasi-maximum occurs on any [j], j≠k, in [1,...,k], if j is even.

Proof. For a quasi-maximum to sit on [j], j even, we must have

(47)

$$\frac{\vec{e}_{jk}^{\,*}}{\vec{e}_{(j-1)k}^{\,*}} = \frac{f_{jk}^{\,*}}{f_{(j-1)k}^{\,*}} = \frac{r^{-j}\sinh Uj}{r^{-(j-1)}\cosh U(j-1)} = \frac{\sinh Uj}{r\cosh U(j-1)} \geq 1\ , \qquad i.e.$$

$$\frac{\sinh Uj}{\cosh U(j-1)} \geq r$$

(48)

$$\frac{\vec{e}_{jk}^{\,*}}{\vec{e}_{(j+1)k}^{\,*}} = \frac{f_{jk}^{\,*}}{f_{(j+1)k}^{\,*}} = \frac{r^{-j}\sinh Uj}{r^{-(j-1)}\cosh U(j+1)} = \frac{r\sinh Uj}{\cosh U(j+1)} \geq 1\ , \qquad i.e.$$

$$r \geq \frac{\cosh U(j+1)}{\sinh Uj}$$

But these conditions can never hold simultaneously. The reason is that

(49) $$\frac{\cosh U(j+1)}{\sinh Uj} > \frac{\sinh Uj}{\cosh U(j-1)}\ , \qquad i.e. \quad \cosh U(j+1)\ \cosh U(j-1) > \sinh^2 Uj$$

To establish this, we invoke the following identity, which is easily verified from the definition of sinh x:

(50) $$\sinh^2 mu = \sinh(m-1)u \ \sinh(m+1)u + \sinh^2 u \ , \quad \textit{for all } m, u \in \mathbb{C}$$

We have, therefore,
(51)

$$\cosh U(j+1) \ \cosh U(j-1) \ >? \ \sinh^2 Uj \ > \ \sinh U(j+1) \ \sinh U(j-1) + \sinh^2 U \ , \quad i.e.$$

$$\cosh U(j+1) \ \cosh U(j-1) - \sinh U(j+1) \ \sinh U(j-1) \ >? \ \sinh^2 U \ , \quad i.e.$$

$$\cosh\,[(j+1)-(j-1)]\,U = \cosh 2U \ >? \ \sinh^2 U \ , \quad i.e.$$

$$\cosh^2 U + \sinh^2 U \ > \ \sinh^2 U$$

Since this last is self-evident, we follow the implicit chain of *iff*'s back up to conclude that inequality (49) is sound. ☐

Corollary. There is a choice of α, β such that a minimum or quasi-minimum can be placed on any even [j] in [1,...,k].

Proof. This follows immediately from inequality (49); an r can always be chosen that fits between l.h.s and r.h.s.

Lemma 2. There is a choice of α,β such that a maximum or quasi-maximum can be placed on any [j], j *odd*, in [1,...,k].

Proof. This follows from a calculation analogous to the one just performed, with sinh and cosh swapped, due to the change in parity of j. To put a maximum or quasi-maximum on [j], j odd, we must have

(52) $$\frac{\cosh Uj}{\sinh U(j-1)} \geq r \geq \frac{\sinh U(j+1)}{\cosh Uj}$$

There will always be such an r if the far l.h.s is always greater than the far r.h.s., which indeed it is. To show this, we need only establish:

(53) $$\cosh^2 Uj \ > \ \sinh U(j-1) \ \sinh U(j+1)$$

But cosh u > |sinh u|, u ∈ ℝ, so that $\cosh^2$ mx > $\sinh^2$ mx, and $\sinh^2$mx > sinh(m+1)x sinh(m−1)x, by identity (50). Observe that the strictness of inequality (53) entails that we can never have [j−1] = [j] = [j+1]. ☐

Lemma 3. *Even-Odd Rule*. If [j−1] ≤ [j] , for some odd j, then [m−1] < [m] for all preceding odd m < j. That is, if we pick r so that there is equality or rise over an even-odd node sequence, then there is a rise over every preceding even-odd sequence.

Equivalently, by contraposition, if [m−1] ≥ [m] for odd m, then [j−1] > [j], m < j, j odd. That is, if there is equality or fall over an even-odd node sequence, then all following even-odd sequences are strictly falling.

Proof. If the assumption holds, then we have:

(54)

$$\frac{\vec{e}_{jk}^{\,*}}{\vec{e}_{(j-1)k}^{\,*}} = \frac{f_{jk}^{\,*}}{f_{(j-1)k}^{\,*}} = \frac{r^{-j}\cosh Uj}{r^{-(j-1)}\sinh U(j-1)} = \frac{\cosh Uj}{r\sinh U(j-1)} \geq 1 \ , \qquad \boldsymbol{i.e.}$$

$$\frac{\cosh Uj}{\sinh U(j-1)} \geq r \ , \qquad j \ odd \, , \, j \leq k$$

We need to show that r satisfies the analogous inequality for all preceding odd m < k. This will happen if

$$(55) \qquad \frac{\cosh Um}{\sinh U(m-1)} > \frac{\cosh Uj}{\sinh U(j-1)} \geq r \ , \qquad j \ odd \, , \, j \leq k \, , \, m < k$$

This relation holds if F(x) = (cosh ux)/sinh u(x−1) decreases strictly with increasing x; which it does, because sinh plays steady catch-up with cosh. More formally,

(56)

$$D\left(\frac{\cosh ux}{\sinh u(x-1)}\right) = \frac{u\sinh u(x-1)\ sinhux \ - \ u\cosh ux\ \cosh u(x-1)}{\sinh^2 u(x-1)}$$

$$= \frac{-u\cosh u}{\sinh u(x-1)}$$

Since this is always negative when defined (x ≠ 1), F(x) is strictly decreasing in the interval we're interested in. ☐

Lemma 4. *Odd-Even Rule*. If [j] ≥ [j+1], for some odd j, then [m] > [m+1] for all odd preceding m < j. That is, if we pick r so that there is equality or a fall over an odd-even node sequence, then there is a fall over every preceding odd-even node sequence.

Equivalently, by contraposition, if [m] ≤ [m+1] for some odd m, then [j] < [j+1] for all following odd j > m. That is, if there is equality or rise over an odd-even node sequence, then all following odd-even sequences are strictly rising.

Proof. If [j] ≥ [j+1], we must have

(57)

$$\frac{\vec{e}_{jk}^{\;*}}{\vec{e}_{(j+1)k}^{\;*}} = \frac{f_{jk}^{\;*}}{f_{(j+1)k}^{\;*}} = \frac{r^{-j}\cosh Uj}{r^{-(j+1)}\sinh U(j+1)} = \frac{r\cosh Uj}{\sinh U(j+1)} \geq 1 \;, \qquad i.e.$$

$$r \geq \frac{\sinh U(j+1)}{\cosh Uj} \;, \qquad j \; odd\,,\; j \leq k$$

We will get [m] > [m+1] for all preceding odd m if we have

(58)
$$r \geq \frac{\sinh U(j+1)}{\cosh Uj} > \frac{\sinh U(m+1)}{\cosh Um}, \qquad j \; odd\,,\; j \leq k\,,\; m < j$$

We have this if G(x) = sinh u(x+1)/cosh ux is strictly increasing, which it clearly is:

(59)

$$\begin{aligned}(\cosh^2 ux)\, G'(x) &= u\cosh ux \;\cosh u(x+1) \;-\; u\sinh u(x+1)\;\sinh ux \\ &= u\cosh 2u \\ &> 0\;, \qquad u > 0\end{aligned}$$

This establishes the result. ☐

Remark. Relation to αβ > 0 cases. The same form of analysis could be applied to the αβ > 0 cases discussed above. Here there would be, of course, no odd/even disparity, and we would find that a rise over [j, j+1] entails a rise over *every* preceding pair, a fall entails continous falls to the left of it, deriving the single-maximum phenomenon.

Theorem 5. If there is a quasi-maximum at [j], j < k (j necessarily odd), then there is a maximum at every preceding [m], m odd, m < j. If there is a quasi-minimum at [j], j < k (j necessarily even), then there is a minimum at every preceding [m], m even, m < j.

Proof. If there is a quasi-maximum at [j] , j odd, then there is equality or fall over the odd-even node pair [j, j+1]. By Lemma 4, there must be a *fall* over every preceding odd-even pair. Likewise, given a quasi-maximum at [j], there must be equality or rise over the even-odd node pair [j−1, j]. By Lemma 3, there is a *rise* over every even-odd node pair preceding [j−1, j]. Therefore every preceding odd node is a local maximum.

A similar argument establishes the claim about minima. □

Theorem 6. For $|\alpha|<1$, the [j] rise monotonically to [k], $1 \le j \le k$. If $|\alpha| = 1$ then [1] = [2] and the monotonic rise begins with [2].

Proof. We prove only the first assertion. First we show that $|\alpha|<1$ entails a rise from [1] to [2]. Then we show, with the aid of Lemmas 1 and 4, that this entails the theorem.

Let us examine the ratio [1]/[2].

(60)

$$\frac{\vec{e}_{1k}^{\,*}}{\vec{e}_{2k}^{\,*}} = \frac{f_{1k}^{\,*}}{f_{2k}^{\,*}} = \frac{r^{-1}\cosh U}{r^{-2}\sinh 2U} = \frac{r\cosh U}{\sinh 2U} = \frac{r\cosh U}{2\sinh U\cosh U}$$

$$= \frac{r}{2\sinh U}$$

Recall that *sinh U* = $|4\alpha\beta|^{-1/2}$. We have a rise here iff

(61)

$$\frac{r}{2\sinh U} > 1 \,, \qquad i.e.$$

$$r > 2\sinh U \,, \qquad i.e.$$

$$\sqrt{\left|\frac{\alpha}{\beta}\right|} > 1/\sqrt{|\alpha\beta|} \,, \qquad i.e.$$

$$|\alpha| > 1$$

Lemma 4 tells us that if there is a rise on [1,2] then there is a rise on every odd-even sequence to its right. This takes care of the [odd, even] node sequences, which are rising.

Now if there were fall or equality over any [even, odd] sequence to the right of [1,2], this would create a quasi-maximum on the even node, which is impossible by Lemma 1. Therefore, the [even, odd] pairs are also strictly rising. □

Theorem 7. For $|\alpha| > 1/(1-|\beta|)$, the [j] fall monotonically from [1] to [k], $1 \le j \le k$.

Proof. First we show that the condition on α entails a fall over node-pair [2,3]. Then we show, with the aid of Lemmas 1 and 3 that this entails a fall throughout the segment [1,....,k].

(62)

$$\frac{\vec{e}_{1k}^{\,*}}{\vec{e}_{2k}^{\,*}} = \frac{f_{1k}^{\,*}}{f_{2k}^{\,*}} = \frac{r^{-2}\ \sinh 2U}{r^{-3}\ \cosh 3U} = \frac{r \sinh 2U}{\cosh 3U} = \frac{2\, r \sinh U\ \cosh U}{4 \cosh^3 U\ -\ 3 \cosh U}$$

$$= \frac{2\, r\ \sinh U}{4 \cosh^2 U\ -\ 3} > 1 \qquad \textit{iff}$$

$$r > \frac{4(\cosh^2 U\ -\ 1)\ +\ 1}{2 \sinh U} = 2 \sinh U + \frac{1}{2 \sinh U} \qquad \textit{iff}$$

$$\sqrt{\left|\frac{\alpha}{\beta}\right|} > \frac{1}{\sqrt{\alpha\beta}} + \sqrt{\alpha\beta} \qquad \textit{iff}$$

$$|\alpha| > 1 + |\alpha\beta|\ , \qquad \textit{to wit}$$

$$|\alpha| > \frac{1}{1-|\beta|}$$

Since $|\beta|>0$, we have $|\alpha| > 1$. Therefore there is a fall over [1,2] which continues over [2,3]. The fall on [2,3] entails, by Lemma 3, a fall over every [even, odd] pair to its right. This takes care of [even,odd].

Now suppose that there is equality or rise over [odd, even]. Since we have shown that [even, odd] is always falling, this would force a quasi-maximum on an even node, an impossibility by Lemma 1. □

We have thus far described the situation in static, implicational terms. The method we have used to derive these descriptions, however, has also given us sufficient information to put together a picture of how network behavior changes as $r = |\alpha\beta|^{1/2}$ changes.

Theorem 8. As $r = |\alpha\beta|^{1/2}$ increases, [1,...,k] is first monotonically increasing, for $|\alpha| < 1$. For $1 < |\alpha| < 1 - |\beta|$, as r increases, first a ripple of minima on even-numbered nodes spreads from node 2 rightward up to the last even-numbered node before k. Then a uniformly falling ramp spreads back from k until, for $|\alpha| > (1 - |\beta|)^{-1}$, the hemi-network [1,...,k] becomes monotonically decreasing.

Proof. We know from the proof of Lemma 3 that a node sequence [even, odd], which we can write as [2p, 2p+1], shows equality or rise just in case

(63) $$\frac{\cosh(2p+1)U}{\sinh(2p)U} \geq r$$

Conversely, the sequence [even, odd] shows equality or fall when

(64) $$r \geq \frac{\cosh(2p+1)U}{\sinh(2p)U}$$

Let's symbolize the ratio term as **C**(2p,2p+1).

We know from the proof of Lemma 4 that a node sequence [odd, even] shows equality or fall just in case

(65) $$r \geq \frac{\sinh(2p)U}{\cosh(2p-1)U}$$

Conversely, the sequence [odd, even] shows equality or rise when

(66) $$\frac{\sinh(2p)U}{\cosh(2p-1)U} \geq r$$

Let's symbolize the ratio term here as **S**(2p−1,2p).

From the proofs of Lemma 3 and Lemma 4 we also know that **C**(x,x+1) is a strictly decreasing function; that **S**(x−1,x) is strictly increasing; from the proof of Lemma 1 and Lemma 2 we know that **S**(x−1,x) < C(x,x+1) and **S**(x+1,x+2) < **C**(x,x+1). This gives us the following strict order:

Odd k:
S(1,2) < **S**(3,4) <...< **S**(k−2,k−1) < **C**(k−1,k) <...< **C**(4,5) < **C**(2,3)

Even k:
S(1,2) < **S**(3,4) <...< **S**(k−1,k) < **C**(k−2,k−1) <...< **C**(4,5) < **C**(2,3)

We can tabulate the effect the r-**C**-**S** relations as follows:

a. *r* < **S**(j−1,j) OE rise over [j−1,j]
b. *r* > **S**(j−1,j) OE fall over [j−1,j]
c. *r* < **C**(j,j+1) EO rise over [j,j+1]
d. *r* > **C**(j,j+1) EO fall over [j,j+1]

With these in hand, it's easy to see what happens as *r* increases. If *r* < **S**(1,2), then *r* is less than everything, and monotonic increase must occur (*cases a,c*).

When r = **S**(j, j+1) we have [j]=[j+1]. When r > **S**(j, j+1) we get a fall from [j] to [j+1], occasioning a minimum at [j+1] (*case b*). We also retain the falls [m, m+1], m < j, that have been occasioned by r passing **S**(m, m+1). In this way, a ripple of minima (and concomitant maxima) spreads rightward from node 1 as r increases, coming to a halt on the last even node before k, which we can call ev_k.

As r rises past the last of the **S**'s, we will get r = **C**(ev_k, ev_k+1), inducing [ev_k] = [ev_k+1] where previously we had [ev_k] < [ev_k+1]). With r > **C**(ev_k, ev_k+1), we now get a fall over [ev(k), ev(k)+1)] (*case b*). Recall that all [odd even] pairs are already falling. Thus, a uniformly falling ramp is extended from [k] back to [1] as r increases, until the whole of [1,...,k] is monotonically decreasing. □

To complete our basic description of $\mathbf{e}_k$* in the condition $\alpha\beta < 0$, we need to integrate the behavior of the left and right hemi-networks [1,...,k] and [k,...,n].

1. $|\alpha| \geq 1$. Then $|\beta|$ <1 since $|\alpha\beta| < ¼$. The left hemi-network shows oscillatory or falling behavior, the right side drops off (rapidly).

2. $|\alpha| < 1$. If we have β <1 too, then [k] is king, and both rise to it.

3. $|\alpha| < 1$ and $|\beta| \geq 1$. Then the right side rises or oscillates, and the left hemi-network falls off (rapidly).

We return below (''Discussion,'' p. 88) to characterization of the gross behavior of the model in the $\alpha\beta < 0$ condition, noting the following three points:

a. the narrow range of oscillatory behavior.

b. the tiny amplitude of oscillations

c. the extreme flatness of activation in oscillatory interval. $\alpha \approx 1$, $\beta \approx \alpha$-1.

****

To conclude the overall discussion, we examine the occurence and properties of absolute equality between neighboring nodes, under any and all conditions of convergence. This has been explicitly dealt with in the treatment of the case $\alpha\beta < 0$, and finessed in the case $\alpha\beta > 0$. Here we assault the problem directly.

Theorem 9. Any pair of adjacent nodes in $\mathbf{e}_k$* can be made exactly equal by appropriate choice of parameters α and β.

Proof. We are looking for [j] = [j+1]. Suppose j < k. From eq. (22), we want

(67)

$$\vec{e}_{jk}^{\,*} = K_{\alpha\beta n} \sqrt{\frac{\alpha}{\beta}}^{\,k-j} \sinh(n-k+1)u \;\sinh ju$$

$$= \vec{e}_{(j+1)k}^{\,*} = K_{\alpha\beta n} \sqrt{\frac{\alpha}{\beta}}^{\,k-j-1} \sinh(n-k+1)u \;\sinh(j+1)u$$

Dividing out the common material leaves this:
(68)

$$\sqrt{\left|\frac{\alpha}{\beta}\right|}\ \sinh ju \ = \ \sinh(j+1)u \ , \qquad i.e.$$

$$\sqrt{\left|\frac{\alpha}{\beta}\right|} \ = \ \frac{\sinh(j+1)u}{\sinh ju} \ , \qquad j<k$$

Recall that u is a function of $\alpha\beta$. But since $|\alpha/\beta|^{1/2}$ can be any positive number whatever, no matter what the value of $\alpha\beta$ is fixed at, this equation can always be satisfied.

Similarly, if $j \geq k$,

(69)
$$\sqrt{\left|\frac{\beta}{\alpha}\right|} \ = \ \frac{\sinh(n-j+2)u}{\sinh(n-j+1)u} \ , \qquad k \leq j \leq n$$

As above, such a $|\beta/\alpha|^{1/2}$ can be always be found. □

An interesting special case arises at either edge.

Corollary.

When $|\alpha| = 1$, then [1] = [2], if $k \neq 1$. That is, the first two nodes have equal activation in all $\mathbf{e}_k$* except $\mathbf{e}_1$*.

When $|\beta| = 1$, then [n] = [n−1] if $k \neq n$. That is, the last two nodes have equal activation in all $\mathbf{e}_k$* except $\mathbf{e}_n$*. □

Proof. Assume k > 1. Then we have

(70)

$$\sqrt{\frac{\alpha}{\beta}} \ = \ \frac{\sinh 2u}{\sinh u} \ = \ \frac{2\ \sinh u\ \cosh u}{\sinh u} \ = \ 2\ \cosh u$$

$$= \ \frac{1}{\sqrt{\alpha\beta}} \quad , \textit{from which}:$$

$$|\alpha| \ = \ 1$$

A similar argument applies to the mirror-image case k < n.

But *two* is the maximum number of adjacent equal nodes in $\mathbf{e}_k$*.

Theorem 10. In $\mathbf{e}_k$*, no more than 2 adjacent nodes may be equal.

Proof. Consider the case 1<j<k. We show that [j] cannot be flanked on both sides by elements equal to it. If [j] = [j+1], we must have

(71) $$\sqrt{\left|\frac{\alpha}{\beta}\right|} = \frac{\sinh(j+1)u}{\sinh ju} , \qquad j<k$$

If [j] = [j−1],we must have

(72) $$\sqrt{\left|\frac{\alpha}{\beta}\right|} = \frac{\sinh ju}{\sinh(j-1)u} , \qquad j<k$$

For these to hold simultaneously we must have

(73)

$$\frac{\sinh ju}{\sinh(j-1)u} \stackrel{?}{=} \frac{\sinh(j+1)u}{\sinh ju} , \quad i.e.$$

$$\sinh^2 ju \stackrel{?}{=} \sinh(j-1)u \ \sinh(j+1)u$$

But we have already seen the following identity:

(74) $$\sinh^2 ju = \sinh(j-1)u \ \sinh(j+1)u + \sinh^2 u , \quad j,u \in \mathbb{C}$$

It follows that eq. (73) can never be satisfied. Eqs. (73) and (74) both hold only when $\sinh^2$ u = 0, i.e. when u = 0 or u = iπ. But u is never 0 and when u=iΘ, Θ < π/(n+1), which means that Θ < π for all n > 0, i.e in all cases. This leaves only the case αβ = ¼, when the parameter *u* is unusable. Here we know from eq.(29) that the relevant condition (asymptotically reached by eq. (73)) is j/(j−1) = (j+1)/j, which can never happen.

The same argument applies in mirror-image form to the case k < j < n.

Suppose now j=k. We must examine the conditions

(75)

$$\frac{\vec{e}_{kk}{}^{*}}{\vec{e}_{(k-1)k}{}^{*}} = \frac{\sinh ku}{\sqrt{|\alpha/\beta|}\ \sinh(k-1)u} = 1 \quad , \ i.e.$$

$$\sqrt{|\alpha/\beta|} = \frac{\sinh ku}{\sinh(k-1)u}$$

(76)

$$\frac{\vec{e}_{kk}{}^{*}}{\vec{e}_{(k+1)k}{}^{*}} = \frac{\sinh(n-k+1)u}{\sqrt{|\beta/\alpha|}\ \sinh(n-k)u} = 1 \quad , \ i.e.$$

$$\sqrt{|\beta/\alpha|} = \frac{\sinh ku}{\sinh(k-1)u} \quad , \ i.e.$$

$$\sqrt{|\alpha/\beta|} = \frac{\sinh(k-1)u}{\sinh ku}$$

Taken together these require

(77)

$$\frac{\sinh(k-1)u}{\sinh ku} \stackrel{?}{=} \frac{\sinh(n-k+1)u}{\sinh(n-k)u} \quad , \ or$$

$$\sinh(k-1)u\ \sinh(n-k)u \stackrel{?}{=} \sinh ku\ \sinh(n-k+1)u \ , \quad u \neq 0$$

This is clearly impossible for u ∈ ℝ, since *sinh* is strictly increasing for real arguments. Note as u → 0, this becomes (k−1)/k = (n−k+1)/(n−k), which requires n=0 and so has no relevant solutions . In the general setting, u ∈ ℂ, we call on the following identity, a generalized form of the one cited above:

(78)

$$\sinh mx\ \sinh nx = \sinh^2 \frac{m+n}{2}x - \sinh^2 \frac{m-n}{2}x$$

Applying this to both sides of the last equation in (77) yields the following:

(79)

$$\sinh^2 \frac{n-1}{2}u - \sinh^2 \frac{n-2k+1}{2}u = \sinh^2 \frac{n+1}{2}u - \sinh^2 \frac{n-2k+1}{2}u \;, \quad i.e.$$

$$\sinh^2 \frac{n-1}{2}u = \sinh^2 \frac{n+1}{2}u \;, \quad i.e.$$

$$\sinh^2 \frac{n+1}{2}u - \sinh^2 \frac{n-1}{2}u = 0$$

We now apply identity (78) once again, to the l.h.s of the last equation in (79):

(80)
$$\sinh nu \,\sinh u = 0$$

This equation holds only when u = 0, for $u \in \mathbb{R}$; but $u \neq 0$. For $u = i\Theta$, we have additional solutions $\Theta = m\pi/n$. But $\Theta < \pi/(n+1)$, so there's no chance of this.□

Theorem 11. If an adjacent pair of elements in $\mathbf{e}_k$* is equal, then no other adjacent pair of elements in $\mathbf{e}_k$* is equal.

Proof. Omitted.

Theorem 12. Suppose [j] = [j+1]. If $\alpha\beta > 0$, then nodes j and nodes (j+1) share the maximum activation value in $\mathbf{e}_k$*. If $\alpha\beta < 0$, then they share some non-maximal value.

Proof. Omitted.

We conclude with an observation that is significant for linguistic applications of the DLM.

Theorem 13. *Theorema Egregium*. The consequences of choice of α, β for the location of maxima in $\mathbf{e}_k$* are independent of the length of the vector $\mathbf{e}_k$*.

Proof. This is entirely clear from the above discussion. Although *n* figures in the basic equations and affects the amplitude of activation and the interval of DLM convergence, none of the formulas derived above for the *location* of maximum depends on *n*. Maxima are located with respect to the edges or fall right on k. Freedom from influence of length holds both for absolute value of activation, on which we have focussed, and for the sign-pattern, which radiates outward from $\mathbf{e}_{kk}$*.

What this means is that the model-defining parameters α,β really define an infinite class of networks of different lengths, which share the significant property of maximum-location. The notion 'model' is thus more abstract than the notion 'network' in just the right way.

Remarks on the Canonical Models and their Relatives.

Here we examine some significant properties of the cases $\alpha\beta = 1/4$, which because of the simplicity of their solution deserve to be called the 'Canonical Models'. These properties are shared by its relatives in the class $\alpha\beta > 0$, and recognizing them can give insight into the entire class. The Canonical Models has two free parameters:

(i) the sign parameter: $\alpha,\beta > 0$, or $\alpha,\beta < 0$, and

(ii) the left-right asymmetry parameter determined by the ratio α/β, which enters into the solution as $r = |\alpha/\beta|^{1/2}$.

We will find that parameter (ii) can be readily reinterpreted to refer directly to the node-string rather than to a low-level property of the model.

The ratio α/β can be chosen put a maximum anywhere on a given $\mathbf{e}_k$*. For example, a straightforward calculation on the Canonical Models shows that for $r = |\alpha/\beta|^{1/2}$ chosen to meet the following restriction

(81) $$\frac{j}{j-1} > r > \frac{j+1}{j}$$

a maximum will be placed on [j] in $\mathbf{e}_k$*, if $j \le k$. (For $j \ge k$, the maximum falls on [k].)

From (42), we know that a maximum falls *exactly* on node j, $j \le k$, when

(82) $$j = \frac{1}{\log\sqrt{\alpha/\beta}} = \frac{1}{\log(2\alpha)} = \frac{-1}{\log(2\beta)}$$

This equation can be solved for α (or β, or r) in terms of j. This is the result:

(83) $$\alpha = \frac{1}{2} e^{1/j}$$ $$\beta = \frac{1}{2} e^{-1/j}$$ $$r = \sqrt{\alpha/\beta} = e^{1/j}$$

This means that the Canonical Models can easily be parametrized in terms of the *node* on which it places the maximum in the basis vectors $\mathbf{e}_k$*. (This is true for the other models as well, but the translation between maximal node and r is not attractive.)

Models for which $\alpha\beta > 0$ can be classified in terms of the sole maximum-node and the edge from which it is calculated. Let us say that we have the j-Model when r is chosen so that a maximum occurs on [j] in $\mathbf{e}_k$* if $j \le k$, and on [k] otherwise. Let us say that we have the (−j)-Model when r is chosen so the maximum occurs on [n−j+1] when $j \ge k$ and on [k] otherwise. In the Canonical (−j)-Model this happens with $r = e^{-1/j}$. This classification calls on two notions: the distance from an edge to the where the maximum wants to sit; and the choice of edge.

When $\alpha = \beta$ so that $r = 1$ and $\log r = 0$, a special condition prevails, since the associated continuous activation functions are monotonic: the maximum is not fixed with respect to an edge, but sits on [k] in $\mathbf{e}_k$*, whatever k may be. Let us call these the A-Models.

It is informative to examine how a particular choice of r affects the ensemble of $\mathbf{e}_k$*'s. To this end, let us contemplate a linguistic interpretation of the model. Think of the $\mathbf{e}_k$*'s as representing words *qua* syllable strings (which in fact needn't be all of the same length — we really need another subscript to indicate the dimension of the vector; we could write, say, ${}_m\mathbf{e}_k$ for the k^{th} canonical basis vector of $\mathbb{R}^m$; but let us put off such refinements). The unit bias in $\mathbf{e}_k$ can be taken to represent a lexical accent on the k^{th} syllable; all other syllables are unaccented. The vector $\mathbf{e}_k$* is the result of processing $\mathbf{e}_k$ through DLM — the phonology. Note that every lexical word is assumed to bear one and only accent.

Suppose we want the canonical 3-Model; $r = e^{1/3}$, from eq.(83). This means the continuous activation function f(x) associated with the initial hemi-network [1,...,k] rises to a maximum at x = 3 and falls therefrom. The corresponding activation function g(x) for the final hemi-network [k,...,n] is falling throughout. What becomes of the various $\mathbf{e}_k$ under this regime? The vector $\mathbf{e}_1$, representing initial lexical accent, has only the degenerate initial hemi-network [1], which therefore gets the maximum in $\mathbf{e}_1$*. The vector $\mathbf{e}_2$ intercepts the rise of f(x) at [2]; since g(x) falls thereafter, [2] bears the maximum in $\mathbf{e}_2$* . The vector $\mathbf{e}_3$* bears its maximum on [3]. So do all other vectors $\mathbf{e}_k$* for k > 3, which host a rise to [3] and fall from it inside the initial hemi-network [1,...,k]. The situation boils down to this: lexical accents on or before node 3 win out, and carry the activation maximum, but after node 3 the lexical accent has no tropism for the maximum, which falls only on [3].

The 3-Model, then, has a kind of barrier at node 3. If the bias in $\mathbf{e}_k$ lies beyond the barrier, it will only have the effect of supplying a reservoir of activation that pumps the maximum on [k]. But any bias which lies before the barrier, *i.e.* in [1,...,j] for the j-Model, gather the maximum to itself. The (−3)-Model, of course, shows the same behavior in mirror-image, counting from the end instead of the beginning.

We find, then, two broad classes of models, the A(ccenting)-Models and the B(arrier)-Models, where the latter fall into j-Models and (−j)-Models, depending on which edge is counted from. For the Canonical Models, the internal parameters of the networks associated with the ±j-Models can be easily determined from ±j itself.

Discussion

FORMAL. The findings reported here suggest certain reformulations of the model.

First, if the updating rule is modified to depend on the *average* of incoming activation rather than the simple sum, a number of cluttering factors of 2 and 4 will disappear from the basic equations. The new rule would look like this:

(84) $$a_k \leftarrow \tfrac{1}{2}\alpha \cdot a_{k+1} + \tfrac{1}{2}\beta \cdot a_{k-1} + b_k \,, \qquad 1 \le k \le n$$

Or, in matrix form:

(85) $$\vec{a} \leftarrow \tfrac{1}{2} \boldsymbol{W}_n \, \vec{a} + \vec{b}$$

This is equivalent to replacing Goldsmith-Larson α, β with $\alpha/2$, $\beta/2$. This done, we find that the DLM converges iff

(86) $$|\alpha\beta| < \frac{1}{\cos^2 \pi/(n+1)}$$

The parameter u is now subject to this condition:

(87) $$\cosh^2 u = \frac{1}{\alpha\beta}$$

Furthermore, the simple model found for $u \to 0$ is now defined by the condition $\alpha\beta = 1$.

A second observation is that α and β appear in the solution principally in the combinations $\alpha\beta$ and α/β (or β/α). This suggests that the basic parameters of the *model* should be conceived in terms of these combinations, with low-level network parameters α and β defined from these more basic elements.

Suppose we identify two parameters: r, a positive quantity measuring left-right asymmetry, and c, a quantity measuring the absolute magnitude of the activation. The following would provide one fairly natural correspondence with the familiar α, β:

(88) $$r^2 = \frac{\alpha}{\beta}$$

$$c^2 = \alpha\beta$$

Now α and β have the following definitions:

(89)
$$\alpha = r\,c \qquad \beta = \frac{c}{r}$$

Since *r* and *c* are positive by stipulation, we need to add another element to the activation rule to incorporate the possibility of sign-alternation . Let $\sigma_R = \pm 1$ be the sign attached to the activation coming in from the right, sgn(α); let $\sigma_L = \pm 1$ be the sign attached to activation coming in from the left, sgn(β). The iteration rule now becomes:

(90)
$$a_k \leftarrow \frac{c}{2}\,[a_{k+1}\sigma_R r + a_{k-1}\sigma_L r^{-1}] + b_k\,, \qquad 1 \le k \le n$$

The solutions for the Canonical Models $c^2 = 1$ now become

(91)

$$\vec{e}_{jk}^{\,*} = \sigma_R^{k-j}\,\frac{2\,(n-k+1)}{n+1}\,j\,r^{k-j}\,, \qquad j \le k$$

$$\vec{e}_{jk}^{\,*} = \sigma_L^{k-j}\,\frac{2k}{n+1}\,(n-j+1)\,r^{k-j}\,, \qquad k \le j$$

The parameter *u* meets the condition $c^2 \cosh^2 u = \sigma_R\sigma_L$ and the DLM converges when $c < \sec \pi/(n+1)$. Another plausible candidate for re-parametrization would take $c^2 = 1/|\alpha\beta|$. We could also contemplate going whole-hog and define α and β in terms of the parameters *a* and *u* which are central to the maxima-locating equations.

****

LINGUISTIC. A variety of linguistically promising characteristics of the DLM have been discussed by Goldsmith (1991). The patterns of sign-alternation are obviously relevant to the model's account of alternating stress patterns; and the single maximum phenomenon can be equated with Praguean *culmination* in stress or accent patterns.

Here we will mention several features of the model which emerge from the kind of analysis attempted in this paper. The goal is not to provide a tendentious catalogue of the model's deficiencies and excellences, but modestly to advance discussion of its properties. Some of these will, of course, call out for further explication, and occasional suggestions will be offered.

1. *Maxima*. We have seen that for the models $\alpha\beta > 0$, the parameters can be chosen put a maximum (more accurately, a barrier) anywhere in $\mathbf{e}_k$*, measuring from the edges, or to place a maximum on [k] in $\mathbf{e}_k$*

This finding has attractive characteristics.

(i) As noted above, Theorem 13, it is independent of n, the length of the string, an important result since string-length figures into the basic equations but not into principles of linguistic form.

(ii) Barrier-placement allows us to model a situation roughly like that seen in a number of accent and stress systems, whereby an accent or stress falls as far from an edge as it can, but no farther than some finite measure, usually three syllables (grossly speaking), with special lexical accents inside the barrier nevertheless being given priority.

However, a couple of remarks:

(i) The linguist's eye will struck by the fact that *j* can be any number, rather than just the few small ones (1, 2, 1+2) familiar from work on prosody. How can this locality property be obtained in a continuous model with an infinite parameter space? Although one might think that the model's parameters will have to be limited ad hoc to a narrow zone, there may well be a more natural approach. In the Canonical Models, for example, we are guaranteed a barrier on node *j* when for $r = (\alpha\beta)^{1/2}$ we have $j/(j-1) > r > (j+1)/j$. This divides up range of *r* in the following way:

Model #	*r Range*	*Length*
1-Model	(∞, 2)	∞
2-Model	(2, 3/2)	1/2
3-Model	(3/2, 4/3)	1/6
4-Model	(4/3, 5/4)	1/12
j-Model	(j/j-1, j+1/j)	1/j(j-1)

The range of *r*'s that will produce the *j*-Model narrows rapidly as *j* increases. (Indeed, it narrows very much more rapidly in the noncanonical models, where the length of the range for putting a maximum on *j* is 1/(sinh j)(sinh j−1), which falls off exponentially as opposed to quadratically.) This suggests that we can get the locality result by limiting the *accuracy* with which parameters can be set. Suppose, for example, that *r* can be specified with no greater accuracy that ±.1. Putting *r* = 3, say, guarantees a 1-Model, because *r* will range from 3.1 to 2.9. Putting *r* = 1.8 similarly guarantees a 2-Model. But since the interval for 3-Models is less that .17 wide, there is no choice within it that ensures stability of the model, if *r* is allowed to wander over a range of .2. A requirement that attained models must be stable, taken with a limitation on accuracy of parameter specification, can give a locality result.

The calculation for the (−j)-Models is similar, but shows an interesting shift, because of the way *r* is defined:

Model #	*r Range*	*Length*
−1	(0, 1/2)	1/2
−2	(1/2, 2/3	1/6
−3	(2/3, 3/4)	1/12
−4	(3/4, 4/5)	1/20
−j	(j-1/j, j/j+1)	1/j(j+1)

Where the j-Models deal with *r*, the (−j)-Models deal with *1/r*. An accuracy of .1, which would allow stable 1- and 2-Models, will only permit a stable (−1)-Model. If anything, the opposite result would be desirable: but this can be achieved by changing the definition of r to depend on β/α. Formulation of the parameter *r* must introduce a directional favoritism (which is the numerator?), which could emerge as descriptively valuable in the context of further assumptions about how the model is interpreted.

A further interesting question arises about the stability of A(ccenting)-Models, for which $r = 1$, exactly. With an accuracy of .09, the Canonical A-Model will be ambiguous between the 11-Model, the (−11)-Model, and the true article. This may be harmless or least invisible in practice. (Or perhaps there is something special about 1, so that it can be hit accurately?)

(ii) Equally striking is the high degree of independence of maximum-placement from the location of the unit input bias *k*. As noted, the Canonical Models can be straightforwardly parametrized in terms of the location of a barrier at *j*, which is measured relative to an edge. What matters is only where $\mathbf{e}_{kk}$, the intrinsic accent, sits with regard to this barrier. Patterns of accent, however, often are calculated in relation to *k* itself, rather than in terms of some absolute measure on the string. One obvious example is pre-accentuation, whereby a mark at position k entails an accent on position (k—1). The recessive accent of Ancient Greek, *e.g.*, though often described in terms reminiscent of the B-Models, is in reality a kind of pre-accenting system (Itô, Mester, and Prince, in prep.). But it is not obvious that there is a way to express this kind of dependency in the model.

2. *Shared Maxima*. We have seen that the models $\alpha,\beta > 0$ can always be configured to place the maximal activation value on two adjacent nodes. This would seem to be the wrong kind of culmination: placing a stress clash instead of a single peak. One might think that this could be interpreted as pre- or post-accentuation. But note that the shared maximum is not a property of the A-Models; the implicit maxima of their continuous activation functions are at $\pm\infty$. The shared maximum is a phenomenon of the B-Models, and is therefore tied to a specific string location. It works out like this: suppose the parameters are chosen so that the maximum is shared between j and j+1 to the left of k ($j<k$, $r>1$); then $\mathbf{e}_k$* gets a single maximum on k if $k \leq j$; but for all $k > j$, we have [j] = [j+1] and there is a double maximum shared across [j, j+1]. Again, it appears that further understanding of how parameters are set is needed to eliminate this possibility.

3. *Sign Pattern*. We have found that the sign pattern of $\mathbf{e}_k$* is determined by the signs of α and β. Consider the case $\alpha\beta < 0$; say w.l.o.g. that $\alpha > 0$, $\beta < 0$. Then the segment [1,...,k] is entirely positive, but in [k,...,n] signs alternate from k: [+,-+,-....]. This configuration stands in need of a linguistic correspondent.

4. *The Ripple*. The ripple of maxima/minima spreading in from the edge in models for which $\alpha\beta < 0$, although of great formal interest, is probably not interpretable linguistically. For one thing the magnitude of the ripples is extremely slight. The very existence of the ripples depends on the difference between cosh u and sinh u, e^{-u}, which hits .1 by u= 2.3, .05 at u=3, and so on. The real significance is probably that region in which they occur is extremely, if of course not perfectly flat. This can be grasped grossly from the structure of network, where the fact that α, say, is near one, and β therefore relatively near 0, means that immediately copied material from the right, the α direction, is far more significant that what comes in, tagged by powers of β, from the left. More precisely, consider what it would mean to have a relatively flat stretch going leftward from $\mathbf{e}_{kk}$*:

(92)
$$\frac{\vec{e}_{kk}^{\;*}}{\vec{e}_{jk}^{\;*}} = \frac{\sinh Uk}{r^{k-j}\sinh Uj} \approx 1 , \qquad j<k$$

Now, sinh u can be approximated by $e^u/2$, with an error of only $½e^{-u}$, so we seek

(93)
$$e^{-(k-j)U} \approx r^{k-j} , \qquad i.e.$$
$$r \approx e^{U}$$

In fact, r can easily be made *exactly* equal to e^U. We have

(94)
$$U = \sinh^{-1}\frac{1}{2\sqrt{|\alpha\beta|}} = \log\left(\frac{1 + \sqrt{1 + 4|\alpha\beta|}}{2\sqrt{\alpha\beta}}\right)$$

We want
(95)

$$r = \sqrt{\left|\frac{\alpha}{\beta}\right|} = e^{U} = \frac{1 + \sqrt{1 + 4|\alpha\beta|}}{2\sqrt{|\alpha\beta|}} , \qquad so\ that$$
$$2|\alpha| - 1 = \sqrt{1 + 4|\alpha\beta|} , \qquad or$$
$$4|\alpha|(|\alpha| - 1) = 4|\alpha\beta| , \qquad i.e.$$
$$|\beta| = |\alpha| - 1$$

So when $|\beta| = |\alpha| - 1$, we have $r = e^U$. Now it is easily seen that e^U stands exactly in the middle of the scale of ratios discussed above near (66); consequently, for this value of *r*, the alternating pattern runs as far in as it can. But we also have, from approximations (92) and (93), that the alternating stretch is quite flat, since r^{-j} essentially cancels out the sinh Uj and cosh Uj terms, as long as e^U is a good approximation of sinh U and cosh U.

A better, grosser characterization of the behavior of the $\alpha\beta < 0$ case with respect to its treatment of $\mathbf{e}_k$, then, would treat it divide it into 3 regions, which we catalogue with left-right orientation, which is subject to mirror-image reversal by replacing α with β:

(i) $|\alpha| < 0$, rise to [k].

(ii) $1<\alpha<(1-|\beta|)^{-1}$, a relatively narrow band of essentially flat activation in [1,...,k], followed by a very rapid fall off in [k,...,n]; and

(iii) $|\alpha| > (1-|\beta|)^{-1}$, fall to [k], followed by extremely rapid fall off in [k,...,n].

In terms of the configuration of absolute value of activation (which is what we are discussing), what distinguishes the $\alpha\beta < 0$ cases from the $\alpha\beta > 0$ cases is the lack of a significant internal maximum or barrier that is placed at some fixed distance from the edge. What depends on measurement from the edge in the $\alpha\beta < 0$ models is the extent of progress of the ripple; if we regard that as essentially flat rather than alternating, then nothing is left that depends absolute string coordinates. The $\alpha\beta < 0$ system is thus very k-dependent, responsive to the position of the unit bias in $\mathbf{e}_k$, and is perhaps suitable for representing effects like spreading of pre-attached tones.

5. *Mirror-Image symmetry*. The model has, of course, no intrinsic directional bias: any phenomenon can be replicated in mirror-image form. Linguistic prosody is known to be asymmetric in fundamental respects (Hayes, 1985, et seq.); but the principles guaranteeing this must be added to the bare-bones time-symmetric formal theory. What's interesting, and perhaps unexpected, is that the full mirror-image symmetry of the model, taken with its linearity, provides considerable powers of global string analysis.

Note first that for *r*=1 there is perfect mirror-image symmetry among the $\mathbf{e}_k$*, so that $\mathbf{e}_k{}^* = R(\mathbf{e}_{n-k+1}{}^*)$, where *R* is reflection about the mid-point (equivalently, string-reversal); algebraically $m \mapsto m-k+1$ for indices *m*, so that $\mathbf{e}_{jk}{}^* = \mathbf{e}_{(n-j+1)(n-k+1)}{}^*$. The symmetry is perfect because the exponential term, which is based on r, remains at 1 and has a uniform (non-)effect on the activation product. The other terms depend on count from either edge, which transforms properly under reflection.

Consider the vector $\Sigma_k \mathbf{e}_k = (1,1,1,...,1)$, call it Σ, which has 1's in every position, a plausible candidate for representing a string of syllables undifferentiated as to quantity. Each $\mathbf{e}_k$* peaks at $[k] = \mathbf{e}_{kk}{}^*$. Since $\mathbf{e}_k{}^* = R(\mathbf{e}_{n-k+1}{}^*)$, the vector Σ* will be mirror-image symmetric about its midpoint. (Summation, to put it in the Saussurean-Jakobsonian manner, projects the paradigmatic symmetry of the $\mathbf{e}_k$* into syntagmatic symmetry in Σ*.)The symmetry is established by the following derivation:

(96) $$\vec{\Sigma}_p^{\,*} = \left(\sum_{k=1}^{n} \vec{e}_{pk}\right)^{*} = \sum_{k=1}^{n} \vec{e}_{pk}{}^{*} = \sum_{k=1}^{n} \vec{e}^{\,*}_{(n-p+1)(n-k+1)} = \vec{\Sigma}^{\,*}_{n-p+1}$$

This leads to a stark differentiation of even-length from odd-length Σ's. If Σ is of length 2q, then [q] and [q+1] are identical. But if Σ is of length 2q+1, [q+1] stands alone.
The properties of the vector Σ* become fully transparent in the Canonical A-Model (r = 1 and $\alpha = \beta = \pm½$), where it has a remarably simple formulation. Let us first consider the case $\alpha = \beta = +½$. Then (it is easy to show that) Σ_j*, the jth node in Σ*, comes out like this:

(97)
$$\Sigma_p^* = p\,(n - p + 1)$$

The value of Σ_j*, then, is just the product of the length of [1,...,j] times the length of [j,...,n].

In Σ of length 2p, the value of Σ_j* rises from Σ_1* = 2p to $\Sigma_p = \Sigma_{p+1} = p^2$ and falls symmetrically back to Σ_{2p} = 2p. In Σ of length 2p+1, the value of Σ_j* rises from Σ_1* = 2p to Σ_{p+1} = p(p+1) and falls symmetrically back to Σ_{2p} = 2p.

The even-length Σ* thus has a high plateau straddling its two halves; the odd-length counterpart has a single peak right in the middle. In this way, the DLM can distinguish odd from even without invoking binary alternation, the standard linguistic source of parity-sensitivity. The DLM can locate the exact center of a string, another distinctly non-linguistic ability of global string analysis. Furthermore, the fact that even-length strings have a shared maximum whereas odd-length strings have a sole peak does not have an obvious linguistic analog. (The fact that the exact length of the string can be read right off the activation values may be, however, of no significance, if it is not the actual value but the relation of values that is empirically interpretable.)

Let us now turn to the case $\alpha=\beta=-½$. Here the distinction between even and odd lengths is perhaps even more striking. A general formula for the activation of Σ_j* can be given, using a parity function π, defined as π(m) = 0 for even m, 1 for odd m:

(98)
$$\vec{\Sigma}_j^* = \vec{\Sigma}^*_{n-j+1} = \frac{j\ \pi(n-j+1)\ +\ (n-j+1)\ \pi(j)}{n+1}\ , \quad j \leq n/2$$

When n, the length of Σ, is odd, this gives the following:

(99) *Length n of Σ odd*

$$\vec{\Sigma}_j^{\,*} = 1 \,, \qquad j \text{ odd.}$$

$$\vec{\Sigma}_j^{\,*} = 0 \,, \qquad j \text{ even.} \quad \textit{More concisely},$$

$$\vec{\Sigma}_j^{\,*} = \pi(j)$$

This is an alternating pattern, with maxima on the odd nodes.

For n even, the following result devolves from eq. (98):

(100) *Length n of Σ even*

$$\vec{\Sigma}_j^{\,*} = \frac{j}{n+1} \,, \qquad j \text{ even.}$$

$$\vec{\Sigma}_j^{\,*} = \frac{n-j+1}{n+1} \,, \qquad j \text{ odd.}$$

Given Σ^* of even overall length, a node j (i.e., Σ_j^*) within it, for j even, has its activation equal to the length of the initial hemi-network containing it, [1,...j], divided by (n+1). A node Σ_k^*, k odd, has its activation equal to the length of the final hemi-network containing it, [k,...,n], divided by (n+1). (The factor (n+1) measures the length of the whole network.)

This generates an alternating pattern, with mirror-image symmetry between the two halves of the vector Σ^*: if its length is 2p, then the mirrored halves are [1,...,p] and [p+1,..., 2p]. It is clear that within the first half the Σ_j^* are maxima for odd j, inasmuch as $(2p-j+1) > j-1$ for $p > j-1$, and $(2p-j+1) > j+1$ for $p > j$. The same pattern alternates backward from node 2p.

Whether node p is a maximum within [1,...,p] depends (therefore) on whether p itself is odd or even. Odd [p] has activation (2p - p +1)/n+1 = p+1/n+1, which is greater than its predecessor [p-1], with activation p-1/n+1. Even-numbered [p] has activation p/n+1, but its predecessor [p-1] has activation (2p -(p-1)+1)/n+1 = p+2/n+1.

The outcome is that for even length 2p, Σ^* *never* shows an alternating pattern. Rather, it divides cleanly into two alternating halves, which meet at the join, putting maximum against maximum, minimum against minimum. If p itself is even, so that Σ^* is of length 4q, then there is a ‘lapse’ consisting of adjacent one-sided minima dead in the middle of the vector. If p itself is odd, then two one-sided *maxima* meet. For even length Σ^*, then, either a stress-clash or a stress-lapse is generated, and generated in the exact center of the string.

Obviously, such parity-dependent effects (and others that could be cited for Σ^* as *r* varies) are quite unlike those in prosodic systems, which 'count' very locally. Furthermore the constraint against clash is thorough-going and lapses are allowed predominantly at edges; certainly there is no tendency to place them *in medias res*, which is not a concept of linguistic form at any rate.

This suggests that the linearity of the Goldsmith-Larson DLM will eventually have to be compromised to achieve greater fidelity to the locality of linguistic structure.

6. *The 0-Models*. The models where $\alpha\beta = 0$ are the simplest and most easily analyzed, but are not for that reason to be regarded as trivial. When α or $\beta = 1$, they are the closest in some ways to established linguistic representations in their behavior, and therefore mark the starting point from which the value of departure should be argued.

Suppose $\alpha = \beta = 0$. Then there is no influence of neighbors upon each other and $\mathbf{e}_k^* = \mathbf{e}_k$. The form derived is the same as the input form or bias vector. For small values of α,β the derived form will be *similar* to the input, though of course the entire range of *r*-dependent behaviors can be evoked, writ in small differences of magnitude of activation.

Suppose $\alpha = 1$, $\beta = 0$. Here, after $(n-1)$ iterations, the end state is reached, in which $\mathbf{b}_k^* = \Sigma_{i\geq k}\, \mathbf{b}_i$; that is, the final derived activation of node k is just the sum of the biases to its right, plus its own innate bias. More generally, $\mathbf{b}_k^* = \Sigma_{i\geq k}\, \mathbf{b}_i \alpha^{k-i}$. For the $\mathbf{e}_k^*$, this means that every node in the the left hemi-network [1,...,k] has the activation 1. (The sum-of-biases expression contains only one non-zero term, the bias of $\mathbf{e}_{kk}$.) Every node in [k,...,n] retains its initial activation of 0. (Note the similarity to the ripple condition in the $\alpha\beta < 0$ models, where $\alpha\approx 1$ and $\beta\approx 0$.) For $|\alpha| < 0$, there is a straight exponential decay back across [1,...,k]; for $|\alpha| > 1$, there is an exponential explosion; in both cases the sequence of node values is $[\alpha^{k-1}],[\alpha^{k-2}],...,[\alpha^{0}]$.

For $\alpha = 1$, the initial plateau on [1,...,k] could be likened to the spreading of a tone back from position k. The role to be played by the exponential rise or fall is less clear, and perhaps requires a commitment to finer-grained phonetic interpretation of activation levels before its utility comes to the fore.

Suppose $\alpha = -1$, $\beta = 0$. Models of this character, with $\alpha < 1$, have been proposed by Goldsmith (1991a) for the analysis of Weri, Warao, and Maranungku. Now $\mathbf{e}_k^*$ will show a simple alternating pattern $...,-1,+1,-1,1_k$, spreading back from node k, with everything to the right of it remaining 0. This then resembles a 'perfect grid' configuration, or half of one. (Fully symmetric alternation spreading outward from node k can only be achieved in the DLM with both α and β negative, and will therefore be accompanied by other effects.) For $\mathbf{e}_1$ and $\mathbf{e}_n$, with intrinsic initial or final stress, this will spread a perfect grid directionally away from the edges, contingent upon the presence of an edge stress, in the manner suggested in Prince(1983:51) and van der Hulst (e.g. 1991).

Suppose now that the input is $\mathbf{e}_1 + \mathbf{e}_n$. If *n* is even (modelling even-syllabled words), then we generate a pattern (0,+1,−1,+1,...,1), since the incoming alternating pattern cancels the intrinsic 1st-node unit bias on $\mathbf{e}_1$, preventing, as it were, a stress-clash from developing. Even more interesting is the effect of exponential decay. Say $\alpha = ½$. Then we have, for $(\mathbf{e}_1 + \mathbf{e}_n)^*$, *n* even, the vector $(1-(½)^{n-1}, ½^{n-2}, -½^{n-3},..., -½, 1)$. In this case, the potential clash is resolved differently: in sufficiently long words, the first maximum falls on [1], the next on [4], with strict alternation thereafter. This is the familiar Tátamagouchi (as opposed to Atcháfalaya) pattern, and related material has been discussed with in Goldsmith (1991b). An interesting feature of the account is the dependence between word length and extent of decay, which is not obviously attested. Note, in particular, that choice of parameters can impose a distinction between short and long words. Say $\alpha = .8$. Here is a table of relevant patterns for $(\mathbf{e}_1 + \mathbf{e}_n)^*$ as *n* increases:

n	*Input Bias*	*Output Pattern*	*Maxima*
2	1, 1	0.2, 1	- +
3	1, 0, 1	1.6, −.8, 1	+ - +
4	1, 0, 0, 1	**.5**, **.6**, −.8, 1	**- +** - +
5	1, 0, 0, 0, 1	1.4, −.5, .6, −.8, 1	+ - + - +
6	1, 0, 0, 0, 0, 1	**.7**, **.4**, −.5, .6, −.8, 1	**+ -** - + - +

Observe that the relation between nodes 1 & 2 is different in the 4-node and 6-node cases. Furthermore, the parameter α can be chosen so that in even-syllabled words of some particular length, the first two syllables are exactly equal, guaranteeing clash; one need merely find the value of α that satisfies the equation $1 - \alpha^{n-1} = \alpha^{n-2}$; such a value always exists. (Perhaps this can be avoided by the limitation on accuracy suggested above, though.) Here again, the DLM shows an expressive capacity that exceeds that of formal prosody, and it remains to be seen whether this is a useful subtlety or not. If it is, and if the the severe locality of standard prosodic theory is an approximation that must be surpassed, one must then inquire as to how the choice of α,β can be limited so as to make most readily accessible the unmarked zones of the parameter space .

Conclusion, with Retrospective Prolepsis.

The Goldsmith-Larson DLM represents a significant departure from familiar methods of combining constraints to arrive at a linguistic description. Although defined in terms of a process of iterative computation, the DLM — a low-dimensional non-nonlinear dynamical system — submits to exact solution. Here the solution has been derived and some initial steps have taken in analytical exploration of the model’s properties, concentrating on its linear structure, omitting consideration of thresholds and other refinements. It is to be hoped that this will, along with the extensive experimental work of Goldsmith and Larson [refs.], lead to a deeper understanding of what makes the model work, of how it compares with and differs from familiar symbol-processing models, and how it may be usefully modified and interpreted (see Part I above for extensions of the analysis in these directions).

★★★★★

Appendix

1. Eigenvector Notes

We have the eigenvalues: $\lambda_m = 2\,(\alpha\beta)^{1/2} \cos\ m\pi/(n+1)$. We want the eigenvectors.
We need to solve, for each λ_m, the set of equations:

$$(W - \lambda_m I)\,\mathbf{x} = 0$$

This amounts to solving the set

$$\beta x_{k-2} - \lambda_m x_{k-1} + \alpha x_k = 0$$

This is nothing more than a recurrence relation cum 2nd order homogeneous difference equation, of the type solved previously ("Convergence").

Let x_{km} be the k^{th} coordinate of the m^{th} eigenvector and $\theta = m\pi/(n+1)$. Set $x_{0m} = 0$, and choose x_{1m} to be $(\beta/\alpha)^{1/2}$. Performing the calculation nets the following result:

$$x_{km} = (\mathrm{sgn}\ \alpha)^{k+1}\,(\beta/\alpha)^{k/2} \sin(k\theta)/\sin\theta$$

For case $\lambda = 0$, which arises with the middle eigenvalue in odd-length nets, for which $m = (n+1)/2$, so that $\theta = \pi/2$, this boils down to

$$x_{km} = (\beta/\alpha)^{k/2} \sin(k\pi/2)$$

where the sin term just says: 0 for even k, +1 for $k \equiv 1 \bmod 4$, -1 for $k \equiv 3 \bmod 4$.
Note that this solution is (necessarily) unique only up to a free multiplicative factor that can be chosen independently for each vector.

The basic fact is that for b_m the mth eigenvector we have:

$$b_m^* = 1/(1-\lambda_m)\ b_m$$

Thus for $\lambda = 0$, the output is an exact copy of the input.
Only for $\alpha\beta > 0$ are the eigenvectors real, and therefore representative of a plausible input vector.

2. Polynomial Expression for the Determinant of $(\mathbf{I}-\mathbf{W}_n)$

In the main body of the text (e.g., p. 59) the determinant of $E_n = (\mathbf{I}-\mathbf{W}_n)$ is given in terms of radicals and is reduced to a more analytically tractable form by a hyperbolic substitution. Direct expansion of the determinant yields a polynomial in $\alpha\beta$, which we display here, without proof:

$$E_n = \sum_{k=0}^{[n/2]} (-1)^k \binom{n-k}{k} \alpha^k \beta^k$$

By $[x]$, we mean the greatest integer not exceeding x.

***

References

Liberman, Mark. 1975. *The Intonational System of English*. Doctoral Dissertation MIT. Published 1979, Garland Press: New York.

Goldsmith, John. 1991a. ‘‘Phonology as an intelligent system.’’ In *Bridges between psychology and linguistics: a Swarthmore festschrift for Lila Gleitman*, eds. Donna Jo Napoli and Judy Kegl. Erlbaum: Hillsdale, NJ.

Goldsmith, John. 1991b. ‘‘Harmonic Phonology,’’ ms. U. of Chicago, to appear in volume of unknown title.

Goldsmith, John. 1991c. ‘‘Dynamic Computational Models: Accent Systems and Sonority,’’ paper delivered at University Illinois conference, ‘‘The Organization of Phonology.’’

Goldsmith, John. 1992. ‘‘Local modelling in phonology,’’ in *Connectionism: Theory and Practice*, ed. Steven Davis. London: Oxford University Press.

Goldsmith, John. In press. ‘‘Dynamic computational models,’’ in *Principles of Organization in Phonology*, eds. Jennifer Cole and Charles Kisseberth. Stanford: CSLI. (Written version of Goldsmith 1991x.)

Goldsmith, John and Gary Larson. 1990. ‘‘Local modelling and syllabification,’’ in *Chicago Linguistic Society Parasession on the Syllable in Phonetics and Phonology 26:2*, eds. Karen Deaton, Manuela Noske, and Michael Ziolkowski. Chicago: CLS.

Hayes, Bruce. 1985. ‘‘Iambic and Trochaic Rhythm in Stress Rules,’’ BLS 11, 429-445.

Itô, Junko, Armin Mester, and Alan Prince. In prep. ‘‘Recession.’’

Kager, René. 1992. ‘‘Stress in windows.’’ Ms. University of Utrecht.

Hulst, Harry van der. 1991. *Notes on the Representation of Stress*, ms. University of Leiden, Dept. of General Linguistics.

Larson, Gary. 1992. *Dynamic computational networks and the representation of phonological information.* Doctoral Dissertation, University of Chicago. Chicago.

Prince, Alan. 1983. ‘‘Relating to the Grid,’’ *Linguistic Inquiry* 14.1, 19-100.

Prince, Alan. 1991a. ‘‘Convergence of the Goldsmith-Larson Dynamic Linear Model of Sonority and Stress Structure,’’ ms. Brandeis University. Waltham, MA.

Prince, Alan. 1991b. ‘‘Closed-form Solution of the Goldsmith-Larson Dynamic Linear Model of Sonority and Stress Structure,’’ ms. Brandeis University. Waltham, MA.

Prince, Alan. 1992. ‘‘Remarks on the Goldsmith-Larson Dynamic Linear Model as a Theory of Stress, with Extension to the Continuous Linear Theory, and Additional Analysis,’’ ms. Brandeis University and Rutgers University Center for Cognitive Science. Issued prematurely as RuCCS-TR-1.

Smolensky, P. 1988. ‘‘On the proper treatment of connectionism,’’ *The Behavioral and Brain Sciences*, 11:1−74.