Eigenvalue solvers based on TT and QTT formats

Olga Lebedeva

Institute of Numerical Mathematics, Russian Academy of Sciences

May 3, 2010

Problem statement

We consider a standard partial eigenproblem

$$MX = X\Lambda \tag{1}$$

for a symmetric positive-definite matrix $M = M^T > 0$.

We wish to approximate $k = 5 \dots 30$ lowest eigenpairs
($X \in \mathbb{R}^{n \times k}$, Λ is a diagonal matrix).

The matrix and its eigenvectors are supposed to posses
an approximate low-parametric structure
based on the QTT-decomposition.

QTT-format

Let $n = 2^d$. For a vector

- Separate indexes:
  replace a "long" index $i = 1 \dots 2^d$
  with a set of "short" indexes $i_1, \dots, i_d = 1, 2$
- Apply the TT-approximation

$$X_{i_1,\dots,i_d} \approx \sum_{\alpha_1,\dots,\alpha_{d-1}}^{r_1,\dots,r_{d-1}} g_1(i_1,\alpha_1) g_2(\alpha_1, i_2, \alpha_2) \dots g_{d-1}(\alpha_{d-2}, i_{d-1}, \alpha_{d-1}) g_d(\alpha_{d-1}, i_d)$$

The same way for a matrix

$$M \approx \sum_{\alpha_1,\dots,\alpha_{d-1}}^{R_1,\dots,R_{d-1}} G_1(i_1, j_1, \alpha_1) G_2(\alpha_1, i_2, j_2, \alpha_2) \dots G_d(\alpha_{d-1}, i_d, j_d).$$

If all mode ranks r_i are bounded the required memory is $O(\log n)$!

Required linear algebra operations

TT-Toolbox *http://spring.inm.ras.ru/osel*

- Matrix-by-vector multiplication

$$r_i(Mx) = R_i(M)r_i(x);$$

- Linear combinations of vectors

$$r_i(\alpha x + \beta y) = r_i(x) + r_i(y);$$

- Scalar multiplication of vectors.

\+ Run-time $O(\log(n))$

\+ QTT-format remains

– Ranks grow!

Compression procedures

For a given vector x in QTT-format to obtain a QTT-approximation with smaller mode ranks.

- Rank minimization with a fixed tolerance

$$\|x - P_\varepsilon x\| < \varepsilon \|x\|$$

- Strict boundaries for mode ranks

$$r_i(P_r x) \leqslant r_{\max}, \ i = 1, \ldots, d-1$$

- Combination

$$P_{r\varepsilon} x = \begin{cases} P_\varepsilon x, \ r_i\left(P_\varepsilon x\right) \leqslant r_{\max} \ \forall\, i \\ P_r x, \ r_q\left(P_\varepsilon x\right) > r_{\max} \end{cases}$$

Also,

- Krylov-based compression procedures for Mx;
- ALS-based compression procedures for $\sum_i x_i$.

TT-Toolbox *http://spring.inm.ras.ru/osel*

The modified iteration method with structure

Suppose we have an iterative method

$$x_{i+1} = F(x_i),\ \|x_0\| = 1,\ (x_0, u) \neq 0 \tag{2}$$

with linear covergence rate

$$\|x_{i+1} - u\| \leqslant \gamma \|x_i - u\|\ (\gamma < 1).$$

We will replace all vectors by their normalized QTT-approximations:

$$\tilde{x}_{i+1} = \frac{P(F(\tilde{x}_i))}{\|P(F(\tilde{x}_i))\|},\ \tilde{x}_0 = x_0. \tag{3}$$

This may lead to divergence.

Convergence theory. Generalization

Let L be some abstract set of structured vectors, that includes the solution u.

P is a projector to L, such that

$$\forall\, x :\ x \neq 0 \ \Rightarrow\ P(x) \neq 0$$

$$\left(\text{for QTT-format } L_r = \{x \,|\, rank_i(x) \leqslant r,\ i = 1 \ldots d-1\}\ \right)$$

'Approximate iterations for structured matrices'
Wolfgang Hackbusch, Boris N. Khoromskij, Eugene E. Tyrtyshnikov

Convergence theory. Generalization

We can also consider $L_{\|\cdot\|} = \{x \,|\, x \in L,\ \|x\| = 1\}$ and a projector $P_{\|\cdot\|}$ to $L_{\|\cdot\|}$.

Lemma

Assume that the inclusion $u \in L$ implies that $\alpha u \in L$ for all α.
Let the norm $\|\cdot\|$ be generated by a scalar product. Then, the relation

$$P_{\|\cdot\|}(x) = \frac{P(x)}{\|P(x)\|}$$

holds for every $x \neq 0$.

The modified method (3) will be represented as

$$\tilde{x}_{i+1} = P_{\|\cdot\|}(F(\tilde{x}_i)),\ \tilde{x}_0 = x_0. \tag{4}$$

The covergence can be proved for several cases.

Convergence theorems

Theorem 1

Let the convergence factor of the original linear method (2) satisfy the condition $\gamma < 1/2$. Assume that the only requirements for the initial approximation x_0 are the normalization condition and the nonorthogonality to the exact solution $u \in L$. Then, with the same initial approximation, the modified iterative method (4) converges to the exact solution linearly with the convergence factor $\tilde{\gamma} \leqslant 2\gamma$.

Corollary

Assume that the original iterative method (2) converges linearly with the convergence factor $\gamma < 1/2$ until the δ-neighborhood of u is attained, where $\delta < \sqrt{2}/2$. Also, assume that the only requirements for the initial approximation x_0 are the normalization condition and the nonorthogonality to the exact solution. Then, with the same initial approximation, the modified iterative method (4) converges to the exact solution u linearly with the convergence factor $\tilde{\gamma} \leqslant 2\gamma$ until the 2δ-neighborhood of the solution is attained.

Convergence theorems

Inexact case: $\rho(u, L) > 0$.

Theorem 2

Let the exact solution u belong to the δ-neighborhood of a structured vector; that is, let $\|P(u) - u\| < \delta$, where $\delta < \sqrt{3}/2$. Assume that the original iterative method (2) converges linearly with the convergence factor $\gamma < 1/3$. Also, assume that the only requirements for the initial approximation x_0 are the normalization condition and the nonorthogonality to the exact solution. Then, with the same initial approximation, the modified iterative method (4) converges to the exact solution u linearly with the convergence factor $\tilde{\gamma} \leqslant 3\gamma$ until the 3δ-neighborhood of u is attained. If the iterations are continued further, the process may diverge; however, the subsequent approximations do not leave the 3δ-neighborhood of the solution.

Convergence theorems

And what if $\gamma > 1/2\,(1/3)$?

- Start with structured x_0;
- Implement $k-1$ non-modified iterations;
- At the k-th iteration perform the appriximation:

$$x_1 = F(x_0),\ \ldots,\ x_{k-1} = F(x_{k-2}),\ x_k = \frac{P(F(x_{k-1}))}{\|P(F(x_{k-1}))\|} \tag{5}$$

- Repeate the approximation periodically at each k-th step .

Convergence theorems

Theorem 3

Assume that the exact solution u belongs to the set L.
Let $k > log_{1/\gamma}2$ ($log_{1/\gamma}3$). Then, method (5) converges linearly with the average (over k steps) convergence factor $\tilde{\gamma} \leqslant 2^{1/k}\gamma$ ($3^{1/k}\gamma$) .

Corollary

Assume that the exact solution u belongs to the δ-neighborhood of a structured vector from the set L; that is, assume $\|u - P(u)\| < \delta$, where $\delta < \sqrt{2}/3$. Let $k > log_{1/\gamma}2$. Then, method (5) with the approximation performed at each k-th step converges linearly with the average (over k steps) convergence factor $\tilde{\gamma} \leqslant 3^{1/k}\gamma$ until the 3δ-neighborhood of the exact solution is attained.

Convergence theorems

Consider operator P_ε that minimizes QTT-ranks under the condition

$$\|P_\varepsilon(x) - x\| \leqslant \varepsilon\|x\|$$

P_ε is not projector for any set L sinse $P_\varepsilon(P_\varepsilon(x)) \neq P_\varepsilon(x)$..

Theorem 4

Assume that original method (2) converges linearly with the convergence factor $\gamma < 1/3$. Also, assume that the only requirements for the initial approximation x_0 are the normalization condition and the nonorthogonality to the exact solution. Let $\varepsilon < \sqrt{2}/3$. Then, with the same initial approximation, the ε-modified iterative method (3) converges linearly with the convergence until the 3ε-neighborhood of u is attained. If the iterations are continued further, the process may diverge; however, the subsequent approximations do not leave the 3ε-neighborhood of the solution.

Convergence theorems

Consider operator $P_{\varepsilon,r}$:

$$P_{\varepsilon,r}(x) = P_r\left(P_\varepsilon(x)\right)$$

Also

$$P_{r\varepsilon}x = \begin{cases} P_\varepsilon x, \;\; r_i\left(P_\varepsilon x\right) \leqslant r_{\max} \;\forall\, i \\ P_r x, \;\; r_q\left(P_\varepsilon x\right) > r_{\max} \end{cases}$$

Theorem 5

Assume that original method (2) converges linearly with the convergence factor $\gamma < 1/3$. Also, assume that the only requirements for the initial approximation x_0 are the normalization condition and the nonorthogonality to the exact solution. Let $\varepsilon < \sqrt{2}/3$. Then, with the same initial approximation, the modified iterative method (3) with projector $P_{\varepsilon,r}(x)$ converges linearly with the convergence until the 3ε-neighborhood of u is attained. If the iterations are continued further, the process may diverge; however, the subsequent approximations do not leave the 3ε-neighborhood of the solution.

The Rayleigh quotient

The scalar case

$$\lambda_{\min} = \min_{x \neq 0} \frac{(Mx, x)}{(x, x)}$$

$$u_{\min} = \arg\min_{x \neq 0} \frac{(Mx, x)}{(x, x)}$$

Subspace approach

$$U = \arg \min_{X^T X = I^{k \times k}} \mathrm{tr}(X^T M X)$$

$$\min_{X^T X = I^{k \times k}} \mathrm{tr}(X^T M X) = \sum_{i=1}^{k} \lambda_i$$

The tensor conjugate gradient method in the scalar case

$$\min_{x \in X_i} \frac{(Mx, x)}{(x, x)}$$

How to construct the subspaces X_i?

The gradient-type methods.

since

$$\nabla \frac{(Mx, x)}{(x, x)} = \frac{2}{(Mx, x)} \left(Mx - \frac{(Mx, x)}{(x, x)} x \right) = \frac{2}{(Mx, x)} r(x)$$

we search for x_{i+1} in $\mathrm{span}\,(x_i, r_i)$:

$$x_{i+1} = arg \min_{x \in \mathrm{span}\,(x_i, r_i)} \frac{(Mx, x)}{(x, x)}$$

The tensor conjugate gradient method in the scalar case

The conjugate gradient-type methods.

Minimization within $\mathrm{span}\,(x_i, r_i, x_i - x_{i-1})$:

$$x_{i+1} = arg \min_{x \in \mathrm{span}\,(x_i, r_i, x_{i-1})} \frac{(Mx, x)}{(x, x)}, \ \lambda_i = \frac{(Mx_{i+1}, x_{i+1})}{(x_{i+1}, x_{i+1})}$$

'Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method' Andrew V. Knyazev

Note that

$$\min_{x \in \mathrm{span}\,(x_i, r_i, x_{i-1})} \frac{(Mx, x)}{(x, x)} \leqslant \min_{x \in \mathrm{span}\,(x_i, r_i)} \frac{(Mx, x)}{(x, x)}$$

The tensor conjugate gradient method. Convergance

Steepest decrease
One-step method

$$x_{i+1} = F(x_i)$$

Modified version

$$\tilde{x}_{i+1} = P(F(\tilde{x}_i))$$

Conjugate gradient
Two-step method

$$x_{i+1} = G(x_i, x_{i-1})$$

Modified version

$$\tilde{x}_{i+1} = P(G(\tilde{x}_i, \tilde{x}_{i-1}))$$

Since $\|G(\tilde{x}_i, \tilde{x}_{i-1}) - u\| \leqslant \|F(\tilde{x}_i) - u\|$ we obtain

$$\tilde{\gamma}^{cg} \leqslant \tilde{\gamma}^{sd} \leqslant 2\gamma^{sd} \ (3\gamma^{sd})$$

The tensor conjugate gradient method. Scalar version

Algorithm

- Choose initial x_0, $||x_0|| = 1$
- $\lambda_0 = (Mx_0, x_0)$, $r_0 = (Mx_0 - \lambda_0 x_0)/||Mx_0 - \lambda_0 x_0||$
- Rayleigh-Ritz procedure for $\mathrm{span}\,(x_0, r_0)$.
  Get the lowest Ritz value λ_1 and the corresponding Ritz vector $x_1 = [x_0, r_0]y$.
- $r_1 = (Mx_1 - \lambda_1 x_1)/||Mx_1 - \lambda_1 x_1||$
- Introduce $p_1 = [0, r_0]y = x_1 - (x_0, x_1)x_0$
- Main loop:
  - Rayleigh-Ritz procedure for $\mathrm{span}\,(x_i, r_i, p_i)$.
    Get the lowest Ritz value λ_{i+1} and the corresponding Ritz vector $x_{i+1} = [x_i, r_i, p_i]y$.
  - $r_{i+1} = (Mx_{i+1} - \lambda_{i+1} x_{i+1})/||Mx_{i+1} - \lambda_{i+1} x_{i+1}||$
  - $p_{i+1} = [0, r_i, p_i]y$

The tensor conjugate gradient method. Block version

Simultanious searching for k lowest eigenpairs.

- If k is rather large, we split the task and use deflations.
- k current approximations of eigenvectors are the columns of an orthogonal $n \times k$ matrix X_i.
- The Rayleigh-Ritz procedure for $3k$-dimensional subspace $\mathrm{span}\,(X_i, R_i, P_i)$.
  Obtaining k Ritz vectors.
- The columns of R_i, P_i should be orthogonalized.
- The columns of R_i, P_i are no longer orthogonal to the columns of X_i.

The tensor conjugate gradient method. Block version

Algorithm

- Choose initial X_0: $X_0^T X_0 = I^{k \times k}$
- $\Lambda_0 = \mathrm{diag}\,(X_0^T M X_0)$, $R_0 = MX_0 - X_0\Lambda_0$
- Orthogonalization of R_0
- Rayleigh-Ritz procedure for $\mathrm{span}\,(X_0, R_0)$.
  Get the k lowest Ritz values Λ_1
  and the corresponding Ritz vectors X_1.
- $R_1 = MX_1 - X_1\Lambda_1$
  Orthogonalize of R_1
- Introduce $P_1 = R_0$
- **Main loop**

The tensor conjugate gradient method. Block version

- **Main loop**
  - Rayleigh-Ritz procedure for $\mathrm{span}\,(X_i, R_i, P_i)$
    Get the k lowest Ritz values Λ_{i+1}
    and the corresponding Ritz vectors $X_{i+1} = [X_i, R_i, P_i]Y$
  - $R_{i+1} = MX_{i+1} - X_{i+1}\Lambda_{i+1}$
    $(R_{i+1})_{(j)} \perp (X_{i+1})_{(j)}$, but $\mathrm{span}\,(R_{i+1})$ and $\mathrm{span}\,(X_{i+1})$ are not orthogonal
  - Apply deflation to R_{i+1}: $R_{i+1} = (I - XX^T)R_{i+1}$
  - Orthogonalize the colunms of R_{i+1}
  - $P_{i+1} = [0, R_i, P_i]Y$
  - Orthogonalize the colunms of P_{i+1}

The tensor conjugate gradient method. Block QTT version

- **Main loop**
  - Rayleigh-Ritz procedure for $\mathrm{span}\,(X_i, R_i, P_i)$
    Get the k lowest Ritz values Λ_{i+1}
    and the Ritz vectors in QTT-format $X_{i+1} = [X_i, R_i, P_i]Y$
    Ranks grow
  - **Compress the columns of X_{i+1}**
  - $R_{i+1} = MX_{i+1} - X_{i+1}\Lambda_{i+1}$
    Ranks grow
  - Apply deflation to R_{i+1}: $R_{i+1} = (I - XX^T)R_{i+1}$ *Ranks grow*
  - Orthogonalize the colunms of R_{i+1}
    Ranks grow
  - **Compress the columns of R_{i+1}**
  - $P_{i+1} = [0, R_i, P_i]Y$ *Ranks grow*
  - Orthogonalize the colunms of P_{i+1}
    Ranks grow
  - **Compress the columns of P_{i+1}**

Block QTT

Block operations such as orthogonalization, block-by-matrix multiplication lead to the additional increase of tensor ranks of vectors.

Label the number of column in the block j an additional dimension of tensor.

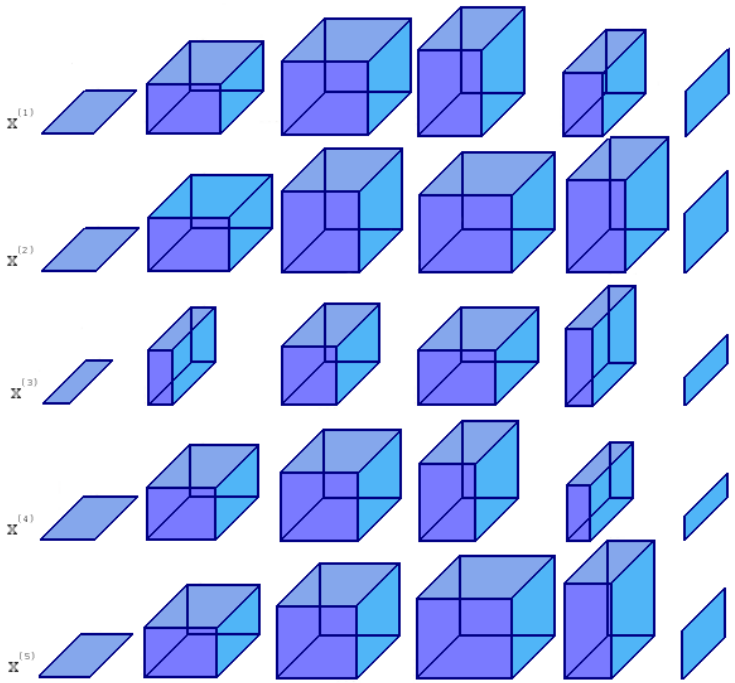
Apply TT-decomposition to this tensor:

$$X_{(i_1,\ldots,i_d),j} \approx \sum_{\alpha_1,\ldots,\alpha_{d-1},\alpha}^{r_1,\ldots,r_{d-1},r} g_1(i_1,\alpha_1)\ldots g_d(\alpha_{d-1},i_d,\alpha)\, g_{d+1}(\alpha,j).$$

Block QTT

$$X_{(i_1,\ldots,i_d),j} \approx \sum_{\alpha} \underbrace{\sum_{\alpha_1,\ldots,\alpha_{d-1}} g_1(i_1,\alpha_1)\ldots g_d(\alpha_{d-1},i_d,\alpha)}_{\textit{common base}}\ g_{d+1}(\alpha,j).$$

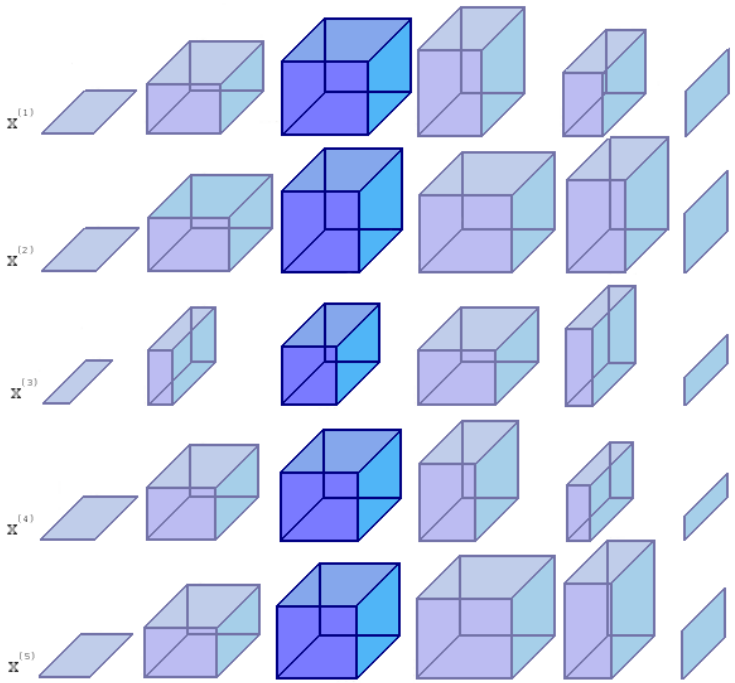
Common base may be composed from the columns of $X = [x^{(1)}, x^{(2)}, \ldots, x^{(k)}]$



Block QTT

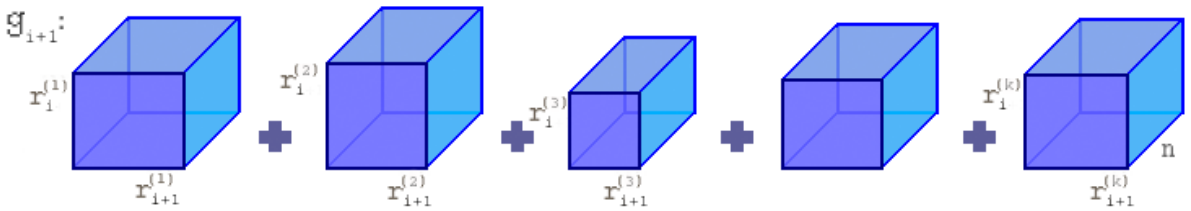
$$X_{(i_1,\ldots,i_d),j} \approx \sum_{\alpha} \underbrace{\sum_{\alpha_1,\ldots,\alpha_{d-1}} g_1(i_1,\alpha_1)\ldots g_d(\alpha_{d-1},i_d,\alpha)}_{\textit{common base}}\ g_{d+1}(\alpha,j).$$

Common base may be composed from the columns of $X = [x^{(1)}, x^{(2)}, \ldots, x^{(k)}]$

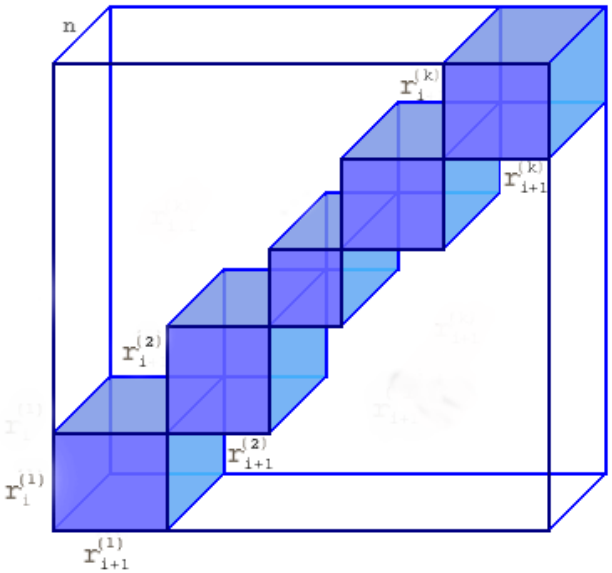


Block QTT

Consider the $(i+1)$-th core of each decomposition $(i = 1 \ldots d-3)$

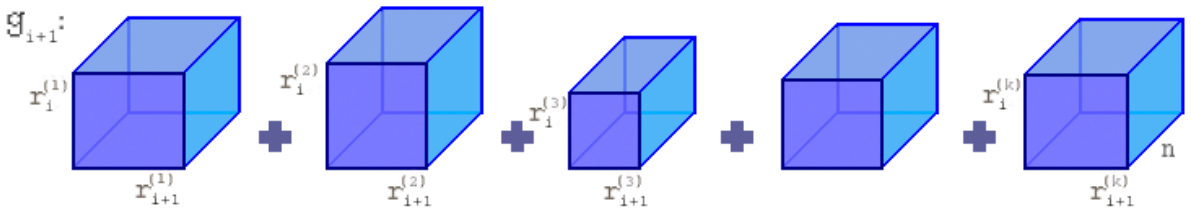


We will put them together in a common core.

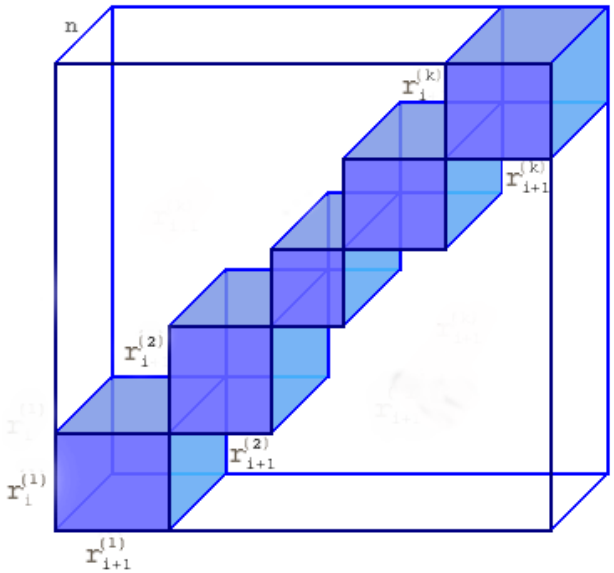


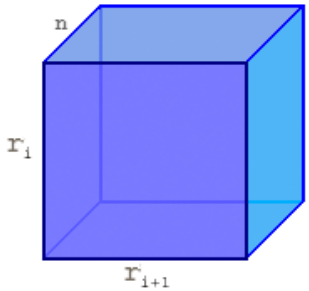
Block QTT

Consider the $(i+1)$-th core of each decomposition $(i = 1 \ldots d-3)$



We will put them together in a common core. And compress.





Block QTT operations

- Quick orthogonalization without ranks growth (left-to-right QR sweep);
- Quick orthogonalization with compression (3 sweeps);
- Linear combinations of the columns

$$X_{(i_1,\ldots,i_d),j} W_{j,l} \approx \sum_{\alpha=1}^{r} \left(\sum_{\alpha_1,\ldots,\alpha_{d-1}}^{r_1,\ldots,r_{d-1}} g_1(i_1,\alpha_1)\ldots g_d(\alpha_{d-1},i_d) \right) g_{d+1}(\alpha,j) W(j,l);$$

- Block operations $X^T Y$, MX;
- Cut a column, add a column, and others.

Stopping criterion

- Residual norm

$$\|r_i^{it}\|/\lambda_i^{it} \leqslant \varepsilon_1$$

(since $|(Mx_i^{it}, x_i^{it})/(x_i^{it}, x_i^{it}) - \lambda_i| \leqslant \|r_i\|$)

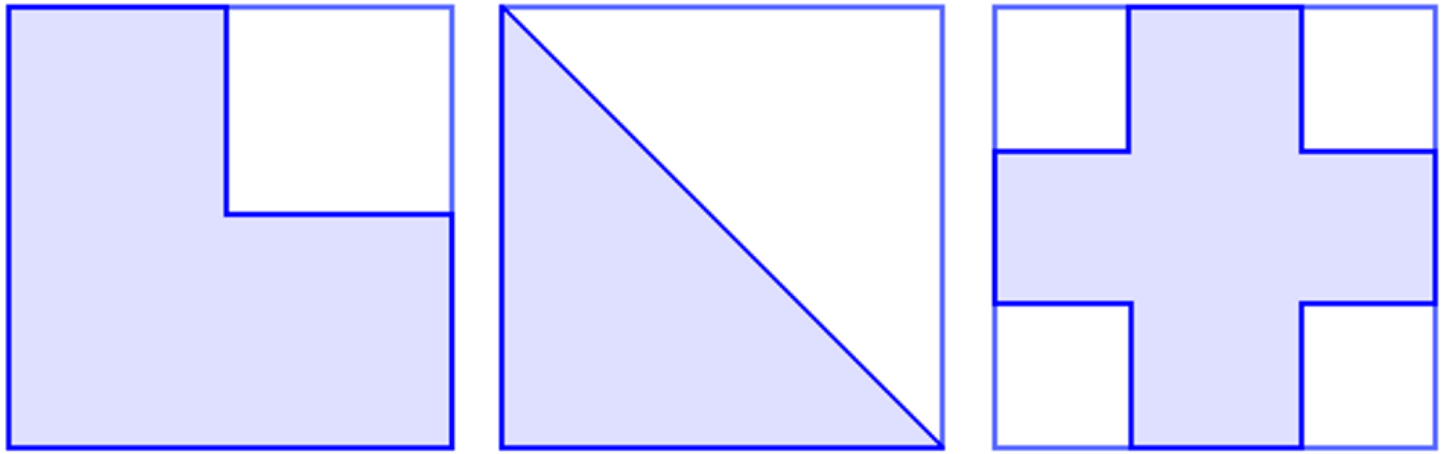
- Eigenvalue difference

$$|\lambda_i^{it+1} - \lambda_i^{it}|/\lambda_i^{it} \leqslant \varepsilon_2.$$

Numerical experiments

Finite difference or finite element discretizations in the following domains

2-dimensional case

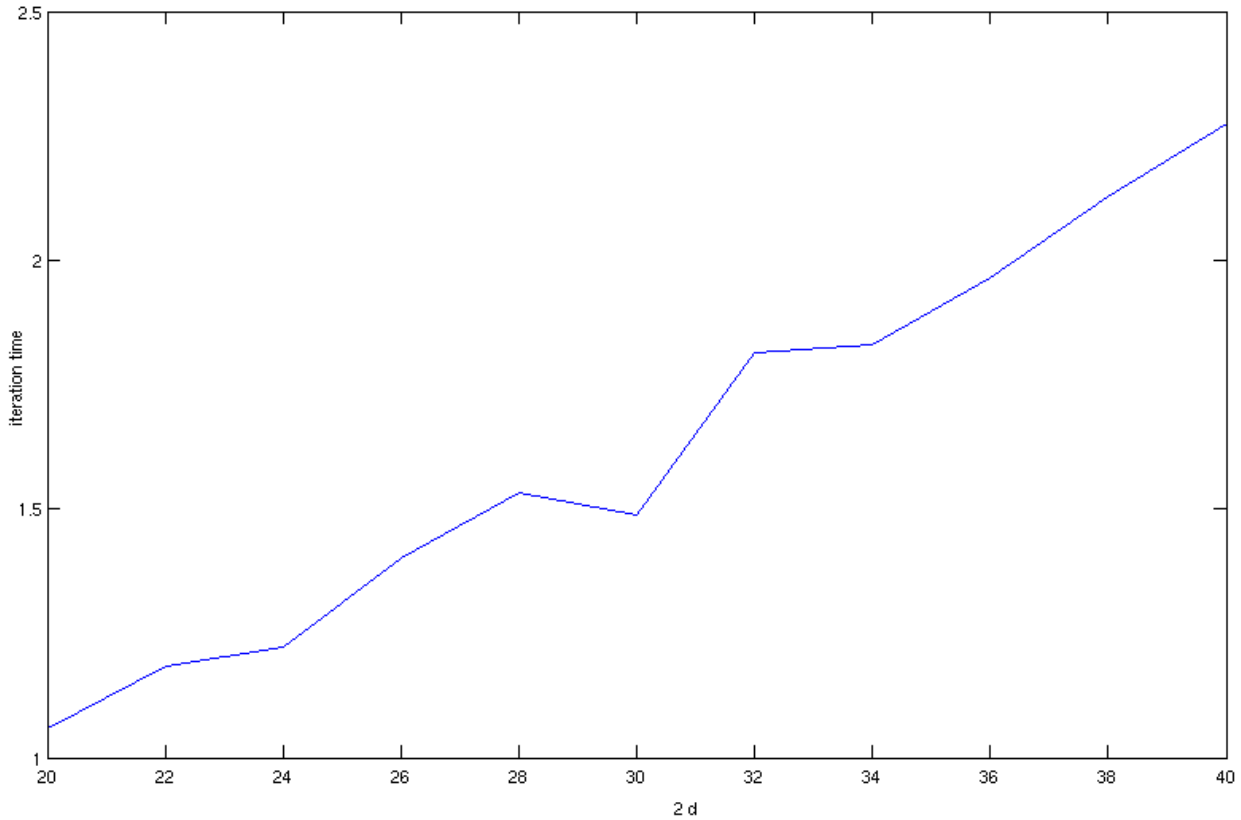


3-dimensional case

Parallelepipeds, polyhedral regions, their unions and intersections

Numerical experiments

Average iteration time



Discrete Laplace operator in 2-dimensional rectangular domain $n \times n$ ($n = 2^d, d = 10, \ldots, 20$)
vector length n^2 ($2^{20}, 2^{22}, \ldots, 2^{40}$), block size 5, $r_{\max} = 25$

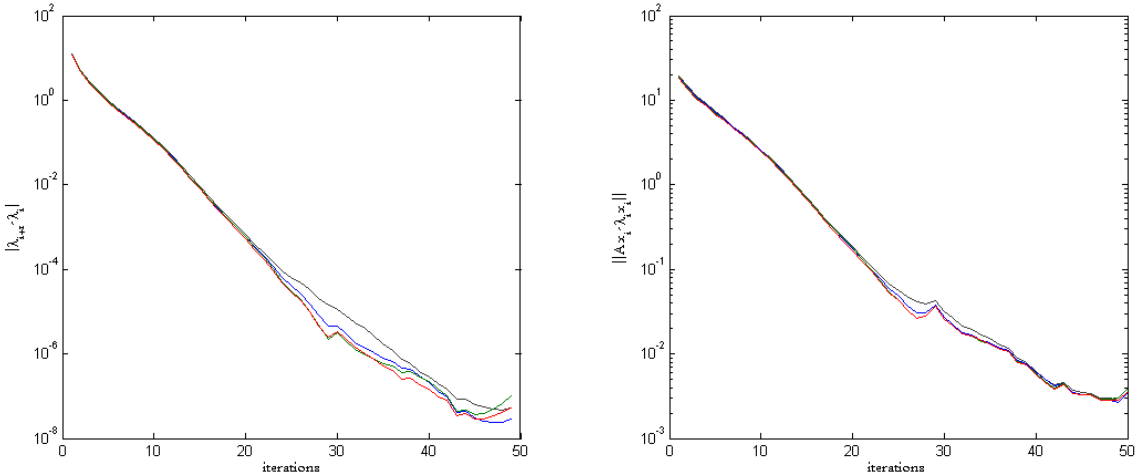
Numerical examples

Harmonic oscillator

$$-\Delta u + \|r\|^2 u = \lambda u$$

in 3-dimensional cube $[0,1] \times [0,1] \times [0,1]$, homogeneous Dirichlet boundary conditions.

$n = 2^5$ points each direction ($N = 2^{15}$)



Left picture: $|\lambda^{it+1} - \lambda^{it}|$; right picture: $\|r_i\|$.

Numerical example

Anharmonic oscillator

$$-\Delta u + (\alpha_1 x^2 + \alpha_2 y^2 + \alpha_3 x^4 + \alpha_4 y^4 + \alpha_5 x^2 y^2)u = \lambda u \ (\alpha_3, \alpha_4, \alpha_5 < \alpha_1, \alpha_2)$$

Discretization in 2-dimensional rectangular domain $[-5, 5] \times [-5, 5]$
$(n = 2^d, d = 6, \ldots, 10)$

Harmonic oscillator

$$-\Delta u + \alpha_1 (x^2 + y^2)u = \lambda u$$

We used the solution of harmonic oscillator as an initial approximation for anharmonic oscillator.

Numerical example

Anharmonic oscillator

$$-\Delta u + (\alpha_1 x^2 + \alpha_2 y^2 + \alpha_3 x^4 + \alpha_4 y^4 + \alpha_5 x^2 y^2)u = \lambda u \ (\alpha_3, \alpha_4, \alpha_5 < \alpha_1, \alpha_2)$$

Discretization in 2-dimensional rectangular domain $[-5, 5] \times [-5, 5]$
$(n = 2^d, d = 6, \ldots, 10)$

Harmonic oscillator

$$-\Delta u + \alpha_1(x^2 + y^2)u = \lambda u$$

We used the solution of harmonic oscillator as an initial approximation for anharmonic oscillator.

Numerical example

Covergence of eigenvalues under the mesh size.

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
$d = 3$	1.9486	3.4531	3.4609	5.8688	5.9841	5.9887
$d = 4$	2.0690	4.1186	4.1200	6.0632	6.1612	6.1688
$d = 5$	2.1142	4.2444	4.2553	6.3544	6.3640	6.4955
$d = 6$	2.1172	4.2698	4.2765	6.4156	6.4212	6.5783
$d = 7$	2.1210	4.2815	4.2879	6.4421	6.4459	6.6039
$d = 8$	2.1222	4.2870	4.2906	6.3371	6.4657	6.6173

Thank you for your attention!