Multigrid Methods for Structured Matrices

Marco Donatelli

Dept. of Science and High Technology – U. Insubria (Italy)

TUM 2016

Outline

Sequences of matrices

- In many applications we are interested in studying the spectral distribution of matrix sequences.
- Discretizing a PDE or an integral equation we obtain a linear system where the size depends on the discretization step.
- The structure of the matrix is fixed by the continuous operator.
- We would use a small discretization step reducing the approximation error.
- In conclusion, we study the matrix sequence

$$\{A_n\}, \qquad A_n \in \mathbb{C}^{d(n)\times d(n)},$$

where $d(n)$ is a function of the number of discretization points n.

Denote by $Circ(\mathbf{a})$ the circulant matrix defined by $\mathbf{a}$, e.g., $Circ(\mathbf{k}) = K_{circ}$, namely

$$Circ(\mathbf{a}) = \begin{bmatrix} a_0 & a_{n-1} & \dots & a_1 \\ a_1 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1} \\ a_{n-1} & \dots & a_1 & a_0 \end{bmatrix} = p_{n-1}(Z) \in \mathbb{P}_{n-1}, \qquad (1)$$

where

$$p_{n-1}(x) = \sum_{j=0}^{n-1} a_j x^j \qquad \text{and} \qquad Z := \left[\begin{array}{ccc|c} & & & 1 \\ \hline 1 & & & \\ & \ddots & & \\ & & 1 & \end{array}\right].$$

Spectral decomposition of Circulant matrices

- Define $\mathbf{y} \in \mathbb{R}^n$ by uniform sampling in $[0, 2\pi]$:

$$y_s = \frac{2\pi s}{n}, \qquad s = 0, \dots, n-1.$$

- The spectral decomposition of Z is

$$Z = F_n \Lambda F_n^{-1}, \qquad \Lambda = \mathrm{diag}(\mathrm{e}^{\mathrm{i}\mathbf{y}}) \tag{2}$$

- Combining (2) with (1), the spectral decomposition of $Circ(\mathbf{a})$ is

$$Circ(\mathbf{a}) = \frac{1}{n} F_n \mathrm{diag}(F_n^H \mathbf{a}) F_n^H \tag{3}$$

  1. the eigenvectors are the column of F_n, i.e, $\mathrm{e}^{-\mathrm{i}j\mathbf{y}}$ the j-th frequency.
  2. the eigenvalues of $Circ(\mathbf{a})$ are

$$\lambda_j = [F_n^H \mathbf{a}]_j = \sum_{s=0}^{n-1} a_s \mathrm{e}^{\mathrm{i}j y_s}, \qquad j = 0, \dots, n-1.$$

- Which is the difference between (3) and (7) in Convolution and Structured Matrices?

For our convolution matrix $K_{\mathrm{circ}} = Circ(\mathbf{k})$ it holds

$$\lambda_j = \sum_{s=0}^{n-1} k_s \mathrm{e}^{\frac{\mathrm{i}2\pi js}{n}} = \sum_{s=0}^{m} k_s \mathrm{e}^{\frac{\mathrm{i}2\pi js}{n}} + \sum_{s=-m}^{-1} k_s \mathrm{e}^{\frac{\mathrm{i}2\pi js}{n}}$$

$$= \sum_{s=-m}^{m} k_s \mathrm{e}^{\frac{\mathrm{i}2\pi js}{n}} = S_m[k](y_j), \qquad j = 0, \ldots, n-1.$$

which is the m-th partial sum of the Fourier series of the function k assuming that $k \in L^1_{[0,2\pi]}$ is 2π-periodic and its Fourier coefficients are

$$k_j = \frac{1}{2\pi} \int_0^{2\pi} k(x) \mathrm{e}^{-\mathrm{i}jx} \mathrm{d}x, \qquad j \in \mathbb{Z}, \qquad k(x) = \sum_{j \in \mathbb{Z}} k_j \mathrm{e}^{\mathrm{i}jx}.$$

Remark

We can construct a sequence of circulant matrices associated to k with increasing size $2m+1$ using $S_m[k](x)$.

The operator of Toeplitz matrices

Definition

Given a function $f : [0, 2\pi] \to \mathbb{C}$, 2π-periodic, $f \in L^1_{[0,2\pi]}$ and with Fourier coefficients

$$a_j = \frac{1}{2\pi}\int_0^{2\pi} f(x)\mathrm{e}^{-\mathrm{i}jx}\mathrm{d}x, \qquad j \in \mathbb{Z},$$

the associated Toeplitz matrix of order n is $T_n = T_n(f) = [a_{i-j}]_{i,j=0}^{n-1}$, namely

$$T_n = \begin{bmatrix} a_0 & a_{-1} & \dots & a_{-(n-1)} \\ a_1 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{-1} \\ a_{n-1} & \dots & a_1 & a_0 \end{bmatrix}$$

Example

$$\begin{cases} u''(x) = g(x) & x \in (0,1) \\ u(0) = u(1) = 0 \end{cases}$$

Finite differences discretization of order 2 $\Rightarrow f(x) = 2 - 2\cos(x)$

Properties of $T_n(f)$

Definition

$$T_n(\cdot) : L^1_{[0,2\pi]} \to \mathbb{C}^{n\times n}$$

Lemma

1. $T_n(\alpha f + \beta g) = \alpha T_n(f) + \beta T_n(g)$
2. *f real $\Longrightarrow T_n(f)$ is a Hermitian matrix,*
3. *$f \geq 0 \Longrightarrow T_n(f)$ is positive semidefinite,*
4. *$f \geq 0$ and $\operatorname{ess\,sup} f > 0 \Longrightarrow T_n(f)$ is positive definite,*
5. *$m_f \leq f \leq M_f$ with $m_f \neq M_f \Longrightarrow \sigma(T_n(f)) \subset (m_f, M_f)$, (spectrum).*

We can replace L^1 with continuous functions and $\operatorname{ess\,sup} f > 0$ with $\max(f) > 0$.

Theorem

Given $f \geq 0$, let α be the maximum order of the zeros, then $\mu_2(T_n(f)) = O(n^\alpha)$.

Definition

Let $f : [0, 2\pi] \to \mathbb{C}$ be $L^1_{[0,2\pi]}$. Let $\{A_n\}$ be a sequence of matrices of size n with eigenvalues $\lambda_j(A_n)$, $j = 1, \ldots, n$.
$\{A_n\}$ is distributed as the pair $(f, [0, 2\pi])$ in the sense of the eigenvalues:

$$\{A_n\} \sim_\lambda (f, [0, 2\pi]),$$

if for all continuous functions F

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} F(\lambda_j(A_n)) = \frac{1}{2\pi} \int_0^{2\pi} F(f(t))\mathrm{d}t.$$

Theorem

Let f be a real and 2π-periodic function. Then

$$\{T_n(f)\} \sim_\lambda (f, [0, 2\pi]), \qquad \text{if } f \text{ is continuous},$$
$$\{C_n(f)\} \sim_\lambda (f, [0, 2\pi]), \qquad \text{if } f \text{ is Lipschitz.}$$

Preconditioning

Given the system (to be solved by a Krylov method)

$$A_n x = b,$$

find P_n and solve $P_n^{-1} A_n x = P_n^{-1} b$ such that

Optimality:

1. the solution of a system with matrix P_n can be performed within the cost of a matrix-vector product with the matrix A_n,
2. $P_n^{-1} A_n$ is a "good spectral approximation" of the identity (constant number of iterations).

$d = 1$

Optimal preconditioner for $T_n(f)$ can be defined in the set of Circulant matrices (or other matrix algebras).

$d > 1$

We should not expect the classical matrix algebra preconditioners to lead to optimal iterative solvers in the ill-conditioned case.

When using matrix Algebra preconditioners, we lose necessarily the optimality and the number of iterations grows like $O(N^{\frac{d-1}{d}})$ [Serra-Capizzano, Tyrtyshnikov, SIMAX '99].

$\Downarrow$

Multigrid methods!

Outline

Algebraic interpretation of Multigrid methods

Multigrid Idea

Project the system in a subspace, solve the resulting system in this subspace and interpolate the solution in order to improve the previous approximation.

Multigrid components

The Multigrid combines two iterative methods:

Smoother: a classic iterative method,

Coarse Grid Correction: projection of the error equation, solution of the restricted problem, interpolation.

Even if the two components have not a good convergence, their combination could results in a very fast iterative method if they are spectrally complementary.

Error equation

We don't have information on the solution but we have information on the error $\Rightarrow$ at the lower level(s) it works on the error equation!

The Algebraic Multigrid

- Algebraic multigrid uses information on the coefficient matrix and no geometric information on the problem.
- Several smoothers have a similar behaviour: in the initial iterations they are not able to reduce effectively the error in the subspace generated by the eigenvectors associated to small eigenvalues (ill-conditioned subspace)

$$\Downarrow$$

- To obtain a fast solver the projector is chosen in order to project the error equation in such subspace.

Galerkin conditions

Define the restriction R, the prolongation P and the coarser matrix A_1 such that the coarse grid correction (CGC) is a projection:

1. $R = P^T$ and full rank,
2. $A_1 = RAP$, coarser matrix.

The algorithm

Two-Grid Methods (TGM)

The j-th iteration for the system $A\mathbf{x} = \mathbf{b}$:

(1) $\tilde{\mathbf{x}} = Smooth(A, \mathbf{x}^{(j)}, \mathbf{b}, \nu)$

(2) $\mathbf{r}_1 = P^T(\mathbf{b} - A\tilde{\mathbf{x}})$

(3) $A_1 = P^T A P$

(4) $\mathbf{e}_1 = A_1^{-1}\mathbf{r}_1$

(5) $\mathbf{x}^{(j+1)} = \tilde{\mathbf{x}} + P\mathbf{e}_1$

Coarse grid correction

The steps (2)–(5) are the coarse grid correction.

Post-smoother

A post smoothing step can be added after the step **(5)**

V-cycle: the step **(4)** becomes a recursive application of the algorithm with the zero vector as initial guess.

Outline

Coarse grid correction

- Iteration matrix of the two-grid method is

$$TGM = CGC * S,$$

where S is the iteration matrix of the pre-smoother and

$$CGC = I - P(P^T AP)^{-1} P^T A.$$

- Properties of CGC
  1. $CGC^2 = CGC$ (is a projector) $\Rightarrow \sigma(CGC) = \{0, 1\}$
  2. $\mathrm{Ker}(CGC) = \mathrm{Im}(P)$
  3. $\mathrm{Im}(CGC) = \mathrm{Ker}(P^T)$

- Let A be positive definite:
  1. $\mathrm{Im}(CGC) \perp_A \mathrm{Im}(P)$
  2. $\|CGC\|_A = 1$

Lemma

Let S be a polynomial of A, then

$$\rho(TGM) \leq \rho(S).$$

Ruge–Stüben theory

J. W. Ruge, K. Stüben, ''Algebraic multigrid''. In: S.F. McCormick (ed.) Multigrid methods, Frontiers Appl. Math., vol. 3, pp. 73--130. SIAM, Philadelphia (1987)

Theorem

Let $A \in \mathbb{R}^{n\times n}$ be s.p.d., $P \in \mathbb{R}^{n\times N}$, $N < n$, with full rank and $TGM = S * CGC$ (post-smoother instead of pre-smoother) defined using the Galerkin conditions. If $\forall\, \mathbf{x} \in \mathbb{R}^n$, $\exists \alpha, \beta > 0$ s.t.

1. $\|S\mathbf{x}\|_A^2 \le \|\mathbf{x}\|_A^2 - \alpha\|\mathbf{x}\|_{A^2}^2$ *(smoothing property)*
2. $\min_{\mathbf{y}\in\mathbb{R}^N} \|\mathbf{x} - P\mathbf{y}\|_2^2 \le \beta\|\mathbf{x}\|_A^2$ *(approximation property)*

Then $\beta > \alpha$ and

$$\|TGM\|_A \le \sqrt{1 - \frac{\alpha}{\beta}}.$$

Smoothing property

- For the TGM convergence it is useful to work with a matrix algebra like circulants.
- Whenever the smoothing property can be also proved directly on Toeplitz matrices.
- We consider the simple Richardson (or Jacobi) iteration.

Lemma

Let $A = T_n(f)$, $f \geq 0$ and $S = I - \tau A$. If $\tau \in (0, \frac{2}{\max(f)})$ then $\exists \alpha > 0$ s.t. that the smoothing property holds.

Proof

Remember that S is a polynomial of A and use the congruence transformation . . .

Projector

Requests for the projector

- Keep the same structure at each level (Toeplitz, . . .).
- Project in the ill-conditioned subspace.

Projector

The projector $P^T = K_n \mathcal{A}_n(p)$, where $\mathcal{A} \in \{C, T\}$ and

- $K_n \in \mathbb{R}^{N \times n}$ is the cutting matrix that preserves the structure at the lower level.
- $p(x, y)$ is the generating function of the projector, which selects the subspace where to project the error equation.

The cutting matrix K_n

circulant	**Toeplitz**
$\begin{bmatrix} 1 & 0 & & & & \\ & & 1 & 0 & & \\ & & & \ddots & \ddots & \\ & & & & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 0 & & & & \\ & & 0 & 1 & 0 & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & 0 & 1 & 0 \end{bmatrix}$

Approximation property for circulant matrices

- Let n be even and F_n be scaled by $\frac{1}{\sqrt{n}}$ such that it is unitary.
- The cutting matrix preserves the circulant structure because

$$K_n F_n = \frac{1}{\sqrt{2}} \left[F_{\frac{n}{2}}, F_{\frac{n}{2}} \right].$$

- Let $A = C_n(f)$ and $P = C_n(p) K_n^T$, then $A_1 = P^T A P$ s.t. $A_1 = C_n(f_1)$ with

$$f_1(x) = \frac{1}{2}\left(p^2 f\left(\frac{x}{2}\right) + p^2 f\left(\frac{x}{2} + \pi\right)\right).$$

Theorem

Let $A = C_n(f)$, n even, $f \geq 0$ trigonometric polynomial and $P = C_n(p) K_n^T$ and p trigonometric polynomial such that for $f(x^0) = 0$ it holds

1. $\limsup_{x \to x^0} \frac{p(x+\pi)^2}{f(x)} = c < +\infty$,
2. $p(x) + p(x+\pi) \neq 0, \qquad \forall x \in \mathbb{C}^n$

Then the approximation property is satisfied.

V-cycle optimality in the d-dimensional case

Definition

The set of all corners of x is

$$\Omega(x) = \{\, y \mid y_j \in \{x_j,\ \pi + x_j\},\ j = 1, \ldots, d\}$$

and the set of the "mirror" points of x is $\mathcal{M}(x) = \Omega(x) \setminus \{x\}$.

Theorem (Aricò, D., NM '07)

Let x_0 be the unique zero of f_i in $[0,\pi]^d$, $\forall x \in [0,\pi]^d$, for $i = 0, \ldots, m-1$ where m is the number of recursion levels, we choose p_i such that

$$\limsup_{x \to x_0} \left| \frac{p_i(y)}{f_i(x)} \right| < +\infty, \qquad y \in \mathcal{M}(x),$$

where

$$0 < \sum_{y \in \Omega(x)} p_i^2(y).$$

Then the V-cycle has a linear convergence rate independent of the dimension.