Conjugate gradient method applied to clustered eigeinvalue matrices

Mele Giampaolo

October 11, 2012

Abstract

This document is the dissertation for the seminars of the professors Bertaccini and Filippone during the Rome Moscow school. The aim is to present the method of the conjugate gradient, to show some properties of that and to prove some results about convergence. At the end it will be given an idea about the fact that the convergence is too fast with clustered eigenvalues.

All the vectors are column vectors in $\mathbb{R}^N$ and all the matrices are in $\mathbb{R}^{N\times N}$

1 Derivation of algorithm

We can describe the conjugate gradient method in this way: suppose we want ro find the minimum of this function

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} + c$$

this means that we have to solve the linear system

$$\boldsymbol{H}\boldsymbol{x} = \boldsymbol{b}$$

We will call $\widehat{\boldsymbol{x}}$ the solution of this problem. Let we suppose H definite positive, we want to find the minimum using this kind of iterations

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \tau_k\boldsymbol{d}_k$$

Definition 1.1. $\boldsymbol{d}$ is called *descendent direction* for f at $\boldsymbol{x}$ if there exists τ_0 such that

$$f(\boldsymbol{x} + \tau\boldsymbol{d}) < f(\boldsymbol{x}) \qquad 0 < \tau \leq \tau_0$$

Theorem 1.1. If $f \in C^1$, let $\boldsymbol{g}(\boldsymbol{x})$ the gradient of f in $\boldsymbol{x}$, then if the vector $\boldsymbol{d}$ satisfies $\boldsymbol{g}^T(\boldsymbol{x})\boldsymbol{d} < 0$ then is a descendent direction for f at $\boldsymbol{x}$

Theorem 1.2. If $f \in C^2$, $\boldsymbol{g}^T(\boldsymbol{x})\boldsymbol{d} = 0$ and $\boldsymbol{d}^T\boldsymbol{H}\boldsymbol{d} < 0$ then $\boldsymbol{d}$ is a descendent direction for f at $\boldsymbol{x}$

The proof of those theorems is very trivial, just a short computation.

Theorem 1.3. Let $f \in C^1$, then among all directions $\boldsymbol{d}$ at some point $\boldsymbol{x}$, that direction in which f descends most rapidly in a neighborhood of $\boldsymbol{x}$ is $\boldsymbol{d} = -\boldsymbol{g}(\boldsymbol{x})$.

If we choose $\boldsymbol{d}^k = \boldsymbol{g}(\boldsymbol{x}^k)$ then we will obtain *the method of stepest descent*, to obtain the conjugate gradient method we have to do a sligthly different choice. We will see. So, from now we will call $\boldsymbol{d}^k$ the *search directions* insted of the descendent directions.

Let suppose given $\boldsymbol{d}$, then we have to choose τ, but if $\boldsymbol{H}$ is definite positive then $\boldsymbol{d}^T\boldsymbol{H}\boldsymbol{d} > 0$ and $f(\boldsymbol{x} + \tau\boldsymbol{d})$ is a parabola respect τ and it is minimized by

$$\tau = -\frac{\boldsymbol{d}^T\boldsymbol{g}(\boldsymbol{x})}{\boldsymbol{d}^T\boldsymbol{H}\boldsymbol{d}}$$

So if we know $\boldsymbol{d}_k$ then we put

$$\tau_k = -\frac{\boldsymbol{d}_k^T\boldsymbol{g}(\boldsymbol{x}^k)}{\boldsymbol{d}_k^T\boldsymbol{H}\boldsymbol{d}_k}$$

Let we denote $\boldsymbol{g}^k := \boldsymbol{g}(\boldsymbol{x}^k)$,it is not difficult to see that

$$\boldsymbol{g}^k = \boldsymbol{H}\boldsymbol{x}^k - \boldsymbol{b}$$

It is important to underline that with this choice of τ_k we minimize $f(\boldsymbol{x}^k + \tau\boldsymbol{d}^k)$ but at the same time we make $\boldsymbol{g}^{k+1}$ (the gradient at $\boldsymbol{x}^{k+1}$) ortogonal to the search direction $\boldsymbol{d}^k$. To see this we observe that

$$\boldsymbol{g}^{k+1} = \boldsymbol{g}^k + \tau_k\boldsymbol{H}\boldsymbol{d}^k$$

and then we multiply for $\boldsymbol{d}^k$

$${\boldsymbol{d}^k}^T\boldsymbol{g}^{k+1} = (\boldsymbol{d}^k)^T\boldsymbol{g}^k + \tau_k(\boldsymbol{d}^k)^T\boldsymbol{H}\boldsymbol{d}^k = 0$$

To see this is sufficient to replace the definition of τ_k.

We will look for search directions of this form

$$\begin{cases} \boldsymbol{d}^{k+1} = -\boldsymbol{g}^{k+1} + \beta_k\boldsymbol{d}^k & k = 1, 2, \dots \\ \boldsymbol{d}^0 = \boldsymbol{g}^0 \end{cases}$$

We have to choose the coefficients β_k.

Observation 1.1. What happen if at same step we obtain a null search direction ($\boldsymbol{d}^k = \boldsymbol{0}$)? Let consider the formula who define $\boldsymbol{d}^{k+1}$ and let us replace k to $k+1$, we obtain $\boldsymbol{d}^{k+1} = -\boldsymbol{g}^{k+1} + \beta_k\boldsymbol{d}^k$, we can do the scalar product with $\boldsymbol{g}^k$ and we obtain

$${\boldsymbol{g}^k}^T\boldsymbol{d}^k = \|\boldsymbol{g}^k\|^2 + \beta_k{\boldsymbol{g}^k}^T\boldsymbol{d}^{k-1}$$

But we know that ${\boldsymbol{g}^k}^T\boldsymbol{d}^{k-1} = 0$, and so, if ${\boldsymbol{g}^k}^T\boldsymbol{d}^k = 0$ then $\|\boldsymbol{g}^k\|^2 = 0$, this means that we reached the point of minimum $\widehat{\boldsymbol{x}}$. With the same argoment we have that if $\boldsymbol{x}^k \neq \widehat{\boldsymbol{x}}$ then $\boldsymbol{d}^k \neq \boldsymbol{0}$.

To choose in the best way β_k we need some definitions.

Definition 1.2. Let now define the following inner product

$$(\boldsymbol{x}, \boldsymbol{y})_{\boldsymbol{H}^{-1}} := \boldsymbol{x}^T \boldsymbol{H}^{-1} \boldsymbol{y}$$

from this we have the norm

$$\|\boldsymbol{x}\|_{\boldsymbol{H}^{-1}} := (\boldsymbol{x}, \boldsymbol{x})_{\boldsymbol{H}^{-1}}^{1/2} = (\boldsymbol{x}^T \boldsymbol{H}^{-1} \boldsymbol{x})^{1/2}$$

We will prefer to work with $\boldsymbol{g}$ insted of $\boldsymbol{x}$, so it will be useful the following lemma.

Lemma 1.1. It holds that $\|\boldsymbol{g}\|_{\boldsymbol{H}^{-1}} = \|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_{\boldsymbol{H}}$

Proof. Let us start with this simple computation

$$\begin{aligned} \boldsymbol{g} &= \boldsymbol{H}\boldsymbol{x} - \boldsymbol{b} \\ &= \boldsymbol{H}\boldsymbol{x} - \boldsymbol{H}\widehat{\boldsymbol{x}} + \boldsymbol{H}\widehat{\boldsymbol{x}} - \boldsymbol{b} \\ &= \boldsymbol{H}\boldsymbol{x} - \boldsymbol{H}\widehat{\boldsymbol{x}} \\ &= \boldsymbol{H}(\boldsymbol{x} - \widehat{\boldsymbol{x}}) \end{aligned}$$

We used that $\boldsymbol{H}\widehat{\boldsymbol{x}} = \boldsymbol{b}$ And so

$$\begin{aligned} \|\boldsymbol{g}\|_{H^{-1}} &= \boldsymbol{g}^T \boldsymbol{H}^{-1} \boldsymbol{g} \\ &= (\boldsymbol{x} - \widehat{\boldsymbol{x}})^T \boldsymbol{H}^T \boldsymbol{H}^{-1} \boldsymbol{H} (\boldsymbol{x} - \widehat{\boldsymbol{x}}) \\ &= (\boldsymbol{x} - \widehat{\boldsymbol{x}})^T \boldsymbol{H} (\boldsymbol{x} - \widehat{\boldsymbol{x}}) \\ &= \|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_H \end{aligned}$$

□

So, looking at the previous lemma, in order to obtain the convergence our goal is to minimize $\|\boldsymbol{g}\|_{\boldsymbol{H}^{-1}}$.

Using the following (already discussed)

$$\begin{cases} \boldsymbol{g}^{k+1} &= \boldsymbol{g}^k + \tau_k \boldsymbol{H} \boldsymbol{d}^k \\ \boldsymbol{d}^{k+1} &= -\boldsymbol{g}^{k+1} + \beta_k \boldsymbol{d}^k \end{cases}$$

We obtain that for any choice of β_k the gradient have this form

$$\boldsymbol{g}^k = \boldsymbol{g}^0 + \sum_{l=1}^{k} \alpha_l^{(k)} \boldsymbol{H}^l \boldsymbol{g}^0$$

where

$$\alpha_k^{(k)} = (-1)^k \prod_{i=1}^{k-1} \tau_i \neq 0$$

Definition 1.3. Let we define

$$\begin{aligned} S_k &= \operatorname{span}\left\{\boldsymbol{H}\boldsymbol{g}^0, \boldsymbol{H}^2\boldsymbol{g}^0, \ldots, \boldsymbol{H}^k\boldsymbol{g}^0\right\} \\ T_k &= \left\{\boldsymbol{g} \in \mathbb{R}^N \text{ such that } \boldsymbol{g} = \boldsymbol{g}^0 + \boldsymbol{h} \text{ with } \boldsymbol{h} \in S_k\right\} \end{aligned}$$

Observation 1.2. S_k is a subspace of $\mathbb{R}^N$ with the dimension equal to the number of lineary indipendent vectors in the set $\left\{\boldsymbol{H}\boldsymbol{g}^0, \boldsymbol{H}^2\boldsymbol{g}^0, \ldots, \boldsymbol{H}^k\boldsymbol{g}^0\right\}$, T_k is a subset of $\mathbb{R}^N$ and not a subspace (it can be consider as affine subspace). Of course $\boldsymbol{g}^k \in T_k$

Theorem 1.4 (The conjugate gradient method)**.** If we impose the condition

$$\|\boldsymbol{g}^k\|_{\boldsymbol{H}^{-1}} = \min_{\boldsymbol{g}\in T_k} \|\boldsymbol{g}\|_{H^{-1}}$$

then

(i) the coefficients are

$$\beta_k = \frac{{\boldsymbol{g}^{k+1}}^T \boldsymbol{H}\boldsymbol{d}^k}{{\boldsymbol{d}^k}^T \boldsymbol{H}\boldsymbol{d}^k}$$

(ii) it holds that

$${\boldsymbol{g}^k}^T \boldsymbol{g}^l = 0 \quad \text{ if } \ l \neq k$$

(iii) it holds that

$${\boldsymbol{d}^k}^T \boldsymbol{H}\boldsymbol{d}^l = 0 \quad \text{ if } \ l \neq k$$

Proof. Let us start proving *(ii)*. It is clear that the condition

$$\|\boldsymbol{g}^k\|_{\boldsymbol{H}^{-1}} = \min_{\boldsymbol{g}\in T_k} \|\boldsymbol{g}\|_{H^{-1}}$$

it's equivalent to

$$\|\boldsymbol{g}^0 + \boldsymbol{h}^k\|_{\boldsymbol{H}^{-1}} = \min_{\boldsymbol{h}\in S_k} \|\boldsymbol{g}^0 + \boldsymbol{h}\|_{H^{-1}}$$

where $\boldsymbol{h}^k = \boldsymbol{g}^k - \boldsymbol{g}^0$.

The idea is to view any $\boldsymbol{h} \in S_k$ as an approximation of $-\boldsymbol{g}^0$ (it can be a terrible approximation, we don't care about that), in this case the error of this approximation is $\boldsymbol{h} - (\boldsymbol{g}^0) = \boldsymbol{h} + \boldsymbol{g}^0$. So now, with this formulation, we want to find in the subspace S_k the vector $\boldsymbol{h}^k$ that most closely approximates $-g^0$ with respect of the norm $\|\cdot\|_{\boldsymbol{H}^{-1}}$. Using the classical theorems of linear algebra we know that this vector exists and is unique, moreover the error $\boldsymbol{g}^0 + \boldsymbol{h}^0$ is ortogonal to all the space S_k, so we have

$$(\boldsymbol{g}^0 + \boldsymbol{h}^k)\boldsymbol{H}^{-1}\boldsymbol{h} = 0 \qquad \forall \boldsymbol{h} \in S_k$$

And so we obtain that

$${\boldsymbol{g}^k}^T \boldsymbol{H}^{-1}\boldsymbol{h} = 0 \qquad \forall \boldsymbol{h} \in S_k$$

So, for any $\boldsymbol{g} \in T_{k-1}$ the vector $\boldsymbol{h} = \boldsymbol{H}\boldsymbol{g}$ belongs to S_k, then we have

$${\boldsymbol{g}^k}^T \boldsymbol{g} = 0 \qquad \forall \boldsymbol{g} \in T_{k-1}$$

For every $l < k$ we have $\boldsymbol{g}^l \in T_l \subseteq T_{k-1}$, so we have

$${\boldsymbol{g}^k}^T \boldsymbol{g}^l = 0 \qquad \forall k < l$$

Let now prove (iii), let assume $l < k$. Using the following relations (already discussed)

$$\begin{cases} \boldsymbol{g}^{k+1} &= \boldsymbol{g}^k + \tau_k \boldsymbol{H}\boldsymbol{d}^k \\ \boldsymbol{d}^{k+1} &= -\boldsymbol{g}^{k+1} + \beta_k \boldsymbol{d}^k \end{cases}$$

and the point (ii), we obtain

$$\begin{aligned}
\boldsymbol{d}^k \boldsymbol{H} \boldsymbol{d}^l &= (\boldsymbol{H}\boldsymbol{d}^k)^T \boldsymbol{d}^l \\
&= \tau_k^{-1} (\boldsymbol{g}^{k+1} - \boldsymbol{g}^k)^T \boldsymbol{d}^l \\
&= \tau_k^{-1} (\boldsymbol{g}^{k+1} - \boldsymbol{g}^k)^T (-\boldsymbol{g}^l + \beta_{l-1} \boldsymbol{d}^{l-1}) \\
&= (\beta_{l-1}/\tau_k)(\boldsymbol{g}^{k+1} - \boldsymbol{g}^k)^T \boldsymbol{d}^{l-1}
\end{aligned}$$

By induction we can prove that

$$\begin{aligned}
\boldsymbol{d}^k \boldsymbol{H} \boldsymbol{d}^l &= \tau_k^{-1} \left(\prod_{i=0}^{l-1} \beta_i \right) (\boldsymbol{g}^{k+1} - \boldsymbol{g}^k)^T \boldsymbol{d}^0 \\
&= 0
\end{aligned}$$

Where we used that $\boldsymbol{d}^0 = -\boldsymbol{g}^0$ and the the ortogonality of the gradients.

Finally we can prove (i), it follows easily from (ii), infact

$$0 = \boldsymbol{d}^{k+1} \boldsymbol{H} \boldsymbol{d}^k = (-\boldsymbol{g}^{k+1} + \beta_k \boldsymbol{d}^k)^T \boldsymbol{H} \boldsymbol{d}^k$$

□

Finally we have the conjugate gradient, we can summarize it

$$\boxed{\begin{aligned}
\boldsymbol{x}^{k+1} &= \boldsymbol{x}^k + \tau_k \boldsymbol{d}_k \\
\tau_k &= -\frac{\boldsymbol{d}_k^T \boldsymbol{g}^k}{\boldsymbol{d}_k^T \boldsymbol{H} \boldsymbol{d}_k} \\
\beta_k &= \frac{\boldsymbol{g}^{k+1^T} \boldsymbol{H} \boldsymbol{d}^k}{\boldsymbol{d}^{k^T} \boldsymbol{H} \boldsymbol{d}^k}
\end{aligned}}$$

2 Convergence analysis

Theorem 2.1 (Finite termination of conjugate gradient method)**.** The conjugate gradient method convergence in $m \leq N$ steps, we mean that $\boldsymbol{x}^m = \widehat{\boldsymbol{x}}$.

Proof. We proved that the gradients $\boldsymbol{g}^k$ are ortogonal. So if the method do not convergence in N steps it means that we have $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^N, \boldsymbol{g}^{N+1}$ with $\boldsymbol{g}^k \neq 0$ vectos mutually ortogonal and this is impossible. □

Definition 2.1. Let Π_k the set of the polynomials p_k of degree k such that $p_k(0) = 1$. Let define the set

$$\tilde{T}_k = \left\{ \boldsymbol{g} \in \mathbb{R}^N \text{ s.t. } \boldsymbol{g} = p_k(\boldsymbol{H})\boldsymbol{g}^0 \ , \ p_k \in \Pi_k \right\}$$

Observation 2.1. It holds that $\tilde{T}_k \subseteq T_k$ and $\boldsymbol{g}^k \in \tilde{T}_k$ infact we showed that

$$\boldsymbol{g}^k = \boldsymbol{g}^0 + \sum_{l=1}^{k} \alpha_l^{(k)} \boldsymbol{H}^l \boldsymbol{g}^0 \qquad \text{with } \alpha_k^{(k)} \neq 0$$

Then we have

$$\begin{aligned}
\|\boldsymbol{g}^k\|_{\boldsymbol{H}^{-1}} &= \min_{\boldsymbol{g}\in\tilde{T}_k} \|\boldsymbol{g}\|_{\boldsymbol{H}^{-1}} \\
&= \min_{p_k\in\Pi_k} \|p_k(\boldsymbol{H})\boldsymbol{g}^0\|_{\boldsymbol{H}^{-1}} \\
&= \min_{p_k\in\Pi_k} \left[{\boldsymbol{g}^0}^T \boldsymbol{H}^{-1} p_k(\boldsymbol{H})^2 \boldsymbol{g}^0\right]^{1/2}
\end{aligned}$$

Theorem 2.2. Let $S \subset \mathbb{R}$ such that S contains all the eigenvalues of $\boldsymbol{H}$ and suppose that for some $M \geq 0$ and for same $p_k \in \Pi_k$ il holds that

$$\max_{\lambda\in S} |p_k(\lambda)| \leq M$$

then

$$\|\boldsymbol{x}^k - \widehat{\boldsymbol{x}}\|_{\boldsymbol{H}} \leq M\|\boldsymbol{x}^0 - \widehat{\boldsymbol{x}}\|_{\boldsymbol{H}}$$

Proof. Let $\{\lambda_i, \boldsymbol{v}_i\}_{i=1}^N$ the eigenvalues and the eigenvectors of $\boldsymbol{H}$ with the ordering $0 < \lambda_1 \leq \cdots \leq \lambda_N$, we can also take this vectors ortogonal, so $\boldsymbol{v}_i^T \boldsymbol{v}_j = \delta_{i,j}$. The initial gradient has the expansion

$$\begin{cases} \boldsymbol{g}^0 &= \sum_{i=1}^N a_i \boldsymbol{v}_i \\ a_i &= \boldsymbol{v}^T \boldsymbol{g}^0 \end{cases}$$

then, doing the computation we obtain

$${\boldsymbol{g}^0}^T \boldsymbol{H}^{-1} p_k(\boldsymbol{H})^2 \boldsymbol{g}^0 = \sum_{i=1}^N a_i^2 \lambda_i^{-1} p_k(\lambda_i)^2$$

using the previous observation we obtain

$$\|\boldsymbol{g}^k\|^2_{\boldsymbol{H}^{-1}} = \min_{p_k\in\Pi_k} \sum_{i=1}^N a_i^2 \lambda_i^{-1} p_k(\lambda_i)^2$$

From the hypothesis of the theorem we have $|p_k(\lambda_i)| \leq M$, then

$$\begin{aligned}
\|\boldsymbol{g}^k\|^2_{\boldsymbol{H}^{-1}} &\leq M^2 \sum_{i=1}^N a_i^2 \lambda_i^{-1} \\
&= M^2 \|\boldsymbol{g}^0\|_{\boldsymbol{H}^{-1}}
\end{aligned}$$

We already proved (Lemma 1.1) that $\|\boldsymbol{x}^k - \widehat{\boldsymbol{x}}\|_{\boldsymbol{H}} = \|\boldsymbol{g}^k\|_{\boldsymbol{H}^{-1}}$ and from this we have the thesis. □

Observation 2.2. If we don't know the eigenvalues distribuition it's natural to choose $S = [\lambda_1, \lambda_N]$ and to look for a polynomial $\tilde{p}_k \in \Pi_k$ such that

$$\max_{\lambda_1 \leq \lambda_k \leq \lambda_N} |\tilde{p}_k(\lambda_k)| = \min_{p_k \in \Pi_k} \max_{\lambda_1 \leq \lambda_k \leq \lambda_N} |p_k(\lambda_k)|$$

The solution of this problem is known, we have to choose

$$\tilde{p}_k(\lambda) = \frac{T_k[(\lambda_N + \lambda_1 - 2\lambda)/(\lambda_N - \lambda_1)]}{T_k[(\lambda_N + \lambda_1)(\lambda_N - \lambda_1)]}$$

where T_k is the Chebyshev polynomial of degree k. Moreover

$$\max_{\lambda_1 \leq \lambda \leq \lambda_N} |\tilde{p}_k(\lambda)| = T_k[(\lambda_N + \lambda_1)/(\lambda_N - \lambda_1)]^{-1}$$

Theorem 2.3 (speed of convergence)**.** From the previous observation we have that

$$\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_{\boldsymbol{H}} \leq T_k[(\lambda_N + \lambda_1)/(\lambda_N - \lambda_1)]^{-1} \|\boldsymbol{x}^0 - \widehat{\boldsymbol{x}}\|$$

Moreover, fixed $\varepsilon > 0$, if k is the smallest integer such that

$$\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_{\boldsymbol{H}}\| \leq \varepsilon \|\boldsymbol{x}^0 - \widehat{\boldsymbol{x}}\| \quad \forall \boldsymbol{x}^0 \in \mathbb{R}^N$$

then

$$\boxed{k \leq \frac{1}{2}\sqrt{K(\boldsymbol{H})}\ln(2/\varepsilon) + 1}$$

where $K(\boldsymbol{H})$ is the condition number.

Observation 2.3. This bound is very important for the theory of the convergence of conjugate gradient but sometimes it can be very bland. For example if $K(\boldsymbol{H}) = 10^4$ and $\varepsilon = 10^{-4}$ then $k \leq 496$, but our matrix can be $\boldsymbol{H} \in \mathbb{R}^{100 \times 100}$ so the algorithm give us the true solution after 100 steps. This is just an example, we will see how to get a better bound in same cases where we know something about eigenvalues distribuition.

3 Convergence with clustered eigenvalues

In the previous section we proved that if $\max_{\lambda \in S} |p_k(\lambda)| \leq M$ then

$$\|\boldsymbol{x}^k - \widehat{\boldsymbol{x}}\|_{\boldsymbol{H}} \leq M \|\boldsymbol{x}^0 - \widehat{\boldsymbol{x}}\|_{\boldsymbol{H}}$$

The idea is to select a set S containing the eigenvalues and to seek a polynomial $\tilde{p}_k \in \Pi_k$ such that $\max_{\lambda \in S} |\tilde{p}_k(\lambda)| \leq M$ is small. So we will do same assumptions on S.

3.1 Example 1

Let assume that

$$S = [\lambda_1, b] \cup [c, \lambda_N]$$

with

$$\begin{cases} 0 < \lambda_1 < b < c < \lambda_N \\ b - \lambda_1 = \lambda_N - c \\ 4b < \lambda_N \end{cases}$$

Thus S consists of two well-separated intervals of equal length.

For any $\gamma \neq 0$ let us consider the parabola

$$p_2(\lambda) = 1 - \gamma\lambda(\lambda_1 + \lambda_N - \lambda)$$

This polynomial is such that $p_2(\lambda_1) = p_2(\lambda_N)$ and $p_2(b) = p_2(c)$. If we require that $p_2(c) = -p_2(\lambda_N)$ we can find γ and we obtain the polynomial

$$\tilde{p}_k(\lambda) = 1 - 2[\lambda_N(c + \lambda_1) - c(c - \lambda_1)]^{-1}\lambda(\lambda_1 + \lambda_N - \lambda)$$

Let now

$$\tilde{p}_{2k}(\lambda) = \frac{T_k\{[\beta + \alpha - 2(1 - \tilde{p}_2(\lambda))]/(\beta - \alpha)\}}{T_k[(\beta + \alpha)/(\beta - \alpha)]} \qquad k = 1, 2, 3, \ldots$$

where $\alpha = 1 - \tilde{p}_2(\lambda_1)$ and $\beta = 1 - \tilde{p}_2(b)$, it can be proved that

$$\max_{\lambda \in S} |\tilde{p}_{2k}(\lambda)| = \min_{p_{2k} \in \Pi_{2k}} \max_{\lambda \in S} |p_{2k}(\lambda)|$$

Moreover we have

$$\max_{\lambda \in S} |\tilde{p}_{2k}(\lambda)| = T_k[(\beta + \alpha)/(\beta - \alpha)]^{-1} \qquad k = 1, 2, 3, \ldots$$

and then, using the previos theorem we have

$$\|\boldsymbol{x}^k - \widehat{\boldsymbol{x}}\|_{\boldsymbol{H}} \leq T_{k/2}[(\beta + \alpha)/(\beta - \alpha)]^{-1}\|\boldsymbol{x}^0 - \widehat{\boldsymbol{x}}\|_{\boldsymbol{H}} \qquad k = 2, 4, 6, \ldots$$

so in this case we obtain that the minimum integer k such that

$$\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_{\boldsymbol{H}}\| \leq \varepsilon\|\boldsymbol{x}^0 - \widehat{\boldsymbol{x}}\| \quad \forall \boldsymbol{x}^0 \in \mathbb{R}^N$$

is

$$k \leq \frac{1}{2}\sqrt{4\beta/\alpha}\ln(2/\varepsilon) + 1$$

Sameone call $4\beta/\alpha$ *the effective spectral condition number.* If $4bc/\lambda_N^2 < 1$ then this bound is better than the general bound.

3.2 Example 2

Let us consider $S = S_1 \cup S_2$ with

$$\begin{cases} S_1 = [\lambda_1, b] \\ S_2 = \displaystyle\bigcup_{n=N-m+1}^{N} \{\lambda_i\} \end{cases}$$

with $1 < m < N$ and $\lambda_{N-m} < b < \lambda_{N-m+1}$. Clearly, all of the eigenvalues are in S. We assume that m and b are small, so that S describes a distribution in which a few of the highest eigenvalues are well separated from the remaining eigenvalues. Let us consider following the polynomial in Π_k

$$\tilde{p}_k(\lambda) = \left[\prod_{i=N-m+1}^{N} \left(1 - \frac{\lambda}{\lambda_i}\right)\right] \frac{T_{k-m}[(b + \lambda_1 - 2\lambda)/(b - \lambda_1)]}{T_{k-m}[(b + \lambda_1)/(b - \lambda_1)]}$$

It is possible to prove that

$$\begin{aligned} \max_{\lambda \in S} |\tilde{p}_k(\lambda)| &= \max_{\lambda \in S_1} |\tilde{p}_k(\lambda)| \\ &\leq \frac{T_{k-m}[(b + \lambda_1 - 2\lambda)/(b - \lambda_1)]}{T_{k-m}[(b + \lambda_1)/(b - \lambda_1)]} \\ &= T_{k-m}\left(\frac{b + \lambda_1}{b - \lambda_1}\right) \end{aligned}$$

We can take this as value of M and so we obtain the bound

$$k \leq \frac{1}{2}\sqrt{2/\varepsilon} + m + 1$$

(As before, k is the minimum number as in the previous example).
If m and b are sufficiently small, then this bound is better then the general bound.

4 Conclusion

The foregoing examples illustrate the fact that the clustering of eigenvalues tends to increase the rate of convergence of the conjugate gradient method. This is related to the property, mentioned earlier, that the value of m in Theorem 2.1 never exceeds the number of distinct eigenvalues. Because of its generality and simplicity, the general bound is very useful and motivates our use of the spectral condition number $K(H)$ to assess the rate of convergence. It should be kept in mind, however, that depending on the distribution of the interior eigenvalues the general boud may be quite pessimistic, hiding the true rate of convergence.