THE INVERSE POWER METHOD (or INVERSE ITERATION)

-- application of the Power method to $A^{-1}$, or (more usually) to $(A - \rho I)^{-1}$, for some fixed constant $\rho$ (which is called a shift).

If the eigenpairs of $A$ are $\{\lambda_i, x^{(i)}\}$, then the eigenpairs of $(A - \rho I)^{-1}$ are

$$\left\{ \frac{1}{\lambda_i - \rho}, x^{(i)} \right\}.$$ 

Let $q^{(0)}$ be the starting vector and (assuming that $A$ is nondefective) suppose that

$$q^{(0)} = \alpha_1 x^{(1)} + \alpha_2 x^{(2)} + \cdots + \alpha_n x^{(n)}.$$ 

Then

$$q^{(1)} = (A - \rho I)^{-1} q^{(0)}$$

$$= \frac{\alpha_1}{\lambda_i - \rho} x^{(1)} + \frac{\alpha_2}{\lambda_2 - \rho} x^{(2)} + \cdots + \frac{\alpha_n}{\lambda_n - \rho} x^{(n)}.$$ 

If $\rho \approx$ some $\lambda_i$, then

$$\frac{1}{|\lambda_i - \rho|} \gg \frac{1}{|\lambda_j - \rho|} \quad \text{for } j \neq i,$$

which implies that the iteration will converge very fast.

NOTE. Shifts are much more effective for the inverse Power method than for the Power method. They enable you to compute any eigenvector, not just those associated with the largest or the smallest eigenvalue.
EXAMPLE

$$A = \begin{bmatrix} -2 & 2 & 2 & 2 \\ -3 & 3 & 2 & 2 \\ -2 & 0 & 4 & 2 \\ -1 & 0 & 0 & 5 \end{bmatrix}$$

has eigenpairs

$$1, \begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}, 2, \begin{bmatrix} 3 \\ 3 \\ 2 \\ 1 \end{bmatrix}, 3, \begin{bmatrix} 2 \\ 2 \\ 2 \\ 1 \end{bmatrix}, 4, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$
Use of the INVERSE POWER METHOD

1. You need a reasonably good approximation to some eigenvalue -- this is obtained by other means.

2. You do not explicitly compute \((A - \rho I)^{-1}\). Instead, for example, you solve the linear system

\[(A - \rho I)q^{(i)} = q^{(0)}\]

for \(q^{(i)}\). You need to compute the LU factorization of \(A - \rho I\) only once, and then use it to solve a linear system in each iteration. If \(A\) is a full matrix, the cost is \(\approx (2/3)n^3\) flops for the LU factorization and \(\approx 2n^2\) flops to compute each vector \(q^{(i)}\). These costs are much cheaper if \(A\) is upper Hessenberg or tridiagonal.

3. As \(\rho \to\) some eigenvalue \(\lambda_i\), \(A - \rho I\) becomes very close to a singular matrix, which suggests that

\[(A - \rho I) q^{(i+1)} = q^{(i)}\]

may be difficult to solve; \(A - \rho I\) will likely be very ill-conditioned.

However, this algorithm works well in practice -- see the discussion on page 324. Using, for example, Gaussian elimination with partial pivoting, the computed approximation to \(q^{(i+1)}\) is the exact solution of some perturbed linear system

\[(A + \delta A - \rho I) \hat{q}^{(i+1)} = q^{(i)}, \text{ where } \frac{\|\delta A\|}{\|A - \rho I\|} \text{ is very small. In this case, } \rho \text{ is very likely also a good approximation to an eigenvalue of } A + \delta A, \text{ and thus } \hat{q}^{(i+1)} \text{ should be very close to an eigenvector of } A + \delta A, \text{ which should then also be very close to an eigenvector of } A. \text{ Thus, inverse iteration is usually very effective.}

From above, how effective it is depends on whether or not a small perturbation \(\delta A\) in \(A - \rho I\) results in only a small perturbation in its eigenvalue and eigenvector. That is, this depends on the condition number of this eigenvalue and eigenvector of \(A - \rho I\) and not on the condition number of \(A - \rho I\) (with respect to solving the above linear system). In other words, the effectiveness of inverse iteration does not depend on the condition number \(\kappa(A - \rho I)\).

THE RAYLEIGH QUOTIENT (page 324)

Given a (complex) vector \(x\) with \(n\) entries and a (complex) \(n \times n\) matrix \(A\), the Rayleigh quotient is
Theorem 5.3.24 (page 325)

Given $A$ and $x$, let

$$r(\mu) = Ax - \mu x.$$  

Then $\|r(\mu)\|_2$ is minimized when $\mu = \frac{x^*Ax}{x^*x}$, the Rayleigh quotient.

Proof.

$$\min_{\mu} \|Ax - \mu x\|_2$$

is a least-squares problem in 1 variable, namely $\mu$. Its associated linear system is the

$n \times 1$ over-determined linear system

$$x\mu = Ax.$$

The solution can be obtained from the normal equations (a $1 \times 1$ linear system), which is

$$(x^*x)\mu = x^*Ax,$$

from which it follows that

$$\mu = \frac{x^*Ax}{x^*x}.$$  

Alternative proof: Given $A$ and $x$, $\{Ax - \mu x\}$ is a line (a 1-dimensional subspace) in the vector space of complex vectors with $n$ entries.
\[
\min_{\mu} \|Ax - \mu x\|_2^2 \quad \text{is obtained when} \quad \mu \quad \text{is such that} \quad Ax - \mu x \quad \text{is orthogonal to} \quad x.
\]

Therefore,
\[
(Ax - \mu x, x) = 0 \\
\Rightarrow x^* Ax - \mu x^* x = 0 \\
\Rightarrow \mu = \frac{x^* Ax}{x^* x}.
\]

NOTE. If \(x\) is an eigenvector of \(A\) (suppose that \(Ax = \lambda x\)), then on multiplying both sides by \(x^*\),
\[
x^* Ax = \lambda x^* x \quad \Rightarrow \quad \lambda = \frac{x^* Ax}{x^* x}.
\]

That is, the Rayleigh quotient with respect to an eigenvector is equal to an eigenvalue.

Thus,
\[
\|Ax - \mu x\|_2 = 0
\]

if and only if \((\mu, x)\) is an eigenpair of \(A\), and if \(x\) is not an eigenvector of \(A\), then
\[
\|Ax - \mu x\|_2 \neq 0 \quad \text{but} \quad \min_{\mu} \|Ax - \mu x\|_2 \quad \text{occurs when} \quad \mu = \frac{x^* Ax}{x^* x}. 
\]

The point of the above discussion: given \(A\) and \(x\) (and thinking of \(x\) as an approximation to an eigenvector), the Rayleigh quotient is the best approximation (in the above sense) to an eigenvalue.

Given a vector \(q\), the following result indicates how good an approximation to an eigenvalue the Rayleigh quotient is, in terms of how close \(q\) is to an eigenvector.

**Theorem 5.3.25** (page 326)

Suppose \(Av = \lambda v\), where \(\|v\|_2 = 1\). Let \(q\) be any vector with \(\|q\|_2 = 1\), and let \(\rho = \frac{q^* Aq}{q^* q} = q^* Aq\). Then
\[
|\lambda - \rho| \leq 2 \|A\|_2 \|v - q\|_2.
\]
Proof.

\[ Av = \lambda v \quad \Rightarrow \quad \lambda = \frac{v^* A v}{v^* v} = v^* A v. \]

Therefore,

\[
\begin{align*}
\lambda - \rho &= v^* A v - q^* A q \\
&= v^* A v - v^* A q + v^* A q - q^* A q \\
&= v^* A(v - q) + (v - q)^* A q
\end{align*}
\]

which implies that

\[
|\lambda - \rho| \leq |v^* A(v - q)| + |(v - q)^* A q|.
\]

Recall the Cauchy-Schwarz inequality:

\[ |(x, y)| \leq \|x\|_2 \|y\|_2, \quad \text{where} \quad (x, y) = y^* x. \]

This gives

\[
|v^* A(v - q)| \leq \|v\|_2 \|A(v - q)\|_2 = \|A(v - q)\|_2 \leq \|A\|_2 \|v - q\|_2.
\]

Similarly,

\[
|(v - q)^* A q| \leq \|v - q\|_2 \|A\|_2.
\]

Thus, from above,

\[
|\lambda - \rho| \leq 2 \|A\|_2 \|v - q\|_2.
\]

EXAMPLE

If \( \|v - q\|_2 < \varepsilon \), then \( |\lambda - \rho| \leq 2 \|A\|_2 \varepsilon \).

That is, if \( \|v - q\|_2 \) is \( O(\varepsilon) \), then \( |\lambda - \rho| \) is also \( O(\varepsilon) \).
THE RAYLEIGH QUOTIENT ITERATION

-- a variation of the inverse Power method
-- a different shift is used at each step, namely the Rayleigh quotient of the
current approximate eigenvector

Algorithm

1. Choose a starting vector $x$ such that $x^*x = 1$ (that is, $\|x\|_2 = 1$).

2. Repeat until convergence:
   2.1 $\rho \leftarrow x^*Ax$
   2.2 apply the inverse Power method: solve $(A - \rho I)\hat{x} = x$ for $\hat{x}$
   2.3 $x \leftarrow \frac{\hat{x}}{\|\hat{x}\|_2}$ (that is, normalize $\hat{x}$)

Because a different shift is used at every step, this algorithm is not guaranteed to
converge to an eigenvector. However, in practice it very seldom fails, and usually
converges very rapidly to an eigenvector of $(A - \rho I)^{-1}$ corresponding to the dominant
eigenvalue of $(A - \rho I)^{-1}$, for some $\rho$. Since $\left\{\rho_j\right\}$ computed in Step 2.1 are Rayleigh
quotients,

$\left\{x\right\} \rightarrow$ an eigenvector of $A$ corresponding to an eigenvalue $\lambda \approx \rho$ of $A$.

The Rayleigh quotient of this computed eigenvector then gives the approximation to $\lambda$. 
If convergent, then the Rayleigh Quotient algorithm converges quadratically (order 2). The proof assumes that the eigenvalue $\lambda_i$ (to which it converges) is simple (multiplicity is 1).

A sketch of the argument:

Suppose that the eigenpairs of $A$ are $(\lambda_i, v_i)$. The Rayleigh Quotient algorithm computes

$\{\rho_i\} \rightarrow \lambda_i$ (assumed to be a simple eigenvalue)
$\{q_i\} \rightarrow$ eigenvector $v_i$

Suppose that $\lambda_k$ is the closest eigenvalue to $\lambda_i$.

Analysis of the Inverse Power method gives

$$\|v_i - q_{j+1}\|_2 \approx r_j \|v_i - q_j\|_2,$$

where $r_j$ is the ratio of the 2 largest eigenvalues of $(A - \rho_j I)^{-1}$. That is,

error of the $(j + 1)^{th}$ approximation to $q_{j+1}$ to $v_i$ is

$$\approx r_j \times \text{error of the } j^{th} \text{ approximation } q_j \text{ to } v_i.$$

For $j$ sufficiently large, since $\frac{1}{\lambda_i - \rho_j}$ is the largest eigenvalue of $(A - \rho_j I)^{-1}$,

$$r_j = \frac{1}{\frac{1}{\lambda_k - \rho_j} - \frac{1}{\lambda_i - \rho_j}} = \frac{1}{\frac{\lambda_i - \rho_j}{\lambda_k - \rho_j}} \approx \frac{|\lambda_i - \rho_j|}{|\lambda_k - \lambda_i|} \leq \frac{2\|A\|_2 \|v_i - q_j\|_2}{|\lambda_k - \lambda_i|}$$

by Thm. 5.3.25.

Therefore,

$$\|v_i - q_{j+1}\|_2 \approx \left(\frac{2\|A\|_2}{|\lambda_k - \lambda_i|}\right) \|v_i - v_j\|_2^2,$$

which is quadratic convergence.

NOTE. If $A$ is a real symmetric (or Hermitian) matrix, the Rayleigh Quotient algorithm is cubically convergent.
$$\|y_i - q_{j+1}\|_2 \approx \text{constant} \times \|y_i - q_j\|_2^3.$$ 

Reason: for Hermitian matrices $A$, the result of Theorem 5.3.25 is stronger -- if $\|y - q\|_2 = O(\varepsilon)$ and $\rho = q^* A q$, then $|\lambda - \rho| = O(\varepsilon^2)$. For a discussion of this, see the middle of page 328.

COST OF THE RAYLEIGH QUOTIENT ALGORITHM (page 329)

$O(n^3)$ flops per iteration for a full matrix, since the $LU$ factorization must be re-computed for each iteration.

However, the cost is only $O(n^2)$ flops per iteration if $A$ is upper Hessenberg. See Exercise 5.3.33.