Chapter 9

PRECONDITIONING

9.1 Preconditioned Krylov Subspace Methods

Saad, Sections 9.0, 9.1

The rate of convergence of a Krylov subspace method for a linear system $Ax = b$ depends on the condition number of the matrix A. Therefore, if we have a matrix M which is a crude approximation to A, $M^{-1}A$ is closer to the identity than is A and should have a smaller condition number, and it would be expected that a Krylov subspace method would converge faster for the "preconditioned" system

$$M^{-1}Ax = M^{-1}b.$$

For example, choosing M to be the diagonal part of A can be quite helpful. This is a useful approach only if computing $M^{-1}v$ for an arbitrary vector v is cheap. Such a matrix M is called a *preconditioner* or, more precisely, a *left preconditioner*.

In the case of a *right preconditioner*, one solves

$$AM^{-1}u = b \quad \text{where } x = M^{-1}u.$$

And if M is in factored form $M = M_{\mathrm{L}}M_{\mathrm{R}}$, one can use this as a *split preconditioner* by solving

$$M_{\mathrm{L}}^{-1}AM_{\mathrm{R}}^{-1}u = M_{\mathrm{L}}^{-1}b \quad \text{where } x = M_{\mathrm{R}}^{-1}u.$$

Note that $M^{-1}A$, AM^{-1} and $M_{\mathrm{L}}^{-1}AM_{\mathrm{R}}^{-1}$ are similar matrices, so they have the same eigenvalues.

For convenience write the given system as

$$A(x - x_0) = r_0 \quad \text{where } r_0 = b - Ax_0,$$

so that $x - x_0$ is the unknown, and consider a split preconditioner $M = M_{\mathrm{L}}M_{\mathrm{R}}$. A left preconditioner is the case $M_{\mathrm{R}} = I$ and a right preconditioner is the case $M_{\mathrm{L}} = I$. Then the preconditioned system is

$$M_{\mathrm{L}}^{-1}AM_{\mathrm{R}}^{-1}u = M_{\mathrm{L}}^{-1}r_0 \quad \text{where } x = x_0 + M_{\mathrm{R}}^{-1}u$$

and the Krylov algorithm is

$x_0 =$ initial guess;
$r_0 = b - Ax_0$;
perform k iterations of
    the Krylov subspace method for $(M_{\mathrm{L}}^{-1}AM_{\mathrm{R}}^{-1})u = (M_{\mathrm{L}}^{-1}r_0)$ with $u_0 = 0$;
$x_k = x_0 + M_{\mathrm{R}}^{-1}u_k$

Hence

$$x_k \in x_0 + M_{\mathrm{R}}^{-1}\mathcal{K}_k(M_{\mathrm{L}}^{-1}AM_{\mathrm{R}}^{-1}, M_{\mathrm{L}}^{-1}r_0) = x_0 + \mathcal{K}_k(M^{-1}A, M^{-1}r_0),$$

and the effect of preconditioning is to change the Krylov subspace to $\mathcal{K}(M^{-1}A, M^{-1}r_0)$. Thus, it is only the product $M = M_{\mathrm{L}}M_{\mathrm{R}}$ that determines the subspace. However, the actual splitting effects the convergence test.

Review questions

1. What two properties should a preconditioner have?
2. Explain how a left preconditioner is used?
3. Explain how a right preconditioner is used?
4. Explain how a split preconditioner is used?
5. How does the spectrum of the preconditioned matrix depend on the splitting of the preconditioner? How does the Krylov subspace depend on the splitting of the preconditioner?
6. If a preconditioner M is used for the system $Ax = b$ with initial guess x_0, what is the modified Krylov subspace from which $x - x_0$ is obtained?

9.2 Preconditioned Conjugate Gradient

Saad, Sections 9.2.0, 9.2.1

Let A be symmetric positive definite. To use the highly effective conjugate gradient method, the preconditioned matrix must also have this property. This can be achieved by using a split preconditioner

$$M = LL^{\mathsf{T}}$$

where the only assumption is that L is nonsingular.

To expand the generic preconditioned Krylov subspace algorithm for the conjugate gradient method, (i) relabel the variables of the conjugate gradient algorithm of Section 8.4 by changing x to u, r to $\bar{r}$, p to $\bar{p}$, and (ii) in that same algorithm substitute $L^{-1}AL^{-\mathsf{T}}$ for A, $L^{-1}r_0$ for b, and 0 for the initial guess:

$x_0 =$ initial guess;
$r_0 = b - Ax_0$;
$u_0 = 0$;
$\bar{r}_0 = L^{-1}r_0$;
$\bar{p}_0 = \bar{r}_0$;
for $i = 0, 1, 2, \ldots$ **do** {
$\quad \alpha_i = \dfrac{\bar{r}_i^{\mathsf{T}}\bar{r}_i}{\bar{p}_i^{\mathsf{T}}L^{-1}AL^{-\mathsf{T}}\bar{p}_i}$;
$\quad u_{i+1} = u_i + \alpha_i\bar{p}_i$;
$\quad \bar{r}_{i+1} = \bar{r}_i - \alpha_i L^{-1}AL^{-\mathsf{T}}\bar{p}_i$;
$\quad \bar{p}_{i+1} = \bar{r}_{i+1} + \dfrac{\bar{r}_{i+1}^{\mathsf{T}}\bar{r}_{i+1}}{\bar{r}_i^{\mathsf{T}}\bar{r}_i}\bar{p}_i$;
}
$x_i = x_0 + L^{-\mathsf{T}}u_i$

It happens that the above algorithm can be rewritten so that L^{-1} and $L^{-\mathsf{T}}$ occur only as the product $M^{-1} = L^{-\mathsf{T}}L^{-1}$: Let $r_i = L\bar{r}_i$ and $p_i = L^{-\mathsf{T}}\bar{p}_i$ and the algorithm becomes

$x_0 =$ initial guess;
$r_0 = b - Ax_0$;
$p_0 = M^{-1}r_0$;
for $i = 0, 1, 2, \ldots$ **do** {
$\quad \alpha_i = \dfrac{r_i^{\mathsf{T}}M^{-1}r_i}{p_i^{\mathsf{T}}Ap_i}$;
$\quad x_{i+1} = x_i + \alpha_i p_i$;
$\quad r_{i+1} = r_i - \alpha_i Ap_i$;
$\quad p_{i+1} = M^{-1}r_{i+1} + \dfrac{r_{i+1}^{\mathsf{T}}M^{-1}r_{i+1}}{r_i^{\mathsf{T}}M^{-1}r_i}p_i$;
}

Hence, one can use a symmetric positive definite matrix M as a split preconditioner without actually factoring it!

Preconditioning has no effect on the definition of the energy norm $|||\cdot|||$. Hence, preconditioned CG minimizes the original energy norm of the error, though in a different affine space.

Review questions

1. What properties should a preconditioner M possess for use with the conjugate gradient method?
2. Which kind of splitting $M = M_{\mathrm{L}}M_{\mathrm{R}}$ is suitable for use with the conjugate gradient method?
3. The conjugate gradient method with split preconditioner $M = LL^{\mathsf{T}}$ can be transformed to a potentially more convenient (and efficient) form. What is this convenience?
4. What is the effect of preconditioning on the definition of the energy norm?

9.3 Preconditioned Generalized Minimal Residual

Saad, Sections 9.3

The effect of preconditioning is not only to change the Krylov subspace but also to *change the residual norm* being minimized—to $\|M_{\mathrm{L}}^{-1}(b - Ax)\|$.

Review question

1. What is the effect of preconditioning with split preconditioner $M = M_{\mathrm{L}}M_{\mathrm{R}}$ on the norm of the residual that is minimized by GMRES?

9.4 Relaxation Method Preconditioners

Saad, Sections 10.2. Note the error in transcribing Equation (4.27) to the middle of page 285.

Recall the theoretical form of relaxation methods

$$\begin{aligned} x^{(k+1)} &= x^{(k)} + D^{-1}(b - Ax^{(k)}), \quad \text{Jacobi,} \\ x^{(k+1)} &= x^{(k)} + (L + \frac{1}{\omega}D)^{-1}(b - Ax^{(k)}), \quad \text{SOR,} \end{aligned}$$

where $A = L + D + U$, in which $M = D$ or $M = L + (1/\omega)D$ serves the role as a crude approximation to A. Hence, these are possible as preconditioners. Within a Krylov subspace iteration, an SOR preconditioner might be implemented as indicated by the theoretical form

with the understanding, of course, that premultiplication by a lower triangular matrix is implemented as forward substitution. When used as a preconditioner, a near optimal choice of the relaxation parameter is not critical, and the Gauss-Seidel choice $\omega = 1$ can be quite effective.

For a symmetric positive definite matrix A, an SOR preconditioner is unsuitable since it is not symmetric. However, SSOR can be used, in which a forward sweep of SOR

$$x^{(k+1/2)} = x^{(k)} + (L + \frac{1}{\omega}D)^{-1}(b - Ax^{(k)})$$

is followed by a backward sweep

$$x^{(k+1)} = x^{(k+1/2)} + (U + \frac{1}{\omega}D)^{-1}(b - Ax^{(k+1/2)}).$$

Note that in a backward sweep the roles of L and U are exchanged. The effective preconditioner M_{SSOR} is obtained by eliminating $x^{(k+1/2)}$ from the equations above. Defining $r^{(k)}$ and $r^{(k+1/2)}$ to be the residuals, we get

$$\begin{aligned}
x^{(k+1/2)} - x^{(k)} &= (L + \frac{1}{\omega}D)^{-1}r^{(k)} \\
r^{(k+1/2)} &= r^{(k)} - A(x^{(k+1/2)} - x^{(k)}) = (L + \frac{1}{\omega}D - A)(x^{(k+1/2)} - x^{(k)}) \\
&= ((\frac{1}{\omega} - 1)D - U)(L + \frac{1}{\omega}D)^{-1}r^{(k+1/2)} \\
x^{(k+1)} - x^{(k+1/2)} &= (U + \frac{1}{\omega}D)^{-1}r^{(k+1/2)} \\
x^{(k+1)} - x^{(k)} &= (U + \frac{1}{\omega}D)^{-1}((\frac{1}{\omega} - 1)D - U + (U + \frac{1}{\omega}D))(L + \frac{1}{\omega}D)^{-1}r^{(k)} \\
&= (U + \frac{1}{\omega}D)^{-1}(\frac{2}{\omega} - 1)D(L + \frac{1}{\omega}D)^{-1}r^{(k)},
\end{aligned}$$

hence

$$M_{\mathrm{SSOR}} = (I + \omega LD^{-1})\frac{1}{\omega(2 - \omega)}D(I + \omega D^{-1}U).$$

Note that this is an LDU factorization.

Review questions

1. What is the preconditioner of the Jacobi iteration for solving a linear system?
2. What is the preconditioner of Gauss-Seidel?
3. In the case where the original matrix is symmetric, what common relaxation methods use a symmetric preconditioner?
4. What is the form of the SSOR preconditioner?

9.5 Incomplete LU Factorization Preconditioners

Saad, Sections 10.1, 10.3.0

The ILU(0) factorization of a matrix A consists of a unit lower triangular matrix L having the same sparsity structure as the lower part of A and an upper triangular matrix U having the same sparsity structure as the the diagonal and upper part of A such that

$$(LU)_{ij} = a_{ij}, \quad \text{for } a_{ij} \neq 0.$$

Hence, there is an equation to satisfy for each unknown element of L and U. These equations can be solved just as in the case of a dense matrix.

For a symmetric positive matrix, an incomplete Cholesky decomposition can be computed.

Review question

1. What conditions determine the ILU(0) factorization of an n by n matrix A? Give one condition for each (i, j) in the range $1 \leq i, j \leq n$.

9.6 Approximate Inverse Preconditioners

Saad, Sections 10.5.0, 10.5.1

ILU preconditioners can be poor for matrices that are indefinite or have large nonsymmetric parts due to large elements in L^{-1} or U^{-1}. For such cases it may be more effective to calculate an approximation N to A^{-1}, which would be the inverse of the preconditioner. For example, for right preconditioning, N is chosen so that it is sparse and

$$\|I - AN\|_{\mathrm{F}}$$

is minimized where F denotes the Frobenius norm. To develop an algorithm, it is helpful to expand this by columns as

$$\|I - AN\|_{\mathrm{F}}^2 = \sum_{j=1}^{n} \|e_j - An_j\|^2$$

where n_j is the jth column of N.

Review questions

1. What two properties determine the inverse N of an approximate inverse preconditioner? State your answer for the case of a right preconditioner.

2. Show how $\|I - AN\|_{\mathrm{F}}^2$ can be expanded in terms of columns n_j of N.