Exponential Integrators using Matrix Functions: Krylov Subspace Methods and Chebyshev Expansion approximations

The HPC Approach

México, February 2018

Marlon Brenes
brenesnm@tcd.ie
https://www.tcd.ie/Physics/research/groups/qusys/people/navarro/

Outline

- Exponential Integrators
- Brief Introduction: Chebyshev expansion for matrix functions
- Brief Introduction: Krylov subspace techniques
- HPC Approach:
    - Relevance and importance of an HPC approach
    - Parallelisation strategy
- Outlook

Exponential Integrators

Matrix functions

The problem

- Consider a problem of the type

$$\frac{d\mathbf{w}(t)}{dt} = \mathbf{A}\mathbf{w}(t),\ t \in [0, T]$$

$$\mathbf{w}(0) = \mathbf{v},\ \text{initial condition}$$

- Its analytic solution is $\mathbf{w}(t) = e^{t\mathbf{A}}\mathbf{v}$

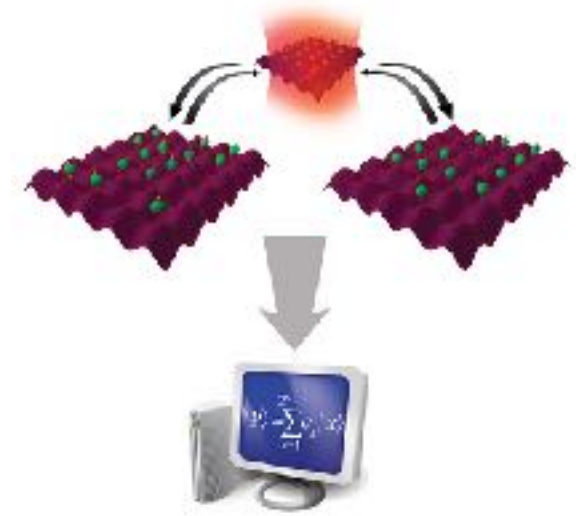
- Things to consider:

    - $\mathbf{A}$ is a matrix

    - The exponential of the matrix is not really required, but merely it's **action** on the vector $\mathbf{v}$

R. B. Sidje, ACM Trans. Math. Softw. 24, 130 (1998).

Exponential Integrators



- Mathematical models of many physical, biological and economic processes systems of linear, constant-coefficient ordinary differential equations

- Growth of microorganisms, population, decay of radiation, control engineering, signal processing…

- More advanced: MHD (magnetohydrodynamics), quantum many-body problems, reaction-advection-diffusion equations…

Moler, C. Van Loan, C. SIAM Rev. 45, 1 (2003).

Numerical approach

- The idea to use exponential functions for the matrix is **not new**
  - Was considered impractical…
- The development of Krylov subspace techniques to the **action** of the matrix exponential substantially changed the landscape
- Different types of solution evaluation for matrix exponentials:
  - ODE methods: numerical integration
  - Polynomial methods
  - Matrix decomposition methods

Moler, C. Van Loan, C. SIAM Rev. 45, 1 (2003).

Hochbruck, M. Lubich, C. Selhofer, H. SIAM J. Sci. Comp. 19, 5 (1998).

Numerical approach

- We're interested in the case where $\mathbf{A}$ is **large** and **sparse**
- A *sole* implementation may not be reliable for *all* types of problems
- Chebyshev expansion approach
- The technique of Krylov subspaces has been proven to be very efficient for **many** classes of problems
- Convergence is faster than applying the solution to linear systems in both techniques

Moler, C. Van Loan, C. SIAM Rev. 45, 1 (2003).

Hochbruck, M. Lubich, C. Selhofer, H. SIAM J. Sci. Comp. 19, 5 (1998).

Take-home message #1:
There's a big number of problems in science and engineering that can be tackled using exponential integrators and matrix functions

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

The Abdus Salam
International Centre
for Theoretical Physics

Polynomial approximation: Chebyshev expansion

A brief introduction

Definition

- We intend to employ a polynomial expansion as an approximation

- Let us start with a definition of the matrix exponential by convergent power series:

$$e^{t\mathbf{A}} = \mathbf{I} + t\mathbf{A} + \frac{t^2\mathbf{A}^2}{2!} + \dots$$

- An effective computation of the **action** of this operator on a vector is the main topic of this talk

Moler, C. Van Loan, C. SIAM Rev. 45, 1 (2003).

Chebyshev expansion

- We intend to employ a **good** converging polynomial expansion
- Explicit computation of $e^{\mathbf{A}t}$ **has** to be avoided
- **Key component**: Efficient matrix-vector product operations
- **Upside**: Efficient parallelisation and vectorisation, extremely simple approach
- **Downside**: Requires computation of two eigenvalues, not so versatile

Chebyshev polynomials

- The polynomials are defined by a three-term recursion relationship in the interval $x \in [-1, 1]$:

$$T_0(x) = 1$$

$$T_1(x) = x$$

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$$

- $T_n(x)$ constitutes an orthogonal basis, therefore one can write:

$$f(x) = \sum_{n=0}^{\infty} b_n T_n(x) \approx \sum_{n=0}^{N} b_n T_n(x)$$

- with $b_n = \frac{2 - \delta_n}{\pi} \int_{-1}^{1} \frac{f(y)T_n(y)}{\sqrt{1 - y^2}} dy$

Bessel coefficients for the particular case of the exponential function

Sharon, N. Shkolniski, Y. arXiv preprint arXiv:1507.03917 (2016)

Kosloff, R. Annu. Rev. Phys. Chem. 45, 145-78. (1994)

Chebyshev polynomials

- We're interested in applying the approximation to the action of the operator on a vector
- First step: Find the extremal eigenvalues of $\mathbf{A}$, more on this later

$$\lambda_{min} \quad \lambda_{max}$$

- Second step: Rescale operator such that it's spectrum is bounded by $[-1, 1]$

$$\mathbf{A}^{'} = 2\frac{\mathbf{A} - \lambda_{min}\mathbf{I}}{\lambda_{max} - \lambda_{min}} - \mathbf{I}$$

- Third step: Use the Chebyshev recursion relation

$$f(t\mathbf{A}')\mathbf{v} = e^{t\mathbf{A}'}\mathbf{v} \approx \sum_{n=0}^{N} b_n T_n(t\mathbf{A}')\mathbf{v}$$

- The recursion can be truncated up to a desired tolerance

Sharon, N. Shkolniski, Y. arXiv preprint arXiv:1507.03917 (2016)

Kosloff, R. Annu. Rev. Phys. Chem. 45, 145-78. (1994)

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

ICTP The Abdus Salam International Centre for Theoretical Physics

Chebyshev recursion relation

- The recursion relation

$$f(t\mathbf{A}')\mathbf{v} = e^{t\mathbf{A}'}\mathbf{v} \approx \sum_{n=0}^{N} b_n T_n(t\mathbf{A}')\mathbf{v}$$

- Goes as follows:

$$\phi_0 = \mathbf{v}$$
$$\phi_1 = t\mathbf{A}'\mathbf{v}$$
$$\phi_{n+1} = 2t\mathbf{A}'\phi_n - \phi_{n-1}$$

- Then:

$$f(t\mathbf{A}')\mathbf{v} \approx \sum_{n=0}^{N} b_n \phi_n$$

- Until desired tolerance

Sharon, N. Shkolniski, Y. arXiv preprint arXiv:1507.03917 (2016)

Kosloff, R. Annu. Rev. Phys. Chem. 45, 145-78. (1994)

Chebyshev polynomials

- Why do we choose the Chebyshev polynomials as basis set? Why not another polynomial set?

- Because of the asymptotic property of the Bessel function!

    - When the order n of the polynomial becomes larger than the argument, the function decays **exponentially** fast

    - This means that in order to obtain a good approximation, an exponentially decreasing amount of terms are required in the expansion as a function of the argument (related to t, λ_{min} and λ_{max})

Sharon, N. Shkolniski, Y. arXiv preprint arXiv:1507.03917 (2016)

Kosloff, R. Annu. Rev. Phys. Chem. 45, 145-78. (1994)

Take-home message #2:
The Chebyshev expansion approach provides a numerically stable and scalable approach at the cost of some restrictions of the problem

Krylov subspace techniques to evaluate the solution

A brief introduction

Krylov subspace techniques

- We intend to employ a combination of a Krylov subspace technique and other known methods for matrix exponential
- Explicit computation of $e^{\mathbf{A}t}$ **has** to be avoided
- **Key component**: Efficient matrix-vector product operations
- **Upside**: Extremely versatile
- **Downside**: Storage of the subspace for large problems, “time scales”

Main idea

- Building a Krylov subspace of dimension m

$$\mathcal{K}_m = \mathrm{span}\{v, \mathbf{A}v, \mathbf{A}^2 v, \ldots, \mathbf{A}^{m-1} v\}$$

- The idea is to approximate the solution to the problem by an element of $\mathcal{K}_m$

- In order to manipulate the subspace, it's convenient to generate an orthonormal basis

$$V_m = [v_1, v_2, \ldots, v_m] \qquad v_1 = v/||v||_2$$

- This can be achieved with the **Arnoldi algorithm**

Gallopoulos, E. Saad, Y. ICS. 89', 17—28 (1989).

Moler, C. Van Loan, C. SIAM Rev. 45, 1 (2003).

Algorithm: Arnoldi

1. *Initialize:* Compute $v_1 = v/\|v\|_2$.
2. *Iterate:* Do $j = 1, 2, ..., m$
   (a) Compute $w := Av_j$
   (b) Do $i = 1, 2, \ldots, j$
      i. Compute $h_{i,j} := (w, v_i)$
      ii. Compute $w := w - h_{i,j}v_i$
   (c) Compute $h_{j+1,j} := \|w\|_2$ and $v_{j+1} := w/h_{j+1,j}$.

- Step 2-b is a modified Gram-Schmidt process.
- **Lanczos** can be applied for the case of symmetric matrices

Gallopoulos, E. Saad, Y. ICS. 89', 17—28 (1989).

Moler, C. Van Loan, C. SIAM Rev. 45, 1 (2003).

Krylov subspace techniques

- The Arnoldi procedure produces a basis V_m of $\mathcal{K}_m$ and an upper Hesseberg matrix $\mathbf{H}_m$ of dimension $m \text{ x } m$ with coefficients h_{ij}

- We start by the relation given by

$$\mathbf{A}V_m = V_m\mathbf{H}_m + h_{m+1,m}v_{m+1}e_m^T$$

- Where v_{m+1} satisfies $V_m^T v_{m+1} = 0$ and $e_m \in I_m$

- From which we obtain

$$\mathbf{H}_m = V_m^T \mathbf{A} V_m$$

- Therefore, $\mathbf{H}_m$ is the projection of the linear transformation $\mathbf{A}$ onto the Krylov subspace $\mathcal{K}_m$ with respect to the basis V_m

Gallopoulos, E. Saad, Y. ICS. 89', 17—28 (1989).

Moler, C. Van Loan, C. SIAM Rev. 45, 1 (2003).

Krylov subspace techniques

- Given that $\mathbf{H}_m$ is a projection of the linear operator, an approximation can be made such that

$$e^{t\mathbf{A}}v \approx ||v||_2 V_m e^{t\mathbf{H}_m} e_1$$

- The approximation is **exact** when the dimension of the Krylov subspace is equal to the dimension of the linear transformation

- Error

$$||e^{t\mathbf{A}}v - ||v||_2 V_m e^{t\mathbf{H}_m} e_1|| \leq 2||v||_2 \frac{(t||A||_2)^m e^{t||A||_2}}{m!}$$

Gallopoulos, E. Saad, Y. ICS. 89’, 17—28 (1989).

Moler, C. Van Loan, C. SIAM Rev. 45, 1 (2003).

Krylov subspace techniques

- With this approach:
  - A **large sparse** matrix problem is approximated by a **small dense** matrix problem
- There are several methods to evaluate the small dense matrix exponential
  - Series methods, ODE methods, diagonalisation, matrix decomposition methods…
  - Padè methods
- This method can be used to compute λ_{min} and λ_{max} for the Chebyshev approach, very effective

Moler, C. Van Loan, C. SIAM Rev. 45, 1 (2003).

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

The Abdus Salam
International Centre
for Theoretical Physics

Take-home message #3:
Krylov subspace methods to evaluate the solution provides a more versatile and less restricted approach, at the expense of higher computational cost and memory consumption

HPC Approach

Relevance and importance

- A platform for numerical calculations
  - Important for current research
- Undertake simulations efficiently
- Quantum physics, CFD, finite-element methods…
- Establish an instance where HPC approach has been used recently in scientific research

Parallelisation strategy

$$y = \mathbf{A}x$$

- The linear operator is **sparse** (low density) and **huge** (big dimension, i.e, large amount of degrees of freedom

- Approach:
  - **Distribute the operator among processing elements**

Take-home message #4:
Your numerical evaluation of the exponential integrator using matrix functions is only going to be **as good as** your implementation of the sparse matrix-vector product

Outlook

- Two different approximation methods to target the solutions to matrix exponentials, which can be interpreted analytically as solutions to a particular set of differential equations
- We will discuss the practical approach using techniques, practices and libraries from the HPC perspective next session…

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

The Abdus Salam
International Centre
for Theoretical Physics

Thank you!