
Andreas Stathopoulos
**Approximating $\text{Tr}(A^{-1})$ with hierarchical probing on
regular stencils**

Department of Computer Science
College of William and Mary
Williamsburg
VA 23187-8795
`andreas@cs.wm.edu`
Jesse Laeuchli
Konstantinos Orginos

The problem of finding the trace of the inverse of a large, sparse matrix A appears in many applications, including lattice quantum chromodynamics (LQCD), data mining, and uncertainty quantification. When the dimension of A , N , is sufficiently small, the problem is trivially solved by direct factorization methods. For large N , where factorization of A is not possible, the problem is surprisingly difficult, and all current approaches resort to a Monte Carlo method, based on the following: Given $x \in \mathbb{Z}_2^N$, i.e., random vectors of $1, -1$,

$$E(x^T A^{-1} x) = \text{Tr}(A^{-1}).$$

A model algorithm generates a list of s vectors, and either solves a series of linear systems of equations, or computes directly a series of Gaussian quadratures ($x^T A^{-1} x$). Because of the rate of Monte Carlo convergence, several hundreds of systems need to be solved to obtain even a single digit of accuracy. The need for algorithmic improvements at any level is obvious.

Our driving application is LQCD, where the matrix structure is based on a four dimensional torus-lattice. In our previous work, we focused on speeding up the systems with multiple right hand sides using deflation of a large number of eigenvectors which are obtained with negligible additional cost to solving the linear systems. Because of the large number of required right hand sides, we showed speedups up to an order of magnitude.

In this research we focus on the orthogonal issue of the quality of the right hand sides. We seek sequences of vectors that provide better approximations than random vectors.

Over the last ten years there has been quite some interest in this topic. Bekas, Kokkiopoulou, and Saad discussed algorithms that use the Hadamard low disparity vectors that work well for matrices with certain diagonal structure. They also discussed probing which is our focus below. Recently, Avron and Toledo did a study of various random vector choices. Tang and Saad, also discussed a method based on probing.

Probing is a method for obtaining the diagonal (or the trace) of a sparse matrix. If the graph of the sparse matrix is k colorable, consider the set of k vectors $X = [x^1, x^2, \dots, x^k]$, where $x_i^j = 1$ if node i has color j , and zero otherwise. It is not hard to see that $\text{Tr}(A) = \text{Tr}(X^T A X)$. For example, in LQCD where the 4-D lattice is 2-colorable, only two vectors would suffice to find its trace. The problem, however, is that we need $\text{Tr}(A^{-1})$, which in general is a dense matrix.

Bekas et al. and Tang et al. exploited the fact that for a large class of PDE based matrices, the elements of the A^{-1} decay exponentially away from the non zero elements of A . Therefore, the largest elements of A^{-1} are on the same sparsity pattern as A , which for the LQCD case is 2-colorable. Thus with 2 vectors we get a rough approximation of $\text{Tr}(A^{-1})$. To improve this, we have to consider a “denser” approximation of A^{-1} , specifically the values of A^{-1} on the sparsity pattern of A^2 . This is known as the distance-2 coloring of the graph of A . If this approximation is not satisfactory, we can continue up to A^m , obtain a k -coloring of its graph, and use the X as defined above to obtain the trace. This is the approach followed by Tang and Saad. However, such approach is wasteful as all computations performed up to step $m - 1$ must be discarded.

In this work, we attempt to avoid this waste; at step m we build on top of what we have already computed in steps $1, \dots, m - 1$. However, we are faced with the problem that the vector subspace $X^{(m-1)}$, computed from the $k^{(m-1)}$ colors needed for A^{m-1} , may not be contained in the subspace $X^{(m)}$ for the next step. Therefore, previous computations cannot be reused in $X^{(m)T} A X^{(m)}$. We address this problem by considering a hierarchical graph coloring of A^m . We start with the bipartite, red-black partitioning of A (for LQCD). Then we consider a subset of the graph of A^2 on the red partition only. We color it with k_r colors. Then, we color the black partition subset of the A^2 with k_b colors. We can now build $X^{(2)}$ with $k_r + k_b$ vectors, which by construction also span $X^{(1)}$. Thus, only the complement space needs to be used. To avoid expensive orthogonalization, we can identify the complement based on a judicious use of Hadamard vectors.

A concern with the above method is that the number of colors for any A^m is larger than the minimum (chromatic number of A^m). We are currently investigating how much larger this can be, in particular for a regular lattice. However, it is unclear whether this is a real disadvantage, as the sparsity structure of A^m is used only as an approximation to A^{-1} . In fact, more vectors yield more accurate results.

We will discuss also two variants of this method. First, it is not necessary to start with a red-black ordering of A , as we know that more vectors will be needed. We can start with a k coloring of A^m (say for $m = 2$ or 3), and then apply the hierarchical graph coloring. Second, we can stop the hierarchical graph coloring at a certain distance- m , i.e., A^m . If more accuracy is needed, each of the vectors in $X^{(m)}$ can be extended with the regular sequence of Hadamard vectors,

or with random vectors, if we want to remove the bias from the deterministic approximation.