Numerical Methods

Recap: We've studied "stationary" iterative methods for solving $A\underline{x} = \underline{b}$,

$$\underline{x}^{k+1} = B\underline{x}^k + \underline{c}$$

meaning that B and $\underline{c}$ are indep. of k.

Examples:

- BSM : $B = I - A$
- Stat. Rich : $B = I - \alpha A$
- Jacobi : $B = I - D^{-1}A = -D^{-1}(L+U)$.
- Gauss-Seidel : $B = I - (L+D)^{-1}A = -(L+D)^{-1}U$.

preconditioned.

Basic theory: $\rho(B) < 1 \iff$ convergence $\forall\, \underline{x}^0$.

Today: non-stationary methods – allow B, $\underline{c}$ to change at each iteration.

+ Friday We'll study 2 methods: "gradient" and "conjugate gradient" methods.

Starting point: stat. Rich iteration: $\underline{x}^{k+1} = \underline{x}^k - \alpha A\underline{x}^k + \alpha\underline{b}$

Increment satisfies $\underline{x}^{k+1} - \underline{x}^k = \alpha(\underline{b} - A\underline{x}^k) = \alpha\underline{r}^k$

where $\underline{r}^k = \underline{b} - A\underline{x}^k$ is the residual.

↖ to get from $\underline{x}^k$ to $\underline{x}^{k+1}$ we move a distance $|\alpha|\|\underline{r}^k\|$ in direction $\dfrac{\underline{r}^k}{\|\underline{r}^k\|}$.

Let's consider a more general update:

$$\boxed{\underline{x}^{k+1} - \underline{x}^k = \alpha_k \underline{p}^k}$$

for some α_k and $\underline{p}^k$, both depending on k.

Check: residual updates by: $\boxed{\underline{r}^{k+1} - \underline{r}^k = -\alpha_k A\underline{p}^k}$.

Gradient method: take $\underline{p}^k = \underline{r}^k$ (as in stat. Rich.)

and choose α_k so as to minimise $\|\underline{e}^{k+1}\|_A$.

↑ A-norm of error at step $k+1$.

(We're assuming henceforth that A is SPD)

(recall: $\|\underline{x}\|_A = \sqrt{\underline{x}^T A\underline{x}} = \sqrt{(A\underline{x}, \underline{x})}$)

We can calculate α_k exactly:

$$\|\underline{e}^{k+1}\|_A^2 = (A\underline{e}^{k+1}, \underline{e}^{k+1})$$

(Recall: $\underline{e}^k = \underline{x} - \underline{x}^k$)

Now use important fact: $\boxed{\underline{r}^k = A\underline{e}^k}$ ← check this!

Hence $\|\underline{e}^{k+1}\|_A^2 = (\underline{r}^{k+1}, A^{-1}\underline{r}^{k+1})$

$$= (\underline{r}^k - \alpha_k A\underline{r}^k, A^{-1}\underline{r}^k - \alpha_k \underline{r}^k)$$

(residual update formula with $\underline{p}^k = \underline{r}^k$)

$$= (\underline{r}^k, A^{-1}\underline{r}^k) - \alpha_k(\underline{r}^k, \underline{r}^k) - \alpha_k(A\underline{r}^k, A^{-1}\underline{r}^k) + \alpha_k^2(A\underline{r}^k, \underline{r}^k)$$

$$= (A\underline{e}^k, \underline{e}^k) - 2\alpha_k \|\underline{r}^k\|_2^2 + \alpha_k^2 \|\underline{r}^k\|_A^2$$

$$(A\underline{r}^k)^T A^{-1}\underline{r}^k = (\underline{r}^k)^T A^T A^{-1}\underline{r}^k = (\underline{r}^k)^T A A^{-1}\underline{r}^k = (\underline{r}^k)^T \underline{r}^k.$$

So
$$\boxed{\|\underline{e}^{k+1}\|_A^2 = \|\underline{e}^k\|_A^2 - 2\alpha_k\|\underline{r}^k\|_2^2 + \alpha_k^2\|\underline{r}^k\|_A^2}$$

This is a quadratic function of α_k, and minimising it gives the optimal step length as

$$\boxed{\alpha_k = \frac{\|\underline{r}^k\|_2^2}{\|\underline{r}^k\|_A^2}}.$$

To summarise, gradient method is:

- choose $\underline{x}^0 \in \mathbb{R}^n$.
- set $\underline{r}^0 = \underline{b} - A\underline{x}^0$
- until convergence, iterate:
  - $\alpha_k = \dfrac{\|\underline{r}^k\|_2^2}{\|\underline{r}^k\|_A^2}$
  - $\underline{x}^{k+1} = \underline{x}^k + \alpha_k \underline{r}^k$
  - $\underline{r}^{k+1} = \underline{r}^k - \alpha_k A\underline{r}^k$.

(until a stopping criterion is reached, e.g. residual-based)

Theorem: If A is SPD then the gradient method converges linearly w.r.t. $\|\cdot\|_A$, with

$$\|\underline{e}^{k+1}\|_A \leq \left(\frac{\kappa(A)-1}{\kappa(A)+1}\right)\|\underline{e}^k\|_A, \quad k = 0, 1, 2, \ldots$$

Same constant as for optimal stat. Richardson!

Difference here is that we don't need to know $\lambda_{min}, \lambda_{max}$.

Proof: Given $\underline{x}^k$, let $\underline{x}_*^{k+1}$ be the result of applying one iteration of stationary Richardson, with optimal $\alpha = \alpha_* = \frac{2}{\lambda_{max} + \lambda_{min}}$.

Then defining $\underline{e}_*^{k+1} = \underline{x} - \underline{x}_*^{k+1}$ we have

$$\|\underline{e}_*^{k+1}\|_A \leq \underbrace{\left(\frac{\kappa(A)-1}{\kappa(A)+1}\right)}_{\rho(B_{\alpha_*})}\|\underline{e}^k\|_A$$

But by def'n of gradient method,

$$\|\underline{e}^{k+1}\|_A \leq \|\underline{e}_*^{k+1}\|_A. \quad \blacksquare$$

Next time: conjugate gradient!