Bernoulli **26**(1), 2020, 1–30
https://doi.org/10.3150/18-BEJ1053

Microscopic path structure of optimally aligned random sequences

RAPHAEL ANDREAS HAUSER[1,2] and HEINRICH MATZINGER[3]

[1]*Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, United Kingdom.*
E-mail: hauser@maths.ox.ac.uk; url: www.maths.ox.ac.uk
[2]*Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, United Kingdom*
[3]*School of Mathematics, Georgia Institute of Technology, 686 Cherry Street, Atlanta, GA 30332-0160, USA.*
E-mail: matzi@math.gatech.edu; url: http://www.math.gatech.edu

Considering optimal alignments of two i.i.d. random sequences of length n, we show that for Lebesgue-almost all scoring functions, almost surely the empirical distribution of aligned letter pairs in all optimal alignments converges to a unique limiting distribution as n tends to infinity. This result helps understanding the microscopic path structure of a special type of last-passage percolation problem with correlated weights, an area of long-standing open problems. Characterizing the microscopic path structure also yields robust alternatives to the use of optimal alignment scores alone for testing the homology of genetic sequences.

Keywords: convex geometry; large deviations; percolation theory; sequence alignment

1. Introduction

1.1. Motivation

Let $x = x_1, \dots, x_n$ and $y = y_1, \dots, y_m$ be two finite sequences consisting of letters from a finite alphabet $\mathcal{A}$. An alignment with gaps of x and y is obtained by introducing an arbitrary number of gaps before, in-between and after the entries of each sequence, subject to the restriction that when gaps are considered to be entries in their own right, denoted by $\mathfrak{G}$, both expanded sequences end up having the same total number of entries, and such that no gap ends up being aligned with another gap. The two sequences are then written above one another and aligned entry by entry.

Defining a score $S(\mathfrak{a}, \mathfrak{b})$ for all possible pairs of aligned letters $(\mathfrak{a}, \mathfrak{b}) \in \mathcal{A}^{*2}$, where $\mathcal{A}^* := \mathcal{A} \cup \{\mathfrak{G}\}$ and $\mathcal{A}^{*2} := \mathcal{A}^* \times \mathcal{A}^* \setminus \{(\mathfrak{G}, \mathfrak{G})\}$, the alignment score under the scoring function S is given as the sum of individual scores of aligned letter pairs. An optimal alignment according to S is an alignment with gaps that maximizes the alignment score.

These concepts, which will be more rigorously defined in Section 1.2, are of standard use in computational genomics. Building on these familiar notions, we associate with each alignment with gaps a concept of empirical distribution over the set of letter pairs. Although such empirical distributions can be defined for sequences of differing lengths $n \neq m$, we consider only sequences of equal length n and propose to study the asymptotics when $n \to \infty$ when the letters of x and y are replaced by i.i.d. random variables. We will discuss the salient concepts more rigorously in Section 1.2. Before we do so, let us comment on how our investigation contributes to the literature in several areas.

1.1.1. *Last-passage percolation*

Consider the set

$$E := \big\{\big\{(z,w),(z,w+1)\big\},\big\{(z,w),(z+1,w)\big\} : z,w \in \mathbb{Z}\big\}$$

of vertical and horizontal edges of unit length incident to points in $\mathbb{Z}^2$, and let a random weight $w(e)$ be associated with each edge. In the classical setting of first-passage percolation, these random weights are taken to be i.i.d. with some fixed distribution. A path of smallest total weight between two points a and $b \in \mathbb{Z}^2$ is then sought, any admissible path having to consist of consecutive adjacent edges $e_1,\ldots,e_n \in E$ with e_1 and e_n incident to a and b, respectively. Interpreting the weights as the time it takes to cross an edge, the total weight $w(e_1)+\cdots+w(e_n)$ of a path corresponds to the time it takes to pass along this path from a to b. Minimum weight paths thus correspond to fastest links between the two points. On directed graphs one can also consider a corresponding notion of maximum weight path or slowest link between two points, and one then speaks of last-passage percolation. Analogous concepts can be defined for other graph topologies and models of random edge weights.

The problem of understanding the structure of optimal paths in first-passage percolation was recognized as being important several decades ago but still remains largely unresolved, see Howard [18]. One open question is to characterize the relative proportions of vertical and horizontal edges in a shortest path from the point $(0,0)$ to $(0,n)$ and their asymptotics as n goes to infinity.

In this paper, we ask a similar question relating to a special last-passage percolation problem, which we shall now describe. Consider the set of oriented edges

$$E' := \big\{\big((z,w),(z,w+1)\big),\big((z,w),(z+1,w)\big),\big((z,w),(z+1,w+1)\big) : z,w \in \mathbb{Z}\big\},$$

let a scoring function S be given, and define random edge weights

$$w(e) = \begin{cases} S(X_{z+1}, Y_{w+1}) & \text{if } e = \big((z,w),(z+1,w+1)\big), \\ S(X_{z+1}, \mathfrak{G}) & \text{if } e = \big((z,w),(z+1,w)\big), \\ S(\mathfrak{G}, Y_{w+1}) & \text{if } e = \big((z,w),(z,w+1)\big). \end{cases}$$

The problem of optimally aligning the random sequences X and Y according to S now becomes a last-passage percolation problem, as there exists a one-one correspondence between paths from $(0,0)$ to (n,n) along the oriented edges from E' and alignments with gaps of X with Y, and as the length of such a path equals the alignment score of the associated alignment with gaps. The optimal alignment score thus equals the weight of a maximum weight path. Furthermore, the empirical distribution of an alignment with gaps reveals the relative proportions of diagonal, horizontal and vertical edges in the associated path. Analogous results for first-passage percolation can be obtained by multiplying the edge weights by -1.

Another type of path microstructure that was investigated in the context of optimal alignments is the local uniqueness of an optimal path, see Hauser and Matzinger [14] and Hirmo, Lember and Matzinger [17] for a definition, theoretical properties and an application of this concept.

We emphasize that optimal alignments with gaps correspond to very special last-passage percolation problems, as edge weights exhibit long-range correlation. As a result, the qualitative behavior of this model is different from other models considered in the literature. For example, Theorem 6.2 below shows that in many cases the fluctuations of optimal alignment scores are of Gaussian order. See also Amsalu, Hauser and Matzinger [2], where strong evidence is provided that this is the generic order. In contrast, other last- and first-passage models are believed to have fluctuations of smaller order.

1.1.2. *Computational genomics*

Consider sections $x = x_1, \ldots, x_n$ and $y = y_1, \ldots, y_m$ of DNA or RNA sequences of two different taxa, each letter representing a site. One would like to decide if x and y have similar biological function and are likely to have evolved from a common ancestral sequence.

Under the assumption of a Markov model of independently evolving sites, a scoring function S is chosen by setting the value $S(\mathfrak{a}, \mathfrak{b})$ at $(\mathfrak{a}, \mathfrak{b}) \in \mathcal{A}^{*2}$ equal to the logarithm of the probability that a letter from the ancestral genome evolved into the letter $\mathfrak{a}$ in the first of the extant taxa and into the letter $\mathfrak{b}$ in the second. In alignments with gaps of x and y, aligned letters are thus interpreted as having evolved from a common ancestral site. Letters aligned with a gap are interpreted as a site in the ancestral genome that became deleted in one of the two extant taxa, or as a new site inserted, by mutation, into the genome of one of the two taxa. The Markov model must therefore also account for the probabilities of insertions and deletions.

Under this choice of scoring function, optimal sequence alignments theoretically correspond to maximum likelihood homologies of genetic sequences, see Karlin and Altschul [20], Waterman [34] and Baxevanis and Ouellette [3]. Naturally, this choice of scoring function depends on how long ago the two taxa got separated on the phylogenetic tree, that is, for how long they evolved without exchange of genetic material. In practice it is however impossible to identify such a clean-cut scoring function, as the Markov model is simplistic and rates of mutation and time since evolutionary separation are not known exactly. Taking feedback of biologists into account, the widely used scoring function that underlies the BLASTZ algorithm (see Baxevanis and Ouellette [3]) has been developed to produce decisions on homology that are as biologically relevant as possible. However, wrong homology classifications may still occur.

A recently pioneered approach to further improve the accuracy of decisions on the homology of two sequences is to exploit properties of the microscopic path structure of optimal alignments. Using the concept of local uniqueness of optimal alignments (see Hauser and Matzinger [14]), Hirmo, Lember and Matzinger [17] found that optimal alignments of homologous and non-homologous sequences have entirely different microscopic structures. Preliminary experiments showed that conducting homology classification on the basis of path micro-structure is competitive with the BLASTZ algorithm of Baxevanis and Ouellette [3], even when an unsophisticated scoring function is used. This suggests that the exploitation of path micro-structure offers robust alternatives to homology classification on the basis of optimal alignments alone. The results of the present paper help taking this work further.

1.1.3. *Monte Carlo simulation of null-models*

Statistical tests for deciding the homology of two sequences under a particular scoring function typically involve a null-model of i.i.d. random sequences. More specifically, two sequences are

deemed to be homologous if their optimal alignment score is significantly higher than typical optimal alignment scores of two random sequences of the same length. Let $L_n(S)$ be the optimal alignment score of two random sequences of length n with i.i.d. letters drawn from some fixed distribution on $\mathcal{A}$. In order to design a statistical test, one typically needs to know $\mathrm{E}[L_n(S)]$ and $\mathrm{VAR}(L_n(S))$, or estimates thereof with guaranteed error bounds.

As will be further discussed in Section 1.9 below, the ratio $\lambda_n(S) := \mathrm{E}[L_n(S)]/n$ is known to converge to a limit $\lambda(S)$ called the Chvàtal–Sankoff constant. While the exact value of this constant is unknown even in the simplest cases and straight-forward Monte Carlo simulation fails to produce estimates of usable accuracy, more sophisticated Monte Carlo simulation methods can be applied to obtain tight estimates, see Hauser, Martinez and Matzinger [13] and Hauser, Matzinger and Dürringer [9]. Having simulated an approximation of $\lambda(S)$, estimates of $\mathrm{E}[L_n(S)]$ are available, since by (1.10) we know that $\lambda_n(S) \leq \lambda(S)$, and since a significant amount of information is known about the difference $\lambda(S) - \lambda_n(S)$: Lemma 3.2 below, proven in Amsalu, Hauser and Matzinger [2], establishes a bound of the form

$$\lambda(S) - \lambda_n(S) \leq \mathcal{O}\left(\frac{\sqrt{\ln n}}{\sqrt{n}}\right). \tag{1.1}$$

Remarkably, the bound (1.1) is independent of the distribution of the letters over $\mathcal{A}$ but only depends on the scoring function and the assumption that the letters of X and Y be i.i.d. In general, the order of the bound (1.1) is tight, as Alexander [1] proved a lower bound of the form

$$\lambda(S) - \lambda_n(S) \geq c\frac{\sqrt{\ln n}}{\sqrt{n}}$$

in the case of the longest common subsequence problem (LCS), which will be further discussed in Example 1.2 of Supplement A [16].

Bounds on $\mathrm{VAR}(L_n(S))$ are less well understood. In the case of the LCS scoring function of Example 1.2 of Supplement A and sequences consisting of i.i.d. Ber(0.5) variables, Chvàtal and Sankoff [8] conjectured that the variance is of order $\mathrm{VAR}(L_n(S)) = o(n^{2/3})$. Steele [30] later proved the bound $\mathrm{VAR}(L_n(S)) \leq 2p(1-p)n$ for the case of Ber(p) variables. Waterman [33] raised the question as to whether or not this bound can be improved. His simulations suggested that for $p < 0.5$ the dependence of $\mathrm{VAR}(L_n(S))$ on n is linear. Boutet de Monvel [6] found that this also applies to the case $p = 0.5$, although the linear growth only sets in for very large n. For the case where p is very small, Lember and Matzinger [23] gave a rigorous proof of the linear order $\mathrm{VAR}(L_n(S)) = \Theta(n)$. Their analysis was based on showing that the manipulation of randomly selecting a letter of specified type from one of the two sequences and changing it into another specified type has a positive biased effect on the optimal alignment score. See also Hauser and Matzinger [14], where this technique was pioneered.

The present paper significantly extends the applicability of this technique to general scoring functions and to random sequences whose distributions are not highly asymmetric. The already mentioned Theorem 6.2 yields a sufficient criterion under which the asymptotic order $\mathrm{VAR}(L_n(S)) = \Theta(n)$ holds.

A related problem was studied in Hauser, Popescu and Matzinger [15], where it was shown that for two randomly sampled symmetric scoring functions S and T, the deviation of the score

relative to T of an optimal alignment relative to S of two random sequences of length n has a deviation of order $\mathcal{O}((\log n)^{1/4} n^{3/4})$. Technically, the cited paper is related to the present discussion, as it also approaches a problem of large deviations via geometric probability and convex analysis.

1.2. Basic concepts and notation

1.2.1. *Alignments with gaps*

An *alignment with gaps* π of $x = x_1, \ldots, x_n$ and $y = y_1, \ldots, y_m$ is defined by two strictly increasing functions $\pi_x : \{1, \ldots, k\} \to \{1, \ldots, n\}$ and $\pi_y : \{1, \ldots, k\} \to \{1, \ldots, m\}$. The $\pi_x(i)$th entry of x is considered aligned with the $\pi_y(i)$th entry of y, and all other entries are considered to be aligned with gaps.

We remark that this definition corresponds to equivalence classes of the less formal notion of alignments with gaps described in Section 1.1, as the introduction of gaps is not fully determined in all cases. For example, the alignments

$$\begin{array}{|c|c|c|c|c|}\hline x & \mathfrak{a} & \mathfrak{G} & \mathfrak{G} & \mathfrak{b} \\ \hline y & \mathfrak{G} & \mathfrak{b} & \mathfrak{a} & \mathfrak{b} \\ \hline \end{array} \quad \text{and} \quad \begin{array}{|c|c|c|c|c|}\hline x & \mathfrak{G} & \mathfrak{a} & \mathfrak{G} & \mathfrak{b} \\ \hline y & \mathfrak{b} & \mathfrak{G} & \mathfrak{a} & \mathfrak{b} \\ \hline \end{array}$$

both correspond to the choice $k = 1$, $\pi_x : 1 \mapsto 2$, and $\pi_y : 1 \mapsto 3$. However, this is inconsequential, since all alignments with gaps from the same equivalence class behave as if they were identical in everything that follows.

1.2.2. *The empirical distribution*

We may assume that an order has been fixed on $\mathcal{A}^*$, so that a lexicographic ordering of $\mathcal{A}^{*2}$ is well-defined. Let π be an alignment with gaps of two sequences x and y of equal length n. For each possible pair of letters $(\mathfrak{a}, \mathfrak{b}) \in \mathcal{A}^{*2}$ let us count the number of times a letter $\mathfrak{a}$ in x is aligned with a letter $\mathfrak{b}$ in y. Divide this number by n and denote the result by $p_{\mathfrak{a}\mathfrak{b}}$. We call the vector $\vec{p}_\pi(x, y)$ of all such ratios collected in lexicographical order the *empirical distribution vector* of π.

We remark that the vector $\vec{p}_\pi(x, y)$ corresponds to an empirical probability distribution in the classical sense that is scaled up by a factor $\tau \geq 1$, the presence of which is due to having divided the frequencies by the length n of the sequences *not counting* the gaps rather than by the length of the sequences *including* the gaps. See Example 1.1 in Supplement A for an example of an empirical distribution of two sequences.

1.2.3. *Two sets of empirical distributions*: $\mathrm{SET}(X, Y)$ *and* SET^n

We denote the set of all empirical distribution vectors associated with x and y by

$$\mathrm{SET}(x, y) := \left\{ \vec{p}_\pi(x, y) : \pi \text{ is an alignment with gaps of } x \text{ and } y \right\}. \tag{1.2}$$

Let us next consider two independent random sequences $X = X_1, \ldots, X_n$ and $Y = Y_1, \ldots, Y_n$, where the random variables X_i and Y_j are all i.i.d., taking values in $\mathcal{A}$ with some fixed distribution. We denote the convex hull of $\mathrm{SET}(X, Y)$ by

$$\mathrm{SET}^n := \mathrm{conv}\big(\mathrm{SET}(X, Y)\big), \tag{1.3}$$

accounting for n because we will be interested in the asymptotics as n tends to infinity.

1.2.4. *Scoring functions and optimal alignments*

A *scoring function* is a function $S : \mathcal{A}^{*2} \to \mathbb{R}$. For an alignment with gaps π of $x = x_1, \ldots, x_n$ and $y = y_1, \ldots, y_m$, let us define the score $S_\pi(x, y)$ under S as the sum of scores of the aligned letter pairs, that is,

$$S_\pi(x, y) = \sum_{i=1}^{k} S(x_{\pi_x(i)}, y_{\pi_y(i)}) + \sum_{j \notin \pi_x(\{1,\ldots,k\})} S(x_j, \mathfrak{G}) + \sum_{j \notin \pi_y(\{1,\ldots,k\})} S(\mathfrak{G}, y_j).$$

We write

$$L_S(x, y) := \max_{\pi} S_\pi(x, y) \tag{1.4}$$

for the optimal alignment score of x and y, where the maximum is taken over all alignments with gaps π of x and y. Any maximizing alignment π^* of (1.4) is called an *optimal alignment according to* S. Multiple optimal alignments may exist for a given pair of sequences and a given scoring function. The dynamic programming approach of Needleman and Wunsch [27] allows to identify an optimal alignment with gaps in in $\mathcal{O}(nm)$ time. Furthermore, if there are k optimal alignments, the algorithm can be easily amended to identify all of these in $\mathcal{O}(nm + k(n + m))$ time.

We remark that, although alignment scores and optimal alignments are well defined for sequences x and y of different lengths, hereafter we will only consider sequences that are of equal length n, as we will be interested in asymptotic results when $n \to \infty$. See Example 2 of Supplement A for examples of optimal alignments.

1.2.5. *Relating* SET^n *to optimal alignment scores*

For any fixed scoring function S, we may define the linear functional

$$\begin{aligned} f_S : \mathbb{R}^{|\mathcal{A}^{*2}|} &\to \mathbb{R}, \\ \vec{x} &\mapsto \sum_{(\mathfrak{a},\mathfrak{b}) \in \mathcal{A}^{*2}} S(\mathfrak{a}, \mathfrak{b}) x_{\mathfrak{a}\mathfrak{b}}. \end{aligned} \tag{1.5}$$

We then have $S_\pi(x, y) = n f_S(\vec{p}_\pi(x, y))$, and consequently,

$$\frac{L_S(x, y)}{n} = \max_{\vec{p} \in \mathrm{SET}(x,y)} f_S(\vec{p}) = \max_{\vec{p} \in \mathrm{conv}(\mathrm{SET}(x,y))} f_S(\vec{p}), \tag{1.6}$$

the last equation following from the fact that the maximum of a linear functional over a set equals its maximum over the convex hull of this set.

The problem of maximizing the alignment score over the set of all alignments with gaps, a purely *combinatorial* optimization problem, can thus be reformulated as a convex optimization problem, a class of particularly well-behaved *continuous* optimization problems. The reformulation is conceptual in that an explicit description of $\mathrm{conv}(\mathrm{SET}(x, y))$ is generally exponentially hard to come by. Hence, this does not provide an avenue for replacing the dynamic programming approach by a continuous algorithm. Nonetheless, the reformulation provides a powerful theoretical tool that will be central to our analysis, making it possible to prove our main results via the machinery of convex analysis. The relevant results will be developed in Section 2. A deep connection between convex optimization and optimal sequence alignments was also exploited in the design of Monte Carlo methods for the simulation of the Chvàtal–Sankoff constant, see Hauser, Martinez and Matzinger [13] and Hauser, Matzinger and Dürringer [9].

1.2.6. *The Chvàtal–Sankoff constant*

From this point onwards we consider only the random sequences X and Y introduced in Section 1.2.3. To account for the dependence of $L_S(X, Y)$ on the common length n of these sequences, we revert to the notations

$$L_n(S) := L_S(X, Y) \tag{1.7}$$

and $\lambda_n(S) = E[L_n(S)]/n$ introduced in Section 1.1.3.

It is easy to see that the sequence $(-L_n(S))_{n\in\mathbb{N}}$ is subadditive. Exploiting this property, Chvàtal and Sankoff [8] showed that

$$\lambda_n(S) \le \lambda_m(S), \quad \forall m = kn, k \in \mathbb{N}, \tag{1.8}$$

$$\lambda(S) := \lim_{n\to\infty} \lambda_n(S) \quad \text{exists}, \tag{1.9}$$

$$\lambda_n(S) \le \lambda(S), \quad \forall n \in \mathbb{N}, \tag{1.10}$$

$$\mathrm{P}\left[\lim_{n\to\infty} \frac{L_n(S)}{n} = \lambda(S)\right] = 1. \tag{1.11}$$

Equation (1.9) follows from Fekete's lemma (see Fekete [10]). See also Steele [32]. All four relations also easily follow from Kingman's Subadditive Ergodic theorem (see Kingman [21] and Steele [31]).

1.2.7. *Another set of empirical distributions*: SET

Another set of empirical distributions of interest is defined by

$$H(S) := \left\{\vec{x} \in \mathbb{R}^{|\mathcal{A}^{*2}|} : f_S(\vec{x}) \le \lambda(S)\right\}, \tag{1.12}$$

$$\mathrm{SET} := \bigcap_S H(S), \tag{1.13}$$

where the intersection in (1.13) is taken over all scoring functions S. Alternatively, since $f_S(\vec{x})$ and $\lambda(S)$ are both homogeneous of degree 1 in S, it suffices to take the intersection over the scoring functions on the unit sphere $\mathrm{S}^{|\mathcal{A}^{*2}|-1}$. It is immediate from (1.13) that SET is a closed convex set. In Lemma 2.4, we will furthermore show that it is compact and nonempty.

1.2.8. *Distance between sets*

The Hausdorff distance (see Hausdorff [12]) between two subsets $A, B \subseteq \mathbb{R}^n$ is defined as follows, where $\|\cdot\|$ denotes the Euclidean norm,

$$\begin{aligned} d(A,B) &= \max\Big(\sup_{x\in A} d(x,B), \sup_{y\in B} d(A,y)\Big), \\ d(x,B) &= \inf\big\{\|x-y\| : y\in B\big\}, \\ d(A,y) &= \inf\big\{\|x-y\| : x\in A\big\}. \end{aligned} \tag{1.14}$$

1.3. A roadmap to the paper

1.3.1. *Contributions*

The following are the main results of this paper: Theorem 4.1 establishes that SET^n almost surely converges to the deterministic set SET in the topology of the Hausdorff distance. We remark that SET only depends on the distribution of the sequences X and Y, but not their realization, while SET^n depends only on the realization. Furthermore, while the definition of SET involves scoring functions, the definition of SET^n does not, as it is solely based on the combinatorial notion of alignments with gaps. Theorem 4.1 establishes the nontrivial fact that the notions of scoring functions and empirical distributions asymptotically become duals of each other. Theorem 5.1 shows that, as n tends to infinity, the empirical distributions of all optimal alignments of X and Y almost surely converge to a deterministic distribution, on condition that the scoring function S be chosen such that f_S has a unique maximizer in SET. Whenever this condition is met, we denote the unique maximiser by $\vec{p}_S$, a vector that depends only on the distribution of X and Y, but not on their realizations. The theorem quantifies the probability that there exists an optimal alignment of X and Y with respect to S with empirical distribution further away than $\epsilon > 0$ from $\vec{p}_S$ as negatively exponentially small in n, where ϵ is an arbitrary small constant independent of n. The condition of Theorem 5.1 is difficult to verify in practice, but Theorem 2.1 shows that the condition is met generically, that is, for Lebesgue-almost every scoring function. As a corollary, we obtain Theorem 5.2, which says that for Lebesgue-almost every scoring function S the empirical distributions of all optimal alignments of X and Y almost surely converge to a deterministic distribution. A further consequence is Theorem 6.2, which provides a sufficient criterion to guarantee that the fluctuation (defined as the standard deviation) of the optimal alignment score is of order $\Theta(\sqrt{n})$. This criterion constitutes a practical tool in the design of a statistical test on the order of fluctuation of the optimal score. A related approach based on Monte Carlo simulation was discussed in Amsalu, Hauser and Matzinger [2].

1.3.2. *Key ideas*

Applying (1.6) to the random sequences X and Y, we have

$$\frac{L_n(S)}{n} = \max_{\vec{p}\in \mathrm{SET}^n} f_S(\vec{p}). \tag{1.15}$$

By (1.11), $L_n(S)/n$ almost surely converges to a deterministic constant $\lambda(S)$ which also appeared in the definition of SET given in (1.12) and (1.13). These equations further imply that

$$\max_{\vec{p}\in \mathrm{SET}} f_S(\vec{p}) \le \lambda(S). \tag{1.16}$$

Lemma 2.4(d) below shows that Inequality (1.16) holds in fact as an equality. Combined with (1.15), this implies

$$\max_{\vec{p}\in \mathrm{SET}^n} f_S(\vec{p}) \overset{n\to\infty}{\longrightarrow} \max_{\vec{p}\in \mathrm{SET}} f_S(\vec{p}) \quad \text{almost surely.} \tag{1.17}$$

At a first pass it is illustrative to see approximate proofs of the main theorems, free of the large deviations complications that will be present in the rigorous arguments we will give later. Theorem 2.1 and Propositions 2.2 and 2.4 from Section 2, provide the crucial insight: By (1.17), the conditions of Proposition 2.4 are approximately met for $C = \mathrm{SET}$ and $C^n = \mathrm{SET}^n$. This makes it plausible that $\mathrm{SET}^n \to \mathrm{SET}$, as claimed in Theorem 4.1. In the rigorous proof, we will use the fact that $L_n(S)/n$ converges to $\lambda(S)$ at a rate of order $\mathcal{O}(\ln n/\sqrt{n})$, which follows directly from the Azuma–Hoeffding Inequality, as we shall see in Section 3. The convergence of SET^n to SET occurs at the same rate. Theorem 2.1 shows that the conditions of Proposition 2.2 are satisfied for Lebesgue-almost every linear functional f_S, and since choosing the scoring function S generically is tantamount to choosing f_S generically, Theorem 5.2 follows.

1.3.3. *Key difficulties*

The random variable $L_n(S)$ is a function of the realizations of the i.i.d. random sequences X and Y. Lemma 3.1 will show that changing the realization of only one entry from either X or Y results in a change of $L_n(S)$ by at most the deterministic constant

$$\max_{(\mathfrak{d},\mathfrak{c}),(\mathfrak{e},\mathfrak{c})\in\mathcal{A}^{*2}} \big|S(\mathfrak{d},\mathfrak{c}) - S(\mathfrak{e},\mathfrak{c})\big|.$$

One can thus apply the Azuma–Hoeffding Inequality to find that, on a scale of $\sqrt{n}$, the tail of $L_n(S)$ decays at least quadratically exponentially fast: In the notation of Lemma 3.3, set $m = 2n$, $Z_i = X_i$ for $(i = 1, \dots, n)$ and $Z_j = Y_{j-n}$ for $(j = n+1, \dots, m)$. We then have $g(Z_1, \dots, Z_m) = L_n(S)$, and setting $\epsilon = t/\sqrt{m}$, the lemma implies that a deviation of $L_n(S)$ from its mean by $t\sqrt{2n}$ is quadratically exponentially rare in t. This powerful tool lends itself to an elegant analysis of the asymptotic convergence of the alignment score and its fluctuations.

In contrast, analyzing the convergence of the empirical distribution of letter pairs in optimal alignments is much harder: upon changing the realization of one of the random letters, it has to be assumed a priori that the entire optimal alignment has changed, and likewise the relative frequencies at which pairs of letters are aligned. As a consequence, the Azuma–Hoeffding

Inequality cannot be applied directly. Luckily, it can be applied indirectly through the optimal alignment scores of different scoring functions, but this comes at the cost of having to deal with additional technicalities.

A further key difficulty is that for the scoring functions S under consideration it is required that f_S be maximized in only one point on SET. This condition would be met if SET were known to be strictly convex everywhere, but this seems very difficult to verify in practice, as the exact shape of SET is unknown: SET corresponds to the asymptotic shape of the wet zone in the first/last passage percolation formulation of our problem, and determining the shape of the corresponding zone in standard first passage percolation is a long-standing open problem in the general case. We get around this problem by showing that if the scoring function S is chosen generically, then there exists a unique maximizer of f_S on SET, see Theorem 2.1 and Lemma 2.4, which result in Theorem 5.2. Counter-Example 1.1 of Supplement A shows that the claim of Theorem 5.2 fails on a null-set of scoring functions, whence the theorem cannot be extended the set of *all* scoring functions.

2. Convex geometry tools

Section 1.2.5 already alluded to the usefulness of convex geometry as an approach to answering the questions raised in this paper. In this section, we will develop the tools required in the proofs of the main theorems. All lemmas and theorems are used in the line of arguments leading to the proofs of the main theorems. Due to the page limit of the journal, we defer some of the proofs of this section to Supplement A. The approach and results presented here may be useful in the analysis of other first- and last-passage percolation models as well.

S^{n-1} denotes the unit sphere in $\mathbb{R}^n$, $\mathrm{B}_\rho(\vec{x})$ the Euclidean ball of radius ρ around $\vec{x} \in \mathbb{R}^n$, d the Hausdorff distance, $\mathrm{conv}(\cdot)$ the convex hull and $\mathrm{cl}(\cdot)$ the closure of a set in the canonical subspace topology inherited from $\mathbb{R}^n$. We say that a convex set $C \subset \mathbb{R}^n$ has dimension k if its affine hull $\mathrm{aff}(C) \subset \mathbb{R}^n$ has dimension k.

The first theorem plays a key role in the proofs of Lemma 2.1, Theorem 5.2 and Proposition 2.3.

Theorem 2.1. *Let $C \subset \mathbb{R}^n$ be nonempty, compact and convex, and let $\vec{S} : \Omega \to \mathrm{S}^{n-1}$ be a random vector that takes values in the unit sphere with uniform distribution, defined on some probability space $(\Omega, \mathscr{A}, \mathrm{P})$. Then for almost all $\omega \in \Omega$, the optimization problem $\arg\max_{\vec{y} \in C} \langle \vec{S}(\omega), \vec{y} \rangle$ has a unique solution.*

Proof. Let us first consider the case where C has nonempty interior. Upon a shift of C we may assume without loss of generality that $\vec{0}$ lies in the interior of C. Then the polar $C^\circ = \{\vec{w} \in \mathbb{R}^n : \langle \vec{w}, \vec{y} \rangle \leq 1, \forall \vec{y} \in C\}$ is also compact convex with nonempty interior. Seen as the claim of the theorem is invariant under positive scaling, we may further assume without loss of generality that $\mathrm{B}_3(\vec{0}) \subset C^\circ \subset \mathrm{B}_\varrho(\vec{0})$.

Next, let $\vec{s} \in \mathrm{S}^{n-1}$ be a given point on the unit sphere and consider the function

$$\tau_{\vec{s}} : \mathrm{T}_{\vec{s}}\, \mathrm{S}^{n-1} \to \mathbb{R},$$

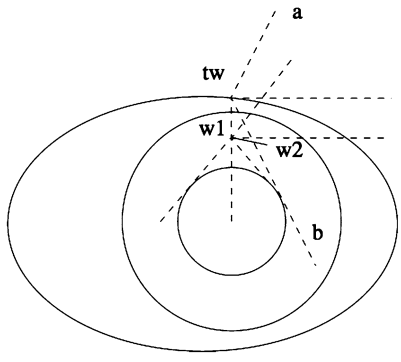

Figure 1. The geometry of the Lipschitz estimate.

$$\vec{w} \mapsto \max\{\tau > 0 : \tau \vec{w} \in C^\circ\}$$

defined on the tangent space at $\vec{s}$. We claim that $\tau_{\vec{s}}$ is Lipschitz continuous on a sufficiently small neighbourhood $\mathscr{V}_{\vec{s}}$ of $\vec{s}$ in $\mathrm{T}_{\vec{s}}\,\mathrm{S}^{n-1} \cap \mathrm{B}_2(\vec{0})$. Let $\vec{w}_1, \vec{w}_2 \in \mathrm{T}_{\vec{s}}\,\mathrm{S}^{n-1} \cap \mathrm{B}_2(\vec{0})$ and $W = \mathrm{span}\{\vec{w}_1, \vec{w}_2\}$. For $(i = 1, 2)$ we then have

$$1 \le \|\vec{w}_i\| < 2, \tag{2.1}$$

$$\tau_{\vec{s}}(\vec{w}_i) = \max\{\tau > 0 : \tau \vec{w}_i \in C^\circ \cap W\}, \tag{2.2}$$

$$1 < \tau_{\vec{s}}(\vec{w}_i)\|\vec{w}_i\| \le \varrho. \tag{2.3}$$

By (2.2), we may assume without loss of generality that $\mathbb{R}^n = W$ for the purposes of proving $|\tau_{\vec{s}}(\vec{w}_1) - \tau_{\vec{s}}(\vec{w}_2)| \le L\|\vec{w}_1 - \vec{w}_2\|$. We refer the reader to Figure 1 for an illustration of the geometric setup. The lines a and b are the tangents from $\tau_{\vec{s}}(\vec{w}_1)\vec{w}_1$ to the unit sphere S^1 in W. Denote the angle between the line $\overline{\vec{w}_1\vec{w}_2}$ and the horizontal at $\vec{w}_1$ by θ, the angle between the horizontal at $\tau_{\vec{s}}(\vec{w}_1)\vec{w}_1$ and the tangents a, b by α, and the angle between the horizontal at $\vec{w}_1$ and the two tangents from $\vec{w}_1$ to S^1 by β. Since the affine hull $\mathrm{aff}(\vec{w}_1, \vec{w}_2)$ cannot enter $\mathrm{B}_1(\vec{0})$, it must lie wedged between the latter two tangents. In combination with (2.1), this implies

$$|\theta| \le \beta = \frac{\pi}{2} - \arcsin\frac{1}{\|\vec{w}_1\|} \le \frac{\pi}{2} - \arcsin\frac{1}{2}. \tag{2.4}$$

Further, (2.3) implies

$$\alpha = \frac{\pi}{2} - \arcsin\frac{1}{\tau_{\vec{s}}(\vec{w}_1)\|\vec{w}_1\|} \le \frac{\pi}{2} - \arcsin\frac{1}{\varrho}. \tag{2.5}$$

Observe that, by convexity, the line segment between the point of tangency of a at S^1 and $\tau_{\vec{s}}(\vec{w}_1)\vec{w}_1$ lies in C°, and further that the definition of $\tau_{\vec{s}}(\vec{w}_1)$ implies $\tau_{\vec{s}}(\vec{w}_1)\vec{w}_1 \in \partial C^\circ$. There-

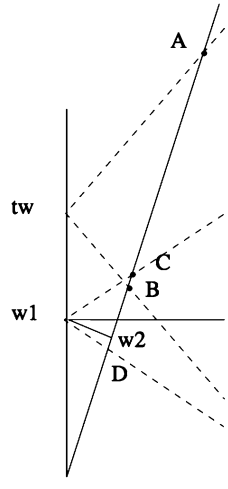

Figure 2. Bounding $\tau_{\vec{s}}(\vec{w}_2)$ by ratios.

fore, the segment of a above $\tau_{\vec{s}}(\vec{w}_1)\vec{w}_1$ lies outside C°, and it follows that

$$\frac{\|B\|}{\|C\|} \leq \tau_{\vec{s}}(\vec{w}_2) \leq \frac{\|A\|}{\|D\|}, \tag{2.6}$$

see Figure 2. Let φ be the angle between $\vec{w}_1$ and $\vec{w}_2$, and let us assume $\varphi < (\pi - 2\theta)/2$, so that the intersection points A, B, C, D exist. This assumption is equivalent to limiting our analysis to a sufficiently small neighbourhood of $\vec{s}$ in $\mathrm{T}_{\vec{s}}\,\mathrm{S}^{n-1}$, as assumed earlier. We can now express the inequalities (2.6) in terms of the angles we introduced,

$$\tau_{\vec{s}}(\vec{w}_1)\frac{1-\tan\varphi\tan\beta}{1+\tan\varphi\tan\alpha} \leq \tau_{\vec{s}}(\vec{w}_2) \leq \tau_{\vec{s}}(\vec{w}_1)\frac{1+\tan\varphi\tan\beta}{1-\tan\varphi\tan\alpha}.$$

This can be simplified by Taylor expansion,

$$\begin{aligned}\big|\tau_{\vec{s}}(\vec{w}_2) - \tau_{\vec{s}}(\vec{w}_1)\big| &\leq \tau_{\vec{s}}(\vec{w}_1)\tan\varphi(\tan\alpha+\tan\beta)\\ &\overset{(2.1),(2.3),(2.4),(2.5)}{\leq} \varrho\tan\varphi\big(\varrho\sqrt{1-\varrho^{-2}}+\sqrt{3}\big),\end{aligned} \tag{2.7}$$

and since $\|\vec{w}_1 - \vec{w}_2\| \geq \|\vec{w}_1\|\tan\varphi$, Equations (2.1) and (2.7) imply

$$\big|\tau_{\vec{s}}(\vec{w}_1) - \tau_{\vec{s}}(\vec{w}_2)\big| \leq L\|\vec{w}_1 - \vec{w}_2\|,$$

with $L = \varrho(\varrho\sqrt{1-\varrho^{-2}}+\sqrt{3})$.

Next, having shown that $\tau_{\vec{s}}$ is Lipschitz continuous on a sufficiently small open neighbourhood $\mathscr{V}_{\vec{s}} \subset \mathrm{T}_{\vec{s}}\,\mathrm{S}^{n-1}$ of $\vec{s}$, Rademacher's theorem (see Rademacher [28] and Gruber [11]) implies that

$\tau_{\vec{s}}$ is Fréchet-differentiable everywhere on $\mathscr{V}_{\vec{s}}$ except on a null-set $\mathscr{B}_{\vec{s}} \subset \mathscr{V}_{\vec{s}}$. We now claim that if the optimization problem

$$\vec{x}(\vec{s}) = \arg\max_{\vec{y}\in C} \langle \vec{s}, \vec{y}\rangle \tag{2.8}$$

has multiple solutions, then $\tau_{\vec{s}}$ is Gâteaux-nondifferentialble at $\vec{s}$. Since $\tau_{\vec{s}}$ is then also Fréchet nondifferentiable at $\vec{s}$, it must be the case that $\vec{s} \in \mathscr{B}_{\vec{s}}$. Let us thus suppose that (2.8) has two different solutions, $\vec{x}_0 \neq \vec{x}_1$. Then $\langle \vec{s}, \vec{x}_1 - \vec{x}_0\rangle = 0$, so that we have $c_1 := \langle \vec{s}, \vec{x}_1\rangle = \langle \vec{s}, \vec{x}_0\rangle$. Furthermore, writing $c_2 := \langle \vec{x}_1 - \vec{x}_0, \vec{x}_0\rangle$ and $c_3 := \langle \vec{x}_1 - \vec{x}_0, \vec{x}_1\rangle$, our assumption that $\vec{x}_0 \neq \vec{x}_1$ implies $c_2 \neq c_3$. Without loss of generality, we may assume that $c_2 < c_3$. For all $\xi \in \mathbb{R}$ let us define $\vec{w}_\xi := \vec{s} + \xi(\vec{x}_1 - \vec{x}_0)$ and consider the restriction $\tau_{\vec{s}}|_{\vec{s}+\mathrm{span}(\vec{x}_1-\vec{x}_0)}$ which we shall denote by $\tau(\xi) := \tau_{\vec{s}}(\vec{w}_\xi)$. Clearly, if $\tau(\xi)$ is nondifferentiable at $\xi = 0$, then $\tau_{\vec{s}}(\vec{w})$ is Gâteaux-nondifferentiable at $\vec{w} = \vec{s}$. The definition of $\tau(\xi)$ implies $\langle \tau(\xi)\vec{w}_\xi, \vec{x}_j\rangle \leq 1$ for $(j = 0, 1)$, so that

$$\begin{aligned}\tau(\xi) &\leq \min\left(\frac{1}{c_1 + c_2\xi}, \frac{1}{c_1 + c_3\xi}\right)\\ &= \frac{1}{c_1}\min\left(1 - \frac{c_2}{c_1}\xi + \mathscr{O}\big(\xi^2\big), 1 - \frac{c_3}{c_1}\xi + \mathscr{O}\big(\xi^2\big)\right).\end{aligned}$$

Furthermore, we have $\tau(0) = 1/c_1$. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}\,\xi_+}\tau(0) = \lim_{\xi\to 0+}\frac{\tau(\xi) - \frac{1}{c_1}}{\xi} \leq -\frac{c_3}{c_1^2} < -\frac{c_2}{c_1^2} \leq \lim_{\xi\to 0-}\frac{\tau(\xi) - \frac{1}{c_1}}{\xi} = \frac{\mathrm{d}}{\mathrm{d}\,\xi_-}\tau(0),$$

showing that $\tau(\xi)$ is nondifferentiable at $\xi = 0$, as claimed.

Next, observe that $\tau_{\vec{s}}$ is Fréchet differentiable at $\vec{w} \in \mathscr{V}_{\vec{s}}$ if and only if the map

$$\begin{aligned}\hat{\tau} : \mathrm{S}^{n-1} &\to \mathbb{R},\\ \vec{z} &\mapsto \max\big\{\tau > 0 : \tau\vec{z} \in C^\circ\big\}\end{aligned}$$

is differentiable at $\hat{w} := \vec{w}/\|\vec{w}\|$ and if and only if $\tau_{\hat{w}}$ is differentiable at $\hat{w}$. Denoting the spherical projections of $\mathscr{V}_{\vec{s}}$ and $\mathscr{B}_{\vec{s}}$ by $\hat{\mathscr{V}}_{\vec{s}}$ and $\hat{\mathscr{B}}_{\vec{s}}$, the compactness of S^{n-1} implies the existence of finitely many points $\vec{s}_1, \ldots, \vec{s}_k \in \mathrm{S}^{n-1}$ such that $\bigcup_{i=1}^k \hat{\mathscr{V}}_{\vec{s}_i} = \mathrm{S}^{n-1}$. Consequently,

$$\mathscr{B} = \bigcup_{i=1}^k \mathscr{B}_{\vec{s}_i}$$

is a null-set with the property that if Problem (2.8) has multiple solutions for a given $\vec{s} \in \mathrm{S}^{n-1}$, then $\vec{s} \in \mathscr{B}$. This proves the claim of the theorem in the case where C has nonempty interior.

Let us now consider the general case. When C consists of a singleton, the claim of the theorem is trivial. We may thus assume that $\dim(C) \geq 1$. Upon a shift we may assume without loss of generality that $\vec{0} \in C$. Let $W = \mathrm{span}(C)$ be the subspace spanned by C, and $W^\perp$ its orthogonal complement under the Euclidean inner product of $\mathbb{R}^n$. We denote the orthogonal projections onto

these spaces by π_W and $\pi_{W^\perp}$ respectively. Finally, let $S_W = S^{n-1} \cap W$ be the unit sphere in W, and

$$\pi_S : S^{n-1} \to S_W,$$
$$\vec{s} \mapsto \frac{\pi_W(\vec{s})}{\|\pi_W(\vec{s})\|}$$

the rescaled projection of S^{n-1} into W.

The condition $\dim(C) \geq 1$ implies $\dim(W^\perp) \leq n-1$, and $\mathscr{B}_{W^\perp} = \{\omega \in \Omega : \vec{S}(\omega) \in W^\perp\}$ is a null-set. Hence, $\pi_S(\vec{s})$ is defined for almost all $\vec{s} \in S^{n-1}$. Further, by isotropy of the uniform distribution on S^{n-1}, the random vector

$$\pi_S(\vec{S}) : \Omega \setminus \mathscr{B}_{W^\perp} \to S_W$$

is uniformly distributed on S_W. Since C has nonempty interior in the subspace topology of W, the case we already settled above applies and implies that

$$\mathscr{B}_W = \left\{\omega \in \Omega \setminus \mathscr{B}_{W^\perp} : \arg\max_{\vec{y} \in C} \langle \pi_S\big(\vec{S}(\omega)\big), \vec{y}\rangle \text{ is nonunique}\right\}$$

is a null-set. Observing that for $\vec{s} \in S^{n-1} \setminus W^\perp$ it is the case that

$$\arg\max_{\vec{y} \in C} \langle \vec{s}, \vec{y}\rangle = \arg\max_{\vec{y} \in C} \langle \pi_S\big(\vec{S}(\omega)\big), \vec{y}\rangle,$$

we find, that $\arg\max_{\vec{y} \in C} \langle \vec{S}(\omega), \vec{y}\rangle$ has a unique solution if and only if ω is not in the null-set $\mathscr{B} = \mathscr{B}_{W^\perp} \cup \mathscr{B}_W$. □

The following notion will play a central role in the sequel.

Definition 2.1. Let $C \subset \mathbb{R}^n$ be convex compact. We say that a boundary point $\vec{x} \in \partial C$ is a point of strict curvature if there exists $\vec{s} \in S^{n-1}$ such that the optimization problem $\max_{\vec{y} \in C} \langle \vec{s}, \vec{y}\rangle$ has $\vec{x}$ as unique maximizer. We denote the set of points of strict curvature by C_{SE}.

Note that if C has a differentiable boundary, then any point where all principal curvatures are nonzero is a point of strict curvature. However, the set of points of strict curvature may be larger. For example, the epigraph of the curve $x \mapsto |x|^3$ has zero curvature at $x = 0$, but under our definition this is a point of strict curvature nonetheless. Furthermore, Definition 2.1 also applies to points where ∂C is nondifferentiable and principal curvatures are not defined. For example, vertices of polytopes are points of strict curvature, while points on edges (1-faces) are not. Definition 2.1 also differs from the related concept of "direction of curvature" used by Howard–Newman (see, e.g., Howard [18], page 139), in that their notion requires a lower bound on the order of curvature, while ours does not. To extend our convergence results of Sections 4 and 5 in developing quantitative bounds on the convergence rate, the notion of direction of curvature would have to be used instead of our weaker notion of point of strict curvature.

The normal cone of C at $\vec{x} \in C$ is defined as $\mathrm{N}_{\vec{x}}\, C = \{\vec{s} \in \mathbb{R}^n : \langle \vec{s}, \vec{x} - \vec{w} \rangle \geq 0, \forall \vec{w} \in C\}$, or equivalently,

$$\begin{aligned} \mathrm{N}_{\vec{x}}\, C &= \left\{ \vec{s} \in \mathbb{R}^n : \vec{x} = \arg\max_{\vec{w} \in C} \langle \vec{s}, \vec{w} \rangle \right\} \\ &= \left\{ \tau \vec{s} : \tau \geq 0, \vec{s} \in \mathrm{S}^{n-1}, \vec{x} = \arg\max_{\vec{w} \in C} \langle \vec{s}, \vec{w} \rangle \right\}. \end{aligned} \tag{2.9}$$

By the dual description of C, it is the case that

$$\mathrm{N}_{\vec{x}}\, C \cap \mathrm{S}^{n-1} \cap \mathrm{span}(C) \neq \varnothing \tag{2.10}$$

if and only if $\vec{x} \in \partial C$, see also Lemma 2.3(c).

The following proposition provides further insight into the notion of points of strict curvature. The proof is given in Supplement A.

Proposition 2.1. *For any $C \subset \mathbb{R}^n$ nonempty, convex and compact, the following hold true*:

(a) *$\vec{x} \in C_{\mathrm{SE}}$ if and only if there exists $\vec{s} \in \mathrm{N}_{\vec{x}}\, C \cap \mathrm{S}^{n-1}$ and sequences $(\delta_k)_{k \in \mathbb{N}}$ and $(\epsilon_k)_{k \in \mathbb{N}} \subset \mathbb{R}_+$ such that $\epsilon_k, \delta_k \to 0$ as k tends to infinity, and such that*

$$\left\{ \vec{y} \in C : \mathrm{N}_{\vec{y}}\, C \cap \mathrm{S}^{n-1} \cap \mathrm{B}_{\delta_k}(\vec{s}) \neq \varnothing \right\} \subset \mathrm{B}_{\epsilon_k}(\vec{x}) \quad \forall k \in \mathbb{N}.$$

(b) $\{\vec{x} \in \partial C : \mathrm{N}_{\vec{x}}\, C \cap \mathrm{N}_{\vec{v}}\, C = \mathrm{span}(C)^{\perp}, \forall \vec{v} \in C \setminus \{\vec{x}\}\} \subset C_{\mathrm{SE}}$.

A point $\vec{x} \in C$ is an extreme point of C if it cannot be written as a convex combination of two points $\vec{y}, \vec{z} \in C \setminus \{\vec{x}\}$, see, for example, Rockafellar [29] or Borwein and Lewis [5]. We denote the set of extreme points of C by C_E.

The following lemma, whose proof is given in Supplement A and which will be used in the proof of Theorem 2.2, shows that points of strict curvature form a dense subset in the set of extreme points. In fact, most of the technical difficulties of this section deal with extending properties of C_E to C_{SE}.

Lemma 2.1. *For any $C \subset \mathbb{R}^n$ nonempty convex compact, it is true that*

$$C_{\mathrm{SE}} \subseteq C_E \subseteq \mathrm{cl}(C_{\mathrm{SE}}).$$

The next lemma forms a key technical tool in the proofs of Theorems 4.1 and 5.1, two of the main theorems of this paper, as well as of Proposition 2.4. The proof is again deferred to Supplement A. The result will be used to establish that for $\vec{x}_0 \in \mathrm{SET}$ a point of strict curvature, SET can be approximated via finitely many inequalities from the dual description of SET in such a way that the only points from the approximating set that lie outside the tangent plane of SET at $\vec{x}_0$ are localized near $\vec{x}_0$.

Lemma 2.2. *Let $C \subset \mathbb{R}^n$ be nonempty, convex and compact, and let $\vec{x}_0 \in C_{\mathrm{SE}}$ and $\vec{s}_0 \in \mathrm{N}_{\vec{x}_0}\, C \cap \mathrm{S}^{n-1}$ be chosen such that $\vec{x}_0$ is the unique maximizer of* $\max_{\vec{y} \in C} \langle \vec{s}_0, \vec{y} \rangle$. *Let $\epsilon > 0$ be given. Then*

there exist finitely many points $\vec{x}_i \in C$ *and normal vectors* $\vec{s}_i \in \mathrm{N}_{\vec{x}_i} C \cap \mathrm{S}^{n-1}$, $(i = 1, \ldots, k)$, *such that*

$$C(\xi_0, \ldots, \xi_k) := \left\{\vec{x} \in \mathbb{R}^n : \langle \vec{s}_0, \vec{x} - \vec{x}_0 \rangle \geq \xi_0\right\} \cap \bigcap_{i=1}^{k} \left\{\vec{x} \in \mathbb{R}^n : \langle \vec{s}_i, \vec{x} - \vec{x}_i \rangle \leq \xi_i\right\}$$

is compact for all $(\xi_0, \ldots, \xi_k) \in \mathbb{R}^{k+1}$, *and* $C(0, \ldots, 0) \subset \mathrm{B}_\varepsilon(\vec{x}_0)$.

The next result is a template for Theorem 5.1 used in Section 1.3.2, presenting the geometric ideas without the complications caused by large deviations. See Supplement A for a proof.

Proposition 2.2. *Let* $C \subset \mathbb{R}^n$ *be nonempty, convex and compact, and let* $C^1, C^2, \ldots$ *be a sequence of compact subsets of* $\mathbb{R}^n$ *such that* $d(C^n, C) \to 0$. *Let* $\vec{s} \in \mathrm{S}^{n-1}$ *be such that the optimization problem*

$$\vec{x}_* = \arg\max_{\vec{x} \in C} \langle \vec{s}, \vec{x} \rangle \tag{2.11}$$

has a unique solution. And finally, for all $n \in \mathbb{N}$ *let* $\vec{x}_n$ *be a solution of*

$$\vec{x}_n = \arg\max_{\vec{x} \in C^n} \langle \vec{s}, \vec{x} \rangle.$$

Then $\vec{x}_n \to \vec{x}_*$ *as n tends to infinity.*

Next, we shall investigate the approximability of compact convex sets by polyhedra and polytopes. Results on outer approximations by polyhedra and algorithms to achieve this in practice are widespread in the literature on the cutting plane approach in numerical optimization, see, for example, Bertsekas and Yu [4]. Similar results for inner approximations by polytopes play a key role in Markov chain Monte Carlo methods for the estimation of the volume of high dimensional convex bodies, see, for example, Jerrum [19]. The literature in both areas is focused on algorithms and relies on separation or membership oracles. As a result, the constructions use outer approximations by cutting planes that do not necessarily touch the boundary of the convex body to be approximated, and likewise, inner approximations use generators that generally do not lie on the boundary either.

In contrast, the approximations required by our analysis have a crucial interplay with the boundary. For outer approximations, we would like cutting hyperplanes to be supported at points of strict curvature. Likewise, we would like inner approximations to be generated as the convex hull of points of strict curvature. Since we are not aware of the required results appearing in the literature, we derive them from first principles.

The following result is key in the proofs of Theorem 4.1, Lemma 2.4, and of Propositions 2.3 and 2.4. See Supplement A for a proof.

Lemma 2.3. *Let* $C \subset \mathbb{R}^n$ *be a set of the form*

$$C = \bigcap_{\vec{s} \in \mathrm{S}^{n-1}} H_{\vec{s}}, \tag{2.12}$$

where $H_{\vec{s}} = \{\vec{x} : \langle \vec{s}, \vec{x}\rangle \le \lambda(\vec{s})\}$ *for some continuous function* $\vec{s} \mapsto \lambda(\vec{s}) \in \mathbb{R}$. *Then the following hold true*:

(a) C *is convex and compact*.
(b) *For any given* $\epsilon > 0$, *there exists a finite collection of points* $\vec{s}_1, \dots, \vec{s}_k \in \mathrm{S}^{n-1}$ *for which*

$$\max_{\vec{x} \in \bigcap_{i=1}^{k} H(\vec{s}_i)} d(\vec{x}, C) \le \epsilon. \tag{2.13}$$

(c) *For every point* $\vec{x} \in \partial C$, *there exists* $s \in \mathrm{S}^{n-1}$ *such that*

$$\langle \vec{s}, \vec{x}\rangle = \max_{\vec{y} \in C} \langle \vec{s}, \vec{y}\rangle = \lambda(\vec{s}).$$

The following is a strengthening of Lemma 2.3(b) of independent interest. The proof is deferred to Supplement A.

Proposition 2.3. *Let* C *be as in Lemma* 2.3 *and nonempty. Then the points* $\vec{s}_i$ *in part* (*b*) *of Lemma* 2.3 *can be chosen so that* $\vec{x}_i = \arg\max_{\vec{y} \in C} \langle \vec{s}_i, \vec{y}\rangle$ *is unique for all* i, *that is*, $\vec{x}_i$ *are points of strict curvature*.

The following result is required for the purposes of the proofs of Theorem 4.1 and of Proposition 2.4.

Theorem 2.2. *Let* $C \subset \mathbb{R}^n$ *be nonempty compact convex. The for all* $\epsilon > 0$ *there exist finitely many points of strict curvature* $\vec{x}_1, \dots, \vec{x}_k \in C_{\mathrm{SE}}$ *such that*

$$\max_{\vec{x} \in C} d\big(\vec{x}, \mathrm{conv}(\vec{x}_1, \dots, \vec{x}_k)\big) \le \epsilon.$$

Proof. Let $\{\vec{x}_1, \dots, \vec{x}_k\} \subset C_{\mathrm{SE}}$ be an ϵ-net on the set C_E of extreme points of C, that is, $\vec{x}_i$ are chosen so that

$$\min_i \|\vec{z} - \vec{x}_i\| \le \epsilon$$

for all $\vec{z} \in C_E$. The existence of such an ϵ-net is established as follows: C being compact, $\mathrm{cl}(C_{\mathrm{SE}})$ is a compact set too, and by the Heine–Borel theorem, we can extract a finite covering by Euclidean balls of radius $\epsilon/2$ around points $\vec{y}_i \in \mathrm{cl}(C_{\mathrm{SE}})$, which by Lemma 2.1 is also a covering of C_E,

$$\bigcup_{i=1}^{k} \mathrm{B}_{\epsilon/2}(\vec{y}_i) \supset C_E.$$

Next, for all i choose $\vec{x}_i \in C_{\mathrm{SE}}$ within distance $\epsilon/2$ of $\vec{y}_i$. It then follows from the triangular inequality that $\{\vec{x}_1, \dots, \vec{x}_k\}$ is the required ϵ-net.

By the theorems of Minkowski (see Minkowski [26] and Gruber [11]: this theorem says that a convex compact set in $\mathbb{R}^n$ is equal to the convex hull of its extreme points; the generalization

to arbitrary topological vector spaces is the Krein–Milman theorem (Krein and Milman [22], Milman [25])) and Carathéodory (see Carathéodory [7]: this result says that if $K = \text{conv}(X)$ for some $X \subset \mathbb{R}^n$, then every point in K is a convex combination of at most $n+1$ elements of X, see also Gruber [11]), any point $\vec{x} \in C$ can be written as a convex combination $\vec{x} = \xi_1 \vec{z}_1 + \cdots + \xi_m \vec{z}_m$ of $m \leq n+1$ extreme points $z_j \in C_E$, and by construction of the ϵ-net, it is then possible to choose $1 \leq i_j \leq k$ such that $\|\vec{z}_j - \vec{x}_{i_j}\| \leq \epsilon$ for all j. Using the triangular inequality once again, we find that

$$\begin{aligned} d\big(\vec{x}, \text{conv}(\vec{x}_1, \ldots, \vec{x}_k)\big) &\leq d\big(\vec{x}, \text{conv}(\vec{x}_{i_1}, \ldots, \vec{x}_{i_m})\big) \\ &\leq d(\xi_1 \vec{z}_1 + \cdots + \xi_m \vec{z}_m, \xi_1 \vec{x}_{i_1} + \cdots + \xi_m \vec{x}_{i_m}) \leq \epsilon, \end{aligned}$$

as claimed. □

The next result, proven in Supplement A, can be seen as an template for Theorem 4.1, exhibiting the main argument without the complications introduced by large deviations.

Proposition 2.4. *Let C be a nonempty convex compact subset of $\mathbb{R}^n$ with dual description (2.12), and let $C^1, C^2, \ldots$ be convex compact subsets of $\mathbb{R}^n$ such that for all linear functionals $f : \mathbb{R}^n \to \mathbb{R}$ it is true that $\max_{\vec{p} \in C^n} f(\vec{p}) \overset{n\to\infty}{\longrightarrow} \max_{\vec{p} \in C} f(\vec{p})$. Then $d(C^n, C) \to 0$.*

The final result shows among other things that the above developed theory is applicable to $C = \text{SET}$, defined in (1.13). This lemma is used in the point-convergence proofs of Theorems 5.1 and 5.2. We once again defer the proof to Supplement A.

Lemma 2.4.

(a) *The function $S \mapsto \lambda(S)$ is continuous.*
(b) SET *is nonempty, convex and compact.*
(c) *For every $\vec{x} \in \partial\text{SET}$ there exists $S_{\vec{x}} \neq 0$ such that $\{\vec{y} : f_{S_{\vec{x}}}(\vec{y}) = \lambda(S_{\vec{x}})\}$ is a tangent plane to* SET *supported at $\vec{x}$.*
(d) $\max_{\vec{x} \in \text{SET}} f_S(\vec{x}) = \lambda(S)$ *holds true for all scoring functions S.*

3. Large deviation tools

Recall the quantities $L_S(x, y)$, $L_n(S)$ and $\lambda_n(S)$ introduced in Section 1.2. From (1.9), we know that that $\lambda_n(S) \to \lambda(S)$. In this section, we will show a stronger result that quantifies the convergence rate as being of order $\mathcal{O}(\sqrt{\ln n / n})$. For this purpose, we introduce the following notation,

$$\|S\|_* = \Big(\max_{(\mathfrak{d},\mathfrak{c}),(\mathfrak{e},\mathfrak{c}) \in \mathcal{A}^{*2}} \big|S(\mathfrak{d},\mathfrak{c}) - S(\mathfrak{e},\mathfrak{c})\big|, \max_{(\mathfrak{c},\mathfrak{d}) \in \mathcal{A}^{*2}} \big|S(\mathfrak{c},\mathfrak{d})\big| \Big).$$

The following two lemmas are required for the purposes of the proof of Theorem 3.1. The first result is proven in Supplement A.

Lemma 3.1. *Let $x = x_1, \ldots, x_m$ and $y = y_1, \ldots, y_n$ be two finite sequences with letters from $\mathcal{A}$, and let S be a scoring function. Let $\mathfrak{x} \in \mathcal{A}$, and consider two amendments of the sequence x, $x^{[i]} = x_1, \ldots, x_{i-1}, \mathfrak{x}, x_{i+1}, \ldots, x_m$, obtained by replacing an arbitrary letter x_i by $\mathfrak{x}$, and $x^{[+]} = x_1, \ldots, x_m, \mathfrak{x}$, obtained by extending x by a letter $\mathfrak{x}$. Then the following hold true,*

$$\left|L_S\left(x^{[i]}, y\right) - L_S(x, y)\right| \leq \|S\|_*, \tag{3.1}$$

$$\left|L_S\left(x^{[+]}, y\right) - L_S(x, y)\right| \leq \|S\|_*. \tag{3.2}$$

Lemma 3.2 (Amsalu, Hauser and Matzinger [2]). *The convergence of $\lambda_n(S)$ to $\lambda(S)$ is governed by the inequality*

$$\lambda_n(S) \leq \lambda(S) \leq \lambda_n(S) + c_n \|S\|_* \frac{\sqrt{\ln n}}{\sqrt{n}} + \frac{2\|S\|_*}{n} \quad \forall n \in \mathbb{N}, \tag{3.3}$$

where

$$c_n := \sqrt{\frac{2\ln 3 + 2\ln(n+2)}{\ln n}}.$$

Note that c_n tends to $\sqrt{2}$ when $n \to \infty$, so that it effectively acts as a constant.
The following result will be required to prove Theorems 3.1 and 6.2.

Lemma 3.3 (McDiarmid's Inequality (McDiarmid [24])). *Let $Z_1, Z_1, \ldots, Z_m$ be i.i.d. random variables that take values in a set D, and let $g : D^m \to \mathbb{R}$ be a function of m variables with the property that*

$$\max_{i=1,\ldots,m} \sup_{z \in D^m, \hat{z}_i \in D} \left|g(z_1, \ldots, z_m) - g(z_1, \ldots, \hat{z}_i, \ldots, z_m)\right| \leq C.$$

Thus, changing a single argument of g changes its image by less than a constant C. Then the following bounds hold,

$$\mathrm{P}\left[g(Z_1, \ldots, Z_m) - \mathrm{E}\left[g(Z_1, \ldots, Z_m)\right] \geq \epsilon \times m\right] \leq \exp\left\{-\frac{2\epsilon^2 m}{C^2}\right\},$$

$$\mathrm{P}\left[\mathrm{E}\left[g(Z_1, \ldots, Z_m)\right] - g(Z_1, \ldots, Z_m) \geq \epsilon \times m\right] \leq \exp\left\{-\frac{2\epsilon^2 m}{C^2}\right\}.$$

The following constitutes a key tool for the proofs of Theorems 4.1 and 5.1.

Theorem 3.1. *For fixed $\epsilon > 0$ and scoring function S there exists $n_\epsilon \in \mathbb{N}$ such that*

$$\mathrm{P}\left[\frac{L_n(S)}{n} \geq \lambda(S) + \epsilon\right] \leq \exp\left\{-\frac{\epsilon^2 n}{4\|S\|_*^2}\right\} \quad \forall n \in \mathbb{N}, \tag{3.4}$$

$$\mathrm{P}\left[\frac{L_n(S)}{n} \leq \lambda_n(S) - \epsilon\right] \leq \exp\left\{-\frac{\epsilon^2 n}{4\|S\|_*^2}\right\} \quad \forall n \in \mathbb{N}, \tag{3.5}$$

$$\mathrm{P}\left[\frac{L_n(S)}{n} \leq \lambda(S) - \epsilon\right] \leq \exp\left\{-\frac{\epsilon^2 n}{4\|S\|_*^2}\right\} \quad \forall n \geq n_\epsilon. \tag{3.6}$$

Proof. We know from Lemma 3.1 that

$$g(X_1, \ldots, X_n, Y_1, \ldots, Y_n) = S(X_1, \ldots, X_n, Y_1, \ldots, Y_n) = L_n(S)$$

satisfies the assumptions of Lemma 3.3 with $m = 2n$ and $C = \|S\|_*$. McDiarmid's Inequality therefore shows

$$\begin{aligned}\mathrm{P}\left[\frac{L_n(S)}{n} \geq \lambda_n(S) + \epsilon\right] &= \mathrm{P}\left[L_n(S) \geq \mathrm{E}\big[L_n(S)\big] + \frac{\epsilon}{2} \times 2n\right] \\ &\leq \exp\left\{-\frac{\epsilon^2}{\|S\|_*^2} \times n\right\},\end{aligned} \tag{3.7}$$

and similarly,

$$\mathrm{P}\left[\frac{L_n(S)}{n} \leq \lambda_n(S) - \epsilon\right] \leq \exp\left\{-\frac{\epsilon^2}{\|S\|_*^2} \times n\right\}. \tag{3.8}$$

Claim (3.5) therefore holds.

Furthermore, Lemma 3.2 established that

$$\lambda_n(S) \leq \lambda(S) \leq \lambda_n(S) + c_n\|S\|_*\frac{\sqrt{\ln n}}{\sqrt{n}} + \frac{2\|S\|_*}{n} \quad \forall n \in \mathbb{N}, \tag{3.9}$$

holds, where $c_n = \sqrt{2\ln 3 + 2\ln(n+2)}/\sqrt{\ln n}$. Using the first inequality from (3.9) in conjunction with (3.7), we find

$$\mathrm{P}\left[\frac{L_n(S)}{n} \geq \lambda(S) + \epsilon\right] \leq \mathrm{P}\left[\frac{L_n(S)}{n} \geq \lambda_n(S) + \epsilon\right] \leq \exp\left\{-\frac{\epsilon^2}{\|S\|_*^2} \times n\right\},$$

which shows Claim (3.4).

Using now the second inequality from (3.9) in conjunction with (3.8), we find

$$\begin{aligned}&\mathrm{P}\left[\frac{L_n(S)}{n} \leq \lambda(S) - \epsilon\right] \\ &\leq \mathrm{P}\left[\frac{L_n(S)}{n} \leq \lambda_n(S) - \left(\epsilon - c_n\|S\|_*\frac{\sqrt{\ln n}}{\sqrt{n}} - \frac{2\|S\|_*}{n}\right)\right] \\ &\leq \exp\left\{-\frac{(\epsilon - c_n\|S\|_*\frac{\sqrt{\ln n}}{\sqrt{n}} - \frac{2\|S\|_*}{n})^2}{\|S\|_*^2} \times n\right\} \\ &\leq \exp\left\{-\frac{\epsilon^2}{4\|S\|_*^2} \times n\right\} \quad \forall n \geq n_\epsilon,\end{aligned}$$

where $n_\epsilon \in \mathbb{N}$ is chosen large enough to satisfy

$$\epsilon - c_n \|S\|_* \frac{\sqrt{\ln n}}{\sqrt{n}} - \frac{2\|S\|_*}{n} > \frac{\epsilon}{2} \quad \forall n \geq n_\epsilon.$$

This establishes Claim (3.6). □

4. Set convergence of empirical distributions

Equipped with the tools of Sections 2 and 3, we are ready to prove the first main theorem of this paper.

Theorem 4.1. *Let* SET *and* SET^n *be as defined in* (1.13) *and* (1.3). *Then*

$$\mathrm{P}\big[d\big(\text{SET}^n, \text{SET}\big) \overset{n\to\infty}{\longrightarrow} 0\big] = 1, \tag{4.1}$$

where d is the Hausdorff distance.

Proof. By the definition of d in (1.14), we need to prove the two identities

$$\mathrm{P}\Big[\max_{\vec{x}\in\text{SET}^n} d(\vec{x}, \text{SET}) \overset{n\to\infty}{\longrightarrow} 0\Big] = 1, \tag{4.2}$$

$$\mathrm{P}\Big[\max_{\vec{x}\in\text{SET}} d\big(\vec{x}, \text{SET}^n\big) \overset{n\to\infty}{\longrightarrow} 0\Big] = 1. \tag{4.3}$$

To prove Equation (4.2), we use Lemma 2.3 which establishes that, given $\epsilon > 0$, there exist finitely many scoring functions $S_1, \ldots, S_k$ such that

$$\max_{\vec{x}\in\bigcap_{i=1}^k H(S_i)} d(\vec{x}, \text{SET}) \leq \epsilon,$$

where the half-spaces

$$H(S_i) := \big\{\vec{x} \in \mathbb{R}^{|\mathcal{A}^*|^2} : f_{S_i}(\vec{x}) \leq \lambda(S_i)\big\},$$

are defined as in Equation (1.12). Let $H_n^+(S_i)$ denote the shifted half-space

$$H_n^+(S_i) = \left\{\vec{x} : f_{S_i}(\vec{x}) \leq \lambda_n(S_i) + \frac{\ln n}{\sqrt{n}}\right\},$$

and let us define the event $\mathscr{A}_n(S_i) = \{\omega \in \Omega : \text{SET}^n(\omega) \subseteq H_n^+(S_i)\}$, where Ω is the probability space over which the random sequences X and Y are defined.

Since the extreme points of SET^n are determined by alignments with gaps of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, the event $\mathscr{A}_n(S_i)^c$ occurs exactly when there exists an alignment with gaps π such that

$$f_{S_i}\big(\vec{p}_\pi\big((X_1, \ldots, X_n), (Y_1, \ldots, Y_n)\big)\big) > \lambda_n(S_i) + \frac{\ln n}{\sqrt{n}},$$

and this in turn occurs if and only if

$$\frac{L_n(S_i)}{n} > \lambda_n(S_i) + \frac{\ln n}{\sqrt{n}}.$$

We thus have

$$\mathrm{P}\big[\mathscr{A}_n(S_i)^c\big] = \mathrm{P}\left[\frac{L_n(S_i)}{n} > \lambda_n(S_i) + \frac{\ln n}{\sqrt{n}}\right] \leq \exp\left\{-\frac{(\ln n)^2}{\|S_i\|_*^2}\right\} = n^{-c_{S_i}\ln n},$$

where the inequality follows by setting $\epsilon = (\ln n)/\sqrt{n}$ in Inequality (3.7) of the proof of Theorem 3.1, and where $c_{S_i} = \|S_i\|_*^{-2} > 0$ does not depend on n, since $\|S_i\|_*$ is as defined in Section 3. It follows that

$$\mathrm{P}\left[\mathrm{SET}^n \subset \bigcap_{i=1}^{k} H_n^+(S_i)\right] \geq 1 - \sum_{i=1}^{k} n^{-c_{S_i}\ln n} \geq 1 - n^{-c\ln n},$$

where $c > 0$ is a constant independent of n. The series $\sum_n n^{-c\ln n}$ being convergent, the Borel–Cantelli lemma implies that almost surely there exists $n_0 \in \mathbb{N}$ such that

$$\mathrm{SET}^n \subset \bigcap_{i=1}^{k} H_n^+(S_i) \quad \forall n \geq n_0.$$

By the definition of the S_i, this implies that for all $n \geq n_0$, we have

$$\max_{\vec{x}\in\mathrm{SET}^n} d(\vec{x}, \mathrm{SET}) \leq C \times \frac{\ln n}{\sqrt{n}} + \epsilon, \tag{4.4}$$

where $C > 0$ is a constant independent of n. This implies that

$$\mathrm{P}\Big[\limsup_{n\to\infty} \max_{\vec{x}\in\mathrm{SET}^n} d(\vec{x}, \mathrm{SET}) \leq \epsilon\Big] = 1 \quad \forall \epsilon > 0.$$

Finally, since this is true for all ϵ rational, Equation (4.2) follows.

To prove Equation (4.3), we employ Theorem 2.2, which establishes that for any given $\epsilon > 0$, there exist points $\vec{x}_1, \ldots, \vec{x}_k \in \mathrm{SET}_{\mathrm{SE}}$, chosen such that

$$d\big(\vec{x}, \mathrm{conv}(\vec{x}_1, \ldots, \vec{x}_k)\big) \leq \epsilon \quad \forall \vec{x} \in \mathrm{SET}.$$

We now claim that for each $\vec{x}_i$ there almost surely exists a sequence of points $\vec{x}_{1,i}, \vec{x}_{2,i}, \vec{x}_{3,i}, \ldots$ such that $\vec{x}_{j,i} \in \mathrm{SET}^j$ for all $j \in \mathbb{N}$ and $\limsup_{j\to\infty} \|\vec{x}_{j,i} - \vec{x}_i\| \leq \epsilon$. By the triangular inequality, our claim implies that almost surely it is the case that

$$\limsup_{n\to\infty} \max_{\vec{x}\in\mathrm{SET}} d\big(\vec{x}, \mathrm{SET}^n\big) \leq 2\epsilon \quad \forall \epsilon > 0,$$

and since this is true for all ϵ rational, Equation (4.3) follows.

It remains to prove our claim. Lemma 2.2 shows that there exist scoring functions $S_{\vec{x}_i}$ and $S_1, \ldots, S_\ell$ such that

$$\vec{x}_i \in C_i := \left\{\vec{x} : f_{S_{\vec{x}_i}}(\vec{x}) \geq \lambda(S_{\vec{x}_i}),\, f_{S_j}(\vec{x}) \leq \lambda(S_j),\, (j = 1, \ldots, \ell)\right\} \subset \mathrm{B}_\epsilon(\vec{x}_i)$$

and such that the sets

$$C_{n,i} := \left\{\vec{x} : f_{S_{\vec{x}_i}}(\vec{x}) \geq \lambda_n(S_{\vec{x}_i}) - \frac{\ln n}{\sqrt{n}},\, f_{S_j}(\vec{x}) \leq \lambda(S_j) + \frac{\ln n}{\sqrt{n}},\, (j = 1, \ldots, \ell)\right\}$$

are compact for all $n \in \mathbb{N}$.

Further, by (1.8), the sets $C_{n,i}$ are nested in the following sense: for any finite set $\{n_1, \ldots, n_k\} \subset \mathbb{N}$ there exist infinitely many integers $n \in \mathbb{N}$ such that $C_{n,i} \subseteq C_{n_j,i}$, $(j = 1, \ldots, k)$. By compactness and (1.9), we then have

$$\limsup_{n\to\infty} d(\vec{x}_i, C_{n,i}) \leq \epsilon. \tag{4.5}$$

We will now show that with high probability $C_{n,i}$ has a nonempty intersection with SET^n. Consider the events

$$\mathscr{B}_{n,j} := \left\{\omega \in \Omega : \frac{L_n(S_j)}{n} \leq \lambda(S_j) + \frac{\ln n}{\sqrt{n}}\right\},$$

$$\mathscr{C}_{n,i} := \left\{\omega \in \Omega : \exists \vec{x} \in \mathrm{SET}^n \text{ s.t. } f_{S_{\vec{x}_i}}(\vec{x}) \geq \lambda_n(S_{\vec{x}_i}) - \frac{\ln n}{\sqrt{n}}\right\}.$$

By Theorem 3.1, we have

$$\mathrm{P}\big[\mathscr{B}^c_{n,j}\big] \leq n^{-K_j \ln n},$$

where $K_j > 0$ is a constant that does not depend on n. Note also that Equation (1.15) implies

$$\mathscr{B}_{n,j} = \left\{\omega \in \Omega : f_{S_j}(\vec{x}) \leq \lambda(S_j) + \frac{\ln n}{\sqrt{n}}, \forall \vec{x} \in \mathrm{SET}^n\right\}.$$

Theorem 3.1 further implies

$$\mathrm{P}\big[\mathscr{C}^c_{n,i}\big] \leq n^{-\ln n}.$$

But note that when the events $\mathscr{C}_{n,i}$ and $\mathscr{B}_{n,1}, \ldots, \mathscr{B}_{n,\ell}$ occur jointly, then $\mathrm{SET}^n \cap C_{n,i} \neq \varnothing$ holds. The probability that the intersection is empty is thus bounded from above by

$$P\big[\mathscr{C}^c_{n,i}\big] + \sum_{i=1}^{\ell} \mathrm{P}\big[\mathscr{B}^c_{n,i}\big] \leq (\ell + 1) n^{-K \ln n},$$

where $K > 0$ is a constant that does not depend on n.

In view of the fact that the series

$$\sum_{n=1}^{\infty}(\ell+1)n^{-K\ln n}$$

converges, the Borel–Cantelli lemma now implies that, almost surely, for all but a finite number of $n \in \mathbb{N}$ there exists $x_{n,i} \in \mathrm{SET}^n \cap C_{n,i}$. In the finitely many cases where $\mathrm{SET}^n \cap C_{n,i} = \varnothing$, we can pick an arbitrary point $x_{n,i} \in \mathrm{SET}^n$ to complete the sequence. In view of (4.5), we thus find that almost surely it is possible to construct a sequence $(x_{n,i})_{n\in\mathbb{N}}$ with the claimed properties. This settles the theorem. □

5. Point convergence

So far we established that the empirical distributions of optimal alignments of random sequences under any scoring function asymptotically lie in SET. We will now show that for a fixed, randomly chosen scoring function S, the empirical distributions of all optimal alignments of X and Y under S converge to a unique point in SET. Recalling the notation $\vec{p}_\pi(x, y)$ introduced in Section 1.2, we write

$$\mathrm{SET}^*(X, Y) = \left\{\vec{p}_\pi(X, Y) : \pi \text{ is an optimal alignment of } X \text{ and } Y\right\}$$

for the set of empirical distributions of optimal alignments of $X = X_1, \ldots, X_n$ and $Y = Y_1, \ldots, Y_n$. Consider the event

$$\mathscr{D}_n(\vec{p}, \epsilon) := \left\{\omega \in \Omega : \mathrm{SET}^*\big(X(\omega), Y(\omega)\big) \setminus \overline{\mathrm{B}}_\epsilon(\vec{p}) \neq \varnothing\right\}$$

that there exists an optimal alignment π of $x = X(\omega)$ and $y = Y(\omega)$ under the scoring function S such that $\|\vec{p}_\pi(x, y) - \vec{p}\| > \epsilon$.

Theorem 5.1. *Let S be a scoring function such that the hyperplane*

$$\left\{\vec{x} : f_S(\vec{x}) = \lambda(S)\right\} \tag{5.1}$$

intersects SET *in a unique point $\vec{p}_S$, and let $\epsilon > 0$ be given. Then there exists a constant K_ϵ such that for all $n \in \mathbb{N}$ it is true that*

$$\mathrm{P}\big[\mathscr{D}_n(\vec{p}_S, \epsilon)\big] \leq e^{-K_\epsilon n}.$$

Furthermore, $\mathrm{SET}^*(X, Y) \to \{\vec{p}_S\}$ *almost surely as n tends to infinity.*

Proof. By Lemma 2.4, SET is a compact convex set with nonempty intersection with the hyperplane (5.1), and by (1.16) all such intersection points are maximizers of the optimization problem $\max_{\vec{y}\in\mathrm{SET}}\langle\vec{s}, \vec{y}\rangle$, where $\vec{s}$ is the normalization of the vector representation of the linear functional f_S defined by the scoring function. It follows that $\vec{p}_S$ satisfies Definition 2.1 of a point of strict

curvature of SET. Lemma 2.2 therefore implies that there exist finitely many scoring functions $S_1, S_2, \ldots, S_k$ and thresholds $\epsilon_0, \ldots, \epsilon_k > 0$ such that

$$\big\{\vec{x} : f_S(\vec{x}) \geq \lambda(S) - \epsilon_0\big\} \cap \bigcap_{i=1}^{k} \big\{\vec{x} : f_{S_i}(\vec{x}) \leq \lambda(S_i) + \epsilon_i\big\} \subset \mathrm{B}_\epsilon(\vec{p}_S). \tag{5.2}$$

Consider now the events

$$\mathscr{E}_{n,i} := \big\{\omega \in \Omega : \mathrm{SET}^n \subset \big\{\vec{x} : f_{S_i}(\vec{x}) \leq \lambda(S_i) + \epsilon_i\big\}\big\}.$$

By (1.15) this is equivalent to requiring that the rescaled optimal alignment score $L_n(S_i)/n$ satisfy $L_n(S_i)/n \leq \lambda(S_i) + \epsilon_i$. By Theorem 3.1, there exists $K_i > 0$ such that

$$\mathrm{P}[\mathscr{E}_{n,i}] \geq 1 - \mathrm{e}^{-K_i n} \quad \forall n. \tag{5.3}$$

Let us further define the event

$$\mathscr{E}_{n,0} := \big\{\omega \in \Omega : \mathrm{SET}^n \cap \big\{\vec{x} : f_{S_i}(\vec{x}) \geq \lambda(S) - \epsilon_0\big\} \neq \varnothing\big\},$$

which is the same as requiring that $L_n(S)/n$ exceed the value $\lambda(S) - \epsilon_0$. Corollary 3.1 once again shows that there exists $K_0 > 0$ such that

$$\mathrm{P}[\mathscr{E}_{n,0}] \geq 1 - \mathrm{e}^{-K_0 n} \quad \forall n. \tag{5.4}$$

Combining all of the above, we now find $\mathscr{D}_n^c \subseteq \bigcup_{i=0}^{k} \mathscr{E}_{n,i}$, so that

$$\mathrm{P}[\mathscr{D}_n] \leq \sum_{i=0}^{k} \mathrm{P}\big[\mathscr{E}_{n,i}^c\big] \leq \sum_{i=0}^{k} \mathrm{e}^{-K_i n} \leq \mathrm{e}^{-K_\epsilon n}$$

for some constant $K_\epsilon > 0$, as claimed.

The last statement follows from the Borel–Cantelli lemma in a similar construction as in the proof of Theorem 4.1. □

The above theorem shows that if $\vec{p}_S$ is the only solution to $f_S(\vec{p}_S) = \lambda_S$, then almost surely it is the case that for all sequences $(\pi_n)_{\mathbb{N}}$, constructed by choosing an optimal alignment π_n of $X = X_1, \ldots, X_n$ and $Y = Y_1, \ldots, Y_n$ for each $n \in \mathbb{N}$, it is true that $\vec{p}_{\pi_n}(X, Y) \to \vec{p}_S$. Note that the convergence rate was not specified. However, our convergence argument, which is based on the Azuma–Hoeffding Inequality, could be made quantitative if a bound on the curvature of SET at $\vec{p}_S$ were known.

Our second and main result of this section shows that the above theorem applies generically.

Theorem 5.2. *For Lebesgue-almost every scoring function S, the following is true: Almost surely it is the case that for all sequences $(\pi_n)_{\mathbb{N}}$, constructed by choosing an optimal alignment π_n of $X = X_1, \ldots, X_n$ and $Y = Y_1, \ldots, Y_n$ for each $n \in \mathbb{N}$, we have $\vec{p}_{\pi_n}(X, Y) \to \vec{p}_S$, where $\vec{p}_S$ is a unique empirical distribution that only depends on S.*

Proof. Consider a scoring function S and denote the normalization of the vector representation of the linear functional f_S by $\vec{S}$. For Lebesgue-almost every S, $\vec{S}$ corresponds to a point where the optimization problem of Theorem 2.1 has a unique optimizer. Together with Lemma 2.4, this implies that the condition that the hyperplane $\{\vec{x} : f_S(\vec{x}) = \lambda(S)\}$ intersect SET in a unique point $\vec{p}_S$ is satisfied, and hence Theorem 5.1 applies, for almost every S. □

6. Fluctuation of the optimal alignment score

Let $X = X_1, \ldots, X_n$ and $Y = Y_1, \ldots, Y_n$ be the random sequences introduced earlier, and let $\mathfrak{a}$ and $\mathfrak{b}$ be two distinct letters from the alphabet $\mathcal{A}$. We define a new random sequence $\tilde{X} = \tilde{X}_1, \ldots, \tilde{X}_n$ via the following compound procedure:

1. Sample a realization $x = x_1, \ldots, x_n$ of X.
2. If $\mathcal{J} := \{i : x_i = \mathfrak{a}\} \neq \varnothing$,
   (a) let J be a random index defined on some probability space $(\tilde{\Omega}, \tilde{\mathrm{P}})$ and taking values with uniform distribution on $\mathcal{J}$,
   (b) select a sample $j = J(\tilde{\omega})$,
   (c) set $\tilde{x}_j = \mathfrak{b}$ and $\tilde{x}_i = x_i$ for $i \neq j$.
3. Else, set $\tilde{x} = x$.

Note that the distribution of $\tilde{X}$ generally differs from the distribution of X, and that, while $X = X_1, \ldots, X_n$ consists of the first n letters of an infinite random sequence $(X_i)_{i\in\mathbb{N}}$, the same is not true for $\tilde{X}$: we only ever sample (at most) one entry of X realized in the form of an $\mathfrak{a}$, independently of n, so that the probability of any given index to be chosen diminishes as n grows.

The following result was proven in Lember and Matzinger [23], where we use the notation

$$\tilde{L}_n(S) := \max_{\pi} S_\pi(\tilde{X}, Y),$$

in analogy to the earlier introduced random variable $L_n(S) = \max_\pi S_\pi(X, Y)$, and where we write $f(n) = \Theta(n)$ if there exist constants $0 < c_1 < c_2$ such that $c_1 n \leq f(n) \leq c_2 n$ for all $n \in \mathbb{N}$.

Theorem 6.1 (Lember and Matzinger [23]). *Let the scoring function S and the distribution of X and Y be chosen so that there exist parameters $\beta, \varepsilon > 0$ for which*

$$\mathrm{P}\big[\mathrm{E}_{\tilde{\mathrm{P}}}\big[\tilde{L}_n(S) - L_n(S) \| X, Y\big] \geq \varepsilon\big] \geq 1 - \mathrm{e}^{-\beta n} \quad \forall n \in \mathbb{N}.$$

Then the variance of the optimal alignment score is of order

$$\mathrm{VAR}\big[L_n(S)\big] = \Theta(n).$$

Up until now, the criterion of Theorem 6.1 could only be verified in a few special cases, notably for random sequences whose letters have highly asymmetrical distribution on $\mathcal{A}$. We will now reduce this criterion to the following two conditions:

C1 The scoring function S is chosen so that the hyperplance $\{\vec{x} : f_S(\vec{x}) = \lambda(S)\}$ intersects SET in a unique point $\vec{p}_S = (p_{\mathfrak{cd}})$.

C2 There exist $\mathfrak{a}, \mathfrak{b} \in \mathcal{A}$ for which it is the case that

$$\sum_{\mathfrak{c} \in \mathcal{A}^*} p_{\mathfrak{ac}}(S_{\mathfrak{bc}} - S_{\mathfrak{ac}}) > 0.$$

These two conditions are much easier to verify, as C1 holds generically by the argument of the proof of Theorem 5.2, and in many cases C2 can be established to high confidence by Monte Carlo simulation.

Theorem 6.2. *Let a scoring function S be chosen so that conditions C1 and C2 apply. Then the variance of the optimal alignment score is of order*

$$\mathrm{VAR}\big[L_n(S)\big] = \Theta(n).$$

Proof. Let $\mathcal{J} = \{i \in \{1, \ldots, n\} : X_i = \mathfrak{a}\}$, $q_{\mathfrak{a}} = \mathrm{P}[X_1 = \mathfrak{a}]$, and let us define the event

$$\mathscr{F}_n := \left\{\omega \in \Omega : \frac{n}{|\mathcal{J}|} \geq \frac{1}{2q_{\mathfrak{a}}}\right\}.$$

Since X has i.i.d. entries, McDiarmid's Inequality (Lemma 3.3) implies that for all $n \in \mathbb{N}$,

$$\mathrm{P}[\mathscr{F}_n] \geq 1 - \mathrm{e}^{-n\frac{q_{\mathfrak{a}}^2}{2}}. \tag{6.1}$$

Next, let

$$\varepsilon := \frac{1}{4q_{\mathfrak{a}}}\big\langle \vec{p}_S, (S_{\mathfrak{bc}} - S_{\mathfrak{ac}})_{\mathfrak{c}}\big\rangle := \frac{1}{4q_{\mathfrak{a}}} \sum_{\mathfrak{c} \in \mathcal{A}^*} p_{\mathfrak{ac}}(S_{\mathfrak{bc}} - S_{\mathfrak{ac}}),$$

where $q_{\mathfrak{a}} = \mathrm{P}[X_1 = \mathfrak{a}]$. Condition C2 shows that $\varepsilon > 0$, and by the continuity of inner products, there exists $\delta > 0$ so that for any $\vec{p} \in \mathrm{B}_\delta(\vec{p}_S)$, we have

$$\frac{1}{2q_{\mathfrak{a}}}\big\langle \vec{p}, (S_{\mathfrak{bc}} - S_{\mathfrak{ac}})_{\mathfrak{c}}\big\rangle \geq \varepsilon. \tag{6.2}$$

Recall now the notations $\mathrm{SET}^*(X, Y)$ and $\mathscr{D}_n(\vec{p}, \epsilon)$ introduced in Section 5. By virtue of Condition C1, Theorem 5.1 applies, which shows that there exists $K_\delta > 0$ such that the probability that all optimal alignments of X and Y have empirical distributions that lie within a distance δ of $\vec{p}_S$ equals

$$\mathrm{P}\big[\mathrm{SET}^*(X, Y) \subseteq \mathrm{B}_\delta(\vec{p}_S)\big] = \mathrm{P}\big[\mathscr{D}_n^c(\vec{p}_S, \delta)\big] \geq 1 - \mathrm{e}^{-K_\delta n} \quad \forall n \in \mathbb{N}. \tag{6.3}$$

But when $\mathscr{D}_n^c(\vec{p}_S, \delta)$ occurs, then for any optimal alignment π_n^* of X and Y, (6.2) holds with $\vec{p} = \vec{p}_{\pi_n^*}(X, Y)$. Denoting the components of $\vec{p}_{\pi_n^*}(X, Y)$ by $p^*_{\mathfrak{cd}}$, where $(\mathfrak{c}, \mathfrak{d}) \in \mathcal{A}^{*2}$, we have

$$\mathrm{P}\big[\mathrm{E}_{\tilde{\mathrm{P}}}\big[\tilde{L}_n(S) - L_n(S) \| X, Y\big] \geq \varepsilon \| \mathscr{D}_n^c(\vec{p}_S, \delta), \mathscr{F}_n\big]$$

$$\geq \mathrm{P}\big[\mathrm{E}_{\tilde{\mathrm{P}}}\big[S_{\pi_n^*}(\tilde{X},Y) - S_{\pi_n^*}(X,Y)\|X,Y\big] \geq \varepsilon \| \mathscr{D}_n^c(\vec{p}_S,\delta), \mathscr{F}_n\big]$$

$$= \mathrm{P}\bigg[\frac{n}{|\mathcal{J}|}\sum_{\mathfrak{c}\in\mathcal{A}^*} p^*_{\mathfrak{ac}}(S_{\mathfrak{bc}} - S_{\mathfrak{ac}}) \geq \varepsilon \| \mathscr{D}_n^c(\vec{p}_S,\delta), \mathscr{F}_n\bigg]$$

$$\overset{(6.2)}{\geq} \mathrm{P}\bigg[\frac{2nq_{\mathfrak{a}}}{|\mathcal{J}|}\varepsilon \geq \varepsilon \| \mathscr{D}_n^c(\vec{p}_S,\delta), \mathscr{F}_n\bigg] = 1 \tag{6.4}$$

Therefore,

$$\mathrm{P}\big[\mathrm{E}_{\tilde{\mathrm{P}}}\big[\tilde{L}_n(S) - L_n(S)\|X,Y\big] \geq \varepsilon\big]$$

$$\geq \mathrm{P}\big[\mathrm{E}_{\tilde{\mathrm{P}}}\big[\tilde{L}_n(S) - L_n(S)\|X,Y\big] \geq \varepsilon \| \mathscr{D}_n^c(\vec{p}_S,\delta), \mathscr{F}_n\big] \times \mathrm{P}\big[\mathscr{D}_n^c(\vec{p}_S,\delta), \mathscr{F}_n\big]$$

$$\overset{(6.1),(6.3),(6.4)}{\geq} 1 - \mathrm{e}^{-n\frac{q_{\mathfrak{a}}^2}{2}} - \mathrm{e}^{-K_\delta n} \quad \forall n \in \mathbb{N}.$$

The conditions of Theorem 6.1 are thus met for $\beta > 0$ small enough, and the claimed order of fluctuation holds true. □

Acknowledgements

Raphael Hauser received support from EPSRC grants EP/H02686X/1 and EP/N510129/1. Heinrich Matzinger was supported under EPSRC grant EP/I01893X/1, IMA Grant SGS29/11, and by Pembroke College Oxford.

Supplementary Material

Omitted examples and proofs (DOI: 10.3150/18-BEJ1053SUPP; .pdf). The supplemental article (Hauser and Matzinger [16]) provides additional examples in support of Section 1, as well as the omitted proofs of Sections 2 and 3.

References

[1] Alexander, K.S. (1994). The rate of convergence of the mean length of the longest common subsequence. *Ann. Appl. Probab.* **4** 1074–1082. MR1304773

[2] Amsalu, S., Hauser, R. and Matzinger, H. (2014). A Monte Carlo approach to the fluctuation problem in optimal alignments of random strings. *Markov Process. Related Fields* **20** 107–144. MR3185558

[3] Baxevanis, A.D. and Ouellette, B.F.F. (2005). *Bioinformatics*: *A Practical Guide to the Analysis of Genes and Proteins*. New York: Wiley.

[4] Bertsekas, D.P. and Yu, H. (2011). A unifying polyhedral approximation framework for convex optimization. *SIAM J. Optim.* **21** 333–360. MR2783219 https://doi.org/10.1137/090772204

[5] Borwein, J.M. and Lewis, A.S. (2000). *Convex Analysis and Nonlinear Optimization*: *Theory and Examples*. *CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC* **3**. New York: Springer. MR1757448 https://doi.org/10.1007/978-1-4757-9859-3

[6] Boutet de Monvel, J. (1999). Extensive simulations for longest common subsequences. *Eur. Phys. J. B* **7** 293–308.
[7] Carathéodory, C. (1911). Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen. *Rend. Circ. Mat. Palermo* (2) **32** 193–217.
[8] Chvatal, V. and Sankoff, D. (1975). Longest common subsequences of two random sequences. *J. Appl. Probab.* **12** 306–315. MR0405531 https://doi.org/10.2307/3212444
[9] Durringer, C., Hauser, R. and Matzinger, H. (2008). Approximation to the mean curve in the LCS problem. *Stochastic Process. Appl.* **118** 629–648. MR2394846 https://doi.org/10.1016/j.spa.2007.05.010
[10] Fekete, M. (1923). Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Math. Z.* **17** 228–249. MR1544613 https://doi.org/10.1007/BF01504345
[11] Gruber, P.M. (1993). History of convexity. In *Handbook of Convex Geometry, Vol. A, B* (P.M. Gruber, and J.M. Wills, eds.) 1–15. Amsterdam: North-Holland. MR1242974
[12] Hausdorff, F. (1949). *Grundzüge der Mengenlehre*. New York, NY: Chelsea Publishing Company. MR0031025
[13] Hauser, R., Martínez, S. and Matzinger, H. (2006). Large deviations-based upper bounds on the expected relative length of longest common subsequences. *Adv. in Appl. Probab.* **38** 827–852. MR2256880 https://doi.org/10.1239/aap/1158685004
[14] Hauser, R. and Matzinger, H. (2013). Letter change bias and local uniqueness in optimal sequence alignments. *J. Stat. Phys.* **153** 512–529. MR3107656 https://doi.org/10.1007/s10955-013-0819-4
[15] Hauser, R., Matzinger, H. and Popescu, I. (2018). An upper bound on the convergence rate of a second functional in optimal sequence alignment. *Bernoulli* **24** 971–992. MR3706783 https://doi.org/10.3150/16-BEJ823
[16] Hauser, R.A. and Matzinger, H. (2020). Supplement to "Microscopic path structure of optimally aligned random sequences." https://doi.org/10.3150/18-BEJ1053SUPP.
[17] Hirmo, E., Lember, J. and Matzinger, H. (2012). Detecting the homology of DNA-sequences based on the variety of optimal alignments: A case study. Available at arXiv:1210.3771.
[18] Howard, C.D. (2004). Probability on discrete structures. In *Encyclopaedia of Mathematical Sciences* (H. Kesten, ed.) **110**. Berlin: Springer. MR2023649 https://doi.org/10.1007/978-3-662-09444-0
[19] Jerrum, M. (2003). *Counting, Sampling and Integrating: Algorithms and Complexity. Lectures in Mathematics ETH Zürich*. Basel: Birkhäuser. MR1960003 https://doi.org/10.1007/978-3-0348-8005-3
[20] Karlin, S. and Altschul, S.F. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. USA* **87** 2264–2268.
[21] Kingman, J.F.C. (1973). Subadditive ergodic theory. *Ann. Probab.* **1** 883–909. MR0356192 https://doi.org/10.1214/aop/1176996798
[22] Krein, M. and Milman, D. (1940). On extreme points of regular convex sets. *Studia Math.* **9** 133–138. MR0004990 https://doi.org/10.4064/sm-9-1-133-138
[23] Lember, J. and Matzinger, H. (2009). Standard deviation of the longest common subsequence. *Ann. Probab.* **37** 1192–1235. MR2537552 https://doi.org/10.1214/08-AOP436
[24] McDiarmid, C. (1989). On the method of bounded differences. In *Surveys in Combinatorics*, 1989 (*Norwich*, 1989). *London Mathematical Society Lecture Note Series* **141** 148–188. Cambridge: Cambridge Univ. Press. MR1036755
[25] Mil'man, D. (1947). Characteristics of extremal points of regularly convex sets. *Dokl. Akad. Nauk SSSR* **57** 119–122. MR0022313
[26] Minkowski, H. (1968). *Geometrie der Zahlen. Bibliotheca Mathematica Teubneriana, Band* 40. New York: Johnson Reprint Corp. MR0249269

[27] Needleman, S.B. and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48** 443–453.
[28] Rademacher, H. (1919). Über partielle und totale differenzierbarkeit von Funktionen mehrerer Variabeln und über die Transformation der Doppelintegrale. *Math. Ann.* **79** 340–359. MR1511935 https://doi.org/10.1007/BF01498415
[29] Rockafellar, R.T. (1997). *Convex Analysis. Princeton Landmarks in Mathematics*. Princeton, NJ: Princeton Univ. Press. MR1451876
[30] Steele, J.M. (1986). An Efron–Stein inequality for nonsymmetric statistics. *Ann. Statist.* **14** 753–758. MR0840528 https://doi.org/10.1214/aos/1176349952
[31] Steele, J.M. (1989). Kingman's subadditive ergodic theorem. *Ann. Inst. Henri Poincaré Probab. Stat.* **25** 93–98. MR0995293
[32] Steele, J.M. (1997). *Probability Theory and Combinatorial Optimization. CBMS-NSF Regional Conference Series in Applied Mathematics* **69**. Philadelphia, PA: SIAM. MR1422018 https://doi.org/10.1137/1.9781611970029
[33] Waterman, M. (1994). Estimating statistical significance of sequence alignments. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **344** 383–390.
[34] Waterman, M. (1995). *Introduction to Computational Biology*. London: Chapman & Hall.

Received June 2016 and revised June 2018