On operator and matrix view to preconditioned Krylov subspace methods

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

Nick Trefethen 2015, Oxford

Motivation

- Infinite dimensional description of the problem (say, formulated using PDEs)
- Discretization
- Finite dimensional iterative matrix computations, here using preconditioned Krylov subspace methods

Standard way. Is it worth a thought?

Notation

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

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

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

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

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

$$\tau : V^{\#} \to V .$$

Weak formulation of the BVP

Let $a(\cdot,\cdot) = V \times V \rightarrow R$ be a bounded and coercive bilinear form. For a fixed $u \in V$ we can see $a(u,\cdot)$ as the bounded linear functional on V and we can write it as

$$\mathcal{A}u \equiv a(u,\cdot) \in V^{\#}\,, \quad \text{i.e.}\,,$$

$$\langle \mathcal{A}u, v\rangle = a(u,v) \quad \text{for all } \; v \in V\,.$$

This defines the bounded and coercive operator

$$\mathcal{A} : V \rightarrow V^{\#}, \qquad \inf_{u \in V,\, \|u\|_V = 1} \langle \mathcal{A}u, u\rangle = \alpha > 0, \; \|\mathcal{A}\| = C\,.$$

The Lax-Milgram theorem ensures that for any $b \in V^{\#}$ there exists a unique solution $x \in V$ of the problem

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

Operator formulations and preconditioning

Equivalently,

$$\langle \mathcal{A}x - b, v \rangle = 0 \quad \text{for all } v \in V\,,$$

which can be written as the equation in $V^{\#}$,

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

Using the Riesz map,

$$(\tau \mathcal{A}x - \tau b, v)_V = 0 \quad \text{for all } v \in V\,.$$

Clearly, the Riesz map τ can be interpreted as transformation of the original problem $\mathcal{A}x = b$ in $V^{\#}$ into the equation in V

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

which is commonly (and inaccurately) called preconditioning.

Outline

1. Vorobyev method of moments and the conjugate gradient method in Hilbert spaces

2. t tight clusters of eigenvalues do not necessarily mean achieving a reasonable accuracy in t steps

3. Operator description of convergence of finite dimensional algebraic solvers

4. Algebraic preconditioning and transformation of the discretization basis

1 Vorobyev method of moments

Let $\mathcal{B}\,(=\tau\mathcal{A})$ be a bounded linear operator on Hilbert space V. Choosing $z_0\,(=\tau r_0) \in V$. Consider the Krylov sequence
$z_0,\; z_1 = \mathcal{B}z_0,\; z_2 = \mathcal{B}z_1 = \mathcal{B}^2 z_0,\; \ldots\;, z_n = \mathcal{B}z_{n-1} = \mathcal{B}^n z_{n-1},\; \ldots$

Determine a sequence of operators $\mathcal{B}_n$ defined on the sequence of nested subspaces $V_n = \text{span}\,\{z_0, \ldots, z_{n-1}\}$, with the projector E_n onto V_n, such that

$$z_1 = \mathcal{B}z_0 = \mathcal{B}_n z_0,$$

$$z_2 = \mathcal{B}^2 z_0 = (\mathcal{B}_n)^2 z_0,$$

$$\vdots$$

$$z_{n-1} = \mathcal{B}^{n-1} z_0 = (\mathcal{B}_n)^{n-1} z_0,$$

$$E_n z_n = E_n \mathcal{B}^n z_0 = (\mathcal{B}_n)^n z_0.$$

1 Model reduction using Krylov subspaces

Using the projection E_n onto V_n we can write for the operators constructed above (here we need the linearity of $\mathcal{B}$)

$$\mathcal{B}_n = E_n \, \mathcal{B} \, E_n \, .$$

The finite dimensional operators $\mathcal{B}_n$ can be used to obtain approximate solutions to various linear problems. The choice of $z_0, z_1, \ldots$ as above gives a sequence of Krylov subspaces that are determined by the operator $\mathcal{B}$ and the initial element z_0 . In this way the Vorobyev method of moments gives the Krylov subspace methods.

Vorobyev (1958, 1965) covers bounded linear operators, bounded self-adjoint operators and some unbounded extensions. He made links to CG, Lanczos, Stieltjes moment problem, work of Markov, Gauss-Christoffel quadrature . . .

1 Bounded self-adjoint operators in V

$$\mathcal{B}\,u \;=\; f \qquad\longleftrightarrow\qquad \omega(\lambda), \qquad \int F(\lambda)\,d\omega(\lambda)$$

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

$$\mathbf{T}_n\,\mathbf{y}_n \;=\; \|f\|_V\,\mathbf{e}_1 \qquad\longleftrightarrow\qquad \omega^{(n)}(\lambda), \;\; \sum_{i=1}^{n} \omega_i^{(n)}\,F\left(\theta_i^{(n)}\right)$$

Using $F(\lambda) = \lambda^{-1}$ gives (assuming coercivity)

$$\int_{\lambda_L}^{\lambda_U} \lambda^{-1}\,d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} \left(\theta_i^{(n)}\right)^{-1} \;+\; \frac{\|u - u_n\|_a^2}{\|f\|_V^2}$$

Stieltjes (1894) and Vorobyev (1958) moment problems for self-adjoint bounded operators reduce to the Gauss-Christoffel quadrature (1814).
No one would consider describing it by contraction.

1 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); Vorobyev (1958, 1965); Karush (1952); Stesin (1954)
Superlinear convergence for (identity + compact) operators.

Outline

1. Vorobyev method of moments and the conjugate gradient method in Hilbert spaces

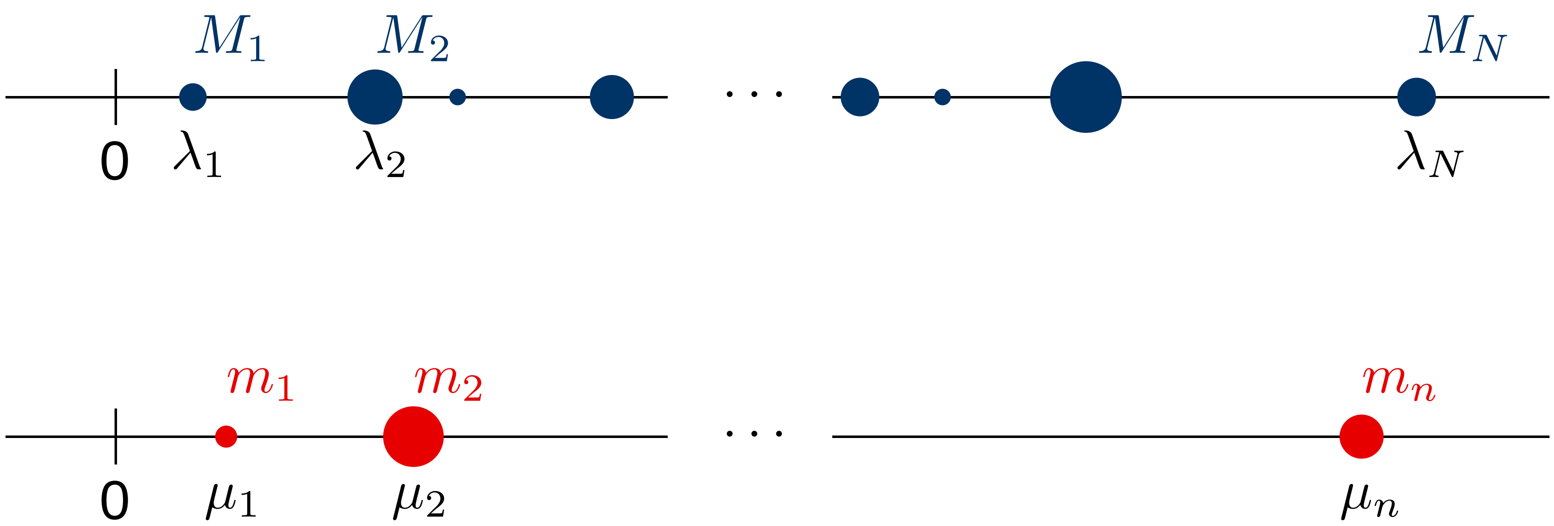
2. t tight clusters of eigenvalues do not necessarily mean achieving a reasonable accuracy in t steps

3. Operator description of convergence of finite dimensional algebraic solvers

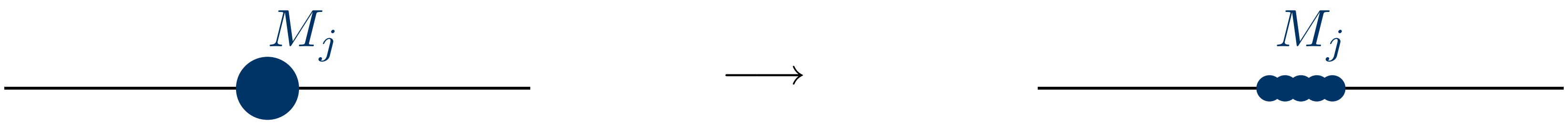
4. Algebraic preconditioning and transformation of the discretization basis

2 Moment problem illustration of exact CG

For a given n find a distribution function with n mass points in such a way that it in a best way captures the properties of the original distribution function determined by the operator τA and the normalized initial residual.



2 Exact CG with tight clusters



single eigenvalue λ_j $\longrightarrow$ many close eigenvalues $\widehat{\lambda}_{j_1}, \widehat{\lambda}_{j_2}, \ldots, \widehat{\lambda}_{j_\ell}$

Replacing single eigenvalues by tight clusters can make a substantial difference Greenbaum (1989); Greenbaum, S (1992); Golub, S (1994). Otherwise CG behaves almost linearly and it can be described by contraction. In such case - is it worth using?

2 Rounding errors can be an important issue

- Good preconditioning can ensure in some problems getting an acceptable solution in a few iterations. Then rounding errors are not of much concern
- However, hard problems do exist, and then we need to consider a considerable number of iterations
- Then rounding errors can not be ignored
- Descriptions of the behaviour of Krylov subspace methods that are based on contraction (condition numbers) are, in general, not descriptive (apart from special cases and cases where the use of Krylov subspace methods should be questioned). Analogy with a-priori and a-posteriori analysis in numerical PDEs

2 Any GMRES convergence with any spectrum

The following statements are equivalent:

1° The spectrum of $\mathbf{A}$ is given by $\{\lambda_1, \ldots, \lambda_N\}$ and GMRES$(\mathbf{A}, \mathbf{b})$ yields residuals with the prescribed nonincreasing sequence

$$\|\mathbf{r}_0\| \geq \|\mathbf{r}_1\| \geq \cdots \geq \|\mathbf{r}_{N-1}\| > \|\mathbf{r}_N\| = 0\,.$$

2° Matrix $\mathbf{A}$ is of the form $\mathbf{A} = \mathbf{W}\mathbf{R}\mathbf{C}\mathbf{R}^{-1}\mathbf{W}^*$ and $\mathbf{b} = \mathbf{W}\mathbf{h}$ where $\mathbf{C}$ is the spectral companion matrix, $\mathbf{W}$ is unitary and $\mathbf{R}$ is a nonsingular upper triangular matrix such that $\mathbf{R}\mathbf{s} = \mathbf{h}$ with $\mathbf{s}$ being the first column of $\mathbf{C}^{-1}$.

Greenbaum, Ptak, Arioli and S (1994 - 98); Liesen (1999); Eiermann and Ernst (2001); Meurant (2012); Meurant and Tebbens (2012, 2014); .....

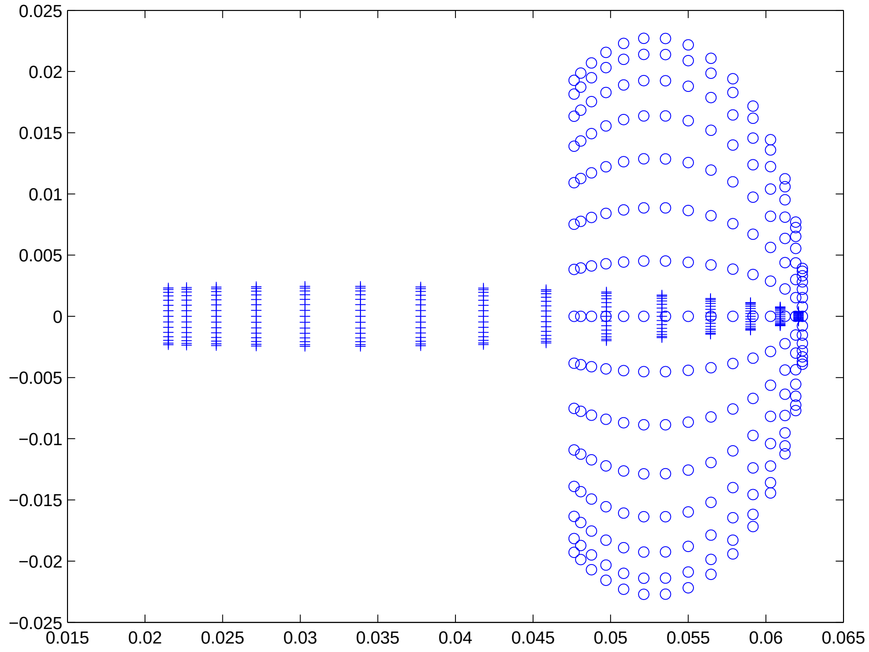
2 Interpretation?

Given the spectrum and the sequence of residual norms, this results gives a complete parametrization of the set of all matrices and right hand sides. The set of problems for which the distribution of eigenvalues alone does not conform to convergence behaviour is not of measure zero and it is not pathological.

Widespread eigenvalues (whatever it means) alone can not be identified with poor convergence, similarly as clustered eigenvalues (whatever it means) alone can not be identified with fast convergence.

Equivalent orthogonal matrices, Greenbaum, S (1994).
Pseudospectrum indication!

2 Convection-diffusion model problem



Quiz: In one case the convergence of GMRES is substantially faster than in the other; for the solution see Liesen, S (2005).

Outline

1. Vorobyev method of moments and the conjugate gradient method in Hilbert spaces

2. t tight clusters of eigenvalues do not necessarily mean achieving a reasonable accuracy in t steps

3. Operator description of convergence of finite dimensional algebraic solvers

4. Algebraic preconditioning and transformation of the discretization basis

3 Basic facts

Consider a bounded linear operator $\mathcal{B}$ on a Hilbert space V, which has a (bounded) inversion, and the problem

$$\mathcal{B}\, u = f \, .$$

- Since the identity operator on an infinite dimensional Hilbert space is not compact and $\mathcal{B}\mathcal{B}^{-1} = \mathcal{I}$, it follows that $\mathcal{B}$ can not be compact.
- A uniform limit (in norm) of finite dimensional (approximation) operators $\mathcal{B}_n$ is a compact operator.
- Results on strong convergence (pointwise limit); for the method of moments see Vorobyev (1958, 1965)

$$\|\mathcal{B}_n\, w - \mathcal{B}\, w\| \to 0 \quad \forall w \in V \, .$$

3 Analysis of convergence of Ksp methods?

Let $\mathcal{Z}_h$ be a numerical approximation of a bounded operator $\mathcal{Z}$ such that

$$\|\mathcal{Z} - \mathcal{Z}_h\| = \mathcal{O}(h).$$

Then we have $[(\lambda - \mathcal{Z})^{-1} - (\lambda - \mathcal{Z}_h)^{-1}] = \mathcal{O}(h)$ uniformly for $\lambda \in \Gamma$, where Γ surrounds the spectrum of $\mathcal{Z}$ with a distance of order $\mathcal{O}(h)$ or more. For any polynomial p

$$p(\mathcal{Z}) - p(\mathcal{Z}_h) = \frac{1}{2\pi i} \int_\Gamma p(\lambda)[(\lambda - \mathcal{Z})^{-1} - (\lambda - \mathcal{Z}_h)^{-1}]\, d\lambda\,.$$

This motivates using operator $\mathcal{Z}$ based description of convergence behaviour of finite dimensional Krylov subspace methods.

But the *assumption* $\|\mathcal{Z} - \mathcal{Z}_h\| = \mathcal{O}(h)$ does not hold for any bounded invertible infinite dimensional operator $\mathcal{Z}$.

Outline

1. Vorobyev method of moments and the conjugate gradient method in Hilbert spaces

2. t tight clusters of eigenvalues do not necessarily mean achieving a reasonable accuracy in t steps

3. Operator description of convergence of finite dimensional algebraic solvers

4. Algebraic preconditioning and transformation of the discretization basis

4 Galerkin discretization gives matrix CG in V_h

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

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

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

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

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

$$\mathbf{z}_n = \mathbf{M}^{-1}\mathbf{r}_n\,,$$ 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

Günnel, Herzog, Sachs (2014); Málek, S (2015)

4 Restriction to a finite dimensional subspace

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

$$\langle f, v \rangle \to \mathbf{v}^* \mathbf{f} \,,$$

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

$$\mathcal{A} u \to \mathbf{A} \mathbf{u} \,, \qquad \mathcal{A} u = \mathcal{A} \Phi_h \mathbf{u} = \Phi_h^{\#} \mathbf{A} \mathbf{u} \,; \quad (\mathbf{A}_{ij}) = (a(\phi_j, \phi_i))_{i,j=1,\dots,N} \,,$$

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

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

4 Observations

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

- Operator preconditioning on a discrete space can be interpreted via orthogonalization of the discretization basis. As a consequence, the resulting matrix of the algebraic system is not sparse!

- Interpretation of the algebraic preconditioning with the preconditioner

$$\widehat{\mathbf{M}} = \widehat{\mathbf{L}}\widehat{\mathbf{L}}^*$$

other than the discretized operator preconditioner $\mathbf{M} = \mathbf{L}\mathbf{L}^*$ given by the Riesz map τ ?

4 Interpretation of the algebraic preconditioning

Transform the discretization bases

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

with the change of the inner product in V_h (recall $(u,v)_V = \mathbf{v}^*\mathbf{M}\mathbf{u}$)

$$(u,v)_{\text{new},V_h} = (\widehat{\Phi}\widehat{\mathbf{u}}, \widehat{\Phi}\widehat{\mathbf{v}})_{\text{new},V_h} := \widehat{\mathbf{v}}^*\widehat{\mathbf{u}} = \mathbf{v}^*\widehat{\mathbf{L}}\widehat{\mathbf{L}}^*\mathbf{u} = \mathbf{v}^*\widehat{\mathbf{M}}\mathbf{u}\,.$$

Then the discretized Hilbert space formulation of CG gives the algebraically preconditioned matrix formulation of CG with the preconditioner $\widehat{\mathbf{M}}$ (in particular, the unpreconditioned CG applied to the algebraically preconditioned discretized system).

4 Sparsity, locality, global transfer of information

Sparsity of matrices of the algebraic systems is always presented as an advantage of the FEM discretizations.

Sparsity means locality of information. In order to solve the problem, we need the appropriate global transfer of information. This can be done, e.g., by globally supported basis functions (cf. hierarchical bases preconditioning, DD with coarse space components, multilevel methods, hierarchical grids ...).

Preconditioning can be interpreted in part as addressing the difficulty related to sparsity (locality of the supports of the basis functions). Coarse grid components, hierarchy of grids etc. helps in efficient handling the transfer of global information.

Conclusions

- Plethora of Krylov subspace methods and their implementations

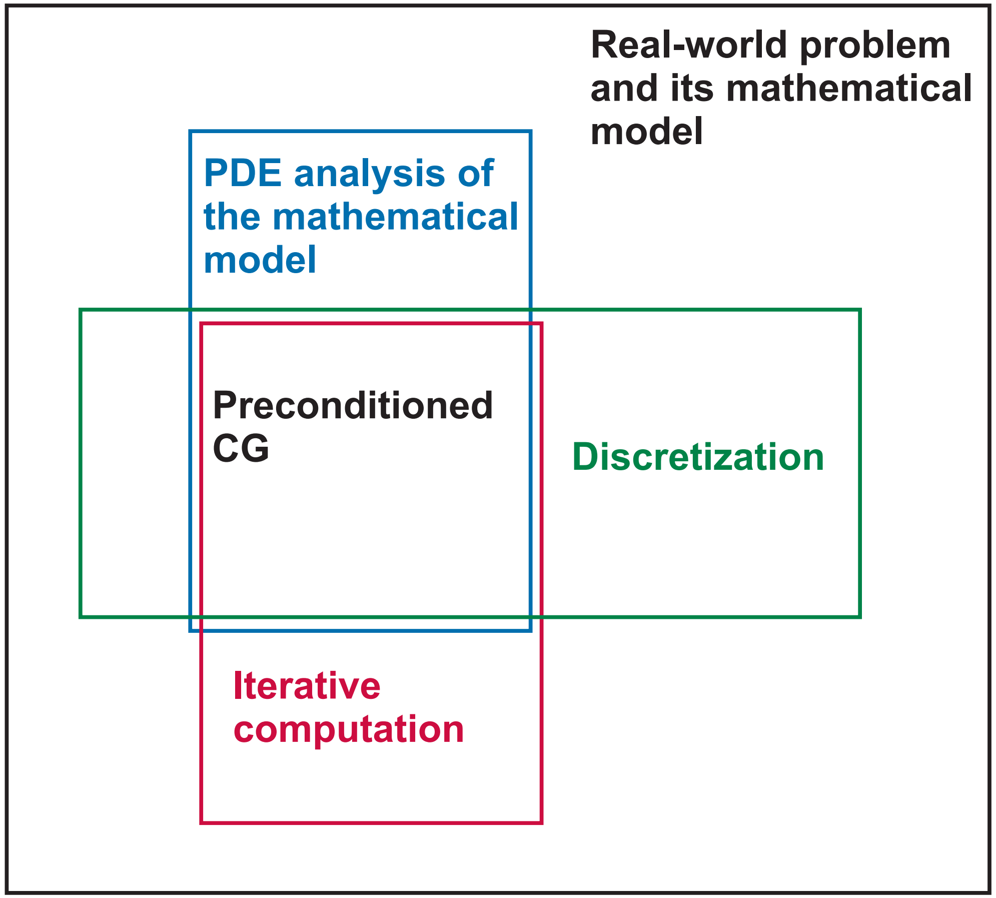
- Understanding of the basic ones, including preconditioning, in the context of the PDE problems to be solved will form a solid ground for further algorithmic developments

- While formulating results of an analysis, attention should be paid to the assumptions restricting their applicability. They offer a valuable guidance on what is yet to be done and which way to go. Generalizing statements lacking a mathematical ground do a lot of harm

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

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

Málek and S, SIAM Spotlight, 2015



References

- J. Málek and Z.S., *Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs.* SIAM Spolight Series, SIAM (2015)
- T. Gergelits and Z.S., *Composite convergence bounds based on Chebyshev polynomials and finite precision conjugate gradient computations,* Numer. Alg. 65, 759-782 (2014)
- J. Papež, J. Liesen and Z.S., *Distribution of the discretization and algebraic error in numerical solution of partial differential equations,* Linear Alg. Appl. 449, 89-114 (2014)
- J. Liesen and Z.S., *Krylov Subspace Methods, Principles and Analysis.* Oxford University Press (2013)
- Z.S. and P. Tichý, *On efficient numerical approximation of the bilinear form $c^* A^{-1} b$,* SIAM J. Sci. Comput., 33 (2011), pp. 565-587

Happy Birthday, Nick!