RESIDUAL CORRECTION

There are two aspects two this idea, the first being a method for improving the accuracy of a solution produced using Gaussian Elimination.

Let $\widehat{x}$ be an approximate solution to $Ax = b$, produced using Gaussian elimination. The latter means that an LU-factorization has been produced for A,

$$A \doteq \widehat{L}\widehat{U} \tag{1}$$

For the error in $\widetilde{x}$, first introduce the residual

$$r = b - A\widehat{x}$$

Then subtract $A\widehat{x} = b - r$ from $Ax = b$, obtaining

$$Ae = r, \qquad e \equiv x - \widehat{x}$$

Solve this equation approximately by using the LU-factorization of (1):

$$e \approx \widehat{e}, \qquad \widehat{L}\widehat{U}\widehat{e} = r$$

$$Ae = r, \qquad e \equiv x - \widehat{x}$$

Solve this equation approximately by using the LU-factorization of (1):

$$e \approx \widehat{e}, \qquad \widehat{L}\widehat{U}\widehat{e} = r$$

Then define an improvement of $\widehat{x}$ using

$$x \approx \widehat{x} + \widehat{e}$$

This leads to a new improved approximation, and with it we can repeat the above process.

We re-write it in iterative form. Let $x^{(0)}$ be an approximate solution to $Ax = b$. Define

$$r^{(0)} = b - Ax^{(0)}$$

Then

$$Ae^{(0)} = r^{(0)}, \qquad e^{(0)} = x - x^{(0)}$$

Solve this approximately using (1), obtaining an approximation $\widehat{e}^{(0)}$. Then define

$$x^{(1)} = x^{(0)} + \widehat{e}^{(0)}$$

We can repeat this process, obtaining a sequence of iterates $\left\{x^{(m)}\right\}$. Given an $x^{(m)}$, define

$$r^{(m)} = b - Ax^{(m)}$$

Solve $Ae^{(m)} = r^{(m)}$ using the LU factorization of (1), obtain an approximate solution $\widehat{e}^{(m)}$. Then define

$$x^{(m+1)} = x^{(m)} + \widehat{e}^{(m)}$$

EXAMPLE

Using 4 digit decimal floating-point arithmetic with rounding. Solve $Ax = b$ with $b = [1, 0, 0]^{\mathsf{T}}$ and

$$A = \begin{bmatrix} 1.000 & .5000 & .3333 \\ .5000 & .3333 & .2500 \\ .3333 & .2500 & .2000 \end{bmatrix}$$

The exact solution is

$$x = [9.062,\ -36.32,\ 30.30]^{\mathsf{T}}$$

With Gaussian elimination, we obtain

$$x^{(0)} = [8.968,\ -35.77,\ 29.77]^{\mathsf{T}}$$

Residual Correction:

$$\begin{aligned} r^{(0)} &= [-.005341,\ -.004359,\ -.005344]^{\mathsf{T}} \\ \widehat{e}^{(0)} &= [.09216,\ -.5442,\ .5239]^{\mathsf{T}} \\ x^{(1)} &= [9.060,\ -36.31,\ 30.29]^{\mathsf{T}} \\ r^{(1)} &= [-.0006570,\ -.0003770,\ -.0001980]^{\mathsf{T}} \\ \widehat{e}^{(1)} &= [.001707,\ -.01300,\ .01241]^{\mathsf{T}} \\ x^{(2)} &= [9.062,\ -36.32,\ 30.30]^{\mathsf{T}} \end{aligned}$$

CALCULATION OF THE RESIDUAL - I

Return to the definition $r = b - A\hat{x}$. In component form,

$$r_i = b_i - \sum_{j=1}^{n} a_{i,j}\hat{x}_j, \qquad i = 1, ..., n$$

We expect the residuals to be small. This implies that calculating r_i involves a loss of significance error. To avoid this, we compute r_i using higher precision arithmetic. If the basic computation is being done in single precision, then each r_i is calculated using double precision; and if the basic computation is being done in double precision, then each r_i is calculated using extended precision (which is possible in hardware with processors based on the IEEE arithmetic standard).

A complete error analysis of the residual correction scheme (or iterative improvement) is given in
N. Higham, *Accuracy and Stability of Numerical Algorithms*, Chap. 11.

(<u>Theorem 11.1</u>) Assume the residuals are computed in extended precision arithmetic, as described above. Introduce

$$\eta \equiv \mathbf{u} \left\| \left| A^{-1} \right| \left| \widehat{L} \right| \left| \widehat{U} \right| \right\|_{\infty}$$

in which $|M|$ denotes the array of order equal to that of M, but composed of the absolute values of the entries in M. Then if η is sufficiently less than 1, residual correction will converge, with the error $\left\| x - x^{(m)} \right\|_{\infty}$ reduced by a factor of approximately η with each iteration. This will eventually lead to an iterate $x^{(*)}$ for which

$$\frac{\left\| x - x^{(*)} \right\|_{\infty}}{\|x\|_{\infty}} \approx \mathbf{u}$$

CALCULATION OF THE RESIDUAL - II

It was widely believed that calculations of the residual in the same precision arithmetic as that used for the remainder of the calculations would not lead to any significant improvement in using residual correction. This was shown to be wrong by R. Skeel in 1980.

(<u>Theorem 11.2</u>) Let the residuals be calculated in the same precision as the Gaussian elimination solution of $Ax = b$, including the calculation of the factorization of (1). Introduce

$$\eta \equiv \mathbf{u} \left\| \left| A^{-1} \right| \left| \widehat{L} \right| \left| \widehat{U} \right| \right\|_{\infty}$$

Then if η is sufficiently less than 1, residual correction will converge, with the error $\left\| x - x^{(m)} \right\|_{\infty}$ reduced by a factor of approximately η with each iteration. This will eventually lead to an iterate $x^{(*)}$ for which

$$\frac{\left\| x - x^{(*)} \right\|_{\infty}}{\|x\|_{\infty}} \approx 2n\mathbf{u} \frac{\left\| |A| \left| A^{-1} \right| |x| \right\|_{\infty}}{\|x\|_{\infty}}$$

This is not as good as the earlier result, but it is an improvement over the use of Gaussian elimination with no residual correction.

RESIDUAL CORRECTION - II

Assume we have an approximation

$$C \approx A^{-1}$$

How can we use this to solve $Ax = b$ approximately? Let $x^{(0)}$ be an initial guess at the solution of $Ax = b$. Then as before, introduce the residual

$$r^{(0)} = b - Ax^{(0)}$$

and obtain

$$Ae^{(0)} = r^{(0)}, \qquad e^{(0)} = x - x^{(0)}$$

To improve upon $x^{(0)}$, use

$$\begin{aligned} e^{(0)} &\approx Cr^{(0)} \\ x^{(1)} &= x^{(0)} + Cr^{(0)} \end{aligned}$$

In general, for any $m \geq 0$, define

$$\begin{aligned} r^{(m)} &= b - Ax^{(m)} \\ x^{(m+1)} &= x^{(m)} + Cr^{(m)} \end{aligned}$$

CONVERGENCE ANALYSIS

$$\begin{aligned} x - x^{(m+1)} &= x - x^{(m)} - Cr^{(m)} \\ &= x - x^{(m)} - C\left[b - Ax^{(m)}\right] \\ &= x - x^{(m)} - C\left[Ax - Ax^{(m)}\right] \\ &= (I - CA)\left[x - x^{(m)}\right] \end{aligned}$$

By induction,

$$\begin{aligned} x - x^{(m)} &= (I - CA)^m \left(x - x^{(0)}\right) \\ \left\|x - x^{(m)}\right\| &\le \left\|(I - CA)^m\right\| \left\|x - x^{(0)}\right\| \end{aligned} \tag{2}$$

For convergence for arbitrary $x^{(0)}$,

$$x^{(m)} \to x \quad \text{if and only if} \quad r_\sigma(I - CA) < 1$$

This will be true if

$$\|I - CA\| < 1 \tag{3}$$

for some matrix norm, and then

$$\left\|x - x^{(m+1)}\right\| \le \|I - CA\| \left\|x - x^{(m)}\right\| \tag{4}$$

Note that

$$r_\sigma\,(I - CA) < 1 \tag{5}$$

implies both A and C are nonsingular. To see this, write

$$\det(C)\det(A) = \det(CA) = \det(I - R)\,, \quad R = I - CA$$

If C or A is singular, then this implies $\det(I - R) = 0$ and 0 is an eigenvalue of $I - R$. But this implies that 1 is an eigenvalue of R, and therefore contradicts (5).

Note also that under the assumption (5), the matrices $I - CA$ and $I - AC$ are similar:

$$I - CA = A\,(I - AC)\,A^{-1}$$

Therefore,

$$r_\sigma\,(I - CA) = r_\sigma\,(I - AC)$$

and

$$\|I - AC\| < 1 \;\Rightarrow\; r_\sigma\,(I - AC) < 1 \;\Rightarrow\; r_\sigma\,(I - CA) < 1$$

EXAMPLE

Let $A(\varepsilon) = A_0 + \varepsilon B$ with

$$A_0 = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix}$$

Solve $A(\varepsilon)x = b$ by the residual correction method with

$$C = A_0^{-1} \approx A(\varepsilon)^{-1}$$

For the convergence,

$$I - CA(\varepsilon) = I - A_0^{-1}\left[A_0 + \varepsilon B\right] = -\varepsilon A_0^{-1} B$$

$$I - CA(\varepsilon) = -\varepsilon \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{4} & -\frac{1}{2} & -\frac{1}{4} \end{bmatrix}$$

$$\|I - CA(\varepsilon)\|_\infty = |\varepsilon|, \qquad r_\sigma\left(I - CA\right) = \frac{|\varepsilon|}{\text{sqrt}(2)}$$

We have convergence if $|\varepsilon| < \text{sqrt}(2)$; and for $|\varepsilon| < 1$, we have from (4) that

$$\left\| x - x^{(m+1)} \right\|_\infty \le |\varepsilon| \left\| x - x^{(m)} \right\|_\infty$$

<u>An improved iteration</u>: For an improved estimate of $A(\varepsilon)^{-1}$ and an improved iteration scheme, use

$$\begin{aligned} A(\varepsilon)^{-1} &= (A_0 + \varepsilon B)^{-1} \\ &= \left(A_0 \left[I + \varepsilon A_0^{-1} B \right] \right)^{-1} \\ &= \left(I + \varepsilon A_0^{-1} B \right)^{-1} A_0^{-1} \\ &\approx \left(I - \varepsilon A_0^{-1} B \right) A_0^{-1} \equiv C_1 \end{aligned}$$

Then

$$\begin{aligned} I - C_1 A(\varepsilon) &= I - \left(A_0^{-1} - A_0^{-1} B A_0^{-1} \right) (A_0 + \varepsilon B) \\ &= \varepsilon^2 \left(A_0^{-1} B \right)^2 \end{aligned}$$

With earlier information,

$$r_\sigma (I - CA) = \frac{\varepsilon^2}{2}$$

Also,

$$I - CA = \frac{\varepsilon^2}{16} \begin{bmatrix} -2 & 0 & -2 \\ 0 & 0 & 0 \\ 4 & 0 & 4 \end{bmatrix}$$

$$\|I - CA\|_\infty = \frac{\varepsilon^2}{2}$$

Thus $\|I - CA\|_\infty < 1$ if and only if $|\varepsilon| <$ sqrt(2). The rate of convergence is now given by

$$\left\|x - x^{(m+1)}\right\|_\infty \leq \frac{\varepsilon^2}{2} \left\|x - x^{(m)}\right\|_\infty$$

CHECKING CONVERGENCE

Checking convergence of an iteration scheme is not always an easy task. In this situation, we have the recursive error formula

$$x - x^{(m+1)} = (I - CA)\left[x - x^{(m)}\right], \qquad m \geq 0 \tag{6}$$

Replace m by $m - 1$, obtaining

$$x - x^{(m)} = (I - CA)\left[x - x^{(m-1)}\right], \qquad m \geq 1$$

Subtract the first from the second, obtaining

$$x^{(m+1)} - x^{(m)} = (I - CA)\left[x^{(m)} - x^{(m-1)}\right], \qquad m \geq 1$$

If $\|I - CA\| < 1$ for some operator matrix norm, we can write

$$\begin{aligned} \left\|x^{(m+1)} - x^{(m)}\right\| &\leq \|I - CA\| \left\|x^{(m)} - x^{(m-1)}\right\| \\ \frac{\left\|x^{(m+1)} - x^{(m)}\right\|}{\left\|x^{(m)} - x^{(m-1)}\right\|} &\leq \|I - CA\| \end{aligned}$$

$$\frac{\left\| x^{(m+1)} - x^{(m)} \right\|}{\left\| x^{(m)} - x^{(m-1)} \right\|} \leq \|I - CA\|$$

We will often attempt to estimate $c \equiv \|I - CA\|$ by using these ratios of the norms of the successive differences. Often, but not always, these ratios are approximately constant as m increases. Thus we use

$$c \approx c_m = \frac{\left\| x^{(m+1)} - x^{(m)} \right\|}{\left\| x^{(m)} - x^{(m-1)} \right\|}$$

or

$$c \approx c_m = \sup_{m-p \leq q \leq m} \frac{\left\| x^{(q+1)} - x^{(q)} \right\|}{\left\| x^{(q)} - x^{(q-1)} \right\|}$$

for some $p > 0$, or perhaps the geometric average of several such successive ratios.

Thus we assume we know $c < 1$ for which (6) is valid. How do we then estimate the error?

$$\begin{aligned}
\left\|x^{(m+1)} - x^{(m)}\right\| &= \left\|\left[x - x^{(m)}\right] - \left[x - x^{(m+1)}\right]\right\| \\
&\ge \left\|\left[x - x^{(m)}\right]\right\| - \left\|\left[x - x^{(m+1)}\right]\right\| \\
&\ge \frac{1}{c}\left\|\left[x - x^{(m+1)}\right]\right\| - \left\|\left[x - x^{(m+1)}\right]\right\| \\
&= \frac{1-c}{c}\left\|\left[x - x^{(m+1)}\right]\right\| \\
\left\|\left[x - x^{(m+1)}\right]\right\| &\le \frac{c}{1-c}\left\|x^{(m+1)} - x^{(m)}\right\|
\end{aligned}$$

In general, this is a very reasonable error bound. We will see it illustrated in a later section with other iteration methods.