A che serve il clustering attorno a 1

Let A be a p.d. matrix and ε, $0 < \varepsilon < 1$, be fixed.

Denote by λ_j^ε the eigenvalues of A outside the interval $[1-\varepsilon, 1+\varepsilon]$ and by r_ε the number of such eigenvalues. Set $S = [1-\varepsilon, 1+\varepsilon] \cup \{\lambda_j^\varepsilon\}$ and let p_q be the polynomial

$$p_q(\lambda) = \prod_{\lambda_j^\varepsilon} \left(1 - \frac{\lambda}{\lambda_j^\varepsilon}\right) \frac{T_{q-r_\varepsilon}((1-\lambda)/\varepsilon)}{T_{q-r_\varepsilon}(1/\varepsilon)}, \ q \geq r_\varepsilon$$

where $T_k(x)$ denotes the chebycev polynomial of degree k. $((b+a-2\lambda)/(b-a) = (1-\lambda)/\varepsilon$, $(b+a)/(b-a) = 1/\varepsilon$, if $a = 1-\varepsilon$, $b = 1+\varepsilon$). Notice that S is a set containing all the eigenvalues of A, and p_q has exactly degree q and $p_q(0) = 1$. Then one can say that if $\mathbf{x}_q$ is the q-th vector generated by the CG method when solving $A\mathbf{x} = \mathbf{b}$, then

$$\|\mathbf{x} - \mathbf{x}_q\|_A \leq (\max_{\lambda \in S} |p_q(\lambda)|)\|\mathbf{x} - \mathbf{x}_0\|_A. \qquad (bound)$$

This bound for $\|\mathbf{x} - \mathbf{x}_q\|_A$ allows a better evaluation of the CG rate of convergence with respect to the well known bound

$$\|\mathbf{x} - \mathbf{x}_q\|_A \leq 2 \left(\frac{\sqrt{\mu_2(A)} - 1}{\sqrt{\mu_2(A)} + 1}\right)^q \|\mathbf{x} - \mathbf{x}_0\|_A, \ \ \mu_2(A) = \frac{\max \lambda(A)}{\min \lambda(A)} \qquad (wkbound)$$

in case it is known that most of (almost all) the eigenvalues of A are in some interval $[1-\varepsilon, 1+\varepsilon]$ where ε is small (almost zero).

If, moreover, the $n \times n$ linear system $A\mathbf{x} = \mathbf{b}$ can be seen as one of a sequence of increasing order linear systems, with the property that $\forall\, \varepsilon > 0 \ \exists\, k_\varepsilon$, n_ε such that for all $n > n_\varepsilon$ outside $[1-\varepsilon, 1+\varepsilon]$ fall no more than n_ε eigenvalues of A, then (bound) allows to prove the superlinear convergence of CG.

(Note that in general CG has a linear rate of convergence, as a consequence of (wkbound)).

Let us prove these assertions, by evaluating $\max_{\lambda \in S} |p_q(\lambda)|$.

$$\begin{aligned} \max_{\lambda \in S} |p_q(\lambda)| &= \max_{\lambda \in [1-\varepsilon, 1+\varepsilon]} |p_q(\lambda)| \\ &\leq (\max_{\dots} \textstyle\prod_{\lambda_j^\varepsilon} \left|1 - \frac{\lambda}{\lambda_j^\varepsilon}\right|)(\max_{\dots} \left|\frac{T_{q-r_\varepsilon}((1-\lambda)/\varepsilon)}{T_{q-r_\varepsilon}(1/\varepsilon)}\right|) \\ &= (\max_{\dots} \textstyle\prod_{\lambda_j^\varepsilon} \left|1 - \frac{\lambda}{\lambda_j^\varepsilon}\right|)\frac{1}{T_{q-r_\varepsilon}(1/\varepsilon)}. \end{aligned}$$

Now first notice that

$$T_{q-r_\varepsilon}\left(\frac{1}{\varepsilon}\right) = T_{q-r_\varepsilon}\left(\frac{\frac{1+\varepsilon}{1-\varepsilon} + 1}{\frac{1+\varepsilon}{1-\varepsilon} - 1}\right) > \frac{1}{2}\left(\frac{\sqrt{\frac{1+\varepsilon}{1-\varepsilon}} + 1}{\sqrt{\frac{1+\varepsilon}{1-\varepsilon}} - 1}\right)^{q-r_\varepsilon}.$$

Then denote by $\hat\lambda_j^\varepsilon$ those eigenvalues λ_j^ε satisfying the inequalities

$$\lambda_j^\varepsilon < 1-\varepsilon, \quad \lambda_j^\varepsilon < \frac{1}{2}(1+\varepsilon)$$

and observe that

$$\begin{aligned} \max_{\lambda \in [1-\varepsilon, 1+\varepsilon]} \textstyle\prod_{\lambda_j^\varepsilon} \left|1 - \frac{\lambda}{\lambda_j^\varepsilon}\right| &\leq \max_{\dots} \textstyle\prod_{\hat\lambda_j^\varepsilon} \left|1 - \frac{\lambda}{\lambda_j^\varepsilon}\right| \\ &= \textstyle\prod_{\hat\lambda_j^\varepsilon} \left(\frac{1+\varepsilon}{\hat\lambda_j^\varepsilon} - 1\right). \end{aligned}$$

So, we have

$$\begin{aligned}\max_{\lambda\in S}|p_q(\lambda)| &\leq \prod_{\hat\lambda_j^\varepsilon}\left(\tfrac{1+\varepsilon}{\hat\lambda_j^\varepsilon}-1\right)2\left(\frac{\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}-1}{\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}+1}\right)^{q-r_\varepsilon}\\ &\leq 2\left(\tfrac{1+\varepsilon}{\min\lambda(A)}-1\right)^{\#\hat\lambda_j^\varepsilon}\left(\frac{\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}-1}{\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}+1}\right)^{q-r_\varepsilon}\\ &\approx \left(\tfrac{1+\varepsilon}{\min\lambda(A)}-1\right)^{\#\hat\lambda_j^\varepsilon}\frac{\varepsilon^q}{\varepsilon^{r_\varepsilon}2^{q-r_\varepsilon-1}},\end{aligned}$$

where in the latter approximation we have used the following Taylor expansion

$$f(\varepsilon)=\frac{\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}-1}{\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}+1}=\frac{\varepsilon}{2}+\frac{\varepsilon^2}{2}f''(0)+\dots.$$

Dubbio su GStrang

If $t^2<1$ then $A\mathbf{x}_i=\lambda_i\mathcal{A}\mathbf{x}_i$, $\lambda_i>0$, for n linearly independent vectors $\mathbf{x}_i$.

Thus, if X denotes the matrix whose columns are the $\mathbf{x}_i$ then $AX=\mathcal{A}XD$, $D=\operatorname{diag}(\lambda_i)\Rightarrow\det(\mathcal{A})\neq 0\Rightarrow\mathcal{A}^{-1}A\mathbf{x}_i=\lambda_i\mathbf{x}_i\Rightarrow$ i λ_i sono gli autovalori di $\mathcal{A}^{-1}A$.

Since $(\mathcal{A}^{-1}A)^{-1}=A^{-1}\mathcal{A}=L^{-T}L^{-1}\mathcal{A}=L^{-T}(L^{-1}\mathcal{A}L^{-T})L^T$, then the eigenvalues of $L^{-1}\mathcal{A}L^{-T}$ are $1/\lambda_i>0$.

$\Rightarrow L^{-1}\mathcal{A}L^{-T}$ is p.d. $\Rightarrow\mathcal{A}$ is p.d. $\Rightarrow\mathcal{A}^{-1}A$ and $E^{-1}AE^{-T}$, $EE^T=\mathcal{A}$, have the same eigenvalues.

But, why the $\mathbf{x}_i$ are independent?

Sui metodi iterativi (H, $\mathbf{u}_k$)

$\mathbf{r}_k=\mathbf{b}-A\mathbf{x}_k$

$H=A^TA$: $\mathbf{x}_{k+1}=\mathbf{x}_k+\frac{\mathbf{r}_k^TA\mathbf{u}_k}{\|A\mathbf{u}_k\|_2^2}\mathbf{u}_k$

becomes Richardson for $\mathbf{u}_k=\mathbf{r}_k$, and a method which converges in one step if $\mathbf{u}_k=A^{-1}\mathbf{r}_k=\mathbf{x}-\mathbf{x}_k$. Choose $\mathbf{u}_k=\mathcal{L}_A^{-1}\mathbf{r}_k$?

For the method

$$\mathbf{x}_{k+1}=\mathbf{x}_k+\frac{(\mathbf{x}-\mathbf{x}_k)^TH\mathbf{u}}{\mathbf{u}^TH\mathbf{u}}\mathbf{u}$$

we know that $|(\mathbf{x}-\mathbf{x}_k)^TH\mathbf{u}|/\sqrt{\mathbf{u}^TH\mathbf{u}}\to 0$, or, in case u are vectors of the canonical basis, $|(\mathbf{x}-\mathbf{x}_k)^TH\mathbf{u}|\to 0$. So, $\mathbf{x}-\mathbf{x}_k$ tends to be orthogonal to $\mathbf{u}$.

Assume $s_k\in\{1,2,\dots,n\}$ are chosen so that $\forall\, i$ we have $s_k=i$ for an infinite number of k. Set $\mathbf{u}=\mathbf{e}_{s_k}$. Then $\mathbf{x}-\mathbf{x}_k$ tends to be orthogonal to all vectors of the canonical basis, and thus tends to be the null vector. It follows that Gauss-Seidel an Southwell are not the only laws leading to the convergence of the scheme with $\mathbf{u}=\mathbf{e}_{s_k}$. Some of these remarks are not correct, why ?

'$\Re(A)>0$' vs 'p.d. of hermitian part of A'

Let A be a $n\times n$ matrix with complex entries.

Set $A_h=\frac{1}{2}(A+A^*))$ (hermitian part of A) and $A_{ah}=\frac{1}{2}(A-A^*)$ (anti-hermitian part of A). Obvioulsly $A=A_h+A_{ah}$.

If $\mathbf{z}^*A_h\mathbf{z}>0$ for all non null complex vectors $\mathbf{z}$, or, equivalently, if A_h is p.d., then $\Re(\lambda(A))>0$.

There exists $\mathbf{x} \neq \mathbf{0}$ such that $\lambda(A) = \mathbf{x}^* A\mathbf{x}$. Thus

$$\Re(\lambda(A)) = \Re(\mathbf{x}^* A_h \mathbf{x} + \mathbf{x}^* A_{ah}\mathbf{x}) = \mathbf{x}^* A_h \mathbf{x} > 0$$

($\overline{\mathbf{x}^* A_{ah}\mathbf{x}} = (\mathbf{x}^* A_{ah}\mathbf{x})^* = \mathbf{x}^* A_{ah}^* \mathbf{x} = -\mathbf{x}^* A_{ah}\mathbf{x}$, so $\mathbf{x}^* A_{ah}\mathbf{x}$ is is a complex number with null real part).

If $\Re(\lambda(A)) > 0$ and A is normal then $\mathbf{z}^* A_h \mathbf{z} > 0$, $\forall\ \mathbf{z} \neq \mathbf{0}$.

Let $\mathbf{x}_k$ be eigenvectors of A ($A\mathbf{x}_k = \lambda_k \mathbf{x}_k$) such that $\mathbf{x}_i^* \mathbf{x}_j$ is 1 if $i = j$ and 0 otherwise. Then

$$\begin{aligned} \mathbf{z}^* A_h \mathbf{z} &= \Re(\mathbf{z}^* A\mathbf{z}) = \Re((\sum \alpha_i \mathbf{x}_i)^* A(\sum \alpha_j \mathbf{x}_j)) \\ &= \Re(\sum |\alpha_k|^2 \lambda_k) = \sum |\alpha_k|^2 \Re(\lambda_k) > 0, \quad \mathbf{z} \neq \mathbf{0}. \end{aligned}$$

The hypothesis $\Re(\lambda(A)) > 0$ alone is not sufficient to assure that $\mathbf{z}^* A_h \mathbf{z} > 0$, $\forall\ \mathbf{z} \neq \mathbf{0}$. For instance, if

$$A = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}, \ a \in \mathbb{R},\ a^2 - 4 \geq 0,$$

then

$$[x\ y]A \begin{bmatrix} x \\ y \end{bmatrix} = [x\ xa + y] \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + axy + y^2$$

is not positive for all $x, y \in \mathbb{R}$, $[x\ y] \neq [0\ 0]$. (We have used the fact that if $\mathbf{z} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$ then $\mathbf{z}^* A_h \mathbf{z} = \mathbf{z}^* A\mathbf{z}$).

On the rate of convergence of some iterative methods in cases where the coefficient matrix has a p.d. hermitian part

For some linear systems iterative solvers (f.i. GMRES), it is known that the residual at step k, $\mathbf{r}_k = \mathbf{b} - A\mathbf{x}_k$, satisfies the inequality

$$\|\mathbf{r}_k\|_2 \leq \left(1 - \frac{\min \lambda(A_h)^2}{\max \lambda(A^* A)}\right)^{\frac{k}{2}} \|\mathbf{r}_0\|_2 \qquad (1)$$

(I'm not sure if the norm is the 2-norm) provided that the hermitian part A_h of A is p.d. (This assertion is surely true in case of systems $A\mathbf{x} = \mathbf{b}$ with A and $\mathbf{b}$ real, see [Saad book]).

Remark. If A is also hermitian or, equivalently, if A is p.d., then the number in the parentheses becomes $1 - 1/\mu_2(A)^2$.

Let us observe that

$$A_h\ p.d. \Rightarrow\ 0 < \frac{\min \lambda(A_h)^2}{\max \lambda(A^* A)} \leq\ 1$$

where the equality (on the right) holds iff $A = \alpha I$, $\alpha > 0$.

Lemma. *If A_h is p.d. then $\Re(\lambda(A)) \geq \min \lambda(A_h)$.*

Proof. There exists a vector $\mathbf{x}$, $\|\mathbf{x}\|_2 = 1$, such that $\Re(\lambda(A)) = \mathbf{x}^* A_h \mathbf{x}$. The thesis follows from the minmax eigenvalues representation theory for hermitian matrices.

By the Lemma, we have

$$\begin{aligned} \min \lambda(A_h) &\leq \Re(\lambda(A)) \leq \sqrt{\Re(\lambda(A))^2 + \Im(\lambda(A))^2} \\ &= |\lambda(A)| \leq \rho(A) \leq \|A\|_2 = \sqrt{\max \lambda(A^* A)}. \end{aligned}$$

Note, more precisely, that the equality $\Re(\lambda(A)) = \rho(A)$ holds for all $\lambda(A)$ iff the eigenvalues $\lambda(A)$ are all equal to an α, $\alpha > 0$, iff there exists T unitary such that

$$T^{-1}AT = \begin{bmatrix} \alpha & * & * \\ & \ddots & * \\ & & \alpha \end{bmatrix}.$$

If at least one of the entries $*$ in the upper triangular part of this matrix is nonzero, then $\|A\|_2 > \alpha = \rho(A)$. It follows that $\min \lambda(A_h) < \sqrt{\max(\lambda(A^*A))}$ unless $A = \alpha I$, $\alpha > 0$.

Let A be a non singular real $n \times n$ matrix. We know that the condition $\Re(\lambda(A)) > 0$ (which is verified f.i. if the hermitian part of A is p.d.) is necessary and sufficient for the existence of $\hat{\omega} > 0$ such that $\forall\, \omega \in (0, \hat{\omega})$ the Richardson iterative scheme

$$\mathbf{x}_0 \in \mathbb{R}^n, \quad \mathbf{x}_{k+1} = \mathbf{x}_k + \omega(\mathbf{b} - A\mathbf{x}_k), \quad k = 0, \dots \qquad (R(\omega))$$

converges to the solution $\mathbf{x} = A^{-1}\mathbf{b}$ of the linear system $A\mathbf{z} = \mathbf{b}$.

Let ω_{ott} be a (it may be not unique) value of ω for which the rate of convergence is maxima, i.e. $\rho(I - \omega_{ott}A) = \min \rho(I - \omega A)$.

Now we prove the following assertion:

If A is real, normal and its eigenvalues are of the form $\alpha + \mathbf{i}\beta_j$, $j = 1, \dots, n$, $\alpha > 0$, β_j zero for at least one j (these hypothesis imply that the hermitian part of A is p.d.), then the error at step k of $R(\omega_{ott})$ satisfies the inequality

$$\|\mathbf{x} - \mathbf{x}_k\|_2 \le \left(1 - \frac{1}{\mu_2(A)^2}\right)^{\frac{k}{2}} \|\mathbf{x} - \mathbf{x}_0\|_2 \qquad (2)$$

where $\mu_2(A) = \sqrt{\alpha^2 + \max \beta_j^2}/\alpha$ is the condition number of A (compare with (1), where A is p.d.).

So, $\mu_2(A)$ small is a sufficient condition to have fast convergence also in cases where A is not p.d..

Example. $A = B + \alpha I$, B real, $B^* = -B$, $\det(B) = 0$, $\alpha > 0$. For instance

$$B = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}.$$

Proof. First recall that if M is a normal matrix then $\rho(M) = \|M\|_2$.

Observe that $\rho(I - \omega A)^2 = \max_k g_k(\omega)$, $g_k(\omega) = \omega^2(\alpha^2 + \beta_k^2) - 2\omega\alpha + 1$. Since $g_k'(\omega) = 2\omega(\alpha^2 + \beta_k^2) - 2\alpha$, we have

$$g_k'\left(\frac{\alpha}{\alpha^2 + \beta_k^2}\right) = 0, \quad g_k'(0) = -2\alpha$$

and thus $g_k'(0)$ is constant with respect to k and negative. It follows that $R(\omega)$ converges iff $\omega \in (0, 2\alpha/(\alpha^2 + \max \beta_k^2))$. Moreover,

$$\omega_{ott} = \frac{\alpha}{\alpha^2 + \max \beta_k^2}, \; \rho(I - \omega_{ott}A)^2 = 1 - \frac{\alpha^2}{\alpha^2 + \max \beta_k^2} = 1 - \frac{1}{\mu_2(A)^2},$$

$$\|\mathbf{x} - \mathbf{x}_k\|_2 \le \|I - \omega_{ott}A\|_2^k \|\mathbf{x} - \mathbf{x}_0\|_2 = \rho(I - \omega_{ott}A)^k \|\mathbf{x} - \mathbf{x}_0\|_2$$

which imply the thesis.