Algebraic preconditioning as transformation of discretization bases

Josef Málek and Zdeněk Strakoš

Nečas Center for Mathematical Modeling

Charles University in Prague and Czech Academy of Sciences

http://www.karlin.mff.cuni.cz/˜strakos

85th GAMM Annual Meeting, March 2014

Numerical solution of partial differential equations

Consider partial differential equation (PDE) problems described in the form of functional equation using the Hilbert space V and its dual $V^{\#}$

$$\mathcal{A}x = b, \quad \mathcal{A} : V \to V^{\#}, \quad x \in V, \quad b \in V^{\#},$$

where the linear bounded and coercive operator $\mathcal{A}$ is self-adjoint with respect to the duality pairing $\langle \cdot, \cdot \rangle$.

In solving boundary-value problems, the state-of-the-art literature on using the preconditioned conjugate gradient method (PCG) proceeds in most cases in the following way

$a(\cdot,\cdot),\ \langle b, \cdot \rangle \ \to\ \mathbf{A}_h, \mathbf{b}_h \ \to$ alg. preconditioning $\to$ PCG with $\mathbf{A}_h \mathbf{x}_h = \mathbf{b}_h$,

where h stands for a discretization parameter. Discretization and preconditioning are often considered separately.

Numerical solution of partial differential equations

Other ways link discretization and preconditioning; see, e.g., the hierarchical basis preconditioning by Yserentant (1985, 1986) and multilevel preconditioning by Axelsson and Vassilevski (1989), with the survey in Vassilevski (2008). If $\mathbf{A}_h \mathbf{x}_h = \mathbf{b}_h$ is the discretized algebraic system obtained by the nodal basis and $\mathbf{S}_h$ is the transformation matrix mapping the hierarchical basis to the nodal basis, then

$$\mathbf{S}_h^T \mathbf{A}_h \mathbf{S}_h \, \mathbf{S}_h^{-1} \mathbf{x}_h \; = \; \mathbf{S}_h^T \mathbf{b}_h$$

is the preconditioned system which corresponds to application of PCG with the hierarchical basis preconditioning, $\mathbf{M} = \mathbf{S}_h^{-T} \mathbf{S}_h^{-1}$.

Here transformation of the discretization basis provides preconditioning. Reversely, any algebraic preconditioning means transformation of the discretization basis; see, e.g., Jarošová, Klawonn and Rheinbach (2012) in the context of FETI-DP. It should be considered that way.

Outline

1. Operator preconditioning
2. CG in Hilbert spaces
3. Finite dimension and matrix formulation
4. Preconditioning as transformation of the basis
5. Conclusions and supplication

1 Basic setting

Let V be a real (infinite dimensional) Hilbert space with the inner product

$$(\cdot,\cdot)_V : V \times V \to R, \quad \text{the associated norm } \|\cdot\|_V ,$$

$V^\#$ be the dual space of bounded (continuous) linear functionals on V with the duality pairing

$$\langle \cdot,\cdot \rangle : V^\# \times V \to R \, .$$

For each $f \in V^\#$ there exists a unique $\tau f \in V$ such that

$$\langle f, v \rangle = (\tau f, v)_V \quad \text{for all } v \in V \, .$$

In this way the inner product $(\cdot,\cdot)_V$ determines the Riesz map

$$\tau : \; V^\# \to V \, .$$

1 Equivalent formulations

Functional equation in the data space $V^{\#}$:

$$\mathcal{A}x = b, \quad , \quad \mathcal{A} : V \rightarrow V^{\#}, \quad x \in V, \quad b \in V^{\#} .$$

Using the bounded and V-elliptic bilinear form $a(\cdot,\cdot) : V \times V \rightarrow R$ defined by $a(u,v) \equiv \langle \mathcal{A}u, v \rangle$ for all $u, v \in V$ gives the weak bilinear form formulation

$$a(x,v) = \langle b, v \rangle \quad \text{for all } v \in V .$$

With the transformation using the the Riesz map,

$$\tau\mathcal{A}x = \tau b, \qquad \tau\mathcal{A} : V \rightarrow V, \quad x \in V, \quad \tau b \in V ,$$

which is commonly (and inaccurately) called preconditioning.
Here $\tau\mathcal{A}$ is self-adjoint with respect to the inner product $(\cdot,\cdot)$.

1 Operator preconditioning references

Klawonn (1995, 1996); Arnold, Falk, and Winther (1997, 1997); Steinbach and Wendland (1998); Mc Lean and Tran (1997); Christiansen and Nédélec (2000, 2000); Powell and Silvester (2003); Elman, Silvester, and Wathen (2005); Hiptmair (2006); Axelsson and Karátson (2009); Mardal and Winther (2011); Kirby (2011); Zulehner (2011); Preconditioning Conference 2013, Oxford; ...

Related ideas can be found also in Faber, Manteuffel and Parter (1990) with references to D'Yakanov (1961) and Gunn(1964, 1965).

The focus is on achieving mesh (model) parameters independence of the condition number-based convergence bounds.

Outline

1. Operator preconditioning
2. CG in Hilbert spaces
3. Finite dimension and matrix formulation
4. Preconditioning as transformation of the basis
5. Conclusions and supplication

2 Krylov manifolds in Hilbert spaces

Using the Riesz map, $\tau\mathcal{A} : V \rightarrow V$. One can therefore form for $g \in V$ the Krylov sequence

$$g,\ \tau\mathcal{A}g\,,\ (\tau\mathcal{A})^2 g\,,\ \ldots \qquad \text{in } V$$

and define Krylov subspace methods in the Hilbert space operator setting (here CG) for solving the functional equation

$$\mathcal{A}x = b, \quad x \in V, \quad b \in V^{\#}$$

such that with $r_0 = b - \mathcal{A}x_0 \in V^{\#}$ the approximations x_n to the solution x , $n = 1, 2, \ldots$ belong to the Krylov manifolds in V

$$x_n \in x_0 + K_n(\tau\mathcal{A}, \tau r_0) \equiv$$
$$x_0 + \text{span}\{\tau r_0, \tau\mathcal{A}(\tau r_0), (\tau\mathcal{A})^2(\tau r_0), \ldots, (\tau\mathcal{A})^{n-1}(\tau r_0)\}\,.$$

2 Self-adjoint $\mathcal{A}$ wrt the duality pairing

The goal is to approximate the solution $x = (\tau\mathcal{A})^{-1}\tau b$ using the initial approximation x_0 and the vectors $\tau r_0, \tau\mathcal{A}(\tau r_0), (\tau\mathcal{A})^2(\tau r_0), \ldots,$ which is different from approximating the operator inverse $(\tau\mathcal{A})^{-1}$ by the operators $I, \tau\mathcal{A}, (\tau\mathcal{A})^2, \ldots$

Looking for the approximate solution minimizing energy leads to

$$\|x - x_n\|_a = \min_{z \in x_0 + K_n} \|x - z\|_a \,,$$

which is equivalent to the Galerkin orthogonality condition

$$\langle b - \mathcal{A}x_n, w\rangle = \langle r_n, w\rangle = 0 \qquad \text{for all } w \in K_n \equiv K_n(\tau\mathcal{A}, \tau r_0)\,.$$

Since K_n is finite dimensional, this provides discretization of the problem by the matching moments model reduction.

2 (Preconditioned) CG in Hilbert spaces

$r_0 = b - \mathcal{A}x_0 \in V^{\#}, \quad p_0 = \tau r_0 \in V$

For $n = 1, 2, \ldots, n_{\max}$

$$\alpha_{n-1} = \frac{\langle r_{n-1}, \tau r_{n-1} \rangle}{\langle \mathcal{A}p_{n-1}, p_{n-1} \rangle} = \frac{(\tau r_{n-1}, \tau r_{n-1})_V}{(\tau \mathcal{A}p_{n-1}, p_{n-1})_V}$$

$$x_n = x_{n-1} + \alpha_{n-1} p_{n-1} \,,$$

stop when the stopping criterion is satisfied

$$r_n = r_{n-1} - \alpha_{n-1} \mathcal{A}p_{n-1}$$

$$\beta_n = \frac{\langle r_n, \tau r_n \rangle}{\langle r_{n-1}, \tau r_{n-1} \rangle} = \frac{(\tau r_n, \tau r_n)_V}{(\tau r_{n-1}, \tau r_{n-1})_V}$$

$$p_n = \tau r_n + \beta_n p_{n-1}$$

End

Hayes (1954); ... ; Glowinski (2003); Axelsson and Karatson (2009);
Mardal and Winther (2011); Günnel, Herzog and Sachs (2013)

2 CG $\equiv$ Gauss-Christoffel quadrature

$$\mathcal{A}, w_1 = \tau r_0 / \|\tau r_0\|_V \quad \longleftrightarrow \quad \omega(\lambda), \quad \int f(\lambda)\, d\omega(\lambda)$$

$$\uparrow \qquad\qquad\qquad\qquad \uparrow$$

$$\mathbf{T}_n, \mathbf{e}_1 \quad \longleftrightarrow \quad \omega^{(n)}(\lambda), \quad \sum_{i=1}^{n} \omega_i^{(n)} f\left(\theta_i^{(n)}\right)$$

$$\omega^{(n)}(\lambda) \longrightarrow \omega(\lambda)$$

2 CG $\equiv$ Gauss-Christoffel quadrature

Using $f(\lambda) = \lambda^{-1}$ we get the quantitative relation between the Gauss-Christoffel quadrature and the energy norm of the error in CG

$$\int \lambda^{-1} \, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} \left(\theta_i^{(n)}\right)^{-1} + \frac{\|x - x_n\|_a^2}{\|\tau r_0\|_V^2}$$

Condition number bounds should always be checked against this!

Outline

1. Operator preconditioning
2. CG in Hilbert spaces
3. Finite dimension and matrix formulation
4. Preconditioning as transformation of the basis
5. Conclusions and supplication

3 Finite dimensional space and the matrix CG

Let $\Phi_h = (\phi_1^{(h)}, \dots, \phi_N^{(h)})$ be the basis of the finite dimensional subspace $V \subset V_h$, $\Phi_h^{\#} = (\phi_1^{(h)\#}, \dots, \phi_N^{(h)\#})$ be the canonical basis of its dual $V_h^{\#}$, (it holds $V_h^{\#} = \mathcal{A} V_h$). Using the coordinates in Φ_h and in $\Phi_h^{\#}$,

$$\langle f, v \rangle \to \mathbf{v}^* \mathbf{f}\,, \quad (u,v)_V \to \mathbf{v}^* \mathbf{M} \mathbf{u}, \quad (\mathbf{M}_{ij}) = ((\phi_j, \phi_i)_V)_{i,j=1,\dots,N}\,,$$

$$\mathcal{A}u \to \mathbf{A}\mathbf{u}\,, \qquad \mathcal{A}u = \mathcal{A}\Phi_h \mathbf{u} = \Phi_h^{\#} \mathbf{A}\mathbf{u}\,;$$

$$(\mathbf{A}_{ij}) = (a(\phi_j, \phi_i))_{i,j=1,\dots,N} = (\langle \mathcal{A}\phi_j, \phi_i \rangle)_{i,j=1,\dots,N}\,,$$

$$\tau f \to \mathbf{M}^{-1}\mathbf{f}\,, \qquad \tau f = \tau \Phi_h^{\#} \mathbf{f} = \Phi_h \mathbf{M}^{-1} \mathbf{f}\,;$$

we get with $b = \Phi_h \mathbf{b}$, $x_n = \Phi_h \mathbf{x}_n$, $p_n = \Phi_h \mathbf{p}_n$, $r_n = \Phi_h^{\#} \mathbf{r}_n$

3 Preconditioned algebraic CG

$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0,$ solve $\mathbf{M}\mathbf{z}_0 = \mathbf{r}_0, \; \mathbf{p}_0 = \mathbf{z}_0$

For $n = 1, \ldots, n_{\max}$

$$\alpha_{n-1} = \frac{\mathbf{z}_{n-1}^* \mathbf{r}_{n-1}}{\mathbf{p}_{n-1}^* \mathbf{A} \mathbf{p}_{n-1}}$$

$$\mathbf{x}_n = \mathbf{x}_{n-1} + \alpha_{n-1} \mathbf{p}_{n-1} \,,$$ stop when the stopping criterion is satisfied

$$\mathbf{r}_n = \mathbf{r}_{n-1} - \alpha_{n-1} \mathbf{A} \mathbf{p}_{n-1}$$

$$M \mathbf{z}_n = \mathbf{r}_n \,,$$ solve for $\mathbf{z}_n$

$$\beta_n = \frac{\mathbf{z}_n^* \mathbf{r}_n}{\mathbf{z}_{n-1}^* \mathbf{r}_{n-1}}$$

$$\mathbf{p}_n = \mathbf{z}_n + \beta_n \mathbf{p}_{n-1}$$

End

Theorem

$$\kappa(\mathbf{M}^{-1}\mathbf{A}) \;\leq\; \frac{\sup_{u,v\in V,\, \|u\|_V=1, \|v\|_V=1} |\langle \mathcal{A}u, v\rangle|}{\inf_{u\in V,\, \|u\|_V=1} \langle \mathcal{A}u, u\rangle}$$

See, e.g., Hiptmair (2006).

Unpreconditioned CG, i.e. $\mathbf{M} = \mathbf{I}$ corresponds
to the basis Φ orthonormal wrt $(\cdot, \cdot)_V$.

Outline

1. Operator preconditioning
2. CG in Hilbert spaces
3. Finite dimension and matrix formulation
4. Preconditioning as transformation of the basis
5. Conclusions and supplication

4 Galerkin (think of FEM) discretization

Consider an N-dimensional discrete solution subspace $V_h \subset V$ with the duality pairing and the inner product as above. Then the restriction to V_h gives an approximation $x_h \in V_h$ to $x \in V$,

$$a(x_h, v) = \langle b, v \rangle \qquad \text{for all} \quad v \in V_h \,.$$

With the basis $\Phi_h = (\phi_1^{(h)}, \dots, \phi_N^{(h)})$ of V_h, this gives the discretized algebraic system

$$\mathbf{A}_h \mathbf{x}_h = \mathbf{b}_h$$

and, with the algebraic preconditioning $\widehat{\mathbf{M}} = \widehat{\mathbf{L}}\widehat{\mathbf{L}}^*$

$$(\widehat{\mathbf{L}}^{-1}\mathbf{A}_h\,(\widehat{\mathbf{L}}^*)^{-1})\,(\widehat{\mathbf{L}}^*\mathbf{x}_h) \;=\; \widehat{\mathbf{L}}^{-1}\mathbf{b}_h \,.$$

4 Preconditioning transforms the basis

Algebraic (unpreconditioned) CG is then applied to the preconditioned system with the substitition to the original approximate solution $\mathbf{x}_h^{(n)}$ and residual $\mathbf{r}_h^{(n)}$, which gives PCG.

Question: How can the algebraic preconditioning of the discretized system $\mathbf{A}_h \mathbf{x}_h = \mathbf{b}_h$ be interpreted within the framework of CG in Hilbert spaces?

Observation:

Let the inner product $(.,.)$ be unchanged, let $\mathbf{M} = \mathbf{L}\mathbf{L}^*$. Then the algebraically preconditioned PCG with the preconditioner $\widehat{\mathbf{M}} = \widehat{\mathbf{L}}\widehat{\mathbf{L}}^*$ is obtained within the Hilbert space formulation of CG with the transformed discretization bases while preserving the approximation subspace

$$\Phi_t = \Phi_h (\mathbf{L}^*)^{-1} \widehat{\mathbf{L}}^*, \quad \Phi_t^{\#} = \Phi_h^{\#} \mathbf{L} \widehat{\mathbf{L}}^{-1} .$$

4 Local discretization and global preconditioning

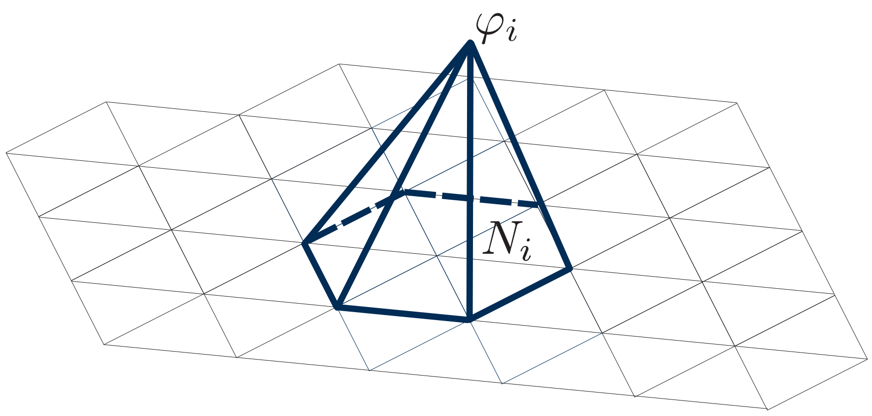
Sparsity of the resulted matrices is always presented as the main advantage of FEM discretizations.

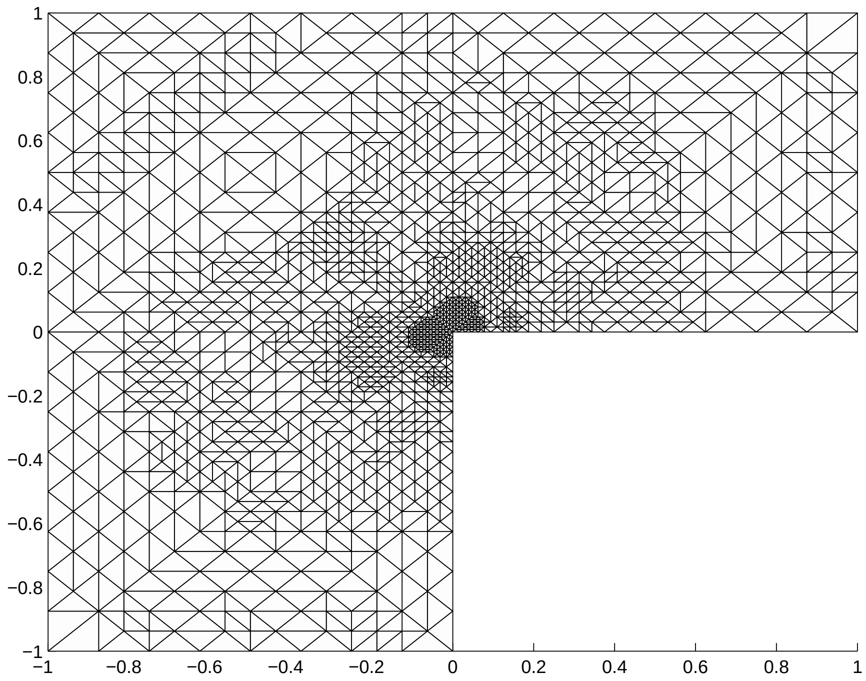
Sparsity means locality of information. In order to solve the problem, we need a global transfer of information. Therefore preconditioning! It is needed on the computational level in order to take care for the trouble caused by the (*computationally*) inconvenient approximation of the mathematical model when the *appropriate globally supported* basis functions are missing (cf. hierarchical bases preconditioning, DD with coarse space components, multilevel methods, ...). Recall Rüde (2009).

Preconditioning can be interpreted in part as addressing the difficulty related to sparsity (locality of the supports of the basis functions).

Sparsity is important for efficiency, but perhaps in a different meaning; see, e.g., Schaeffer, Caflisch, Hauck and Osher (2013), .

4 Local discretisation and global computation





Outline

1. Operator preconditioning

2. CG in Hilbert spaces

3. Finite dimension and matrix formulation

4. Preconditioning as transformation of the basis

5. Conclusions and supplication

5 Discretization via Krylov subspaces

- Coarse grid components, inverted dense blocks etc. means handling global information. The focus on locality of FEM bases?

- Changing the approximation subspace.

- What if the n-th Krylov subspace K_n is taken as the finite dimensional subspace $V_h \subset V$ in

$$\{\mathcal{A}, \tau\} \;\rightarrow\; \{\tau\mathcal{A}_n \,:\, K_n \rightarrow K_n\} \;\rightarrow\; \text{PCG with } \{\mathbf{A}_h, \mathbf{M}_h\}\ ?$$

5 Please help with explaining the myths

- It is not true that CG (or other Krylov subspace methods used for solving systems of linear algebraic equations with symmetric matrices) applied to a matrix with t distinct well separated tight clusters of eigenvalues produces in general a large error reduction after t steps; see Sections 5.6.5 and 5.9.1 of Liesen, S (2013). This myth has been disproved more than 20 years ago; see Greenbaum (1989); S (1991); Greenbaum, S (1992). Still it is persistently repeated in an authoritative way in literature.

- With no information on the structure of invariant subspaces it is not true that distribution of eigenvalues provides insight into the asymptotic behavior of Krylov subspace methods (such as GMRES) applied to systems with generally nonsymmetric matrices; see Sections 5.7.4, 5.7.6 and 5.11 of Liesen, S (2013). As before, the relevant results Greenbaum, S (1994); Greenbaum, Pták, S (1996) and Arioli, Pták, S (1998) are (almost) twenty years old.

Recent references

- J. Liesen and Z.S., Krylov Subspace Methods, Principles and Analysis. Oxford University Press (2013)
- T. Gergelits and Z.S., Composite convergence bounds based on Chebyshev polynomials and finite precision conjugate gradient computations, Numerical Algorithms (2013) (DOI 10.1007/s11075-013-9713-z)
- J. Papez, J. Liesen and Z.S., On distribution of the discretization and algebraic error in numerical solution of partial differential equations, to appear in LAA (2014)
- M. Arioli, J. Liesen, A. Miedlar, and Z.S., Interplay between discretization and algebraic computation in adaptive numerical solution of elliptic PDE problems, GAMM Mitteilungen 36, 102-129 (2013)
- J. Málek and Z.S., From PDEs through functional analysis to iterative methods, or there and back again. Preprint MORE/2014/02.

Thank you very much for kind patience!