Krylov subspace methods for linear systems with tensor product structure

Christine Tobler

Seminar for Applied Mathematics,
ETH Zürich

19. August 2009

Outline

Outline

Linear system with tensor product structure

We consider the linear system

$$\mathcal{A}x = b$$

with

$$\mathcal{A} = \sum_{s=1}^{d} I_{n_1} \otimes \cdots \otimes I_{n_{s-1}} \otimes A_s \otimes I_{n_{s+1}} \otimes \cdots \otimes I_{n_d},$$

$$b = b_1 \otimes \cdots \otimes b_d,$$

$$A_s \in \mathbb{R}^{n_s \times n_s} \text{ positive definite, } b_s \in \mathbb{R}^{n_s}.$$

Example for 3 dimensions:

$$(A_1 \otimes I \otimes I + I \otimes A_2 \otimes I + I \otimes I \otimes A_3)x = b_1 \otimes b_2 \otimes b_3$$

Linear system with tensor product structure

We consider the linear system

$$\mathcal{A}x = b$$

with

$$\mathcal{A} = \sum_{s=1}^{d} I_{n_1} \otimes \cdots \otimes I_{n_{s-1}} \otimes A_s \otimes I_{n_{s+1}} \otimes \cdots \otimes I_{n_d},$$

$$b = b_1 \otimes \cdots \otimes b_d,$$

$$A_s \in \mathbb{R}^{n_s \times n_s} \text{ positive definite, } b_s \in \mathbb{R}^{n_s}.$$

Example for 3 dimensions:

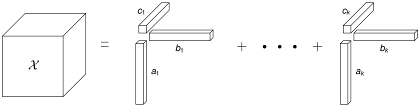
$$(A_1 \otimes I \otimes I + I \otimes A_2 \otimes I + I \otimes I \otimes A_3)x = b_1 \otimes b_2 \otimes b_3$$

Tensor decompositions

CP decomposition:

$$\mathrm{vec}(\mathcal{X}) = \sum_{r=1}^{k} a_r \otimes b_r \otimes c_r$$

$$a_r \in \mathbb{R}^m, b_r \in \mathbb{R}^n, c_r \in \mathbb{R}^p$$



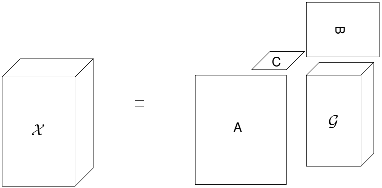
Tucker decomposition:

$$\mathrm{vec}(\mathcal{X}) = \sum_{i=1}^{\tilde{m}} \sum_{s=1}^{\tilde{n}} \sum_{l=1}^{\tilde{p}} \mathcal{G}_{ijl}\, a_i \otimes b_j \otimes c_l$$

$$= (A \otimes B \otimes C)\,\mathrm{vec}(\mathcal{G})$$

$$\mathcal{G} \in \mathbb{R}^{\tilde{m} \times \tilde{n} \times \tilde{p}},$$

$$a_i \in \mathbb{R}^m, b_j \in \mathbb{R}^n, c_l \in \mathbb{R}^p$$



About this system

The eigenvalues of the matrix $\mathcal{A}$ are given by all possible sums

$$\lambda_{i_1}^{(1)} + \lambda_{i_2}^{(2)} + \ldots + \lambda_{i_d}^{(d)}$$

where $\lambda_{i_s}^{(s)}$ denotes an eigenvalue of A_s. For A_s positive definite, the system has a unique solution.

Note that x and b are vector representations of tensors in $\mathbb{R}^{n_1 \times \cdots \times n_d}$, and b represents a rank-one tensor.

A tensor arising from the discretization of a sufficiently smooth function f can be approximated by a short sum of rank-one tensors. By superposition, we can insert such a tensor as right-hand side into our system.

About this system

The eigenvalues of the matrix $\mathcal{A}$ are given by all possible sums

$$\lambda_{i_1}^{(1)} + \lambda_{i_2}^{(2)} + \ldots + \lambda_{i_d}^{(d)}$$

where $\lambda_{i_s}^{(s)}$ denotes an eigenvalue of A_s. For A_s positive definite, the system has a unique solution.

Note that x and b are vector representations of tensors in $\mathbb{R}^{n_1 \times \cdots \times n_d}$, and b represents a rank-one tensor.

A tensor arising from the discretization of a sufficiently smooth function f can be approximated by a short sum of rank-one tensors. By superposition, we can insert such a tensor as right-hand side into our system.

About this system

The eigenvalues of the matrix $\mathcal{A}$ are given by all possible sums

$$\lambda_{i_1}^{(1)} + \lambda_{i_2}^{(2)} + \dots + \lambda_{i_d}^{(d)}$$

where $\lambda_{i_s}^{(s)}$ denotes an eigenvalue of A_s. For A_s positive definite, the system has a unique solution.

Note that x and b are vector representations of tensors in $\mathbb{R}^{n_1 \times \cdots \times n_d}$, and b represents a rank-one tensor.

A tensor arising from the discretization of a sufficiently smooth function f can be approximated by a short sum of rank-one tensors. By superposition, we can insert such a tensor as right-hand side into our system.

Tensorized Krylov subspaces

A Krylov subspace is defined as

$$\mathcal{K}_k(A, b) = \operatorname{span}\{b,\ Ab,\ \dots,\ A^{k-1}b\}.$$

We define a tensorized Krylov subspace as

$$\mathcal{K}^{\otimes}_{\mathfrak{K}}(\mathcal{A}, b) := \operatorname{span}\big(\mathcal{K}_{k_1}(A_1, b_1) \otimes \cdots \otimes \mathcal{K}_{k_d}(A_d, b_d)\big).$$

Note that

$$\mathcal{K}_{k_0}(\mathcal{A}, b) \subset \mathcal{K}^{\otimes}_{\mathfrak{K}}(\mathcal{A}, b)$$

for $\mathfrak{K} = (k_0, \dots, k_0)$.
$\Rightarrow$ Tensorized Krylov subspaces are richer than standard Krylov subspaces.

Tensorized Krylov subspaces

A Krylov subspace is defined as

$$\mathcal{K}_k(A, b) = \operatorname{span}\{b,\ Ab,\ \ldots,\ A^{k-1}b\}.$$

We define a tensorized Krylov subspace as

$$\mathcal{K}^{\otimes}_{\mathfrak{K}}(\mathcal{A}, b) := \operatorname{span}\big(\mathcal{K}_{k_1}(A_1, b_1) \otimes \cdots \otimes \mathcal{K}_{k_d}(A_d, b_d)\big).$$

Note that

$$\mathcal{K}_{k_0}(\mathcal{A}, b) \subset \mathcal{K}^{\otimes}_{\mathfrak{K}}(\mathcal{A}, b)$$

for $\mathfrak{K} = (k_0, \ldots, k_0)$.
$\Rightarrow$ Tensorized Krylov subspaces are richer than standard Krylov subspaces.

Tensorized Krylov subspaces

A Krylov subspace is defined as

$$\mathcal{K}_k(A, b) = \text{span}\{b,\ Ab,\ \ldots,\ A^{k-1}b\}.$$

We define a tensorized Krylov subspace as

$$\mathcal{K}^{\otimes}_{\mathfrak{K}}(\mathcal{A}, b) := \mathrm{span}\big(\mathcal{K}_{k_1}(A_1, b_1) \otimes \cdots \otimes \mathcal{K}_{k_d}(A_d, b_d)\big).$$

Note that

$$\mathcal{K}_{k_0}(\mathcal{A}, b) \subset \mathcal{K}^{\otimes}_{\mathfrak{K}}(\mathcal{A}, b)$$

for $\mathfrak{K} = (k_0, \ldots, k_0)$.

$\Rightarrow$ Tensorized Krylov subspaces are richer than standard Krylov subspaces.

Outline

Reminder: CG method ($Ax = b$)

The best approximation of x in $\mathcal{K}_k(A, b)$ is defined by:

$$\|x_k - x\|_A = \min_{\tilde{x} \in \mathcal{K}_k(A,b)} \|\tilde{x} - x\|_A.$$

Find U_k with (orthonormal) columns that span the Krylov subspace $\mathcal{K}_k(A, b)$. Set $x_k = U_k y$, then y is the solution of the compressed system

$$H_k y = \tilde{b}, \qquad H_k = U_k^\top A U_k, \tilde{b} = U_k^\top b.$$

Convergence bound ($\kappa = \kappa_2(A)$)

$$\|x_k - x\|_A \leq C(A, b, \kappa)\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^k.$$

Reminder: CG method ($Ax = b$)

The best approximation of x in $\mathcal{K}_k(A, b)$ is defined by:

$$\|x_k - x\|_A = \min_{\tilde{x} \in \mathcal{K}_k(A,b)} \|\tilde{x} - x\|_A.$$

Find U_k with (orthonormal) columns that span the Krylov subspace $\mathcal{K}_k(A, b)$. Set $x_k = U_k y$, then y is the solution of the compressed system

$$H_k y = \tilde{b}, \qquad H_k = U_k^\top A U_k, \tilde{b} = U_k^\top b.$$

Convergence bound ($\kappa = \kappa_2(A)$)

$$\|x_k - x\|_A \leq C(A, b, \kappa)\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^k.$$

Reminder: CG method ($Ax = b$)

The best approximation of x in $\mathcal{K}_k(A, b)$ is defined by:

$$\|x_k - x\|_A = \min_{\tilde{x} \in \mathcal{K}_k(A,b)} \|\tilde{x} - x\|_A.$$

Find U_k with (orthonormal) columns that span the Krylov subspace $\mathcal{K}_k(A, b)$. Set $x_k = U_k y$, then y is the solution of the compressed system

$$H_k y = \tilde{b}, \qquad H_k = U_k^\top A U_k, \tilde{b} = U_k^\top b.$$

Convergence bound ($\kappa = \kappa_2(A)$)

$$\|x_k - x\|_A \leq C(A, b, \kappa)\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^k.$$

Tensorized Krylov: $\mathcal{A}\,\mathrm{vec}(\mathcal{X}) = b$

Applying the Arnoldi method to A_s, b_s results in U_s, H_s with

$$U_s^\top A_s U_s = H_s,$$

U_s column-orthogonal, H_s upper Hessenberg matrix.

The columns of U_s span $\mathcal{K}_{k_s}(A_s, b_s)$, and similarly the columns of $\mathcal{U} := U_1 \otimes \cdots \otimes U_d$ span $\mathcal{K}_{\mathfrak{K}}^{\otimes}(\mathcal{A}, b)$.

Solve the compressed system

$$\mathcal{H}y = \tilde{b}$$

with $x_{\mathfrak{K}} = \mathcal{U}y$, $\tilde{b} = \mathcal{U}^\top b$ and $\mathcal{H} = \mathcal{U}^\top \mathcal{A}\mathcal{U}$.

Tensorized Krylov: $\mathcal{A}\,\mathrm{vec}(\mathcal{X}) = b$

Applying the Arnoldi method to A_s, b_s results in U_s, H_s with

$$U_s^\top A_s U_s = H_s,$$

U_s column-orthogonal, H_s upper Hessenberg matrix.

The columns of U_s span $\mathcal{K}_{k_s}(A_s, b_s)$, and similarly the columns of $\mathcal{U} := U_1 \otimes \cdots \otimes U_d$ span $\mathcal{K}_{\mathfrak{K}}^{\otimes}(\mathcal{A}, b)$.

Solve the compressed system

$$\mathcal{H} y = \tilde{b}$$

with $x_{\mathfrak{K}} = \mathcal{U} y$, $\tilde{b} = \mathcal{U}^\top b$ and $\mathcal{H} = \mathcal{U}^\top \mathcal{A} \mathcal{U}$.

Tensorized Krylov: $\mathcal{A}\,\mathrm{vec}(\mathcal{X}) = b$

Applying the Arnoldi method to A_s, b_s results in U_s, H_s with

$$U_s^\top A_s U_s = H_s,$$

U_s column-orthogonal, H_s upper Hessenberg matrix.

The columns of U_s span $\mathcal{K}_{k_s}(A_s, b_s)$, and similarly the columns of $\mathcal{U} := U_1 \otimes \cdots \otimes U_d$ span $\mathcal{K}^{\otimes}_{\mathfrak{K}}(\mathcal{A}, b)$.

Solve the compressed system

$$\mathcal{H}y = \tilde{b}$$

with $x_{\mathfrak{K}} = \mathcal{U}y$, $\tilde{b} = \mathcal{U}^\top b$ and $\mathcal{H} = \mathcal{U}^\top \mathcal{A}\mathcal{U}$.

Compressed system

The structure of $\mathcal{H}y = \tilde{b}$ corresponds to that of $\mathcal{A}x = b$:

$$\mathcal{H} = \sum_{s=1}^{d} I_{k_1} \otimes \cdots \otimes I_{k_{s-1}} \otimes H_s \otimes I_{k_{s+1}} \otimes \cdots \otimes I_{k_d}$$

$$\tilde{b} = \tilde{b}_1 \otimes \cdots \otimes \tilde{b}_d = \|b\|_2 \, (e_1 \otimes \cdots \otimes e_1)$$

For dense core tensor y, $x_{\mathfrak{K}}$ is a tensor in Tucker decomposition:

$$x_{\mathfrak{K}} = (U_1 \otimes \cdots \otimes U_d)y$$

Outline

Convergence, s.p.d. case (1)

Similarly to CG, we have:

$$\|x_{\mathfrak{K}} - x\|_{\mathcal{A}} = \min_{\tilde{x} \in \mathcal{K}^{\otimes}_{\mathfrak{K}}(\mathcal{A}, b)} \|\tilde{x} - x\|_{\mathcal{A}}$$

Every vector in a Krylov subspace $\mathcal{K}_{k_s}(A_s, b_s)$ can be seen as $p(A_s)b_s$, with p a polynomial of order at most $k_s - 1$. A similar thought reduces the bound calculation to the min-max problem

$$E_{\Omega}(\mathfrak{K}) := \min_{p \in \Pi^{\otimes}_{\mathfrak{K}}} \|p(\lambda_1, \dots, \lambda_d) - \frac{1}{\lambda_1 + \cdots + \lambda_d}\|_{\Omega},$$

where $\Pi^{\otimes}_{\mathfrak{K}}$ is a space of multivariate polynomials, and Ω contains the eigenvalues $(\lambda_1, \dots, \lambda_d)$, with $\lambda_i \in [\lambda_{\min}(A_i), \lambda_{\max}(A_i)]$.

Convergence, s.p.d. case (2)

Inserting an upper bound for $E_\Omega(\mathfrak{K})$, we find

$$\|x_{\mathfrak{K}} - x\|_{\mathcal{A}} \leq \sum_{s=1}^{d} C(\mathcal{A}, b, \kappa_s)\left(\frac{\sqrt{\kappa_s} - 1}{\sqrt{\kappa_s} + 1}\right)^{k_s},$$

with $\kappa_s = 1 + \frac{\lambda_{\max}(A_s) - \lambda_{\min}(A_s)}{\lambda_{\min}(\mathcal{A})}$.

For the case A_s, k_s constant:

$$\|x_{\mathfrak{K}} - x\|_{\mathcal{A}} \leq C(\mathcal{A}, b, d)\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^{k},$$

with $\kappa = \frac{d-1}{d} + \frac{\kappa_2(\mathcal{A})}{d}$.

Note that the convergence rate improves with increasing dimension.

Convergence, non-symmetric positive definite case

General convergence bound:

$$\|x_{\mathfrak{K}} - x\|_2 \leq \sum_{s=1}^{d} \int_0^{\infty} e^{-\hat{\alpha}_s t} \, \|U_s e^{-tH_s} e_1 - e^{-tA_s} b_s\|_2 \, \mathrm{d}t$$

with $\hat{\alpha}_s := \sum_{j \neq s} \alpha_j$ and $\alpha_j = \lambda_{\min}(A_j + A_j^\top)/2$.

The bound on $\|U_s e^{-tH_s} e_1 - e^{-tA_s} b_s\|_2$ will depend on additional knowledge on A_s.

For example, when the field of values of each matrix A_s is contained in a known ellipse, an explicit convergence bound can be found.

Outline

Grasedyck's method, system $\mathcal{H}y = \tilde{b}$

For $\mathcal{H}$ positive definite:

$$\mathcal{H}^{-1} = \int_0^\infty \exp(-t\mathcal{H})dt$$

The exponential of $\mathcal{H}$ has a Kronecker product structure, too:

$$\exp(-t\mathcal{H}) = \exp(-t\sum_{s=0}^{d} \hat{H}_s) = \prod_{s=1}^{d} \exp(-t\hat{H}_s) = \bigotimes_{s=1}^{d} \exp(-tH_s),$$

Approximation y_m:

$$y_m = \sum_{j=1}^{m} \omega_j \bigotimes_{s=1}^{d} \exp\Big(-\alpha_j H_s\Big)\tilde{b}_s,$$

with certain coefficients α_j, ω_j.

Coefficients of the exponential sum (1)

The coefficients α_j, ω_j should minimize

$$\sup_{z \in \Lambda(\mathcal{H})} \left| \frac{1}{z} - \sum_{j=1}^{m} \omega_j e^{-\alpha_j z} \right|$$

Case 1: $\mathcal{H}$ symmetric and condition number known
The eigenvalues of $\mathcal{H}$ are real: $\Lambda(\mathcal{H}) \subset [\lambda_{\min}, \lambda_{\max}]$. There are coefficients α_j, ω_j s.t.

$$\|y - y_m\|_2 \le C(\mathcal{H}, \tilde{b}) \exp\left(\frac{-m\pi^2}{\log(8\kappa_2(\mathcal{H}))} \right).$$

These coefficients may be found using a variant of the Remez algorithm. Tabellated values have been made available by Hackbusch.

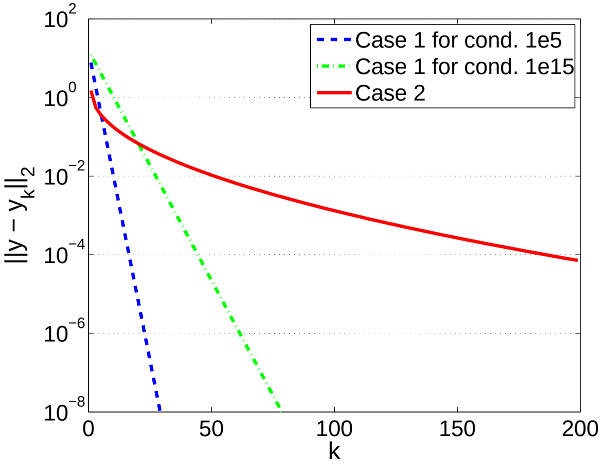
Coefficients of the exponential sum (2)

Case 2: Nonsymmetric $\mathcal{H}$ and/or unknown condition number
There is an explicit formula for α_j, ω_j, where the eigenvalues of $\mathcal{H}$ only need to have positive real part.

$$\|y - y_{2m+1}\|_2 \leq C(\mathcal{H}, \tilde{b}) \exp(\mu/\pi) \exp(-\sqrt{m}),$$
$$\text{where } \mu = \max\{|\Im m(\Lambda(\mathcal{H}))|\}.$$

The convergence is significantly slower than for Case 1.

Theoretical convergence bounds for the coefficients



Outline

Symmetric example: Poisson Equation

Finite difference discretization of the Poisson equation in d dimensions:

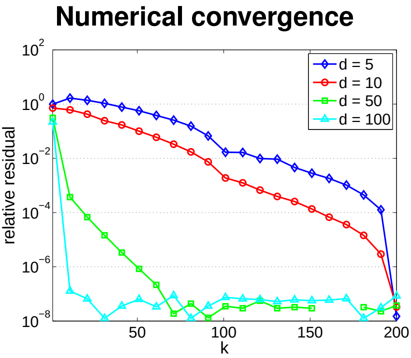
$$\begin{aligned} \Delta u &= f \quad \text{in } \Omega = [0,1]^d \\ u &= 0 \quad \text{on } \Gamma := \partial\Omega, \end{aligned}$$

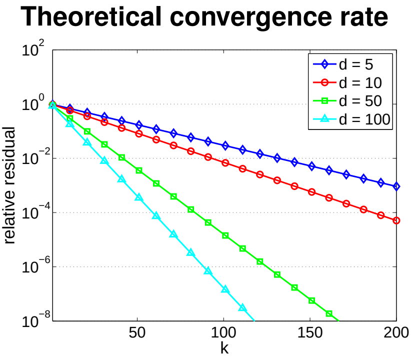
where the right-hand side f is a separable function.

$$A_s = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}, \; b_s : \text{random numbers.}$$

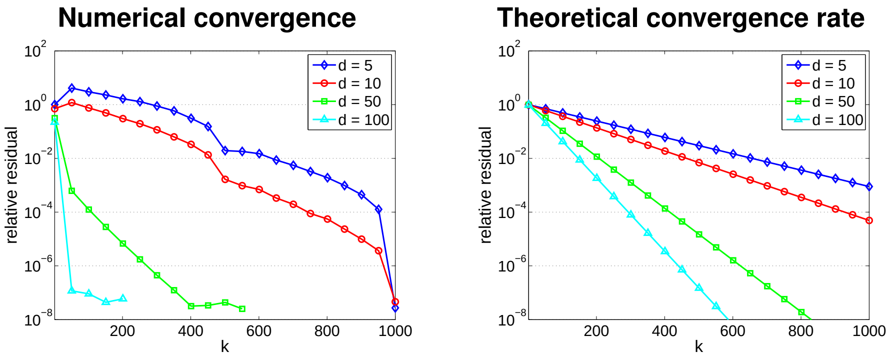
The approximation error is measured by *relative residual*, $\frac{\|\mathcal{A}x_{\mathfrak{R}} - b\|_2}{\|b\|_2}$.

Convergence for system size 200^d





Convergence for system size 1000^d



Extended Krylov subspaces

Using *extended Krylov subspaces*

$$\widetilde{\mathcal{K}}_{k_s}(A_s, b_s) := \mathrm{span}(\mathcal{K}_{k_s}(A_s, b_s) \cup \mathcal{K}_{k_s+1}(A_s^{-1}, b_s)),$$

the algorithm works analogously.
Convergence bound (A_s, k_s constant):

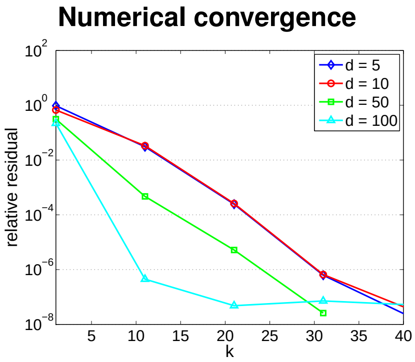
$$\|x_{\mathfrak{K}} - x\|_2 \leq C(\mathcal{A}, b, d)\left(\frac{\sqrt{\tilde{\kappa}} - 1}{\sqrt{\tilde{\kappa}} + 1}\right)^k.$$

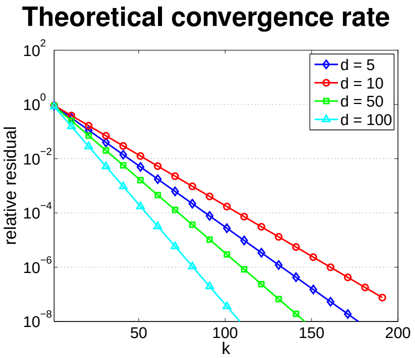
The convergence rate depends on $\tilde{\kappa}$:

$$\tilde{\kappa} \approx \frac{d-1}{d} + \frac{1}{d^2}(d-1)^{\frac{d-1}{d}} \kappa_2(\mathcal{A})^{\frac{d-1}{d}}$$

$$d = 2 : \tilde{\kappa} = \frac{1}{2} + \frac{\sqrt{\kappa_2(\mathcal{A})}}{4}, \qquad d \gg 0 : \tilde{\kappa} \approx \frac{d-1}{d} + \frac{\kappa_2(\mathcal{A})}{d}.$$

Extended Krylov, system size 200^d





Non-symmetric example

Finite difference discretization in d dimensions of the convection-diffusion equation

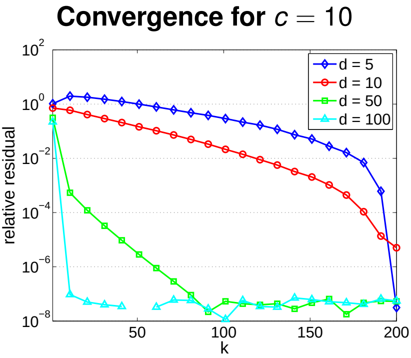
$$\begin{aligned} \Delta u - (c, \ldots, c)\nabla u &= f \quad \text{in } \Omega = [0,1]^d \\ u &= 0 \quad \text{on } \Gamma := \partial\Omega, \end{aligned}$$

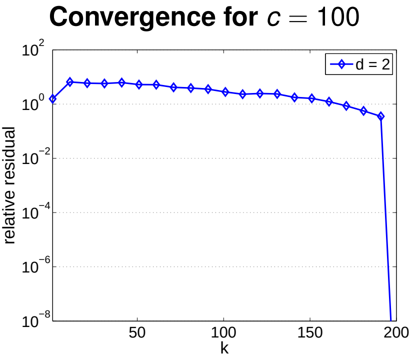
where f is again a separable function.

$$A_s = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} + \frac{c}{4h}\begin{pmatrix} 3 & -5 & 1 & & \\ 1 & 3 & -5 & \ddots & \\ & \ddots & \ddots & \ddots & 1 \\ & & 1 & 3 & -5 \\ & & & 1 & 3 \end{pmatrix}.$$

For $c = 10$, all eigenvalues of $\mathcal{A}$ are real. For $c = 100$, this is not the case, and the convergence is significantly worsened.

Non-symmetric case, system size 200^d





Conclusions

- An efficient algorithm to calculate a low-rank approximation of the solution tensor
- The computational complexity is linear in the number of dimensions
- Only matrix-vector operations with the full system matrices are required
- A theoretical convergence bound was found.

Conclusions

- An efficient algorithm to calculate a low-rank approximation of the solution tensor
- The computational complexity is linear in the number of dimensions
- Only matrix-vector operations with the full system matrices are required
- A theoretical convergence bound was found.

Thank you for your attention!

Literature

D. Kressner, C. Tobler, “Krylov subspace methods for linear systems with tensor product structure”, *Research report No. 2009-16, SAM, ETH Zurich* , 2009.

L. Grasedyck, “Existence and computation of low Kronecker-rank approximations for large linear systems of tensor product structure”, *Computing*, 72(3-4):247-265, (2004).

D. Braess, W. Hackbusch, “Approximation of 1/x by exponential sums in $[1, \infty)$”, *IMA J. Numer. Anal.*, 25(4):685-697, (2005).