A Method to Calculate Conserved Currents and Fermionic Force for the Lanczos Approximation to the Overlap Dirac Operator

Matthias Puhr, Pavel Buividovich

Unterstützt von / Supported by

Alexander von Humboldt
Stiftung / Foundation

Motivation

- Derivative of lattice Dirac operator
  - Fermionic force in HMC
  - Conserved Currents
- Overlap Dirac operator:

$$\mathrm{D}_{\mathrm{ov}} = \frac{1}{a}\left(\mathbb{1} + \gamma_5 \mathrm{sgn}[\gamma_5 \mathrm{D_W}]\right)$$

- Numerically challenging
  - Polynomial/partial fraction approximation
  - Krylov subspace methods (finite μ)

Motivation

- Study anomalous transport in dense QCD
- Overlap Dirac operator at finite chemical potential[a]

$$\gamma_5 \, \mathrm{D}_{\mathrm{ov}}(\mu)\gamma_5 = \mathrm{D}_{\mathrm{ov}}(-\mu)$$

- Conserved currents $\leftrightarrow$ derivatives over (background) gauge fields:

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

- Compute approximation to sign function and its derivative for large matrices

---

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

Matrix Sign Function

- Sign function for complex numbers: $\mathrm{sgn}(z) = \frac{z}{\sqrt{z^2}} = \mathrm{sgn}(\Re(z))$
- Matrix sign function:
  - Spectral form: (λ_i eigenvalues of $\mathbf{A}$)

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

  - Roberts iteration:

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

    (Newton's method for $\mathbf{X}^2 - \mathbf{1} = \mathbf{0}$)
- Both methods numerically very expensive!

Two-sided Lanczos algorithm

- Compute approximation to $\vec{y} = f(\mathbf{A})\vec{x}, \quad \mathbf{A} \in \mathbb{C}^{n\times n}$
- Krylov subspace method:
  $\mathcal{K}_k(\mathbf{A},\vec{x}) := \mathrm{span}(\vec{x}, \mathbf{A}\vec{x}, \mathbf{A}^2\vec{x}, \ldots, \mathbf{A}^{k-1}\vec{x})$
- Approximate $f(\mathbf{A})$ by a polynomial of degree $k-1$
- Information about $\vec{x}$ taken into account
- Construct (biorthonormal) matrices $\mathbf{V}_k$ and $\mathbf{W}_k \in \mathbb{C}^{n\times k}$ such that
  $$\mathbf{A} = \mathbf{V}_k \mathbf{T}_k \mathbf{W}_k^\dagger, \qquad \mathbf{T}_k \in \mathbb{C}^{k\times k} \text{ tridiagonal}$$
- Compute
  $$\vec{y} \approx \mathbf{V}_k f(\mathbf{T}_k) \mathbf{W}_k^\dagger \vec{x}$$

function LANCZOS($\mathbf{A}$,$\vec{x}$,k,f)

$\vec{v}_1 \leftarrow \vec{x}/\|\vec{x}\|$

$\vec{w}_1 \leftarrow \vec{x}/\|\vec{x}\|$

for $i \leftarrow 1$ **to** $k-1$ **do**

$\mathbf{T}_{ii} \leftarrow \vec{w}_i^\dagger \mathbf{A} \vec{v}_i$

$\vec{v}_{i+1} \leftarrow (\mathbf{A} - \mathbf{T}_{ii})\vec{v}_i$

$\vec{w}_{i+1} \leftarrow (\mathbf{A}^\dagger - \mathbf{T}_{ii}^*)\vec{w}_i$

if $i > 1$ **then**

$\vec{v}_{i+1} \leftarrow \vec{v}_{i+1} - \mathbf{T}_{(i-1)i}\vec{v}_{i-1}$

$\vec{w}_{i+1} \leftarrow \vec{w}_{i+1} - \mathbf{T}_{i(i-1)}^*\vec{w}_{i-1}$

end if

$\mathbf{T}_{(i+1)i} \leftarrow \|\vec{v}_{i+1}\|$

$\mathbf{T}_{i(i+1)} \leftarrow \vec{w}_{i+1}^\dagger \vec{v}_{i+1}$

$\mathbf{T}_{i(i+1)} \leftarrow \mathbf{T}_{i(i+1)}/\mathbf{T}_{(i+1)i}$

$\vec{v}_{i+1} \leftarrow \vec{v}_{i+1}/\mathbf{T}_{(i+1)i}$

$\vec{w}_{i+1} \leftarrow \vec{w}_{i+1}/\mathbf{T}_{i(i+1)}^*$

end for

$\mathbf{T}_{kk} \leftarrow \vec{w}_k^\dagger \mathbf{A} \vec{v}_k$

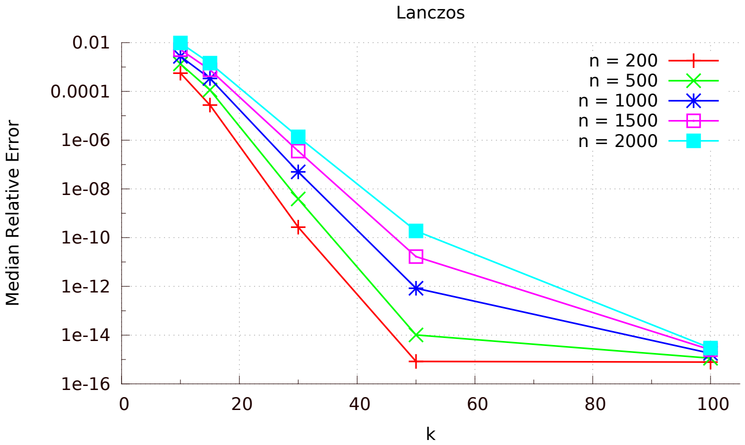
return $\|\vec{x}\| \mathbf{V} f(\mathbf{T}) \hat{e}_1$

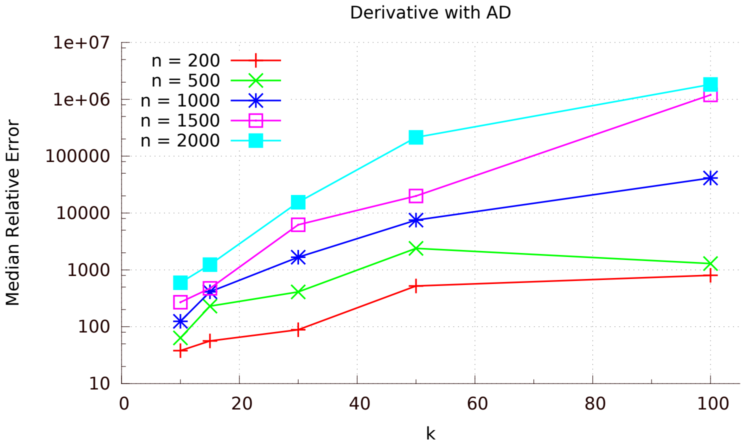
end function

- $\mathbf{A}$ only in matrix-vector multiplication
- Approximation depends on $\vec{x}$
- Compute f only for $\mathbf{T} \in \mathbb{C}^{k\times k}$
- Since $\vec{w}_1 = \vec{x}/\|\vec{x}\|$:

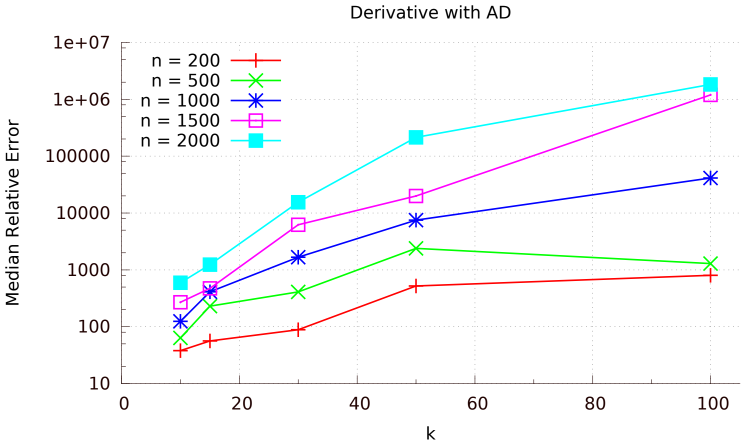
$$\mathbf{W}^\dagger \vec{x} = \hat{e}_1$$

- Operations: $\mathcal{O}(nk) + \mathcal{O}(k^3)$


Lanczos
Median Relative Error
0.01
0.0001
1e-06
1e-08
1e-10
1e-12
1e-14
1e-16
0
20
40
60
80
100
k
n = 200
n = 500
n = 1000
n = 1500
n = 2000



Derivative with AD
Median Relative Error
k
n = 200
n = 500
n = 1000
n = 1500
n = 2000
1e+07
1e+06
100000
10000
1000
100
10
0
20
40
60
80
100



- Large errors even for small matrix sizes
- Lanczos AD numerically unstable

A hint from A. Frommer:

Theorem (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 containded 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:

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

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

A hint from A. Frommer:

Theorem (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 containded 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:

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

We can compute the derivative of $f(\mathbf{A})$ without knowing $f'(\mathbf{A})$!

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

For polynomials

Let $p_n(\bar{\mathbf{A}}) = \bar{\mathbf{A}}^n$

$n = 2$:

$$p_2(\bar{\mathbf{A}}) = \begin{bmatrix} \mathbf{A} & \dot{\mathbf{A}} \\ 0 & \mathbf{A} \end{bmatrix}^2 = \begin{bmatrix} \mathbf{A}^2 & \mathbf{A}\dot{\mathbf{A}} + \dot{\mathbf{A}}\mathbf{A} \\ 0 & \mathbf{A}^2 \end{bmatrix} = \begin{bmatrix} p_2(\mathbf{A}) & \dot{p}_2(\mathbf{A}) \\ 0 & p_2(\mathbf{A}) \end{bmatrix}$$

$n \to n+1$:

$$p_{n+1}(\bar{\mathbf{A}}) = \begin{bmatrix} p_n(\mathbf{A}) & \dot{p}_n(\mathbf{A}) \\ 0 & p_n(\mathbf{A}) \end{bmatrix} \begin{bmatrix} \mathbf{A} & \dot{\mathbf{A}} \\ 0 & \mathbf{A} \end{bmatrix} = \begin{bmatrix} p_{n+1}(\mathbf{A}) & \dot{p}_{n+1}(\mathbf{A}) \\ 0 & p_{n+1}(\mathbf{A}) \end{bmatrix}$$

General proof

Let $\varepsilon \neq 0$ and $\mathbf{U} = \begin{bmatrix} \mathbb{1} & \varepsilon^{-1}\mathbb{1} \\ 0 & \mathbb{1} \end{bmatrix}$.

Since $f(\mathbf{X}\mathbf{A}\mathbf{X}^{-1}) = \mathbf{X}f(\mathbf{A})\mathbf{X}^{-1}$ we get:

$$
\begin{aligned}
f\left(\begin{bmatrix} \mathbf{A}(0) & \frac{\mathbf{A}(\varepsilon)-\mathbf{A}(0)}{\varepsilon} \\ 0 & \mathbf{A}(\varepsilon) \end{bmatrix}\right) &= \mathbf{U}f\left(\mathbf{U}^{-1}\begin{bmatrix} \mathbf{A}(0) & \frac{\mathbf{A}(\varepsilon)-\mathbf{A}(0)}{\varepsilon} \\ 0 & \mathbf{A}(\varepsilon) \end{bmatrix}\mathbf{U}\right)\mathbf{U}^{-1} \\
&= \mathbf{U}f\left(\begin{bmatrix} \mathbf{A}(0) & 0 \\ 0 & \mathbf{A}(\varepsilon) \end{bmatrix}\right)\mathbf{U}^{-1} \\
&= \mathbf{U}\begin{bmatrix} f(\mathbf{A}(0)) & 0 \\ 0 & f(\mathbf{A}(\varepsilon)) \end{bmatrix}\mathbf{U}^{-1} \\
&= \begin{bmatrix} f(\mathbf{A}(0)) & \frac{f(\mathbf{A}(\varepsilon))-f(\mathbf{A}(0))}{\varepsilon} \\ 0 & f(\mathbf{A}(\varepsilon)) \end{bmatrix}
\end{aligned}
$$

Take $\lim_{\varepsilon\to 0}$ on both sides.

- Useful for numerical calculations?
- Advantage:
  - Compute derivative with Lanczos algorithm:

$$\mathrm{sign}\left(\begin{bmatrix} \mathbf{A} & \dot{\mathbf{A}} \\ 0 & \mathbf{A} \end{bmatrix}\right)\begin{pmatrix} 0 \\ \vec{x} \end{pmatrix} = \begin{pmatrix} \frac{d}{dt}\,\mathrm{sign}(\mathbf{A})\vec{x} \\ \mathrm{sign}(\mathbf{A})\vec{x} \end{pmatrix}$$

- Disadvantage:
  - Size of linear space doubles
- Convergence of Lanczos $\rightarrow$ spectrum of $\bar{\mathbf{A}} := \begin{bmatrix} \mathbf{A} & \dot{\mathbf{A}} \\ 0 & \mathbf{A} \end{bmatrix}$?

- Useful for numerical calculations?
- Advantage:
  - Compute derivative with Lanczos algorithm:

$$\mathrm{sign}\left(\begin{bmatrix} \mathbf{A} & \dot{\mathbf{A}} \\ 0 & \mathbf{A} \end{bmatrix}\right)\begin{pmatrix} 0 \\ \vec{x} \end{pmatrix} = \begin{pmatrix} \frac{d}{dt}\,\mathrm{sign}(\mathbf{A})\vec{x} \\ \mathrm{sign}(\mathbf{A})\vec{x} \end{pmatrix}$$

- Disadvantage:
  - Size of linear space doubles
- Convergence of Lanczos $\rightarrow$ spectrum of $\bar{\mathbf{A}} := \begin{bmatrix} \mathbf{A} & \dot{\mathbf{A}} \\ 0 & \mathbf{A} \end{bmatrix}$?

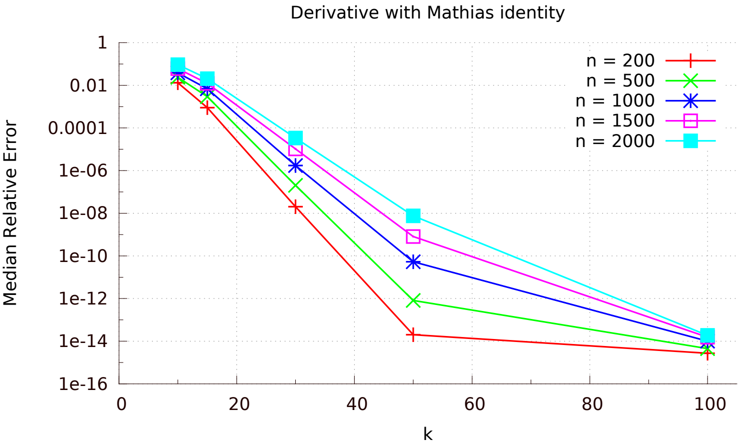
Properties of $\bar{\mathbf{A}}$

- $\det \bar{\mathbf{A}} = (\det \mathbf{A})^2$
- λ eigenvalue of $\bar{\mathbf{A}} \Leftrightarrow \lambda$ eigenvalue of A
- $\vec{x} := \begin{pmatrix} \vec{x}_1 \\ \vec{x}_2 \end{pmatrix}$ eigenvector of $\bar{\mathbf{A}}$ to eigenvalue λ, then:

$\vec{x}_2 \equiv 0$ and $\vec{x}_1$ eigenvector of $\mathbf{A}$ to eigenvalue λ

or

$\vec{x}_2$ eigenvector of $\mathbf{A}$ to eigenvalue λ and $\vec{x}_1 = \frac{\partial}{\partial t}\vec{x}_2$


Derivative with Mathias identity
Median Relative Error
1
0.01
0.0001
1e-06
1e-08
1e-10
1e-12
1e-14
1e-16
0
20
40
60
80
100
k
n = 200
n = 500
n = 1000
n = 1500
n = 2000

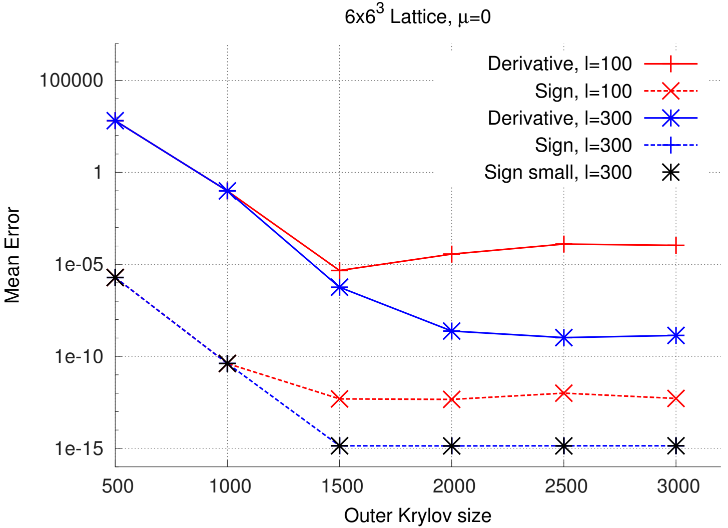
Derivative of the Overlap operator

- Lattice size $6x6^3$
- SU(3) configurations[c] with improved action
  - $\beta = 5.95$, $\rho = 1.4$
- No deflation (yet)
- Nested Lanczos algorithm [d] $\rightarrow$ inner Krylov size l fixed
- Results for

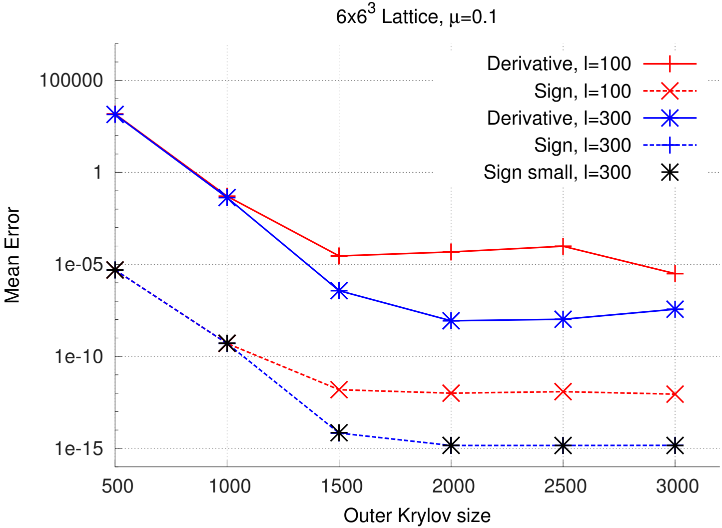
$$\mathrm{sgn}(\gamma_5 D_W) \quad \text{and} \quad \frac{\partial}{\partial \theta_\mu(x)} \mathrm{sgn}(\gamma_5 D_W)$$

[c] Provided by Oleg Kochetkov

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



$6x6^3$ Lattice, μ=0
Derivative, l=100
Sign, l=100
Derivative, l=300
Sign, l=300
Sign small, l=300
100000
1
1e-05
1e-10
1e-15
Mean Error
500
1000
1500
2000
2500
3000
Outer Krylov size



6x6^3 Lattice, μ=0.1
Derivative, l=100
Sign, l=100
Derivative, l=300
Sign, l=300
Sign small, l=300
100000
1
1e-05
1e-10
1e-15
Mean Error
500
1000
1500
2000
2500
3000
Outer Krylov size

Summary and Outlook

- Summary
  - A method to compute derivatives of the Overlap operator
  - Test on small lattices
  - First results are very promising
- Outlook
  - Implement deflation
  - Generalization to higher derivatives
  - Application to physical problems