32 Conjugate Gradient Method (CG)

(*These notes where heavily revised after the class*).
In the final section of this course, we will look at one of the most popular iterative methods for solving linear systems of equations when the system matrix is s.p.d.

32.1 Example

The CG method is more complicated to state and analyse than other iterative methods we have studied, such has the Gauss-Seidel method. So, even before we state the method, we begin by demonstrating that it is worth the effort: it is *much* faster than Gauss-Seidel.

Suppose we apply the Gauss-Seidel method and CG to the linear system that arises from solving the Laplace equation on a 16×16 grid. So A has 256 rows and columns. In Figure 32.1 below we show the residual reduction for both methods. It is clear that CG is converging *much* faster that Gauss-Seidel.

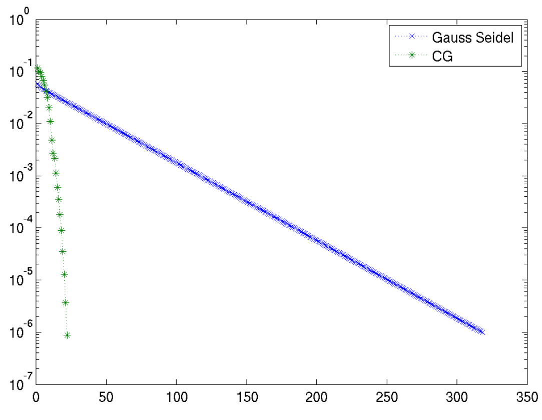

Figure 32.1: Gauss-Seidel and CG residuals

Looking at just the residual reduction in CG, as shown in Figure 32.2, one could be convinced that, unlike Gauss-Seidel, we might actually achieve the exact solution in a finite number of steps.

32.2 Conjugates

We say that two vectors, x and y, are *conjugate*, with respect to the matrix A, if $x^T A y = 0$. When A is s.p.d., the bilinear form

$$(x, y)_A := (Ax, y) = x^T A y,$$

is an inner product. So, being *conjugate with respect to* A is the same as being orthogonal with respect to the inner product $(\cdot, \cdot)_A$. So the expression "A-orthogonal" is also used for conjugate w.r.t. A.

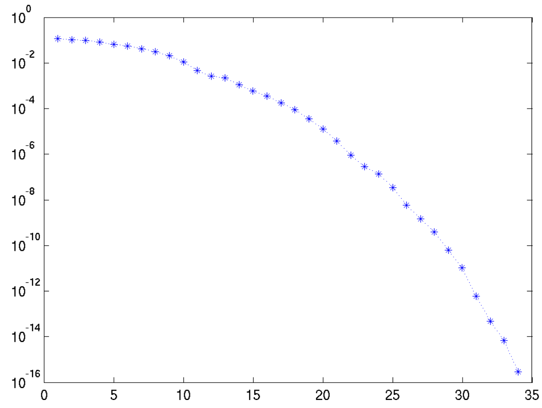

Figure 32.2: CG residuals

32.3 The CG algorithm

Choose , $\mathbf{x}^{(0)}$. Set $\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$ and $\mathbf{p}^{(0)} = \mathbf{r}^{(0)}$. For $k = 1, 2, \dots,$ compute: $A\mathbf{p}^{(k-1)}$ and then set

$$a_{k-1} = \frac{(\mathbf{r}^{(k-1)}, \mathbf{r}^{(k-1)})}{(\mathbf{p}^{(k-1)}, A\mathbf{p}^{(k-1)})}. \tag{32.1}$$

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + a_{k-1}\mathbf{p}^{(k-1)}. \tag{32.2}$$

$$\mathbf{r}^{(k)} = \mathbf{r}^{(k-1)} - a_{k-1}A\mathbf{p}^{(k-1)} \tag{32.3}$$

$$b_{k-1} = \frac{(\mathbf{r}^{(k)}, \mathbf{r}^{(k)})}{(\mathbf{r}^{(k-1)}, \mathbf{r}^{(k-1)})}. \tag{32.4}$$

$$\mathbf{p}^{(k)} = \mathbf{r}^{(k)} + b_{k-1}\mathbf{p}^{(k-1)} \tag{32.5}$$

Important: There are other possible expressions for a_k and b_k. See Exercise 32.1.

32.4 Analysis (Initial ideas)

Usually, we define the residual, $\mathbf{r}^{(k)}$ as $\mathbf{r}^{(k)} := b - Ax^{(k)}$. However, we seem to be using a different definition in (32.3). So we spent a new minutes in class, showing that these formulations are equivalent.

Next we observed that the method is well-defined. It would only fail to be well-defined in the case there the denominator in (32.1) or (32.4) was zero. However, since A is s.p.d., this can only happen if $\mathbf{r}^{(k-1)} = 0$, in which case $x = x^{(k+1)}$, and we are done.

32.5 Exercises

Exercise 32.1. Show that the following expressions are equivalent to those in (32.1) and (32.4)

$$a_{k-1} = \frac{(\mathbf{r}^{(k-1)}, \mathbf{p}^{(k-1)})}{(\mathbf{p}^{(k-1)}, A\mathbf{p}^{(k-1)})}, \tag{32.6}$$

$$b_{k-1} = -\frac{(\mathbf{r}^{(k-1)}, A\mathbf{p}^{(k-1)})}{(\mathbf{p}^{(k-1)}, A\mathbf{p}^{(k-1)})} \tag{32.7}$$

Exercise 32.2. Write a Matlab implementation of the CG algorithm. Use it to generate the diagrams in Figure 32.1.