An Estimator for the Diagonal of a Matrix *

C. Bekas, E. Kokiopoulou, Y. Saad

Computer Science & Engineering Dept.
University of Minnesota, Twin Cities
{beaks,kokobeh,saad}@cs.umn.edu

June 1, 2005

Abstract

A number of applications require to compute an approximation of the diagonal of a matrix when this matrix is not explicitly available but matrix-vector products with it are easy to evaluate. In some cases, it is the trace of the matrix rather than the diagonal that is needed. This paper describes methods for estimating diagonals and traces of matrices in these situations. The goal is to obtain a good estimate of the diagonal by applying only a small number of matrix-vector products, using selected vectors. We begin by considering the use of random test vectors and then explore special vectors obtained from Hadamard matrices. The methods are tested in the context of computational materials science to estimate the diagonal of the density matrix which holds the charge densities. Numerical experiments indicate that the diagonal estimator may offer an alternative method that in some cases can greatly reduce computational costs in electronic structures calculations.

1 Introduction and motivation

The problem of computing the diagonal of a matrix, or its trace, when the matrix is known only via its actions on arbitrary vectors, arises in a number of applications. In regularized solutions of least-squares problems solved in image restoration, one is required to estimate a certain regularization parameter θ. The Generalized Cross-Validation approach to this problem (see [8]) consists of seeking a value θ which minimizes a certain function of the form

$$\frac{\|(I - A(\theta))g\|_2}{tr(I - A(\theta))} \quad \text{with} \quad A(\theta) \equiv I - D(D^T D + \theta L L^T)^{-1} D^T,$$

where D is the blurring operator and L is the regularization operator. The trace involved in the above expression is difficult to compute as it involves the inverse of a matrix. In [8] a method based on statistical arguments was proposed for estimating this trace.

Approximating traces of operators is also needed in a few methods employed in the completely different arena of electronic structures calculations. One such method proposed in the 1990s for

*Work supported by NSF grants ITR-0082094, ACE-0305120, by DOE under Grants DE-FG02-03ER25585, DE-FG02-03ER15491, and by the Minnesota Supercomputers Institute.

computing density of states exploits Chebychev moments of a Hermitian matrix A, scaled to have eigenvalues in $[-1,\ 1]$, i.e.,

$$\mu_k = tr(C_k(A)) = \int_{-1}^{1} C_k(t)\rho(t)\ dt,$$

where C_k is the k-th degree Chebychev polynomial of the first kind, and $\rho(t)$ represents the density of states $\rho(t) \equiv \sum_{\lambda_i} \delta(\lambda - \lambda_i)$. Statistical methods are again used to estimate the trace $tr(C_k(A))$ for many values of k, and these are used to recover the density of states ρ. Still in the same application area, one is often required to compute the diagonal of a certain eigen-projector associated with the smallest m eigenvalues of a Hamiltonian matrix. The projector is a matrix of the form $P = VV^T$, where the columns of V are eigenvectors associated with the smallest m eigenvalues. However, P can also be viewed as a function of A, specifically, $P = f(A)$ where f is a step function which has the value one in the interval containing the lowest m eigenvalues and zero elsewhere. The goal now is to estimate the diagonal of this matrix without computing the eigenvectors.

Both of these problems, estimating the diagonal or the trace of a matrix, are addressed in this paper as they require similar techniques. Clearly, estimating the diagonal is more complex than estimating the trace and this problem is not as common in the literature. However, a good estimator of the diagonal of a projector may lead to exceptional improvements in density functional theory – especially for systems with many atoms. Our goal in this paper is not as much to show that methods based on diagonal estimators are competitive as it is to suggest a number of possible options. Some of these methods have already been used in the literature, but comparisons with standard approaches have not yet been made. For example, methods based on stochastic estimations are, relatively speaking, very expensive because they tend to converge slowly. When some structure of the matrix under consideration is known, then as will seen, estimating the diagonal can become quite inexpensive. Both approaches are considered in turn.

2 Approximating the diagonal of a matrix

2.1 The stochastic framework

In [8] Hutchinson described an unbiased stochastic estimator for the trace $\mathtt{tr}(I - A)$, where I is the identity matrix and A is the influence matrix associated with the calculation of Laplacian smoothing splines. Since A involves the inverse of another matrix, its trace is not readily available. Hutchinson extended ideas of Girard (see [5]) and presented an estimator of minimum variance that is based on simulations of the discrete variable which assumes the values -1 and 1 with equal probability 1/2. In particular, to estimate the trace $\mathtt{tr(A)}$ of the matrix A, take random vectors $v_k, i = 1, .., s$ with entries ± 1 and then compute the average over the sample of $v_k^\top A v_k$

$$\mathtt{tr}(A) \approx \frac{1}{s} \sum_{k=1}^{s} v_k^\top A v_k. \tag{1}$$

The above idea can be easily extended for the problem of estimating the diagonal of a matrix A. Consider again a sequence of vectors $v_1, \ldots, v_s$. For reasons that will become clear in later

sections, the entries of these vectors do not have to be ± 1. Then, $\texttt{diag}(A)$, the diagonal of A written as an n-dimensional vector can be estimated by the following vector sequence:

$$D^s = \left[\sum_{k=1}^{s} v_k \odot Av_k\right] \oslash \left[\sum_{k=1}^{s} v_k \odot v_k\right], \tag{2}$$

where $\odot$ represents componentwise multiplication of vectors, and similarly $\oslash$ represents componentwise division of vectors. Figure 1 shows an algorithmic description of the stochastic estimator.

Diagonal Estimator

(* *Input* *) Function $\mathcal{A}(v) : \mathbb{R}^n \to \mathbb{R}^n$ such that $x = Av$,
where A is the unknown matrix, number of steps s
(* *Output* *) Approximate diagonal D^s of the matrix A
1. $D^0 = 0$
2. **for** $k = 1 \ldots s$
   2.1 Generate current "random" vector v_k
   2.2 $t_k = t_{k-1} + (\mathcal{A}(v_k) \odot v_k)$
   2.3 $q_k = q_{k-1} + (v_k \odot v_k)$
   2.3 $D^k = t_k \oslash q_k$
3. **end**

Figure 1: Estimator for the diagonal of an unknown matrix A which is available only through matrix-vector products.

If the vectors v_k have entries ± 1, then the trace of D^s is nothing but the Hutchinson trace estimation of A. A careful look at the $i-$th entry of the approximation D^s to the diagonal of A will provide some insight. Again, superscripts refer to components and subscripts refer to the vector index within the sample.

$$\begin{aligned} D_i^s &= \frac{\sum_{k=1}^{s} v_k^i \sum_{j=1}^{n} \alpha_{ij} v_k^j}{\sum_{k=1}^{s} \left(v_k^i\right)^2} \\ &= \alpha_{ii} + \frac{\sum_{k=1}^{s} v_k^i \sum_{j\neq i} \alpha_{ij} v_k^j}{\sum_{k=1}^{s} \left(v_k^i\right)^2} \\ &= \alpha_{ii} + \sum_{j\neq i} \alpha_{ij} \frac{\sum_{k=1}^{s} v_k^i v_k^j}{\sum_{k=1}^{s} \left(v_k^i\right)^2}. \end{aligned} \tag{3}$$

On the average, the coefficient of α_{ij} in the above expansion will converge to zero provided that the components of the vectors v_k have balanced $\pm$ signs. This is true if the components of v_k are equiprobably drawn between the values of 1 and -1, justifying this choice in Hutchinson's trace estimator.

One problem with statistical techniques, such as the one presented here, is that they tend to be slow. In order to get more effective algorithms, we may consider selecting the vectors v_k more carefully, rather than using random vectors. In order to justify the methods to be considered later, we next examine conditions under which the diagonal estimator D^s is exact.

2.2 Sufficient conditions for exactness

We now present a general sufficient condition which will ensure that the diagonal estimator given by formula (2) is exact. Denoting by A_i the $i-$th column of A, and by v_k^i the i-component of the k-th vector v_k, we can write:

$$\begin{aligned}\sum_{k=1}^{s}(v_k \odot Av_k) &= \sum_{k=1}^{s}\left(v_k \odot \sum_{i=1}^{n}\left(v_k^i A_i\right)\right)\\ &= \sum_{k=1}^{s}\sum_{i=1}^{n}(v_k^i v_k \odot A_i)\\ &= \sum_{i=1}^{n}\sum_{k=1}^{s}(v_k^i v_k \odot A_i)\\ &= \sum_{i=1}^{n} A_i \odot \sum_{k=1}^{s}(v_k^i v_k). \end{aligned} \tag{4}$$

Observe that $\sum_{k=1}^{s}(v_k^i v_k)$ is the $i-$th column of the matrix $VV^\top$, where the vectors $v_k, k = 1,\ldots,s$ are the columns of the matrix $V \in \mathbb{R}^{n\times s}$. Therefore, the diagonal estimator (2) can be rewritten as :

$$D^s = \left[(A \odot VV^\top)\mathbf{e}\right] \oslash \left[(V \ \odot \ V)\mathbf{e}\right], \tag{5}$$

where $\mathbf{e}$ is the vector of all ones.

The nonzero pattern of the matrix $VV^\top$ will clearly play an important role in the convergence of the diagonal estimator. If the matrix V has mutually orthogonal rows, then the matrix $VV^\top$ will be diagonal, in which case all off-diagonal elements of A will be excluded from the bracketed expression of the estimator (5). The same conclusion can be reached by looking at the expression (3), since $\sum_{k=1}^{s} v_k^i v_k^j$ is the inner product of two rows of the matrix V. More generally we have the following proposition.

Proposition 2.1 *Let $V \in \mathbb{R}^{n\times s}$ be a matrix the columns of which are used in the diagonal estimator. If the i-th row of V is orthogonal to all those rows j of V for which $a_{ij} \neq 0$, then the diagonal estimator will yield an exact result for a_{ii}, the i-th diagonal entry of A.*

Proposition 2.1 clearly suggests that if the non-zero structure of a sparse matrix A is known, then it may be possible to select the vectors v_k in such a way that the off-diagonal elements of A are eliminated from (3). This idea will be explored further in Section 2.3.

On the other hand, should nothing be known about the structure of the matrix A, then according to equation (5) the best that we can do is to select the vectors v_k in such a way that the matrix $VV^\top$ is as close to a diagonal matrix as possible. Of course, if the number of vectors v_k is smaller than the size of the matrix A ($s < n$), as is desirable in practice, then the rows of the matrix V cannot be selected to be mutually orthogonal, and therefore the matrix $VV^\top$ cannot, in general, be diagonal. Thus, the question is how to select the vectors v_k such that the matrix $VV^\top$ is the closest possible to being diagonal. In terms of the rows of the matrix V we would like to minimize

$$E_{rms} = \sqrt{\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{j'\neq j}^{n}|V_j V_{j'}^\top|^2}, \tag{6}$$

where V_j is the j-th row of the matrix V. This is just the root mean square magnitude of the off-diagonal entries of the matrix $VV^\top$. If the rows are scaled to have unit-norm, then the above measure is just a scaled Frobenius norm of $I - VV^T$. In addition, we may be interested in minimizing the maximum absolute value of the off-diagonal elements in $VV^\top$

$$E_{max} = \max_{1 \le j' < j < n} |V_j V_{j'}^\top|. \tag{7}$$

Because of symmetry, only the strict upper (or lower) triangular part of $VV^\top$ need to be considered.

2.2.1 Grassmannian spaces and code books

Problems similar to the one discussed above arise in various other fields. A prominent example refers to line packing in Grassmannian spaces [3]. The Grassmannian[1] space $G(M, N)$ is the set of all N-dimensional subspaces of the real Euclidean space $\mathbb{R}^M$. Consider, for example, the problem of best separating n lines that pass through the origin in $\mathbb{R}^3$. In other words, we are interested in finding the best arrangement of these lines so that the acute angle between any two lines is maximized. In the setting of the diagonal estimator the equivalent problem is to maximize the acute angle between any two rows V_i, V_j of the matrix V in (5). The Grassmannian space then is $G(s, 1)$ and the number of lines is $n >> s$.

Another manifestation of the same problem is in communications and refers to the design of code books that minimize the maximal cross-correlation amplitude (see [13], [18] and references therein). In particular, the rows of the matrix V are codewords, which are taken to have unit norm. Then, the length of the code is s and the number of codewords is n. In this case, we are either interested in minimizing E_{max} or E_{rms}. The following bounds are known [16]

$$E_{rms} \ge \sqrt{\frac{n-s}{(n-1)s}}, \tag{8}$$

with equality if and only if $\sum_{j=1}^n V_j V_j^\top = \frac{n}{s} I$ and

$$E_{max} \ge \sqrt{\frac{n-s}{(n-1)s}}, \tag{9}$$

with equality if and only if

$$|V_j V_{j'}^\top| = \sqrt{\frac{n-s}{(n-1)s}}, \ \forall j \ne j'. \tag{10}$$

Designing codes that meet the optimality bound (9) for E_{max} with equality is a difficult problem that is still open for the case of general pairs (n, s), while it has found elegant closed form solutions for special values of (n, s) and complex valued unit norm codewords [18]. On the other hand, designing real valued codes that meet the E_{rms} bound with equality is considered to be an easy problem. Examples include binary codes such as a truncated Hamming codes with size $M = 2n$, distance d and code length s.

[1] See also the web site: `http://www.research.att.com/~njas/grass/index.html`

2.3 Probing techniques for diagonal estimation

When the matrix A is sparse and its non-zero structure is known, efficient algorithms are known (see [2] and [4]) which utilize matrix-vector multiplications to approximate the entries of A. These algorithms exploit the structure of the matrix A in order to use only a small number of matrix-vector products to generate the matrix A. They have been quite successful in the context of estimating sparse Jacobians of a linear mapping $F : \mathbb{R}^n \to \mathbb{R}^m$. In particular, these methods are commonplace in standard nonlinear equation solvers, when it is difficult to evaluate Jacobian matrices explicitly.

A special case of the above methods applies to banded matrices and is known as *probing*. Even when the matrix A is not banded, but rather the magnitude of its entries exhibits a decaying behavior away from the main diagonal, probing techniques can still be used to yield approximations to the matrix. The decay of off-diagonal entries away from the main diagonal occurs in a many applications where locality plays a central role. For example, in molecular systems of insulators, particles that are far apart are frequently found to be only very weakly coupled. This induces an "almost" banded structure of the density matrix in electronic structure calculations [6], [7].

Next, we briefly outline the main idea behind probing. For details, see [2], for example. Consider a tridiagonal matrix $T \in \mathbb{R}^{n\times n}$, which is the simplest case of a banded matrix with no trivial upper and lower bandwidths,

$$T = \begin{pmatrix} a_1 & b_2 & & & & \\ c_2 & a_2 & b_3 & & & \\ & c_3 & a_3 & b_4 & & \\ & & \ddots & \ddots & \ddots & \\ & & & c_{n-1} & a_{n-1} & b_n \\ & & & & c_n & a_n \end{pmatrix}. \tag{11}$$

The main observation is that the column set of T consists of three subsets:

$$S_1 = \{1, 4, 7, \ldots, 1+3i, ...\} \quad S_2 = \{2, 5, 8, \ldots, 2+3i, ...\} \quad S_3 = \{3, 6, 9, \ldots, 3+3i, ...\}$$

each of which has the characteristic of not sharing any common rows with the other two sets. Thus, all columns from the set S_1 can be computed by multiplying T by the vector v_1 which has the entry one in positions 1, 4, $\cdots$, $1+3i$, $\cdots$, and zero elsewhere. Similarly, the columns in S_2 and S_3 can be extracted by multiplying T by the vectors

$$v_2 = [0, 1, 0, 0, 1, 0, 0, \ldots]^\top \quad v_3 = [0, 0, 1, 0, 0, 1, 0, \ldots]^\top,$$

respectively. The above idea can be extended to general sparse matrices with an arbitrary sparsity pattern used a method based on coloring the adjacency graph of the sparse matrix [2], [4].

In the case of a banded matrix A, the stochastic diagonal estimator presented earlier (see Fig. 1), can yield an efficient and accurate approach for estimating the diagonal of A. The key observation comes from equation (3) which we re-write as

$$d_i^s = \alpha_{ii} + \sum_{j\neq i} \alpha_{ij} \frac{\sum_{k=1}^s v_k^i v_k^j}{\sum_{k=1}^s \left(v_k^i\right)^2}.$$

Let the columns of the matrix $V = [v_1, v_2, \ldots, v_s]$ be the pseudorandom vectors used in the stochastic estimator. The computed approximation d_i^s to the i-th diagonal entry will be exact (i.e., equal to α_{ii}) when either $\alpha_{ij} = 0$, for $i \neq j$, or the i-th row of V is orthogonal to the j-th row of V. Note that the computation involved in the above stochastic estimator is mathematically equivalent to multiplying the matrix A componentwise by the matrix $VV^\top$ and taking the sum of the columns of the resulting matrix. Let us now consider a banded matrix B with upper bandwidth b_u and lower bandwidth b_l. Then the application of the stochastic estimator will be successful provided that we ensure that the matrix $V \in \mathbb{R}^{n\times s},\ s = \max\{b_l, b_u\}$ will have orthonormal rows. In this case only s products are required with the matrix to obtain an exact diagonal. In contrast, probing can also be used to compute the whole matrix, but it requires $b_l + b_u - 1$ matrix-vector products.

2.4 Using optimal codes and Hadamard vectors

In light of the discussion in Section 2.2 we now study the use of vector sequences $v_k,\ k = 1, \ldots, s$ with binary entries. In particular, the following important property will be exploited [13]:

Proposition 2.2 *There exist a sequence of s vectors v_k with binary entries that achieves the bound (8), if and only if s equals 2 or is a multiple of 4.*

Hadamard matrices are a special class of matrices which satisfy the above assumptions.

Definition 2.1 *An $n \times n$ matrix $H = [h_{ij}]$ is a Hadamard matrix of order n if the entries of H are either +1 or -1 and if $HH^\top = nI$, where I is the identity matrix of order n.*

Essentially, Hadamard matrices are scaled unitary matrices with entries ± 1. Interestingly, they are known to achieve both bounds (8) and (9). A few examples of Hadamard matrices are

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

It is easy to build Hadamard matrices of high dimensions in a recursive way by exploiting Kronecker products of matrices. This is the result of the following theorem [15].

Theorem 2.1 *Given Hadamard matrices H_1 of order n and H_2 of order m, the Kronecker product of these two matrices, represented by*

$$\begin{bmatrix} h_{11}H_2 & h_{12}H_2 & \ldots & h_{1n}H_2 \\ h_{21}H_2 & h_{22}H_2 & \ldots & h_{2n}H_2 \\ \ldots & \ldots & \ldots & \ldots \\ h_{n1}H_2 & h_{n2}H_2 & \ldots & h_{nn}H_2 \end{bmatrix}$$

is a Hadamard matrix of order $n \cdot m$.

In using Theorem 2.1 it is not required to store a resulting large Hadamard matrix H, since it is possible to construct its columns on demand. In particular, when these columns are to be used in the diagonal estimator of Figure 1, we can construct them from the entries of two smaller

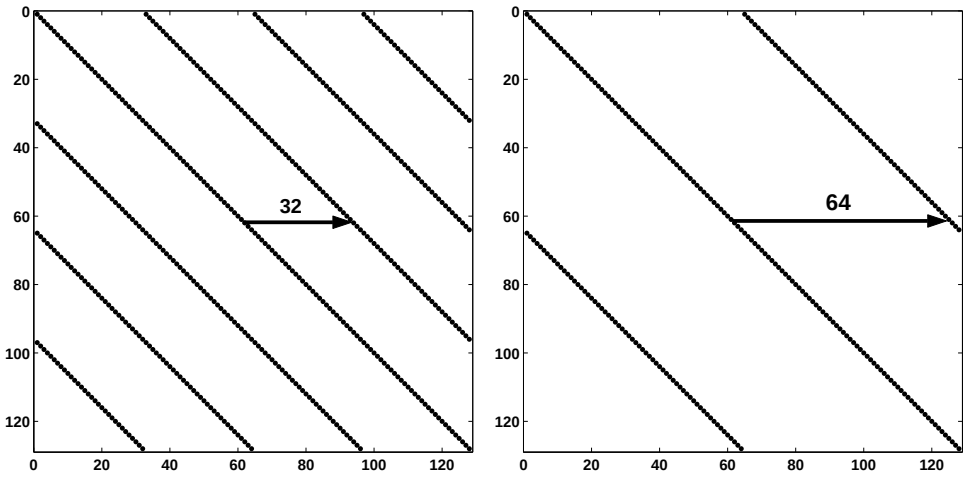

Figure 2: Pattern of the non-zero elements of the matrix $VV^\top$, where the columns of v_k of V are the first $s = 32$ (left) and $s = 64$ (right) rows of the Hadamard matrix of dimension $n = 128$.

Hadamard matrices H_1 and H_2. All we need to ensure is that $m \cdot n$ is larger than the maximum allowed number of matrix-vector products s.

Proposition 2.2 suggests that if we were to use all rows of the Hadamard matrix as the vectors v_k for the diagonal estimator, then $VV^\top = nI$. Thus, the estimator will yield exactly the diagonal of the unknown matrix. However, this would be too expensive. Hadamard rows become interesting only when we employ a few number, say s of them, where $s \ll n$. Figure 2 illustrates the non-zero pattern of the matrix $VV^\top$ when using $s = 32$ (left) and $s = 64$ (right) Hadamard rows. In the diagonal estimator the error will be induced by off-diagonal entries of the original matrix A that occupy the diagonals illustrated in plots of Figure 2. Clearly, as the number s of Hadamard rows increases we have fewer undesired diagonals.

If the matrix A is banded, with bandwidth $b = b_u + b_l$, then we need to use $2^{\lceil \log_2(\max\{b_l, b_u\}) \rceil + 1}$ Hadamard rows. Furthermore, if the off-diagonal entries of A exhibit a decaying behavior away from the main diagonal, then it is reasonable to expect that a small number s of Hadamard rows will be sufficient to yield very good accuracy.

3 Applications to computational material science

A particularly interesting case for the application of the diagonal estimator involves matrices which have a decay property, in the sense that their entries decay rapidly away from the main diagonal. Matrices of this type which arise in Density Functional Theory (DFT) calculations are "density matrices", and one critical computation is precisely to compute approximations to their diagonals. Density matrices hold the key to virtually all important properties of atomic systems in a DFT approach. In particular, its diagonal entries are equal to charge densities of the electronic distribution. The traditional way of computing this diagonal is to obtain eigenvectors of the Hamiltonian of the system associated with the n_o occupied states. This approach is reliable but expensive. In recent years, significant research efforts have been devoted

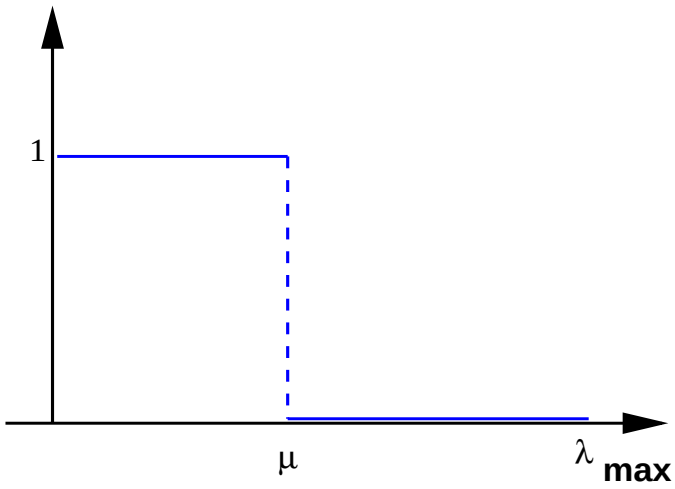

Figure 3: The Heaviside function for the eigenvalues of the Hamiltonian.

to developing methods whose cost scales linearly with the number of atoms considered. All these so-called Order-N methods (see for example [6]) are based on the decay properties of the density matrix.

A different approach is to approximate the density matrix by expanding the Fermi-Dirac operator of the Hamiltonian on a basis of Chebychev polynomials (see [10]). The density matrix is then available in the form of matrix-vector products. Therefore, estimating the diagonal of the density matrix, and thus the charge densities, is a typical application for the diagonal estimator. Furthermore, the decaying properties of density matrices can be efficiently exploited in the context of the estimator with Hadamard rows. In the following we briefly introduce the polynomial approximation of density matrices and present the application of the diagonal estimator in this case.

3.1 Computing charge densities without diagonalization

The charge density at a point r in space is commonly computed from the eigenvectors Ψ_i of the Hamiltonian matrix via the formula

$$\rho(r) = \sum_{j=1}^{n_o} |\Psi_j(r)|^2, \tag{12}$$

where the summation is taken over all occupied states. The multiplicative constant 2 which corresponds to the spin of the electrons is dropped from the above formula for simplicity.

In order to use (12), eigenvectors of the Hamiltonian are normally required. However, it is possible to compute $\rho(r)$ in a few different ways without eigenvectors. When the eigenfunction $\psi_j(r)$ is discretized with respect to r, then the charge density at a point r_i is $\rho(r_i) \equiv \rho_{ii}$, which is the diagonal entry of the functional density matrix

$$P = VV^\top \quad \text{with} \quad V = [\psi_1, \ldots, \psi_{n_o}]. \tag{13}$$

Observe that P is a projector on the subspace spanned by the eigenvectors corresponding to occupied states. Thus, any orthogonal basis V of the same space can be used. Order-N methods

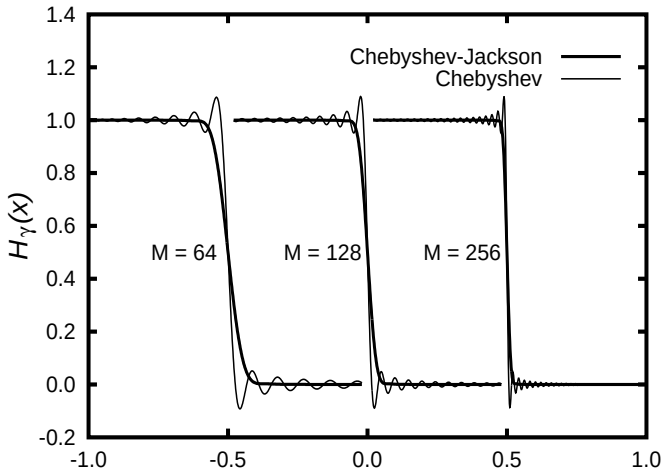

Figure 4: Approximation of the Heaviside function using Chebychev expansion with (thick lines) and without (thin lines) Jackson dampening, for increasing number of Chebychev factors $M = 64, 128$ and 256. Note that for illustration purposes we use a different chemical potential μ (which is effectively the cutoff point in the Heaviside function) for the different values of M.

are based on finding an approximation to P by exploiting some of its properties, such as near bandedness (in planewave bases), and a few relations between P and the Hamiltonian matrix.

The density matrix P can be characterized in functional form as

$$P = f(H), \tag{14}$$

where H is the Hamiltonian of the system and

$$f(\epsilon) = \frac{1}{1 + \exp(\frac{\epsilon - \mu}{k_B T})}. \tag{15}$$

The function f is the Fermi-Dirac distribution, k_B is Boltzmann's constant, μ is the chemical potential and T is the temperature. Here we are interested in the case that $T \to 0$ and thus f is just the Heaviside (or step) function, illustrated in Figure 3.

In [10] the density matrix was approximated by directly approximating the Heaviside functional by means of expansion in a basis of Chebychev polynomials. The undesired phenomenon of Gibbs oscillations in the neighborhood of λ_{n_o}, caused by the discontinuity of the original function being approximated by polynomials, can be avoided by Jackson smoothing (see Figure 4) [9, 12, 14].

The Chebychev expansion of order M, with Jackson smoothing, of the function $h : [-1, 1] \to \mathbb{R}$, is defined by

$$h(x) \approx \frac{\alpha_0}{2} + \sum_{m=1}^{M} g_{m,M} \alpha_m T_m(x), \tag{16}$$

where $g_{m,M}$ and α_m are the Jackson and the Chebychev factors respectively and $T_m(x) = \cos(m \arccos(x))$ are the Chebychev polynomials of the first kind. The density matrix (13) is

essentially a projector on the subspace spanned by eigenvectors corresponding to a number of the smallest eigenvalues of the Hamiltonian. In the density matrix case, the projection is essentially defined by the Heaviside function of H. In general, we could define projectors on arbitrary, even disjoint, parts of the spectrum. All we need is to approximate the corresponding scalar function on a basis of Chebychev polynomials (using an expansion similar to (16)) and then approximate the projection using this expansion on the matrix at hand.

It was shown in [10] that for the Heaviside function there exist explicit formulas for the factors $g_{m,M}$ and α_m, which depend on μ. Furthermore, the Chebychev polynomials $T_m(x)$ can be recursively defined according the standard three term recurrence (see [1])

$$T_{m+1}(x) = 2xT_m(x) - T_{m-1}(x), \quad T_0 = 1, \quad T_1 = x. \tag{17}$$

Let us now assume that the extreme eigenvalues $\lambda_{\min}$, $\lambda_{\max}$ of the Hamiltonian H, as well as the chemical potential μ are known. Therefore, we can shift and scale the Hamiltonian, in order to move its eigenvalues into the interval $[-1, 1]$, as well as the chemical potential, such that

$$\hat{H} = cI + dH, \quad \hat{\mu} = c + d\mu \quad \text{where} \quad c = -\frac{\lambda_{\max} + \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}} \quad \text{and} \quad d = \frac{2}{\lambda_{\max} - \lambda_{\min}}.$$

Shifting and scaling the Hamiltonian H will not affect its eigenvectors, and so the density matrix will remain the same. Applying the Chebychev expansion for the new Hamiltonian $\hat{H}$ results in the following approximation for the density matrix

$$P \approx \frac{\hat{\alpha}_0}{2} I + \sum_{m=1}^{M} \hat{g}_{m,M} \hat{\alpha}_m T_m(\hat{H}), \tag{18}$$

where I is the identity matrix. Since (18) is a polynomial expansion of P, multiplying a vector v with P results in a sequence of matrix vector products with the Chebychev polynomials. In Figure 5 we illustrate an algorithm (Algorithm 2) for the multiplication of the approximated P by a vector v. The vector $T_{k-1}(\hat{H})x$ denotes the product of the input vector x with the Chebychev polynomial matrix $T_{k-1}(\hat{H})$. Clearly, the 3-term recurrence (17) of the Chebyshev polynomials is exploited for this calculation. At each step k, where $k = 2 : M$, we need to multiply the Hamiltonian with the vector $T_{k-1}(\hat{H})x$. The computational cost of Algorithm 2 amounts to $(M-2)$ matrix vector multiplications and $(M-1)$ `xAXPY` operations (with each of them costing $O(N)$ operations).

Due to the low cost of Algorithm 2, approximating the diagonal of the density matrix, and thus the charge densities, appears to be a good candidate for the proposed diagonal estimator which requires only matrix-vector products. Furthermore, when the density matrix exhibits significant decay properties, one may expect that satisfactory accuracy to the diagonal entries can be achieved by applying only a limited number of matrix-vector products. Clearly, the potential for cost reduction compared to diagonalization of the Hamiltonian is significant and this provided the initial motivation for this research.

3.2 Using Chebychev vectors

Recall that M is the number of Chebychev polynomials in the expansion (18) and s the number of vectors employed in the diagonal estimator. Using Algorithm 2 in the diagonal estimator (Fig. 1, line 2.2), will require $O(M \times s)$ matrix-vector products with the Hamiltonian matrix.

Algorithm 2

(* *Input* *) Hamiltonian $\hat{H}$, coefficients $\hat{\alpha}_k, g_k, k = 1, \ldots, M$
input vector x
(* *Output* *) Approximate vector v such that $v \approx Px$,
where P is the density matrix

1. $T_0(\hat{H})x = x$
2. $T_1(\hat{H})x = \hat{H}x$
3. $v = \frac{1}{2}\alpha_0 x + \hat{g}_1\hat{\alpha}_1 T_1(\hat{H})x$
4. **for k=2:M**
   4.1 $T_k(\hat{H})x = 2\hat{H}T_{k-1}(\hat{H})x - T_{k-2}(\hat{H})x$
   4.2 $v = v + \hat{g}_k\hat{\alpha}_k T_k(\hat{H})x$
5. **end**

Figure 5: Algorithm for multiplication of the Chebychev approximation of the density matrix P with a vector x.

If convergence is slow, meaning that s needs to be large, then the total number of matrix-vector products with the Hamiltonian could become unacceptably large.

One remedy to this problem is to exploit the recurrence relations of Chebychev polynomials, and to use as v_k's in the diagonal estimator the vectors of the Chebychev basis itself. Specifically, assume for simplicity that the eigenvalues of the Hamiltonian are between -1 and 1 so that the expansion (18) for the density matrix P holds.

We first recall a well known property of Chebychev polynomials which is a simple consequence of elementary trigonometric identities:

$$\begin{aligned} T_m(x)T_k(x) &= \cos(m\arccos(x))\ \cos(k\arccos(x)) \\ &= \frac{1}{2}(\cos((m+k)\arccos(x)) + \cos((m-k)\arccos(x))) \\ &= \frac{1}{2}(\cos((m+k)\arccos(x)) + \cos((|m-k|)\arccos(x))) \\ &= \frac{1}{2}(T_{m+k}(x) + T_{|m-k|}(x)). \end{aligned} \tag{19}$$

Consider now the following sequence of vectors,

$$v_k = T_k(\hat{H})v_0, \quad k = 1, \ldots, s,$$

where T_k is the Chebychev polynomial of order k and v_0 is some initial vector. If Pv^k is computed for the purpose of estimating the diagonal of P, we find that

$$Pv_k \approx \frac{\hat{\alpha}_0}{2}v_k + \sum_{m=1}^{M} \hat{g}_{m,M}\hat{\alpha}_m T_m(\hat{H})v_k \tag{20}$$

$$= \frac{\hat{\alpha}_0}{2}v_k + \sum_{m=1}^{M} \hat{g}_{m,M}\hat{\alpha}_m T_m(\hat{H})T_k(\hat{H})v_0 \tag{21}$$

$$= \frac{\hat{\alpha}_0}{2} v_k + \frac{1}{2} \sum_{m=1}^{M} \hat{g}_{m,M} \hat{\alpha}_m [T_{m+k}(\hat{H}) v_0 + T_{|m-k|}(\hat{H}) v_0] \quad (22)$$

$$= \frac{\hat{\alpha}_0}{2} v_k + \frac{1}{2} \sum_{m=1}^{M} \hat{g}_{m,M} \hat{\alpha}_m [v_{m+k} + v_{|m-k|}]. \quad (23)$$

For each (approximate) product Pv_k, only $M+1$ vector additions (`xAXPY` operations) are required, which is far cheaper than the cost of Algorithm 2 which multiplies a general (random) vector with the density matrix P. Of course, we still have to compute the Chebychev basis $[v_0, \ldots, v_s]$. Once more, we can take advantage of the three term recurrence of Chebychev polynomials and reduce the computational cost, since

$$v_k = T_k(\hat{H}) v_0 = 2\hat{H} T_{k-1}(\hat{H}) v_0 - T_{k-2}(\hat{H}) v_0, \quad k = 1, \ldots, s. \quad (24)$$

When a large number s of Chebychev vectors are needed in the diagonal estimator, then computing these vectors *a priori* will quickly become intractable due to memory requirements, especially when the Hamiltonian H is large. Thus, we need to compute the Chebychev vectors only when we need them. In order to compute the k-th product Pv_k, when $k > M$ only vectors $k - M$ through $k + M$ are required. Therefore, storage requirements for each product Pv_k can be limited to $2M+1$ Chebychev vectors. Also when $k \leq M$, then we need a few of the vectors v_0 to v_{k+M}, the number of which does not exceed $2M + 1$. The procedure is outlined in Algorithm 3 (see Figure 6). For the sake of clarity and simplicity, the stochastic estimator described in Figure 6 starts from the k-th Chebychev vector where $k = M + 1$.

A final note on the method described in this section is that the whole process can be restarted. In particular, when the maximum number of steps s has been reached, we compute a new set of initial Chebychev vectors $v_k, k = 0, \ldots M$, starting with a new random vector v_0. However, it is necessary to keep the partial sums in vectors q_k and r_k (see Algorithm 3). The rationale behind the restarting strategy is to avoid possible stagnation of the convergence of the estimator, since the three term relation between the Chebychev vectors suggests that they are correlated and not completely random.

4 Experimental evaluation

In this section we test the performance of the diagonal estimator in a variety of situations. The experiments were performed using `MATLAB 6.5`.

4.1 Experiments with sample sparse banded test matrices

This section investigates the effectiveness of the proposed diagonal estimator with random vectors and Hadamard rows, using test matrices from the Matrix Market[2]. In particular, we used the following matrices: `GRE_512`, `MHD416A`, `AF23560`, `NOS6`, `BCSSTK07` and `ORSREG_1`. A few details on these matrices are given in Table 1. They are all banded as well as sparse (in particular, they are not dense within the band.) This property readily suggests that the diagonal estimator with Hadamard rows should give very accurate results with a small number of matrix-vector products.

[2] `http://math.nist.gov/MatrixMarket`

Algorithm 3: Diagonal Estimator with Chebychev vectors for the density matrix P

(* *Input* *) Hamiltonian $\hat{H}$, coefficients $\hat{\alpha}_m, g_m, m = 1, \ldots, M$,
initial Chebychev vectors $v_k, k = 0, \ldots, M$,
total number of steps s
(* *Output* *) Approximate diagonal D^s of density the matrix P

1. Generate the first M Chebychev vectors via the three term recurrence (17)
2. $d_0 = 0,\ q_0 = 0,\ r_0 = 0$
3. **for k=M+1 : s**
   3.1 Generate current Chebychev vector $v_k = 2\hat{H}v_{k-1} - v_{k-2}$
   3.2 $t_{k,0} = \frac{\hat{\alpha}_0}{2} v_k$
   3.3 **for m=1 : M**
      3.3.1 $t_{k,m} = t_{k,m-1} + \frac{1}{2}\hat{g}_{m,M}\hat{\alpha}_m(v_{k+m} + v_{k-m})$
   3.4 **end**
   3.5 $q_k = q_{k-1} + (t_{k,M} \odot v_k)$
   3.6 $r_k = r_{k-1} + (v_k \odot v_k)$
   3.7 $d_k = q_k \oslash r_k$
4. **end**
5. Set $D^s = d_s$

Figure 6: Algorithm 3. Diagonal estimator for the density matrix using Chebychev vectors.

In our first experimental set, random vectors v_k are used for the diagonal estimator. The entries of the vectors were drawn from a normal distribution with zero mean, variance one and standard deviation one (we used the `MATLAB` routine `randn`). In Table 2 we illustrate the mean relative error for the diagonal entries of these matrices, error $= \frac{1}{n}\sum_{i=1}^{n} |(d_i - \hat{d}_i)/d_i|$. We employed an increasing number of random vectors. Note that there are matrices for which we used a number of vectors that is larger than the size of the matrix. However, as it is pointed out in the introduction of this paper, for a matrix with n entries on the main diagonal only n vectors (canonical basis) are required. In these cases we did not restrict the number of vectors used to $s = n$, but rather we preferred to use the same number of vectors across all test matrices. We note that the diagonal estimator quickly yields somewhat accurate approximations for a very small number of vectors ($s = 10$). However, convergence, although steady in most cases, appears to be slow for subsequent steps.

We next investigate the diagonal estimator with Hadamard rows. Table 3 illustrates the results. The second column contains again the mean relative error. As expected, the Hadamard estimator requires far fewer matrix-vector products than the stochastic estimator. In some cases four vectors already suffice to yield exact results. For example, although the total bandwidth of the matrix `NOS6` is sixty-two (62), we are able to accurately compute its diagonal using only four Hadamard rows. This is due to the fact that the "forbidden diagonals" of the matrix $VV^\top$ coincide only with zero entries of the matrix at hand.

Matrix	size	nnz	lower b/w	upper b/w
GRE_512	512×512	4140	48	48
MHD416A	416×416	8562	25	65
AF23560	23560×23560	484256	305	305
NOS6	675×675	1465	31	31
BCSSTK07	420×420	4140	48	48
ORSREG_1	2205×2205	14133	442	442

Table 1: Characteristics of test matrices from Matrix Market. The matrices are sparse and banded. The third column illustrates the number of non-zero elements, while the fourth and fifth columns illustrate the upper and lower bandwidth of the matrices respectively.

4.2 Matrices with decaying off-diagonal entries

We now experiment with matrices whose off-diagonal elements exhibit a decaying behavior. We use the diagonal estimator with Hadamard rows.

We consider the following family of (dense) matrices:

$$A \in \mathbb{R}^{n\times n},\ \alpha_{ij} \quad = \quad \begin{cases} 1 & i = j \\ \frac{1}{|i-j|^k} & i \neq j \end{cases} \tag{25}$$

for $k = 1, 2, 3$ and 4. In addition, we also consider sparse versions of the above matrices: We randomly set a percentage c of the off-diagonal entries of the matrix A to zero.

We start with the dense versions of the matrix A and consider $n = 3000$ (experiments with different values for n gave very similar results). The left plot in Figure 7 illustrates the mean error $\frac{1}{n}\sum_{i=1}^{n} |(d_i - \hat{d}_i)|$ (note that we do not divide with d_i since it is equal to 1). For the sparse case, we experimented with $n = 10000$ and an average `nnz` ≈ 100 non-zero elements per row. Again the magnitude non-zero elements of the test matrix are defined according to (25). The results are illustrated in the right plot of Figure 7. In both cases it is seen that the stronger the decay, the better the accuracy achieved by the diagonal estimator with Hadamard rows. Our experience with sparse matrices with decaying off-diagonal elements suggests that the proposed diagonal estimator can yield quite satisfactory results. In many instances, 32 Hadamard rows or fewer are enough to get 3-4 digits of accuracy.

4.3 Experiments with density matrices

We now experiment with density matrices and Hadamard rows in our diagonal estimator. In particular, we use a parametrized model Hamiltonian that has been well studied in the context of evaluating iterative algorithms for the approximation of charge densities [11, 17]. In this model there are N_b bands and N_s substates per band. The diagonal entries of the Hamiltonian $(i = i', j = j')$

$$H_{ij,ij} = (i-1)\Delta + (j-1)\delta, \quad \delta \ll \Delta. \tag{26}$$

For the intraband coupling terms $(i = i', j \neq j')$

$$H_{ij,ij'} = Ce^{-|j-j'|} \tag{27}$$

AF23560		GRE_512		MHD416A	
♯ vectors	error	♯ vectors	error	♯ vectors	error
10	0.28	10	0.60e-1	10	4.3
20	0.19	20	0.41e-1	20	5.2
30	0.16	30	0.33e-1	30	2.9
40	0.13	40	0.27e-1	40	2.6
50	0.12	50	0.25e-1	50	2.9
100	0.084	100	0.19e-1	100	2.2
200	0.059	200	0.11e-1	200	1.1
500	0.037	500	0.71e-1	500	0.7
1000	0.027	1000	0.49e-2	1000	0.77
ORSREG_1		BCSSTK07		NOS6	
♯ vectors	error	♯ vectors	error	♯ vectors	error
10	0.22	10	0.32e-2	10	1.00e-4
20	0.15	20	0.19e-2	20	0.72e-4
30	0.12	30	0.16e-2	30	0.58e-4
40	0.10	40	0.13e-2	40	0.45e-4
50	0.091	50	0.11e-2	50	0.40e-4
100	0.065	100	0.86e-3	100	0.30e-4
200	0.046	200	0.65e-3	200	0.21e-4
500	0.029	500	0.39e-3	500	0.13e-4
1000	0.020	1000	0.26e-3	1000	0.10e-5

Table 2: Mean absolute relative errors of the stochastic diagonal estimator using increasing number of random vectors. The error (columns 1,3 and 5) is computed as error $= \frac{1}{n}\sum_{i=1}^{n} |(d_i - \hat{d}_i)/d_i|$, where n is the dimension of the matrix.

and for the interband coupling terms $(i \neq i', j \neq j')$

$$H_{ij,i'j'} = \frac{C}{n_{od}(|i-i'|+1)}\ e^{-|j-j'|}, \tag{28}$$

where in all of the above $i = 1, \ldots, N_b$ and $j = 1, \ldots, N_s$. Table 4 summarizes the parameters used (based on those used in [11, 17]). Since this model Hamiltonian is a dense matrix, its size is kept relatively low, specifically we take $N_b \times N_s = 2000$. For the parameter n_{od} which affects the decaying properties of the off-diagonal entries of the density matrix we used the values $5, 50, 500$ and 5000.

We emphasize that when calculating charge densities we employ an approximation at two levels. First, we rely on the Chebychev expansion (18) to approximate the density matrix P and then we employ the diagonal estimator in order to approximate the charge densities (diagonal of P). In this paper we are primarily interested in studying the approximating qualities of the diagonal estimator. Therefore in what follows we test its results compared to the (exact) diagonal of the approximated density matrix $\tilde{P}$ (right hand side of the polynomial expansion (18)) rather than the exact charge densities.

In order to approximate the density matrix P we have used a Chebychev expansion of degree

AF23560		GRE_512		MHD416A	
♯ vectors	error	♯ vectors	error	♯ vectors	error
4	0.99	4	0.14	4	0.56
8	0.5	8	0.064	8	0.50
16	0.0028	16	0.049	16	0.0039
32	0	32	0.015	32	0
64	0	64	0.012	64	0
128	0	128	0	128	0
256	0	256	0	256	0
512	0	512	0	512	0
1024	0	1024	0	1024	0
ORSREG_1		BCSSTK07		NOS6	
♯ vectors	error	♯ vectors	error	♯ vectors	error
4	0	4	0.81e-3	4	0
8	0	8	0.26e-3	8	0
16	0	16	0	16	0
32	0	32	0	32	0
64	0	64	0	64	0
128	0	128	0	128	0
256	0	256	0	256	0
512	0	512	0	512	0
1024	0	1024	0	1024	0

Table 3: Mean absolute relative errors of the diagonal estimator using increasing number of Hadamard rows. The error (columns 1,3 and 5) is computed as error $= \frac{1}{n}\sum_{i=1}^{n}|(d_i - \hat{d}_i)/d_i|$, where n is the dimension of the matrix.

$M = 32$. In all cases of n_{od} we are interested in the density matrix associated with the 50 smallest eigenvalues of the Hamiltonian.

Table 5 illustrates the results. We point out that larger values of the parameter n_{od} tend to increase the diagonal dominance of the resulting density matrix by decreasing the magnitude of its elements away form the main diagonal. It is clear, that at least 128 Hadamard rows are required for the case $n_{od} = 5$, while only 16 Hadamard rows are required to achieve roughly the same accuracy for the Hamiltonian with $n_{od} = 5000$. Since we approximate the density matrix on a set of $M = 32$ Chebychev polynomials the latter case results to $128 \times 32 = 4096$ matrix-vector products with the Hamiltonian, which induces a cost similar to that of direct diagonalization of the Hamiltonian. However, one must appreciate that the diagonal estimator with Hadamard rows (or with random vectors) has very limited memory requirements, which can be considered to be negligible compared to the memory requirements of direct diagonalization.

On the other hand, for $n_{od} = 5000$ only $16 \times 32 = 512$ matrix-vector products are required, leading to a much more modest cost. It is exactly for such cases that we can expect the diagonal estimator to provide a competitive alternative to standard methods which rely on explicit eigenvalue calculations.

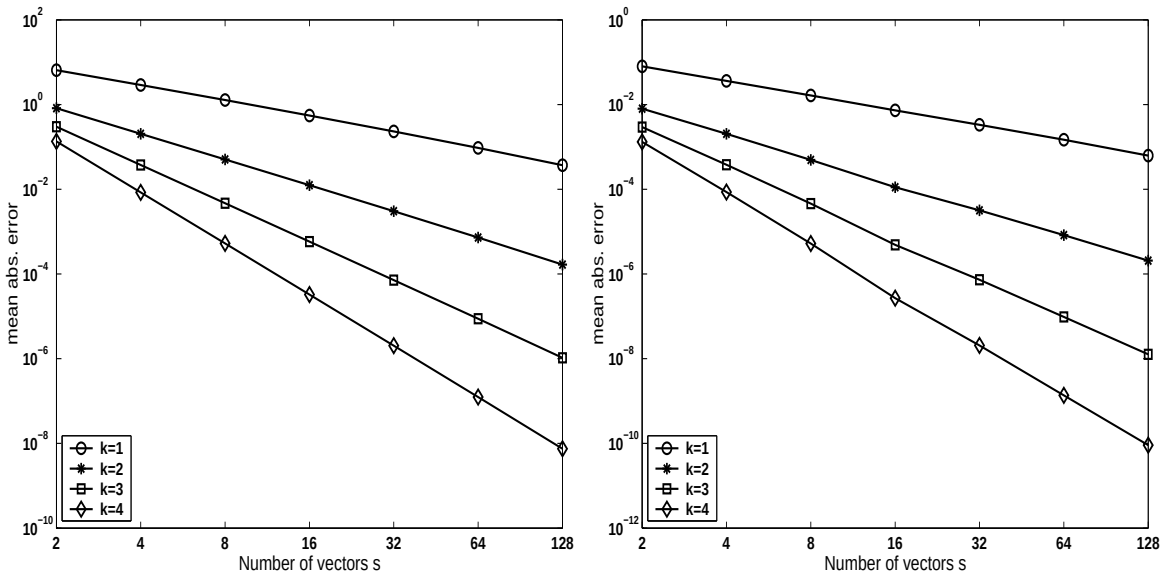

Figure 7: Mean absolute error for the family of matrices (25). Left: dense case with $n = 3000$. Right: Sparse case with $n = 10000$ and an average of 100 non-zero elements per row.

N_b	N_s	δ	Δ	C	n_{od}
10	200	10^{-4}	10^{-1}	10^{-1}	$*$

Table 4: Parameters used for the model Hamiltonian (26)-(28). For n_{od} we used the values $5, 50, 500$ and 5000.

4.3.1 Diagonal estimator with Chebychev vectors

When a large number s of vectors is required in the diagonal estimator and at the same time we need a long Chebychev expansion (large M) then, clearly the overall cost rapidly increases as $s \times M$. One can anticipate this to pose significant problems when very large Hamiltonians are considered. In order to avoid this problem, we suggested using Chebychev vectors in the diagonal estimator (see Sec. 3.2). Clearly, these vectors are correlated and thus convergence of the diagonal estimator is anticipated to be slow. However, an inspection of Algorithm 3 (see Fig. 6) shows that at each step only $2M + 1$ vectors are needed in memory rendering the method particularly suitable for very long simulations. In Fig. 8 we illustrate the resulting diagonal using Chebychev vectors for the model Hamiltonian (26)-(28) with $n_{od} = 5000$. We used 30 restarts and $s = 1024$ Chebychev vectors at each restart. Again $M = 32$ and we approximated the diagonal of the density matrix associated with the 50 smallest eigenvalues of the Hamiltonian. Our goal in this experiment is not to show that this setting is competitive but rather to illustrate that using Chebychev vectors can achieve convergence. Clearly, for large values of the charge density the estimator yields satisfactory results, with early converge in the simulation. For small values of charge densities we witness a significant error. Our experience has showed that very small values of charge densities require very long simulations in order to be captured accurately. On the other hand, very small values of charge densities can be safely

	Number of Hadamard rows					
n_{od}	4	8	16	32	64	128
5	21.7	24.5	12.8	7.2	3.9	1.6e-2
50	3.2	3.6	1.6	8.6e-1	4.5e-1	2.1e-3
500	3.4e-1	3.5e-1	1.7e-1	8.7e-2	4.6e-2	2.21e-4
5000	6.3e-2	2.64e-2	1.8e-2	8.8e-3	4.6e-3	2.28e-5

Table 5: Mean absolute relative errors of the diagonal estimator using increasing number of Hadamard rows and increasing parameter n_{od} for the model Hamiltonian (26)-(28). The error (columns 2-7, rows 3-6) is computed as $= \frac{1}{n}\sum_{i=1}^{n} |(d_i - \hat{d_i})/d_i|$, where n is the dimension of the matrix.

treated as being zero, and thus it is only the order of the magnitude of the charge density that is of interest, which is captured by the estimator.

5 Conclusion

We presented a few algorithms for the approximation of the diagonal of a matrix in the case where the matrix is not readily available but it is easy to compute its product with an arbitrary vector.

We extended ideas based on statistical arguments, due to Hutchinson [8], Girard [5] and others, and described a "stochastic" diagonal estimator which uses random vectors. The analysis of the convergence properties of the estimator led us to make connections between the problem at hand with the problem of line packing in Grassmannian spaces and that of the design of optimal codebooks in communications. This suggested utilizing rows of Hadamard matrices, instead of random vectors, that can effectively take advantage of special properties of the matrix at hand, such as bandedness or strong decaying of the magnitude of off-diagonal elements of the matrix. We should point out that we did not perform an exhaustive search for "optimal" sequences of vectors, the existence of which is predicted by the Welch bounds of Section 2.2.1. We only showed that Hadamard rows can yield significantly better results than the simple use of random vectors.

In Density Function Theory, a primary method used in electronic structures calculations, one is faced with the problem of computing electronic charge densities. In matrix terms, these are the diagonal entries of a projector associated with a number of the smallest eigenvalues of the Hamiltonian. Previous work by several authors has shown that this density matrix can be accurately expanded in polynomial bases, rendering it an ideal candidate for the diagonal estimator presented in this paper. Our claim and hope is that in several interesting situations which lead to sufficient decay of the off-diagonal entries, the diagonal estimators discussed in this paper can be used as effective and inexpensive alternatives to standard eigenvalue-based methods, in part because of their very small memory requirements and their simplicity.

References

[1] C. C. Cheney. *Introduction to approximation theory.* McGraw-Hill, New York, 1966.

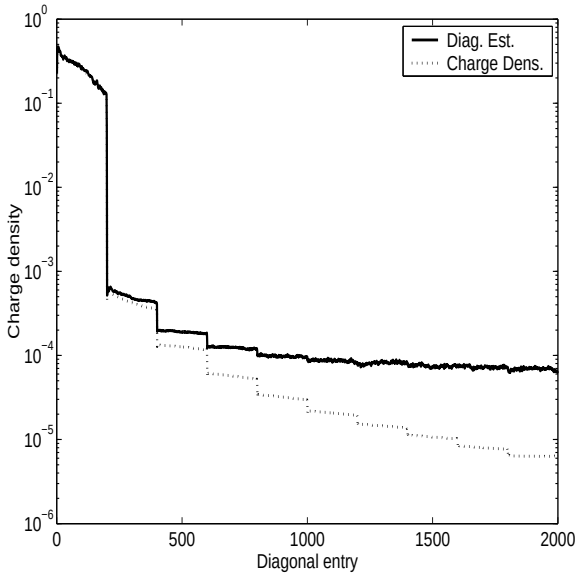

Figure 8: Approximate charge densities for the model Hamiltonian (26)-(28) with $n_{od} = 5000$. Solid line: Approximated diagonal using the diagonal estimator with Chebychev vectors (Algorithm 3). Dotted line: Exact diagonal of the approximated density matrix.

[2] Thomas F. Coleman and Jorge J. Moré. Estimation of sparse Jacobian matrices and graph coloring problems. *SIAM J. Numer. Anal.*, 20(1):187–209, 1983.

[3] J. H. Conway, R. H. Hardin, and N. J. A. Sloane. Packing lines, planes etc.: Packings in Grassmannian spaces. *Exper. Math.*, 5(2):139–159, 1996.

[4] A. R. Curtis, M. J. D. Powel, and J. K. Reid. On the estimation of sparse jacobian matrices. *J. Inst. MAth. Appl.*, 13:117–119, 1974.

[5] D. Girard. Un algorithme simple et rapide pour la validation croissee generalisee sur des problemes de grande taillee. *RR 669-M, Grenoble, France: Informatique et Mathématiques Appliquées de Grenoble.*, 1987.

[6] S. Goedecker. Linear scaling electronic structure methods. *Reviews of Modern Physics*, 71:1085–1123, 1999.

[7] S. Goedecker and M Teter. Tight-binding electronic-structure calculations and tight-binding molecular dymanics with localized orbitals. *Phys. Rev. B*, 51:19455, 1995.

[8] M. F. Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *J. Commun. Statist. Simula.*, 19(2):433–450, 1990.

[9] D. Jackson. *The theory of approximation*, volume XI of Amer. Math. Soc. Aolloq. Publ. Amer. Math. Soc., Providence, RI, 1930.

[10] L. O. Jay, H. Kim, Y. Saad, and J. R. Chelikowsky. Electronic structure calculations in plane-wave codes without diagonalization. *Comput. Phys. Comm.*, 118:21–30, 1998.

[11] G. A. Parket, W. Zhu, Y. Huang, D. K. Hoffman, and D. J. Kouri. Matrix pseudo-spectroscopy: iterative calculation of matrix eigenvalues and eigenvectors of large matrices using a polynomial expansion of the dirac delta function. *Comp. Phys. Comm.*, 96:27–35, 1996.

[12] T. J. Rivlin. *An Introduction to the Approximation of Functions*. Dover, 1969.

[13] D. V. Sarwate. *Metting the Welch bound with equallity*, pages 79–102. Springer, 1999.

[14] R. N. Silver, H. Roeder, A. F. Voter, and J. D. Kress. Kernel polynomial approximations for densities and spectral functions. *J. Comput. Phys.*, 124:115–130, 1996.

[15] W.D. Wallis, A. Penfold Street, and J. S. Wallis. *Combinatorics: Room Squares, Sum-Free Sets, Hadamard Matrices*. Lecture Notes in Mathematics, 292. Springer-Verlag, 1972.

[16] L. Welch. Lower bounds on the maximum cross correlation of signals. *IEEE Transactions on Information Theory*, 20(3):397–399, May 1974.

[17] R. E. Wyatt. Matrix spectroscopy: Computation of interior eigenstates of large matrices using layered iteration. *Phys. Rev. E*, 51:3643–3658, 1995.

[18] P. Xia, S. Zhou, and G. B. Giannakis. Achieving the Welch bound with difference sets. *To appear in IEEE Trans. on Information Theory*, 2005.