ML MC Forks: Closed Form Analytic Solutions*

Benny Chor† Sagi Snir‡

March 12, 2004

Abstract

Maximum likelihood (ML) is increasingly used as an optimality criterion for selecting evolutionary trees (Felsenstein, 1981), but finding the global optimum is a hard computational task. Because no general analytic solution is known, numeric techniques such as hill climbing or expectation maximization (EM), are used in order to find optimal parameters for a given tree. So far, analytic solutions were derived only for the simplest model - *three* taxa, two state characters, under a molecular clock (MC). Quoting Ziheng Yang (2000), who initiated the analytic approach, "*this seems to be the simplest case, but has many of the conceptual and statistical complexities involved in phylogenetic estimation*".

In this work, we give general analytic solutions for a family of trees with *four* taxa, two state characters, under a molecular clock. The change from three to four taxa incurs a major increase in the complexity of the underlying algebraic system, and requires novel techniques and approaches. We start by presenting the general maximum likelihood problem on phylogenetic trees as a constrained optimization problem, and the resulting system of polynomial equations. In full generality, it is infeasible to solve this system, therefore specialized tools for the MC case are developed.

Four taxa rooted trees have two topologies – the *fork* (two subtrees with two leaves each) and the *comb* (one subtree with three leaves, the other with a single leaf). We combine the ultrametric properties of MC fork trees with the Hadamard conjugation (Hendy and Penny, 1993) to derive a number of topology dependent identities. Employing these identities, we substantially simplify the system of polynomial equations for the fork. We finally employ symbolic algebra software to obtain *closed form* analytic solutions (expressed parametrically in the input data). In general, four taxa trees can have multiple ML points (Steel, 1994, Chor *et. al.*, 2001). In contrast, we can now prove that each fork topology has a *unique* (local and global) ML point.

*Research supported by ISF grant 418/00. Part of these results were presented at the RECOMB03 conference in Berlin.

†School of Computer Science, Tel-Aviv University, Tel-Aviv 39040 Israel. benny@cs.tau.ac.il

‡Computer Science dept., Technion, Haifa 32000, Israel. ssagi@cs.technion.ac.il

1 Introduction

The study of evolution and the construction of phylogenetic (evolutionary) trees are classical subjects in biology. DNA sequences from a variety of organisms are rapidly accumulating, providing the data to a number of sequence based approaches for phylogenetic trees reconstruction.Given a set of n aligned *sequences*, the goal is to find the best explanation for the data within the model space. Among tree reconstruction approaches, maximum likelihood (Felsenstein, 1981) is increasingly used as an optimality criterion for inferring trees. In the phylogeny context, this usually means a weighted tree (the weights are parameters of the substitution model for each edge) that maximizes the likelihood of generating the observed sequences.

Maximum likelihood (ML) algorithms are computationally intensive, but for tractable cases ML is the method of choice. Because no general analytical solution is available, numeric techniques (such as hill climbing or expectation maximization), are used in order to find optimal likelihood values for any given tree. The first to consider *analytical solutions* for simple substitution models with a small number of taxa was Yang (2000), who worked on three taxa with two state characters under molecular clock [13]. Yang calls this "the simplest phylogeny estimation problem", but adds that it "has many of the conceptual and statistical complexities involved in phylogenetic estimation". The solution of Yang was generalized and its derivation was simplified by Chor, Hendy and Penny [2] using the Hadamard Conjugation of Hendy, Penny, and Steel 1994 [5, 6], together with convexity arguments.

In this work we retain the symmetric two states model of Neyman [9], as used in [13, 2] under molecular clock, but increase the number of taxa to four. The change from three to four taxa incurs a major increase in the complexity of the underlying system of polynomial equations, and requires novel techniques and approaches. Our starting point, like [1], is to formulate the ML problem as one of constrained optimization, and express it in terms of Lagrange coefficients. We use the Hadamard conjugation [5, 6] to simplify the resulting system of polynomial equations. This yields a system of nine degree 5 polynomials in nine variables. This system is substantially more complex than the three taxa system [2], and is not solvable by current techniques. (The analytical solutions in [1] were all for special cases where at least 2 out of the 7 input parameters are 0.)

There are two families of rooted topologies on four taxa: Topologies with two taxa in each subtree of the root, which we call *fork* topologies, and topologies where one subtree of the root has three taxa, which we call *comb* topologies. Under molecular clock, the distance from each of the four leaves to the root is the same (Figure 1). In this work we focus on the fork topology. We first derive identities that help simplify the general system of equations for the fork topology. By using these identities, the system of equations becomes

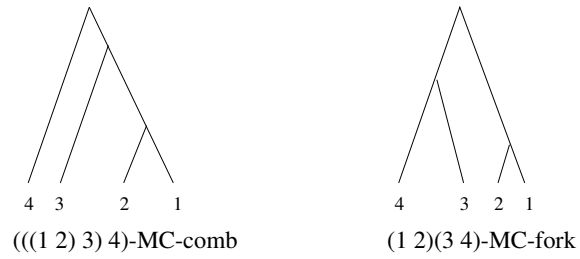

Figure 1: The fork and comb – two rooted topologies on four taxa.

simple enough it is now solvable by computer algebra tools (*e.g.* `Maple`). This leads to the derivation of *closed form* analytical solutions, expressed as rational functions in the input parameters. This solution is unique, implying a single local and global ML point. It is important to understand that we express the problem and solve it using the so called *sequence spectrum* representation [5]. This representation is closely related to the more standard edge representation. However there are rare cases where a seemingly innocent sequence spectrum gives rise to an *unreasonable tree*: Either a tree with a *negative edge*, or a tree where the internal edge is too short, so the length of one external edge (leading to a leaf) is larger than the length of another external edge *plus* the length of the internal edge of the tree. Such a tree may arise as a global ML point, and if this is the case the ML solution found by maximizing over the the *edge space* will produce solutions that are *on the boundary*. (Length 0 edge(s) or an external edge whose length equals the sum of lengths of another external edge and the internal edge). In non pathological cases, though, the global ML point gives rise to a reasonable tree (with positive edges, and long enough internal edge) that resides in the interior of the solution space.

Even in cases where it is feasible to derive them, analytical solutions will most probably *not* replace numeric approaches in ML based tree reconstruction packages. But the analytic solutions do reveal properties of the maximum likelihood points that are not obtainable numerically. For example, we show that every molecular clock fork tree has a unique (global and local) ML point. Without the molecular clock hypothesis, this uniqueness does not hold, as proved in [10, 1].

2 Definitions, Notations, and the Hadamard Conjugation

In this section we define the model of substitution we use, introduce useful notations, and briefly describe the Hadamard conjugation.

2.1 Definitions and Notations

We start with a tree labeling notation that will be useful for the rest of the work. For simplicity we use four taxa, but the definitions extend to any n. A *split* of the species is any partition of $\{1,2,3,4\}$ into two disjoint subsets. We will identify each split by the subset which does not contain 4 (in general n), so that for example the split $\{\{1,2\},\{3,4\}\}$ is identified by the subset $\{1,2\}$. For brevity, to label objects subscribed by a split we concatenate the members of the split. Each edge e of a phylogenetic tree T induces a split of the taxa, i.e. the cut induced by removing e. We denote the edge e by the cut it induces. For instance the central edge of the tree $T = (12)(34)$ induces the split $\{\{1,2\},\{3,4\}\}$, that is identified by the subset $\{1,2\}$ and therefore this edge is denoted e_{12}. Thus $E(T') = \{e_1, e_2, e_{12}, e_3, e_{123}\}$ (see Figure 2).

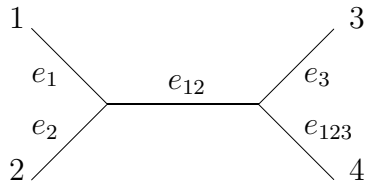

Figure 2: The tree $T' = (12)(34)$ and its edges

In Neyman 2 states model [9], each character at a species admits one out of two states, without loss of generality $\{\mathtt{x},\mathtt{y}\}$. Hence, a character evolving along an evolutionary tree T with n leaves, induces a split pattern between the leaves admitting the state $\mathtt{x}$ and $\mathtt{y}$.

In the 2 states model, The length of an edge q_e, $e \in E(T)$ in the tree T is defined as the expected number of substitutions (changes) per site along that edge. Given the edge lengths of T: $\mathbf{q} = [q_e]_{e\in E(T)}$ $(0 \le q_e < \infty)$, the probability of generating an α-split pattern $(\alpha \subseteq \{1,\ldots,n-1\})$ is well defined. Denote this probability by $s_\alpha = Pr(\alpha\text{-split}|T,\mathbf{q})$. Using the same indexing scheme as above, we define the *expected sequence spectrum* (expected spec) $\mathbf{s} = [s_\alpha]_{\alpha\subseteq\{1,\ldots,n-1\}}$.

The *edges lengths* spectrum (edges spec) of a tree T with n leaves is the 2^{n-1} dimensional vector $\mathbf{q} = [q_\alpha]_{\alpha\subseteq\{1,\ldots,n-1\}}$, defined for any subset $\alpha \subseteq \{1,\ldots,n-1\}$ by

$$q_\alpha = \begin{cases} q_e & \text{if } e \in E(T) \text{ induces the split } \alpha\ , \\ -\sum_{e\in E(T)} q_e & \text{if } \alpha = \emptyset\ , \\ 0 & \text{otherwise.} \end{cases}$$

2.2 Hadamard Conjugation

The Hadamard conjugation [5, 6] is an invertible transformation that specifies a relation between the expected sequence spectrum $\mathbf{s}$ and the edge lengths spectrum $\mathbf{q}$ of the tree. In other words, the transformation links the probabilities of site substitutions on edges of an evolutionary tree T to the probabilities of obtaining each possible combination of characters. The Hadamard conjugation is applicable to a number of site substitution models: Neyman 2 state model, Jukes–Cantor model [7], and Kimura 2ST and 3ST models [8] (the last two are applicable to "normal", four states DNA). For these models, the transformation yields a powerful tool which greatly simplifies and unifies the analysis of phylogenetic data, and in particular the analytical approach to ML.

Definition 1 *A* Hadamard matrix *of order ℓ is an $\ell \times \ell$ matrix A with ± 1 entries such that $A^t A = \ell I_\ell$.*

We will use a special family of Hadamard matrices, called Sylvester matrices in MacWilliams and Sloan (1977, p. 45), defined inductively for $n \geq 0$ by $H_0 = [1]$ and $H_{n+1} = \begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix}$. For example,

$$H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{ and } H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

It is convenient to index the rows and columns of H_n by lexicographically ordered subsets of $\{1, \ldots, n\}$. Denote by $h_{\alpha,\gamma}$ the (α, γ) entry of H_n, then $h_{\alpha,\gamma} = (-1)^{|\alpha \cap \gamma|}$. This implies that H_n is symmetric, namely $H_n^t = H_n$, and thus by the definition of Hadamard matrices $H_n^{-1} = \frac{1}{2^n} H_n$.

Proposition 1 *(Hendy and Penny 1993) Let T be a phylogenetic tree on n leaves with finite edge lengths ($q_e < \infty$ for all $e \in E(T)$). Assume that sites mutate according to a symmetric substitution model, with equal rates across sites. Let $\mathbf{s}$ be the expected sequence spectrum. Then*

$$\mathbf{s} = \mathbf{s}(\mathbf{q}) = H_{n-1}^{-1} \exp(H\mathbf{q}) \ ,$$

where the exponentiation function $exp(x) = e^x$ is applied element wise to the vector $\rho = H\mathbf{q}$. That is, for $\alpha \subseteq \{1, \ldots, n-1\}$, $s_\alpha = 2^{-(n-1)} \sum_\gamma h_{\alpha,\gamma} \left(\exp\left(\sum_\delta h_{\gamma\delta} q_\delta\right)\right)$.

This transformation is called the *Hadamard conjugation*.

Definition 2 *A vector $\hat{\mathbf{s}} \in \mathcal{R}^{2^{n-1}}$ satisfying $\sum_{\alpha \subseteq \{1,\ldots,n-1\}} \hat{\mathbf{s}}_\alpha = 1$ and $H\hat{\mathbf{s}} > \mathbf{0}$ is called* conservative.

For conservative data $\hat{\mathbf{s}}$, the Hadamard conjugation is invertible, yielding :

$$\gamma = \gamma(\hat{\mathbf{s}}) = H_{n-1}^{-1} \ln(H\hat{\mathbf{s}})$$

where the ln function is applied element-wise to the vector $H\hat{\mathbf{s}}$. We note that γ is not necessarily the edge length spectrum of any tree. On the other hand, the expected sequence spectrum of any tree T is always conservative.

3 Maximum Likelihood on Four Taxa Trees

In this section, we describe how the system of equations is set up, and how MC is used to simplify it. We begin with the formulation of the general maximum likelihood problem as a constrained optimization problem, and the resulting system of polynomial equations. Then we use the molecular clock model properties together with the Hadamard conjugation to derive a number of identities relevant to the fork. Using the derived identities, the system is substantially simplified in both cases, to the point where analytic closed form solutions can be derived.

3.1 General ML System

Given an input data ψ of n aligned, two states sequences, every column in ψ induces a split. Let $\hat{s}_\alpha$ be the number of columns in ψ inducing the split α ($\alpha \subseteq \{1, \ldots, n-1\}$). The vector $\hat{s} = [\hat{s}_\alpha]_{\alpha \subseteq \{1,\ldots,n-1\}}$, indexed analogously to the expected sequence spectrum, is called the *observed sequence spectrum* (observed spec). The likelihood of producing the observed spec $\hat{\mathbf{s}}$ given the expected spec $\mathbf{s}$ equals

$$L(\hat{\mathbf{s}}|\mathbf{s}) = \prod_{\alpha \subseteq \{1,\ldots,n-1\}} Pr(\alpha\text{-split} \mid \mathbf{s})^{\hat{s}_\alpha} = \prod_{\hat{s}_\alpha > 0} s_\alpha^{\hat{s}_\alpha} \ .$$

In the specific case of a four taxa *unrooted* tree:

$$L(\mathbf{s}|\hat{\mathbf{s}}) = s_\emptyset^{\hat{s}_\emptyset} \cdot s_1^{\hat{s}_1} \cdot s_2^{\hat{s}_2} \cdot s_{12}^{\hat{s}_{12}} \cdot s_3^{\hat{s}_3} \cdot s_{13}^{\hat{s}_{13}} \cdot s_{23}^{\hat{s}_{23}} \cdot s_{123}^{\hat{s}_{123}} \ .$$

Without loss of generality, we describe the systems for the unrooted trees corresponding to (12)(34) MC-fork. Figure 3 shows the (12)(34)-MC-fork on the left and on the right its equivalent unrooted tree. Both versions have e_{12} as their "central" edge. The expected spec $\mathbf{s}$ of these trees can be represented as a point in $\mathbb{R}^8$ whose edge lengths satisfy

- $q_\alpha(s) \geq 0$ for all $\alpha \in E(T)$.

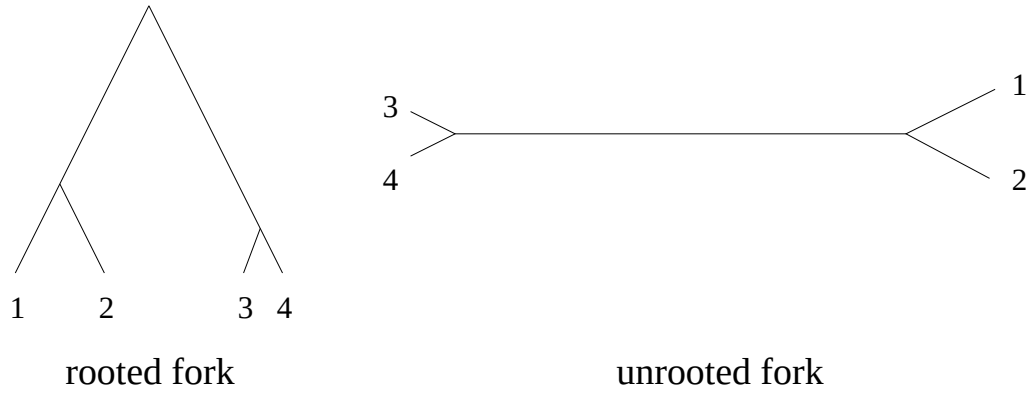

Figure 3: Rooted layout of the (12)(34)-MC-fork (left) and its unrooted version (right)

- $q_\alpha(s) = 0$ for all $\alpha \notin E(T)$.

Thus q_{13} and q_{23} must equal zero, and we can formulate the problem of maximizing the likelihood function as a constrained maximization problem: Find the maximum value of L under the constraints $q_{13}(\mathbf{s}) = 0$ and $q_{23}(\mathbf{s}) = 0$. (We can ignore the "out of boundary" requirements when maximizing the likelihood, provided we eventually verify that the resulting ML tree is indeed reasonable, namelty inside the boundary.) The approach taken in [1] is to initially express the set of critical points using Lagrange multipliers. By Proposition 1, every q_α is a function of the expected spec $\mathbf{s}$, so we seek the point or points where

$$\nabla L = \lambda_1 \nabla q_{13}(\mathbf{s}) + \lambda_2 \nabla q_{23}(\mathbf{s}) \ .$$

This gives rise to a system of ten degree 5 polynomial equations in ten variables: The eight s_α variables, and two additional"Lagrange" variables (λ_1 and λ_2). We emphasize that the

eight $\hat{s}_\alpha$ are not variables - they are parameters detemined by the four *input* sequences. (For brevity, we will use q_α and not $q_\alpha(\mathbf{s})$ in the sequel.)

The resulting system, in its full generality, is beyond the reach of current computer algebra techniques. The key to deriving analytical solutions is to combine the Hadamard conjugation with the molecular clock structure in order to derive a number of identities on the fork topology. Using the derived identities, the system is substantially simplified. For the fork, we have $q_1 = q_2$ and $q_3 = q_{123}$ (see Figure 4). We emphasise that the system of equations does not take explicitly into account *inequalities* like $q_{12} \geq 0$ or $q_{12} + min(q_{123}, q_1) > max(q_1, q_{123})$. The system is hard enough to solve as it is. Of course if the final ML point (solution) does not satisfy the relevant inequlities, it would not correspond to a "real" phylogenetic tree.

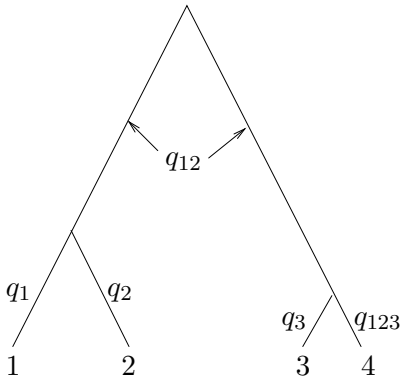

Figure 4: In the (12)(34)-MC-fork, $q_1 = q_2$ and $q_3 = q_{123}$

3.2 Simplifying Identities

The key to our simplifications is the use of lengths relations among the edges (the $\mathbf{q}$ variables), which follow from MC, in order to derive identities on the expected spec variables (the $\mathbf{s}$ variables). The following relation on the expected spec variables is proved in [2].

Theorem 1 *[2] Let i and j be sister taxa in a phylogenetic tree T with n leaves and edge weights $\mathbf{q}$. Let $\mathbf{s}$ be the expected spec, such that $\mathbf{s} = H^{-1}\ln(H\mathbf{q})$, then $q_i = q_j$ implies $s_i = s_j$ and $q_i > q_j$ implies $s_i > s_j$.*

Under a molecular clock, the four taxa MC-fork has two pairs of sister taxa, each with equal edge lengths on the tree. The next theorem is a generalization of the previous one, yielding one additional identity for the MC-fork.

Theorem 2 *Consider a tree T on n leaves, with two sister taxa i and j such that $q_i = q_j$ (see figure 5). Let $\mathbf{s}$ be the expected spec, such that $\mathbf{s} = H^{-1}\ln(H\mathbf{q})$. Then for every $\alpha \subseteq$*

$\{1, 2, ..., n-1\} \setminus \{i, j\}$, $s_{\alpha\cup\{i\}} = s_{\alpha\cup\{j\}}$.

Proof. The argument is essentially a symmetry argument, saying that i and j are interchangeable. More formally, since i and j are sister taxa in T and $q_i = q_j$, then for every other taxa k in T ($k \neq i, j$) the length of the tree paths from i to k and from j to k is the same. By the definition of the edge spec, this implies that for every $\beta \subseteq \{1, 2, ..., n-1\} \setminus \{i, j\}$, $q_{\beta\cup\{i\}} = q_{\beta\cup\{j\}}$. This means that any function of the edge spec, $\mathbf{q}$, is invariant under interchange of i and j. In particular, $s_{\alpha\cup\{i\}} = s_{\alpha\cup\{j\}}$. ■

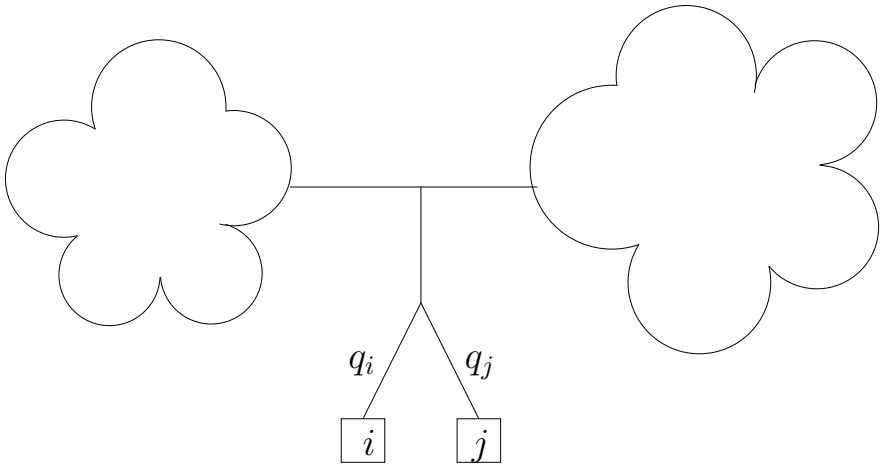

Figure 5: A general tree with two sister taxa i and j s.t. $q_i = q_j$

Claim 1 *Let T be a $(1,2)(3,4)$-MC-fork. Then $s_3 = s_{13}(1 - 2s_1 - 2s_{13})/(2s_1 + 2s_{13})$*

Proof. Substituting $s_1 = s_2$, $s_{123} = s_3$ and $s_{13} = s_{23}$ in Proposition 1, we get

$$q_{13} = -\frac{1}{8}\left(\ln(1 - 4s_1 - 4s_{13}) + \ln(1 - 4s_3 - 4s_{13}) - \ln(1 - 4s_1 - 4s_3)\right) .$$

By equating q_{13} to zero, taking exponent, and multiplying through by the denominator , we get $(1 - 4s_1 - 4s_3) = (-1 + 4s_1 + 4s_{13})(-1 + 4s_3 + 4s_{13})$. Arithmetic manipulation yields the claimed equality. ■

The next technical claim deals with conservative points $\mathbf{s} \in \mathbb{R}^8$ (namely $H\mathbf{s} > 0$) satisfying $s_1 = s_2$, $s_{13} = s_{23}$, and $\sum_{\alpha\subseteq\{1,2,3\}} s_\alpha = 1$. (These points need not be the expected spec of a tree.) This technical claim will be useful in simplifying the system of polynomial equations that we solve in Section 4.

Claim 2 *Let $\mathbf{s} = (s_\emptyset, s_1, s_2, s_{12}, s_3, s_{13}, s_{23}, s_{123}) \in \mathbb{R}^8$ be a conservative point satisfying $s_1 = s_2$ and $s_{13} = s_{23}$, and let $\mathbf{q} = H^{-1} \ln H\mathbf{s}$. Then $\mathbf{s}$ satisfies $q_{13}(\mathbf{s}) = q_{23}(\mathbf{s})$.*

Proof. By the Hadamard conjugation we get:

$$\begin{aligned} 4(q_{13}-q_{23}) &= \ln(1-2\,s_{1}-2\,s_{12}-2\,s_{123}-2\,s_{13})-\ln(1-2\,s_{2}-2\,s_{12}-2\,s_{23}-2\,s_{123})- \\ &\quad \ln(1-2\,s_{1}-2\,s_{12}-2\,s_{3}-2\,s_{23})+\ln(1-2\,s_{2}-2\,s_{12}-2\,s_{3}-2\,s_{13}) \end{aligned}$$

By the assumption $s_1 = s_2$ and $s_{13} = s_{23}$, so $(1-2\,s_{1}-2\,s_{12}-2\,s_{123}-2\,s_{13})$ (the first term) equals $(1-2\,s_{2}-2\,s_{12}-2\,s_{23}-2\,s_{123})$ (the second term) and $(1-2\,s_{1}-2\,s_{12}-2\,s_{3}-2\,s_{23})$ (third term) equals $(1-2\,s_{2}-2\,s_{12}-2\,s_{3}-2\,s_{13})$ (fourth term). This implies $q_{13}(\mathbf{s}) - q_{23}(\mathbf{s}) = 0$. ■

4 Solving the MC-fork

In this section we develop the analytic solutions for the ML MC-fork.

Theorem 3 *Let* $\hat{\mathbf{s}} = (\hat{s}_{\emptyset}, \hat{s}_1, \hat{s}_2, \hat{s}_{12}, \hat{s}_3, \hat{s}_{13}, \hat{s}_{23}, \hat{s}_{123})$ *be the observed spec, and let* $\sum_\alpha \hat{s}_\alpha = c$. *Then the expected spec of the ML* (12)(34)*-MC-fork equals:*

$$\begin{aligned} s_{13}=s_{23} &= (\hat{s}_{23}^2+\hat{s}_{1}\,\hat{s}_{23}+\hat{s}_{2}\,\hat{s}_{23}+\hat{s}_{123}\,\hat{s}_{23}+2\,\hat{s}_{13}\,\hat{s}_{23}+\hat{s}_{3}\,\hat{s}_{23}+\hat{s}_{13}\,\hat{s}_{3}+ \\ &\quad \hat{s}_{13}^2+\hat{s}_{13}\,\hat{s}_{123}+\hat{s}_{2}\,\hat{s}_{3}+\hat{s}_{2}\,\hat{s}_{123}+\hat{s}_{2}\,\hat{s}_{13}+\hat{s}_{1}\,\hat{s}_{3}+\hat{s}_{1}\,\hat{s}_{123}+\hat{s}_{1}\,\hat{s}_{13})/2c^2, \\ s_{12} &= \frac{(c-\hat{s}_{2}-\hat{s}_{1}-\hat{s}_{23}-\hat{s}_{13})\,(c-\hat{s}_{3}-\hat{s}_{23}-\hat{s}_{123}-\hat{s}_{13})\,\hat{s}_{12}}{c^2\,(c-\hat{s}_{1}-\hat{s}_{2}-\hat{s}_{3}-\hat{s}_{123}-\hat{s}_{13}-\hat{s}_{23})}, \\ s_{1}=s_{2} &= -(\hat{s}_{23}^2-c\,\hat{s}_{23}+\hat{s}_{2}\,\hat{s}_{23}+\hat{s}_{123}\,\hat{s}_{23}+\hat{s}_{3}\,\hat{s}_{23}+\hat{s}_{1}\,\hat{s}_{23}+2\,\hat{s}_{13}\,\hat{s}_{23}-c\,\hat{s}_{13}+\hat{s}_{13}\,\hat{s}_{3}+ \\ &\quad \hat{s}_{13}^2+\hat{s}_{13}\,\hat{s}_{123}+\hat{s}_{2}\,\hat{s}_{3}+\hat{s}_{2}\,\hat{s}_{123}-\hat{s}_{2}\,c+\hat{s}_{2}\,\hat{s}_{13}+\hat{s}_{1}\,\hat{s}_{3}+\hat{s}_{1}\,\hat{s}_{123}-\hat{s}_{1}\,c+\hat{s}_{1}\,\hat{s}_{13})/2c^2 \\ s_{3}=s_{123} &= -(\hat{s}_{13}\,\hat{s}_{1}+\hat{s}_{23}\,\hat{s}_{1}+\hat{s}_{3}\,\hat{s}_{1}+\hat{s}_{123}\,\hat{s}_{1}-\hat{s}_{23}\,c-\hat{s}_{123}\,c-\hat{s}_{3}\,c-\hat{s}_{13}\,c+ \\ &\quad \hat{s}_{123}\,\hat{s}_{23}+\hat{s}_{123}\,\hat{s}_{13}+\hat{s}_{3}\,\hat{s}_{23}+\hat{s}_{3}\,\hat{s}_{13}+\hat{s}_{123}\,\hat{s}_{2}+\hat{s}_{23}\,\hat{s}_{2}+\hat{s}_{23}^2+\hat{s}_{13}\,\hat{s}_{2}+ \\ &\quad 2\,\hat{s}_{13}\,\hat{s}_{23}+\hat{s}_{13}^2+\hat{s}_{3}\,\hat{s}_{2})/2c^2 \end{aligned}$$

Proof. The (12)(34)-MC-fork satisfies $q_1 = q_2$ and $q_3 = q_{123}$, therefore by Theorem 1, $s_1 = s_2$ and $s_3 = s_{123}$. By Theorem 2, $q_1 = q_2$ implies $s_{13} = s_{23}$. Substituting $s_2 = s_1$, $s_{23} = s_{13}$ and $s_{123} = s_3$, our likelihood function becomes:

$$L(\mathbf{s}|\hat{\mathbf{s}}) = s_{\emptyset}^{\hat{s}_{\emptyset}} \cdot s_1^{\hat{s}_1+\hat{s}_2} \cdot s_{12}^{\hat{s}_{12}} \cdot s_3^{\hat{s}_3+\hat{s}_{123}} \cdot s_{13}^{\hat{s}_{13}+\hat{s}_{23}}.$$

By Claim 1, $q_{13} = 0$ implies $s_3 = s_{13}(1-2s_1-2s_{13})/(2s_1+2s_{13})$, so after substituting the resulting value for s_3, q_{13} becomes identically zero. Now since $s_1 = s_2$ and $s_{13} = s_{23}$ and

by definition $\mathbf{s}$ is conservative, so $\mathbf{s}$ satisfies the conditions of Claim 2, and thus $q_{13} = q_{23}$, so $q_{23} = 0$ as well. This means that after making this series of substitutions the two constraints are satisfied, so in order to look for critical points of L we should just equate the partial derivatives of L to 0.

The final step is to use the fact that the s_α variables add up to 1 and substitute $s_\emptyset = 1 - \sum_{\alpha \subseteq \{1,2,3\} \setminus \emptyset} s_\alpha$. The likelihood function becomes

$$L(\mathbf{s}|\hat{\mathbf{s}}) = \left(\frac{s_1 - 2\,{s_1}^2 - 2\,s_1\,s_{13} - s_{12}\,s_1 - s_{12}\,s_{13}}{s_1 + s_{13}} \right)^{(c - \hat{s}_1 - \hat{s}_2 - \hat{s}_3 - \hat{s}_{123} - \hat{s}_{12} - \hat{s}_{13} - \hat{s}_{23})}$$
$${s_{12}}^{\hat{s}_{12}}\,{s_{13}}^{(\hat{s}_{13} + \hat{s}_{23})}\,{s_1}^{(\hat{s}_1 + \hat{s}_2)} \left(\frac{s_{13}\,(1 - 2\,s_1 - 2\,s_{13})}{2(s_1 + s_{13})} \right)^{(\hat{s}_3 + \hat{s}_{123})}$$

where $c = \sum_{\alpha \subseteq \{1,\ldots,3\}} \hat{s}_\alpha$.
This is a polynomial in three free variables s_1, s_{12}, s_{13} and eight (given) parameters. As we argued before, each ML point will be a critical point if $\nabla L = \mathbf{0}$, namely will satisfy the three polynomial equations:

$$\frac{\partial L}{\partial s_1} = 0\ ,\ \frac{\partial L}{\partial s_{12}} = 0\ ,\ \frac{\partial L}{\partial s_{13}} = 0\ .$$

The resulting system of polynomial equations is written down in the appendix. It is too involved to solve manually, so we applied Maple (version 6) to solve it. The solutions for the three variables s_1, s_{12}, s_{13} appear in the statement of the theorem. The other variables are obtained by back substitutions. ■

The question of uniqueness of the ML point for phylogenetic analysis has raised considerable interest in the past [4, 12, 10, 1] . It is now known that even four taxa ML trees exhibit datasets giving rise to multiple ML points [10, 1]. In contrast, our result implies uniqueness for the ML MC-fork.

Corollary 4 *Each MC-fork topology has a unique local and global maximum likelihood point.*

We remark that in certain, rare cases, the ML solution (in the sequence spectrum representation) translates to a solution in (the edge space $\mathbf{q}$) where one entry is negative. The resulting tree is unrealistic. When this happens, the maxima of ML for "realistic forks" will be attained on the boundaries of the "realistic fork manifold" (one or more edge length equal 0).

While our results were stated for the ML (12)(34)-MC-fork, they clearly apply to the other two MC-forks as well.

5 Additional Results

In the previous section we showed how to find, for a given observed data and a particular fork, the tree (in the sequence spectrum representation) that attain the maximum likelihood. Naturally, the next step is to determine, given some observed data, which of the three forks attains the maximum likelihood. The straightforward way for this is, for each of the three forks:

- Find the edge lengths by the formulae in Theorem 3.
- Substitute and obtain the likelihood of each fork.
- Choose the fork that achieves the maximum likelihood.

It would be desirable to bypass this process and directly infer the ML MC-fork topology among the three possibilities. We remark that for three taxa under MC, this can be done directly [13, 2]. This task in general is substantially more demanding computationally, and seems infeasible using currently available tools. Nevertheless, as was mentioned before, our goal in this work is not to replace numerical methods for ML, but rather to point out special cases and establish rigorously intuitive behaviors. Indeed, for certain restricted families of inputs, we succeeded in finding a closed form solution for this problem. We begin with a definition:

Definition 3 *An observed sequence spectrum* $\mathbf{h}$ *over the set of species* X *is said to be reasonable if* $\hat{s}_0 > \sum_{\alpha \in X-0} \hat{s}_\alpha$. *In other words, the data is reasonable if the number of constant sites is bigger than the total number of variable sites. For homologous genes and n=4 taxa, this requirement is realistic.*

The assumption of a reasonable data guarantees that the data is conservative and that a reasonable data $\hat{\mathbf{s}}$ such that $\sum_\alpha \hat{s}_\alpha = c$ implies $c \geq 2 \sum_{\alpha \in X \setminus \emptyset} \hat{s}_\alpha$.

5.1 The Singletons Case

Here we deal with a family of inputs where all non singletons entries $(\hat{s}_{12}, \hat{s}_{13}, \hat{s}_{23})$ are equal, and only the singletons entries $\hat{s}_1, \hat{s}_2, \hat{s}_3, \hat{s}_{123}$ may vary.

5.1.1 Pairwise Equal Singletons

We start with a specific case of the singletons family that will be later extended to all types of inputs in the family. In this case, all pairwise splits are equal. The singletons are grouped

in two sets of size two $\hat{s}_2 = \hat{s}_{123} = b$ and $\hat{s}_1 = \hat{s}_3 = b+\epsilon$. As can be seen, this is a special case of the general singletons case. Intuitively, the best ML MC fork should be (13)(24). The following claim shows that this is always the case.

Claim 3 *Let* $\hat{\mathbf{s}} = (\hat{s}_0, \hat{s}_1, \hat{s}_2, \hat{s}_{12}, \hat{s}_3, \hat{s}_{13}, \hat{s}_{23}, \hat{s}_{123})$ *such that* $\sum_\alpha \hat{s}_\alpha = c$*, be a reasonable observed sequence spectrum. Let all non singleton entries in* $\hat{\mathbf{s}}$ *have the same value* $\hat{s}_{12} = \hat{s}_{13} = \hat{s}_{23} = a$*. Now partition the singleton entries into two pairs, say* $\{1,3\}$ *and* $\{2,4\}$*. Suppose* $\hat{s}_2 = \hat{s}_{123} = b$ *and* $\hat{s}_1 = \hat{s}_3 = b + \varepsilon$*. Then the ML MC-fork is the* (13)(24)*-MC-fork.*

Proof. By definition, the observed sequence spectrum $\hat{\mathbf{s}}$ is: $(\hat{s}_0, \hat{s}_1 = b+\varepsilon, \hat{s}_2 = b, \hat{s}_{12} = a, \hat{s}_3 = b+\varepsilon, \hat{s}_{13} = a, \hat{s}_{23} = a, \hat{s}_{123} = b)$.
Let L_{12} be the likelihood of the ML (12)(34)-MC-fork and let L_{13} and L_{23} be defined analogously. Now let $\mathbf{s}^{(12)}$ be the expected sequence spectrum attaining L_{12} (i.e. maximizing the likelihood of the (12)(34)-MC-fork) and let $\mathbf{s}^{(13)}$ and $\mathbf{s}^{(23)}$ be defined analogously. We will show that for every $\varepsilon \neq 0$, if the observed input is reasonable $r = \frac{L_{13}}{L_{12}} > 1$. Since $L_{12} = \prod_{\alpha\subseteq\{1,2,3\}} s_\alpha^{(12)^{\hat{s}_\alpha}}$, we can write

$$r = \frac{\prod_{\alpha\subseteq\{1,2,3\}} s_\alpha^{(13)^{\hat{s}_\alpha}}}{\prod_{\alpha\subseteq\{1,2,3\}} s_\alpha^{(12)^{\hat{s}_\alpha}}} = \prod_{\alpha\subseteq\{1,2,3\}} \left(\frac{s_\alpha^{(13)}}{s_\alpha^{(12)}}\right)^{\hat{s}_\alpha}.$$

Substituting in every $s_\alpha^{(ij)}$ the values we obtained in Theorem 3 we get:

$$\begin{aligned} r = \quad & \left(\frac{(c-2\,a-2\,b)\,(c-2\,b-2\,\varepsilon-2\,a)}{(\mathrm{c}-2\,b-\varepsilon-2\,a)^2}\right)^{(c-2\,\varepsilon-4\,b-3\,a)} \left(2\,\frac{(c-2\,a-2\,b)\,(a+b+\varepsilon)}{(\mathrm{c}-2\,b-\varepsilon-2\,a)\,(2\,a+2\,b+\varepsilon)}\right)^{2\,(b+\varepsilon)} \\ & \left(2\,\frac{(a+b)\,(2\,b+2\,\varepsilon+2\,a-c)}{(2\,b+\varepsilon-\mathrm{c}+2\,a)\,(2\,a+2\,b+\varepsilon)}\right)^{2\,b} \left(2\,\frac{(a+b+\varepsilon)\,(a+b)\,(c-4\,b-2\,\varepsilon-2\,a)}{(\mathrm{c}-2\,b-\varepsilon-2\,a)^2\,a}\right)^{a}. \\ & \left(2\,\frac{(-2\,a-2\,b+c)\,a\,(2\,b+2\,\varepsilon+2\,a-c)}{(2\,a+2\,b+\varepsilon)^2\,(4\,b+2\,\varepsilon+2\,a-c)}\right)^{a} \left(4\,\frac{(a+b+\varepsilon)\,(a+b)}{(2\,a+2\,b+\varepsilon)^2}\right)^{a} \end{aligned}$$

It can easily be seen that $\varepsilon = 0$ implies $r = 1$ and indeed, such symmetric input has no preference among the three forks. The derivative $\frac{\partial r}{\partial \varepsilon}$ admits two critical points:

1. $\varepsilon_1 = 0$
2. $\varepsilon_2 = 1/2c - 2a - b$

For reasonable data, The second derivative $\frac{\partial^2 r}{\partial \varepsilon^2}$ at the point ε_1 is always positive, implying that at $\varepsilon_1 = 0$ r is at local minimum ($r = 1$). The second critical point coerce data outside the positive quadrant of $\mathbb{R}^8$ and therefore is illegal.

It follows that for our family of inputs $\hat{s}_\varepsilon$, there is a single local minimum of $r(\varepsilon)$ at the point $\varepsilon_1 = 0$ and hence this is also a global minimum. Since $r(\varepsilon_1) = 1$ it follows that for all $\varepsilon \neq 0$, $r(\varepsilon) > 1$. It is clear that the same applies for $\varepsilon < 0$ as well. ■

We now extend the previous result to the general case of the singletons where all singletons can differ from one another. This case is more complicated since instead of a single free variable ε we have here three so the number of free variables increases substantially.

Claim 4 *Let* $\hat{\mathbf{s}} = (\hat{s}_0, \hat{s}_1, \hat{s}_2, \hat{s}_{12}, \hat{s}_3, \hat{s}_{13}, \hat{s}_{23}, \hat{s}_{123})$ *such that* $\sum_\alpha \hat{s}_\alpha = c$, *be a reasonable observed sequence spectrum. Let all non singleton entries in* $\hat{\mathbf{s}}$ *have the same value* $\hat{s}_{12} = \hat{s}_{13} = \hat{s}_{23} = a$. *Without loss of generality, let the singleton entries be as follows:*

- $\hat{s}_2 = \hat{s}_1 + \varepsilon_1$
- $\hat{s}_3 = \hat{s}_2 + \varepsilon_2 = \hat{s}_1 + \varepsilon_2 + \varepsilon_1$
- $\hat{s}_{123} = \hat{s}_3 + \varepsilon_3 = \hat{s}_2 + \varepsilon_3 + \varepsilon_2 = \hat{s}_1 + \varepsilon_3 + \varepsilon_2 + \varepsilon_1$

where ε_1, ε_2, $\varepsilon_3 \geq 0$. *Then the ML MC-fork is the* (12)(34)*-MC-fork.*

Proof. We use the same notation as in Claim 3 with

$$r = \frac{L_{13}}{L_{12}} = \frac{\prod_{\alpha\subseteq\{1,2,3\}} {s_\alpha^{(13)}}^{\hat{s}_\alpha}}{\prod_{\alpha\subseteq\{1,2,3\}} {s_\alpha^{(12)}}^{\hat{s}_\alpha}} = \prod_{\alpha\subseteq\{1,2,3\}} \left(\frac{s_\alpha^{(13)}}{s_\alpha^{(12)}}\right)^{\hat{s}_\alpha}.$$

But substituting every $s_\alpha^{(ij)}$ with the values we obtained in Theorem 3 we get:

$$
\begin{aligned}
r = \quad & \left(\frac{(-\mathrm{c}+2\,\hat{s}_1+\varepsilon_1+2\,\hat{s}_{12})\,(2\,\hat{s}_1+2\,\varepsilon_1+2\,\varepsilon_2+\varepsilon_3+2\,\hat{s}_{12}-\mathrm{c})}{(2\,\hat{s}_1+\varepsilon_1+\varepsilon_2-\mathrm{c}+2\,\hat{s}_{12})\,(2\,\hat{s}_{12}+2\,\hat{s}_1+2\,\varepsilon)}\right)^{(c-4\,\hat{s}_1-3\,\varepsilon_1-2\,\varepsilon_2-\varepsilon_3-3\,\hat{s}_{12})} \\
& \left(\frac{(2\,\hat{s}_1+2\,\varepsilon_1+2\,\varepsilon_2+\varepsilon_3+2\,\hat{s}_{12}-\mathrm{c})\,(2\,\hat{s}_{12}+2\,\hat{s}_1+\varepsilon_1)}{(2\,\hat{s}_{12}+2\,\hat{s}_1+2\,\varepsilon)\,(2\,\hat{s}_{12}+2\,\hat{s}_1+\varepsilon_1+\varepsilon_2)}\right)^{\hat{s}_1} \\
& \left(\frac{(2\,\hat{s}_1+2\,\varepsilon_1+2\,\varepsilon_2+\varepsilon_3+2\,\hat{s}_{12}-\mathrm{c})\,(2\,\hat{s}_{12}+2\,\hat{s}_1+\varepsilon_1)}{(2\,\hat{s}_1+\varepsilon_1+\varepsilon_2-\mathrm{c}+2\,\hat{s}_{12})\,(2\,\hat{s}_{12}+2\,\hat{s}_1+2\,\varepsilon 1+\varepsilon_2+\varepsilon_3)}\right)^{(\hat{s}_1+\varepsilon_1)} \\
& \left(-2\,\frac{(2\,\hat{s}_1+2\,\varepsilon_1+2\,\varepsilon_2+\varepsilon_3+2\,\hat{s}_{12}-\mathrm{c})\,\hat{s}_{12}\,(-\mathrm{c}+2\,\hat{s}_1+\varepsilon_1+2\,\hat{s}_{12})}{(4\,\hat{s}_1+3\,\varepsilon_1+2\,\hat{s}_{12}+2\,\varepsilon_2+\varepsilon_3-c)\,(2\,\hat{s}_{12}+2\,\hat{s}_1+\varepsilon_1+\varepsilon_2)\,(2\,\hat{s}_{12}+2\,\hat{s}_1+2\,\varepsilon 1+\varepsilon_2+\varepsilon_3)}\right)^{\hat{s}_{12}} \\
& \left(\frac{(-\mathrm{c}+2\,\hat{s}_1+\varepsilon_1+2\,\hat{s}_{12})\,(2\,\hat{s}_{12}+2\,\hat{s}_1+2\,\varepsilon 1+2\,\varepsilon_2+\varepsilon_3)}{(2\,\hat{s}_{12}+2\,\hat{s}_1+2\,\varepsilon)\,(2\,\hat{s}_{12}+2\,\hat{s}_1+\varepsilon_1+\varepsilon_2)}\right)^{(\hat{s}_1+\varepsilon 1+\varepsilon_2)} \\
& \left(-\frac{1}{2}\,\frac{(2\,\hat{s}_{12}+2\,\hat{s}_1+\varepsilon_1)\,(2\,\hat{s}_{12}+2\,\hat{s}_1+2\,\varepsilon 1+2\,\varepsilon_2+\varepsilon_3)\,(4\,\hat{s}_1+3\,\varepsilon_1+2\,\hat{s}_{12}+2\,\varepsilon_2+\varepsilon_3-c)}{(2\,\hat{s}_{12}+2\,\hat{s}_1+2\,\varepsilon)\,\hat{s}_{12}\,(2\,\hat{s}_1+\varepsilon_1+\varepsilon_2-\mathrm{c}+2\,\hat{s}_{12})}\right)^{\hat{s}_{12}} \\
& \left(\frac{(2\,\hat{s}_{12}+2\,\hat{s}_1+\varepsilon_1)\,(2\,\hat{s}_{12}+2\,\hat{s}_1+2\,\varepsilon 1+2\,\varepsilon_2+\varepsilon_3)}{(2\,\hat{s}_{12}+2\,\hat{s}_1+\varepsilon_1+\varepsilon_2)\,(2\,\hat{s}_{12}+2\,\hat{s}_1+2\,\varepsilon 1+\varepsilon_2+\varepsilon_3)}\right)^{\hat{s}_{12}} \\
& \left(\frac{(-\mathrm{c}+2\,\hat{s}_1+\varepsilon_1+2\,\hat{s}_{12})\,(2\,\hat{s}_{12}+2\,\hat{s}_1+2\,\varepsilon 1+2\,\varepsilon_2+\varepsilon_3)}{(2\,\hat{s}_1+\varepsilon_1+\varepsilon_2-\mathrm{c}+2\,\hat{s}_{12})\,(2\,\hat{s}_{12}+2\,\hat{s}_1+2\,\varepsilon 1+\varepsilon_2+\varepsilon_3)}\right)^{(\hat{s}_1+\varepsilon_1+\varepsilon_2+\varepsilon_3)}
\end{aligned}
$$

.

Again, for $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 0$ the input is symmetric and we get $r = 1$.

Observation 1 *Let T be some ML tree for some fixed topology L on some input data $\hat{s}$. Let $\hat{s}'$ be obtained from $\hat{s}$ as follows: swap between the values of $\hat{s}_{i\cup\alpha}$ and $\hat{s}_{j\cup\alpha}$ for all α and retain the other entries of $\hat{s}$ as they are. Let L' be the same topology as L with swapping between taxa i and j. Then the ML tree for topology L' on input $\hat{s}'$ will have the same branches lengths (and naturally the same likelihood).*

Observation 2 *For $\varepsilon_2 = 0$ and for any ε_1, $\varepsilon_3 \Rightarrow r = 1$.*

Proof. $\varepsilon_2 = 0$ means $s_2 = s_3$ and by assumption $s_{12} = s_{13}$. Actually, we are performing the swaps indicated in Observation 1, but remain with the same input and therefore for that input the ML (13)(24))-MC-fork and the ML (12)(34))-MC-fork attain the same likelihood. ■

It remains to show that an increase in ε_2 will increase r. The derivative $\frac{\partial r}{\partial \varepsilon_2}$ has several critical points:

1. $\{\varepsilon_1 = -2\,\hat{s}_{12} - 2\,\hat{s}_1,\, c = c,\, \hat{s}_1 = \hat{s}_1,\, \hat{s}_{12} = \hat{s}_{12},\, \varepsilon_3 = \varepsilon_3,\, \varepsilon_2 = \varepsilon_2\}$
2. $\{\hat{s}_1 = \hat{s}_1,\, \hat{s}_{12} = \hat{s}_{12},\, \varepsilon_1 = \varepsilon_1,\, \varepsilon_3 = \varepsilon_3,\, \varepsilon_2 = \varepsilon_2,\, c = 0\}$
3. $\{\varepsilon_2 = 0,\, c = c,\, \hat{s}_1 = \hat{s}_1,\, \hat{s}_{12} = \hat{s}_{12},\, \varepsilon_1 = \varepsilon_1,\, \varepsilon_3 = -\varepsilon_1\}$
4. $\{\hat{s}_1 = \hat{s}_1,\, \hat{s}_{12} = \hat{s}_{12}, \varepsilon_1 = \varepsilon_1,\, \varepsilon_3 = \varepsilon_3,\, \varepsilon_2 = \varepsilon_2,$
$$c = (4\,\varepsilon_2^3 + \varepsilon_3^3 + 6\,\varepsilon_2^2\,\varepsilon_3 + 4\,\varepsilon_2\,\varepsilon_3^2 + 4\,\varepsilon_3^2\,\hat{s}_1 + 5\,\varepsilon_3^2\,\varepsilon_1$$
$$+ 12\,\hat{s}_1\,\varepsilon_1\,\varepsilon_3 + 4\,\varepsilon_3^2\,\hat{s}_{12} + 8\,\varepsilon_1^2\,\varepsilon_3 + 8\,\hat{s}_1\,\hat{s}_{12}\,\varepsilon_1 + 8\,\hat{s}_1\,\hat{s}_{12}\,\varepsilon_3 + 4\,\hat{s}_{12}^2\,\varepsilon_3$$
$$+ 12\,\varepsilon_1\,\hat{s}_{12}\,\varepsilon_3 + 4\,\varepsilon_1\,\hat{s}_{12}^2 + 8\,\varepsilon_1^2\,\hat{s}_{12} + 8\,\hat{s}_1\,\varepsilon_1^2 + 4\,\hat{s}_1^2\,\varepsilon_1 + 4\,\hat{s}_1^2\,\varepsilon_3 + 4\,\varepsilon_1^3$$
$$+ 12\,\hat{s}_{12}\,\varepsilon_2\,\varepsilon_3 + 14\,\varepsilon_1\,\varepsilon_2\,\varepsilon_3 + 20\,\hat{s}_1\,\varepsilon_1\,\varepsilon_2 + 12\,\hat{s}_{12}\,\varepsilon_2^2 + 12\,\varepsilon_1\,\varepsilon_2^2 + 12\,\varepsilon_1^2\,\varepsilon_2$$
$$+ 12\,\hat{s}_1\,\varepsilon_2\,\varepsilon_3 + 12\,\hat{s}_1\,\varepsilon_2^2 + 8\,\hat{s}_1^2\,\varepsilon 2 + 8\,\hat{s}_{12}^2\,\varepsilon_2 + 16\,\hat{s}_1\,\hat{s}_{12}\,\varepsilon_2 + 20\,\varepsilon_1\,\hat{s}_{12}\,\varepsilon_2)\Big/($$
$$2\,\hat{s}_1\,\varepsilon_1 + 2\,\varepsilon_1^2 + 5\,\varepsilon_1\,\varepsilon_2 + 3\,\varepsilon_2^2 + 3\,\varepsilon_3\,\varepsilon_1 + 2\,\varepsilon_3\,\hat{s}_1 + 4\,\hat{s}_1\,\varepsilon_2 + 2\,\varepsilon_3\,\hat{s}_{12} + 2\,\varepsilon_1\,\hat{s}_{12}$$
$$+ \varepsilon_3^2 + 3\,\varepsilon_2\,\varepsilon_3 + 4\,\hat{s}_{12}\,\varepsilon_2)\}.$$

The first three points correspond to invalid data (either negative values or $c = 0$). It remains to check the fourth point. Under reasonable data assumption the input satisfies $c \geq 2\sum_{\alpha\neq\emptyset}\hat{s}_\alpha$ and in the the current context, the above implies $c \geq 2(4\hat{s}_1+\varepsilon_1+\varepsilon_2+\varepsilon_3)+6\hat{s}_{12}$. It is easy to see that the solution does not satisfy this assumption. We can conclude that the derivative does not change its sign in the valid region.

It is only left to see that the derivative is indeed positive. Since at $\varepsilon_2 = 0$, $r = 1$, it's enough to show that for $\varepsilon_2 > 0$ **and** $\varepsilon_1, \varepsilon_3 \geq 0$, $r > 1$ and indeed this is the case.

The proof that $L_{12} \geq L_{23}$ is essentially the same so we omit the equations and describe the procedure. Similarly to Observation 2, it's easy to see that for any ε_1 and ε_2, $\varepsilon_3 = 0$, $L_{12} = L_{23}$. Next, we show that $\left.\frac{\partial \frac{L_{12}}{L_{23}}}{\partial \varepsilon_2}\right|_{\varepsilon_3=0} \geq 0$, and eventually we show that $\frac{\partial \frac{L_{12}}{L_{23}}}{\partial \varepsilon_3} \geq 0$. That completes the proof. ■

The following corollary is a consequence of the proof itself.

Corollary 5 *For any reasonable data where all non-singletons entries are equal, the ML MC-fork is the $(ij)(k\ell)$-MC-fork where the singletons entries satisfy $\hat{s}_i \geq \hat{s}_j \geq \hat{s}_k \geq \hat{s}_\ell$.*

5.2 The Non Singletons Case

In this subsection we analyze the case where all singletons are equal and only the non singletons differ. In particular, we consider the case where two of the three non singletons,

say $\hat{s}_{12}$ and $\hat{s}_{23}$ are equal and the other, $\hat{s}_{13}$, is bigger by some non negative ε. We would expect that, similar to the singletons case, (13)(24)-MC-fork will always be more likely than the (12)(34)-MC-fork, as the similarity between taxa 1 and 3 is higher than the similarity between any other pair. However, this case is more subtle than the singletons case since there are observed data that are reasonable but give rise to ML points with *negative* edge length. Under the condition that the edge lengths are positive, the likelihood of the (13)(24)-MC-fork is strictly higher than the likelihood of the (12)(34)-MC-fork.

Observation 3 *Suppose*

$$c > \left(6\,\hat{s}_{12} + \varepsilon + 2\,\hat{s}_1 + \sqrt{4\,\hat{s}_{12}^2 + 4\,\hat{s}_{12}\,\varepsilon - 8\,\hat{s}_1\,\hat{s}_{12} + \varepsilon^2 + 4\,\hat{s}_1\,\varepsilon + 4\,\hat{s}_1^2}\right)(2\,\hat{s}_{12} + \varepsilon + 2\,\hat{s}_1)/4\hat{s}_{12}$$

then all edges in the ML (12)(34)*-MC-fork and the ML* (13)(24)*-MC-fork are positive.*

Claim 5 *Let* $\hat{\mathbf{s}} = (\hat{s}_0, \hat{s}_1, \hat{s}_2, \hat{s}_{12}, \hat{s}_3, \hat{s}_{13}, \hat{s}_{23}, \hat{s}_{123})$ *such that* $\sum_\alpha \hat{s}_\alpha = c$*, be a reasonable observed sequence spectrum. Suppose all singleton entries in* $\hat{\mathbf{s}}$ *have the same value* $\hat{s}_2 = \hat{s}_{123} = \hat{s}_1 = \hat{s}_3$ *and let two non singleton entries be equal, say* $\hat{s}_{12} = \hat{s}_{23}$*, and the other* $\hat{s}_{13} = \hat{s}_{12} + \varepsilon_0$*. Then, under the condition that the* (12)(34)*-MC-fork has a non negative internal edge (i.e.* $q_{12} > 0$*), the ML MC-fork is the* (13)(24)*-MC-fork.*

Proof. We set the problem as before:

$$r = \frac{\prod_{\alpha\subseteq\{1,2,3\}} {s_\alpha^{(13)}}^{\hat{s}_\alpha}}{\prod_{\alpha\subseteq\{1,2,3\}} {s_\alpha^{(12)}}^{\hat{s}_\alpha}} = \prod_{\alpha\subseteq\{1,2,3\}} \left(\frac{s_\alpha^{(13)}}{s_\alpha^{(12)}}\right)^{\hat{s}_\alpha}.$$

and seek to prove that $r \geq 1$.

In the sequel, since q_{13} in every (12)(34)-MC-fork is identically zero (and in particular in the ML (12)(34)-MC-fork), when we write q_{13} we refer to q_{13} in the ML (13)(24)-MC-fork (and similarly with q_{12} refers to the ML (12)(34)-MC-fork).

We first show that for reasonable data, the fact that q_{12} is positive implies that c is big enough. The equation for q_{12} is:

$$q_{12} = \ln\left(\sqrt{\frac{c\,(c - 4\,\hat{s}_{12} - 2\,\varepsilon - 4\,\hat{s}_1)\,(c - 4\,\hat{s}_1 - 2\,\hat{s}_{12} - \varepsilon)}{(c - 4\,\hat{s}_{12} - 4\,\hat{s}_1 - \varepsilon)\,(c - 2\,\hat{s}_{12} - \varepsilon - 2\,\hat{s}_1)^2}}\right).$$

This expression has two roots (as a function of c):

1. $$c = \frac{(6\,\hat{s}_{12} + \varepsilon + 2\,\hat{s}_1 + \sqrt{4\,\hat{s}_{12}^2 + 4\,\hat{s}_{12}\,\varepsilon - 8\,\hat{s}_1\,\hat{s}_{12} + \varepsilon^2 + 4\,\hat{s}_1\,\varepsilon + 4\,\hat{s}_1^2})(2\,\hat{s}_{12} + \varepsilon + 2\,\hat{s}_1)}{4\hat{s}_{12}}$$

2.

$$c = \frac{(6\,\hat{s}_{12} + \varepsilon + 2\,\hat{s}_1 - \sqrt{4\,\hat{s}_{12}^2 + 4\,\hat{s}_{12}\,\varepsilon - 8\,\hat{s}_1\,\hat{s}_{12} + \varepsilon^2 + 4\,\hat{s}_1\,\varepsilon + 4\,\hat{s}_1^2})(2\,\hat{s}_{12} + \varepsilon + 2\,\hat{s}_1)}{4\hat{s}_{12}}$$

The second is outside the reasonable region so we are interested only in the first one. This is a function on the input variables $\hat{s}_1$, $\hat{s}_{12}$ and ε. Let c' be the value of this function on the input variables $\hat{s}_1$, $\hat{s}_{12}$ and ε.

Observation 4 *If $c > c'$ then $q_{12} > 0$.*

Proof. This can be easily observed since the fraction inside the root of q_{12} goes to 1 from above as c goes to infinity. ■

We also note that the biological reasoning for that is that as the relative number of invariable sites grows, the difference between the likelihood of different topologies decreases.

Corollary 6 *if $q_{12}(\hat{s}) > 0$ then $c > c'$ where c' is as defined above.*

The next argument we show, is that for all $c > c'$, $\frac{\partial r}{\partial \varepsilon} > 0$. Let $\alpha = \frac{\partial r}{\partial \varepsilon}$. Then α has two roots:

1.

$$c = \frac{(6\,\hat{s}_{12} + 2\,\hat{s}_1 + 5\,\varepsilon + \sqrt{4\,\hat{s}_{12}^2 - 8\,\hat{s}_1\,\hat{s}_{12} + 4\,\hat{s}_{12}\,\varepsilon + 4\,\hat{s}_1^2 - 12\,\hat{s}_1\,\varepsilon + \varepsilon^2})(2\,\hat{s}_{12} + \varepsilon + 2\,\hat{s}_1)}{4(\,\hat{s}_{12} + \varepsilon)}$$

2.

$$c = \frac{(6\,\hat{s}_{12} + 2\,\hat{s}_1 + 5\,\varepsilon - \sqrt{4\,\hat{s}_{12}^2 - 8\,\hat{s}_1\,\hat{s}_{12} + 4\,\hat{s}_{12}\,\varepsilon + 4\,\hat{s}_1^2 - 12\,\hat{s}_1\,\varepsilon + \varepsilon^2})(2\,\hat{s}_{12} + \varepsilon + 2\,\hat{s}_1)}{4(\,\hat{s}_{12} + \varepsilon)}$$

We are interested in the first (bigger) one - root number 1. This is another function of $\hat{s}_1$, $\hat{s}_{12}$ and ε and let c'' be the value of that function for the input variables $\hat{s}_1$, $\hat{s}_{12}$ and ε. We show that c' is bigger than c''. The expression $c' - c''$ has no solutions in the positive quadrant (of $\hat{s}_1$, $\hat{s}_{12}$ and ε) but has at least one positive value. therefore:

Corollary 7 *For reasonable input $\hat{s} = [\hat{s}_0, \hat{s}_1, \hat{s}_1, \hat{s}_{12}, \hat{s}_1, \hat{s}_{13} + \varepsilon, \hat{s}_{23}, \hat{s}_1]$ such that $\sum_\alpha \hat{s}_\alpha = c$, if $q_{12} > 0$ (i.e. $c > c'$), then $\frac{\partial r}{\partial \varepsilon} > 0$ ($c > c''$).*

Proof. By the last argument we know that for all reasonable data, $c' > c''$ and therefore, by Corollary 6 we can conclude that $c > c''$. By the fact that the function $\frac{\partial r}{\partial \varepsilon}$ has only two roots and we consider the bigger one, the corollary follows. ■

Figure 6 depicts the situation. The x-axis corresponds to c. The solid line corresponds to the function $q_{12}^{(12)}$ (as a function of c). The point where the line crosses the x-axis is c'. The dashed line corresponds to the function $\frac{\partial r}{\partial \varepsilon}$ (as a function of c). The point where the line crosses the x-axis

is c''. As explained above, it can be noted that $c' > c''$, and for all $c_0 > c'$, the dashed line is in the positive quadrant. This particular example is for the values $\hat{s}_1 = 40$, $\hat{s}_{12} = 7$, $\varepsilon = 3$ and $c = 700$, so $c'' = 465$, $c' = 676$. The dotted line corresponds to r (again, as a function of c). For clarity this line is the logarithm of r so the point where it crosses the x-axis is where $r = 1$. the stretch left to that point is the reasonable input where $r < 1$. However, as the theorem states, we can see that for $c > c'$, $r > 1$.

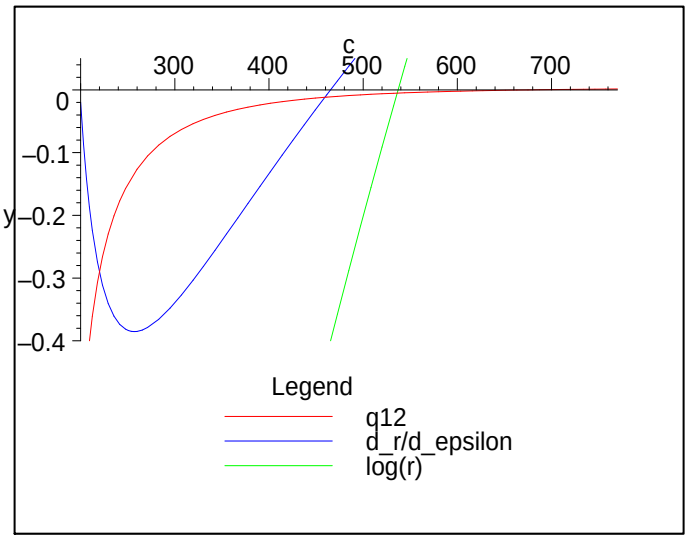

Figure 6: $\frac{\partial r}{\partial \varepsilon}$, log(r), q_{12} as function of c.

It is left now to show that indeed in the point $\hat{s}$, $r > 1$ (The fact that $\frac{\partial r}{\partial \varepsilon} > 0$ does not guarantee that $r > 1$, since, as can be shown in the Figure 6, $\frac{\partial r}{\partial \varepsilon}$ is positive from $c = 465$ but $r < 1$ up to point $c = 537$). For this we exploit the fact that for all given input data, increasing ε increases the point c':

$$\frac{\partial c'}{\partial \varepsilon} = \frac{1}{4}\frac{\left(1+\frac{\frac{1}{2}\left(2\,\varepsilon+4\,\hat{s}_1+4\,\hat{s}_{12}\right)}{\sqrt{\varepsilon^2+4\,\hat{s}_1\,\varepsilon+4\,\hat{s}_{12}\,\varepsilon+4\,\hat{s}_1^2-8\,\hat{s}_1\,\hat{s}_{12}+4\,\hat{s}_{12}^2}}\right)\left(2\,\hat{s}_{12}+\varepsilon+2\,\hat{s}_1\right)}{\hat{s}_{12}} + \frac{\frac{1}{4}\left(\varepsilon+2\,\hat{s}_1+6\,\hat{s}_{12}+\sqrt{\varepsilon^2+4\,\hat{s}_1\,\varepsilon+4\,\hat{s}_{12}\,\varepsilon+4\,\hat{s}_1^2-8\,\hat{s}_1\,\hat{s}_{12}+4\,\hat{s}_{12}^2}\right)}{\hat{s}_{12}}$$

and as can be seen, both components are positive.
So if we start from our input point $\hat{s}$, maintain all variables except ε constant, and decrease ε, by corollary 7 r should decrease. By the facts that c' decreased, c is fixed and by corollary 6, $\frac{\partial r}{\partial \varepsilon}$ is still positive. So we can continue decreasing ε. We repeat this process until $\varepsilon = 0$. We got a monotonic increasing function $r(\varepsilon)$ in the range $[0, \varepsilon]$ with $r(0) = 1$ (see Figure 7).

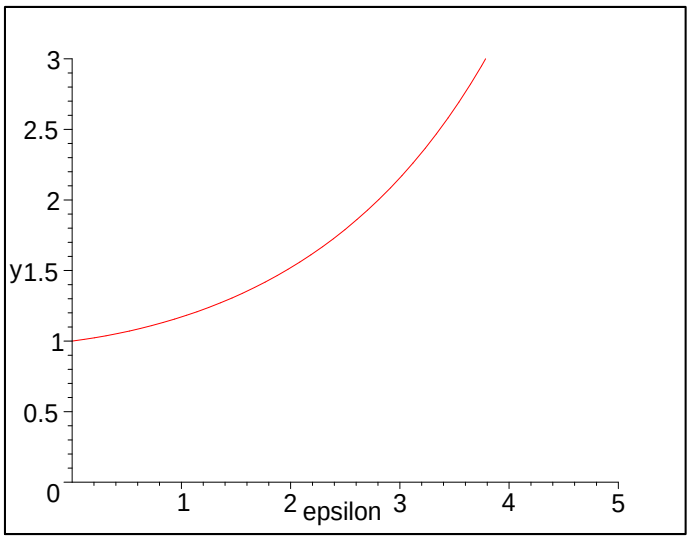

Figure 7: r as a function of ε.

Corollary 8 $r(\varepsilon_0) > 1$.

■

6 Experimental Results

In order to evaluate our method we performed one computational experiment on genomic sequences. Since our methods are designed for four species, we took four homologous sequence from species where it is believed that MC approximately holds. We aligned the sequences using CLUSTALW [11]. We then converted the sequences to two states (purines – A,G and pyrimidines – C,T) and applied our formulae to get the best ML MC-fork. We then compared this tree to the one obtained by running Phylip [3] under MC for the "regular" four states DNA sequences. We took the sequences NK cell receptor D gene of the species Human, Chimpanzee (pan), Rat (rattus), and mouse (mus) (accession numbers AF260135, AF259063, AF009511, and AF030313, respectively). After conversion, the resulting observed spec is $\hat{s} = [539, 0, 0, 79, 9, 0, 0, 15]$. The trees inferred by our method and Phylip dnamlk were almost identical, and are shown in Figure 6.

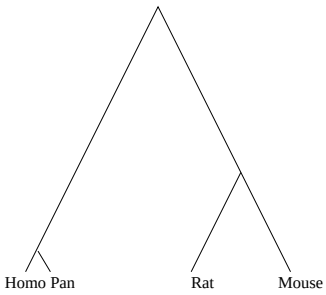

Figure 8: ML MC tree for NK cell receptor D. gene

7 Concluding Remarks

By applying novel algebraic techniques, we extended the state of the art in analytic solutions of ML to four taxa under molecular clock in the 2-states model. We believe this is a significant step in this area of research. We remark that the use of Hadamard conjugation in this context seems necessary. Not only does it allow variable elimination, but the resulting system is substantially simplified. Alternatively, it is possible to formulate the (12)(34) molecular clock fork as a system with only three variables, p_1 for the two pendant edges of taxa 1 and 2, p_3 for 3 and 4, and p_{12} for the central edge (see Figure 3). However, this formulation yields a polynomial system of total degree 12, with up to 550 monomials per equation. This should be contrasted with the system of total degree 5 and up to 63 monomials that we get for the molecular clock fork. It is this difference that enabled us to bring the problem to that degree of complexity that is solvable, when employing symbolic algebra tools. These techniques also enabled us to find the ML molecular clock fork for two families of observed data (the so called singletons and non-singletons). Even for these relatively simple families, the solution exhibited technical difficulties, and in some cases anomalies (in the form of negative edges) occur.

The next step in this line of research is to extend this result to the other molecular clock four taxa topology – the so called comb. In this case we "lose" one of the simple equalities $q_i = q_j$, and have a more complicated identity instead. This small change makes the resulting system substantially harder to solve.

Acknowledgments

We would like to thank Mike Hendy, Mike Steel and Ziheng Yang for very helpful discussions and for comments on earlier versions of this work.

References

[1] B. Chor, M. Hendy, B. Holland, and D. Penny. Multiple maxima of likelihood in phylogenetic trees: An analytic approach. *MBE*, 17(10):1529–1541, 2000. Earlier version appeared in RECOMB 2000.

[2] B. Chor, M. Hendy, and D. Penny. Analytic solutions for three taxon mlmc trees with variable rates across sites. In *WABI 2001*, 2001.

[3] J. Felsenstein. PHYLIP - phylogenetic inference package, (version 3.2). *Cladistics*, 5:164–166, 1989.

[4] K. Fukami and Y. Tateno. On the uniqueness of the maximum likelihood method for estimating molecular trees: Uniqueness of the likelihood point. *J. Mol. Evol.*, 28:460–464, 1989.

[5] M. D. Hendy and D. Penny. Spectral analysis of phylogenetic data. *J. Classif.*, 10:5–24, 1993.

[6] M. D. Hendy, D. Penny, and M.A. Steel. Discrete fourier analysis for evolutionary trees. *Proc. Natl. Acad. Sci. USA.*, 91:3339–3343, 1994.

[7] T.H. Jukes and C.R. Cantor. Evolution of protein molecules. In H.N. Munro, editor, *Mammalian Protein Metabolism*, pages 21–132. Academic Press, New York, 1969.

[8] M. Kimura. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge, 1983.

[9] J. Neymann. Molecular studies of evolution: A source of novel statistical problems. In S. Gupta and Y. Jackel, editors, *Statistical Decision Theory and Related Topics*, pages 1–27. Academic Press, New York, 1971.

[10] M. Steel. The maximum likelihood point for a phylogenetic tree is not unique. *Syst. Biology*, 43(4):560–564, 1994.

[11] J.D. Thompson, D.G. Higgins, and T.J. Gibson. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalty and weight matrix choice. *Nucleic Acids Research*, 22:4673–4780, 1994.

[12] E.R.M. Tillier. Maximum likelihood with multiparameter models of substitution. *J. Mol. Evol.*, 39:409–417, 1994.

[13] Z. Yang. Complexity of the simplest phylogenetic estimation problem. *Proc. R. Soc. Lond. B*, 267:109–119, 2000.

A The Gradient of the Likelihood Function

$$
\begin{aligned}
\frac{\partial L}{\partial s_1} &= (s_1 - 2\,{s_1}^2 - 2\,s_1\,s_{13} - s_{12}\,s_1 - s_{12}\,s_{13})^{(c-h_1-h_2-h_3-h_{123}-h_{12}-h_{13}-h_{23})} \\
&\quad (\frac{1}{2} - s_1 - s_{13})^{(h_3+h_{123})}(2\,h_{12}\,{s_1}^3 - 4\,h_{12}\,{s_1}^4 - 2\,h_1\,s_1\,s_{12}\,s_{13} \\
&\quad + 6\,h_1\,{s_1}^2\,s_{12}\,s_{13} + 6\,h_2\,{s_1}^2\,s_{12}\,s_{13} + 6\,h_1\,s_1\,s_{12}\,{s_{13}}^2 + 6\,h_2\,s_1\,s_{12}\,{s_{13}}^2 \\
&\quad - 2\,h_2\,s_1\,s_{12}\,s_{13} - h_{12}\,s_1\,s_{13} - h_{13}\,s_1\,s_{13} - h_{23}\,s_1\,s_{13} + 12\,c\,{s_1}^3\,s_{13} \\
&\quad + 12\,c\,{s_1}^2\,{s_{13}}^2 - h_1\,{s_1}^2\,s_{12} + 2\,h_1\,{s_1}^3\,s_{12} - h_2\,{s_1}^2\,s_{12} + 2\,h_2\,{s_1}^3\,s_{12} \\
&\quad - 12\,h_{12}\,{s_1}^3\,s_{13} - 12\,h_{12}\,{s_1}^2\,{s_{13}}^2 + h_1\,{s_1}^2 + h_2\,{s_1}^2 - h_3\,{s_1}^2 - h_{123}\,{s_1}^2 \\
&\quad + c\,s_1\,s_{13} - h_3\,s_1\,s_{13} - h_{123}\,s_1\,s_{13} + h_3\,s_{12}\,{s_1}^2 + h_{123}\,s_{12}\,{s_1}^2 \\
&\quad - 12\,h_{123}\,{s_1}^2\,{s_{13}}^2 - 12\,h_3\,{s_1}^3\,s_{13} - 12\,h_3\,{s_{13}}^2\,{s_1}^2 - 12\,h_{123}\,{s_1}^3\,s_{13} - 4\,h_3\,{s_1}^4 \\
&\quad - 4\,h_{123}\,{s_1}^4 + 4\,h_3\,{s_1}^3 + 4\,h_{123}\,{s_1}^3 - h_2\,s_{12}\,{s_{13}}^2 + 2\,h_2\,s_{12}\,{s_{13}}^3 \\
&\quad + 4\,h_{13}\,s_1\,{s_{13}}^2 + 4\,h_{23}\,s_1\,{s_{13}}^2 + 4\,h_{123}\,s_1\,{s_{13}}^2 + 4\,h_{12}\,{s_{13}}^2\,s_1 - 4\,c\,{s_{13}}^2\,s_1 \\
&\quad + 8\,h_3\,{s_1}^2\,s_{13} + 4\,h_3\,{s_{13}}^2\,s_1 + 8\,h_{123}\,{s_1}^2\,s_{13} - 12\,h_{13}\,{s_1}^3\,s_{13} - 12\,h_{13}\,{s_1}^2\,{s_{13}}^2 \\
&\quad - 12\,h_{23}\,{s_1}^3\,s_{13} - 12\,h_{23}\,{s_1}^2\,{s_{13}}^2 + 6\,h_{23}\,{s_1}^2\,s_{13} - h_1\,s_{12}\,{s_{13}}^2 + 2\,h_1\,s_{12}\,{s_{13}}^3 \\
&\quad + 6\,h_{12}\,{s_1}^2\,s_{13} + 6\,h_{13}\,{s_1}^2\,s_{13} - 2\,h_1\,{s_1}^2\,s_{13} - 6\,c\,{s_1}^2\,s_{13} - 2\,h_2\,{s_1}^2\,s_{13} \\
&\quad + 2\,h_{13}\,{s_1}^3 - 4\,h_{13}\,{s_1}^4 + 2\,h_{23}\,{s_1}^3 - 4\,h_{23}\,{s_1}^4 - 2\,c\,{s_1}^3 + 4\,c\,{s_1}^4 - 2\,h_1\,{s_1}^3 \\
&\quad - 2\,h_2\,{s_1}^3 - 4\,s_1\,h_{123}\,{s_{13}}^3 - 4\,s_1\,h_{12}\,{s_{13}}^3 + 4\,s_1\,c\,{s_{13}}^3 - 4\,s_1\,h_3\,{s_{13}}^3 \\
&\quad - 4\,s_1\,h_{13}\,{s_{13}}^3 - 4\,s_1\,h_{23}\,{s_{13}}^3 + s_1\,h_3\,s_{12}\,s_{13} + s_1\,h_{123}\,s_{12}\,s_{13})
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial s_{12}} &= -h_{12}\,s_1 + 2\,h_{12}\,{s_1}^2 + 2\,h_{12}\,s_1\,s_{13} + c\,s_{12}\,s_1 + c\,s_{12}\,s_{13} - h_1\,s_{12}\,s_1 - h_1\,s_{12}\,s_{13} \\
&\quad - h_2\,s_{12}\,s_1 - h_2\,s_{12}\,s_{13} - h_3\,s_{12}\,s_1 - h_3\,s_{12}\,s_{13} - h_{123}\,s_{12}\,s_1 - h_{123}\,s_{12}\,s_{13} \\
&\quad - h_{13}\,s_{12}\,s_1 - h_{13}\,s_{12}\,s_{13} - h_{23}\,s_{12}\,s_1 - h_{23}\,s_{12}\,s_{13}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial s_{13}} \;=\;& (s_1 - 2\,{s_1}^2 - 2\,s_1\,s_{13} - s_{12}\,s_1 - s_{12}\,s_{13})^{(c-h_1-h_2-h_3-h_{123}-h_{12}-h_{13}-h_{23})} \\
& (\frac{1}{2} - s_1 - s_{13})^{(h_3+h_{123})}(2\,h_{23}\,{s_{13}}^3\,s_{12} + 2\,h_{13}\,{s_{13}}^3\,s_{12} + 2\,h_3\,s_{12}\,{s_{13}}^3 \\
& + 2\,h_3\,s_{12}\,{s_1}^3 + 6\,h_3\,s_{12}\,{s_1}^2\,s_{13} + 6\,h_3\,s_{12}\,s_1\,{s_{13}}^2 + 2\,h_{123}\,s_{12}\,{s_{13}}^3 \\
& + 2\,h_{123}\,s_{12}\,{s_1}^3 + 6\,h_{123}\,s_{12}\,{s_1}^2\,s_{13} + 6\,h_{123}\,s_{12}\,s_1\,{s_{13}}^2 + 6\,h_{13}\,s_1\,s_{12}\,{s_{13}}^2 \\
& - 2\,h_{13}\,s_1\,s_{12}\,s_{13} + 6\,h_{13}\,{s_1}^2\,s_{12}\,s_{13} - 2\,h_{23}\,s_1\,s_{12}\,s_{13} + 6\,h_{23}\,{s_1}^2\,s_{12}\,s_{13} \\
& + h_{12}\,s_1\,s_{13} + 2\,h_{13}\,s_1\,s_{13} + 2\,h_{23}\,s_1\,s_{13} + h_3\,{s_1}^2 + h_{123}\,{s_1}^2 - c\,s_1\,s_{13} \\
& + h_1\,s_1\,s_{13} + h_2\,s_1\,s_{13} + h_3\,s_1\,s_{13} + h_{123}\,s_1\,s_{13} + h_{23}\,{s_1}^2 + h_{13}\,{s_1}^2 \\
& + 6\,h_{23}\,s_1\,s_{12}\,{s_{13}}^2 - h_3\,s_{12}\,{s_1}^2 - h_{123}\,s_{12}\,{s_1}^2 + 12\,h_{123}\,{s_1}^2\,{s_{13}}^2 \\
& + 12\,h_3\,{s_1}^3\,s_{13} + 12\,h_3\,{s_{13}}^2\,{s_1}^2 + 12\,h_{123}\,{s_1}^3\,s_{13} + 4\,h_3\,{s_1}^4 + 4\,h_{123}\,{s_1}^4 \\
& - 4\,h_3\,{s_1}^3 - 4\,h_{123}\,{s_1}^3 - 2\,h_2\,s_1\,{s_{13}}^2 - 6\,h_{13}\,s_1\,{s_{13}}^2 - h_{13}\,s_{12}\,{s_{13}}^2 \\
& - 6\,h_{23}\,s_1\,{s_{13}}^2 - h_{23}\,s_{12}\,{s_{13}}^2 - 4\,h_{123}\,s_1\,{s_{13}}^2 - 2\,h_{12}\,{s_{13}}^2\,s_1 + 2\,c\,{s_{13}}^2\,s_1 \\
& - 8\,h_3\,{s_1}^2\,s_{13} - 4\,h_3\,{s_{13}}^2\,s_1 - 8\,h_{123}\,{s_1}^2\,s_{13} + 12\,h_{13}\,{s_1}^3\,s_{13} + 12\,h_{13}\,{s_1}^2\,{s_{13}}^2 \\
& - h_{13}\,{s_1}^2\,s_{12} + 2\,h_{13}\,{s_1}^3\,s_{12} + 12\,h_{23}\,{s_1}^3\,s_{13} + 12\,h_{23}\,{s_1}^2\,{s_{13}}^2 - h_{23}\,{s_1}^2\,s_{12} \\
& + 2\,h_{23}\,{s_1}^3\,s_{12} - 2\,h_1\,s_1\,{s_{13}}^2 - 10\,h_{23}\,{s_1}^2\,s_{13} - 2\,h_{12}\,{s_1}^2\,s_{13} - 10\,h_{13}\,{s_1}^2\,s_{13} \\
& - 2\,h_1\,{s_1}^2\,s_{13} + 2\,c\,{s_1}^2\,s_{13} - 2\,h_2\,{s_1}^2\,s_{13} - 4\,h_{13}\,{s_1}^3 + 4\,h_{13}\,{s_1}^4 - 4\,h_{23}\,{s_1}^3 \\
& + 4\,h_{23}\,{s_1}^4 + 4\,s_1\,h_{123}\,{s_{13}}^3 + 4\,s_1\,h_3\,{s_{13}}^3 + 4\,s_1\,h_{13}\,{s_{13}}^3 + 4\,s_1\,h_{23}\,{s_{13}}^3 \\
& - s_1\,h_3\,s_{12}\,s_{13} - s_1\,h_{123}\,s_{12}\,s_{13})
\end{aligned}
$$