A Monte Carlo Approach to Computing the Trace of $A^{-1}$ Using Scattering Amplitude Theory

Motivation
In electronic structures problems, it often arises that one has to compute the diagonal elements $(A^{-1})_{ii}$ of the inverse of a given matrix $A$. This problem can be solved with direct sparse methods, or iterative methods, or the method we are proposing: the scattering amplitude algorithm.

Problem Definition
Scattering amplitude is intensity of a flux through a fixed point in space, and is defined as $c^TA^{-1}b$ for any non-singular matrix $A$, and two vectors (sender/receiver) $b$ and $c$. (Notation: $Z(A, b, c) = Z = c^TA^{-1}b$)

With the correct choice of $b$ and $c$, the diagonal elements of the inverse $A^{-1}_{ii}$ can be easily computed using the scattering amplitude in the following way:

1. create two linear systems: $Ax = b$ and $A^*y = c$, where $A^*$ is the transpose conjugate of $A$.
2. solve the systems using Krylov methods: $x^{(k)} \rightarrow x$ and $y^{(k)} \rightarrow y$.
3. stop the iterative method when $\|x-x^{(k)}\| \rightarrow 0$, and $\|y-y^{(k)}\| \rightarrow 0$.
4. use the following formula for $Z$, with complex dot product defined as: $(a, b) = a^*b$.

$$\langle (A^*y^{(k)} - c)A^{-1}(Ax^{(k)} - b) \rangle = \langle (A^*y^{(k)} - c)(x^{(k)} - A^{-1}b) \rangle$$

$$= (y^{(k)})^*Ax^{(k)} - (y^{(k)})^*b - c^*x^{(k)} + c^*A^{-1}b$$

$$= \langle A^*y^{(k)}, x^{(k)} \rangle - \langle y^{(k)}, b \rangle - \langle c, x^{(k)} \rangle + \langle c, x \rangle$$

Then: $Z = \langle c, x \rangle = -\langle A^*y^{(k)}, x^{(k)} \rangle + \langle y^{(k)}, b \rangle + \langle c, x^{(k)} \rangle + \mathcal{O}(\varepsilon_k^2)$.

Convergence
The convergence of the scattering amplitude method is quadratic. Let $w$ be defined as the following inner product: $w = \langle A^*y^{(k)} - c, A^{-1}(Ax^{(k)} - b) \rangle$ and $w \rightarrow 0$.

Then: $|w| \leq \|A^{-1} \| \|A^*y^{(k)} - c \| \|Ax^{(k)} - b \|$.

And also let’s define $\varepsilon_k = max\{\varepsilon_y^{(k)}, \varepsilon_x^{(k)}\}$, where $\varepsilon_y^{(k)} = \|Ay^{(k)} - c\|$ and $\varepsilon_x^{(k)} = \|Ax^{(k)} - b\|$.

Then $|w| \leq \mathcal{O}(\varepsilon_k^2)$. QED.

Results
The scattering amplitude algorithm can be used to compute the trace of the inverse of a matrix, by calculating the individual diagonal entries on the inverse of the matrix and then summing them up. The most obvious choice for $b$ and $c$ would be the column vectors of the identity, since $A^{-1}_{ii} = e_i^TA^{-1}e_i$; however this means using the algorithm $n$ times, where $n$ is the number of columns in the matrix.

Another choice is using a Monte Carlo approach, in which the trace is computed by summing up only a sample of the diagonal elements of the inverse. And the samples are computed with random right-hand side vectors that have elements uniformly distributed on the unit circle. Although less accurate, this approach is more effective, as it can be seen from Figure 1. The Krylov method of choice is the bi-conjugate gradient method or BiCG. The tolerance of the method is fixed (less $10^{-6}$ and controlled by a maximum number of iterations, which also determines the number of matvecs). BiCG requires two matvecs per iteration. The matrix used is non-symmetric and complex (identified as Cu686), with dimension $10976 \times 10976$. Figure 2 is similar to Figure 1, except that more samples are used. However, using more samples, does not increase the accuracy by too much (5 samples give a relative error of $10^{-3}$, while to get to an accuracy of $10^{-4}$, 100 samples are enough; in our tests, we have not tried more than 400 samples, but probably 1000 samples would give a relative error of $10^{-5}$). A 100 samples on a regular desktop took about 300 seconds.

If the trace needs to be computed for blocks of a matrix, let’s say of dimension $m$, the same algorithm works with the following modification: the right-hand side vectors should be initialized with zero, and then only $m$ elements that correspond to the block for which the trace needs to be computed should be uniformly distributed random numbers from the unit circle. Preliminary tests have shown that one or two samples compute the trace of the block to within a $10^{-3}$ relative error.

The scattering amplitude algorithm presented in this report can be easily parallelized, by assigning several blocks to each processor. And thus, relative to the errors incurred in the BiCG and for small blocks (e.g. $m = 16$), the trace can be computed exactly (using elementary vectors).
Figure 1: The relative error for different number of samples as the total number of matvecs are predetermined by a fixed tolerance and maximum number of iterations.

Figure 2: The relative error for different number of samples as the total number of matvecs are predetermined by a fixed tolerance and maximum number of iterations.