

---

# Tree-Independent Dual-Tree Algorithms

---

Ryan R. Curtin  
William B. March  
Parikshit Ram  
David V. Anderson  
Alexander G. Gray  
Charles L. Isbell, Jr.

RCURTIN@CC.GATECH.EDU  
MARCH@GATECH.EDU  
P.RAM@GATECH.EDU  
ANDERSON@GATECH.EDU  
AGRAY@CC.GATECH.EDU  
ISBELL@CC.GATECH.EDU

Georgia Institute of Technology, 266 Ferst Drive NW, Atlanta, GA 30332 USA

## Abstract

*Dual-tree algorithms* are a widely used class of branch-and-bound algorithms. Unfortunately, developing dual-tree algorithms for use with different trees and problems is often complex and burdensome. We introduce a four-part logical split: the tree, the traversal, the point-to-point base case, and the pruning rule. We provide a meta-algorithm which allows development of dual-tree algorithms in a tree-independent manner and easy extension to entirely new types of trees. Representations are provided for five common algorithms; for  $k$ -nearest neighbor search, this leads to a novel, tighter pruning bound. The meta-algorithm also allows straightforward extensions to massively parallel settings.

## 1. Introduction

In large-scale machine learning applications, algorithmic scalability is paramount. Hence, much study has been put into fast algorithms for machine learning tasks. One commonly-used approach is to build trees on data and then use branch-and-bound algorithms to minimize runtime. A popular example of a branch-and-bound algorithm is the use of the trees for nearest neighbor search, pioneered by Bentley (1975), and subsequently modified to use two trees (“dual-tree”) (Gray & Moore, 2001). Later, an optimized tree structure, the cover tree, was designed (Beygelzimer et al., 2006), giving provably linear scaling in the number of queries (Ram et al., 2009)—a significant improvement over the quadratically-scaling brute-force algorithm.

Asymptotic speed gains as dramatic as described above are common for dual-tree branch-and-bound algorithms. These types of algorithms can be applied to a class of problems referred to as ‘ $n$ -body problems’ (Gray & Moore, 2001). The  $n$ -point correlation,

important in astrophysics, is an  $n$ -body problem and can be solved quickly with trees (March et al., 2012). In addition, Euclidean minimum spanning trees can be found quickly using tree-based algorithms (March et al., 2010). Other dual-tree algorithms include kernel density estimation (Gray & Moore, 2003a), mean shift (Wang et al., 2007), Gaussian summation (Lee & Gray, 2006), kernel density estimation, fast singular value decomposition (Holmes et al., 2008), range search, furthest-neighbor search, and many others.

The dual-tree algorithms referenced above are each quite similar, but no formal connections between the algorithms have been established. The types of trees used to solve each problem may differ, and in addition, the manner in which the trees are traversed can differ (depending on the problem or the tree). In practice, a researcher may have to implement entirely separate algorithms to solve the same problems with different trees; this is time-consuming and makes it difficult to explore the properties of tree types that make them more suited for particular problems. Worse yet, parallel dual-tree algorithms are difficult to develop and appear to be far more complex than serial implementations; yet, both solve the same problem. We make these contributions to address these shortcomings:

- A **representation** of dual-tree algorithms as **four separate components**: a space tree, a traversal, a base case, and a pruning rule.
- A **meta-algorithm** that produces dual-tree algorithms, given those four separate components.
- Base cases and pruning rules for a **variety of dual-tree algorithms**, which can be used with **any** space tree and **any** traversal.
- A **theoretical framework**, used to prove the correctness of these meta-algorithms and develop a new, tighter bound for  $k$ -nearest neighbor search.
- Implications of our representation, including **easy creation of large-scale distributed dual-tree algorithms** via our meta-algorithm.

## 2. Overview of Meta-Algorithm

In other works, dual-tree algorithms are described as standalone algorithms that operate on a query dataset  $S_q$  and a reference dataset  $S_r$ . By observing commonalities in these algorithms, we propose the following logical split of any dual-tree algorithm into four parts:

- A *space tree* (a type of data structure).
- A *pruning dual-tree traversal*, which visits nodes in two space trees, and is parameterized by a `BaseCase()` and a `Score()` function.
- A `BaseCase()` function that defines the action to take on a combination of points.
- A `Score()` function that determines if a subtree should be visited during the traversal.

We can use this to define a **meta-algorithm**:

*Given a type of space tree, a pruning dual-tree traversal, a `BaseCase()` function, and a `Score()` function, use the pruning dual-tree traversal with the given `BaseCase()` and `Score()` functions on two space trees  $\mathcal{T}_q$  (built on  $S_q$ ) and  $\mathcal{T}_r$  (built on  $S_r$ ).*

In Sections 3 and 4, space trees, traversals, and related quantities are rigorously defined. Then, Sections 5–8 define `BaseCase()` and `Score()` for various dual-tree algorithms. Section 9 discusses implications and future possibilities, including large-scale parallelism.

## 3. Space Trees

To develop a framework for understanding dual-tree algorithms, we must introduce some terminology.

**Definition 1.** A *space tree* on a dataset  $S \in \mathbb{R}^{N \times D}$  is an undirected, connected, acyclic, rooted simple graph with the following properties:

- Each node (or vertex), holds a number of points (possibly zero) and is connected to one parent node and a number of child nodes (possibly zero).
- There is one node in every space tree with no parent; this is the root node of the tree.
- Each point in  $S$  is contained in at least one node.
- Each node  $\mathcal{N}$  has a convex subset of  $\mathbb{R}^D$  containing each point in that node and also the convex subsets represented by each child of the node.

Notationally, we use the following conventions:

- The set of child nodes of a node  $\mathcal{N}_i$  is denoted  $\mathcal{C}(\mathcal{N}_i)$  or  $\mathcal{C}_i$ .

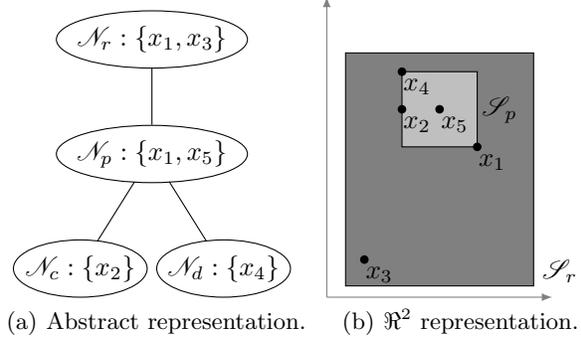


Figure 1. An example space tree.

- The set of points held in a node  $\mathcal{N}_i$  is denoted  $\mathcal{P}(\mathcal{N}_i)$  or  $\mathcal{P}_i$ .
- The convex subset represented by node  $\mathcal{N}_i$  is denoted  $\mathcal{S}(\mathcal{N}_i)$  or  $\mathcal{S}_i$ .
- The set of descendant nodes of a node  $\mathcal{N}_i$ , denoted  $\mathcal{D}^n(\mathcal{N}_i)$  or  $\mathcal{D}_i^n$ , is the set of nodes  $\mathcal{C}(\mathcal{N}_i) \cup \mathcal{C}(\mathcal{C}(\mathcal{N}_i)) \cup \dots$ <sup>1</sup>.
- The set of descendant points of a node  $\mathcal{N}_i$ , denoted  $\mathcal{D}^p(\mathcal{N}_i)$  or  $\mathcal{D}_i^p$ , is the set of points  $\{p : p \in \mathcal{P}(\mathcal{D}^n(\mathcal{N}_i)) \cup \mathcal{P}(\mathcal{N}_i)\}$ <sup>2</sup>.
- The parent of a node  $\mathcal{N}_i$  is denoted  $\text{Par}(\mathcal{N}_i)$ .

An abstract representation of an example space tree on a five-point dataset in  $\mathbb{R}^2$  is shown in Figure 1(a). In this illustration,  $\mathcal{N}_r$  is the root node of the tree; it has no parent and it contains the points  $x_3$  and  $x_1$  (that is,  $\mathcal{P}_r = \{x_1, x_3\}$ ). The node  $\mathcal{N}_p$  contains points  $x_1$  and  $x_5$  and has children  $\mathcal{N}_c$  and  $\mathcal{N}_d$  (which each have no children and contain points  $x_2$  and  $x_4$ , respectively). In Figure 1(b), the points in the tree and the subsets  $\mathcal{S}_r$  (darker rectangle) and  $\mathcal{S}_p$  (lighter rectangle) are plotted.  $\mathcal{S}_c = \{x_2\}$  and  $\mathcal{S}_d = \{x_4\}$  are not labeled.

**Definition 2.** The *minimum distance* between two nodes  $\mathcal{N}_i$  and  $\mathcal{N}_j$  is defined as

$$d_{\min}(\mathcal{N}_i, \mathcal{N}_j) = \min \{ \|p_i - p_j\| \mid \forall p_i \in \mathcal{D}_i^p, p_j \in \mathcal{D}_j^p \}.$$

**Definition 3.** The *maximum distance* between two nodes  $\mathcal{N}_i$  and  $\mathcal{N}_j$  is defined as

$$d_{\max}(\mathcal{N}_i, \mathcal{N}_j) = \max \{ \|p_i - p_j\| \mid \forall p_i \in \mathcal{D}_i^p, p_j \in \mathcal{D}_j^p \}.$$

**Definition 4.** The *maximum child distance* of a node  $\mathcal{N}_i$  is defined as the maximum distance between the centroid  $C_i$  of  $\mathcal{S}_i$  and each point in  $\mathcal{P}_i$ :

$$\rho(\mathcal{N}_i) = \max_{p \in \mathcal{P}_i} \|C_i - p\|.$$

<sup>1</sup>By  $\mathcal{C}(\mathcal{C}(\mathcal{N}_i))$  we mean all the children of the children of node  $\mathcal{N}_i$ :  $\mathcal{C}(\mathcal{C}(\mathcal{N}_i)) = \{\mathcal{C}(\mathcal{N}_c) : \mathcal{N}_c \in \mathcal{C}(\mathcal{N}_q)\}$ .

<sup>2</sup>The meaning of  $\mathcal{P}(\mathcal{D}^n(\mathcal{N}_i))$  is similar to  $\mathcal{C}(\mathcal{C}(\mathcal{N}_i))$ .

**Definition 5.** The *maximum descendant distance* of a node  $\mathcal{N}_i$  is defined as the maximum distance between the centroid  $C_i$  of  $\mathcal{S}_i$  and points in  $\mathcal{D}_i^p$ :

$$\lambda(\mathcal{N}_i) = \max_{p \in \mathcal{D}_i^p} \|C_i - p\|.$$

It is straightforward to show that  $kd$ -trees, octrees, metric trees, ball trees, cover trees (Beygelzimer et al., 2006), R-trees, and vantage-point trees all satisfy the conditions of a space tree. The quantities  $d_{\min}(\mathcal{N}_q, \mathcal{N}_r)$ ,  $d_{\max}(\mathcal{N}_q, \mathcal{N}_r)$ ,  $\lambda(\mathcal{N}_i)$ , and  $\rho(\mathcal{N}_i)$  are easily derived (or bounded, which in many cases is sufficient) for each of these types of trees.

For a  $kd$ -tree,  $d_{\min}(\mathcal{N}_i, \mathcal{N}_j)$  is bounded below by the minimum distance between  $\mathcal{S}_i$  and  $\mathcal{S}_j$ ;  $d_{\max}(\mathcal{N}_i, \mathcal{N}_j)$  is bounded above similarly by the maximum distance between  $\mathcal{S}_i$  and  $\mathcal{S}_j$ . Both  $\rho(\mathcal{N}_i)$  and  $\lambda(\mathcal{N}_i)$  are bounded above by  $d_{\max}(C_i, \mathcal{N}_i)$ .

For the cover tree (Beygelzimer et al., 2006), each node  $\mathcal{N}_i$  contains only one point  $p_i$  and has ‘scale’  $s_i$ .  $d_{\min}(\mathcal{N}_i, \mathcal{N}_j)$  is bounded below by  $d(p_i, p_j) - 2^{s_i+1} - 2^{s_j+1}$  and  $d_{\max}(\mathcal{N}_i, \mathcal{N}_j)$  is bounded above by  $d(p_i, p_j) + 2^{s_i+1} + 2^{s_j+1}$ . Because  $p_i$  is the centroid of  $\mathcal{S}_i$ ,  $\rho(\mathcal{N}_i) = 0$ .  $\lambda(\mathcal{N}_i)$  is simply  $2^{s_i+1}$ .

#### 4. Tree Traversals

In general, the nodes of each space tree can be traversed in a number of ways. However, there have been no attempts to formalize tree traversal. Therefore, we introduce several definitions which will be useful later.

**Definition 6.** A *single-tree traversal* is a process that, given a space tree, will visit each node in that tree once and perform a computation on any points contained within the node that is being visited.

As an example, the standard depth-first traversal or breadth-first traversal are single-tree traversals. From a programming perspective, the computation in the single-tree traversal can be implemented with a simple callback `BaseCase(point)` function. This allows the computation to be entirely independent of the single-tree traversal itself. As an example, a simple single-tree algorithm to count the number of points in a given tree would increment a counter variable each time `BaseCase(point)` was called. However, this concept by itself is not very useful; without pruning branches, no computations can be avoided.

**Definition 7.** A *pruning single-tree traversal* is a process that, given a space tree, will visit nodes in the tree and perform a computation to assign a score to that node. If the score is above some bound, the node is ‘pruned’ and none of its descendants will be visited; otherwise, a computation is performed on any points

---

**Algorithm 1** `DepthFirstTraversal`( $\mathcal{N}_q, \mathcal{N}_r$ ).

---

```

if Score( $\mathcal{N}_q, \mathcal{N}_r$ ) =  $\infty$  then return
for each  $p_q \in \mathcal{P}_q, p_r \in \mathcal{P}_r$  do
    BaseCase( $p_q, p_r$ )

for each  $\mathcal{N}_{qc} \in \mathcal{C}_q, \mathcal{N}_{rc} \in \mathcal{C}_r$  do
    DepthFirstTraversal( $\mathcal{N}_{qc}, \mathcal{N}_{rc}$ )

```

---

contained within that node. If no nodes are pruned, then the traversal will visit each node in the tree once.

Clearly, a pruning single-tree traversal that does not prune any nodes is a single-tree traversal. A pruning single-tree traversal can be implemented with two callbacks: `BaseCase(point)` and `Score(node)`. This allows both the point-to-point computation and the scoring to be entirely independent of the traversal. Thus, single-tree branch-and-bound algorithms can be expressed in a tree-independent manner. Extensions to the dual-tree case are given below.

**Definition 8.** A *dual-tree traversal* is a process that, given two space trees  $\mathcal{T}_q$  (query tree) and  $\mathcal{T}_r$  (reference tree), will visit every combination of nodes ( $\mathcal{N}_q, \mathcal{N}_r$ ) once, where  $\mathcal{N}_q \in \mathcal{T}_q$  and  $\mathcal{N}_r \in \mathcal{T}_r$ . At each visit ( $\mathcal{N}_q, \mathcal{N}_r$ ), a computation is performed between each point in  $\mathcal{N}_q$  and each point in  $\mathcal{N}_r$ .

**Definition 9.** A *pruning dual-tree traversal* is a process which, given two space trees  $\mathcal{T}_q$  (query tree) and  $\mathcal{T}_r$  (reference tree), will visit combinations of nodes ( $\mathcal{N}_q, \mathcal{N}_r$ ) such that  $\mathcal{N}_q \in \mathcal{T}_q$  and  $\mathcal{N}_r \in \mathcal{T}_r$  no more than once, and perform a computation to assign a score to that combination. If the score is above some bound, the combination is pruned and no combinations ( $\mathcal{N}_{qc}, \mathcal{N}_{rc}$ ) such that  $\mathcal{N}_{qc} \in \mathcal{D}_q^n$  and  $\mathcal{N}_{rc} \in \mathcal{D}_r^n$  will be visited; otherwise, a computation is performed between each point in  $\mathcal{N}_q$  and each point in  $\mathcal{N}_r$ .

Similar to the pruning single-tree traversal, a pruning dual-tree algorithm can use two callback functions `BaseCase( $p_q, p_r$ )` and `Score( $\mathcal{N}_q, \mathcal{N}_r$ )`. An example implementation of a depth-first pruning dual-tree traversal is given in Algorithm 1. The traversal is started on the root of the  $\mathcal{T}_q$  and the root of  $\mathcal{T}_r$ .

Algorithm 1 provides only one example of a commonly-used pruning dual-tree traversal. Other possibilities not explicitly documented here include breadth-first traversals and the unique cover tree dual-tree traversal described by Beygelzimer et al. (2006), which can be adapted to the generic space tree case.

The rest of this work is devoted to using these concepts to represent existing dual-tree algorithms in the four parts described in Section 2.

## 5. $k$ -Nearest Neighbor Search

$k$ -nearest neighbor search is a well-studied problem with a plethora of algorithms and results. The problem can be stated as follows:

Given a query dataset  $S_q \in \mathbb{R}^{n \times d}$ , a reference dataset  $S_r \in \mathbb{R}^{m \times d}$ , and an integer  $k : 0 < k < m$ , for each point  $p_q \in S_q$ , find the  $k$  nearest neighbors in  $S_r$  and their distances from  $p_q$ . The list of nearest neighbors for a point  $p_q$  can be referred to as  $N_{p_q}$  and the distances to nearest neighbors for  $p_q$  can be referred to as  $D_{p_q}$ . Thus, the  $k$ -th nearest neighbor to point  $p_q$  is  $N_{p_q}[k]$  and  $D_{p_q}[k] = \|p_q - N_{p_q}[k]\|$ .

This can be solved using a brute-force approach: compare every possible point combination and store the  $k$  smallest distance results for each  $p_q$ . However, this scales poorly –  $O(nm)$ ; hence the importance of fast algorithms to solve the problem. Many existing algorithms employ tree-based branch-and-bound strategies (Beygelzimer et al., 2006; Cover & Hart, 1967; Friedman et al., 1977; Fukunaga & Narendra, 1975; Gray & Moore, 2001; Ram et al., 2009).

We unify all of these branch-and-bound strategies by defining methods  $\text{BaseCase}(p_q, p_r)$  and  $\text{Score}(\mathcal{N}_q, \mathcal{N}_r)$  for use with a pruning dual-tree traversal.

At the initialization of the tree traversal, the lists  $N_{p_q}$  and  $D_{p_q}$  are empty lists for each query point  $p_q$ . After the traversal is complete, for a query point  $p_q$ , the set  $\{N_{p_q}[1], \dots, N_{p_q}[k]\}$  is the ordered set of  $k$  nearest neighbors of the point  $p_q$ , and each  $D_{p_q}[i] = \|p_q - N_{p_q}[i]\|$ . If we assume that  $D_{p_q}[i] = \infty$  if  $i$  is greater than the length of  $D_{p_q}$ , we can formulate  $\text{BaseCase}()$  as given in Algorithm 2<sup>3</sup>.

With the base case established, only the pruning rule remains. A valid pruning rule will, for a given query node  $\mathcal{N}_q$  and reference node  $\mathcal{N}_r$ , prune the reference subtree rooted at  $\mathcal{N}_r$  if and only if it is known that there are no points in  $\mathcal{D}_r^p$  that are in the set of  $k$  nearest neighbors of any points in  $\mathcal{D}_q^p$ . Thus, at any point in the traversal, we can prune the combination  $(\mathcal{N}_q, \mathcal{N}_r)$  if and only if  $d_{\min}(\mathcal{N}_q, \mathcal{N}_r) \geq B_1(\mathcal{N}_q)$ , where

$$B_1(\mathcal{N}_q) = \max_{p \in \mathcal{D}_q^p} D_p[k].$$

Now, we can describe this bound recursively. This is important for implementation; a recursive function

<sup>3</sup> In practice,  $k$ -nearest-neighbors is often run with identical reference and query sets. In that situation it may be useful to modify this implementation of  $\text{BaseCase}()$  so that a point does not return itself as the nearest neighbor (with distance 0).

---

### Algorithm 2 $k$ -nearest-neighbors $\text{BaseCase}()$

---

**Input:** query point  $p_q$ , reference point  $p_r$ , list of  $k$  nearest candidate points  $N_{p_q}$  and  $k$  candidate distances  $D_{p_q}$  (both ordered by ascending distance)

**Output:** distance  $d$  between  $p_q$  and  $p_r$

$d \leftarrow \|p_q - p_r\|$

**if**  $d < D_{p_q}[k]$  **and**  $\text{BaseCase}(p_q, p_r)$  not yet called **then**

insert  $d$  into ordered list  $D_{p_q}$  and truncate list to length  $k$

insert  $p_r$  into  $N_{p_q}$  such that  $N_{p_q}$  is ordered by distance and truncate list to length  $k$

**return**  $d$

---

can cache previous calculations for large speedups.

$$\begin{aligned} B_1(\mathcal{N}_q) &= \max \left\{ \max_{p \in \mathcal{D}_q^p} D_p[k], \max_{p \in \mathcal{D}_q^p, p \notin \mathcal{D}_q^p} D_p[k] \right\} \\ &= \max \left\{ \max_{p \in \mathcal{D}_q^p} D_p[k], \max_{\mathcal{N}_c \in \mathcal{C}_q} \left\{ \max_{p \in \mathcal{D}_c^p} D_p[k] \right\} \right\} \\ &= \max \left\{ \max_{p \in \mathcal{D}_q^p} D_p[k], \max_{\mathcal{N}_c \in \mathcal{C}_q} B_1(\mathcal{N}_c) \right\} \end{aligned}$$

Suppose we have, at some point in the traversal, two points  $p_0, p_1 \in \mathcal{D}_q^p$  for some node  $\mathcal{N}_q$ , with  $D_{p_0}[k] = \infty$  and  $D_{p_1}[k] < \infty$ . This means there exist  $k$  points  $\{p_r^1, \dots, p_r^k\}$  in  $S_r$  such that  $d(p_1, p_r^i) \leq D_{p_1}[k]$  for  $i = \{1, \dots, k\}$ . Because  $p_0, p_1 \in \mathcal{D}_q^p$ , we can apply the triangle inequality to see that  $d(p_0, p_1) \leq 2\lambda(\mathcal{N}_q)$ . Therefore,  $d(p_0, p_r^i) \leq D_{p_1}[k] + 2\lambda(\mathcal{N}_q)$  for  $i = \{1, \dots, k\}$ . Using this observation we can construct an alternate bound function  $B_2(\mathcal{N}_q)$ :

$$B_2(\mathcal{N}_q) = \min_{p \in \mathcal{D}_q^p} D_p[k] + 2\lambda(\mathcal{N}_q)$$

which can, like  $B_1(\mathcal{N}_q)$ , be rearranged to provide a recursive definition. In addition, if  $p_0 \in \mathcal{D}_q^p$  and  $p_1 \in \mathcal{D}_q^p$ , we can bound  $d(p_0, p_1)$  more tightly with  $\rho(\mathcal{N}_q) + \lambda(\mathcal{N}_q)$  instead of  $2\lambda(\mathcal{N}_q)$ . These observations yield

$$B_2(\mathcal{N}_q) = \min \left\{ \min_{p \in \mathcal{D}_q^p} (D_p[k] + \rho(\mathcal{N}_q) + \lambda(\mathcal{N}_q)), \min_{\mathcal{N}_c \in \mathcal{C}_q} (B_2(\mathcal{N}_c) + 2(\lambda(\mathcal{N}_q) - \lambda(\mathcal{N}_c))) \right\}.$$

Both  $B_1(\mathcal{N}_q)$  and  $B_2(\mathcal{N}_r)$  provide valid pruning rules. We can combine both to get a tighter pruning rule by taking the tighter of the two bounds. In addition,  $B_1(\mathcal{N}_q) \geq B_1(\mathcal{N}_c)$  and  $B_2(\mathcal{N}_q) \geq B_2(\mathcal{N}_c)$  for all  $\mathcal{N}_c \in \mathcal{C}_q$ . Therefore, we can prune  $(\mathcal{N}_q, \mathcal{N}_r)$  if  $d_{\min}(\mathcal{N}_q, \mathcal{N}_r) \geq \min\{B_1(\text{Par}(\mathcal{N}_q)), B_2(\text{Par}(\mathcal{N}_q))\}$ .

---

**Algorithm 3**  $k$ -nearest-neighbors `Score()`


---

**Input:** query node  $\mathcal{N}_q$ , reference node  $\mathcal{N}_r$   
**Output:** a score for the node combination  $(\mathcal{N}_q, \mathcal{N}_r)$ , or  $\infty$  if the combination should be pruned  
**if**  $d_{\min}(\mathcal{N}_q, \mathcal{N}_r) < B(\mathcal{N}_q)$  **then**  
     **return**  $d_{\min}(\mathcal{N}_q, \mathcal{N}_r)$   
**return**  $\infty$

---

These observations are combined for a better bound:

$$B(\mathcal{N}_q) = \min \left\{ \max \left\{ \max_{p \in \mathcal{P}_q} D_p[k], \max_{\mathcal{N}_c \in \mathcal{C}_q} B(\mathcal{N}_c) \right\}, \right. \\ \min_{p \in \mathcal{P}_q} (D_p[k] + \rho(\mathcal{N}_q) + \lambda(\mathcal{N}_q)), \\ \left. \min_{\mathcal{N}_c \in \mathcal{C}_q} \left( B(\mathcal{N}_c) + 2(\lambda(\mathcal{N}_q) - \lambda(\mathcal{N}_c)) \right) \right. \\ \left. B(\text{Par}(\mathcal{N}_q)) \right\}.$$

As a result of this bound function being expressed recursively, previous bounds can be cached and used to calculate the bound  $B(\mathcal{N}_q)$  quickly. We can use this to structure `Score()` as given in Algorithm 3.

Applying the meta-algorithm in Section 2 with any tree type and any pruning dual-tree traversal gives a correct implementation of  $k$ -nearest neighbor search. Proving the correctness is straightforward; first, a (non-pruning) dual-tree traversal which uses `BaseCase()` as given in Algorithm 2 will give correct results for any space tree. Then, we already know that  $B(\mathcal{N}_q)$  is a bounding function that, at any point in the traversal, will not prune any subtrees which could contain better nearest neighbor candidates than the current candidates. Thus, the true nearest neighbors for each query point will always be visited, and the results will be correct.

We now show that this algorithm is a generalization of the standard  $kd$ -tree  $k$ -NN search, which uses a pruning dual-tree depth-first traversal. The archetypal algorithm for all-nearest neighbor search ( $k$ -nearest neighbor search with  $k = 1$ ) given for  $kd$ -trees in Alex Gray’s Ph.D. thesis (2003) is shown here in Algorithm 4 with converted notation.  $\delta_q^{nn}$  is the bound for a node  $\mathcal{N}_q$  and is initialized to  $\infty$ ;  $D_{p_q}$  represents the nearest distance for a query point  $p_q$ , and  $N_{p_q}$  represents the nearest neighbor for a query point  $p_q$ .  $\mathcal{N}_q.\text{left}$  represents the left child of  $\mathcal{N}_q$  and is defined to be  $\mathcal{N}_q$  if  $\mathcal{N}_q$  has no children;  $\mathcal{N}_q.\text{right}$  is similarly defined.

The structure of the algorithm matches Algorithm 1; it is a dual-tree depth-first recursion. Because this is a depth-first recursion,  $\delta_q^{nn} = \infty$  for a node  $\mathcal{N}_q$  if no descendants of  $\mathcal{N}_q$  have been recursed into. Otherwise,

---

**Algorithm 4** `AllNN`( $\mathcal{N}_q, \mathcal{N}_r$ ) (Gray, 2003)
 

---

**if**  $d_{\min}(\mathcal{N}_q, \mathcal{N}_r) \geq \delta_q^{nn}$ , **then return**  
**if**  $\mathcal{N}_q$  is leaf **and**  $\mathcal{N}_r$  is leaf **then**  
     **for each**  $p_q \in \mathcal{P}_q, p_r \in \mathcal{P}_r$  **do**  
          $d_{qr} \leftarrow \|p_q - p_r\|$ .  
         **if**  $d_{qr} < D_{p_q}$  **then**  $D_{p_q} = d_{qr}; N_{p_q} = p_r$   
         **if**  $d_{qr} < \delta_q^{nn}$  **then**  $\delta_q^{nn} \leftarrow d_{qr}$   
     `AllNN`( $\mathcal{N}_q.\text{left}$ , `closer-of`( $\mathcal{N}_r.\text{left}$ ,  $\mathcal{N}_r.\text{right}$ ))  
     `AllNN`( $\mathcal{N}_q.\text{left}$ , `farther-of`( $\mathcal{N}_r.\text{left}$ ,  $\mathcal{N}_r.\text{right}$ ))  
     `AllNN`( $\mathcal{N}_q.\text{right}$ , `closer-of`( $\mathcal{N}_r.\text{left}$ ,  $\mathcal{N}_r.\text{right}$ ))  
     `AllNN`( $\mathcal{N}_q.\text{right}$ , `farther-of`( $\mathcal{N}_r.\text{left}$ ,  $\mathcal{N}_r.\text{right}$ ))  
      $\delta_q^{nn} = \min(\delta_q^{nn}, \max(\delta_{q.\text{left}}^{nn}, \delta_{q.\text{right}}^{nn}))$

---

$\delta_q^{nn}$  is the maximum of  $D_{p_q}$  for all  $p_q \in \mathcal{P}_q$ . That is,  $\delta_q^{nn} = B_1(\mathcal{N}_q)$ . Thus, the comparison in the first line of Algorithm 4 is equivalent to Algorithm 3 with  $B_1(\mathcal{N}_q)$  instead of  $B(\mathcal{N}_q)$ .

The first two lines of the inside of the **for each** loop (the base case) are equivalent to Algorithm 2 with  $k = 1$ .  $kd$ -trees only hold points in leaves; therefore, the base case is called for all combinations of points in each node combination, identical to the depth-first traverser (Algorithm 1).

A  $kd$ -tree is a space tree and the dual depth-first recursion is a pruning dual-tree traversal. Also, we showed the equivalency of the pruning rule (that is, the `Score()` function) and the equivalency of the base case. So, it is clear that Algorithm 4 is produced using our meta-algorithm with these parameters. In addition, because  $B(\mathcal{N}_q)$  is always less than  $B_1(\mathcal{N}_q)$ , Algorithm 3 provides a **tighter bound** than the pruning rule in Algorithm 4.

This algorithm is also a generalization of the standard cover tree  $k$ -NN search (Beygelzimer et al., 2006). The cover tree search is a pruning dual-tree traversal where the query tree is traversed depth-first while the reference tree is simultaneously traversed breadth-first. The pruning rule (after simple adaptation to the  $k$ -nearest-neighbor search problem instead of the nearest-neighbor search problem) is equivalent to

$$d_{\min}(\mathcal{N}_q, \mathcal{N}_r) \geq D_{p_q}[k] + \lambda(\mathcal{N}_q)$$

where  $p_q$  is the point contained in  $\mathcal{N}_q$  (remember, each node of a cover tree contains one point). This is equivalent to  $B_2(\mathcal{N}_q)$  because  $\rho(\mathcal{N}_q) = 0$  for cover trees. The transformation from the algorithm given by Beygelzimer et al. (2006) to our representation is made clear in Appendix A (supplementary material) and in the  $k$ -nearest neighbor search implementation of the C++ library MLPACK (Curtin et al., 2011); this is implemented in terms of our meta-algorithm.

Specific algorithms for ball trees, metric trees, VP trees, octrees, and other space trees are trivial to create using the `BaseCase()` and `Score()` implementation given here (and in `MLPACK`). Note also that this implementation will work in any metric space.

An extension to  $k$ -furthest neighbor search is straightforward. The bound function must be ‘inverted’ by changing ‘max’ to ‘min’ (and vice versa); in addition, the distances  $D_{p_q}[i]$  must be initialized to 0 instead of  $\infty$ , and the lists  $D$  and  $N$  must be sorted by descending distance instead of ascending distance. Lastly, the comparison  $d < D_{p_q}[k]$  must be changed to  $d > D_{p_q}[k]$ . With these simple changes, we have easily solved an entirely different problem using our meta-algorithm. An implementation using our meta-algorithm for both  $kd$ -trees and cover trees is also available in `MLPACK`.

## 6. Range Search

Range search is another popular neighbor searching problem related to  $k$ -nearest neighbor search. In addition to being a fairly standard machine learning task, it has numerous uses in applications such as databases and geographic information systems (GIS). A treatise on the history of the problem and solutions is given by Agarwal & Erickson (1999). The problem is:

Given query and reference datasets  $S_q, S_r$  and a range  $[\delta_1, \delta_2]$ , for each point  $p_q \in S_q$ , find all points in  $S_r$  such that  $\delta_1 \leq \|p_q - p_r\| \leq \delta_2$ . As with  $k$ -nearest neighbor search, refer to the list of neighbors for each query point  $p_q$  as  $N_{p_q}$  and the corresponding distances as  $D_{p_q}$ . These lists are not sorted in any particular order, and at initialization time, they are empty.

In different settings, the problem of range search may not be stated identically; however, our results are easily adaptable. A `BaseCase()` implementation is given in Algorithm 5, and a `Score()` implementation is given in Algorithm 6. The only bounds to consider are  $[\delta_1, \delta_2]$ , so no complex bound handling is necessary.

While range search is sometimes mentioned in the context of dual-tree algorithms (Gray & Moore, 2001), the focus is usually on  $k$ -nearest neighbor search. As a result, we cannot find any explicitly published dual-tree algorithms to generalize; however, a single-tree algorithm was proposed by Bentley and Friedman (1979). Thus, the `BaseCase()` and `Score()` proposed here can be used with our meta-algorithm to produce entirely novel range search implementations; `MLPACK` has  $kd$ -tree and cover tree implementations.

## 7. Borůvka’s Algorithm

Finding a Euclidean minimum spanning tree has been a relevant problem since Borůvka’s algorithm was

---

### Algorithm 5 Range search `BaseCase()`.

---

**Input:** query point  $p_q$ , reference point  $p_r$ , neighbor list  $N_{p_q}$ , distance list  $D_{p_q}$

**Output:** distance  $d$  between  $p_q$  and  $p_r$

$d \leftarrow \|p_q - p_r\|$

**if**  $\delta_1 \leq d \leq \delta_2$  **and** `BaseCase`( $p_q, p_r$ ) not yet called **then**

$N_{p_q} \leftarrow N_{p_q} \cup \{p_r\}$

$D_{p_q} \leftarrow D_{p_q} \cup \{d\}$

**return**  $d$

---



---

### Algorithm 6 Range search `Score()`.

---

**Input:** query node  $\mathcal{N}_q$ , reference node  $\mathcal{N}_r$

**Output:** a score for  $(\mathcal{N}_q, \mathcal{N}_r)$ , or  $\infty$  if the combination should be pruned

**if**  $\delta_1 \leq d_{\min}(\mathcal{N}_q, \mathcal{N}_r) \leq \delta_2$  **then**

**return**  $d_{\min}(\mathcal{N}_q, \mathcal{N}_r)$

**return**  $\infty$

---

proposed in 1926. Recently, a dual-tree version of Borůvka’s algorithm was developed (March et al., 2010) for  $kd$ -trees and cover trees. We unify these two algorithms and generalize to other types of space tree by formulating `BaseCase()` and `Score()` functions.

For a dataset  $S_r \in \mathfrak{R}^{N \times D}$ , Borůvka’s algorithm connects each point to its nearest neighbor, giving many ‘components’. For each component  $c$ , the nearest point in  $S_r$  to any point of  $c$  that is not part of  $c$  is found. The points are connected, combining those components. This process repeats until only one component—the minimum spanning tree—remains.

During the algorithm, we maintain a list  $F$  made up of  $i$  components  $F_i : \{E_i, V_i\}$  where  $E_i$  is the list of edges and  $V_i$  is the list of vertices in the component  $F_i$  (these are points in  $S_r$ ). Each point in  $S_r$  belongs to only one  $F_i$ . At initialization,  $|F| = |S_r|$  and  $F_i = \{\emptyset, \{p_i\}\}$  for  $i = \{1, \dots, |S_r|\}$ , where  $p_i$  is the  $i$ ’th point in  $S_r$ . For  $p \in S_r$  we define  $F(p) = F_i$  if  $F_i$  is the component containing  $p$ . During the algorithm, we maintain  $N(F_i)$  as the candidate nearest neighbor of component  $F_i$  and  $p_c(F_i)$  as the point in component  $F_i$  nearest to  $N(F_i)$ . Then,  $D(F_i) = \|p_c(F_i) - N(F_i)\|$ . Remember that  $F(N(F_i)) \neq F_i$ .

To run Borůvka’s algorithm with a space tree  $\mathcal{T}_r$  built on the set of points  $S_r$ , a pruning dual-tree traversal is run with `BaseCase()` as Algorithm 7, `Score()` as Algorithm 8,  $\mathcal{T}_r$  as *both* of the trees, and  $F$  as initialized before. Note that `Score()` uses  $B(\mathcal{N}_q)$  from Section 5 with  $k = 1$ . Upon traversal completion, we have a list  $N(F_i)$  of nearest neighbors of each component  $F_i$ . The

---

**Algorithm 7** Borůvka’s algorithm `BaseCase()`.

---

**Input:** query point  $p_q$ , reference point  $p_r$ , nearest candidate point  $N(F(p_q))$  and distance  $D(F(p_q))$

**Output:** distance  $d$  between  $p_q$  and  $p_r$

```

if  $p_q = p_r$  then
  return 0
if  $F(p_q) \neq F(p_r)$  and  $\|p_q - p_r\| < D(F(p_q))$  then
   $D(F(p_q)) \leftarrow \|p_q - p_r\|$ 
   $N(F(p_q)) \leftarrow p_r; p_c(F(p_q)) \leftarrow p_q$ 
return  $\|p_q - p_r\|$ 

```

---



---

**Algorithm 8** Borůvka’s algorithm `Score()`.

---

**Input:** query node  $\mathcal{N}_q$ , reference node  $\mathcal{N}_r$

**Output:** a score for the node combination  $(\mathcal{N}_q, \mathcal{N}_r)$ , or  $\infty$  if the combination should be pruned

```

if  $d_{\min}(\mathcal{N}_q, \mathcal{N}_r) < B(\mathcal{N}_q)$  then
  if  $F(p_q) = F(p_r) \forall p_q \in \mathcal{D}_q^p, p_r \in \mathcal{D}_r^p$  then
    return  $\infty$ 
  return  $d_{\min}(\mathcal{N}_q, \mathcal{N}_r)$ 
return  $\infty$ 

```

---

edge  $(N(F_i), p_c(F_i))$  is added to  $F_i$  for each  $F_i$ . Then, any components in  $F$  with shared edges are merged into a new list  $F'$  where  $|F'| < |F|$ . The pruning dual-tree traversal is then run again with  $F = F'$  and the traversal-merge process repeats until  $|F| = 1$ . When  $|F| = 1$ , then  $F_1$  is the minimum spanning tree of  $S_r$ . To prove the correctness of the meta-algorithm, see Theorem 4.1 in March et al. (2010). That proof can be adapted from  $kd$ -trees to general space trees. Our representation is a generalization of their algorithms; our meta-algorithm produces their  $kd$ -tree and cover tree implementations with a tighter distance bound  $B(\mathcal{N}_q)$ . Our meta-algorithm produces a provably correct dual-tree algorithm with any type of space tree.

## 8. Kernel Density Estimation

Much work has been produced regarding the use of dual-tree algorithms for kernel density estimation (KDE), including by Gray & Moore (2001; 2003b) and later by Lee et al. (2006; 2008). KDE is an important machine learning problem with a vast range of applications, such as signal processing to econometrics. The problem is, given query and reference sets  $S_q, S_r$ , to estimate a probability density  $f_q$  at each point  $p_q \in S_q$  using each point  $p_r \in S_r$  and a kernel function  $K_h$ . The exact probability density at a point  $p_q$  is the sum of  $K(\|p_q - p_r\|)$  for all  $p_r \in S_r$ .

In general, the kernel function is some zero-centered probability density function, such as a Gaussian. This means that when  $\|p_q - p_r\|$  is very large, the contri-

bution of  $K$  to  $f_q$  is very small. Therefore, we can approximate small values using a dual-tree algorithm to avoid unnecessary computation; this is the idea set forth by Gray & Moore (2001). Because  $K$  is a function which is decreasing with distance, the maximum difference between  $K$  values for a given combination  $(\mathcal{N}_q, \mathcal{N}_r)$  can be bounded above with

$$B_K(\mathcal{N}_q, \mathcal{N}_r) = K(d_{\min}(\mathcal{N}_q, \mathcal{N}_r)) - K(d_{\max}(\mathcal{N}_q, \mathcal{N}_r)).$$

The algorithm takes a parameter  $\epsilon$ ; when  $B_K(\mathcal{N}_q, \mathcal{N}_r)$  is less than  $\epsilon/|S_r|$ , the kernel values are approximated using the kernel value of the centroid  $C_r$  of the reference node. The division by  $|S_r|$  ensures that the total approximation error is bounded above by  $\epsilon$ . The base case on  $p_q$  and  $p_r$  merely needs to add  $K(\|p_q - p_r\|)$  to the existing density estimate  $f_q$ . When the algorithm is initialized,  $f_q = 0$  for all query points. `BaseCase()` is Algorithm 10 and `Score()` is Algorithm 9.

Again we emphasize the flexibility of our meta-algorithm. To our knowledge cover trees, octrees, and ball trees have never been used to perform KDE in this manner. Our meta-algorithm can produce these implementations with ease.

## 9. Discussions

We have now shown five separate algorithms for which we have taken existing dual-tree algorithms and constructed a `BaseCase()` and `Score()` function that can be used with any space tree and any dual-tree traversal. Single-tree extensions of these four methods are straightforward simplifications.

This modular way of viewing tree-based algorithms has several useful immediate applications. The first

---

**Algorithm 9** KDE `Score`( $\mathcal{N}_q, \mathcal{N}_r$ ).

---

**Input:** query node  $\mathcal{N}_q$ , reference node  $\mathcal{N}_r$

**Output:** a score for  $(\mathcal{N}_q, \mathcal{N}_r)$  or  $\infty$  if the combination should be pruned

```

if  $B_K(\mathcal{N}_q, \mathcal{N}_r) \geq \frac{\epsilon}{|S_r|}$  then
  for each  $p_q \in \mathcal{D}_q^p$  do
     $f_q \leftarrow f_q + |\mathcal{D}_r^p| K(\|p_q - C_r\|)$ 
  return  $\infty$ 
return  $d_{\min}(\mathcal{N}_q, \mathcal{N}_r)$ 

```

---



---

**Algorithm 10** KDE `BaseCase`( $p_q, p_r$ ).

---

**Input:** query point  $p_q$ , reference point  $p_r$ , density estimate  $f_q$

**Output:** distance between  $p_q$  and  $p_r$

```

if BaseCase( $p_q, p_r$ ) already called then return
 $f_q \leftarrow f_q + K_h(\|p_q - p_r\|)$ 
return  $\|p_q - p_r\|$ 

```

---

is implementation. Given a tree implementation and a dual-tree traversal implementation, all that is required is `BaseCase()` and `Score()` functions. Thus, code reuse can be maximized, and new algorithms can be implemented simply by writing two new functions. More importantly, the code is now modular. MLPACK (Curtin et al., 2011), written in C++, uses templates for this. One example is the `DualTreeBoruvka` class, which implements the meta-algorithm discussed in Section 7, and has the following arguments:

```
template<typename MetricType,
        typename TreeType,
        typename TraversalType>
class DualTreeBoruvka;
```

This means that any class satisfying the constraints of the `TreeType` template parameter can be designed without any consideration or knowledge of the `DualTreeBoruvka` class or of the `TraversalType` class; it is *entirely independent*. Then, assuming a `TreeType` and `TraversalType` without bugs, the dual-tree Borůvka’s algorithm is guaranteed to work. An immediate example of the advantage of this is that cover trees were implemented for MLPACK for use with  $k$ -nearest neighbor search. This cover tree implementation could, without any additional work, be used with `DualTreeBoruvka`—which was never an intended goal during the cover tree implementation but still a particularly valuable result!

Of course, the utility of these abstractions are not limited to implementation details. Each of the papers cited in the previous sections describe algorithms in terms of one specific tree type. March et al. (2010) discuss implementations of Borůvka’s Algorithm on both  $kd$ -trees and cover trees and give algorithms for both. Each algorithm given is quite different and it is not easy to see their similarities. Using our meta-algorithm, any of these tree-based algorithms can be expressed with less effort—especially for more complex trees like the cover tree—and in a more general sense.

In addition, correctness proofs for our algorithms tend to be quite simple. These proofs for each algorithm here can be given in two simple sub-proofs: (1) prove the correctness of `BaseCase()` when no prunes are made, and (2) prove that `Score()` does not prune any subtrees on which the algorithm’s correctness depends.

The logical split of base case, pruning rule, tree type, and traversal can also be advantageous. A strong example of this is the function  $B(\mathcal{N}_q)$  devised in Section 5, which is a novel, tighter bound. When not considering a particular tree, the path to a superior algorithm can often be simpler (as in that case).

## 9.1. Parallelism

Nowhere in this paper has parallelism been discussed in any detail. In fact, all of the given algorithms seem to be suited to serial implementation. However, the pruning dual-tree traversal is entirely separate from the rest of the dual-tree algorithm; therefore, a parallel pruning dual-tree traversal can be used without modifying the rest of the algorithm.

For instance, consider  $k$ -nearest neighbor search. Most large-scale parallel implementations of  $k$ -NN do not use space trees but instead techniques like LSH for fast (but inexact) search. To our knowledge, no freely available software exists that implements distributed dual-tree  $k$ -nearest neighbor search.

As a simple (and not necessarily efficient) proof-of-concept idea for a distributed traversal, suppose we have  $t^2$  machines and a “master” machine for some  $t > 0$ . Then, for a query tree  $\mathcal{T}_q$  and a reference tree  $\mathcal{T}_r$ , we can split  $\mathcal{T}_q$  into  $t$  subtrees and one “master tree”  $\mathcal{T}_{qm}$ . The reference tree  $\mathcal{T}_r$  is split the same way. Each possible combination of query and reference subtrees is stored on one of the  $t^2$  machines, and the master trees are stored on the master machine. The lists  $D$  and  $N$  can be stored on the master machine and can be updated or queried by other machines.

The traversal starts at the roots of the query tree and reference tree and proceeds serially on the master. When a combination in two subtrees is reached, `Score()` and `BaseCase()` are performed on the machine containing those two subtrees and that subtree traversal continues in parallel. This idea satisfies the conditions of a pruning dual-tree traversal; thus, we can use it to make any dual-tree algorithm parallel.

Recently, a distributed dual-tree algorithm was developed for kernel summation (Lee et al., 2012); this work could be adapted to a generalized distributed pruning dual-tree traversal for use with our meta-algorithm.

## 10. Conclusion

We have proposed a tree-independent representation of dual-tree algorithms and a meta-algorithm which can be used to create these algorithms. A dual-tree algorithm is represented as four parts: a space tree, a pruning dual-tree traversal, a `BaseCase()` function, and a `Score()` function. We applied this representation to generalize and extend five example dual-tree algorithms to different types of trees and traversals. During this process, we also devised a novel bound for  $k$ -nearest neighbor search that is tighter than existing bounds. Importantly, this abstraction can be applied to help approach the problem of parallel dual-tree algorithms, which currently is not well researched.

## References

- Agarwal, P.K. and Erickson, J. Geometric range searching and its relatives. *Contemporary Mathematics*, 223:1–56, 1999.
- Bentley, J.L. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.
- Bentley, J.L. and Friedman, J.H. Data structures for range searching. *ACM Computing Surveys (CSUR)*, 11(4):397–409, 1979.
- Beygelzimer, A., Kakade, S.M., and Langford, J. Cover trees for nearest neighbor. In *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, volume 23, pp. 97, 2006.
- Cover, T.M. and Hart, P.E. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.
- Curtin, R.R., Cline, J.R., Slagle, N.P., Amidon, M.L., and Gray, A.G. MLPACK: a scalable C++ machine learning library. In *BigLearning: Algorithms, Systems, and Tools for Learning at Scale*, 2011.
- Friedman, J.H., Bentley, J.L., and Finkel, R.A. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software (TOMS)*, 3(3):209–226, 1977.
- Fukunaga, K. and Narendra, P.M. A branch and bound algorithm for computing k-nearest neighbors. *Computers, IEEE Transactions on*, 100(7):750–753, 1975.
- Gray, A.G. *Bringing Tractability to Generalized N-Body Problems in Statistical and Scientific Computation*. PhD thesis, Carnegie Mellon University, 2003.
- Gray, A.G. and Moore, A.W. ‘N-Body’ problems in statistical learning. In *Advances in Neural Information Processing Systems 14 (NIPS 2001)*, volume 4, pp. 521–527, 2001.
- Gray, A.G. and Moore, A.W. Rapid evaluation of multiple density models. In *Advances in Neural Information Processing Systems 16 (NIPS 2003)*, volume 2003 of *NIPS '03*, 2003a.
- Gray, A.G. and Moore, A.W. Nonparametric density estimation: Toward computational tractability. In *SIAM International Conference on Data Mining (SDM)*. Citeseer, 2003b.
- Holmes, M.P., Gray, A.G., and Isbell Jr., C.L. QUIC-SVD: Fast SVD using cosine trees. *Advances in Neural Information Processing Systems (NIPS)*, 21, 2008.
- Lee, D. and Gray, A.G. Faster Gaussian summation: Theory and experiment. In *In Proceedings of the Twenty-second Conference on Uncertainty in Artificial Intelligence (UAI)*, 2006.
- Lee, D. and Gray, A.G. Fast high-dimensional kernel summations using the monte carlo multipole method. *Advances in Neural Information Processing Systems (NIPS)*, 21, 2008.
- Lee, D., Gray, A.G., and Moore, A.W. Dual-tree fast Gauss transforms. In *Advances in Neural Information Processing Systems 18*, pp. 747–754. MIT Press, Cambridge, MA, 2006.
- Lee, D., Vuduc, R., and Gray, A.G. A distributed kernel summation framework for general-dimension machine learning. In *SIAM International Conference on Data Mining*, volume 2012, pp. 5, 2012.
- March, W.B., Ram, P., and Gray, A.G. Fast Euclidean minimum spanning tree: algorithm, analysis, and applications. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '10)*, pp. 603–612, 2010.
- March, W.B., Connolly, A.J., and Gray, A.G. Fast algorithms for comprehensive n-point correlation estimates. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1478–1486. ACM, 2012.
- Ram, P., Lee, D., March, W.B., and Gray, A.G. Linear-time algorithms for pairwise statistical problems. *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, 23, 2009.
- Wang, P., Lee, D., Gray, A.G., and Rehg, J.M. Fast mean shift with accurate and stable convergence. In *Workshop on Artificial Intelligence and Statistics (AISTATS)*, 2007.