Multiple extremal eigenpairs by the power method

J.E. Gubernatis [a,*], T.E. Booth [b]

[a] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[b] *Applied Physics Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

A R T I C L E   I N F O

A B S T R A C T

We report the production and benchmarking of several refinements of the power method that enable the computation of multiple extremal eigenpairs of very large matrices. In these refinements we used an observation by Booth that has made possible the calculation of up to the 10th eigenpair for simple test problems simulating the transport of neutrons in the steady state of a nuclear reactor. Here, we summarize our techniques and efforts to-date on determining mainly just the two largest or two smallest eigenpairs. To illustrate the effectiveness of the techniques, we determined the two extremal eigenpairs of a cyclic matrix, the transfer matrix of the two-dimensional Ising model, and the Hamiltonian matrix of the one-dimensional Hubbard model.

© 2008 Published by Elsevier Inc.

1. Introduction

Computing eigenpairs of large matrices is a ubiquitous problem in computational physics. In this paper, we present several refinements of the basic power method that enable the efficient and accurate computation of multiple extremal eigenvalues of very large matrices. Ultimately, our objective is producing Monte Carlo versions of such methods for matrices whose orders are so large that even the eigenvectors cannot be stored in computer memory. For such problems, the computation of a basic vector quantity as the inner product is generally either very inefficient or impractical. It can be impractical, for instance, because the nature of Monte Carlo sampling means most components of these vectors are unknown. Here, we focus on the basic algorithms developed to date, noting they work well when used deterministically. Novel will be the illustration of how the power method can be expanded to compute several extremal eigenpairs simultaneously rather than just one at a time. While various versions of the power method often compute very well the dominant eigenvalue λ_1, the one with largest absolute value, computing subdominant eigenvalues $\lambda_2, \lambda_3, \ldots$ has often proven much more difficult and is much less frequently attempted.

The algorithms to be presented use some recent insights of Booth [1,2] that were developed for Monte Carlo simulations of steady state neutron transport in nuclear reactors. Initially, he proposed a novel modification of the power method that has produced up to 10 eigenpairs for simple test problems. Here, we present refinements of these insights and focus on determining just λ_1 and λ_2, plus their eigenvectors. The convergence of the power method is well known to slow as the ratio λ_2/λ_1, sometimes called the dominance ratio, approaches unity. As such, this ratio is an indicator of solution difficulty and acceptability. As we will illustrate, an advantage of computing two dominant eigenpairs simultaneously is often improved convergence to the first one. An advantage of the present techniques is the ease in getting both eigenfunctions along with their eigenvalues.

It is important to note that various areas of science and engineering seek multiple eigenpairs for reasons other than algorithmic gains. In nuclear engineering, a dominance ratio distinct from unity is an acceptance qualifier for various nuclear

---

* Corresponding author. Tel.: +1 505 667 6727; fax: +1 505 665 4063.
 E-mail address: jg@lanl.gov (J.E. Gubernatis).

criticality safety assessments and nuclear reactor designs [3]. In statistical physics, a dominance ratio nearing unity, on the other hand, is often a condition sought. Near a continuous phase transition, $\lambda_2 \to \lambda_1$, and λ_2/λ_1 controls the microscopic spatial correlations among *physical* degrees of freedom [4]. Today, an important topic in quantum statistical mechanics is quantum critical phenomena, phase transitions driven by zero-point motion at zero temperature [5]. Here, it is the two smallest eigenvalues of the Hamiltonian matrix describing the physical system that are of interest. The quantum critical phenomenon construct, while supplemented by a few exact solutions to some very simple problems, is largely phenomenological in part because of the inability to compute λ_2 for models of direct physical relevance.

In the next section, Section 2, we summarize some basic notions about the power method and our refined procedures. For simplicity, we will assume the two largest extremal eigenpairs are sought. Also we restrict attention to systems with real eigenvectors and eigenvalues, but our methods can be applied to complex systems as well. In Section 3, we apply these techniques to determination of a few eigenpairs of three problems. The first is the cyclic matrix that results from the discretization of the gradient operator on a circle, the second is the transfer matrix of the two-dimensional Ising model, and the third is the Hamiltonian matrix of the one-dimensional Hubbard model. For the first and third problems, we determine the smallest two eigenpairs (ground-state and first excited state) instead of the largest ones to illustrate the flexibility of the techniques. The second and third problems counterpose in their computational challenges in a number of ways: The transfer matrix for the Ising model is non-symmetric, positive, and dense. Its eigenvalues are known analytically and all its matrix elements follow a single simple analytic expression. The Hamiltonian matrix for the Hubbard model is symmetric, indefinite, and sparse. Its two smallest eigenvalues are not known analytically, and its matrix elements, while easy to compute, lack a simple expression. In Section 4, the final section, we will discuss extensions of the techniques to broader classes of problems, including those involving continuous operators.

2. Methodology

We first summarize the power method, and then we discuss ways to refine it so convergence is to the two largest extremal eigenpairs simultaneously. We conclude this section with two refinements of the power method: one is necessary for the reduction of round-off error and the other improves the convergence rate to the dominant eigenpair while simultaneously calculating the second extremal eigenpair.

2.1. A. Power method basics

For some real-valued $N \times N$ matrix A, not necessarily symmetric, we will be concerned with the N eigenpairs (λ_i, ψ_i) satisfying

$$A\psi_i = \lambda_i \psi_i \tag{1}$$

In the simplest application of the power method [6], an iteration is started with some ψ, normalized in some convenient, but otherwise relatively arbitrary, manner and consists of iterating two steps

$$\begin{aligned} \phi &= A\psi \\ \psi &= \phi/\|\phi\| \end{aligned} \tag{2}$$

If we write

$$\psi = \sum_{i=1}^{N} \alpha_i \psi_i$$

then after n iterations

$$A^n \psi = \lambda_1^n \left[\alpha_1 \psi_1 + \sum_{i=2}^{N} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^n \psi_i \right] \tag{3}$$

If $|\lambda_1| > |\lambda_2| \geqslant |\lambda_3| \geqslant \cdots \geqslant |\lambda_N|$, then for $\alpha_1 \neq 0$,

$$\begin{aligned} \psi &\to \psi_1/\|\psi_1\| \\ \|\phi\| &\to \lambda_1 \end{aligned}$$

Hence, the dominant eigenpair is simultaneously determined. For the norm of the vector ϕ whose components are ϕ_i, a frequent choice is

$$\|\phi\| \equiv \|\phi\|_\infty = \max_i |\phi_i|$$

This is the choice adopted here.

Clearly, if $|\lambda_2/\lambda_1| \simeq 1$ convergence of the iteration is slow. Often it can be improved by the replacement $A \to A - \sigma I$ which shifts the value of each eigenvalue by a constant amount σ but does not change the associated eigenvector. Besides poten-

tially accelerating convergence, the shift also enables the determination of the smallest, instead of the largest, eigenpair. In particular, if A and all the λ_i are real, no matter how σ is chosen, either $\lambda_1 - \sigma$ or $\lambda_N - \sigma$ will be the converged eigenvalue. Most often, σ is chosen to be independent of iteration step. In this case, for convergence to λ_1, the optimal choice for σ is $\frac{1}{2}(\lambda_2 + \lambda_N)$; for convergence to λ_N, the choice is $\frac{1}{2}(\lambda_1 + \lambda_{N-1})$ [6].

If the dominant eigenvalue is degenerate, for example, doubly degenerate with $\lambda_1 = \lambda_2$, or degenerate in magnitude, for example, doubly degenerate with $|\lambda_1| = |\lambda_2|$, then the power method, as most iterative methods, cannot determine a unique eigenvector. As can be seen from Eq. (3), in these situations the iteration converges to

$$A^n \psi = \lambda_1^n \left[\alpha_1 \psi_1 + \text{sign}\left(\frac{\lambda_2}{\lambda_1}\right) \alpha_2 \psi_2 + \sum_{i=3}^{N} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^n \psi_i \right]$$

The eigenvalue estimators will converge to the correct values of λ_1 and λ_2 but the eigenvector estimate corresponding to the dominant eigenvalue will be some linear combination of ψ_1 and ψ_2. A similar situation will can occur for convergence to the first subdominant eigenvalue if for example $|\lambda_2| = |\lambda_3|$. In this case ψ_1 can be determined but ψ_2 cannot.

If a few dominant eigenpairs, say M, are desired, one of two approaches are tried. One approach is to use the power method to determine the dominant eigenpair, use deflation to project out this state out of the matrix, and then reuse the power method on the deflated matrix. To determine several eigenpairs simultaneously, the power method can be generalized to

$$\Phi = A\Psi$$

where Φ and Ψ are $M \times N$ matrices whose columns are orthogonalized to each other. This orthogonality needs maintenance throughout the computation or else all M vectors, represented by the columns of the initial Ψ, will converge to the same one. This algorithm is called the simultaneous iteration method [7].

2.2. B. Observation

Booth's refinement of the power method [1,2] uses the observation that for any eigenpair (λ, ψ) and for each non-zero *component* of the eigenvector, the eigenvalue equation $A\psi = \lambda\psi$ can be rewritten as

$$\lambda = \frac{\sum_\beta A_{\alpha\beta} \psi_\beta}{\psi_\alpha} \tag{4}$$

and that similar equations can also be written for any number of groupings of components,

$$\lambda = \frac{\sum_{\alpha \in R_1} \sum_\beta A_{\alpha\beta} \psi_\beta}{\sum_{\alpha \in R_1} \psi_\alpha} = \frac{\sum_{\alpha \in R_2} \sum_\beta A_{\alpha\beta} \psi_\beta}{\sum_{\alpha \in R_2} \psi_\alpha} = \cdots = \frac{\sum_{\alpha \in R_N} \sum_\beta A_{\alpha\beta} \psi_\beta}{\sum_{\alpha \in R_N} \psi_\alpha} \tag{5}$$

where the R_i are rules for different groupings. The groups can overlap. In addition, any two groupings, say 1 and 2, imply

$$\sum_{\alpha \in R_2} \psi_\alpha \sum_{\alpha \in R_1} \sum_\beta A_{\alpha\beta} \psi_\beta = \sum_{\alpha \in R_1} \psi_\alpha \sum_{\alpha \in R2} \sum_\beta A_{\alpha\beta} \psi_\beta \tag{6}$$

The eigenvalue estimator (4) is a special case of what is often called a mixed estimator [8]

$$\lambda = \frac{\langle \phi | A | \psi \rangle}{\langle \phi | \psi \rangle}$$

In the present case, the component ϕ_i of the vector ϕ is unity if $i \in R$; otherwise, it is zero. From N groupings of the components, Booth constructs N estimators for the Nth eigenvalue and forces them to become equal by adjusting certain parameters at each iteration step. Several ways to do this have been devised, and we will now sketch the most recent ones for obtaining two extremal eigenpairs simultaneously.

For almost any starting point $\psi = \sum_i \alpha_i \psi_i$, the power method will converge to (λ_1, ψ_1). The same would be true for almost any other two normalized, but not necessarily orthogonal, starting points $\psi' = \sum_i b_i \psi_i$ or $\psi'' = \sum_i a_i \psi_i$. We will in fact choose two such starting points and at each step apply A to them individually, but at each step we will adjust the relationship between them to prevent the collapse of their *sum* to the dominant eigenfunction.

Formally, we start the iteration with $\psi = \psi' + \eta\psi''$ and assume that after a large number of steps just the two dominant eigenpairs remain significant. Then we have

$$A^n \psi = A^n \sum_{i=1}^{2} \alpha_i \psi_i = \sum_{i=1}^{2} (a_i + \eta b_i) \lambda_i^n \psi_i \tag{7}$$

To determine η, we define two groupings of the components of $A^n\psi$, R_1 and R_2, and let κ_j be the eigenvalue estimate for the jth grouping. Then from Eq. (5) we find that

$$\kappa_1 = \frac{(a_1 + \eta b_1)\lambda_1^n \sum_{\alpha \in R_1}\psi_{1,\alpha} + (a_2 + \eta b_2)\lambda_2^n \sum_{\alpha \in R_1}\psi_{2,\alpha}}{(a_1 + \eta b_1)\lambda_1^{n-1}\sum_{\alpha \in R_1}\psi_{1,\alpha} + (a_2 + \eta b_2)\lambda_2^{n-1}\sum_{\alpha \in R_1}\psi_{2,\alpha}}$$

$$\kappa_2 = \frac{(a_1 + \eta b_1)\lambda_1^n \sum_{\alpha \in R_2}\psi_{1,\alpha} + (a_2 + \eta b_2)\lambda_2^n \sum_{\alpha \in R_2}\psi_{2,\alpha}}{(a_1 + \eta b_1)\lambda_1^{n-1}\sum_{\alpha \in R_2}\psi_{1,\alpha} + (a_2 + \eta b_2)\lambda_2^{n-1}\sum_{\alpha \in R_2}\psi_{2,\alpha}}$$

If we require $\kappa_1 = \kappa_2$, a quadratic equation for η results. If one solution of this equation is chosen to guide the iteration to $a_1 + \eta b_1 = 0$, then $\kappa_1 = \kappa_2 = \lambda_2$. If the solution on the other hand is chosen to guide the iteration towards $a_2 + \eta b_2 = 0$, then $\kappa_1 = \kappa_2 = \lambda_1$.

In practice, we find the coefficients of this quadratic equation in the following manner: Suppose at the nth step, ψ' and ψ'' have iterated to $\hat{\psi}'$ and $\hat{\psi}''$, then at the $(n+1)$th step we require

$$\frac{\sum_{\alpha \in R_1}\sum_\beta A_{\alpha\beta}\hat{\psi}'_\beta + \eta \sum_{\alpha \in R_1}\sum_\beta A_{\alpha\beta}\hat{\psi}''_\beta}{\sum_{\alpha \in R_1}\hat{\psi}'_\alpha + \eta \sum_{\alpha \in R_1}\hat{\psi}''_\alpha} = \frac{\sum_{\alpha \in R_2}\sum_\beta A_{\alpha\beta}\hat{\psi}'_\beta + \eta \sum_{\alpha \in R2}\sum_\beta A_{\alpha\beta}\hat{\psi}''_\beta}{\sum_{\alpha \in R_2}\hat{\psi}'_\alpha + \eta \sum_{\alpha \in R_2}\hat{\psi}''_\alpha} \tag{8}$$

which leads to $q_2\eta^2 + q_1\eta + q_0 = 0$ with

$$q_2 = \sum_{\alpha \in R_2}\hat{\psi}''_\alpha \sum_{\alpha \in R_1}\sum_\beta A_{\alpha\beta}\hat{\psi}''_\beta - \sum_{\alpha \in R_1}\hat{\psi}''_\alpha \sum_{\alpha \in R2}\sum_\beta A_{\alpha\beta}\hat{\psi}''_\beta$$

$$q_1 = \sum_{\alpha \in R_2}\hat{\psi}''_\alpha \sum_{\alpha \in R_1}\sum_\beta A_{\alpha\beta}\hat{\psi}'_\beta - \sum_{\alpha \in R_1}\hat{\psi}''_\alpha \sum_{\alpha \in R_2}\sum_\beta A_{\alpha\beta}\hat{\psi}'_\beta + \sum_{\alpha \in R_2}\hat{\psi}'_\alpha \sum_{\alpha \in R_1}\sum_\beta A_{\alpha\beta}\hat{\psi}''_\beta - \sum_{\alpha \in R_1}\hat{\psi}'_\alpha \sum_{\alpha \in R_2}\sum_\beta A_{\alpha\beta}\hat{\psi}''_\beta \tag{9}$$

$$q_0 = \sum_{\alpha \in R_2}\hat{\psi}'_\alpha \sum_{\alpha \in R_1}\sum_\beta A_{\alpha\beta}\hat{\psi}'_\beta - \sum_{\alpha \in R_1}\hat{\psi}'_\alpha \sum_{\alpha \in R_2}\sum_\beta A_{\alpha\beta}\hat{\psi}'_\beta$$

The strategy is to apply A repeatedly until two real solutions for η exist. One solution will then guide the iteration to (λ_1, ψ_1); the other, to (λ_2, ψ_2). Typically, this procedure would be used only if the simultaneous convergence to two pairs is desired or if the convergence to just the second eigenpair is desired. In some cases, however, accelerated convergence to the first pair occurs.

2.3. C. First refinement

For simplicity, we focus on the determination of the *second largest* eigenpair [2] and note that one additional improvement is necessary for finite precision computers. As both $\hat{\psi}'$ and $\hat{\psi}''$ are converging to the first eigenfunction, only their sum, for proper choices of η, is converging to the second one. Eventually, when η is the root, say η_2, guiding $\hat{\psi}''$ toward the second eigenvector ψ_2, the determination of ψ_2 is limited by the accuracy of the sum of $\hat{\psi}'$ and $\eta_2\hat{\psi}''$. To mitigate this situation, we modify the iteration by making the replacements $\hat{\psi}' \leftarrow \hat{\psi}'$ and $\hat{\psi}'' \leftarrow \hat{\psi}'' + \eta_2\hat{\psi}'$ before moving to the $(n+1)$th step, and then in the $(n+1)$th step we find the new η from the quadratic equation and subtract from it the η_2 from the nth step. Formally, this is equivalent to rewriting the coefficients of the ψ_i in Eq. (7) as

$$a_i + b_i\eta = (a_i + b_i\eta_2) + b_i(\eta - \eta_2) \tag{10}$$

making the replacements

$$a_i \leftarrow (a_i + b_i\eta_2)$$
$$\eta \leftarrow \eta - \eta_2$$

and then in the next iteration solving the quadratic equation for the shifted η.

What does this procedure accomplish? We note that near convergence, when only ψ_1 and ψ_2 remain significant, the current best estimate of ψ_1 is contaminated with ψ_2 and vice versa. Denoting these estimates by $\psi_1 + \epsilon\psi_2$ and $\psi_2 + \delta\psi_1$ and introducing the adjustable parameter η, we can write another estimate of $\psi_2(\eta)$ as

$$\hat{\psi}_2(\eta) = (\psi_2 + \delta\psi_1) + \eta(\psi_1 + \epsilon\psi_2)$$

and so with the application of A to move to the $(n+1)$th step, the new estimate of the second eigenfunction becomes

$$\psi_{2\text{new}} = A\hat{\psi}_2(\eta) = (\lambda_2\psi_2 + \delta\lambda_1\psi_1) + \eta(\lambda_1\psi_1 + \epsilon\lambda_2\psi_2)$$

If at this step η_2 is the choice that guides to ψ_2, then we define η'' by

$$\eta = \eta'' + \eta_2$$

so that

$$\psi_{2\text{new}}(\eta'') = (\lambda_2\psi_2 + \delta\lambda_1\psi_1) + (\eta'' + \eta_2)(\lambda_1\psi_1 + \epsilon\lambda_2\psi_2)$$
$$= [(\lambda_2\psi_2 + \delta\lambda_1\psi_1) + \eta_2(\lambda_1\psi_1 + \epsilon\lambda_2\psi_2)] + \eta''(\lambda_1\psi_1 + \epsilon\lambda_2\psi_2)$$

We observe that $\lambda_1 \psi_1 + \epsilon \lambda_2 \psi_2$ is this step's power iteration estimate for the first eigenfunction so that the second eigenfunction, the term in brackets, is essentially being corrected by an attempt to remove the remaining contamination $\delta \lambda_1 \psi_1$ from the first eigenfunction. If we define the new eigenfunction iterates as

$$\psi_{2new} = \lambda_2 \psi_2 + \delta \lambda_1 \psi_1$$
$$\psi_{1new} = \lambda_1 \psi_1 + \epsilon \lambda_2 \psi_2$$

then

$$\psi_{2new}(\eta) \leftarrow (\psi_{2new} + \eta_2 \psi_{1new}) + \eta \psi_{1new}$$

Thus, the effect of Eq. (10) is promoting the convergence of $\hat{\psi}'$ to the first eigenfunction in the normal power method way whereas $\hat{\psi}''$ (the second eigenfunction estimate) is being corrected at each step by adding (removing) a little of the first eigenfunction estimate. Convergence is reached when $\eta_2 \to 0$; that is, when the second eigenfunction needs no correction from the first eigenfunction.

The above analysis leads to a simple numerical algorithm. The basic steps are

Step 1: Initialize
      1. Set convergence parameter ϵ to a small value.
      2. Choose initial estimates $\psi' \approx \psi_1$ and $\psi'' \approx \psi_2$
Step 2: Reset
      1. Normalize $\psi' \leftarrow \psi' / \|\psi'\|$ and $\psi'' \leftarrow \psi'' / \|\psi''\|$
Step 3: Execute power step
      1. Apply A to ψ' and ψ'' and solve resulting quadratic balance condition (Eq. (8))
      2. If the roots are real, assign the roots η_1 and η_2 to correspond to the largest and smallest (in magnitude) eigenvalue estimates respectively and then update via

$$\psi' \leftarrow \psi'$$
$$\psi'' \leftarrow A\psi'' + \eta_2 A\psi'$$

      else update via

$$\psi' \leftarrow A\psi'$$
$$\hat{\psi}'' \leftarrow A\hat{\psi}''$$

Step 4: Test for convergence
      1. If $|\eta_2| > \epsilon$, go to Step 2
Step 5: Terminate

Eigenvalue estimates can be made by placing ψ' and ψ'' in Eq. (5) for the same or different R_i. These R_i can be the same or different from the two used to compute the q_i. When the roots are complex, an alternative to Step 3.2 is the use of complex arithmetic. If it is used, then the ψ' and ψ'' estimates in Step 3.2 are updated with complex η_i.

2.4. D. Second refinement

Because Eq. (7) will be almost true for large n, it yields a way to estimate the ψ_i. If C_i are normalizing constants and η_2 is the root that gives λ_2, then from Eq. (7) the $(n+1)$th guess at ψ_2 is

$$\psi_2 \approx \psi_2^{(n+1)} = C_2 [A^n \psi]_{\eta=\eta_2}$$

If $\eta_1 \approx -a_2/b_2$ is the root that gives λ_1, then from Eq. (7) the $(n+1)$th guess at ψ_1 is

$$\psi_1 \approx \psi_1^{(n+1)} = C_1 [A^n \psi]_{\eta=\eta_1}$$

These two estimates suggest using

$$\psi^{(n+1)} = \psi_2^{(n+1)} + \eta \psi_1^{(n+1)}$$

as the next iteration guess. Next, we insert

$$\hat{\psi}' = \psi_2^{(n+1)}$$
$$\hat{\psi}'' = \psi_1^{(n+1)}$$

into Eq. (8) and solve it for the two η roots. Now we take for new estimates (with the C_i providing normalization)

$$\psi_2 \approx \psi_2^{(n+2)} = C_2 [A\psi^{(n+1)}]_{\eta=\eta_2} \tag{11}$$

$$\psi_1 \approx \psi_1^{(n+2)} = C_1 [A\psi^{(n+1)}]_{\eta=\eta_1} \tag{12}$$

We note that Eq. (11) is the same adjustment to the second eigenfunction estimate as in the first refinement, which uses the best estimate of ψ_2. Eq. (12) uses the best estimate of ψ_1 instead of the power iterated estimate used in first refinement. Empirically, this second refinement simultaneously produces estimates of ψ_1 converging as λ_3/λ_1 and estimates of ψ_2 converging as λ_3/λ_2. In the Appendix we demonstrate these rates of convergence for non-degenerate states.

Incorporating this refinement requires only replacing Step 3 of the algorithm for the previous with

Step 3: Execute power step
1. Apply A to ψ' and ψ'' and solve resulting quadratic balance condition (Eq. (8))
2. If the roots are real, assign the roots η_1 and η_2 to correspond to the largest and smallest (in magnitude) eigenvalue estimates respectively and then update via

$$\psi' \leftarrow A\psi'' + \eta_1 A\psi'$$
$$\psi'' \leftarrow A\psi'' + \eta_2 A\psi'$$

else update via

$$\psi' \leftarrow A\psi'$$
$$\hat{\psi}'' \leftarrow A\hat{\psi}''$$

Step 4: Test for convergence
1. If $|\eta_2| > \epsilon$, go to Step 2
Step 5: Terminate

2.5. E. Practical algorithm

In an actual implementation of these algorithms, monitoring convergence by $|\eta_2| < \epsilon$ is not the only choice. The more common way would be monitoring successive estimates of the λ_i plus monitoring the residuals $\|A\psi' - \lambda_1\psi'\|$ and $\|A\psi'' - \lambda_2\psi''\|$. We also note the following alternative: As ψ' converges to ψ_1 and ψ'' converges to ψ_2, q_0 and q_2 converge to zero. In short, multiple criteria exist, leading to cross checks. Some recycle already computed quantities and are consequently quite efficient. Here is an algorithm for the second refinement suitable for implementation:

Step 1: Initialize
1. Set convergence parameters ϵ_0 and ϵ_2 to small values.
2. Initialize iteration index $n = 0$,
3. Choose initial estimates for ψ' and ψ'',
4. Choose the rules R_1 and R_2 for grouping of iterated vector components.
Step 2: Reset
1. Normalize $\psi' \leftarrow \psi'/\|\psi'\|$ and $\psi'' \leftarrow \psi''/\|\psi''\|$
Step 3: Execute power step
1. Apply A to ψ' and ψ'',
2. Solve resulting quadratic balance condition (Eq. 8),
3. Estimate eigenvalues using either rule (region) R_1 or R_2,
4. If the roots are real, assign the roots η_1 and η_2 to correspond to the largest and smallest (in magnitude) eigenvalue estimates. For example, we will have

$$\lambda_1 = \frac{\sum_{\alpha \in R_1} \sum_{\beta} A_{\alpha\beta} \psi'_{\beta} + \eta_1 \sum_{\alpha \in R_1} \sum_{\beta} A_{\alpha\beta} \psi''_{\beta}}{\sum_{\alpha \in R_1} \psi''_{\alpha} + \eta_1 \sum_{\alpha \in R_1} \psi''_{\alpha}}$$

$$\lambda_2 = \frac{\sum_{\alpha \in R_1} \sum_{\beta} A_{\alpha\beta} \psi'_{\beta} + \eta_2 \sum_{\alpha \in R_1} \sum_{\beta} A_{\alpha\beta} \psi''_{\beta}}{\sum_{\alpha \in R_1} \psi''_{\alpha} + \eta_2 \sum_{\alpha \in R_1} \psi''_{\alpha}}$$

then update via

$$\psi' \leftarrow A\psi'' + \eta_1 A\psi'$$
$$\psi'' \leftarrow A\psi'' + \eta_2 A\psi'$$

else update via

$$\psi' \leftarrow A\psi'$$
$$\hat{\psi}'' \leftarrow A\hat{\psi}''$$

   5. Go to Step 2
Step 4: Test for convergence
   1. If either $|q_0| > \epsilon_0$ or $|q_2| > \epsilon_2$, increment the iteration index, $n \leftarrow n + 1$ and go to Step 2.
Step 5: Terminate.

When the roots are complex, an alternative to Step 3.2 is the use of complex arithmetic. If it is used, then the eigenvector estimates in Step 3.2 are updated with complex η_i. The choice of rules is quite flexible. A rule may use one vector component selected randomly, a small number of components selected randomly, all odd or even components, the first or second half of the vector, etc. We note that q_2 goes to zero faster than q_0, and when q_2 becomes very small, then the quadratic equation numerically reduces to $q_1 \eta + q_0 = 0$ which is solved to get η_2. Essentially $q_2 = 0$ means that the dominant eigenpair is known to machine accuracy, so it cannot be improved on further iteration.

3. Applications

3.1. A. Cyclic matrix

To illustrate the effectiveness of the *first* refinement, we applied it to the symmetric $N \times N$ matrix

$$A = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 & -1 \\ -1 & 2 & -1 & \ddots & & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & & \ddots & -1 & 2 & -1 \\ -1 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix}$$

whose eigenvalues for any N are

$$\gamma_n = 2 - 2\cos k_n = 4\sin^2 \frac{k_n}{2}$$

where $k_n = \frac{2\pi n}{N}$ with $n = 0, 1, 2, \ldots, N - 1$. Physically, the matrix represents the N point discretization of the second derivative defined on a circle. We note that for N odd, all but the minimal eigenstate ($n = 0$) are doubly degenerate, while for N even, all but the minimal ($n = 0$) and maximal ($n = N/2$) ones are doubly degenerate. Accordingly for even N,

$$0 = \gamma_0 < \gamma_1 = \gamma_{N-1} < \cdots < \gamma_{N/2-1} = \gamma_{N/2+1} < \gamma_{N/2} = 4$$

Table 1 reports the results of a deterministic computation of the second smallest eigenpair for a sequence of even N. To generate it, all the eigenvalues of the matrix were shifted by subtracting four times the identity matrix and then getting the two largest magnitude eigenvalues of $A - 4I$. For the shifted matrix $\lambda_1 = \gamma_0 = -4$ and $\lambda_2 = \gamma_1 = -4\cos^2(\pi/N)$. λ_1 is thus seen as being independent of N and is not reported. The λ_2's in Table 1 are 4 plus the power method's computation of second largest magnitude eigenvalue of $A - 4I$. The iteration was stopped when the absolute value of the maximum difference between any component of the eigenvector in successive iterations was less than 10^{-10}. We see remarkable agreement between the values determined by the power method and the exact analytic value is obtained even for largest possible N on our desktop computer. We converged accurately to the second smallest eigenvalue even though it is approaching the smallest one as N is increased and is itself degenerate. By it being degenerate, our eigenvector estimate is a linear combination of two eigenstates that depends on the starting conditions for the iteration. It is not something we can benchmark but it does approximate well $A\psi = \lambda_2\psi$. For R_1 and R_2, we used the first and second half of the vector components. For N up to about 1000, starting vectors whose components were set randomly worked well. For starting vectors at $N > 1000$ we used the eigenvectors found at $N/2$ injected into the higher dimension via $a(2j) \leftarrow 0.75b(j) + 0.25b(j+1)$ and $a(2j+1) \leftarrow 0.25b(j) + 0.75b(j+1)$. The coefficients were chosen to adjust for the fact that $a(2j)$ is 1 unit from $b(j)$ and 3 units from $b(j+1)$ and $a(2j+1)$ is 1 unit from $b(j+1)$ and 3 units from $b(j)$.

3.2. B. Two-dimensional Ising model

The two-dimensional Ising model is one of the few two-dimensional models of a system of many interacting degrees of freedom that has an exact solution for its thermodynamic properties. This solution, first constructed by Onsager [9], shows that in the thermodynamic limit the model has a phase transition between a magnetically ordered (ferromagnetic) state at low temperatures and a magnetically disordered state (paramagnetic) at high temperatures. Onsager succeeded in calculating many of the properties of the model exactly, including the temperature T_c at which the transition occurs. Key to his calculations was expressing the partition function of the model in terms of its transfer matrix [10], finding the dominant eigenvalue of this matrix, and showing in the thermodynamic limit (letting the area of the model approach infinity) that this eigenvalue implies the onset of long-range ordering among the spin variables of the model.

Table 1
For the cyclic matrix, the exact and first refinement calculations for the sub-dominant eigenvalue λ_2, plus their difference

N	Exact	PM1	Difference
100	0.0039465433630297	0.0039465431649277	1.98E−10
200	0.0009868793234571	0.0009868792721619	5.13E−11
400	0.0002467350504105	0.0002467351362063	−8.58E−11
800	0.0000616847138537	0.0000616847980273	−8.42E−11
1600	0.0000154212379171	0.0000154212887717	−5.09E−11
3200	0.0000038553131951	0.0000038553389818	−2.58E−11
6400	0.0000009638285310	0.0000009638405936	−1.21E−11
12,800	0.0000002409571473	0.0000002409625672	−5.42E−12
25,600	0.0000000602392878	0.0000000602416170	−2.33E−12
51,200	0.0000000150598221	0.0000000150608281	−1.01E−12
102,400	0.0000000037649555	0.0000000037654755	−5.20E−13
204,800	0.0000000009412389	0.0000000009414407	−2.02E−13
409,600	0.0000000002353098	0.0000000002353788	−6.91E−14
819,200	0.0000000000588274	0.0000000000589155	−8.82E−14
1,638,400	0.0000000000147069	0.0000000000148614	−1.55E−13
3,276,800	0.0000000000036766	0.0000000000036855	−8.88E−15

We will consider the model for finite area, that is, an $m \times n$ model defined with periodic boundary conditions in one direction and open boundary conditions in the other. Because of the one open boundary, the transfer matrix will thus be non-symmetric. In the absence of an applied magnetic field, the model's energy is

$$E\{\mu\} = -J\sum_{i=1}^{m-1}\sum_{j=1}^{n}\mu_{i,j}\mu_{i,j-1} - J\sum_{i=1}^{m}\sum_{j=1}^{n}\mu_{i,j}\mu_{i,j+1}$$

Here, (i,j) are the coordinates of a lattice site. The Ising spin variable $\mu_{i,j}$ on each site has the value of $\pm 1, J > 0$, and $\mu_{i,n+1} = \mu_{i,1}$. A column configuration of Ising spins will be denoted by

$$\sigma_j = (\mu_{1,j}, \mu_{2,j}, \ldots, \mu_{m,j})$$

and there are 2^m possible configurations for each column.

The definition of the transfer matrix follows from the expression for the partition function [4]

$$
\begin{aligned}
Z(m,n) &= \sum_{\{\mu\}} \exp[-\beta E(\{\mu\})] \\
&= \sum_{\sigma_1,\ldots,\sigma_n} \exp\left[-\beta\left(\sum_{j=1}^{n}\{V_1(\sigma_j) + V_2(\sigma_j, \sigma_{j+1})\}\right)\right] \\
&= \sum_{\sigma_1,\ldots,\sigma_n} L(\sigma_1,\sigma_2)L(\sigma_2,\sigma_3)\cdots L(\sigma_{n-1},\sigma_n)L(\sigma_n,\sigma_1) \\
&= \sum_{\sigma_1} L^n(\sigma_1,\sigma_1)
\end{aligned}
$$

where

$$V_1(\sigma_j) = -v\sum_{i=1}^{m-1}\mu_{i,j}\mu_{i+1,j}$$

is the interaction energy of the jth column and

$$V_2(\sigma_j, \sigma_{j+1}) = -v\sum_{i=1}^{m}\mu_{i,j}\mu_{i,j+1}$$

is the interaction energy between the jth and $(j+1)$th columns, $v = J/k_B T, k_B$ is Boltzmann's constant, and $L(\sigma,\sigma')$ is the transfer matrix of order $N = 2^m \times 2^n$ whose elements are

$$L(\sigma,\sigma') = \exp\left(v\sum_{k=1}^{m-1}\mu_k\mu_{k+1}\right)\exp\left(v\sum_{k=1}^{m}\mu_k\mu_k'\right)$$

More succinctly,

$$Z(m,n) = \mathrm{Tr}(L^n) = \sum_{j=1}^{2^m}\lambda_j^n$$

Onsager found analytic expressions for all the eigenvalues of the transfer matrix. Since we needed to form this matrix, we found it as convenient to compute them numerically. Spot checks produce excellent agreement between the the two approaches. In the thermodynamic limit, when $T \to T_c, \lambda_2 \to \lambda_1$. Here, although we chose $T = T_c$ and the order of or matrix became quite large, we were still reasonably far away for this critical point. Table 2 presents a comparison of the two largest eigenvalues of the transfer matrix as determined by our second refinement of the power method and those determined by the EISPACK eigensolver RG [11]. For cases except $m = 11$, we simply used as the solution the results after 100 iterations. For $m = 11$, we used 1000 iterations. The larger number of iterations was necessary to obtain the same level of accuracy.

3.3. C. One-dimensional Hubbard model

The Hubbard Hamiltonian was originally proposed as a model for metallic ferromagnetism [12]. Most recently, its two-dimensional version has been the subject of intense scrutiny as a possible model for electronic superconductivity. In one-dimension, a variant of it, called the Pariser–Parr–Popple Hamiltonian is frequently used to model conjugated cyclic molecules. Other variants model one-dimensional organic conductors. Because of the enormous amount of computer memory required by deterministic methods, precise specification of the ground state (zero temperature) properties of these models has often been hampered by techniques limited to relatively small system sizes. The memory requirements scale as 4^N, where N is the number of lattice sites.

The Hamiltonian operator for the Hubbard model is

$$\widehat{H} = -t \sum_{\langle i,j \rangle, \sigma} (\hat{c}_{i,\sigma}^\dagger \hat{c}_{j,\sigma} + \hat{c}_{j,\sigma}^\dagger \hat{c}_{i,\sigma}) + U \sum_i \hat{n}_{i,\uparrow} \hat{n}_{i,\downarrow}$$

where the summation is over nearest-neighbor pairs of lattice sites i and j and electron spin σ; t and U are the hopping amplitude and repulsive Coulomb parameters; $\hat{c}_{i,\sigma}^\dagger, \hat{c}_{i,\sigma}$ and $\hat{n}_{i,\sigma} = \hat{c}_{i,\sigma}^\dagger \hat{c}_{i,\sigma}$ are the creation, destruction, and number operators for an electron at site i with spin σ. Usually a Fock basis is used to represent the Hamiltonian operator as a matrix

$$h_{ij} = \langle i|\widehat{H}|j \rangle$$

where

$$|i\rangle = |n_{1,\uparrow}, n_{2,\uparrow}, \ldots, n_{N,\uparrow}\rangle |n_{1,\downarrow}, n_{2,\downarrow}, \ldots, n_{N,\downarrow}\rangle$$

with $n_i^\sigma = 0, 1$ being the eigenvalues of the number operator. As a representative of an Hermitian operator whose matrix elements are real, the resulting matrix is symmetric. Various symmetries are usually used to block diagonalize the matrix and then obtain the ground state for each block. We will consider only blocks that have a specific value of the z-component of the total electron spin. The size of the Hilbert space and hence the order of the matrix is

Table 2
Comparison of the two largest eigenvalues of the transfer matrix for the 2D Ising model as a function of the lattice edge size m computed by the EISPACK routine RG and the second refinement. $N = 2^m$ is the order of the matrix. Also shown is the fractional difference $(\lambda_{RG} - \lambda_{PM2})/\lambda_{RG}$ between the different estimates. We choose $m = n$ and $T = T_c$

m	N	RG	PM2	FD
1	4	3.41421355573626E+00	3.41421355573626E+00	0.00E+00
		1.41421355573626E+00	1.41421355573626E+00	0.00E+00
2	16	7.46410158611908E+00	7.46410158611907E+00	3.57E−16
		4.82842709270073E+00	4.82842709270073E+00	−1.29E−15
3	64	1.78770541980345E+01	1.78770541980345E+01	−7.95E−16
		1.35518083939891E+01	1.35518083939891E+01	3.93E−16
4	256	4.41298558292434E+01	4.41298558292434E+01	4.83E−16
		3.60398703210879E+01	3.60398703210878E+01	5.91E−16
5	1024	1.10192319565854E+02	1.10192319565854E+02	9.03E−16
		9.38962258961220E+01	9.38962258961221E+01	−4.54E−16
6	4096	2.76599914093667E+02	2.76599914093667E+02	−8.22E−16
		2.42266413140723E+02	2.42266413140723E+02	0.00E+00
7	16,384	6.96269201662783E+02	6.96269201662782E+02	1.47E−15
		6.21748520715910E+02	6.21748520715909E+02	1.65E−15
8	65,536	1.75565374661531E+03	1.75565374661531E+03	1.30E−16
		1.59043428137424E+03	1.59043428136461E+03	6.05E−12
9	262,144	4.43180239838645E+03	4.43180239838646E+03	−1.44E−15
		4.05958858259757E+03	4.05958858259756E+03	1.79E−15
10	1,048,576	1.11957434253463E+04	1.11957434253463E+04	1.46E−15
		1.03466429299731E+04	1.03466429299731E+04	8.79E−16
11	4,194,304	2.82985308867953E+04	2.82985308867954E+04	−3.60E−15
		2.63419326613631E+04	2.63419326613632E+04	−3.87E−15

$$N = \frac{N!}{N_\uparrow!(N - N_\downarrow)!}$$

where $N_\uparrow$ is the number of up spin electrons and being and $N_\downarrow$ is the number of down spin electrons.

For a given lattice site i the maximum number of non-zero values of h_{ij} is $2zN$ where z is the number of nearest neighbors of the chosen lattice. Typically, $z \ll N$; hence, the matrix is very sparse. Here, we will consider the model in one-dimension where $z = 2$. In one-dimension the model has an exact solution. Obtaining the ground or first excited state for these solution is not as straightforward as for the two previous test cases. We chose to obtain them numerically and compare the effectiveness of our second refinement of the power method to that of several standard eigenpair methods. Our emphasis is on how well degenerate states are captured.

We will compute the two largest and two smallest eigenvalues of the sparse, potentially hugely dimensioned, matrix H representing the operator $\hat{H}$ in one dimension with periodic boundary conditions. In this case and if $N_\uparrow = N_\downarrow$ and $N_\uparrow$ is odd, the model satisfies the following version of the Perron–Frobenius Theorem [13]: If a matrix is irreducible and all off-diagonal elements are non-positive, the state corresponding to smallest eigenvalue is real and non-degenerate. We will study the model on a 10 site lattice with $U = 4$ and $t = 1$. For this lattice size and filling half or less, the theorem applies to cases $(N_\uparrow, N_\downarrow) = (1,1), (3,3)$, and $(5,5)$. They are called closed shell cases and the result of our calculations for them are shown in Table 3. Because of various symmetries, features of the smallest states are reflected in those of the largest ones.

In Table 3, three methods where used to get the eigenvalues. One method used was the LAPACK routine DSYEV [14]. This double precision routine returns all the eigenvalues and eigenvectors of a symmetric matrix. At large orders computer memory became insufficient for its use. Accordingly, we supplemented our results with those obtained by using the DNLASO double precision subroutine [15] which is a block Lanczos method with selective reorthogonalization [7]. The components of the starting vectors for the Lanczos iteration are selected randomly and uniformly on the interval $(-0.5, 0.5)$. The quality of the results is controlled by specifying the block size (the number of starting states), the number of significant figures for the convergence of the eigenvalues, and the maximum number of iterations. We found a block size of 1 gave estimates for the second and third eigenvalues that became progressively poorer as N increased. This is reasonable. A size of 2 produced cases where the sub-dominant eigenvalue was consistently returned as the dominant. For a size of 6, convergence was very slow if at all. For a size of 8, memory soon became insufficient. A size of 4 was used for the data in the table. We found lack of convergence for many cases if the precision was requested to be larger than 8 decimal places. Typically, a few hundred iterations were needed, but the computation times were a few tens of seconds. Table 3 shows excellent agreement between all three methods. We note that the excited state was at least doubly degenerate.

The next set of results are for electron fillings where the eigenstates are subjected to Kramers degeneracies. Kramers's Theorem [16] says all energy levels of a system containing an odd number of electrons must at least be at least doubly degenerate provided there are no magnetic fields present to remove time-reversal symmetry. In Table 4, we present several Kramers cases where $(N_\uparrow, N_\downarrow) = (3,2), (4,3)$, and $(5,4)$. For the standard software packages, we listed the three largest and three smallest eigenvalues they estimated to see the accuracy to which they determined the degenerate ground state. We see very good agreement between all three methods in estimating the eigenvalue of degenerate largest and smallest state. All three however lack the precision necessary to differentiate between a true degeneracy and a very near one. The power method in particular is less than adequate for this purpose.

In Table 4, the three largest and smallest eigenvalues are presented to provide extra information about the degeneracies. Sixteen significant figures were printed to indicate how well degeneracies are captured. Basically we do not know the precision of the eigenvalues other then it is no more than difference between eigenvalues that should exactly be degenerate. All the expected features of the lowest eigenstates with regard to degeneracies are exhibited. Because the model has particle-hole symmetry similar features also exists for its largest eigenstates.

Table 3

For closed shell cases, comparison of the eigenvalues of a 10 site 1D Hubbard model computed by the eigenpair routine DSYEV, the block Lanczos routine DNLASO, and the second refinement. For the first two methods, the three largest and three smallest eigenvalues were computed to measure their consistency, effectiveness, and accuracy

$N_\uparrow$	$N_\downarrow$	N	DSYEV	DNLASO	PM2
1	1	100	0.5657693716217906E+01	0.5657693716217901E+01	0.5657693716217914E+01
			0.5519554669107880E+01	0.5519554669107876E+01	0.5519554669107137E+01
			−0.3862202348191250E+01	−0.3862202348191248E+01	−0.3862202348191251E+01
			−0.3618033988749895E+01	−0.3267468797160054E+01	−0.3618033988603501E+01
3	3	14,400	0.1656339684606611E+02	0.1656339684376816E+02	0.1656339684606624E+02
			0.1617312172182284E+02	0.1617312172136987E+02	0.1617312172191405E+02
			−0.8262531385370846E+01	−0.8262531383972004E+01	−0.8262531385370927E+01
			−0.7599976793651736E+01	−0.7599976793264113E+01	−0.7599976793831864E+01
5	5	63,504		0.2583432263352126E+02	0.2583432263577081E+02
				0.2543485463377173E+02	0.2543485464252857E+02
				−0.5834322635176973E+01	−0.5834322635773042E+01
				−0.5434854632148166E+01	−0.5434830052960784E+01

Table 4
For Kramers degeneracies cases, comparison of the eigenvalues of a 10 site 1D Hubbard model computed eigenpair routine DSYEV, the block Lanczos program DNLASO, and the second refinement of the power method. For the first two methods, the three largest and three smallest eigenvalues were computed to measure their consistency, effectiveness, and accuracy in computing degenerate eigenvalues

$N_\uparrow$	$N_\downarrow$	N	DSYEV	DNLASO	PM2
3	2	5400	0.1306499556833340E+02	0.1306499556833341E+02	0.1306499556833335E+02
			0.1306499556833336E+02	0.1306499554321960E+02	0.1306499556833335E+02
			0.1282579739183819E+02	0.1282579739163641E+02	
			−0.7511951740365890E+01	−0.7511951740281242E+01	−0.7511951740365513E+01
			−0.7511951740365851E+01	−0.7511951731564561E+01	−0.7511951740365509E+01
			−0.7249884543021683E+01	−0.7249884543021723E+01	
4	3	25,200	0.1816344283994604E+02	0.1816344283994595E+02	0.1816344283994610E+02
			0.1816344283994604E+02	0.1816344283994549E+02	0.1816344283994610E+02
			0.1771746494384758E+02	0.1771746494296794E+02	
			−0.8030089029893539E+01	−0.8030089029893475E+01	−0.8030089030622399E+01
			−0.8030089029893492E+01	−0.8030089029485101E+01	−0.8030089030532327E+01
			−0.7521441552342070E+01	−0.7521441551092671E+01	
5	4	52,920		0.2285321122055267E+02	0.2285321221296537E+02
				0.2285321121726444E+02	0.2285321166226352E+02
				0.2221382316719896E+02	
				−0.6853211221744196E+01	−0.6853211215825024E+01
				−0.6853211221310334E+01	−0.6853211214979680E+01
				−0.6213823170572150E+01	

Table 5
For open shell cases, comparison of the eigenvalues of a 10 site 1D Hubbard model computed by the LAPACK routine DSYEV, the Netlib program DNLASO, and the second refinement of the power method. For the first two methods, the three largest and three smallest eigenvalues were computed to measure their consistency, effectiveness, and accuracy in computing degenerate eigenvalues

$N_\uparrow$	$N_\downarrow$	N	DSYEV	DNLASO	PM2
2	2	2025	0.1121466372028744E+02	0.1121466372028743E+02	0.1121466372028747E+02
			0.1096186919469933E+02	0.1096186919469927E+02	0.1096186919053599E+02
			0.1096186919469928E+02	0.1096186919469929E+02	
			−0.6601239688910290E+01	−0.6431629846631373E+01	−0.6601239688910274E+01
			−0.6431629846631359E+01	−0.6424903541072491E+01	−0.6431629865616197E+01
			−0.6431629846631350E+01	−0.5854101965765123E+01	
4	4	44,100		0.2143485463460406E+02	0.2143485463565059E+02
				0.2106806509093548E+02	0.2106806509410040E+02
				0.2106806508923771E+02	
				−0.7647179205940428E+01	−0.7647179208191599E+01
				−0.7538791441444121E+01	−0.7538797509518616E+01
				−0.7538791440796984E+01	

For other electron fillings *a priori* exact information about degeneracies is lacking. What is known is that when $U = 0$ the ground state for most fillings is degenerate. These fillings typically are called open shell cases. When $U \neq 0$, the degeneracies are typically lifted, the degree to which it is however depends on the closeness of the nearest unoccupied eigenstate. For small systems, the lifting might be minor, and for limited precision calculations this can make distinguishing degenerate and nearly degenerate states difficult. For the results in Table 5, this discussion applies to the cases $(N_\uparrow, N_\downarrow) = (2, 2)$ and (4,4) illustrated there. The qualitative character of the results are similar to those of the Kramers's cases.

The results of the three tables indicate that this test case is nontrivial. We judge our preliminary results as indicating that the block Lanczos and the second refinement of the power method have comparable effectiveness. The version of the power method used here lacks the ability to return more than two eigenvalues in contrast to the block Lanczos used that should estimate well at least four.

We comment that the Lanczos method is not a black box. To ensure that the result is the minimum as opposed to some excited state, the calculation usually needs to be run multiple times with different random number seeds or some other means to change the starting vectors, and then the results need to be studied to identifiy those to be regarded as estimates of the ground state [17,18]. An error is usually estimated from the variance of the average of the ground state estimates. We restarted the Lanczos calculations for a few of the cases multiple times. As our interest was qualitative, we did not perform an error analysis but instead presented representative results.

The power method is also not a black box. It used the same sparse matrix as the block Lanczos method. From it we only show the two largest and two smallest as our double precision code did not incorporate a procedure to allow the determination of the third eigenpair. The components of one starting state were selected uniformly and randomly over (0,1); the other from (−0.5,0.5). We defined our regions in the following manner. First we performed a random permutation

of the vector components. Region 1 was the first $N/2 + 1$ of these permuted components; region 2, the last $N/2 + 1$. Initially, we used $\epsilon_0 = 10^{-13}$ and $\epsilon_2 = 10^{-10}$. Multiple starts with changing random number seeds were used to ensure consistency. We first computed the largest two eigenvalues and used the largest value, truncated to two significant figures, as the shift to compute the two smallest. Cases where the lowest or the largest had a near degeneracy converged only so far. Often with $|q_0| < \epsilon_0$ being satisfied while the ϵ_2 being too small for q_2 to converge. Instead of adjusting these stopping criteria, we found that simply stopping the iteration after some fixed number of iterations and the choosing the result by locating when the residual $\|A\psi'' - \lambda_2\psi''\|$ ceased decreasing was very effective.

4. Concluding remarks

We presented two refinements of the power method that enable the simultaneous determination of two extremal eigenpairs of a matrix. We illustrated their effectiveness by benchmarking them on three quite distinct but physically challenging problems. For the cyclic matrix, we exactly knew the eigenvalues and their degeneracies. We showed we could determine the two smallest extremal eigenvalues to nearly machine precision. For the transfer matrix of the two-dimensional Ising model, we knew the exact values of the dominant eigenvalues. The two-dimensional Hubbard model was more challenging: As a function of the electron filling various degeneracies exist. Here, we choose to compare the effectiveness of the second refinement with two standard numerical determinations of the ground and first excited state, illustrating the limitations of all three methods especially when the dominant state is degenerate or very nearly so. In general, the second refinement appears as effective as a readily available implementation of the block Lanczos method with selective reorthogonalization.

All our test cases involved real matrices, but preliminary testing indicates that the techniques presented also work well for complex matrices and non-symmetric matrices with complex eigenvalues. The techniques presented are easily adapted to the determination of just the dominant or just the sub-dominant eigenpair. Convergence to the dominant one is accelerated as it is controlled by λ_3/λ_1 is instead of λ_2/λ_1. Generalizations to more than two extremal pairs are possible. Preliminary testing for up to four have been promising.

One advantage of our refinements is that they maintain the simplicity of the basic power method. Another is their adaptability to Monte Carlo implementations. In a number of fields of physics and chemistry, the power method is the core of the Monte Carlo methods for determining the ground state of models whose complexity grow exponentially with physical size. Here, the ground state energy is estimated from samplings of the ground state wavefunction. Such samplings may involve only a small fraction of all possible components of the state, and a mixed estimator [8] for the energy is often used. Here we pushed the use of such estimators a step further with novel consequences. In two other papers, we will describe Monte Carlo implementations of our refinements and their applications to the transfer matrix of the two-dimensional Ising model and to Hamiltonian matrix of the two-dimensional Hubbard model [19].

Appendix

We now show that if the eigenstates are non-degenerate, our second refinement of the power method simultaneously produces estimates of ψ_1 converging as λ_3/λ_1 and estimates of ψ_2 converging as λ_3/λ_2. Suppose that the estimates of ψ_1 and ψ_2 are very good with only a small mixtures of other components; that is, there is some small v such that

$$\psi_1 \approx (1 + dv)\psi_1 + ev\psi_2 + fv\psi_3 \tag{A.1}$$
$$\psi_2 \approx av\psi_1 + (1 + bv)\psi_2 + cv\psi_3 \tag{A.2}$$

Hence

$$\psi = (av\psi_1 + (1 + bv)\psi_2 + cv\psi_3) + x((1 + dv)\psi_1 + ev\psi_2 + fv\psi_3) \tag{A.3}$$

Define the total ith eigenfunction component in region j as

$$N_{ij} = \sum_{\alpha \in R_j} \psi_{i,\alpha} \tag{A.4}$$

We now apply the balance condition for equal λ's. For region R_1 the eigenvalue estimate is

$$\frac{N_{11}\lambda_1(av + x(1 + dv)) + N_{21}\lambda_2((1 + bv) + xev) + N_{31}\lambda_3(cv + xfv)}{N_{11}(av + x(1 + dv)) + N_{21}((1 + bv) + xev) + N_{31}(cv + xfv)}$$

while for region R_2 it is

$$\frac{N_{12}\lambda_1(av + x(1 + dv)) + N_{22}\lambda_2((1 + bv) + xev) + N_{32}\lambda_3(cv + xfv)}{N_{12}(av + x(1 + dv)) + N_{22}((1 + bv) + xev) + N_{32}(cv + xfv)}$$

Set the above two eigenvalue estimates equal and cross multiply to clear the denominators to obtain a quadratic equation in x. We now collect powers of x. Terms involving x^0 are

$$L_1 = -a\lambda_1 N_{12}N_{21}v + a\lambda_2 N_{12}N_{21}v + a\lambda_1 N_{11}N_{22}v - a\lambda_2 N_{11}N_{22}v \tag{A.5}$$
$$L_2 = -c\lambda_2 N_{22}N_{31}v + c\lambda_3 N_{22}N_{31}v + c\lambda_2 N_{21}N_{32}v - c\lambda_3 N_{21}N_{32}v$$
$$L_3 = v^2(-ab\lambda_1 N_{12}N_{21} + ab\lambda_2 N_{12}N_{21} + ab\lambda_1 N_{11}N_{22})$$
$$L_4 = -c\lambda_2 N_{22}N_{31}v + c\lambda_3 N_{22}N_{31}v + c\lambda_2 N_{21}N_{32}v - c\lambda_3 N_{21}N_{32}v$$
$$L_5 = v^2(-bc\lambda_2 N_{22}N_{31} + bc\lambda_3 N_{22}N_{31} + ac\lambda_1 N_{11}N_{32})$$
$$L_6 = v^2(-ac\lambda_3 N_{11}N_{32} + bc\lambda_2 N_{21}N_{32} - bc\lambda_3 N_{21}N_{32})$$

Terms involving x^1 are

$$L_7 = -f\lambda_2 N_{22}N_{31}v + f\lambda_3 N_{22}N_{31}v + f\lambda_2 N_{21}N_{32}v - f\lambda_3 N_{21}N_{32}v \tag{A.6}$$
$$L_8 = v^2(-ae\lambda_1 N_{12}N_{21} + ae\lambda_2 N_{12}N_{21})$$
$$L_9 = v^2(ae\lambda_1 N_{11}N_{22} - ae\lambda_2 N_{11}N_{22} - af\lambda_1 N_{12}N_{31})$$
$$L_{10} = v^2(af\lambda_3 N_{12}N_{31} - ce\lambda_2 N_{22}N_{31} - bf\lambda_2 N_{22}N_{31})$$
$$L_{11} = v^2(ce\lambda_3 N_{22}N_{31} + bf\lambda_3 N_{22}N_{31} + af\lambda_1 N_{11}N_{32})$$
$$L_{12} = v^2(-af\lambda_3 N_{11}N_{32} + ce\lambda_2 N_{21}N_{32} + bf\lambda_2 N_{21}N_{32})$$
$$L_{13} = v^2(-ce\lambda_3 N_{21}N_{32} - bf\lambda_3 N_{21}N_{32})$$
$$L_{14} = (1 + dv)(-\lambda_1 N_{12}N_{21} + \lambda_2 N_{12}N_{21} + \lambda_1 N_{11}N_{22} - \lambda_2 N_{11}N_{22})$$
$$L_{15} = v(1 + dv)(-b\lambda_1 N_{12}N_{21} + b\lambda_2 N_{12}N_{21} + b\lambda_1 N_{11}N_{22} - b\lambda_2 N_{11}N_{22})$$
$$L_{16} = v(1 + dv)(-c\lambda_1 N_{12}N_{31} + c\lambda_3 N_{12}N_{31} + c\lambda_1 N_{11}N_{32} - c\lambda_3 N_{11}N_{32})$$

Terms involving x^2 are

$$L_{17} = v^2 ef(-\lambda_2 N_{22}N_{31} + \lambda_3 N_{22}N_{31} + \lambda_2 N_{21}N_{32} - \lambda_3 N_{21}N_{32}) \tag{A.7}$$
$$L_{18} = ev(1 + dv)(-\lambda_1 N_{12}N_{21} + \lambda_2 N_{12}N_{21} + \lambda_1 N_{11}N_{22} - \lambda_2 N_{11}N_{22})$$
$$L_{19} = fv(1 + dv)(-\lambda_1 N_{12}N_{31} + \lambda_3 N_{12}N_{31} + \lambda_1 N_{11}N_{32} - \lambda_3 N_{11}N_{32})$$

Finally we can write

$$q_0 = L_1 + L_2 + L_3 + L_4 + L_5 + L_6 \tag{A.8}$$
$$q_1 = L_7 + L_8 + L_9 + L_{10} + L_{11} + L_{12} + L_{13} + L_{14} + L_{15} + L_{16} \tag{A.9}$$
$$q_2 = L_{17} + L_{18} + L_{19} \tag{A.10}$$

and note that we seek the smallest magnitude root of

$$q_2 x^2 + q_1 x + q_0 = 0 \tag{A.11}$$

Note that x will be very small when $v \approx 0$, so that the quadratic part can be ignored, leading to

$$x = -q_0/q_1 \tag{A.12}$$

Also the terms involving v^2 can be ignored compared to terms involving v. Thus

$$\begin{aligned} q_0 &= L_1 + L_2 \\ &= v(-a\lambda_1 N_{12}N_{21} + a\lambda_2 N_{12}N_{21} + a\lambda_1 N_{11}N_{22} - a\lambda_2 N_{11}N_{22} \\ &\quad - c\lambda_2 N_{22}N_{31} + c\lambda_3 N_{22}N_{31} + c\lambda_2 N_{21}N_{32} - c\lambda_3 N_{21}N_{32}) \end{aligned}$$

Now consider q_1. Every term except L_{14} has at least v^1 in it, which will be small compared to the v^0 term in L_{14}. Thus using the v^0 term in L_{14},

$$q_1 = (-\lambda_1 N_{12}N_{21} + \lambda_2 N_{12}N_{21} + \lambda_1 N_{11}N_{22} - \lambda_2 N_{11}N_{22})$$

so from Eq. (A.12) we have that

$$x = -av - cv\frac{(-\lambda_2 N_{22}N_{31} + \lambda_3 N_{22}N_{31} + \lambda_2 N_{21}N_{32} - \lambda_3 N_{21}N_{32})}{(-\lambda_1 N_{12}N_{21} + \lambda_2 N_{12}N_{21} + \lambda_1 N_{11}N_{22} - \lambda_2 N_{11}N_{22})}$$

In what follows, ψ is replaced by ψ_2 because the root corresponding to ψ_2 has been selected.

$$A\psi_2 = \lambda_1(av + x(1 + dv))\psi_1 + \lambda_2((1 + bv) + xev)\psi_2 + \lambda_3(cv + xfv)\psi_3$$

Both x and v are small, so the product xv is ignored yielding

$$A\psi_2 = \lambda_1(av + x)\psi_1 + \lambda_2((1 + bv) + xev)\psi_2 + \lambda_3(cv + xfv)\psi_3$$

Substituting for x, we rewrite this equation as

$$A\psi_2 = \lambda_1 \frac{-cv(-\lambda_2 N_{22}N_{31} + \lambda_3 N_{22}N_{31} + \lambda_2 N_{21}N_{32} - \lambda_3 N_{21}N_{32})}{(-\lambda_1 N_{12}N_{21} + \lambda_2 N_{12}N_{21} + \lambda_1 N_{11}N_{22} - \lambda_2 N_{11}N_{22})}\psi_1 + \lambda_2(1+bv)\psi_2 + \lambda_3(cv)\psi_3$$

Dividing by $\lambda_2(1+bv)$ and keeping terms to order v yields

$$\frac{1}{\lambda_2}A\psi_2 = \frac{-cv(-N_{22}N_{31} + (\lambda_3/\lambda_2)N_{22}N_{31} + N_{21}N_{32} - (\lambda_3/\lambda_2)N_{21}N_{32})}{(-N_{12}N_{21} + (\lambda_2/\lambda_1)N_{12}N_{21} + N_{11}N_{22} - (\lambda_2/\lambda_1)N_{11}N_{22})}\psi_1 + \psi_2 + \frac{\lambda_3}{\lambda_2}(cv)\psi_3$$

We note that the ψ_3 component has dropped by the ratio λ_3/λ_2 after the iteration. Note that the ψ_1 component in the above is proportional to the ψ_3 component (cv) in the estimate of ψ_2 at the beginning of the iteration. Therefore, both the ψ_1 and ψ_3 components drop out of the ψ_2 estimate as λ_3/λ_2.

Now we will look at the estimated ψ_1 component. We note that dividing the iterate by a constant does not affect the eigenvalue estimates so we divide (A.3) by x and then label the estimate as ψ_1 because the root corresponding to ψ_1 will be chosen.

$$\psi_1 = (av\psi_1 + (1+bv)\psi_2 + cv\psi_3)\frac{1}{x} + ((1+dv)\psi_1 + ev\psi_2 + fv\psi_3)$$

The eigenvalue balance equation is the same, but instead of Eq. A.11 we have

$$q_2 + q_1\frac{1}{x} + q_0\frac{1}{x^2} = 0$$

When estimating ψ_1, note that $1/x$ will be very small when $v \approx 0$, so that the inverse quadratic part of this equation can be ignored, leading to

$$\frac{1}{x} = -q_2/q_1 \tag{A.13}$$

After all v^2 terms in q_2 are dropped, (A.10) and (A.7) become

$$q_2 = ev(-\lambda_1 N_{12}N_{21} + \lambda_2 N_{12}N_{21} + \lambda_1 N_{11}N_{22} - \lambda_2 N_{11}N_{22}) + fv(-\lambda_1 N_{12}N_{31} + \lambda_3 N_{12}N_{31} + \lambda_1 N_{11}N_{32} - \lambda_3 N_{11}N_{32})$$

and after all v and v^2 terms in q_1 are dropped, (A.9) and (A.6) become

$$q_1 = (-\lambda_1 N_{12}N_{21} + \lambda_2 N_{12}N_{21} + \lambda_1 N_{11}N_{22} - \lambda_2 N_{11}N_{22})$$

From (A.13)

$$\frac{1}{x} = -ev - \frac{fv(-\lambda_1 N_{12}N_{31} + \lambda_3 N_{12}N_{31} + \lambda_1 N_{11}N_{32} - \lambda_3 N_{11}N_{32})}{(-\lambda_1 N_{12}N_{21} + \lambda_2 N_{12}N_{21} + \lambda_1 N_{11}N_{22} - \lambda_2 N_{11}N_{22})}$$

We now note that

$$A\psi_1 = (av\lambda_1\psi_1 + (1+bv)\lambda_2\psi_2 + cv\lambda_3\psi_3)\frac{1}{x} + ((1+dv)\lambda_1\psi_1 + ev\lambda_2\psi_2 + fv\lambda_3\psi_3)$$

If the small terms associated with v^2 and $v\frac{1}{x}$ are dropped, then dividing the equation by λ_1 yields (to first order in v)

$$\frac{1}{\lambda_1}A\psi_1 = \frac{\lambda_2}{\lambda_1}\psi_2\frac{1}{x} + \left(\psi_1 + ev\frac{\lambda_2}{\lambda_1}\psi_2 + fv\frac{\lambda_3}{\lambda_1}\psi_3\right)$$

Substituting for $\frac{1}{x}$,

$$\frac{1}{\lambda_1}AT_1 = \psi_1 + fv\frac{\lambda_3}{\lambda_1}\psi_3 + fv\frac{(\lambda_2/\lambda_1)(-\lambda_1 N_{12}N_{31} + \lambda_3 N_{12}N_{31} + \lambda_1 N_{11}N_{32} - \lambda_3 N_{11}N_{32})}{(-\lambda_1 N_{12}N_{21} + \lambda_2 N_{12}N_{21} + \lambda_1 N_{11}N_{22} - \lambda_2 N_{11}N_{22})}\psi_2$$

We see that the ψ_3 component is decreasing as λ_3/λ_1 and that the ψ_2 component is proportional to the ψ_3 component (fv) at the beginning of the iteration. Thus the ψ_2 component also should be falling as λ_3/λ_1. Thus the procedure is converging to the first eigenfunction at the accelerated rate λ_3/λ_1.

References

[1] T.E. Booth, Nucl. Sci. Eng. 143 (2003) 291.
[2] T.E. Booth, Nucl. Sci. Eng. 154 (2006) 48.
[3] J. Spanier, E.M. Gelbard, Monte Carlo Principles and Neutron Transport Problems, Addison-Wesley, Reading MA, 1969.
[4] C.J. Thompson, Mathematical Statistical Mechanics, Princeton University Press, Princeton, 1972.
[5] S. Sachdev, Quantum Phase Transitions, Cambridge University Press, Cambridge, 1999.
[6] J.H. Wilkinson, The Algebraic Eigenvalue Problem, Oxford University Press, Oxford, 1965.
[7] G.H. Golub, C.F. van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, 1989.
[8] B.L. Hammond, W.A. Lester Jr., P.J. Reynolds, Monte Carlo Methods in Ab Initio Chemistry, World Scientific, Singapore, 1994.
[9] L. Onsager, Phys. Rev. 65 (1944) 117.
[10] E.W. Montroll, J. Chem. Phys. 9 (1941) 706.

[11] B.T. Smith, J.M. Boyle, J.J. Dongarra, B.S. Garbow, Y. Ikebe, V.C. Klema, C.B. Moler, Matraix Eigensystems Routines – EISPACK Guide, Springer-Verlag, Berlin, 1976.
[12] A. Montorsi (Ed.), The Hubbard Model, World Scientific, Singapore, 1992.
[13] H.S. Wilf, Mathematics for the Physical Sciences, Dover, New York, 1978.
[14] E. Anderson et al, LAPACK Users' Guide, third ed., SIAM, Philadelphis, 1999.
[15] Subroutine DNLASO.f, authors D.S. Scott and B.N. Parlett, <www.netlib.org>.
[16] M. Tinkham, Group Theory and Quantum Mechanics, Dover, New York, 1964.
[17] H.-Q. Lin, J.E. Gubernatis, Comput. Phys. 7 (1998) 400.
[18] Y.-Q. Wang, H.-Q. Lin, J.E. Gubernatis, Commun. Comp. Phys. 1 (2006) 575.
[19] T.E. Booth, J.E. Gubernatis, unpublished.