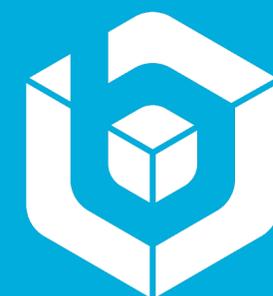# Non-Hermitian Eigensolvers for Lattice Gauge Theory

Peter Boyle, Chulwoo Jung, Patrick Oare, Shuhei Yamamoto
February 13th, 2026

Brookhaven
National Laboratory

Lattice Dirac Operators

- Each action S_F for lattice fermions is characterized by its **Dirac operator** D.

$$S_F = \sum_{x,y \in \Lambda} \overline{\psi}_a^\alpha(x) D_{ab}^{\alpha\beta}(x,y) \psi_b^\beta(y) = \overline{\psi} D \psi$$

Vector-matrix-vector product

  - D is the lattice analogue of the continuum Dirac operator, realized as a large, sparse complex matrix.

  - Typical sizes of D range from $\approx 10^8 \times 10^8$ to $\approx 10^{10} \times 10^{10}$.

Lattice Dirac Operators

- Each action S_{F} for lattice fermions is characterized by its **Dirac operator D**.

$$S_F = \sum_{x,y \in \Lambda} \overline{\psi}_a^{\alpha}(x) D_{ab}^{\alpha\beta}(x,y) \psi_b^{\beta}(y) = \overline{\psi} D \psi$$

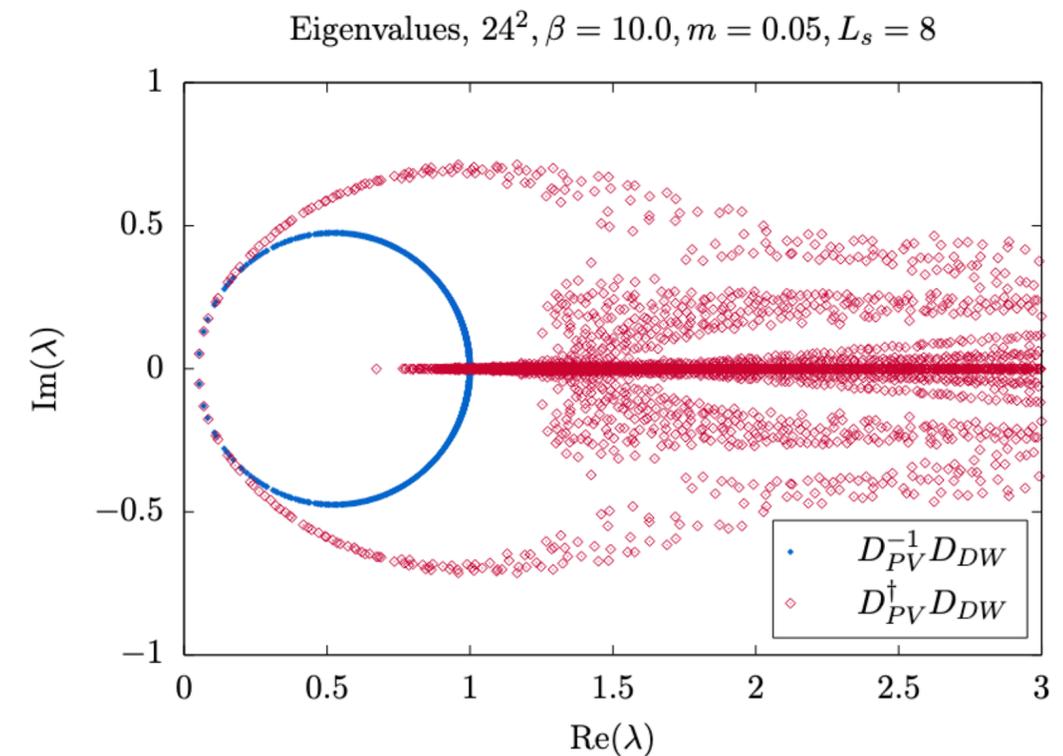Vector-matrix-vector product

  - D is the lattice analogue of the continuum Dirac operator, realized as a large, sparse complex matrix.

  - Typical sizes of D range from $\approx 10^8 \times 10^8$ to $\approx 10^{10} \times 10^{10}$.

- The eigenspectrum of D informs us about properties of the operator.

  - Low modes of $D \implies$ areas where iterative methods take longer to converge.

  - Index theorem $\implies$ relates topological charge to number of left- and right-handed zero modes.

Lattice Dirac Operators

- Hermitian Dirac operators are easier to understand:

  ▸ They have a real spectrum.

  ▸ Their eigenvectors are guaranteed to be orthogonal.

  ▸ Can compute spectra of Hermitian operators with **Lanczos algorithm**.

Lattice Dirac Operators

- Hermitian Dirac operators are easier to understand:

  ‣ They have a real spectrum.

  ‣ Their eigenvectors are guaranteed to be orthogonal.

  ‣ Can compute spectra of Hermitian operators with **Lanczos algorithm**.

- We want to understand the spectrum of non-Hermitian Dirac operators, which will have complex eigenvalues and a non-orthogonal eigenbasis.

Eigenvalues, $24^2, \beta = 10.0, m = 0.05, L_s = 8$

$D_{PV}^{-1} D_{DW}$

$D_{PV}^{\dagger} D_{DW}$

Brower *et. al.*, hep-lat/2004.07732

Agenda

I. Spectrum of the Wilson Operator

II. Eigensolver Algorithms

III. Numerical Examples

I: Spectrum of the Wilson Operator

The Wilson Operator

- The naïve discretization of the QCD action features unphysical modes on the edge of the Brillouin zone called "doublers".

- The Wilson-Dirac operator D_W gaps out the doublers to remove them from the discretized theory.

$$D_W(m)_{x,y} = (4+m) - \frac{1}{2}\sum_{\mu=1}^{4}(1-\gamma_\mu)U_\mu(x)\delta_{x+\hat\mu,y} + (1+\gamma_\mu)U_\mu^\dagger(y)\delta_{x-\hat\mu,y}$$

The Wilson Operator

- The naïve discretization of the QCD action features unphysical modes on the edge of the Brillouin zone called "doublers".

- The Wilson-Dirac operator D_W gaps out the doublers to remove them from the discretized theory.

$$D_W(m)_{x,y} = (4 + m) - \frac{1}{2} \sum_{\mu=1}^{4} (1 - \gamma_\mu) U_\mu(x) \delta_{x+\hat{\mu},y} + (1 + \gamma_\mu) U_\mu^\dagger(y) \delta_{x-\hat{\mu},y}$$

- D_W is a non-Hermitian operator, hence has a complex spectrum.

  ▸ However, it satisfies γ_5**-hermiticity**:

$$\gamma_5 D_W \gamma_5 = D_W^\dagger$$

The Wilson Operator

- The naïve discretization of the QCD action features unphysical modes on the edge of the Brillouin zone called "doublers".

- The Wilson-Dirac operator D_W gaps out the doublers to remove them from the discretized theory.

$$D_W(m)_{x,y} = (4+m) - \frac{1}{2}\sum_{\mu=1}^{4}(1-\gamma_\mu)U_\mu(x)\delta_{x+\hat{\mu},y} + (1+\gamma_\mu)U_\mu^\dagger(y)\delta_{x-\hat{\mu},y}$$

- D_W is a non-Hermitian operator, hence has a complex spectrum.

  ▸ However, it satisfies γ_5-**hermiticity**:

  $$\gamma_5 D_W \gamma_5 = D_W^\dagger$$

- Its Hermitian counterpart $H_W(m)$ has a real spectrum.

$$H_W(m) = \gamma_5 D_W(m) \qquad\qquad H_W(m) = H_W^\dagger(m)$$

Spectrum of D_W

- Consider the characteristic polynomial $p(\lambda)$ of D_W:

$$p(\lambda) = \det(D_W - \lambda) = \det(\gamma_5^2(D_W - \lambda))$$
$$= \det(\gamma_5 D_W \gamma_5 - \lambda) = \det(D_W^\dagger - \lambda) = \det(D_W - \lambda^*)^*$$
$$= p(\lambda^*)^*$$

Spectrum of D_W

- Consider the characteristic polynomial $p(\lambda)$ of D_W:

$$p(\lambda) = \det(D_W - \lambda) = \det(\gamma_5^2(D_W - \lambda))$$
$$= \det(\gamma_5 D_W \gamma_5 - \lambda) = \det(D_W^\dagger - \lambda) = \det(D_W - \lambda^*)^*$$
$$= p(\lambda^*)^*$$

- If $p(\lambda) = 0$, then either:

  1. $\lambda \in \mathbb{R}$ is a real eigenvalue of D_W, or

  2. $p(\lambda^*) = 0$ as well, hence λ^* is also an eigenvalue of D_W.

Spectrum of D_W

- Consider the characteristic polynomial $p(\lambda)$ of D_W:

$$p(\lambda) = \det(D_W - \lambda) = \det(\gamma_5^2(D_W - \lambda))$$
$$= \det(\gamma_5 D_W \gamma_5 - \lambda) = \det(D_W^\dagger - \lambda) = \det(D_W - \lambda^*)^*$$
$$= p(\lambda^*)^*$$

- If $p(\lambda) = 0$, then either:

  1. $\lambda \in \mathbb{R}$ is a real eigenvalue of D_W, or

  2. $p(\lambda^*) = 0$ as well, hence λ^* is also an eigenvalue of D_W.

> The eigenvalues of a γ_5-hermitian Dirac operator are either real or come in complex-conjugate pairs.

Zero modes of $D_W(m)$

- We are interested in the spectrum of D_W at a fixed mass $m_0 = 0$.

- The massive Dirac operator with mass m, $D_W(m)$, is related to $D_W(0)$ by a shift of $D_W(m) = D_W(0) + m$.

  ‣ Likewise, this holds for $H_W(m) = \gamma_5 D_W(m)$.

Zero modes of $D_W(m)$

- We are interested in the spectrum of D_W at a fixed mass $m_0 = 0$.

- The massive Dirac operator with mass m, $D_W(m)$, is related to $D_W(0)$ by a shift of $D_W(m) = D_W(0) + m$.

  ▸ Likewise, this holds for $H_W(m) = \gamma_5 D_W(m)$.

- Zero modes of $D_W(m)$ correspond to zero modes of $H_W(m)$: if $D_W(m)|\psi\rangle = 0$, then $H_W(m)|\psi\rangle = 0$ as well.

  ▸ Suppose $D_W(0)$ has a real eigenvalue λ. Then $D_W(-\lambda) = D_W(0) - \lambda$ has a zero mode, and so does $H_W(\lambda)$.

Zero modes of $D_W(m)$

- We are interested in the spectrum of D_W at a fixed mass $m_0 = 0$.

- The massive Dirac operator with mass m, $D_W(m)$, is related to $D_W(0)$ by a shift of $D_W(m) = D_W(0) + m$.

  ▸ Likewise, this holds for $H_W(m) = \gamma_5 D_W(m)$.

- Zero modes of $D_W(m)$ correspond to zero modes of $H_W(m)$: if $D_W(m)|\psi\rangle = 0$, then $H_W(m)|\psi\rangle = 0$ as well.

  ▸ Suppose $D_W(0)$ has a real eigenvalue λ. Then $D_W(-\lambda) = D_W(0) - \lambda$ has a zero mode, and so does $H_W(\lambda)$.

> If $H_W(-m)$ has a zero mode, $D_W(0)$ has a real eigenvalue at m.

Spectral Flow Method

- We can compute the **real spectrum of** D_W by varying m and computing the spectrum of $H_W(m)$.

  ▸ When $H_W(-m)$ has a zero mode $\lambda^H(-m) = 0$, this corresponds to a real eigenvalue of D_W at m.

Spectral Flow Method

- We can compute the **real spectrum of** D_W by varying m and computing the spectrum of $H_W(m)$.

  ▸ When $H_W(-m)$ has a zero mode $\lambda^H(-m) = 0$, this corresponds to a real eigenvalue of D_W at m.

- For each real eigenpair $(\lambda_k^D, |k\rangle)$ of D_W, the chirality of the mode is given by the derivative of the zero crossing,

$$\langle k|\gamma_5|k\rangle = \left.\frac{d\lambda_k^H}{dm}\right|_{m=-\lambda_k^D}$$

Spectral Flow Method

- We can compute the **real spectrum of** D_W by varying m and computing the spectrum of $H_W(m)$.

  ▸ When $H_W(-m)$ has a zero mode $\lambda^H(-m) = 0$, this corresponds to a real eigenvalue of D_W at m.

- For each real eigenpair $(\lambda_k^D, |k\rangle)$ of D_W, the chirality of the mode is given by the derivative of the zero crossing,

$$\langle k|\gamma_5|k\rangle = \frac{d\lambda_k^H}{dm}\bigg|_{m=-\lambda_k^D}$$
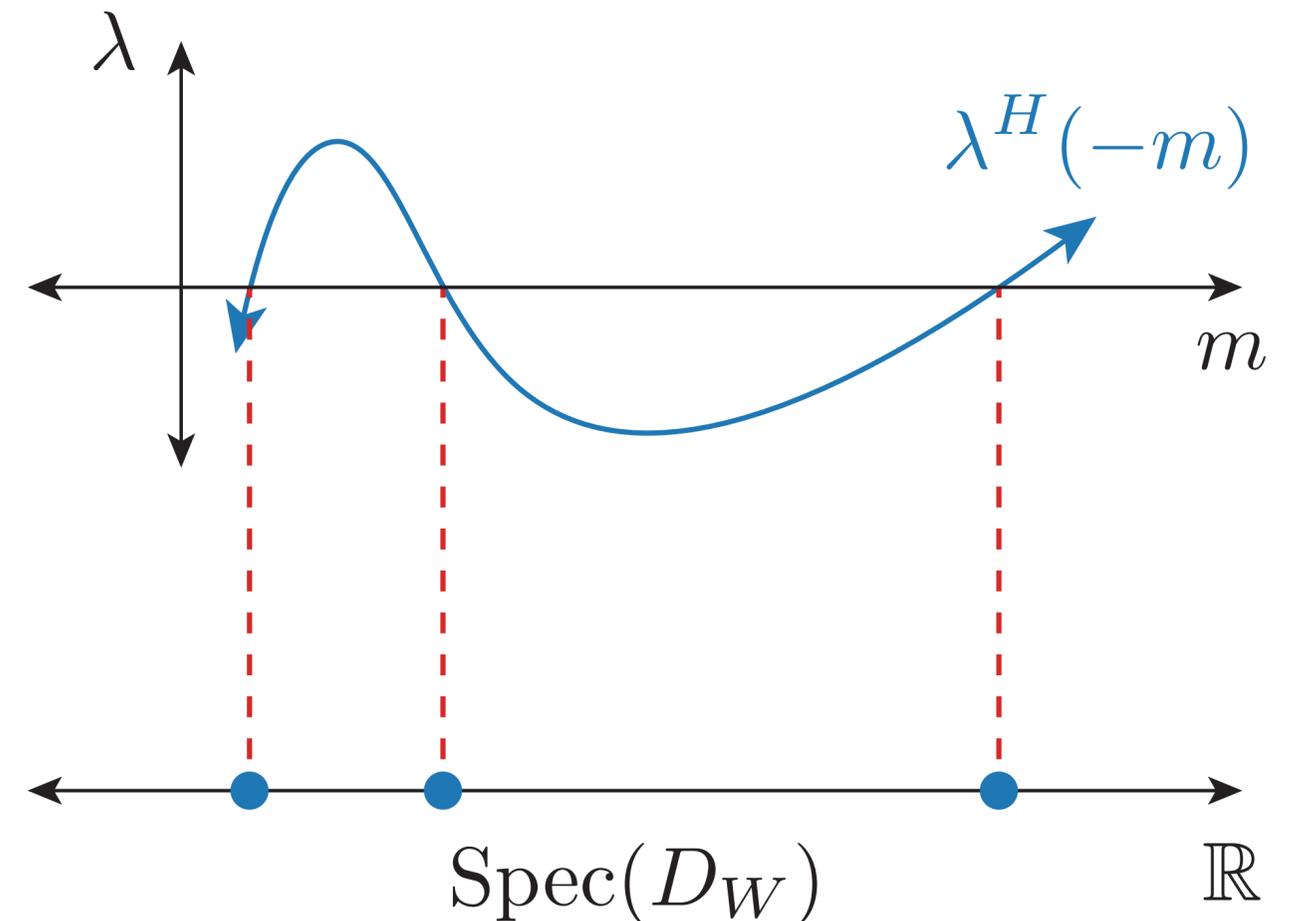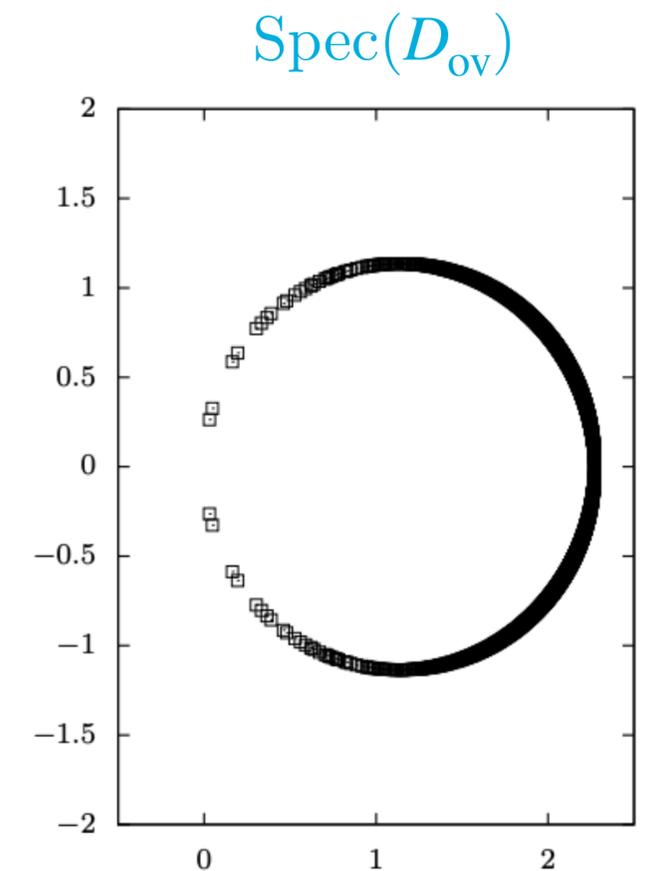
Topological Charge on the Lattice

- Topological charge is well-defined in continuum QFT because the base manifold is smooth.

  ▸ On the lattice there is no unique definition: different definitions must only agree in the continuum limit.

Topological Charge on the Lattice

- Topological charge is well-defined in continuum QFT because the base manifold is smooth.

  ‣ On the lattice there is no unique definition: different definitions must only agree in the continuum limit.

- One appealing definition is in terms of the **overlap operator**,

$$D_{\mathrm{ov}} \equiv \left(1 - \frac{A}{\sqrt{A^\dagger A}} \right)$$

$$A \equiv 1 + s - D_W$$

  ‣ s is a free parameter used to tune the overlap operator.

Topological Charge on the Lattice

- Topological charge is well-defined in continuum QFT because the base manifold is smooth.

  ‣ On the lattice there is no unique definition: different definitions must only agree in the continuum limit.

- One appealing definition is in terms of the **overlap operator**,

$$D_{\text{ov}} \equiv \left(1 - \frac{A}{\sqrt{A^\dagger A}} \right)$$

$$A \equiv 1 + s - D_W$$

  ‣ s is a free parameter used to tune the overlap operator.

- The topological charge is defined as the difference in the number of negative (n_-) and positive (n_+) chirality zero modes of D_{ov}:

$$Q_{\text{top}} \equiv n_- - n_+$$

$\text{Spec}(D_{\text{ov}})$

Topological Charge (Wilson Operator)

- The low, real modes of the Wilson operator are projected onto the zero modes of the overlap operator.

  ▸ A real mode $\theta_W \in \mathbb{R}$ with $\theta_W < 1 + s$ is mapped to the 0 mode of D_{ov}.

$$Q = \sum_{\theta_k^W \in \mathbb{R}} \mathrm{sign} \langle k | \gamma_5 | k \rangle$$

$\mathrm{Spec}(D_W(m = 0))$

hep-lat/1410.7170

Topological Charge (Wilson Operator)

- The low, real modes of the Wilson operator are projected onto the zero modes of the overlap operator.

  ▸ A real mode $\theta_W \in \mathbb{R}$ with $\theta_W < 1 + s$ is mapped to the 0 mode of D_{ov}.

$$Q = \sum_{\theta_k^W \in \mathbb{R}} \mathrm{sign}\,\langle k | \gamma_5 | k \rangle$$

$\mathrm{Spec}(D_W(m = 0))$

hep-lat/1410.7170
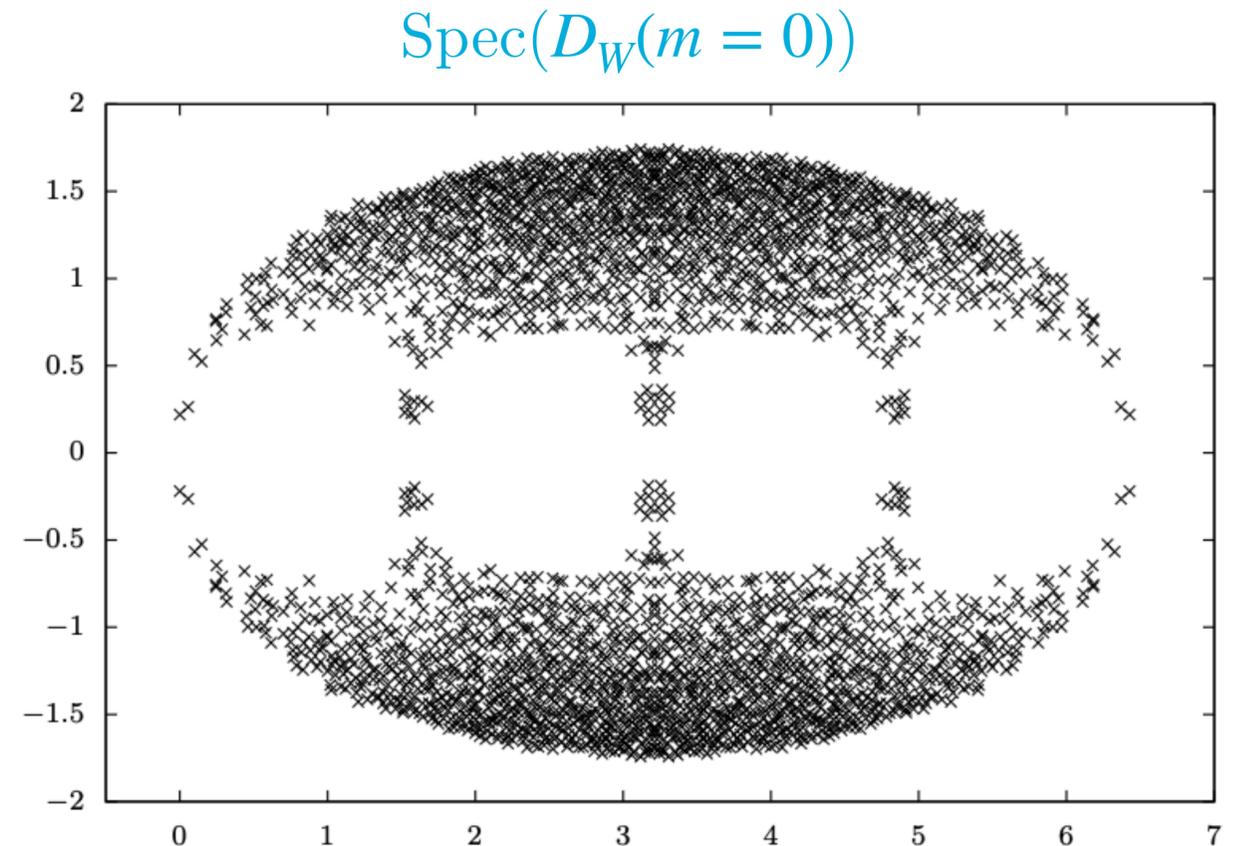
- Topology can change if the net chirality of the real modes changes.

  ▸ Requires the annihilation of a complex-conjugate eigenvalue pair.
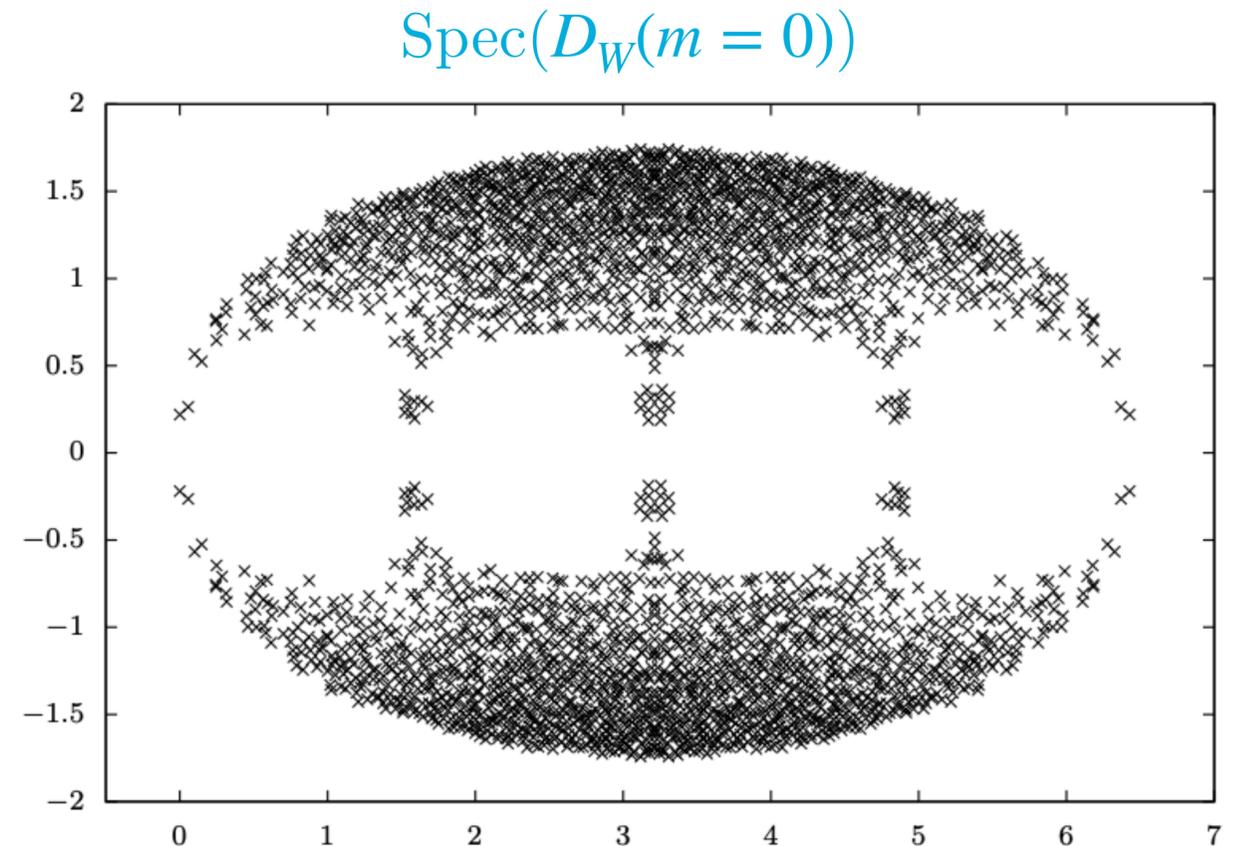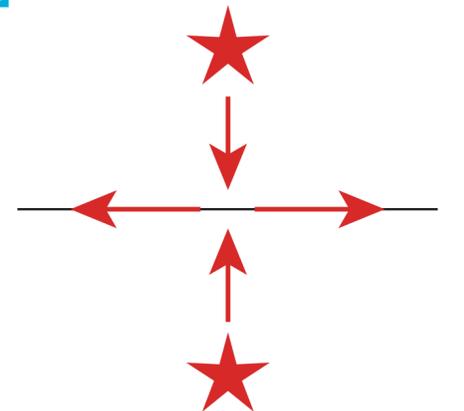
II: Eigenvalue algorithms for $D\psi = b$

Eigenvalue Algorithms

- Direct calculation of the eigenspectrum of a large $M \times M$ matrix D through the characteristic polynomial $p(\lambda) = \det(D - \lambda I_M)$ is not feasible.

  ▸ Polynomials with $\deg p \geq 5$ are not necessarily solvable in radicals.

  ▸ Rootfinding algorithms must be iterative.

Eigenvalue Algorithms

- Direct calculation of the eigenspectrum of a large $M \times M$ matrix D through the characteristic polynomial $p(\lambda) = \det(D - \lambda I_M)$ is not feasible.

  - Polynomials with $\deg p \geq 5$ are not necessarily solvable in radicals.

  - Rootfinding algorithms must be iterative.

- Right: spectrum of the free theory domain-wall Dirac operator.

- The spectrum is extremely complicated and dense, and is a good example of the types of spectrum we are interested in computing.
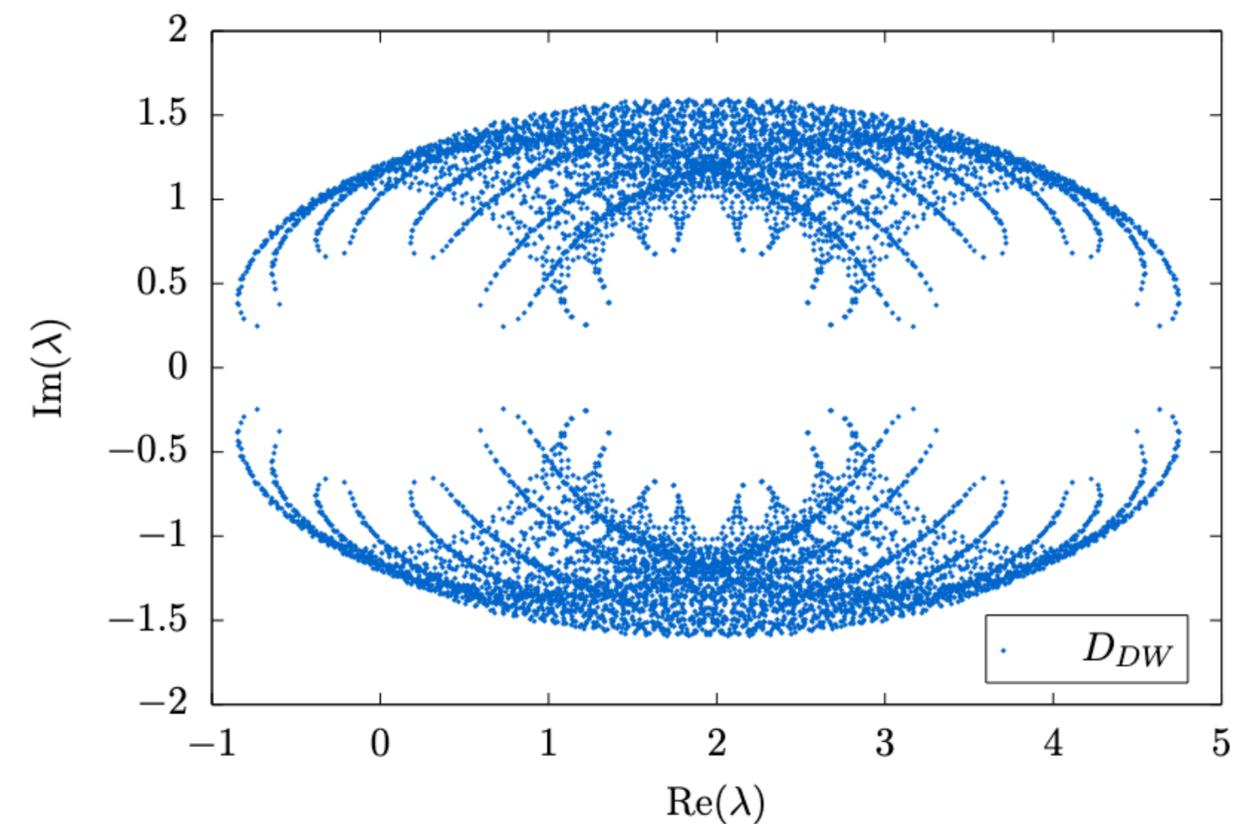
Brower *et. al.*, hep-lat/2004.07732

The Power Method

$$D = \sum_\lambda \lambda |\lambda\rangle\langle\lambda|$$
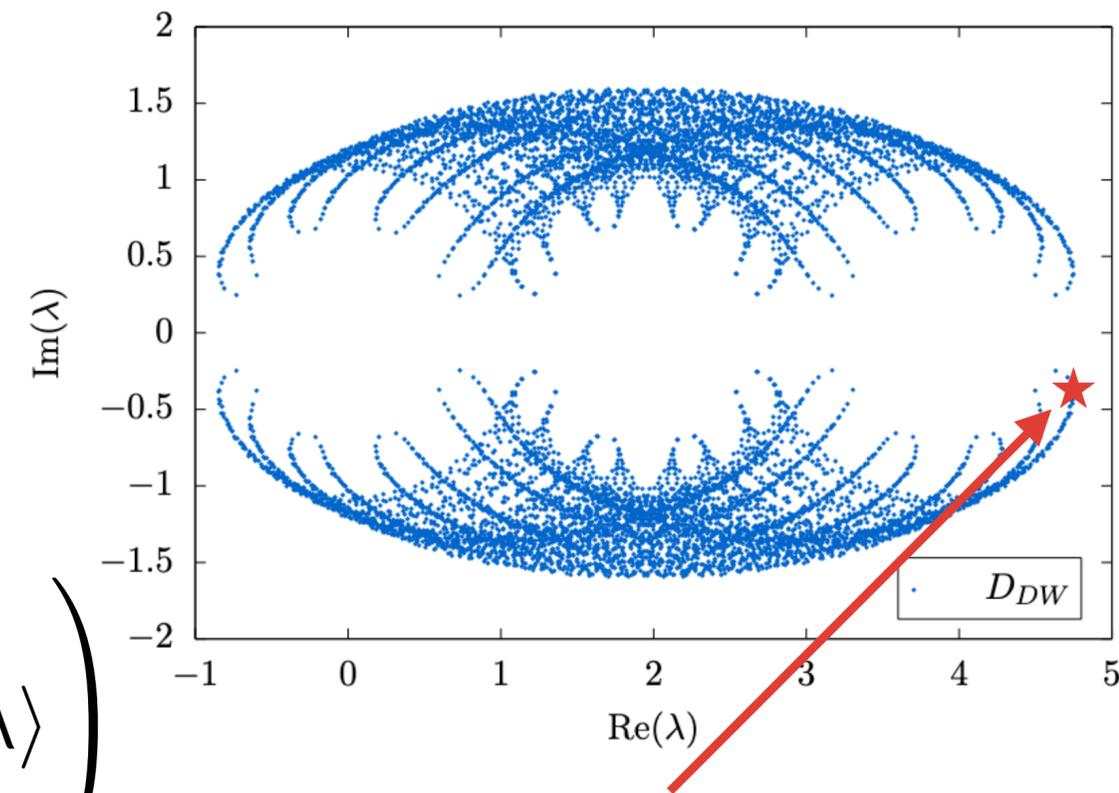
- To extract the largest eigenvalue of D, consider applying D repeatedly to a vector $|v_0\rangle$. The initial vector $|v_0\rangle$ can be expressed in the eigenbasis of D as

$$|v_0\rangle = \sum_\lambda c_\lambda |\lambda\rangle$$

The Power Method

$$D = \sum_\lambda \lambda |\lambda\rangle\langle\lambda|$$

- To extract the largest eigenvalue of D, consider applying D repeatedly to a vector $|v_0\rangle$. The initial vector $|v_0\rangle$ can be expressed in the eigenbasis of D as

$$|v_0\rangle = \sum_\lambda c_\lambda |\lambda\rangle$$

- Repeated applications of D to $|v_0\rangle$ filters out the eigenvector with largest eigenvalue $|\Lambda|$:

$$D^n |v_0\rangle = \sum_\lambda c_\lambda \lambda^n |\lambda\rangle = c_\Lambda \Lambda^n \left(|\Lambda\rangle + \sum_{\lambda \neq \Lambda} \frac{c_\lambda}{c_\Lambda} \left(\frac{\lambda}{\Lambda} \right)^n |\lambda\rangle \right)$$

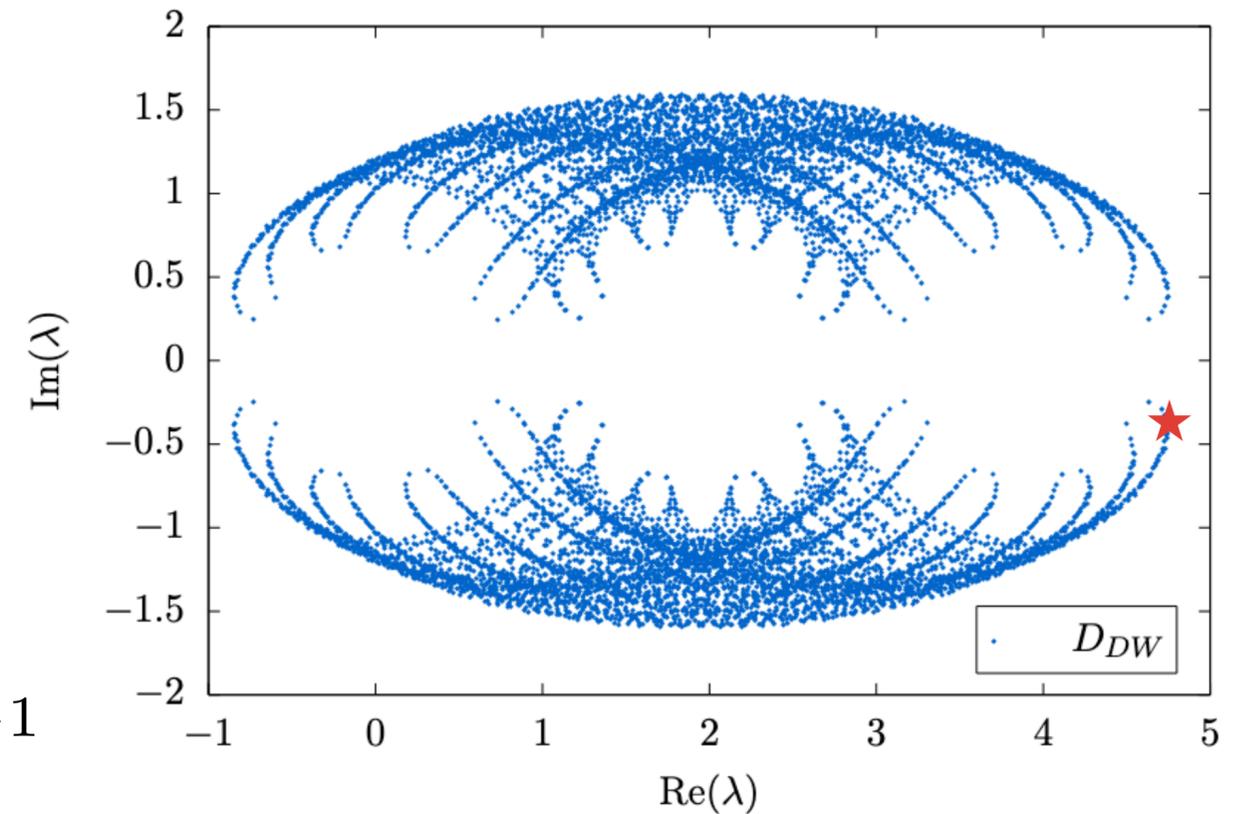$$\xrightarrow{n\to\infty} c_\Lambda \Lambda^n |\Lambda\rangle$$

$\frac{\lambda}{\Lambda} < 1$ by assumption

The Power Method

$$D = \sum_\lambda \lambda |\lambda\rangle\langle\lambda|$$

- To extract the largest eigenvalue of D, consider applying D repeatedly to a vector $|v_0\rangle$. The initial vector $|v_0\rangle$ can be expressed in the eigenbasis of D as

$$|v_0\rangle = \sum_\lambda c_\lambda |\lambda\rangle$$

- Repeated applications of D to $|v_0\rangle$ filters out the eigenvector with largest eigenvalue $|\Lambda|$:

$$D^n|v_0\rangle = \sum_\lambda c_\lambda \lambda^n |\lambda\rangle = c_\Lambda \Lambda^n \left(|\Lambda\rangle + \sum_{\lambda \neq \Lambda} \frac{c_\lambda}{c_\Lambda} \left(\frac{\lambda}{\Lambda}\right)^n |\lambda\rangle \right)$$

$$\xrightarrow{n\to\infty} c_\Lambda \Lambda^n |\Lambda\rangle$$

Example eigenvalue the power method can compute
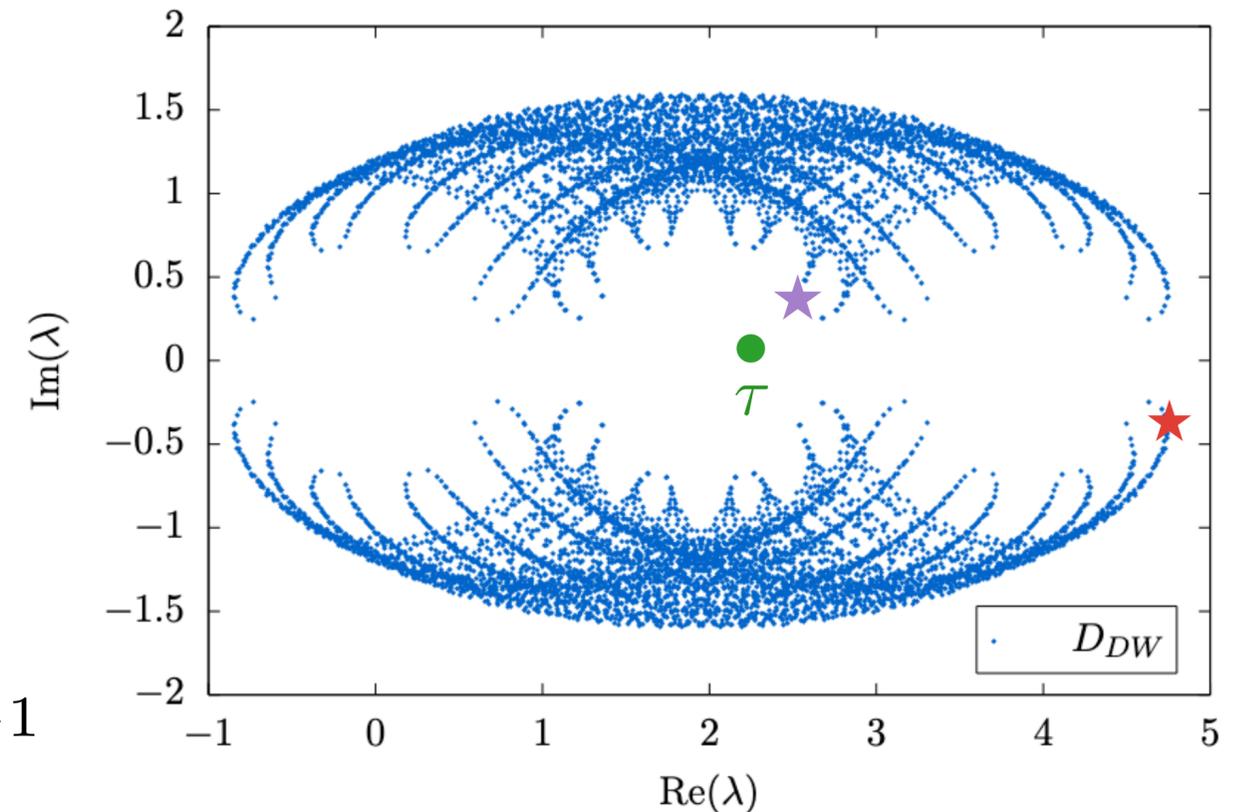
$\frac{\lambda}{\Lambda} < 1$ by assumption

Shift-and-Invert Operator

- Power method only allows you to isolate the eigenpair of A with the largest-norm eigenvalue.

- Fix $\tau \in \mathbb{C}$ and consider the operator,

$$\Theta_\tau(D) \equiv (D - \tau I)^{-1}$$

Shift-and-Invert Operator

- Power method only allows you to isolate the eigenpair of A with the largest-norm eigenvalue.

- Fix $\tau \in \mathbb{C}$ and consider the operator,

$$\Theta_\tau(D) \equiv (D - \tau I)^{-1}$$

- An eigenvalue λ of D corresponds to the eigenvalue $\theta_\tau(\lambda) = (\lambda - \tau)^{-1}$ of $\Theta_\tau(D)$ with the same eigenvector.

  ‣ The largest norm $\theta_\tau(\lambda)$ corresponds to the eigenvalue λ closest to τ.

  ‣ The power method applied to $\Theta_\tau(D)$ yields the closest eigenvalue of D to τ.

Shift-and-Invert Operator

- Power method only allows you to isolate the eigenpair of A with the largest-norm eigenvalue.

- Fix $\tau \in \mathbb{C}$ and consider the operator,
$$\Theta_\tau(D) \equiv (D - \tau I)^{-1}$$

- An eigenvalue λ of D corresponds to the eigenvalue $\theta_\tau(\lambda) = (\lambda - \tau)^{-1}$ of $\Theta_\tau(D)$ with the same eigenvector.

  ▸ The largest norm $\theta_\tau(\lambda)$ corresponds to the eigenvalue λ closest to τ.

  ▸ The power method applied to $\Theta_\tau(D)$ yields the closest eigenvalue of D to τ.

- Major downside: **matrix inversion is expensive**. Can use this to compute a few eigenvalues in isolation, but **not** the whole spectrum.
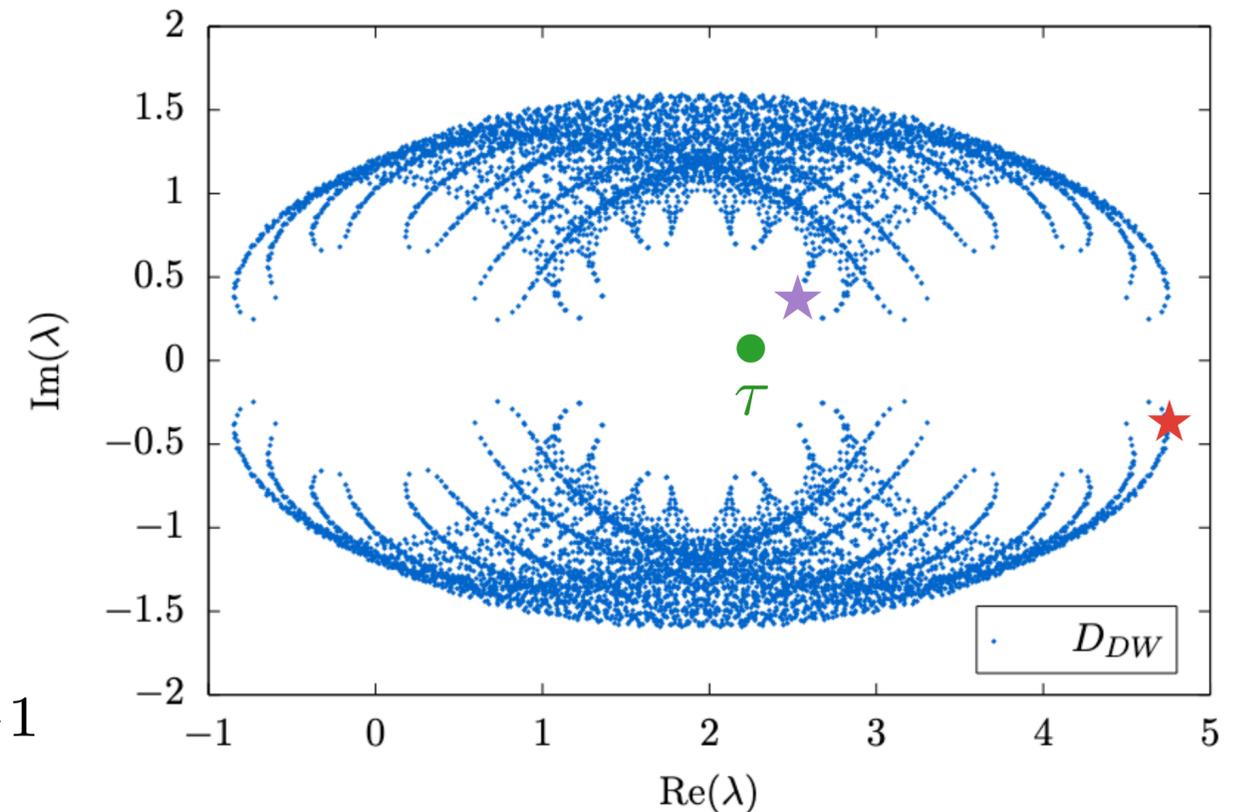
Krylov Methods

- Repeatedly applying D to $|v_0\rangle$ imparts information about the spectrum of D.

- Given an initial vector $|v\rangle$, what other eigenvalue information can we learn from the **Krylov space**,

$$K_N(D, |v\rangle) = \mathrm{span}\{|v\rangle, D|v\rangle, D^2|v\rangle, ..., D^{N-1}|v\rangle\}$$

Krylov Methods

- Repeatedly applying D to $|v_0\rangle$ imparts information about the spectrum of D.

- Given an initial vector $|v\rangle$, what other eigenvalue information can we learn from the **Krylov space**,

$$K_N(D, |v\rangle) = \mathrm{span}\{|v\rangle, D|v\rangle, D^2|v\rangle, ..., D^{N-1}|v\rangle\}$$

- Restricting D to $K_N(D, |v\rangle)$ yields a truncated operator on the Krylov space that equals the full operator when $K_N(D, |v\rangle)$ spans the entire space ($N = \dim D$).

  ▸ This operator should still contain approximate information about the spectrum of D, which becomes a better approximation as $N \longrightarrow M$.

  ▸ Krylov methods work best on extremal parts of the spectrum (largest or smallest eigenvalues).

The Arnoldi Method

- Applying Gram-Schmidt to $\{|v\rangle, D|v\rangle, \ldots, D^{N-1}|v\rangle\}$ produces an orthonormal basis $\{|u_1\rangle, |u_2\rangle, \ldots, |u_N\rangle\}$ for $K_N(D, |v\rangle)$.

  $M = \dim A$

  ▸ Assemble $\{|u_k\rangle\}_{k=1}^N$ into an $M \times N$ matrix $U_N := \mathrm{col}\{|u_1\rangle, \ldots, |u_N\rangle\}$.

The Arnoldi Method

- Applying Gram-Schmidt to $\{|v\rangle, D|v\rangle, \dots, D^{N-1}|v\rangle\}$ produces an orthonormal basis $\{|u_1\rangle, |u_2\rangle, \dots, |u_N\rangle\}$ for $K_N(D, |v\rangle)$.

  $M = \dim A$

  ▸ Assemble $\{|u_k\rangle\}_{k=1}^N$ into an $M \times N$ matrix $U_N := \mathrm{col}\{|u_1\rangle, \dots, |u_N\rangle\}$.

$$DU_N = U_N Q_N + f_N e_N^T$$

$(|f_N\rangle \perp |u_1\rangle, \dots, |u_N\rangle)$

$Q_N =$ Hessenberg matrix

$$Q_N = \begin{pmatrix} * & * & * & * & \dots & * \\ * & * & * & * & \dots & * \\ 0 & * & * & * & \dots & * \\ 0 & 0 & * & * & \dots & * \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & * \end{pmatrix}$$

The Arnoldi Method

- Applying Gram-Schmidt to $\{|v\rangle, D|v\rangle, \ldots, D^{N-1}|v\rangle\}$ produces an orthonormal basis $\{|u_1\rangle, |u_2\rangle, \ldots, |u_N\rangle\}$ for $K_N(D, |v\rangle)$.

  $M = \dim A$

  - Assemble $\{|u_k\rangle\}_{k=1}^N$ into an $M \times N$ matrix $U_N := \mathrm{col}\{|u_1\rangle, \ldots, |u_N\rangle\}$.

$$\boxed{DU_N = U_N Q_N + f_N e_N^T}$$

  $(|f_N\rangle \perp |u_1\rangle, \ldots, |u_N\rangle)$

  $Q_N = $ Hessenberg matrix

- The **Rayleigh quotient** $Q_N = U_N^\dagger D U_N$ is the projection of D onto the Krylov space.

$$Q_N = \begin{pmatrix} * & * & * & * & \ldots & * \\ * & * & * & * & \ldots & * \\ 0 & * & * & * & \ldots & * \\ 0 & 0 & * & * & \ldots & * \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & * \end{pmatrix}$$

  - Easy to compute the spectrum of Q_N.

The Arnoldi Method

- Applying Gram-Schmidt to $\{|v\rangle, D|v\rangle, \ldots, D^{N-1}|v\rangle\}$ produces an orthonormal basis $\{|u_1\rangle, |u_2\rangle, \ldots, |u_N\rangle\}$ for $K_N(D, |v\rangle)$.

  $M = \dim A$

  ▸ Assemble $\{|u_k\rangle\}_{k=1}^N$ into an $M \times N$ matrix $U_N := \mathrm{col}\{|u_1\rangle, \ldots, |u_N\rangle\}$.

$$DU_N = U_N Q_N + f_N e_N^T$$

$(|f_N\rangle \perp |u_1\rangle, \ldots, |u_N\rangle)$

$Q_N = \text{Hessenberg matrix}$

$$Q_N = \begin{pmatrix} * & * & * & * & \ldots & * \\ * & * & * & * & \ldots & * \\ 0 & * & * & * & \ldots & * \\ 0 & 0 & * & * & \ldots & * \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & * \end{pmatrix}$$

- The **Rayleigh quotient** $Q_N = U_N^\dagger D U_N$ is the projection of D onto the Krylov space.

  ▸ Easy to compute the spectrum of Q_N.

How well does $\mathbf{Spec}\, Q_N$ approximate $\mathbf{Spec}\, D$?

Ritz Estimates

- An eigenpair (θ, s) of Q_N, $Q_N s = \theta s$ is called a **Ritz pair**.

  ▸ This is not necessarily an eigenpair of the full operator D, but we can compute its deviation from being an eigenpair. $s \in \mathbb{C}^N$, so Us is a vector of size $\dim D$

  ▸ This deviation is called the **Ritz estimate**,

$$R(s, \theta) = ||D(Us) - \theta(Us)||$$

Ritz Estimates

- An eigenpair (θ, s) of Q_N, $Q_N s = \theta s$ is called a **Ritz pair**.

  ▸ This is not necessarily an eigenpair of the full operator D, but we can compute its deviation from being an eigenpair. $s \in \mathbb{C}^N$, so Us is a vector of size $\dim D$

  ▸ This deviation is called the **Ritz estimate**,

  $$R(s, \theta) = ||D(Us) - \theta(Us)||$$

- We say (θ, s) has **converged** to an eigenpair of D if $R(s, \theta) < \epsilon$ for some tolerance ϵ (typically $\epsilon \approx 10^{-8}$).

  ▸ Note if $N \to M$, then $R(s, \theta) \to 0$ for all eigenpairs of Q_N because $Q_N \to D$.

Ritz Estimates

- An eigenpair (θ, s) of Q_N, $Q_N s = \theta s$ is called a **Ritz pair**.

  ▸ This is not necessarily an eigenpair of the full operator D, but we can compute its deviation from being an eigenpair. $s \in \mathbb{C}^N$, so Us is a vector of size $\dim D$

  ▸ This deviation is called the **Ritz estimate**,

$$R(s, \theta) = ||D(Us) - \theta(Us)||$$

- We say (θ, s) has **converged** to an eigenpair of D if $R(s, \theta) < \epsilon$ for some tolerance ϵ (typically $\epsilon \approx 10^{-8}$).

  ▸ Note if $N \to M$, then $R(s, \theta) \to 0$ for all eigenpairs of Q_N because $Q_N \to D$.

- The Ritz pairs satisfy the **Galerkin condition**,

$$U^\dagger [D(Us) - \theta(Us)] = 0$$

(Residual $(D - \theta)(Us)$ is $\perp$ to the Krylov space)

Ritz Estimates

- An eigenpair (θ, s) of Q_N, $Q_N s = \theta s$ is called a **Ritz pair**.

  ▸ This is not necessarily an eigenpair of the full operator D, but we can compute its deviation from being an eigenpair. $s \in \mathbb{C}^N$, so Us is a vector of size $\dim D$

  ▸ This deviation is called the **Ritz estimate**,
  $$R(s, \theta) = ||D(Us) - \theta(Us)||$$

- We say (θ, s) has **converged** to an eigenpair of D if $R(s, \theta) < \epsilon$ for some tolerance ϵ (typically $\epsilon \approx 10^{-8}$).

  ▸ Note if $N \to M$, then $R(s, \theta) \to 0$ for all eigenpairs of Q_N because $Q_N \to D$.

- The Ritz pairs satisfy the **Galerkin condition**,
  $$U^\dagger [D(Us) - \theta(Us)] = 0$$

  (Residual $(D - \theta)(Us)$ is $\perp$ to the Krylov space)

Restarting Arnoldi

- **Implicitly restarting** an eigensolver allows us to retain information about the Krylov space but work with $N \ll M$ vectors which are constantly updated.

Restarting Arnoldi

- **Implicitly restarting** an eigensolver allows us to retain information about the Krylov space but work with $N \ll M$ vectors which are constantly updated.

- Keep $k < N$ of eigenpairs of Q_N which are close to the region of the spectrum you are interested in.

  ▸ Truncate the Arnoldi expansion $DU_k = U_k Q_k + f_k e_k^T$ to keep only these eigenpairs.

  ▸ Restart the Arnoldi process and generate $N - k$ new Krylov vectors to furnish the expansion $DU_N = U_N Q_N + f_N e_N^T$.

  ▸ Repeat this, computing the Ritz estimates of all eigenpairs, until you have enough converged eigenpairs.

N = 10, k = 4, want Spec D with smallest real part

- Run Arnoldi to generate 10 approximate eigenpairs.
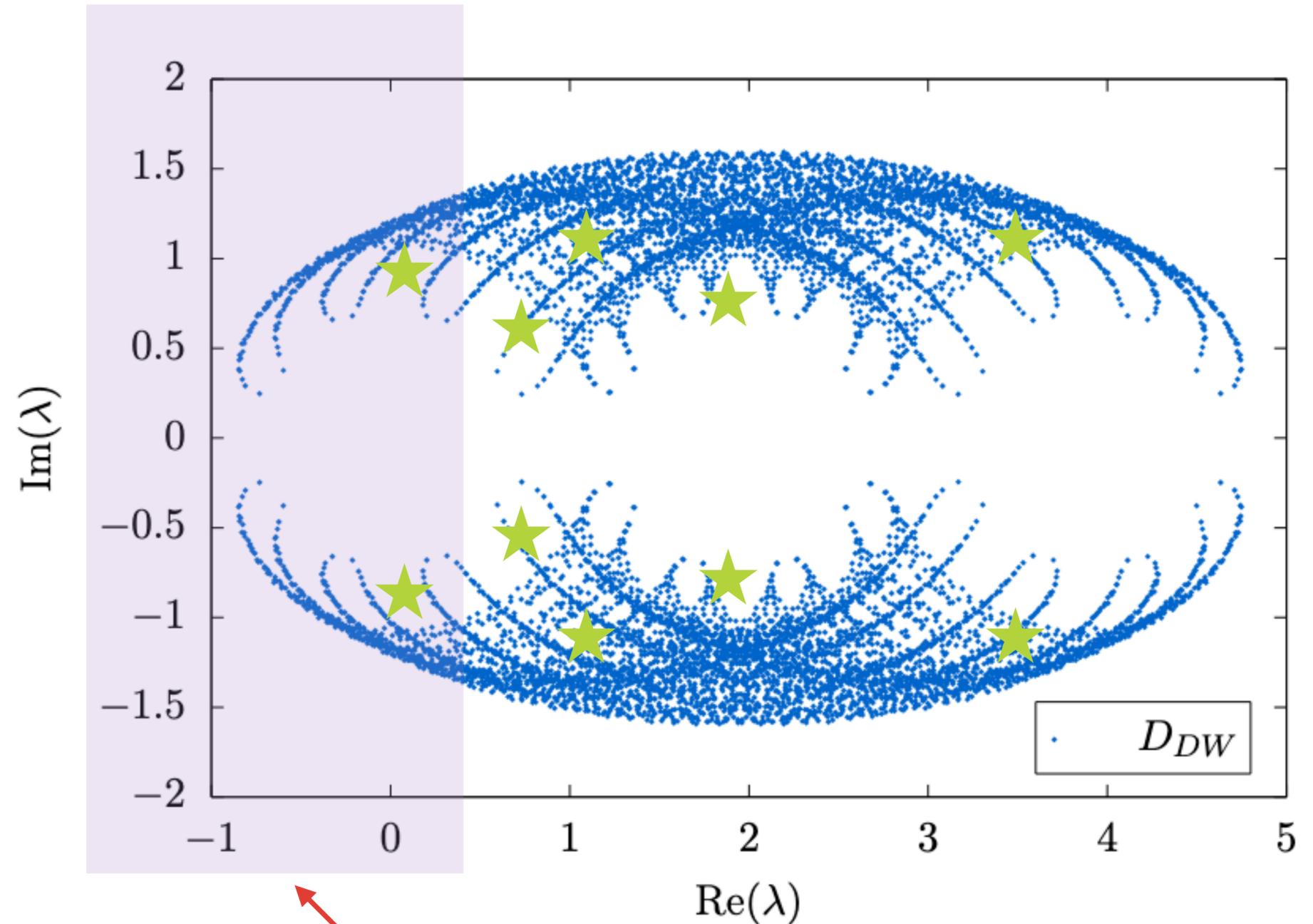
Want to find spectrum in this region

$N = 10, k = 4,$ want Spec D with smallest real part
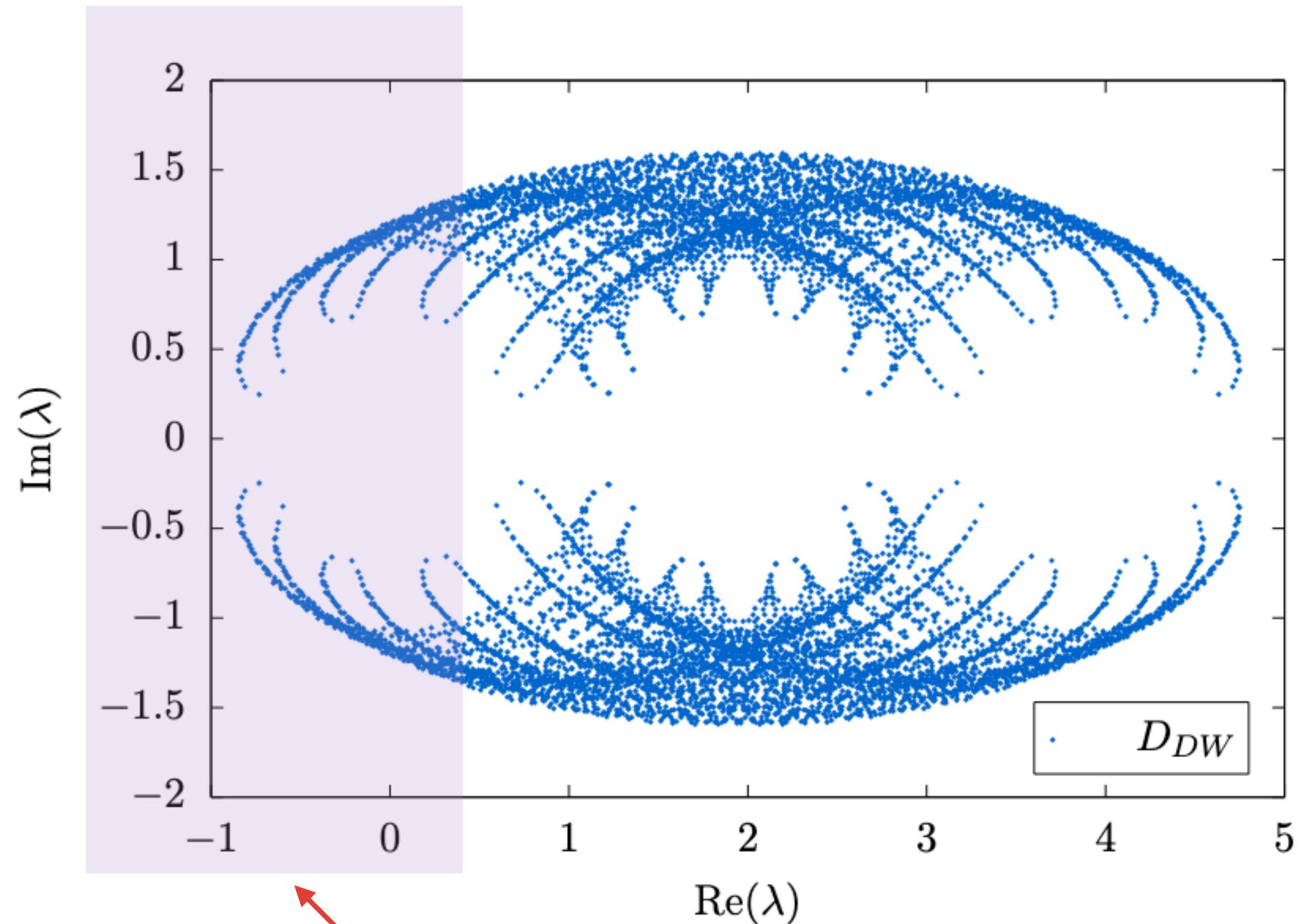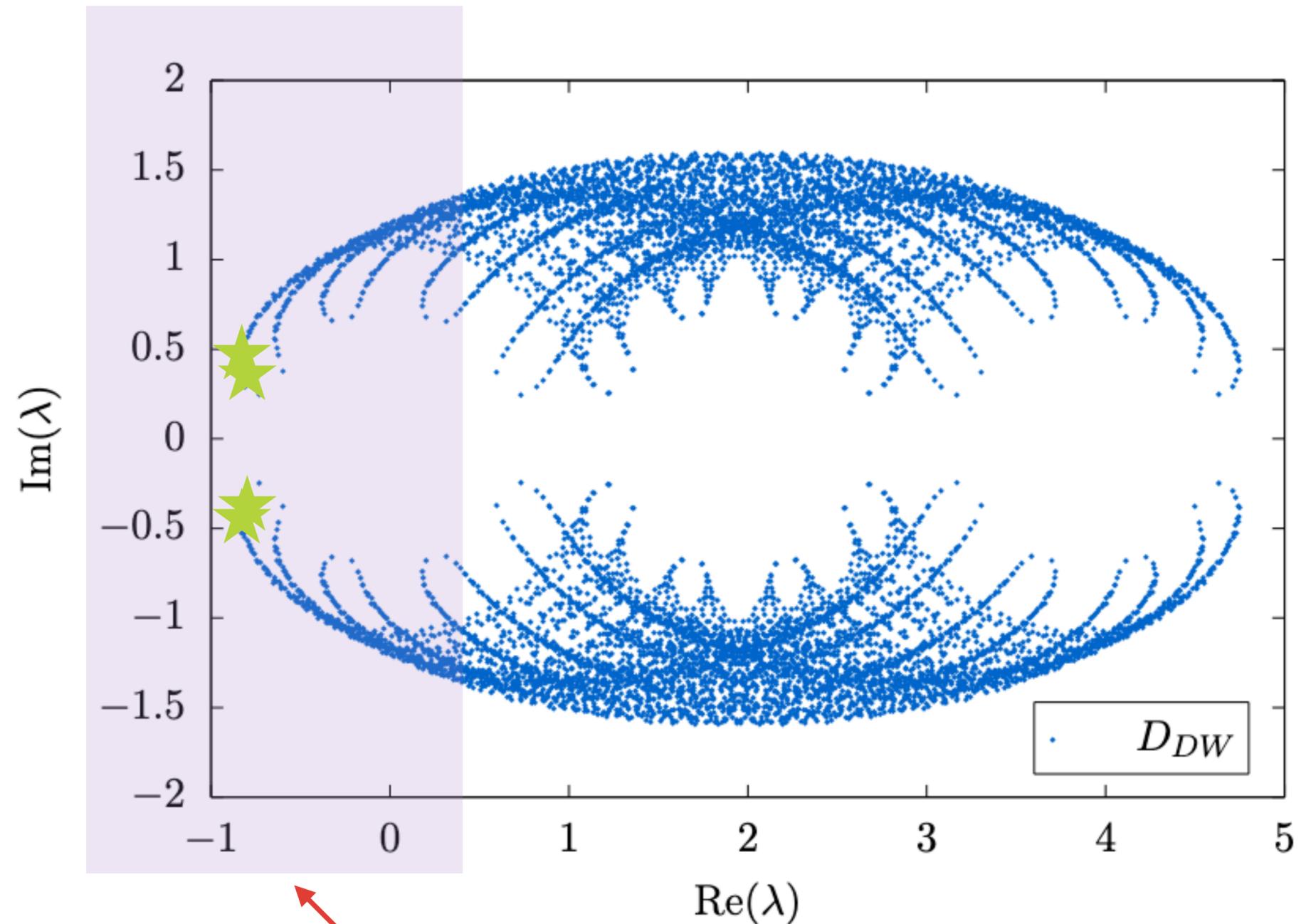
- Run Arnoldi to generate 10 approximate eigenpairs.

- Keep 4 evals with smallest real part, truncate the rest.

Want to find spectrum in this region

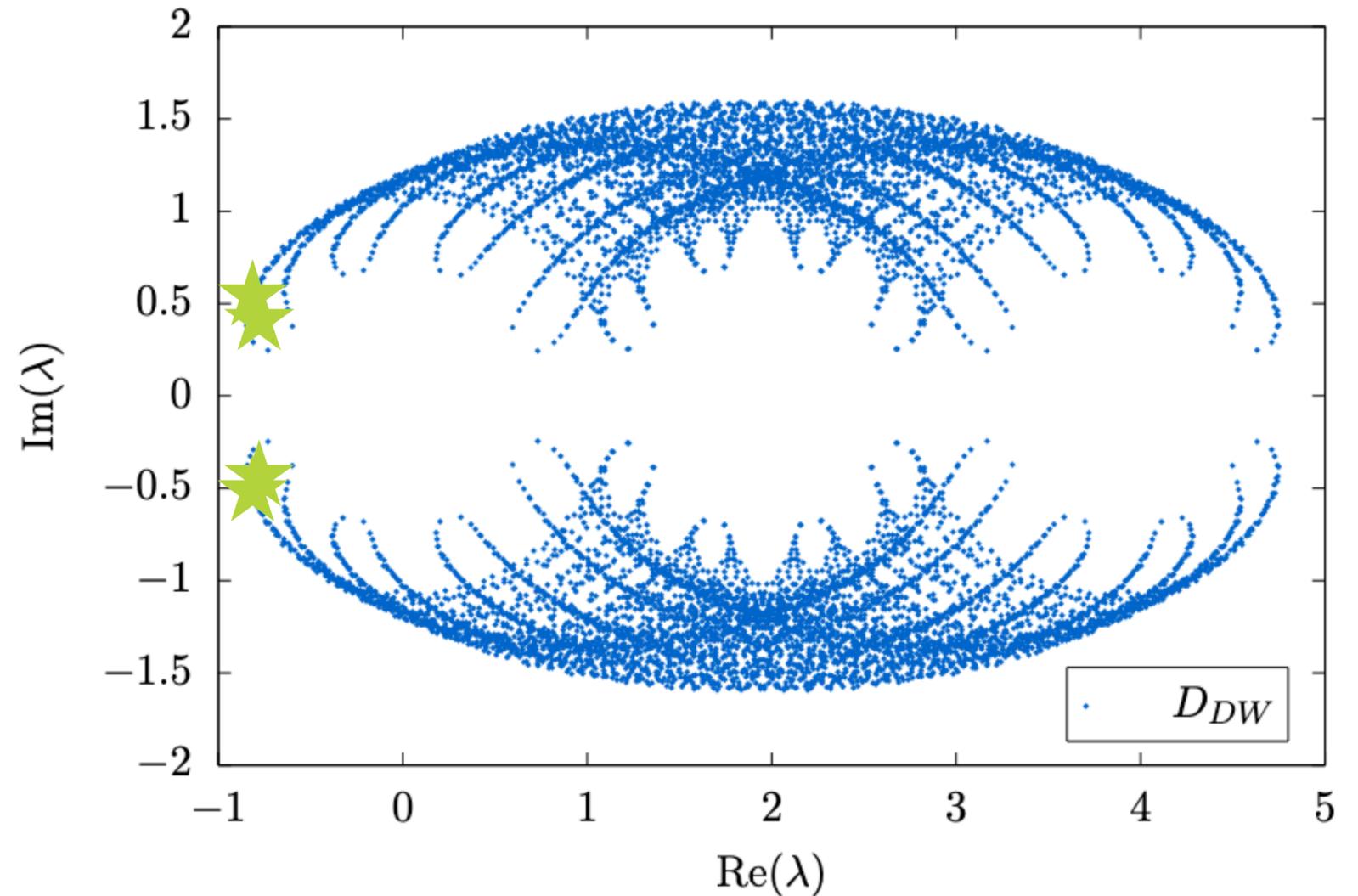# N = 10, k = 4, want Spec D with smallest real part

- Run Arnoldi to generate 10 approximate eigenpairs.

- Keep 4 evals with smallest real part, truncate the rest.

- Restart the Arnoldi process and generate 6 new eigenpairs.

Want to find spectrum in this region

N = 10, k = 4, want Spec D with smallest real part

- Run Arnoldi to generate 10 approximate eigenpairs.

- Keep 4 evals with smallest real part, truncate the rest.

- Restart the Arnoldi process and generate 6 new eigenpairs.

- Repeat this until convergence at a given tolerance.

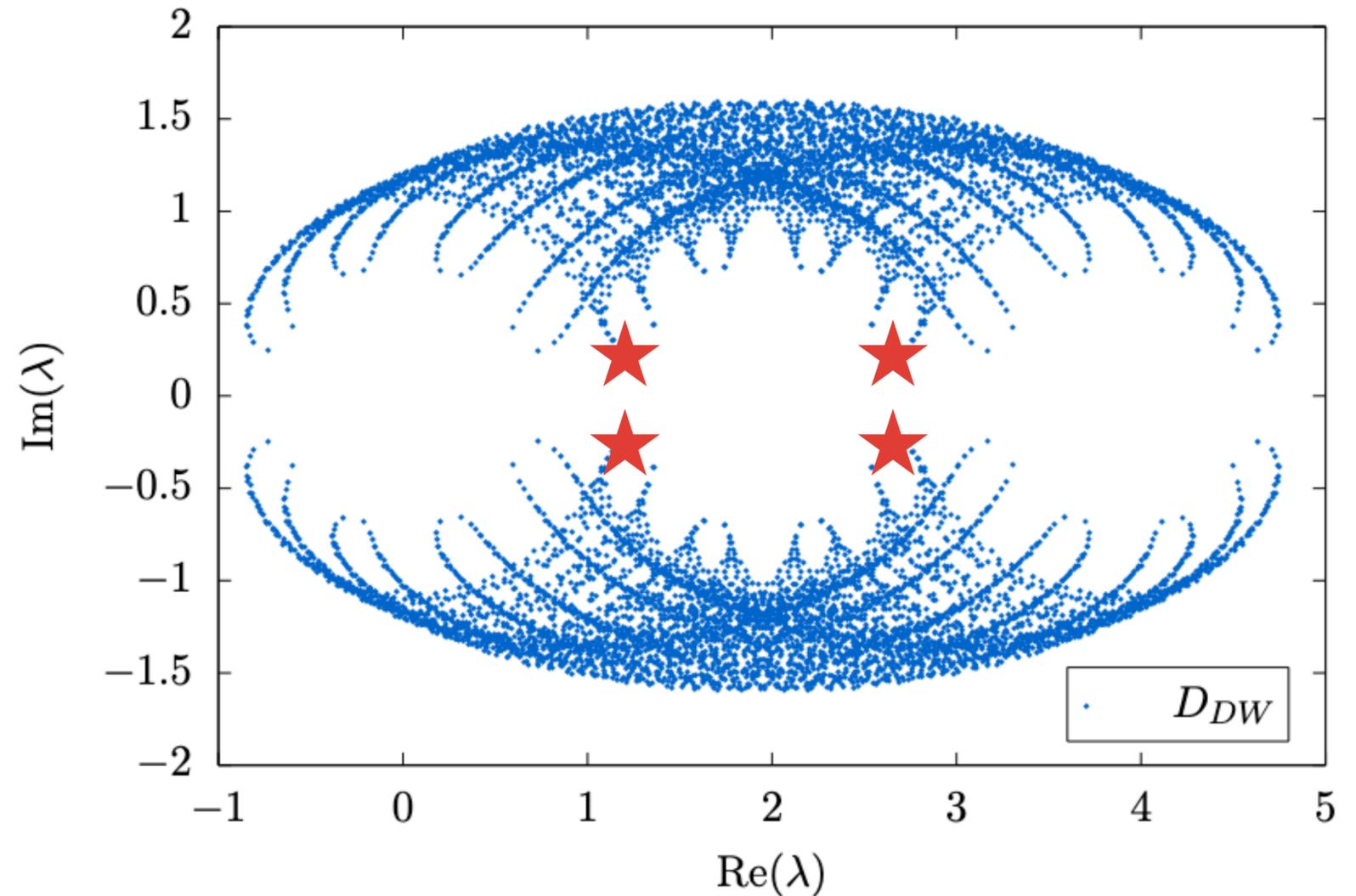Want to find spectrum in this region

$\mathbf{N = 10, \; k = 4}$, want $\mathrm{Spec}\,D$ with smallest real part

- Run Arnoldi to generate 10 approximate eigenpairs.

- Keep 4 evals with smallest real part, truncate the rest.

- Restart the Arnoldi process and generate 6 new eigenpairs.

- Repeat this until convergence at a given tolerance.

Want to find spectrum in this region

$\mathbf{N = 10, \, k = 4,}$ want $\mathrm{Spec} \, D$ with smallest real part

- Run Arnoldi to generate 10 approximate eigenpairs.

- Keep 4 evals with smallest real part, truncate the rest.

- Restart the Arnoldi process and generate 6 new eigenpairs.

- Repeat this until convergence at a given tolerance.

Want to find spectrum in this region

N = 10, k = 4, want Spec D with smallest real part

- Run Arnoldi to generate 10 approximate eigenpairs.

- Keep 4 evals with smallest real part, truncate the rest.

- Restart the Arnoldi process and generate 6 new eigenpairs.

- Repeat this until convergence at a given tolerance.

Want to find spectrum in this region

Interior Spectrum

- Krylov methods are great at accessing the spectrum near its boundary.
  - ▸ If one tries to retain the eigenvalues deep in the interior of the spectrum, the solver will often not converge.
  - ▸ Ex: keeping smallest $|\mathrm{Im}\,\theta|$ each restart will not converge.

Interior Spectrum

- Krylov methods are great at accessing the spectrum near its boundary.

  ▸ If one tries to retain the eigenvalues deep in the interior of the spectrum, the solver will often not converge.

  ▸ Ex: keeping smallest $|\mathrm{Im}\,\theta|$ each restart will not converge.

Interior Spectrum

- Krylov methods are great at accessing the spectrum near its boundary.

  ‣ If one tries to retain the eigenvalues deep in the interior of the spectrum, the solver will often not converge.

  ‣ Ex: keeping smallest $|\mathrm{Im}\,\theta|$ each restart will not converge.

- The only method we've discussed which can well-approximate interior eigenpairs is the shift-and-invert power method.

  ‣ Approximate the shift-and-invert operator within the confines of a Krylov solver: **Harmonic Krylov-Schur algorithm**
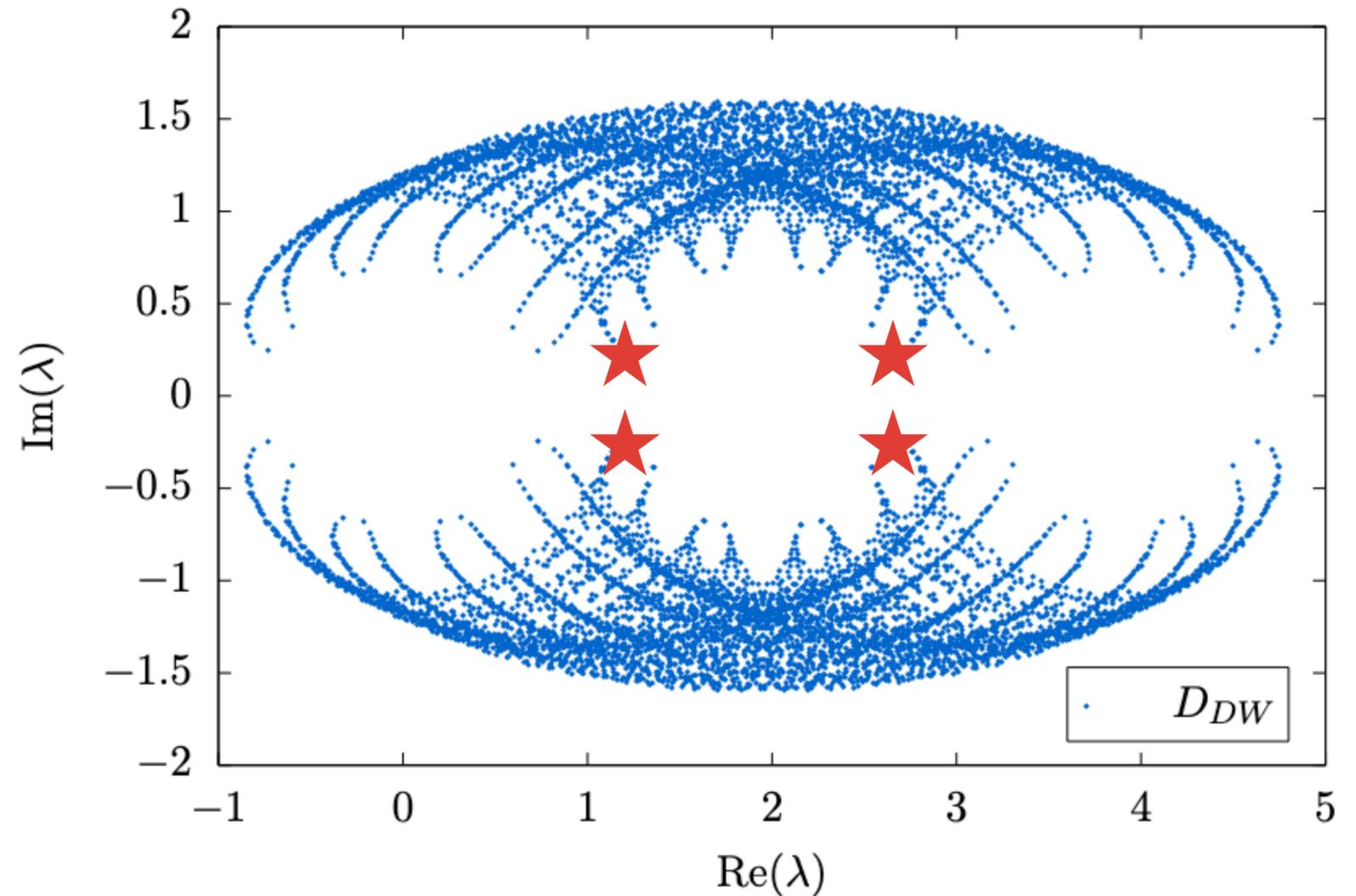
Harmonic Krylov-Schur (HKS)

- The **Krylov-Schur algorithm** loosens the restrictions in the Arnoldi factorization for easier restarting.

  ‣ More efficient and robust than Arnoldi.

Harmonic Krylov-Schur (HKS)

- The **Krylov-Schur algorithm** loosens the restrictions in the Arnoldi factorization for easier restarting.

  ▸ More efficient and robust than Arnoldi.

- **Harmonic Krylov-Schur**: approximate the Galerkin condition for the shift-and-invert operator $\Theta_\tau(D) = (D - \tau I)^{-1}$ by augmenting the standard Krylov space as $V = (D - \tau I)U$,
$$V^\dagger[\Theta_\tau(D)(Vz) - \widetilde{\theta}_\tau(Vz)] = 0$$

- Modifies Krylov-Schur **without using any additional matrix inversions**, only translations.

Numerical Examples

Ensemble Generation

- Shuhei has generated domain-wall gauge configurations of size 32^4 at $m = -1.6$, $M_5 = 1.8$.

  ▸ "Generated" mean that they are evolved in Monte Carlo time τ to simulate dynamical QCD evolution.

- They are observed to undergo a topology change around MD time $\tau \approx 3$.

Ensemble Generation

- Shuhei has generated domain-wall gauge configurations of size 32^4 at $m = -1.6$, $M_5 = 1.8$.

  ‣ "Generated" mean that they are evolved in Monte Carlo time τ to simulate dynamical QCD evolution.

- They are observed to undergo a topology change around MD time $\tau \approx 3$.

- On these configurations, Chulwoo has run:

  ‣ Spectral flow to determine the real modes of D_W.

  ‣ Harmonic Krylov-Schur to determine part of the spectrum of D_W.

- Dimensions: $M = \dim D \approx 10^8$, number of evals $N = 250$.

Convergence History at $\tau = 3.000000$
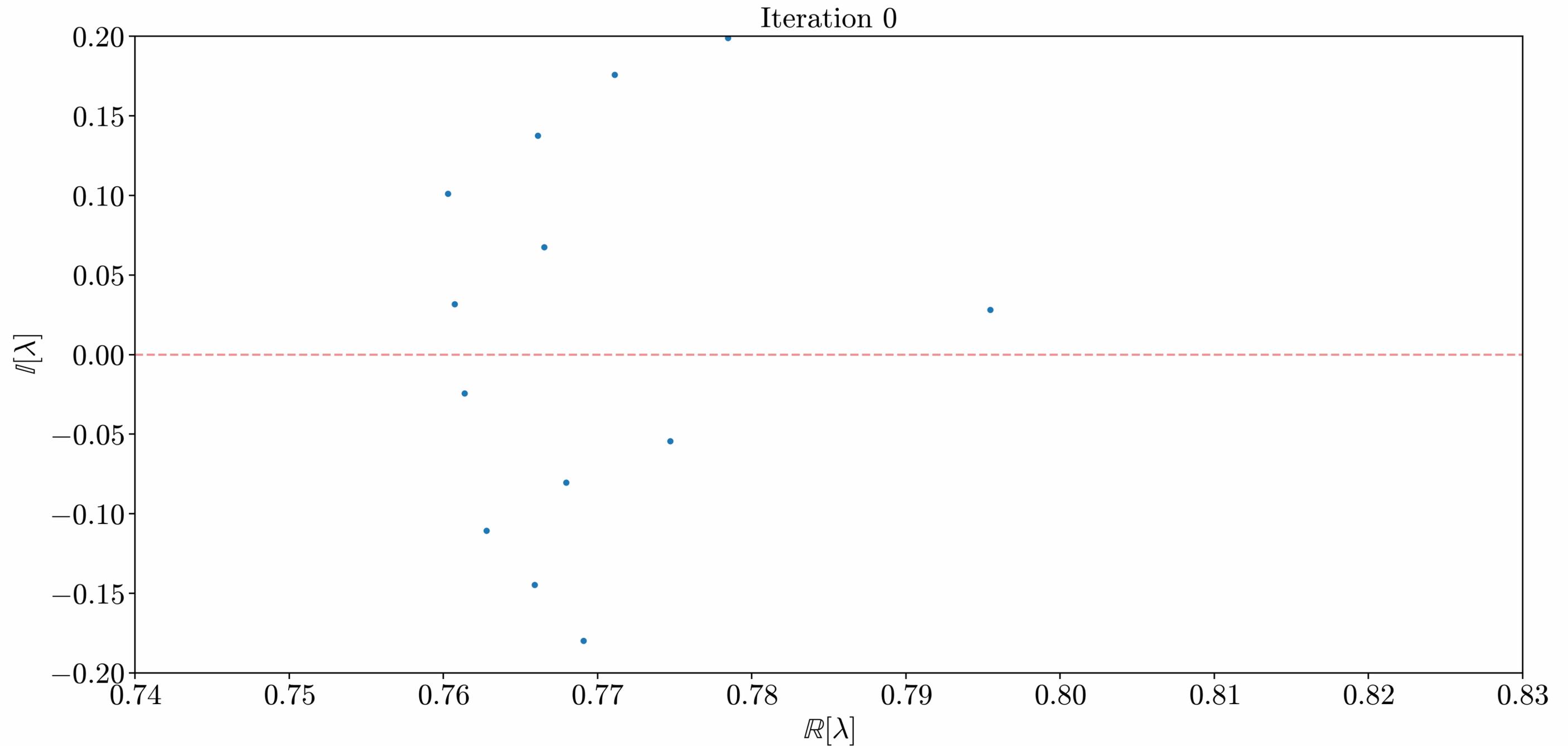
Convergence History at $\tau = 3.000000$
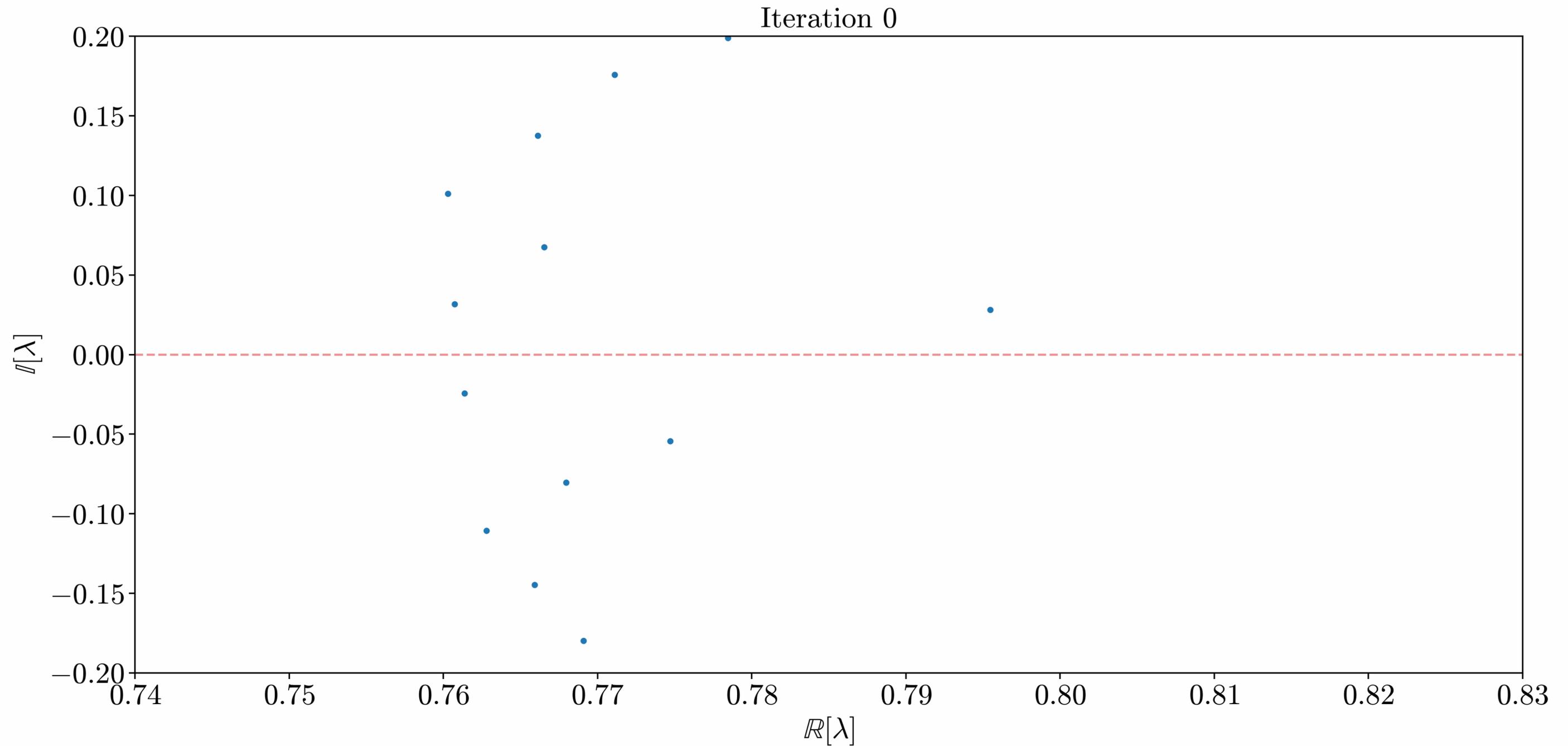
Spectral Flow

Spectrum Evolution from HKS
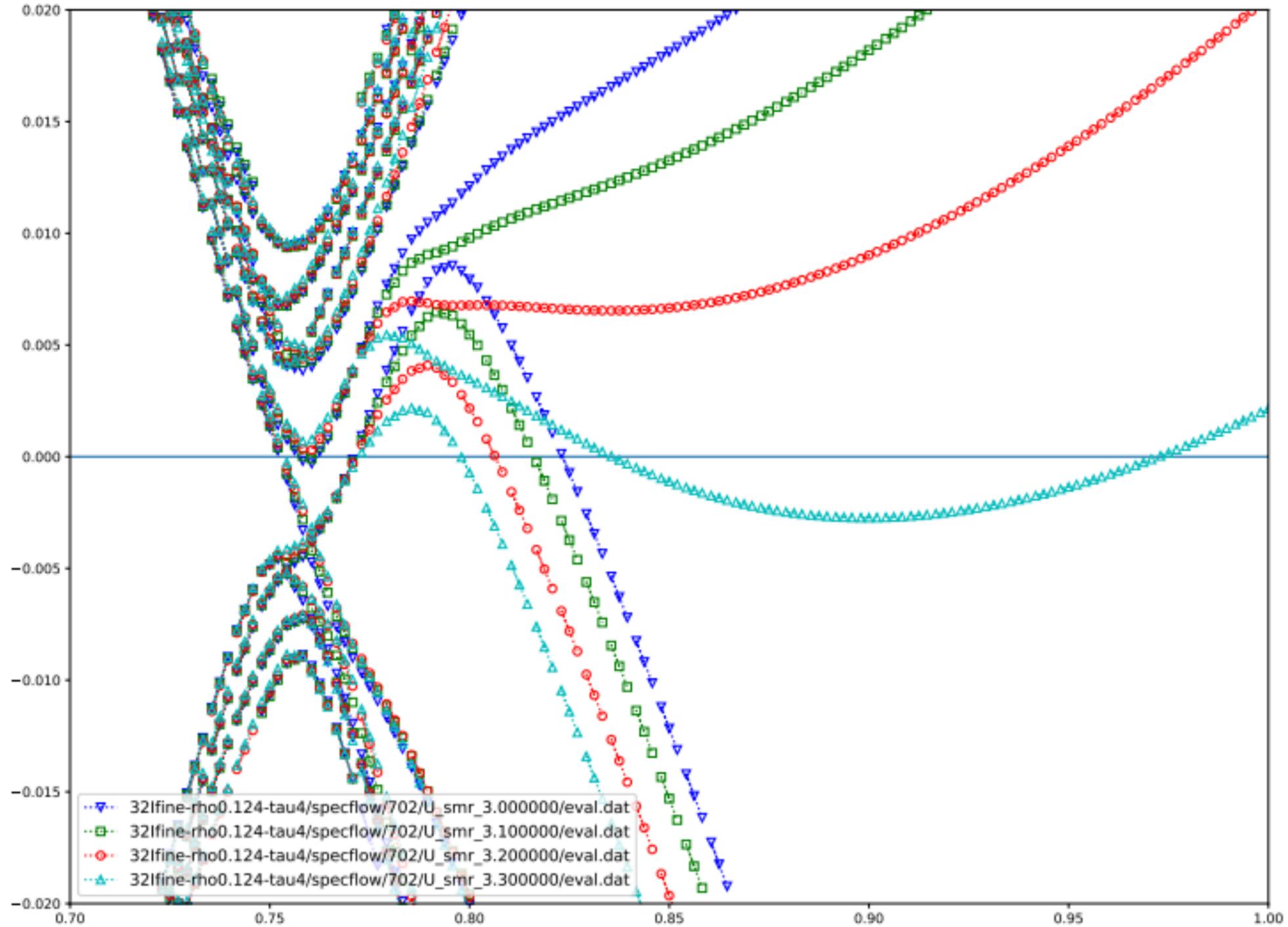
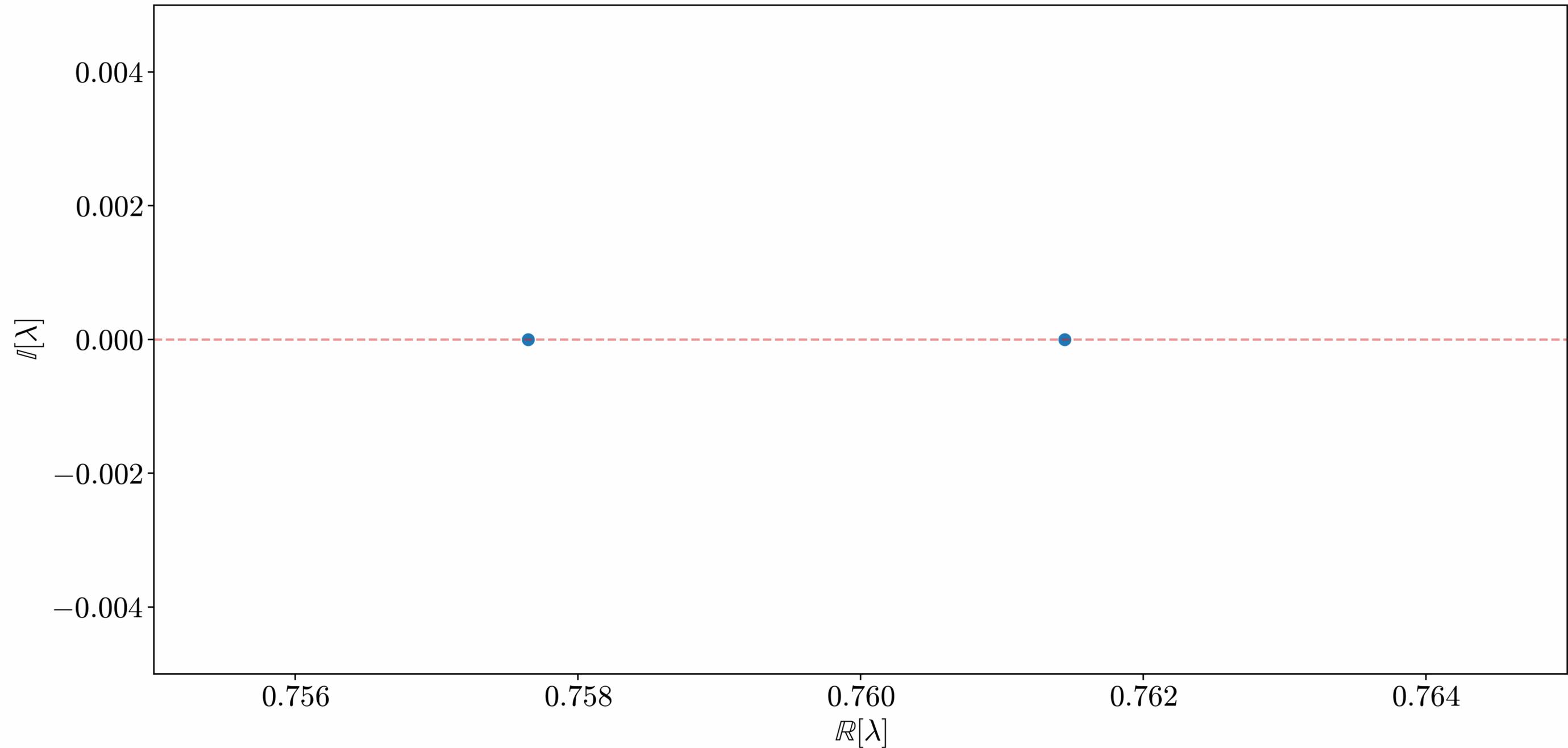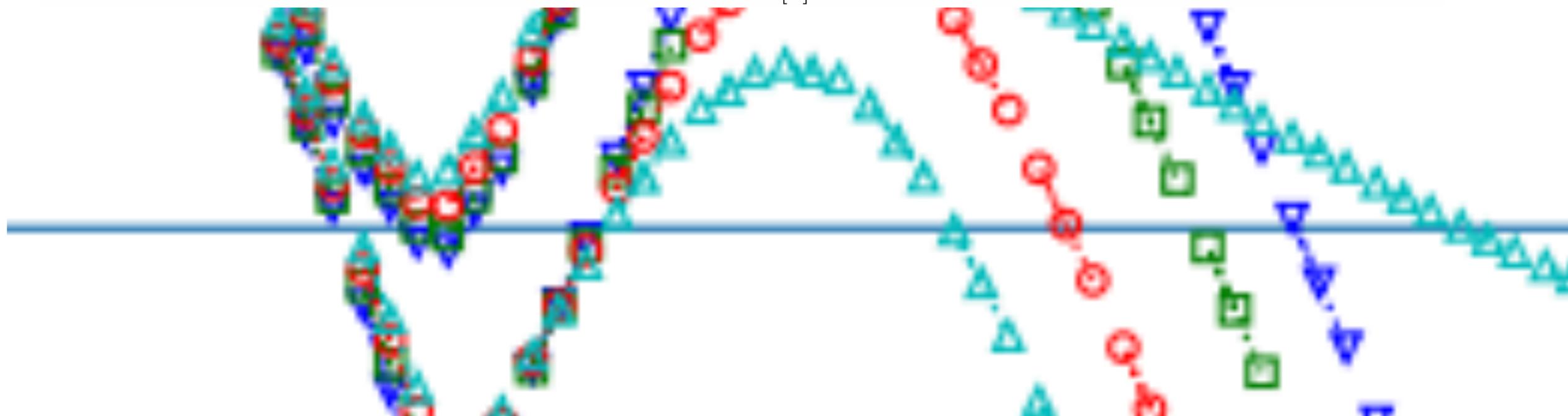# Spectrum Evolution from HKS

$\tau = 3.000000$
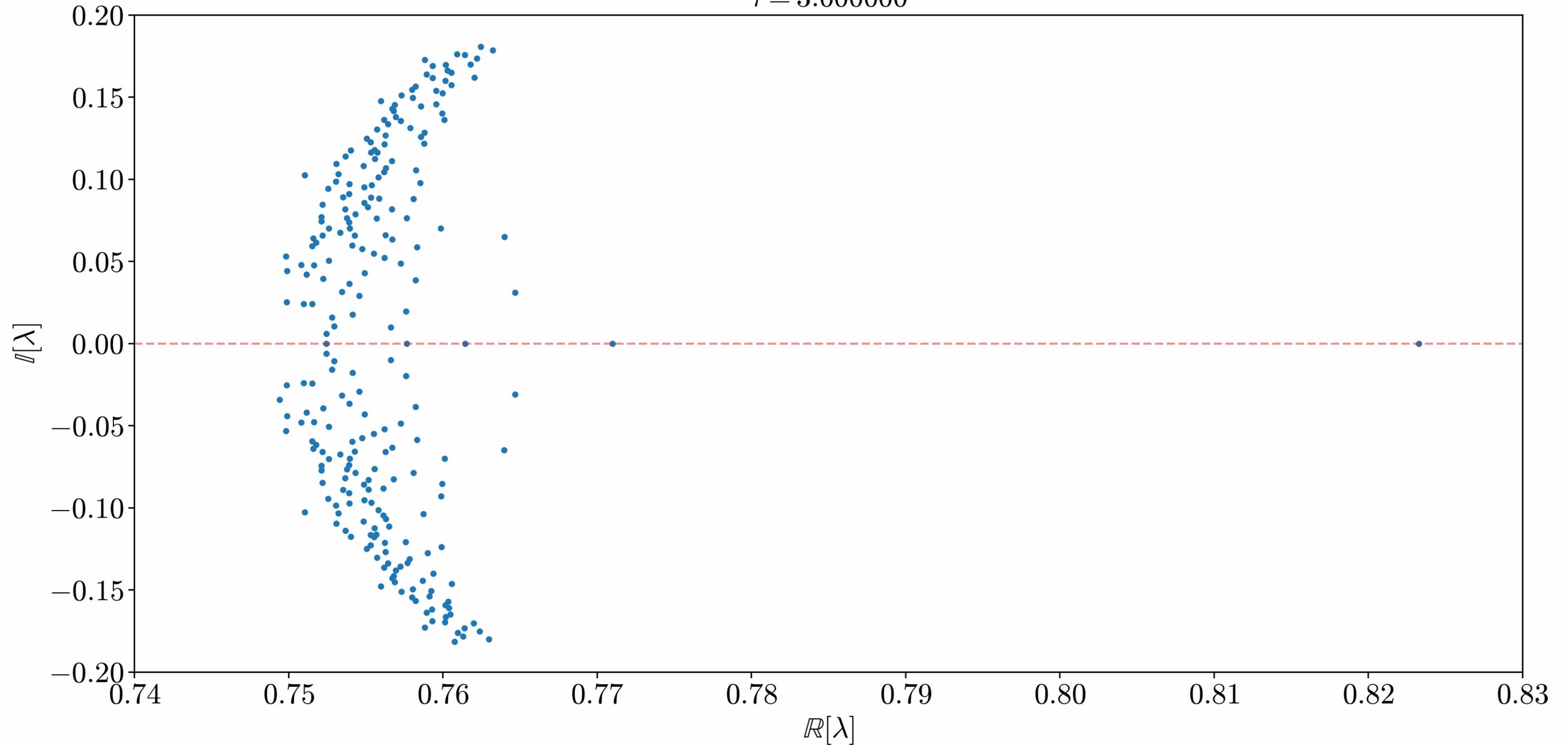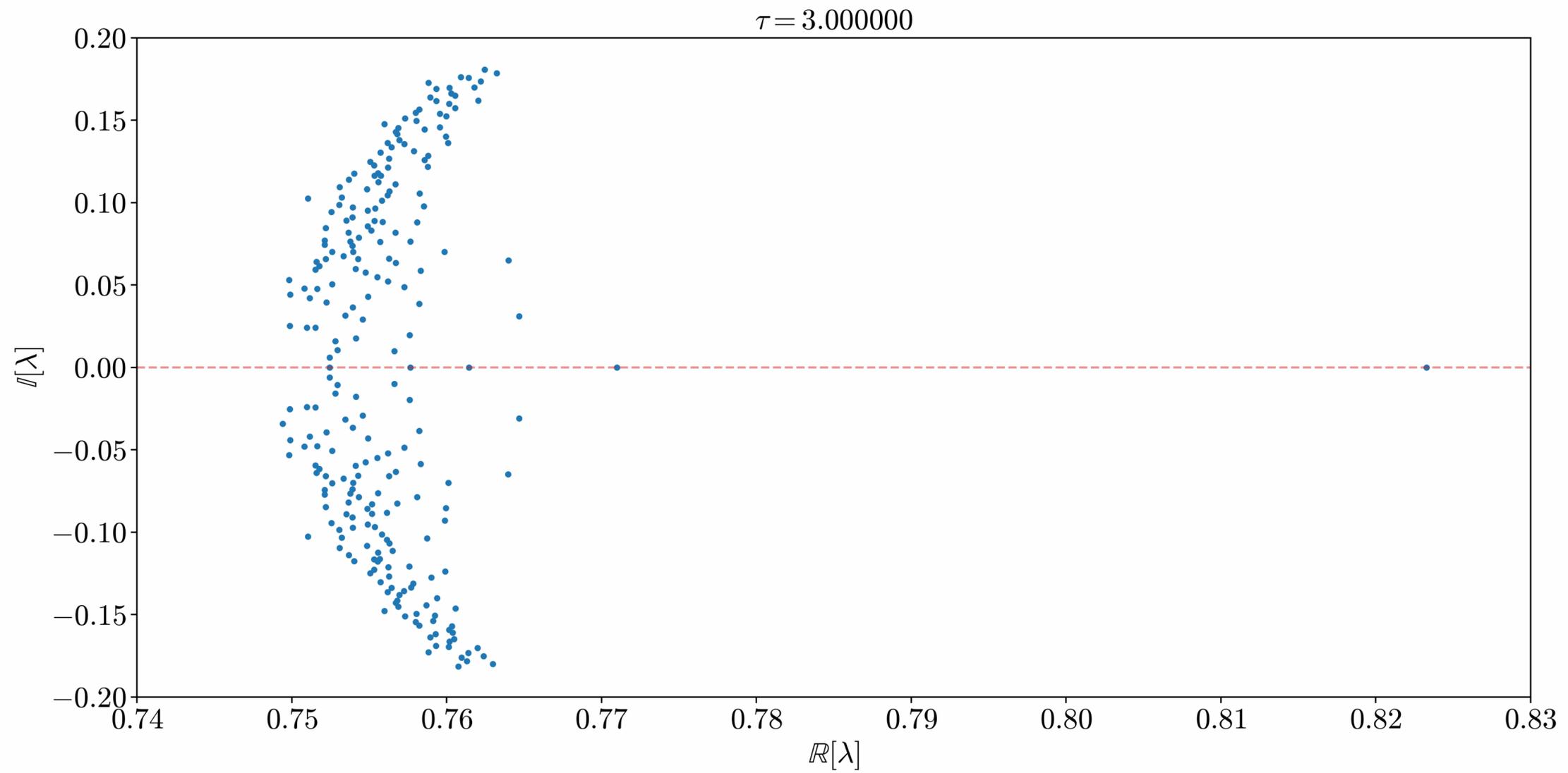
Eigenpair creation
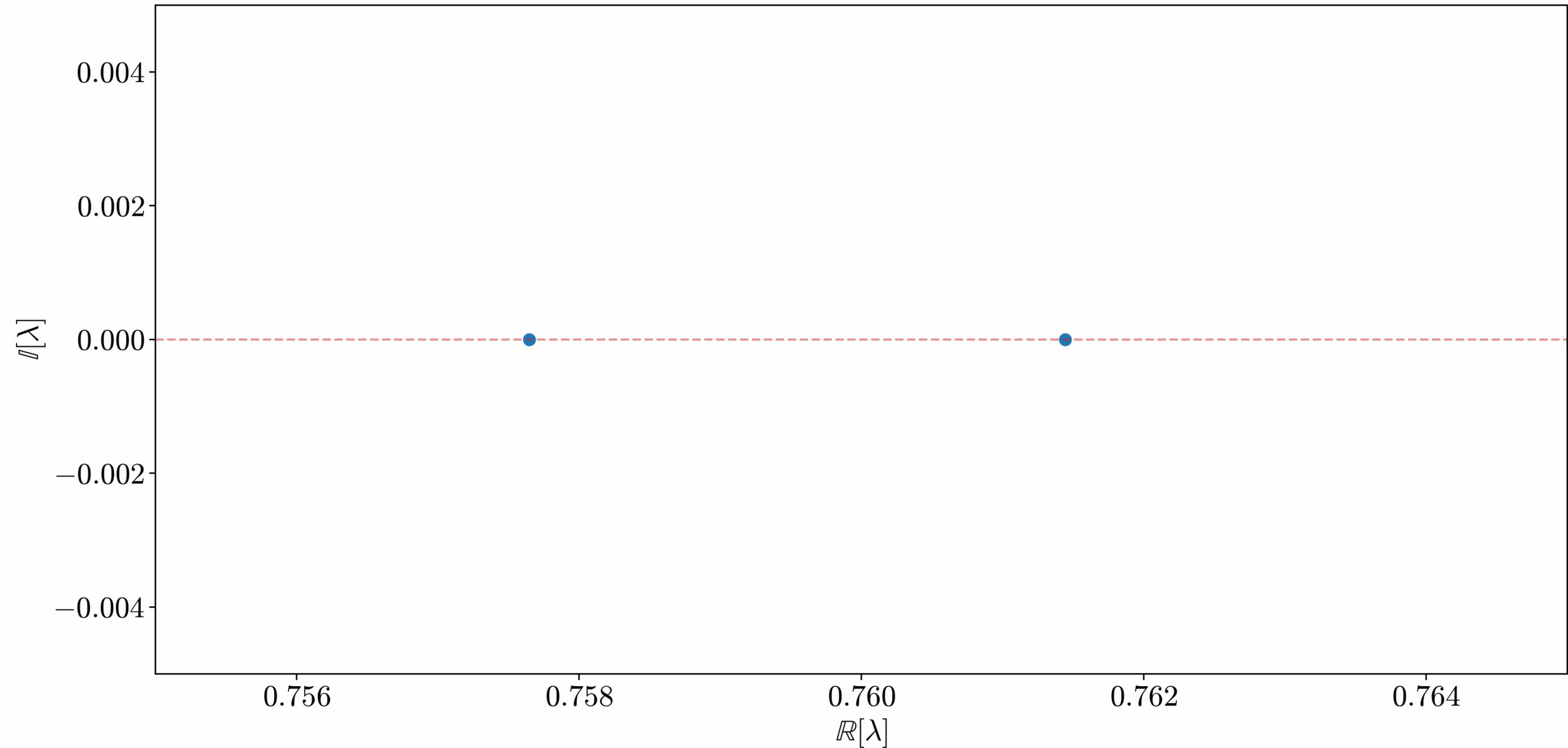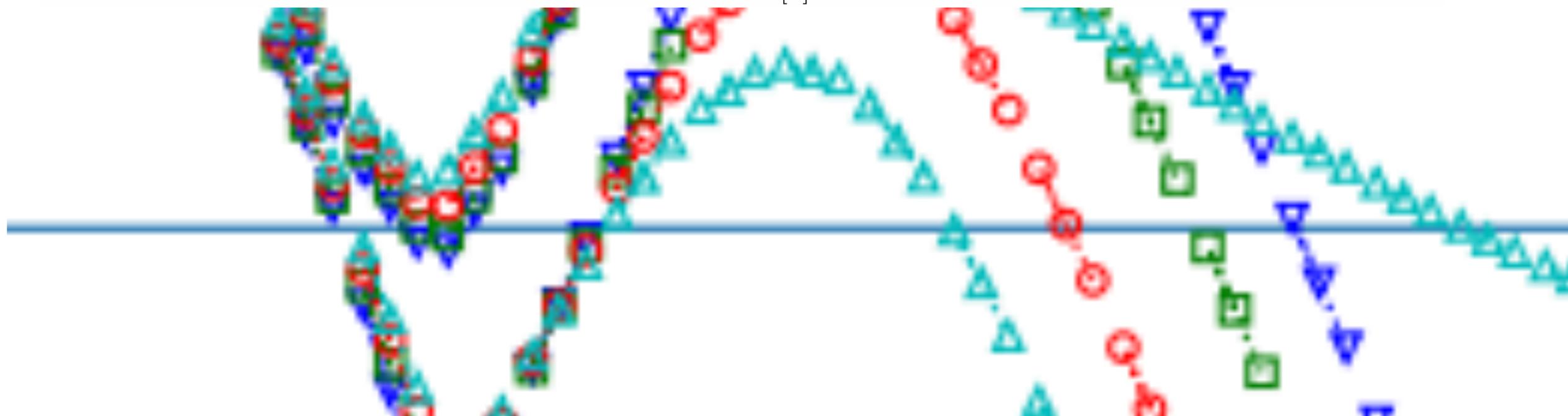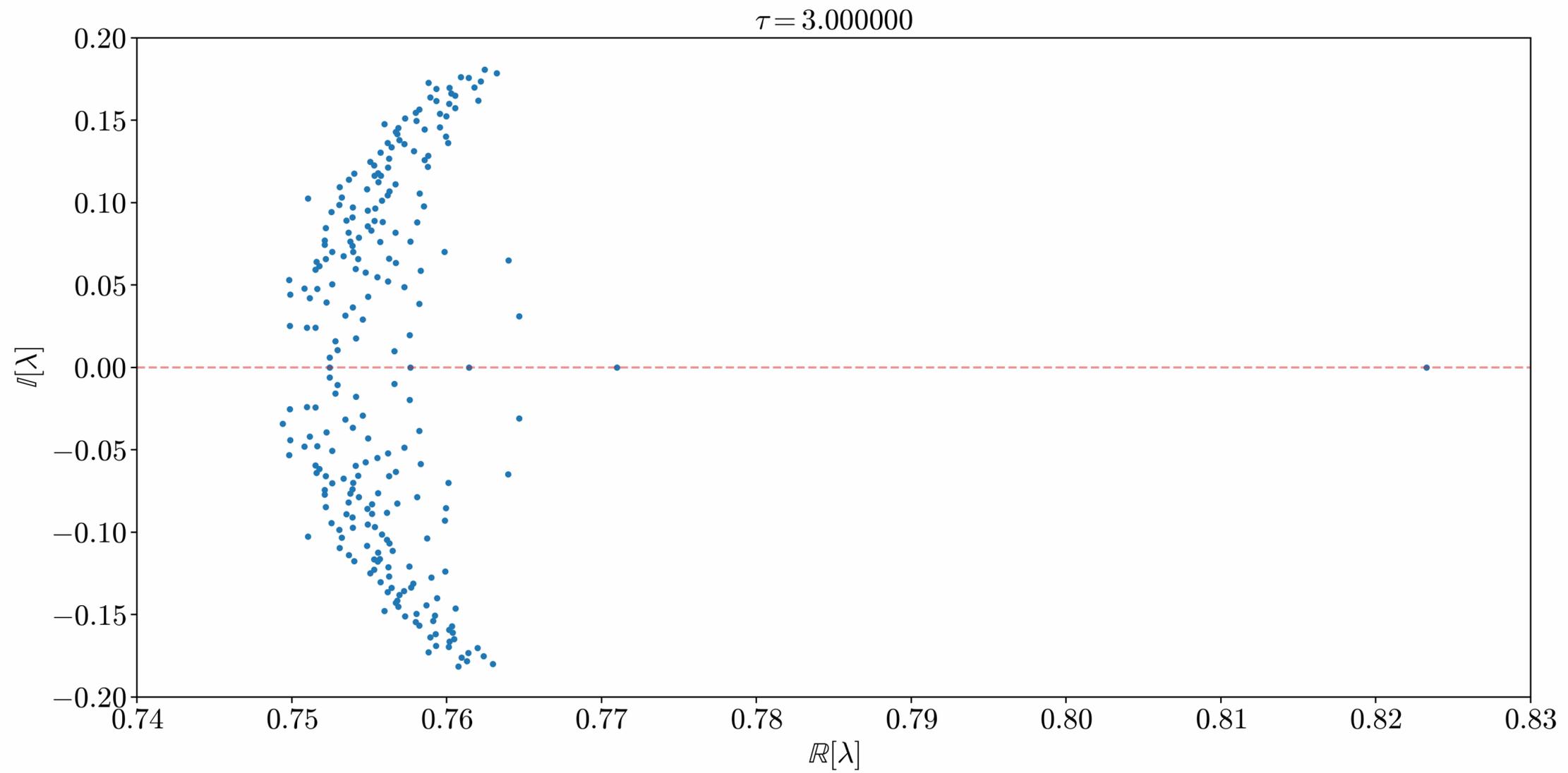

$$\tau = 3.000000$$

Eigenpair creation

$\tau = 3.000000$

$\tau = 3.000000$

Conclusion

- Knowledge of the spectrum of lattice Dirac operators is vital for understanding their physics.

  ‣ Real modes of the Wilson-Dirac operator yield the topological charge.

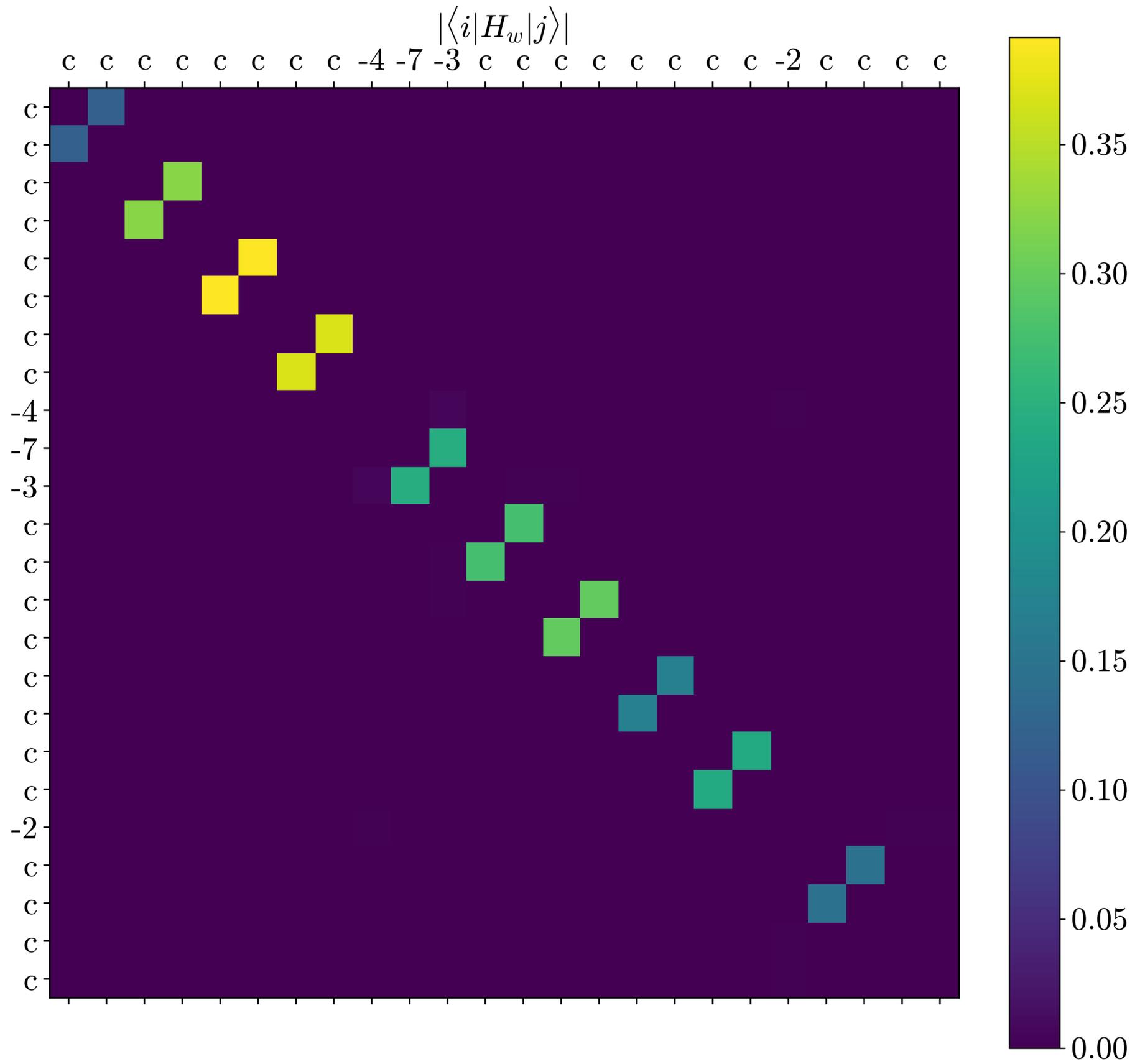# Conclusion

- Knowledge of the spectrum of lattice Dirac operators is vital for understanding their physics.

  ‣ Real modes of the Wilson-Dirac operator yield the topological charge.

- Modes in the interior of the spectrum are difficult to determine numerically.

  ‣ The HKS algorithm allows us to target the spectrum's interior efficiently, but we are still understanding its capabilities.

Conclusion

- Knowledge of the spectrum of lattice Dirac operators is vital for understanding their physics.
  - ‣ Real modes of the Wilson-Dirac operator yield the topological charge.
- Modes in the interior of the spectrum are difficult to determine numerically.
  - ‣ The HKS algorithm allows us to target the spectrum's interior efficiently, but we are still understanding its capabilities.
- The low modes of the spectrum are also closely related to the slow modes of convergence for iterative linear solvers.
- We are looking into using HKS to improve and better understand our multigrid setup for solving the Dirac equation.

Backup Slides

31

$|\langle i|H_w|j\rangle|$

32

$$|\langle \psi' | D_W | \psi \rangle|$$