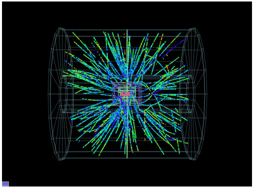
A numerical method to compute derivatives of functions of large complex matrices

Matthias Puhr, Pavel Buividovich

Universität Regensburg

Motivation



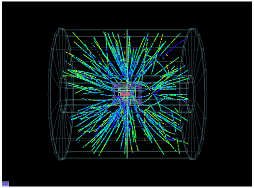
Copyright: CERN

- Conserved lattice currents for chiral fermions
- Study anomalous transport in dense QCD
- Example:
  Chiral Separation Effect

$$j_i^A = \sigma_{\text{CSE}} B_i, \qquad \boxed{\sigma_{\text{CSE}} = \frac{1}{2\pi^2}\mu}$$

- Important to have
  - Dirac operator that preserves chiral symmetry
  - Finite density (quenched approximation)

Motivation



- Conserved lattice currents for chiral fermions
- Study anomalous transport in dense QCD
- Example:
  Chiral Separation Effect

$$j_i^A = \sigma_{\mathrm{CSE}} B_i, \qquad \boxed{\sigma_{\mathrm{CSE}} = \frac{1}{2\pi^2}\mu} \quad ???$$

- Important to have
  - Dirac operator that preserves chiral symmetry
  - Finite density (quenched approximation)

Motivation

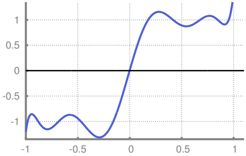
- Overlap Dirac operator at finite chemical potential μ[a]:

$$D_{ov} = \frac{1}{a}\left(\mathbb{1} + \gamma_5 \operatorname{sgn}[\gamma_5 D_W(\mu)]\right)$$

- Wilson Dirac operator:

$$\gamma_5 D_W(\mu)\gamma_5 = D_W^\dagger(-\mu)$$

- Sign function is numerically challenging
  - Polynomial/partial fraction approximation
  - Krylov subspace methods (finite μ)



[a] J. Bloch and T. Wettig, Phys.Rev.Lett.97:012003,2006

Motivation

Why derivatives?

- Conserved lattice currents
  - Anomalous transport, conductivity, charge diffusion, ...
  - Derivatives over (background) gauge fields:

$$j_\nu^V(x) = \langle \bar{\psi} \frac{\partial \, \mathrm{D}_{\mathrm{ov}}}{\partial \theta_\nu(x)} \psi \rangle \qquad \dots \quad \theta_\nu \text{ background g.f.}$$

- Dynamical HMC Simulations ($\mu = 0$)
  - Derivatives of Dirac operator needed to evaluate the fermionic force

Matrix Sign Function

- Sign function for complex numbers: $\operatorname{sgn}(z) = \frac{z}{\sqrt{z^2}} = \operatorname{sgn}(\operatorname{Re}(z))$
- Generalisation to matrices:
  - **Spectral form:** (λ_i eigenvalues of $\mathbf{A}$)

$$\operatorname{sgn}(\mathbf{A}) = \mathbf{U}\operatorname{sgn}(\mathbf{\Lambda})\mathbf{U}^{-1}, \quad \operatorname{sgn}(\mathbf{\Lambda}) := \operatorname{diag}(\operatorname{sgn}(\lambda_1), \ldots, \operatorname{sgn}(\lambda_n))$$

  - **Roberts' iteration**:

$$\mathbf{X}_{k+1} := \frac{1}{2}\left(\mathbf{X}_k + \mathbf{X}_k^{-1}\right), \quad \mathbf{X}_0 = \mathbf{A}$$

- Approximation necessary
- Derivative of the approximation algorithm?

Numerical derivatives of matrix functions

- Theorem by R. Mathias[b]:
  Let $\mathbf{A}(t) \in \mathbb{C}^{n\times n}$ be differentiable at $t=0$ and assume that the spectrum of $\mathbf{A}(t)$ is contained in an open subset $\mathcal{D} \subset \mathbb{C}$ for all t in some neighbourhood of 0. Let f be $2n-1$ times continuously differentiable on $\mathcal{D}$. We then have:

$$f\left(\begin{bmatrix} \mathbf{A}(0) & \partial_{\mathbf{t}}\mathbf{A}(0) \\ 0 & \mathbf{A}(0) \end{bmatrix}\right) \equiv \begin{bmatrix} f(\mathbf{A}(0)) & \frac{\partial}{\partial t}\Big|_{t=0} f(\mathbf{A}(t)) \\ 0 & f(\mathbf{A}(0)) \end{bmatrix}$$

- Compute the derivative of $f(\mathbf{A})$ without knowing $f'(\mathbf{A})$ explicitly!
- Works also for implicit approximation algorithms (Krylov subspace methods)

[b] R. Mathias, SIAM J. Matrix Anal. Appl., 17(3):610-620,1996

Numerical derivatives of matrix functions

- Useful for efficient numerical calculations?
- Advantage:
  - Compute derivative with any approximation algorithm:

$$\operatorname{sgn}\left(\begin{bmatrix} \mathbf{A} & \partial_{\mathbf{t}}\mathbf{A} \\ 0 & \mathbf{A} \end{bmatrix}\right)\begin{pmatrix} 0 \\ |\boldsymbol{\psi}\rangle \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial t}\operatorname{sgn}(\mathbf{A})\,|\boldsymbol{\psi}\rangle \\ \operatorname{sgn}(\mathbf{A})\,|\boldsymbol{\psi}\rangle \end{pmatrix}$$

- Disadvantage:
  - Size of linear space doubles
- Efficiency of approximation $\rightarrow$ spectrum of $\bar{\mathbf{A}} := \begin{bmatrix} \mathbf{A} & \partial_{\mathbf{t}}\mathbf{A} \\ 0 & \mathbf{A} \end{bmatrix}$

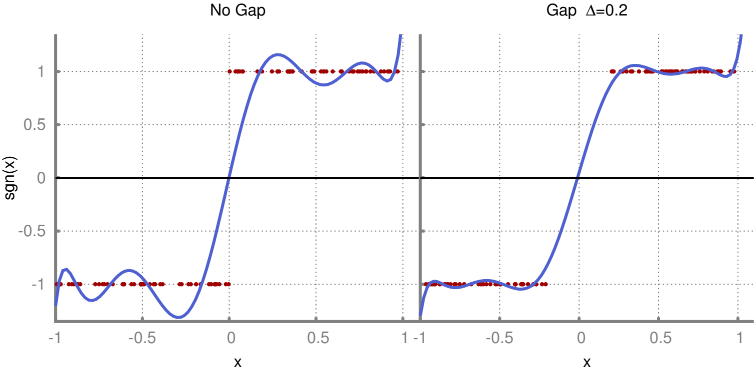
Numerical derivatives of matrix functions

- Useful for efficient numerical calculations?
- Advantage:
  - Compute derivative with any approximation algorithm:

$$\operatorname{sgn}\left(\begin{bmatrix} \mathbf{A} & \partial_{\mathbf{t}}\mathbf{A} \\ 0 & \mathbf{A} \end{bmatrix}\right)\begin{pmatrix} 0 \\ |\boldsymbol{\psi}\rangle \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial t}\operatorname{sgn}(\mathbf{A})|\boldsymbol{\psi}\rangle \\ \operatorname{sgn}(\mathbf{A})|\boldsymbol{\psi}\rangle \end{pmatrix}$$

- Disadvantage:
  - Size of linear space doubles
- Efficiency of approximation $\rightarrow$ spectrum of $\bar{\mathbf{A}} := \begin{bmatrix} \mathbf{A} & \partial_{\mathbf{t}}\mathbf{A} \\ 0 & \mathbf{A} \end{bmatrix}$

Polynomial interpolation (order 10) of the sign function for 100 random points in $(-1,-\Delta)\cup(\Delta,1)$



Left: $\Delta = 0.0$

Right: $\Delta = 0.2$

Deflation

- Idea: "Increase gap around 0"
- $\mathbf{A}$ diagonalisable, right/left eigenvectors $|R_i\rangle$ and $\langle L_i|$

$$f(\mathbf{A})\,|\psi\rangle = \sum_{i=1}^{n} f(\lambda_i)\,|R_i\rangle\,\langle L_i|\,\psi\rangle$$

- In practical calculations, with $\mathbf{P}_m^n := \sum_{m+1}^{n} |R_i\rangle\,\langle L_i|$:

$$\mathrm{sgn}(A)\,|\psi\rangle = \underbrace{\sum_{i=1}^{m} \mathrm{sgn}(\lambda_i)\,|R_i\rangle\,\langle L_i|\,\psi\rangle}_{\text{exact}} + \underbrace{\mathrm{sgn}(\mathbf{A})\mathbf{P}_m^n\,|\psi\rangle}_{\text{approximation}}$$

Deflation for derivative computation

- Problem: $\bar{\mathbf{A}}$ in general not diagonalisable
- Jordan decomposition:

$$\bar{\mathbf{A}} = \mathbf{X}\mathbf{J}\mathbf{X}^{-1}, \quad \mathbf{J} = \mathrm{diag}\left(\mathbf{J}_1, \cdots, \mathbf{J}_n\right), \quad \mathbf{J}_i = \begin{pmatrix} \lambda_i & 1 \\ 0 & \lambda_i \end{pmatrix}$$

- Generalisation of the spectral form of $f(\bar{\mathbf{A}})$:

$$f(\bar{\mathbf{A}}) = \mathbf{X}\,\mathrm{diag}(f(\mathbf{J}_1), \cdots, f(\mathbf{J}_n))\mathbf{X}^{-1}, \quad f(\mathbf{J}_i) = \begin{pmatrix} f(\lambda_i) & f'(\lambda_i) \\ 0 & f(\lambda_i) \end{pmatrix}$$

Deflation for derivative computation

- Problem: $\bar{\mathbf{A}}$ in general not diagonalisable
- Jordan decomposition:

$$\bar{\mathbf{A}} = \mathbf{X}\mathbf{J}\mathbf{X}^{-1}, \quad \mathbf{J} = \operatorname{diag}\left(\mathbf{J}_1, \cdots, \mathbf{J}_n\right), \quad \mathbf{J}_i = \begin{pmatrix} \lambda_i & 1 \\ 0 & \lambda_i \end{pmatrix}$$

- Generalisation of the spectral form of $f(\bar{\mathbf{A}})$:

$$f(\bar{\mathbf{A}}) = \mathbf{X}\operatorname{diag}(f(\mathbf{J}_1), \cdots, f(\mathbf{J}_n))\mathbf{X}^{-1}, \quad f(\mathbf{J}_i) = \begin{pmatrix} f(\lambda_i) & f'(\lambda_i) \\ 0 & f(\lambda_i) \end{pmatrix}$$

- Derivative of sign function vanishes $\Rightarrow$ $\operatorname{sgn}(\mathbf{J})$ is diagonal!

Deflation for derivative computation

- Analogous to diagonalisable case:

$$\operatorname{sgn}(\bar{\mathbf{A}})\,|\bar{\psi}\rangle = \sum_{i=1}^{2n} \operatorname{sgn}(\lambda(i))\,|\bar{R}_i\rangle\,\langle\bar{L}_i|\,\bar{\psi}\rangle\,, \quad \lambda(i) := \mathbf{J}_{ii}$$

- $|\bar{R}_i\rangle \leftrightarrow$ columns of $\mathbf{X}$ and $\langle\bar{L}_i| \leftrightarrow$ rows of $\mathbf{X}^{-1}$
- Known in terms of eigenvectors $\left|R_j\right\rangle$ of $\mathbf{A}$ and their derivatives $\left|\partial_t R_j\right\rangle$:

$$\left|\bar{R}_{(2j-1)}\right\rangle = \begin{pmatrix} \left|R_j\right\rangle \\ 0 \end{pmatrix} \qquad \left|\bar{R}_{(2j)}\right\rangle = \frac{1}{\partial_t \lambda_j} \begin{pmatrix} \left|\partial_t R_j\right\rangle \\ \left|R_j\right\rangle \end{pmatrix}$$

The method

- Use Mathias' theorem to compute the derivative:

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \partial_t \mathbf{A} \\ 0 & \mathbf{A} \end{bmatrix}, \qquad \mathrm{sgn}\left(\bar{\mathbf{A}}\right) \begin{pmatrix} 0 \\ |\psi\rangle \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial t} \mathrm{sgn}(\mathbf{A})\,|\psi\rangle \\ \mathrm{sgn}(\mathbf{A})\,|\psi\rangle \end{pmatrix}$$

- In practical calculations:

$$\mathrm{sgn}(\bar{\mathbf{A}}) \begin{pmatrix} 0 \\ |\psi\rangle \end{pmatrix} = \underbrace{\sum_{i=1}^{m} \mathrm{sgn}(\lambda_i) \left\{ \begin{pmatrix} |R_i\rangle \\ 0 \end{pmatrix} \langle \partial_t L_i | \psi \rangle + \begin{pmatrix} |\partial_t R_i\rangle \\ |R_i\rangle \end{pmatrix} \langle L_i | \psi \rangle \right\}}_{\text{exact}} + \underbrace{\mathrm{sgn}(\bar{\mathbf{A}}) \bar{\mathbf{P}}_{2m}^{2n} \begin{pmatrix} 0 \\ |\psi\rangle \end{pmatrix}}_{\text{approximation}}$$

- Technical details and pseudo-code implementation:
  MP, P. Buividovich [1604.08057]

Results

- Quenched $SU(3)$ configurations with Lüscher-Weisz action
  - $\{\beta = 8.45,\ V = 6 \times 18^3\}$, $\{\beta = 8.1,\ V = 14 \times 14^3\}$
  - Above and below T_c for deconfinement transition
- Nested Lanczos algorithm [c]
  - Outer Krylov size $\leftrightarrow$ order of interpolating polynomial
- Deflation for sgn (n_{sgn}) and derivative (n_{D})
- Error estimate:

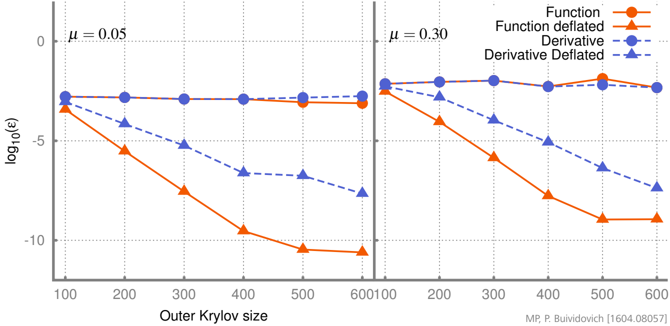
$$\mathrm{sgn}(A)^2 = \mathbb{1} \quad \rightarrow \quad \varepsilon := \frac{\| \mathrm{sgn}(A)^2 \, |\psi\rangle - |\psi\rangle \, \|}{2\|\psi\|}$$

[c] J. Bloch and S. Heybrock, Comput.Phys.Commun.182:878-889,2011

Results

Effect of deflation:

- $V = 14 \times 14^3$, $n_{\text{sgn}} = 40$, $n_{\text{D}} = 8$:



MP, P. Buividovich [1604.08057]

- Significant improvement of error even for small n_{D} !

Results

Chiral separation effect:

- Result for free chiral quarks:

$$j_i^A = \sigma_{\text{CSE}}^{\text{free}} \, B_i, \qquad \sigma_{\text{CSE}}^{\text{free}} = \frac{N_C \mu}{2\pi^2}$$

- Prediction for interacting theory[d]:
  - Free result if chiral symmetry restored
  - Chiral symmetry broken:

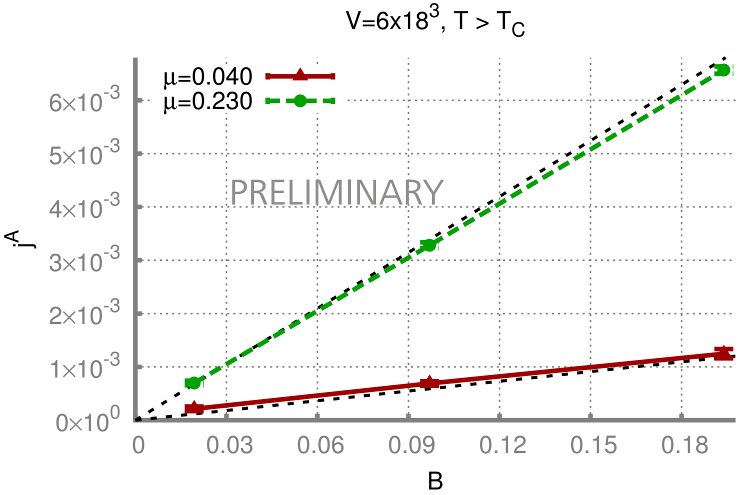
$$\sigma_{\text{CSE}} = \frac{N_C \mu}{2\pi^2} \left(1 - g_{\pi^0\gamma\gamma}\right) \qquad g_{\pi^0\gamma\gamma} : \text{"}\pi^0 \to 2\gamma \text{ amplitude"}$$

- On the lattice:

$$j_\nu^A(x) = \mathrm{tr}\left(\mathrm{D}_{\text{ov}}^{-1} \, \frac{\partial \, \mathrm{D}_{\text{ov}}}{\partial \Theta_\nu(x)} \gamma_5 \right)$$

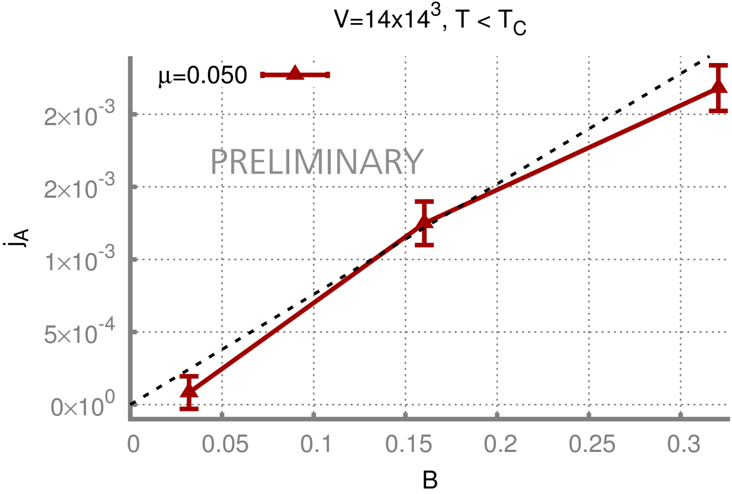
[d] G. M. Newman and D. T. Son, PRD73, 045006 (2006)

Results



V=6x18^3, T > T$_C$
μ=0.040
μ=0.230
PRELIMINARY
6×10^{-3}
5×10^{-3}
4×10^{-3}
3×10^{-3}
2×10^{-3}
1×10^{-3}
0×10^0
j^A
0
0.03
0.06
0.09
0.12
0.15
0.18
B

Results



V=14x14^3, T < T_C
μ=0.050
PRELIMINARY
j_A
2×10^-3
2×10^-3
1×10^-3
5×10^-4
0×10^0
0
0.05
0.1
0.15
0.2
0.25
0.3
B

Summary and Outlook

- **Summary**
  - Numerical derivatives of matrix functions
  - Deflation method for non-diagonalisable matrices
  - Test with the overlap Dirac operator
  - Efficiency significantly improved by deflation
- **Outlook**
  - Application to physical problems
  - Work on chiral separation effect in progress
  - Increase statistics to measure deviations from $\sigma_{\mathrm{CSE}}^{\mathrm{free}}$ at $T < T_C$