Conjugate Gradients

Positive Definite \( A = A' \). To solve \( Az = b \) for \( z = A^{-1}b \), choose \( x_{-1} := x_0 \) arbitrarily and, for \( n = 0, 1, 2, 3, \ldots \) in turn, compute \( x_{n+1} := x_n + \beta_n(b - Ax_n) + \mu_n(x_n - x_{n-1}) \) with scalars \( \beta_n \) and \( \mu_n \) chosen to minimize \( (x_{n+1} - z)'A(x_{n+1} - z) = x_{n+1}'Ax_{n+1} - 2b'x_{n+1} + z'Az \). Simplify typography by dropping subscripts from \( r_n := b - Ax_n, \quad d_n := x_n - x_{n-1} \), \( \beta_n \) and \( \mu_n \) to find that 

\[
[r, d]'A[r, d] = [r, d]'r.
\]

Solve this equation for \( \beta \) and \( \mu \). At least one solution always exists if \( A \) is positive definite.

In the absence of roundoff, successive residuals \( r_n := b - Ax_n \) turn out to be orthogonal; in fact 

\[
r_m'r_n = 0 = d_m'd_n
\]

for all \( m > n \geq 0 \). Consequently \( r_m = 0 \) and \( x_m = z \) at least as soon as \( m \) equals the dimension of \( A \). But the point of the iteration is not to iterate that many times when the dimension is huge. Instead, take advantage of the tendency of residuals \( r_n \) and increments \( d_n \) to dwindle as \( n \) increases, and stop iterating when they both become small enough.

Overrelaxation

Positive Definite \( A = A' = -L + V - L' \) in which \( V := \text{Diag}(A) = V' \) is Positive Definite too, and \( -L = \text{Subdiag}(A) \). Given \( A \) and \( b \) we seek \( z := A^{-1}b \). Starting from an arbitrary initial guess \( x_0 \), for \( n = 0, 1, 2, 3, \ldots \) in turn, ordinary Gauss-Seidel iteration solves 

\[
V(x_{n+1} - x_n) = (V - L)^{-1}(b + Lx_n).
\]

Then \( (x_{n+1} - z) = E(x_n - z) \) where \( E = (V - L)^{-1}L' \) can be shown to have eigenvalues all with magnitudes less than 1, though not necessarily much less unless \( \|L\| \) is rather smaller than the smallest eigenvalue of \( V \).

To accelerate convergence, consider using an Over/Underrelaxation parameter \( \delta \) confined to \(-1 < \delta < 1\) for reasons to be explained later. To solve \( Az = b \) for \( z = A^{-1}b \), choose \( x_0 \) arbitrarily and, for \( n = 0, 1, 2, 3, \ldots \) in turn, solve 

\[
V(x_{n+1} - x_n) = (1+\delta)(b + Lx_n).
\]

Then \( (x_{n+1} - z) = E(x_n - z) \) where \( E = (V - (1+\delta)L)^{-1}(-\delta V + (1+\delta)L') \). Note that, because \( L \) lies strictly below the diagonal, (product of all \( E \)'s eigenvalues) = \( \det(E) = \det(-\delta I) \). Therefore at least one eigenvalue of \( E \) has magnitude at least as big as \( |\delta| \). This is why we keep \(-1 < \delta < 1\).

\( E \) and its eigenvalues depend upon \( \delta \), as well as \( A \), but the dependence is obscure except in special cases. An important special case arises when \( A = \begin{bmatrix} 1 & -B^2 \\ -B & 1 \end{bmatrix} \). In this case every eigenvalue \( \alpha \) of \( A \) has the form \( \alpha = 1 \pm \beta \) where \( \beta \) is either a singular value of \( B \) or, if \( B \) is not square, 0. And in this case every eigenvalue \( \varepsilon \) of \( E \) can be shown easily to satisfy 

\[
(\varepsilon + \delta) = \pm \delta(1+\delta)\sqrt{\varepsilon}; \quad \varepsilon = \left( (1+\delta)\sqrt{\varepsilon}/2 \pm \sqrt{(1+\delta)^2 - \delta} \right)^2.
\]

The largest of the magnitudes of eigenvalues \( \varepsilon \) is minimized when \( \delta = (\|B\|/(1 + \sqrt{(1-\|B\|^2)})^2 \), and then every eigenvalue \( \varepsilon \) has the same magnitude \( \delta \).