29 Non-stationary methods

29.1 Motivation, via preconditioners

In recent lectures, we studied the Gauss-Seidel and Jacobi (and, in an exercises, SOR) methods for solving linear systems of the form

$$A\mathbf{x} = \mathbf{b}.$$

We call a matrix M a *preconditioner* for A if $M^{-1}A \approx I$. Then, instead of solving $A\mathbf{x} = vec\mathrm{b}$, we solve

$$M^{-1}A\mathbf{x} = M^{-1}\mathbf{b}.$$

If indeed $M^{-1}A \approx I$, this should be a relatively easy task. For the Jacobi method, we took $M = \mathrm{diag}(A)$, and then we iterated as

$$x^{(k+1)} = M^{-1}(M - A)x^{(k)} + M^{-1}\mathbf{b}.$$

This can be rearranged as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + M^{-1}(\mathbf{b} - A\mathbf{x}^{(k)}). \tag{29.1}$$

Several observations come from this

(a) Recall that the *residual* is $\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)}$. So we can write (29.1) as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + M^{-1}\mathbf{r}^{(k)}. \tag{29.2}$$

(b) Suppose that, in fact, $\mathbf{x}^{(k)} = \mathbf{x}$. Since the iteration should always give that $\|\mathbf{x} - x^{(k+1)}\| \leqslant \|\mathbf{x} - x^{(k)}\|$, then we need $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$. This is achieved when $\|\mathbf{r}^{(k)}\| = 0$. Since this is not achievable (realistically), when instead want to minimise $\|M^{-1}\mathbf{r}^{(k)}\|$.

(c) Different methods give different choices of M. Why not allow the freedom for this value to change at each iteration, in order to maximise the error reduction at each step?

The topic of choosing a preconditioner, M, is a deep one, and not really part of this course (it will be part of the follow-on workshop in May). But here we use it as a motivation.

29.2 Orthomin(1)

See [Gre97, Chap. 2]. As a variation on (29.1), let's consider methods of the form

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + a_k(\mathbf{b} - A\mathbf{x}^{(k)}), \tag{29.3}$$

and where the goal is to choose the scalar a_k in the best way possible. As discussed above, this could be by trying to minimise the 2-norm of the residual, $\mathbf{r}^{(k+1)}$. With a little work (take notes!) this can be done by picking

$$a_k = \frac{(\mathbf{r}^{(k)}, A\mathbf{r}^{(k)})}{(A\mathbf{r}^{(k)}, A\mathbf{r}^{(k)})}. \tag{29.4}$$

This method is known (among other things) as *Orthomin(1)*.

29.3 Steepest Descent

Suppose that A is s.p.d. Then $\|\mathbf{x}\|_A := \sqrt{\mathbf{x}^T A\mathbf{x}}$ is a norm. We could choose a_k in (29.3) to minimise $\mathbf{e}^{(k+1)} = \mathbf{x} - \mathbf{x}^{(k)}$ in this norm. This leads to

$$a_k = \frac{(\mathbf{r}^{(k)}, \mathbf{r}^{(k)})}{(\mathbf{r}^{(k)}, A\mathbf{r}^{(k)})}, \tag{29.5}$$

which is called the method of *steepest descent*. See Exercise 29.1

29.4 Exercises

Exercise 29.1. Suppose that we want to choose a_k in (29.3) to minimise $\|\mathbf{e}^{(k+1)}\|_A$. Recalling that $\mathbf{e}^{(k+1)} = \mathbf{e}^{(k)} - a_k\mathbf{r}^{(k)}$, show that this leads to the formula in (29.5). *Hint: see [Gre97, §1.3.2].*