Appl. Comput. Harmon. Anal. • • • (• • • •) • • •–• • •

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Far-field compression for fast kernel summation methods in high dimensions

William B. March *, George Biros

Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Article history:
Received 11 February 2015
Received in revised form 14 September 2015
Accepted 23 September 2015
Available online xxxx
Communicated by Leslie F. Greengard

Keywords:
Kernel independent fast multipole methods
Fast summation
Randomized matrix approximation
Interpolative decomposition
Matrix sampling

ABSTRACT

We consider fast kernel summations in high dimensions: given a large set of points in d dimensions (with $d \gg 3$) and a pair-potential function (the *kernel* function), we compute a weighted sum of all pairwise kernel interactions for each point in the set. Direct summation is equivalent to a (dense) matrix–vector multiplication and scales quadratically with the number of points. Fast kernel summation algorithms reduce this cost to log-linear or linear complexity.

Treecodes and Fast Multipole Methods (FMMs) deliver tremendous speedups by constructing approximate representations of interactions of points that are far from each other. In algebraic terms, these representations correspond to low-rank approximations of blocks of the overall interaction matrix. Existing approaches require an excessive number of kernel evaluations with increasing d and number of points in the dataset.

To address this issue, we use a randomized algebraic approach in which we first sample the rows of a block and then construct its approximate, low-rank interpolative decomposition. We examine the feasibility of this approach theoretically and experimentally. We provide a new theoretical result showing a tighter bound on the reconstruction error from uniformly sampling rows than the existing state-of-the-art. We demonstrate that our sampling approach is competitive with existing (but prohibitively expensive) methods from the literature. We also construct kernel matrices for the Laplacian, Gaussian, and polynomial kernels—all commonly used in physics and data analysis. We explore the numerical properties of blocks of these matrices, and show that they are amenable to our approach. Depending on the data set, our randomized algorithm can successfully compute low rank approximations in high dimensions. We report results for data sets with ambient dimensions from four to 1,000.

1. Introduction

Given n source points x_j with densities q_j, m target points y_i, and a kernel function $\mathcal{K}$, we seek to evaluate the *kernel sum*

* Corresponding author.
E-mail address: march@ices.utexas.edu (W.B. March).

http://dx.doi.org/10.1016/j.acha.2015.09.007
1063-5203/

$$u_i = \sum_{j=1}^{n} \mathcal{K}(y_i, x_j) q_j = \sum_{j=1}^{n} K_{ij} q_j \tag{1.1}$$

for each target y_i, with $K_{ij} = \mathcal{K}(y_i, x_j)$. Computing $u \in \mathbb{R}^m$ is equivalent to a matrix–vector multiplication, $u = Kq$, $K \in \mathbb{R}^{m \times n}$, and it requires $\mathcal{O}(nm)$ work. It is prohibitively expensive for large m and n. Fast kernel summation algorithms (also known as generalized N-body problems) aim to provide an approximate solution with guaranteed error using only $\mathcal{O}(n + m)$ kernel evaluations. They do so by identifying and approximating blocks of K that have low-rank structure.

Fast kernel summations are a fundamental operation in computational physics. They are related to the solution of partial differential equations in which $\mathcal{K}$ is the corresponding Green's function. Examples include the 3D Laplace potential (reciprocal distance kernel) and the heat potential (Gaussian kernel).

Kernel summations are also fundamental to non-parametric statistics and machine learning tasks such as density estimation, regression, and classification. Linear inference methods such as support vector machines [68] and dimension reduction methods such as principal components analysis [62] can be efficiently generalized to non-linear methods by replacing inner products with kernel evaluations [9]. Problems in statistics and machine learning are often characterized by very high-dimensional inputs.

Existing fast algorithms for the kernel summation problem hinge on the construction of efficient approximations of interactions[1] between groups of sources and targets when these groups are far apart or *well separated* (see Section 2). In the physics/PDE community, they are known as far-field approximations. From a linear algebraic point-of-view, they correspond to low-rank decompositions of blocks of the matrix K. These approximations can be roughly grouped in three categories: **analytic**, **semi-analytic**, and **algebraic**.

In analytic methods, Taylor or kernel-dependent special function expansions are used to approximate the far-field. The Fast Multipole Method (FMM) [38] is one of these. Semi-analytic methods rely only on kernel evaluations, but the low-rank constructions use the analytical properties of the underlying kernels. For example, the kernel-independent fast multipole method [81] requires that the underlying kernel is the Green's function of a PDE. Finally, algebraic methods (e.g. [59]) also only use kernel evaluations, and the only necessary condition is the existence a low-rank block structure for K.

In high dimensions, most existing methods fail. There are two main reasons for the lack of scalability of analytic and semi-analytic methods. First, all existing schemes require too many terms for the kernel approximation. Analytic and semi-analytic schemes can deliver approximations to arbitrary accuracy (in practice all the way to machine precision) with $\mathcal{O}(n + m)$ kernel evaluations, but the constant can be very large. For p terms in the series expansion, they require $p = c^d$ or $p = c^{d-1}$ terms to deliver error that decays exponentially in $c > 1$. Variants that can scale reasonably well beyond three dimensions scale as $p = d^c$ and deliver error that decays algebraically in c. For sufficiently large d and $c > 1$, either of these methods becomes too expensive [40].

The second reason for lack of scalability of existing schemes is that they do not take advantage of any lower-dimensional structures that may be present in the data. For example, the data may be embedded in a low-dimensional manifold. This is mostly relevant in data analysis applications in which often the important dimension is not the *ambient* one but instead an *intrinsic* dimension that depends on the distribution of the source and target points.

Algebraic approximations [59] are a promising direction for scalable methods in high dimensions. These approximations are based on the observation that Equation (1.1) is a matrix–vector product and certain blocks of the matrix have low-rank structure. Algebraic methods are useful only if the approximation can be computed efficiently. Efficient methods for low dimensions do exist, but in high-dimensions they fail because the number of kernel evaluations required exceeds the cost of the direct summation.

[1] We use the term *interaction* between two points to refer to the value of the kernel $\mathcal{K}$.

Please cite this article in press as: W.B. March, G. Biros, Far-field compression for fast kernel summation methods in high dimensions, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.09.007