Distance-driven adaptive trees in biological metric spaces: uninformed accretion does not prevent convergence

BY YANNICK LOUIS KERGOSIEN1,2,*

1Département de Sciences de l’Information, Université de Cergy-Pontoise, 196 rue des Rabats, 92160 Antony, France
2LIM&BIO (EA 3969), UFR SMBH, Université Paris 13, France

We present several variants of a stochastic algorithm which all evolve tree-structured sets adapted to the geometry of general target subsets in metric spaces, and we briefly discuss their relevance to biological modelling. In all variants, one repeatedly draws random points from the target (step 1), each time selecting from the tree to be grown the point which is closest to the point just randomly drawn (step 2), then adding to the tree a new point in the vicinity of that closest point (step 3 or accretion step). The algorithms differ in their accretion rule, which can use the position of the target point drawn, or not. The informed case relates to the early behaviour of self-organizing maps that mimic somatotopy. It is simple enough to be studied analytically near its branching points, which generally follow some unsuccessful bifurcations. Further modifying step 2 leads to a fast version of the algorithm that builds oblique binary search trees, and we show how to use it in high-dimensional spaces to address a problem relevant to interventional medical imaging and artificial vision. In the case of an uninformed accretion rule, some adaptation also takes place, the behaviour near branching points is computationally very similar to the informed case, and we discuss its interpretations within the Darwinian paradigm.

Keywords: tree; adaptive; evolution; self-organizing; fractal; stochastic

1. Introduction

Branching structures and phenomena are observed in different biological contexts. Anatomical tree-like structures such as vascular networks stand in physical space, their existence is not questionable and their formation generally has some adaptive properties for which some guiding signals can often be found. The morphogenesis of these structures has some analogy with purely physical phenomena such as the growth of crystals or the formation of sparks and material fractures, and some models borrowed from physics, such as diffusion limited aggregation, have been used as a first approach to biological modelling, even if some elaboration and complexification are needed for further realism, e.g. in the form of reaction–diffusion models involving several diffusible species or mechanisms.

*y.l.kergosien@gmail.com

One contribution of 17 to a Theme Issue ‘From biological and clinical experiments to mathematical models’.
for signal integration across several cells. Self-organization can also be studied at higher levels of organization, such as in neuroscience, ecology or evolution. At these levels, phenomena take place in abstract spaces such as spaces of biological parameters and, there, diffusion can be less important even if it is relevant at the lower levels. Distances can be more effective there, for instance, in the form of correlations between activation patterns in neuroscience or biometric distances in evolution. Models can use such distances within a differential equations formalism and describe forces acting on different components in the model. It is tempting to ask whether some general properties of the growth of trees can be studied at that level of abstraction and to build a framework where this would be possible. A schematic drawing of an evolutionary tree was included by Sir Charles Darwin as an illustration to ch. 4 of *On The Origin of Species* (Darwin 1859). Darwin’s view of evolution as a natural phenomenon calls for an attempt at finding forces on evolutionary trees, ideally deduced from natural forces in population dynamics, that would to a certain amount explain some growth characteristics of these trees, especially if some adaptation is involved. We describe a unified framework in which distance-driven growth of some tree-like sets, according to an algorithm that specifies a stochastic process, is shown to adapt to very general target sets in metric spaces. We shall borrow vocabulary from plant biology and angiogenesis, thinking of a vascular tree that grows towards an organ or a tumour, which we call the ‘target’, still calling ‘seed’ the first state of the growing set. After describing a common framework for the algorithm, we shall describe the informed accretion version of this process, where the small modifications—or accretions—of the growing set use information about the target to directly produce adaptation-enhancing accretions, then a modification of this informed process that produces search trees relevant to technological applications in artificial vision and interventional medical imaging, and finally the uninformed accretion process where accretions can be compared with Darwinian mutations as their directions are blind to the target and have to be fit enough to result in survival.

2. Common setting of the algorithms

Let $E$ be a metric space, i.e. a set on which a distance $d$ is defined, and $T$, the target, a subset of $E$ which is the support of a probability measure $P$ (Parthasarathy 1967), i.e. such that for any $c$ in $T$, any real number $r > 0$, and any open ball $B(c, r)$ centred at $c$ with radius $r$, $P(B(c, r)) > 0$. We are also given an initial finite subset, the seed, often a single point, which constitutes the growing set at time 0 and is thus denoted $G_0$. For each further instant $t$ taken in $\{1, 2, \ldots\}$, we shall use $P$ to randomly draw (i.e. select) from $T$ the point $c_t$, and compute a set $G_t$ depending on $G_{t-1}$ and $c_t$. In our examples, we shall often consider the simpler case where $E$ is the Euclidean space $\mathbb{R}^n$ and $T$ is a closed subset of it with a probability distribution on it.

The algorithms have the following general structure:

(i) set $t = 0$

(ii) repeat:
   — increment $t$
   — randomly draw a point $c_t$ from $T$ using $P$
— compute the point $a_t$ in $G_{t-1}$ which minimizes $d(c_t, a_t)$, possibly using a randomized rule to avoid ties
— set $G_t = G_{t-1} \cup \{b_t\}$, where point $b_t$ is computed from $a_t$ and possibly
$c_t$ or a random variable, according to the accretion rule. Call $a_t$ the parent of $b_t$
(iii) until some condition on $t$ or $G_t$ is met.

3. The informed algorithm and its adaptive properties

(a) Algorithm

In the informed accretion algorithm, the accretion rule is

$$b_t = (1 - \epsilon)a_t + \epsilon c_t,$$

where $\epsilon$ is a small real positive number and $b_t$ is, in spaces where this can be
defined, the barycentre of $a_t$ and $c_t$ with $(1 - \epsilon)$ and $\epsilon$ as respective weights.
We shall use the same notation in the more general metric spaces where it is
possible to find $b_t$ aligned with $a_t$ and $c_t$, i.e. $d(a_t, b_t) + d(b_t, c_t) = d(a_t, c_t)$, and
such that $d(a_t, b_t) = \epsilon d(a_t, c_t)$. Here again some randomized rule can be applied
to pick a particular $b_t$ among a choice of possibilities. Riemannian manifolds are
metric spaces where such generalization can be used, $b_t$ being found on a geodesic
from $a_t$ to $c_t$; they are common in biometric applications where shape spaces and
Procrustes methods are used (Kendall 1984). Accretion of $b_t$ thus takes place in
the vicinity of $a_t$ and in the direction of $c_t$, the latter fact making the modification
informed about the target.

The informed algorithm is thus

(i) set $t = 0$
(ii) repeat:
— increment $t$
— randomly draw a point $c_t$ from $T$ using $P$
— compute the point $a_t$ in $G_{t-1}$ which minimizes $d(c_t, a_t)$, possibly using
a randomized rule to avoid ties
— set $b_t = (1 - \epsilon)a_t + \epsilon c_t$ and $G_t = G_{t-1} \cup \{b_t\}$, call $a_t$ the parent of $b_t$
(iii) until some condition on $t$ or $G_t$ is met.

Figure 1 shows the evolution of growing sets in $\mathbb{R}^2$ from a single point to a
segment for different values of $\epsilon$. In figure 2, a dual target is made of a circle and
a disc. Successive branchings lead the tips of the branches close to the target.
More precisely, notice that the parenthood relationship gives $G_t$ the structure
of a tree if $G_0$ is a singleton; call tips its leaves, and say that branching occurs
at time $t$ if the parent $a_t$ of $b_t$ already had a child in $G_{t-1}$. In that case $a_t$ is
called a branching point. The branches of the network at time $t$ are the connected
components of the tips once the branching points have been removed from $G_t$. In
order to analyse branching, let us call active at time $t$ the points of $G_t$ which have
a non-zero probability of being $a_{t+1}$. It is also useful to define as the target share
of a point $a \in G_t$ the set $S_a = \{c \in T | \forall a' \in G_t, \ d(c, a) \leq d(c, a')\}$. Notice that once it
Figure 1. Informed accretion: growing a tree towards a line segment in $\mathbb{R}^2$. (a) $\epsilon = 0.01$, 2000 steps. (b) $\epsilon = 0.001$, 50000 steps.

Figure 2. Informed accretion: growing a tree towards a dual target (not drawn) made of (a) a circle and (b) a disc in $\mathbb{R}^2$ ($\epsilon = 0.01$, 80000 steps).

*Phil. Trans. R. Soc. A* (2009)
becomes inactive, a point cannot grow anymore and cannot be active again since
the points which collectively ‘dominate’ it will never be removed. Also, each \( b_t \) is
active at time \( t \) owing to the informed accretion rule. For \( G_t \) to have two active
points, it must have two elements \( x \) and \( y \), such that the bisector hyperplane of
the segment \((x, y)\) cuts the target. Starting from a singleton seed far from the
target, each \( b_t \) will be the unique active point until some \((b_{t-1}, b_{t})\) segment is
such that \( b_{t-1} \) is active at time \( t \), which requires that \( b_{t-1} \) ‘sees’ the target with
an angle at least equal to \( \pi/2 \), and thus be close to the target. After the first
branching, the target is split by the bisector hyperplane of the segment joining
the two tips. Accretion at each tip is due to the random points that are drawn
from its target share, with a probability depending on the weight of that region,
which in turn moves the separating hyperplane and gives a small advantage to
one of the tips. This competition between the tips requires a quantitative analysis
postponed to §3c, but it should be clear that splitting the target is likely to
decrease the maximum angle just mentioned and demands repeating the process
of approaching the target share before splitting can occur again.

As the tree grows, the third instruction in the loop of the algorithm gets
computationally more costly if all the points of \( G_{t-1} \) are examined to find \( a_t \). It is
important to periodically remove the inactive points from the list of points to be
checked. Simple targets such as in figure 1 afford fast algorithms for such garbage
collection, and for complex cases, one can at least approximate the target by a
finite set and exhaustively scan it to tag inactive points. In higher dimensional
spaces, computational complexity depends on the distance function, which in
Euclidean space remains very tractable (figure 3).
(b) Convergence

A convergence result has been obtained for this algorithm (Kergosien 2003). We provide here a stronger convergence result and a proof that will also readily adapt, in §5, to the new case of uninformed accretion. We shall often use the following theorem (Parthasarathy 1967) and its corollary in the present context.

**Theorem 3.1 (Parthasarathy (1967), theorem 1.1).** Let $(E, d)$ be a metric space. For any subset $F \subseteq E$ and any $x$ in $E$, define $d(x, F) = \inf_{y \in F} d(x, y)$. Then, for any $x, x'$ in $E$, $|d(x, F) - d(x', F)| \leq d(x, x')$.

**Corollary 3.2.** Let $A$ be a subset of the target $T$ such that for any $c, c'$ in $A$, $d(c, c') < \delta$. For any $t$ in $\mathbb{N}$, and any $c \in T$, define $\Delta_t(c) = \min_{x \in G_t} d(c, x)$, the distance from $c$ to $G_t$. Then for any $t$ in $\mathbb{N}$, any $c, c'$ in $A$, $|\Delta_t(c) - \Delta_t(c')| < \delta$.

**Lemma 3.3.** For any point $c$ in $A$, define $\Gamma_t(c) = \{ a \in G_t \mid d(a, c) = \Delta_t(c) \}$, the set of points active for $c$ at time $t$, and for any $A \subseteq T$, set $\Gamma_t(A) = \bigcup_{c \in A} \Gamma(c)$ and call it the set of points active for $A$ at time $t$. Let $A$ be a subset of $T$ such that for any $c, c'$ in $A$, $d(c, c') < \delta$, and $t$ a time such that accretion of $b_t$ happened at $a_t$ after $c_t$ was drawn from $A$. If $\Delta_{t-1}(c_t) - \Delta_t(c_t) > 2\delta$ then $\Gamma_t(A) = \{ b_t \}$, i.e. the point just accreted is the unique point active for $A$ at time $t$. We shall call any accretion occurring with this property an efficient $A$-step, and any $t$ such that $c_t$ is in $A$ an $A$-time.

**Proof.** Take any $c$ in $A$. Elements of $\Gamma_t(c)$ can be either $b_t$ or elements of $\Gamma_{t-1}(c) = \Gamma_t(c) \cap G_{t-1}$. Suppose $x$ is in $\Gamma_t(c) \cap \Gamma_{t-1}(c)$. We know that $a_t$ is in $\Gamma_{t-1}(c_t)$ and, using theorem 3.1, that $|d(c, x) - d(c_t, a_t)| < \delta$; in particular $d(c, x) > d(c_t, a_t) - \delta$. But, by the hypothesis

$$d(c_t, a_t) = \Delta_{t-1}(c_t) > \Delta_t(c_t) + 2\delta = d(c_t, b_t) + 2\delta,$$

which, combined with the former inequality, gives

$$d(c, x) > d(c_t, b_t) + \delta.$$

To conclude, we use the hypothesis on $A$ and the triangle inequality:

$$\delta + d(c_t, b_t) > d(c, c_t) + d(c_t, b_t) \geq d(c, b_t)$$

to get $d(c, x) > d(c, b_t)$, so that $x$ is not active at time $t$ for $c$. This proves that whatever $c$ in $A$, the only active point at time $t$ for $c$ is $b_t$. ■

**Corollary 3.4.** Let $A \subseteq T$ such that for any $c, c'$ in $A$, $d(c, c') < \delta$. If $k$ efficient $A$-steps occurred at times $t_1, \ldots, t_k$, then

$$d(c_{t_k}, b_{t_k}) < d(c_{t_1}, b_{t_1}) - (k - 1)\delta < \sup_{c \in T} d(c, G_0) - k\delta.$$

**Proof.** As step at time $t_k$ is efficient and from a drawing in $A$

$$d(c_{t_k}, b_{t_k}) < d(c_{t_k}, a_{t_k}) - 2\delta.$$

But we also have, since $t_{k-1} + 1 \leq t_k$, from the monotony of $\Delta$ and corollary 3.2

$$\Delta_{t_k}(c_{t_k}) = d(c_{t_k}, a_{t_k}) \leq \Delta_{t_{k-1}+1}(c_{t_k}) \leq \Delta_{t_{k-1}+1}(c_{t_{k-1}}) + \delta = d(c_{t_{k-1}}, b_{t_{k-1}}) + \delta.$$

We thus get

$$d(c_{t_k}, b_{t_k}) < d(c_{t_{k-1}}, b_{t_{k-1}}) - \delta.$$
Iterating such inequalities and using efficiency at time $t_i$
\[d(c_{t_k}, b_{t_k}) < d(c_{t_i}, b_{t_i}) - (k - 1)\delta < d(c_{t_i}, a_{t_i}) - (k + 1)\delta\]
and using again corollary 3.2
\[d(c_{t_k}, b_{t_k}) < d(c_{t_i}, b_{t_i}) - (k - 1)\delta < d(c_{t_i}, a_{t_i}) - (k + 1)\delta < \sup_{c \in T} d(c, G_0) - k\delta.\]

**Theorem 3.5.** Let $(E, d)$ be a metric space, $T$ a compact subset of $E$ which is the support of a probability measure $P$ on $E$ and $G_0$ a finite subset of $E$. The following property is $P$-almost sure: for any real $\xi > 0$ there exists a finite time $t$ from which the sets $G_{t'}$, $t' \geq t$, grown from $G_0$ according to the informed adaptive tree algorithm will simultaneously have the two properties

(i) for any $c$ in $T$, there exists $x$ in $G_{t'}$ such that $d(c, x) < \xi$, i.e. no point of the target is more than $\xi$ away from $G_{t'}$ and

(ii) for any $x$ in $G_{t'}$, for any $c$ in $T$, if $x$ is in $\Gamma_{t'}(c)$ then $d(x, c) < \xi$, i.e. active points are not further than $\xi$ away from any point of their target shares.

**Proof.** First notice that once property (i) is true, it remains true because for any $c$ in $T$, since no point is ever removed from the growing set, the function $\Delta_i(c)$ cannot increase with $t$.

Property (ii) is an immediate consequence of property (i), owing to the properties of $\Delta$ and $\Gamma$: if $x$ in $G_t$ is in $\Gamma_i(c)$ for some $c$ in $T$ with $d(x, c) \geq \xi$, as $c$ is, by property (i), at a distance $d(c, y) < \xi$ from another element $y$ of $G_t$, $x$ cannot be active at $c$ at time $t$, a contradiction.

We must now prove that property (i) is almost surely reached in finite time.

Let us first consider a subset $A$ of $T$. Adequately choosing the size of $A$, we can ensure that all steps for which the target point is drawn from $A$ are efficient, at least as long as our objective of getting all the active points close enough to $A$ is not met. The fact that the dispersion of the $\Delta(c)$’s is never more than the diameter of $A$ is essential. We must get specific about the algorithm used and bundle that variable part of the proof in a last lemma.

**Lemma 3.6.** Let $A$ be a subset of $T$ contained in an open ball, denoted $B(e, \delta/2)$, of centre $e \in E$ and radius $\delta/2$. Almost surely there exists $k \in \mathbb{N}$ such that for any $n \geq k$ and any $c$ in $A$, $\Delta_n(c) < \delta(7/2 + 2/\epsilon)$ (remember that $\epsilon$ is defined with the accretion rule).

**Proof.**

(i) Notice that for any $c$ in $B(e, \delta/2)$ and any $x$ not in $B(e, \delta/2 + R)$, where $R > 0$ will be computed later, $d(c, x) > R$, since
\[\frac{\delta}{2} + d(c, x) \geq d(e, c) + d(c, x) \geq d(e, x) \geq \frac{\delta}{2} + R.\]

This guarantees that each accretion at a $t$ such that $c_t \in A$ will produce a decrease $\Delta_{t-1}(c_t) - \Delta_t(c_t) = \epsilon d(c_t, a_t) \geq \epsilon R$ as long as $a_t$ lies outside $B(e, \delta/2 + R)$ (we used the alignment property of the informed accretion rule).
If we can prove that such steps eventually bring the unique active point $a_t$ at some $t$ inside $B(e, \delta/2 + R)$, we shall still need a bound on the distance of such active point to any another point of $A$.

(ii) For any $c$ in $B(e, \delta/2)$ and any $x$ in $B(e, \delta/2 + R)$, $d(c, x) < R + \delta$, since $d(c, x) \leq d(c, e) + d(e, x) < \delta/2 + R + \delta/2 = R + \delta$.

(iii) We can check that taking $R = 2\delta/\epsilon$ brings at the same time

— $\epsilon R \geq 2\delta$ to ensure efficient steps each time $c_t$ is drawn from $A$ and $a_t$ lies outside $B(e, \delta/2 + R)$,

— $d(c, x) < R + \delta = \delta(1 + 2/\epsilon)$ whenever $c \in A$ and $x \in B(e, \delta/2 + R)$.

(iv) Let $B_1 = B(e, \delta/2)$, $B_2 = B(e, \delta/2 + R)$, $B_3 = B(e, \delta/2 + R + 5\delta/2)$. For any $c, c' \in B_1$, $x' \in B_2$, $x \notin B_3$, we can first write

$$\frac{\delta}{2} + d(c, x) > d(e, c) + d(c, x) \geq d(e, x) \geq R + \frac{5\delta}{2}$$

to get $d(c, x) > R + 5\delta/2 - \delta/2 = R + 2\delta$, which combined to

$$d(c', x') \leq d(c', e) + d(e, x') < \frac{\delta}{2} + \frac{\delta}{2} + R = R + \delta$$

implies

$$d(c, x) - d(c', x') > \delta.$$

This shows, using diam$(A) < \delta$ and corollary 3.2, that $\Gamma_t(A)$ cannot have points both in $B_2$ and outside $B_3$.

The event we are thus going to wait for, and which implies the conclusion of lemma 3.6, i.e. $\Delta_n(c) < \delta(7/2 + 2/\epsilon)$, is $\Gamma_t(A) \subset B_3$. It implies the conclusion of lemma 3.6 because for any $x$ in $B_3$, $d(e, x) < 3\delta + 2\delta/\epsilon$ and thus, for any $c$ in $A$, $d(c, x) \leq d(c, e) + d(e, x) < \delta/2 + 3\delta + 2\delta/\epsilon = \delta(7/2 + 2/\epsilon)$, so that if the active points relative to $A$ are in $B_3$, $\Delta_n(c)$ is less than $\delta(7/2 + 2/\epsilon)$.

The nested open balls $B_1$, $B_2$ and $B_3$ have the property that as long as $\Gamma_t(A)$ has any element outside $B_3$, all its elements are outside $B_2$, making certain that the next drawing of $c_t$ from $A$ will result in an efficient step, whatever the resulting $a_t$. On the other hand, if all the elements of $\Gamma_t(A)$ are in $B_3$, a sufficient condition for this being that at least one element of $\Gamma_t(A)$ is in $B_2$, then the inequality we want to reach is met.

(v) Let $(X_t)_{t \in \mathbb{N}^+}$ be the sequence of random variables such that $X_t = 1$ if $c_t$ is drawn from $A$, and $X_t = 0$ otherwise. These random variables are independent identically distributed with $P(X_t = 1) = P(A) > 0$ for any $t$, obviously making the series $\sum_{t=1}^{\infty} P(X_t = 1)$ diverge, and the Borel–Cantelli lemma tells us that with a probability equal to 1, i.e. almost surely, an infinite number of such events occur.

(vi) Let us show that from time $t = 0$, at which $\sup_{c \in A}(\Delta_0(c)) < \infty$, only a finite number of efficient steps are needed in order to observe an element of $\Gamma(A)$ inside $B_2$. Using corollary 3.4, let us compute the smallest integer $k$ such that $\sup_{c \in A}(\Delta_0(c)) - k\delta < R$. We know that whatever $k$ we shall almost surely observe such a number of $A$-times $t_1, \ldots, t_k$. After these, either the
corresponding steps will all have been efficient $A$-steps, and the inequality, becoming true, will also imply that $b_{tk}$ is in $B_2$ since $d(e, b_{tk}) \leq d(e, c_{tk}) + d(c_{tk}, b_{tk}) < \delta/2 + d(c_{tk}, b_{tk})$, or some active point will have earlier entered $B_2$, and thus all the active points will have entered $B_3$, which will also enforce $\Delta_n(c) < \delta(7/2 + 2/\epsilon)$.

Back to the proof of the theorem, remember that since $T$ is the support of the probability measure $P$, any open ball of positive radius around an element of $T$, however small, has a positive probability. Consider for every $c$ in $T$ the open ball of diameter $\delta$ centred at $c$, where $\delta = \xi/(7/2 + 2/\epsilon)$. These balls of course cover $T$, and using the compacity of $T$, extract from this family a finite covering of $T$. The intersections of these balls with $T$ provide us with a finite number of subsets of $T$: $A_1, \ldots, A_m$ (possibly overlapping), of positive probabilities, each of which not containing any $c, c'$ such that $d(c, c') \geq \delta$. From lemma 3.6, we get $m$ properties which read ‘there exists $k_i$ such that for any $n \geq k_i$ and any $c$ in $A_i \Delta_n(c) < \xi$’, each holding with a probability equal to 1. The conjunction of these properties also holds with a probability equal to 1 and we only have to take $t = \max(k_1, \ldots, k_m)$ to get a time from which properties (i) and (ii) are true with $\xi$. Now take a sequence of positive $\xi_1, \xi_2, \ldots$ which tends to 0. For each $\xi_j$, there exists almost surely an $N_j$ such that properties 1 and 2 are true for that $\xi_j$, and almost surely all these facts hold simultaneously. As a consequence, with a probability equal to 1, for any positive $\xi$ there exists a $\xi_j$ which is less than $\xi$ and makes points 1 and 2 of the conclusion true for $\xi$.

(c) Behaviour near branching

We recall here some results about the phenomenon of unsuccessful branchings (or abortive branching; see Kergosien 1988, 2003) and provide some analysis for it before we observe another form of it for the uninformed algorithm. Decreasing $\epsilon$ makes the tree more regular (compare the two cases in figure 1), but close inspection of the branches before they bifurcate shows (figure 4) that even with small $\epsilon$, branching happens only after a sequence of unsuccessful branchings. This phenomenon can be studied analytically in the following simple case. Let us consider in $E = \mathbb{R}^2$ a target $T = \{(x, y) \in \mathbb{R}^2 | x = e, y_2 \leq y \leq y_1\}$, i.e. the line segment extending from $M_1 = (e, y_1)$ to $M_2 = (e, y_2)$, with uniform probability distribution. We suppose a growing set at a time when it has only two tips $A$ and $B$ which will both remain very close to the origin during all the time period we study ($\epsilon$ being taken small enough). The target shares of $A$ and $B$, i.e. the subsets of points of $T$ which are closer to $A$ or to $B$, are delimited by $N = (e, y_n)$ which is such that $d(N, A) = d(N, B)$ and is thus on the line, call it $L$, which
bisects the segment $AB$ orthogonally. One of the target shares disappears, with the corresponding tip dying, whenever $L$ crosses one end of the target. Under the approximating hypotheses, it is enough to study when the vector $X = A - B$ gets normal to one of the critical positions of $L$ associated to the deaths of the tips. Setting $X = (x, y)$ and assuming that $A$ and $B$ are attracted, respectively, by the upper part and the lower part of the target, we compute the expected vector field that acts on $X = A - B$:

$$\frac{d(A - B)}{dt} = \frac{k(y_1 - y_N)}{y_1 - y_2} \cdot \frac{M_1 + N}{2} - \frac{k(y_N - y_2)}{y_1 - y_2} \cdot \frac{N + M_2}{2},$$

and thus

$$\frac{dx}{dt} = e(y_1 + y_2 - 2y_N) \quad \frac{dy}{dt} = \frac{(y_1 - y_N)(y_1 + y_N) - (y_N - y_2)(y_N + y_2)}{y_1 - y_2}.$$

Computing $y_N$ from $x$ and $y$ leads to

$$\frac{dx}{dt} = \frac{ey_1 + ey_2 + 2e^2x/y}{y_1 - y_2} \quad \frac{dy}{dt} = \frac{y_1^2 + y_2^2 - 2e^2x^2/y^2}{2(y_1 - y_2)},$$

which we shall write

$$\frac{dx}{dt} = c \left( \frac{x}{y} \right) + d$$

and

$$\frac{dy}{dt} = a \left( \frac{x}{y} \right)^2 + b.$$

To study the qualitative properties of that system, let us look for its eigen-directions, i.e. the directions defined by $dx/dy = x/y$

$$a \left( \frac{x}{y} \right)^3 + (b - c) \left( \frac{x}{y} \right) - d = 0.$$
(d) Biological relevance of distance-based informed adaptation

Developing biological structures commonly use information provided by diverse growth factors to guide their adaptation to targets in physical space. For these signals, the notions of concentration and diffusion are essential. Distance-based processing of signals, on the other hand, must be searched at higher levels of organization, like in neural structures, where, for instance, distances between activation patterns come in the form of correlations. It is in the context of signal processing and neural modelling of somatotopic maps that Kohonen’s self-organizing maps (SOMs) were first described (see references in Kohonen 1997), and several features of the informed accretion algorithm, in particular the random point selections and the accretion rule, are analogous to it. Indeed, the behaviour of the very first steps of the evolution of SOMs, when the net unfolds from an initially clustered position, is very similar to that of the informed accretion algorithm, with some branching patterns often apparent on proper visualizations. Distance-based adaptation in abstract spaces could also be sought in spaces of molecular shapes. In these contexts, however, less signal integration could be available and population-based mechanisms might be more likely. We shall see in §5 that uninformed accretion algorithms have very similar adaptation properties while affording a Darwinian interpretation. Distance-driven adaptation could thus find uninformed population-based implementations where informed accretion algorithms require too much integrative processing.

4. Technological aside: adaptive search trees for interventional medical imaging and vision

Before studying the algorithm with uninformed accretion in the next section, we summarize here the results of a modification of the informed accretion algorithm, still of an informed kind, which applies to problems in artificial vision and interventional medical imaging (see Kergosien (2005) for details). This will show how the formalism of metric spaces can apply to concrete situations, but it will also show the dangers of giving up the conditions for convergence.
This algorithm produces binary search trees that quickly retrieve, from a point in the target, its closest active point. This permits using the tree for tasks related to classification and searching, and also speeds up the growth algorithm.

The principle is to require that after each branching with tips \( b \) and \( b' \) the target shares of \( b \) and \( b' \) are frozen. Each branching thus splits the target share of its branching point in two subtargets, each definitively assigned to one of the future two subtrees. This is equivalent to changing the proximity criterion; then, given \( c \) in \( T \), in order to look for the point in \( G \) closest to \( c \) we first compare the distances from \( c \) to the two tips of the first branching point to know in which subtree we must continue the search, then compare the distances of \( c \) with the two tips of the next branching point on that subtree and so on until we reach a terminal branch, after which we use the same rules as for the standard informed algorithm. These operations use classical binary search tree data structures and fast algorithms, with oblique separating hyperplanes orthogonally bisecting the segments \( bb' \) joining the tips present at each bifurcation. Using a function \( d: T \times E \rightarrow \mathbb{R}^+ \) which is not a distance anymore, and which has to be updated after each branching, we come close to the algorithmic framework of §2:

(i) set \( t = 0, \quad d_{t+1} = d \) (i.e. start with the original distance)
(ii) repeat:

| increment \( t \) |
| randomly draw a point \( c_t \) from \( T \) using \( P \) |
| compute the point \( a_t \) in \( G_t \) which minimizes \( d_t(c_t, a_t) \), possibly using a randomized rule to avoid ties |

| set \( G_t = G_{t-1} \cup \{ b_t \} \), where \( b_t \) is computed like in algorithm 1: \( b_t = (1 - \epsilon)a_t + \epsilon c_t \). |

Call \( a_t \) the parent of \( b_t \)

| if no branching occurred at \( a_t \), set \( d_{t+1} = d_t \), otherwise update \( d_{t+1} \) with the following rule: let \( a \) be the branching point and \( b_1, b_2 \) its two children (one of them is \( b_t \)). For any \( x \) in \( T \) such that \( d_t(x, b_1) < d_t(x, b_2) \) and any \( y \) in \( E \), set \( d_{t+1}(x, y) = d_t(x, y) \) if \( d_t(y, b_1) < d_t(y, b_2) \), otherwise set \( d_{t+1}(x, y) = \infty \). Swapping \( b_1 \) and \( b_2 \) completes the rule. A test and a randomized rule should care for possible ties |

(iii) until some condition on \( t \) or \( G_t \) is met.

The functions \( d_t \) built in the loop have simple meaning and use. At each branching, the hyperplane that bisects segment \( (b_1, b_2) \) also splits \( E \) and \( T \). Our rule requires that if a point has been drawn from the target on one side of that hyperplane, no point from the other side will be considered in searching for a closest active point, even at later steps. These successive requirements combine in function \( d_t \) and restrict the search to a convex region of \( E \) which will be further split when the next branching in that region occurs. Practically, it is enough to use classical data structures and algorithms for binary search trees, comparing the target point to the two children of the first branching point, then to the

Phil. Trans. R. Soc. A (2009)
Adaptive trees

Figure 6. (a) Growing a search tree in $\mathbb{R}^2$ towards a dual target made of a circle (top) and a disc (bottom). (b) Growing a search tree in $\mathbb{R}^6$ toward a 5-hypercube.

two children of the next branching point of the next permitted dead branch thus selected, and so on until the unique active branch is reached.

(b) Simulation

The behaviour of the growing set under this algorithm is very similar to what we saw before, as seen in figure 6. One notable exception is the absence of unsuccessful branchings, which might appear as an advantage. The algorithm is fast and does not need garbage collection.

(c) Application to interventional imaging and vision

We want to quickly retrieve the pose of a polyhedron from the outline of its projection, a problem of artificial vision and interventional medical imaging. We first implement a distance between two such outlines using a Procrustes method. Two parametrized curves in $\mathbb{R}^2$ being given as continuous periodic functions $\sigma_0, \sigma_1: [0, 2\pi] \to \mathbb{R}^2$ where the parameter is supposed to be proportional to the curvilinear abscissas, for each phase $\phi$ in $[0, 2\pi]$ and each displacement $D$ of $\mathbb{R}^2$ we can compute a numerical approximation to the expression

$$d_{D,\phi}(\sigma_0, \sigma_1) = \sqrt{\int_0^{2\pi} (D(\sigma_1(u + \phi)) - \sigma_0(u))^2 \, du},$$

where we used the periodicity of $\sigma_1$. As a distance between $\sigma_0$ and $\sigma_1$, we set

$$d(\sigma_0, \sigma_1) = \inf_{D,\phi} d_{D,\phi}(\sigma_0, \sigma_1).$$

The set of outlines observed for all possible poses of a polyhedron is taken as a target inside the space of all possible $n$-gons, where all outline polygons are approximated by $n$-gons, represented by points in $\mathbb{R}^{2n}$. We thus grow a search tree from a seed, which is arbitrarily chosen to be a circle, towards the target obtained by computing a great number of outlines of the known polyhedron in random attitudes.

*Phil. Trans. R. Soc. A* (2009)
Figure 7. Search tree (root on the right, leaves on the left) grown towards the set of polygonal outlines of a human clavicle polyhedral model, starting from a circle taken as the seed. Only branching points and leaves of the tree are represented, each of them being replaced by a drawing of the corresponding polygon (the representation leads to some overlap). Notice the progressive differentiation towards the possible outlines of a clavicle. Parameters: 5000 steps, $\epsilon = 0.02$, giving 422 nodes and 202 leaves.

Figure 7 was obtained by growing a small tree towards the set of outlines of a model of a human right clavicle. It shows how successive branchings of the tree bring the progressive differentiation from the original circle to the various possible outlines. Very similar patterns are observed with the standard informed accretion algorithm, though taking more time and not having so direct applications to...
Adaptive trees

4981

Figure 8. Computer interface for real-time recognition of the pose of a clavicle polyhedral model. The clavicle polyhedral model can be rotated interactively in the right window; the left window displays the same polyhedral model in the position retrieved using the search tree; the middle window permits checking the match of the two outlines, and also displays as a third polygon the outline retrieved before the plane rotation was applied.

searching. Nodes of the tree are associated to polygons, and for the leaves, one also records the pose of the target polyhedron that produced the polygon that led to its accretion, together with the best in-plane rotation found while computing the Procrustes distance. Once the tree is computed, it is very quick to retrieve, given an outline of the polyhedron in an unknown pose, the closest outline among the leaves of the tree and the associated pose and plane rotation, from which one deduces the pose of the polyhedron which produces the best approximation of the given outline. Figure 8 shows the computer interface where the right window displays the clavicle polyhedral model that can be rotated interactively; the left window displays the same polyhedral model in the position retrieved in real time using the search tree; and the middle window permits checking the match of the two outlines, and also displays as a third polygon the outline retrieved before the plane rotation was applied.

(d) Discussion

This application shows how one can grow a self-organizing search tree in a high-dimensional space of curve shapes. It should be noticed that the adaptive search tree algorithm is completely generic and that no specialized programming was necessary beyond the implementation of the distance. The use of such a tree can be compared with Breiman’s popular classification and regression trees (Breiman et al. 1984) or, better, to its extensions to oblique classification trees, as the separating hyperplanes generated by the adaptive search tree algorithm are oblique. We were not able to prove nor disprove convergence for this algorithm. The results of §3, lemma 3.3, in particular, are not applicable here as we do not have a distance. Moreover, there are situations (figure 9) where a first piece of target generates many separating hyperplanes which the growing set has to respect during further growth, so that to reach some other target parts the active points have to remain in parallel corridors that prevent an active point to win over the others if located in other corridors, with the effect that neighbouring parts of the target cannot cooperate to bring the tree to them. Even if convergence
Figure 9. Growing a search tree in $\mathbb{R}^2$ towards a dual target made of two line segments. While reaching the first piece of target, separatrices are generated that confine the growth to the second piece of target to narrow corridors.

happened—for instance, in high dimension—this phenomenon might delay it and give a poor structure to the search tree, enough to be a major drawback for the algorithm.

5. Uninformed adaptive trees and their Darwinian interpretation

(a) Algorithm

The algorithm differs from the informed algorithm only by the accretion rule, which we call uninformed. We shall study the behaviour of a particular algorithm in $\mathbb{R}^n$ but shall state the convergence theorem for a wider range of settings and algorithm variants, including algorithms on Riemannian manifolds. The algorithm is as follows:

(i) set $t = 0$
(ii) repeat:
   — increment $t$
   — randomly draw a point $c_t$ from $T$ using $P$
   — compute the point $a_t$ in $G_{t-1}$ which minimizes $d(c_t, a_t)$, possibly using a randomized rule to avoid ties
   — set $G_t = G_{t-1} \cup \{b_t\}$, where $b_t$ is computed from $a_t$ and an independent random variable according to the uninformed accretion rule: draw randomly $b_t$ using a uniform distribution on the sphere of radius $\epsilon$ centred at $a_t$
(iii) until some condition on $t$ or $G_t$ is met.
A major difference brought by this new rule is the fact that an accreted point $b_t$ is not necessarily active at the time of its accretion. Indeed, it happens that $d(c_t, b_t) > d(c_t, a_t)$ and when implementing the algorithm one should check and eliminate at once those points which are born inactive. However, the probability for an accretion to be efficient is not less than a certain value which depends here, where $E$ is Euclidean, on the dimension $n$. This property is essential to the convergence of the process and to extend our convergence result to more general spaces and accretion rules, we need only require a minimal probability for accretions to be efficient.

(b) Numerical results

Figure 10 shows the growth of the uninformed accretion process towards a line segment (figure 10a) and the same dual target (figure 10b) as in figure 2.

(c) Convergence

Let us say that an uninformed accretion rule, where $b_t(\omega)$ is a random variable depending only on $a_t$, has property $P_1(\alpha, p, \epsilon)$ if there exist positive real numbers $\alpha$, $p$ and $\epsilon < \alpha$ such that, for any $c$ in $T$ and any $a_t$ in $E$ with $d(c, a_t) \geq \alpha$, the probability that $d(c, b_t) < d(c, a_t) - \epsilon$ is not less than $p$. For instance, one can check that property $P_1(2\epsilon, 1/3, \epsilon/4)$ is true in the case where $E = \mathbb{R}^2$ and $b_t$ is drawn from a uniform distribution on the circle of centre $a_t$ and radius $\epsilon$.

**Theorem 5.1.** Let $(E, d)$ be a metric space, $G_0$ a finite subset of $E$ and $T$ a compact subset of $E$ which is the support of a probability measure $P_T$ on $E$, with $c_t$, $t$ in $\mathbb{N}$, random points independently drawn from $T$ according to $P_T$. Let also
be given, for each $t$ in $\mathbb{N}$ and each $a_t$ in $E$, a random variable $b_t$ independent, conditionally to $a_t$, of the $c_i$’s and the random variables used to avoid distance ties. Let us assume that property $P_1(\alpha, p, \epsilon)$ holds, i.e. that there exist positive real numbers $\alpha$, $p$ and $\epsilon < \alpha$ such that, for any $c$ in $T$ and $a_t$ in $E$ with $d(c, a_t) > \alpha$, the probability that $d(c, b_t) < d(c, a_t) - \epsilon$ is not less than $p$. Then almost surely there exists a finite time $t$ from which the sets $G_t$, $t' \geq t$, grown from $G_0$ according to the uninformed adaptive tree algorithm will simultaneously have the two properties

(i) for any $c$ in $T$, there exists $x$ in $G_t$ such that $d(c, x) < \alpha + 5\epsilon/4$, i.e. no point of the target is more than $\alpha + 5\epsilon/4$ away from $G_t$ and

(ii) for any $x$ in $G_t$, for any $c$ in $T$, if $x$ is in $\Gamma_t(c)$, then $d(x, c) < \alpha + 5\epsilon/4$, i.e. active points are not further than $\alpha + 5\epsilon/4$ away from any point of their target shares.

Proof. The lemmas and consequences stated in §3 for theorem 3.5 are still true here except lemma 3.6 which we replace by the following.

Lemma 5.2. Let $A \subset T$ be contained in the open ball $B(e, \delta/2)$ of centre $e$ and radius $\delta/2$. We assume that the accretion process satisfies: there exist positive real numbers $\alpha > 2\delta$ and $p$ such that, for any $c$ in $T$ and $a_t$ in $E$ with $d(c, a_t) > \alpha$, the probability that $d(c, b_t) < d(c, a_t) - 2\delta$ is not less than $p$. Then almost surely there exists $N$ in $\mathbb{N}$ such that for any $n > N$

$$\sup_{c \in A} d(c, G_n) < \alpha + \frac{5\delta}{2}.$$ 

Proof. Define the open balls $B_1 = B(e, \delta/2)$, $B_2 = B(e, \delta/2 + \alpha)$, $B_3 = B(e, \alpha + 3\delta)$. One can then check, as in the proof of lemma 3.6, that

(i) for any $c$ in $B_1$, for any $x$ not in $B_2$, $d(c, x) > \alpha$, thus if all the points active for $A$ are outside $B_2$, the $A$-steps have a probability not less than $p$ of being efficient;

(ii) for any $c$ in $B_1$, for any $x$ in $B_3$, $d(c, x) < \alpha + 7\delta/2$, so in order to reach the conclusion of lemma 5.2, it is enough that all the active points for $A$ are in $B_3$;

(iii) for any $c, c'$ in $B_1$, for any $x$ not in $B_3$, for any $x'$ in $B_2$, $d(c, x') - d(c', x') > \delta$, hence one active point being in $B_2$ is enough for all the active points for $A$ to be in $B_3$; and

(iv) for any $c$ in $B_1$, for any $x$ in $E$, $d(c, x) < \alpha$ implies that $x$ is in $B_2$; thus, for $x$ to be in $B_2$, it is enough that $d(c, x) < \alpha$ for some $c$.

Find the smallest integer $k$ such that $\sup_{c \in T} d(c, G_0) - k\delta < \alpha$. We know from corollary 3.4 that at most $k$ efficient steps are necessary to achieve

$$d(c_{t_k}, b_{t_k}) < \sup_{c \in T} d(c, G_0) - k\delta < \alpha$$

and thus to have $b_{t_k}$ in $B_2$. At that point, either $b_{t_k}$ is active for $A$ or some other point active for $A$ is in $B_2$, so that all the points active for $A$ are in $B_3$; hence, for any $c$ in $A$ and any $x$ in $B_3$, $d(c, x) < \alpha + 7\delta/2$. 

Phil. Trans. R. Soc. A (2009)
Any $A$-step has a probability not less than $p$ of being efficient, and the efficiencies of these steps depend only on the values of $b_t$ given $a_t$, which are independent. Let $(X_t)_{t \in \mathbb{N}}$ be the sequence of random variables such that $X_t = 1$ if $c_t$ is drawn from $A$ and the corresponding step is efficient, and $X_t = 0$ otherwise. These random variables are independent identically distributed with $P(X_t = 1) > P(A) \cdot p > 0$ for any $t$, obviously making the series $\sum_{t=1}^{\infty} P(X_t = 1)$ diverge. The Borel–Cantelli lemma tells us that almost surely an infinite number, and in particular at least $k$, of these events occur.

To prove the theorem, again cover $T$ with open balls of radius $\delta/2 = \epsilon/4$ and using the compacity of $T$ select from them $B(e_1, \delta/2), \ldots, B(e_m, \delta/2)$ which still cover $T$, then set for $0 < i < m$, $A_i = T \cap B(e_i, \delta/2)$. For any $0 < i < m$, $P_t(A_i)$ is positive as $T$ is the support of $P_T$, and for any $c, c'$ in $A_i$, $d(c, c') < \delta$. Lemma 5.2 shows that for any of these $A_i$, there almost surely exists $N_i$ such that for $n > N_i$, $\sup_{c \in A} d(c, G_n) < \alpha + 5\delta/2 = \alpha + 5\epsilon/4$. The conjunction of this finite set of almost sure properties is still almost sure and we can take $N = \max(N_1, \ldots, N_m)$.

(d) Behaviour near branching

Simulations show, like in figure 11, that the uninformed accretion process generates unsuccessful branchings analogous to those seen for the informed accretion process. One difference, however, is that the competition usually involves, instead of the two competing tips in the informed accretion, two clusters of several active points.

(e) Interpretations, Darwinian evolution

It has been a strong requirement in the neo-Darwinian paradigm that genetic variations—such as mutations—blindly modify the genome, after which selection decides which individuals or populations (or more complex entities) survive and thrive. We propose that uninformed accretions of the process just described model such genetic variations: once the $a_t$ has been chosen for the next attempt at an accretion, the tentative $b_t$ is drawn at random in its vicinity. We experimented with ‘mutants’ drawn from a sphere or a ball centred at $a_t$, but a large range of possibilities are permitted by the proof of the convergence theorem. Even a small imbalance still permits adaptive convergence, and indeed the proof of the convergence theorem does not require symmetry.

We must also examine how the other step—drawing a random $c_t$ from $T$ and deducing $a_t$ where the next tentative accretion will take place—can be given an evolutionary interpretation to justify the whole algorithm in that context, especially since that step uses some information about the target and could be
seen as a breach in the neo-Darwinian paradigm. Let us suppose that the active points correspond to some populations fitted to a manifold of niches constituting the target set, and that species are more adapted to a point in the target if their representative points are closer to it. We can assume that the size reached by the population of a thriving species is roughly proportional to the amount of target where it dominates all other species (its target share in our formalism). We can also assume that more genetic variations arise from large populations, citing Darwin: ‘we have also seen that the species, which are the commonest and the most widely diffused, vary more than do the rare and restricted species’ (Darwin 1859). Thus, drawing a point from the target and finding the closest member of the growing set, which gives more mutation opportunities to active points with a large target share, can be viewed as a way to express in the model that species that are winners for a great extent of target give rise to more numerous populations which consequently have more opportunities for mutations. The former interpretation is still very simplistic in terms of population genetics. Providing a complete population model compatible with these assumptions would be a way to base more firmly these ideas. However, we keep as an objective to develop explanatory mechanisms at that intermediary description level of trees.

These ideas also apply to the biological phenomena we referred to for the informed algorithm and could extend the range of applicability of some informed models to settings like neural Darwinism (Edelman 1987) where selection does not necessarily follow from classical population dynamics: even if some informed accretion process needs too complex integration mechanisms to be implemented, some uninformed version of it, with lighter requirements, could be postulated.

References


