Algorithms for overlap fermions

Stefan Schaefer

NIC, DESY

Lattice Practice 2006

Overview

1. Chiral symmetry
2. Construction of the overlap operator
3. Solution of linear equations

Chiral symmetry in the continuum

$$D(m=0)\gamma_5 + \gamma_5 D(m=0) = 0$$

- spontaneously broken by interactions: $\langle \Sigma \rangle = -\langle \bar{q} q \rangle \neq 0$
- explicitly broken by quark mass
- index theorem

Important for understanding of low-energy physics

- light meson spectrum
- chiral perturbation theory
- instantons

$\Longrightarrow$ Want to keep it on the lattice

Chiral symmetry on the lattice

Ginsparg, Wilson '82

$$D_0\gamma_5 + \gamma_5 D_0 = \frac{a}{R_0} D_0\gamma_5 D_0$$

D_0 : massless Dirac operator
a : lattice spacing
R_0 : radius of the GW–circle

- correct chiral and flavor symmetries
- $\mathcal{O}(a)$ improvement
- topology can be defined by index of the Dirac operator
- simplified renormalization
- strictly positive fermion determinant

Solutions to the Ginsparg–Wilson Equation

- overlap fermions (Neuberger)
- Domain Wall fermions (Kaplan, Shamir)
- Perfect Action (Hasenfratz, Niedermayer,..)
- approximate solutions (Gattringer et al, Bietenholz)

Use the overlap if you want chiral symmetry to machine precision.

Neuberger's overlap operator

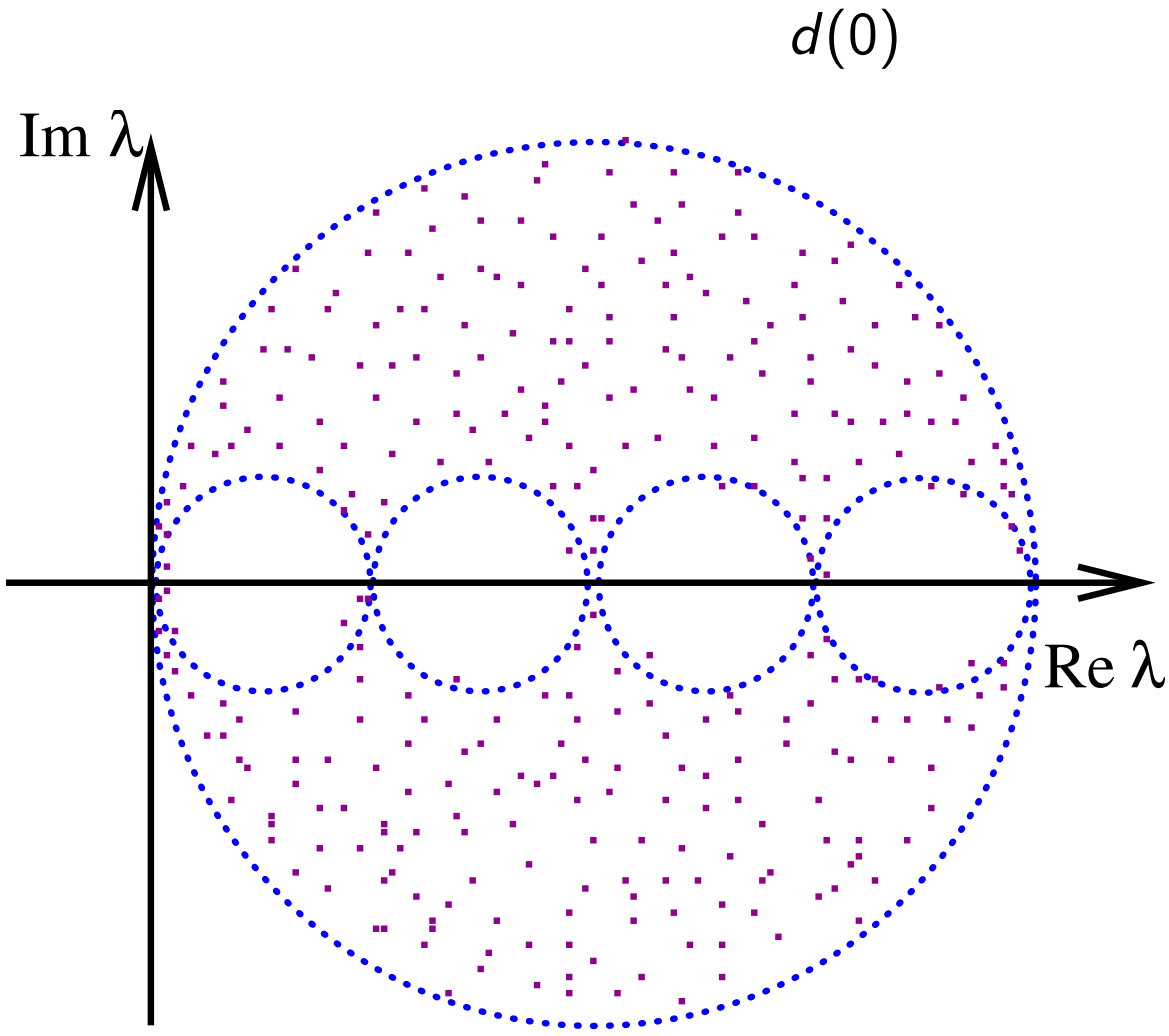
Neuberger '97, '98

$$D_{ov}(m=0) = R_0\Big[\ 1\ +\ d(-R_0)\frac{1}{\sqrt{d^\dagger d(-R_0)}}\ \Big] \qquad (1)$$
$$= R_0\Big[\ 1\ +\ \gamma_5\ \epsilon(h(-R_0))\ \Big]$$

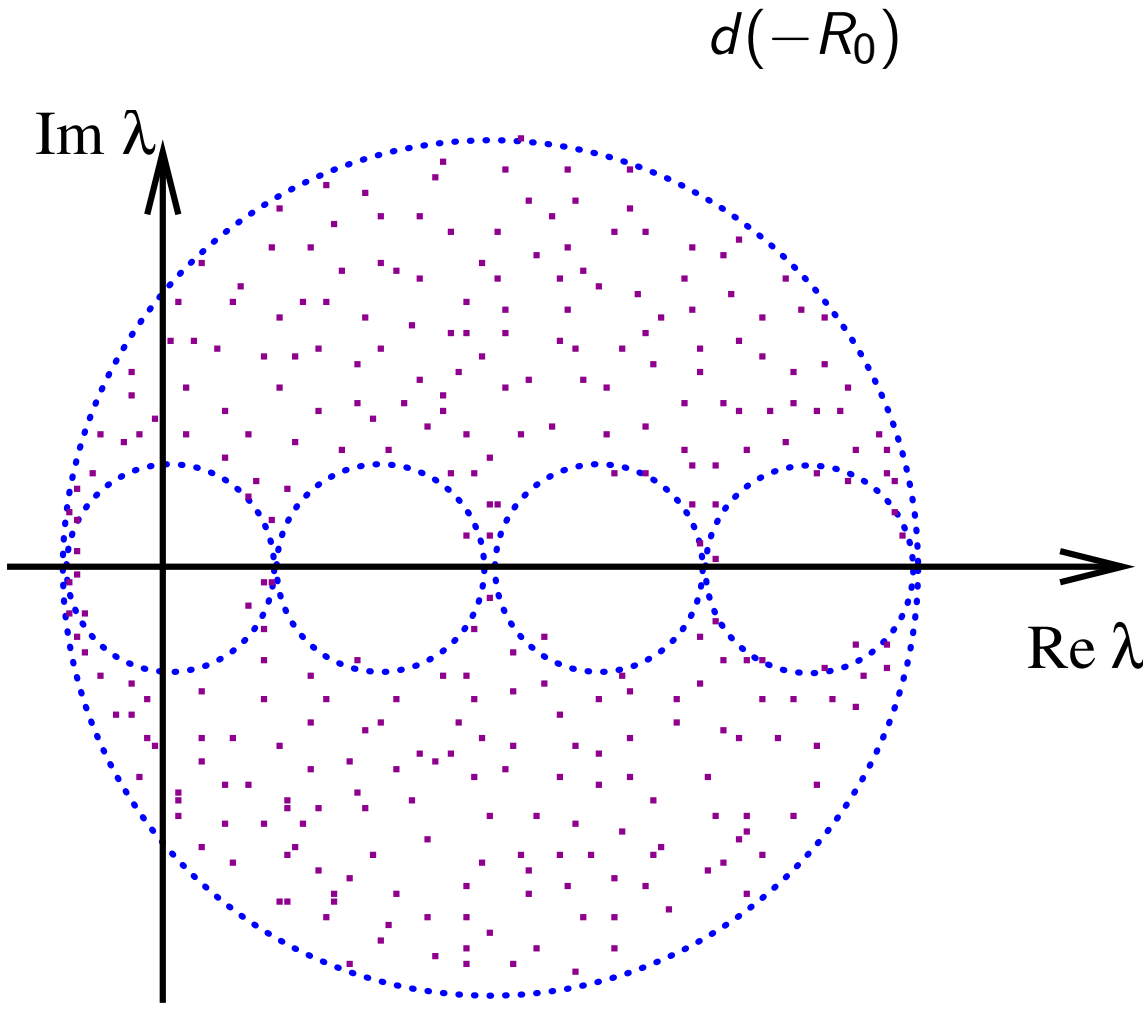
with

- $h = \gamma_5 d$ Hermitian (Wilson) Dirac operator
- $-R_0$ negative mass shift $\propto 1/a$
- ϵ is the matrix sign function

Construction of the overlap

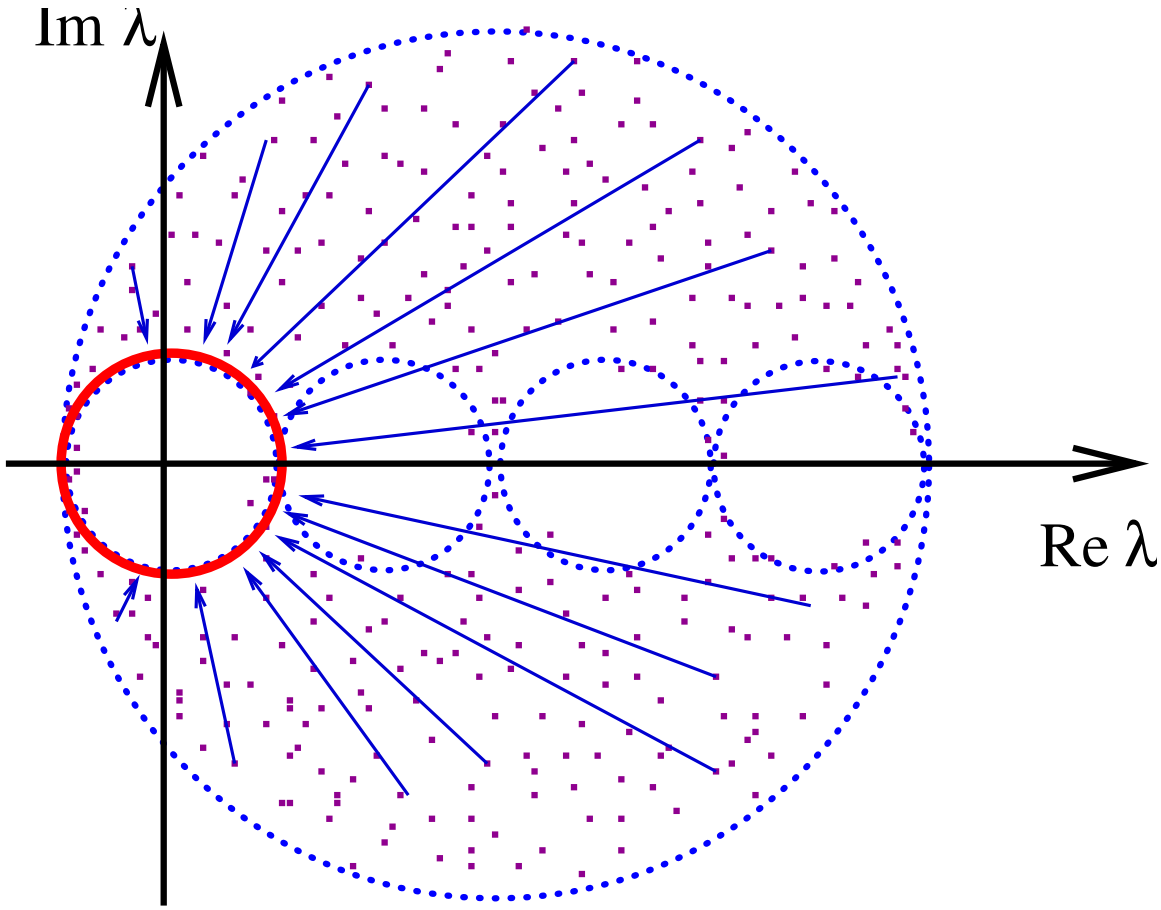


Construction of the overlap



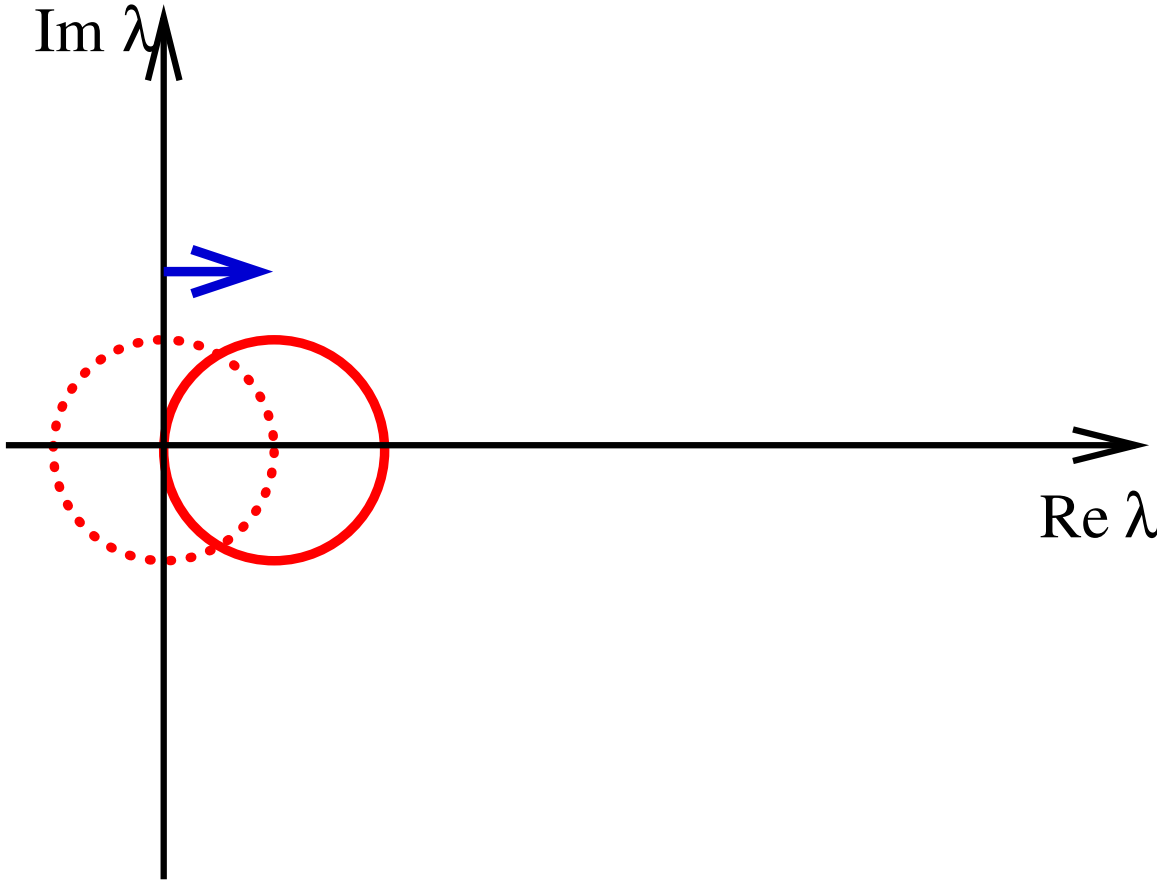
Construction of the overlap

$$d(-R_0)\frac{1}{\sqrt{d(-R_0)^\dagger d(-R_0)}} = \gamma_5 \frac{h(-R_0)}{\sqrt{h(-R_0)^2}} = \gamma_5 \epsilon(h(-R_0))$$



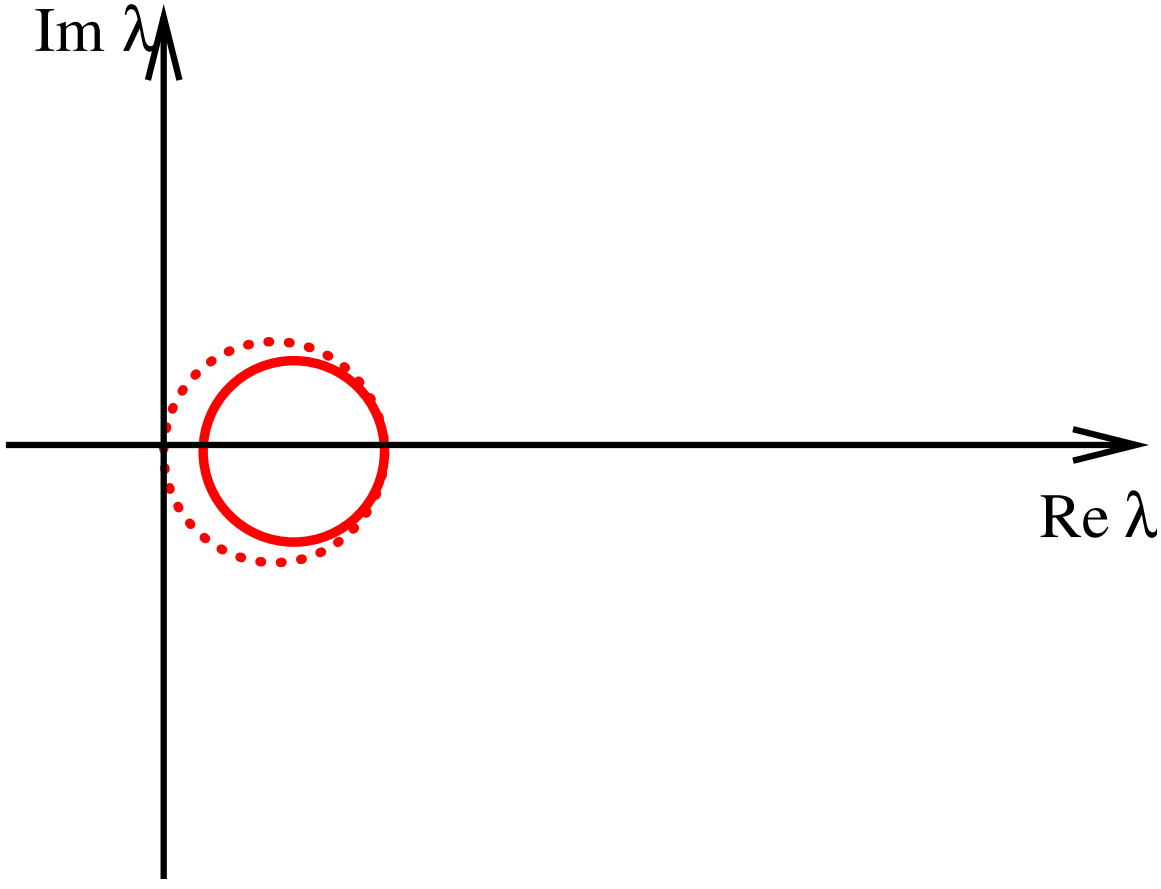
Construction of the overlap

$$D(m=0) = R_0(1+\gamma_5\epsilon(h(-R_0))$$



Construction of the overlap

$$D(m) = (R_0 - \frac{m}{2})(1 + \epsilon(h(-R_0)) + m$$

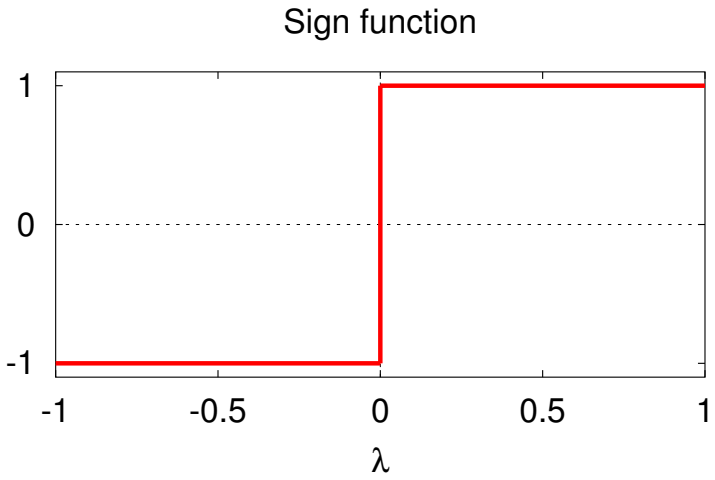


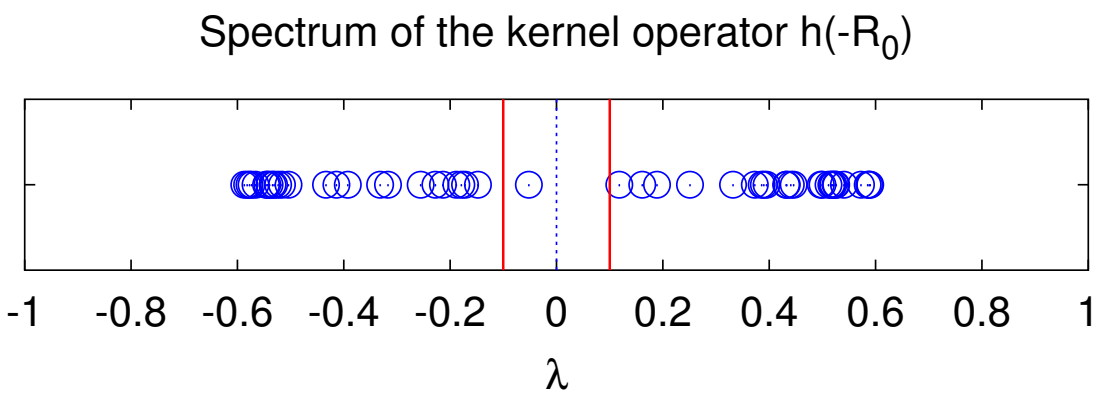
Definition of the matrix sign function

$$\epsilon(h) = \frac{h}{\sqrt{h^2}} = \sum_i \mathrm{sign}(\lambda_i)|\lambda_i\rangle\langle\lambda_i|$$

λ_i : eigenvalues of h

$|\lambda_i\rangle$: associated eigenvectors

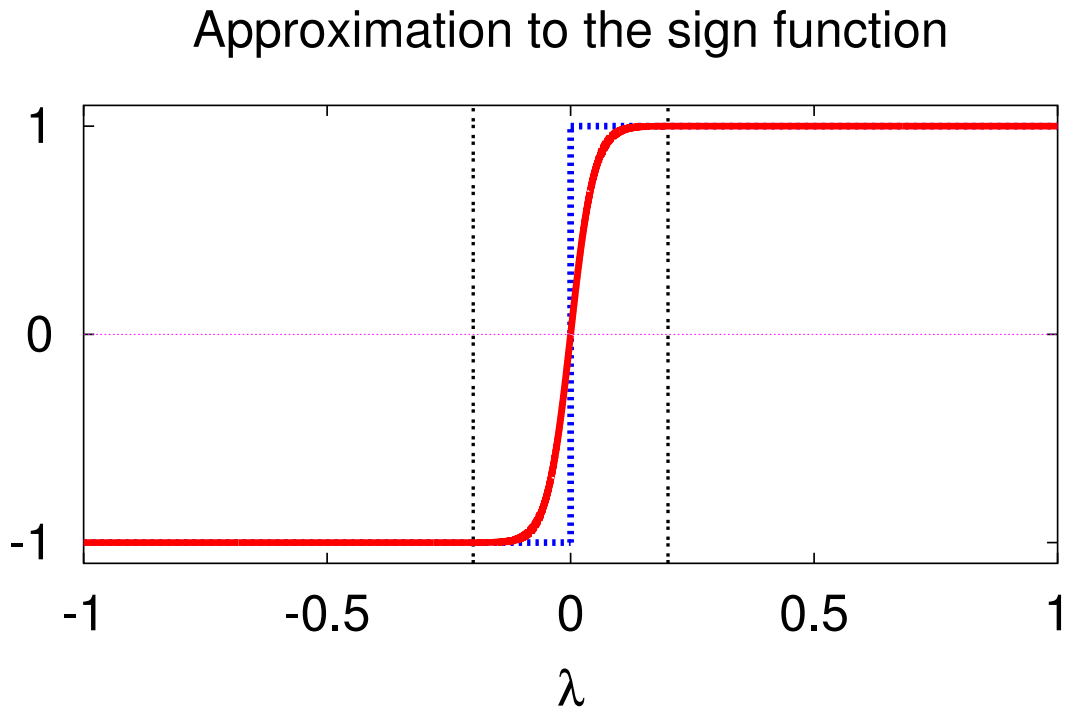




Implementation of the sign function

- Use spectral representation for $|\lambda| < |\lambda_{\min}|$
- Use approximation for the rest of the spectrum.
- Cost of approximation increases for smaller $\lambda_{\min}$

$$\epsilon(h) = \sum_{n=1}^{N} \mathrm{sign}(\lambda_i)|\lambda_i\rangle\langle\lambda_i| + (1 - \sum_{n=1}^{N} |\lambda_i\rangle\langle\lambda_i|)\epsilon_{\mathrm{app}}(h)$$

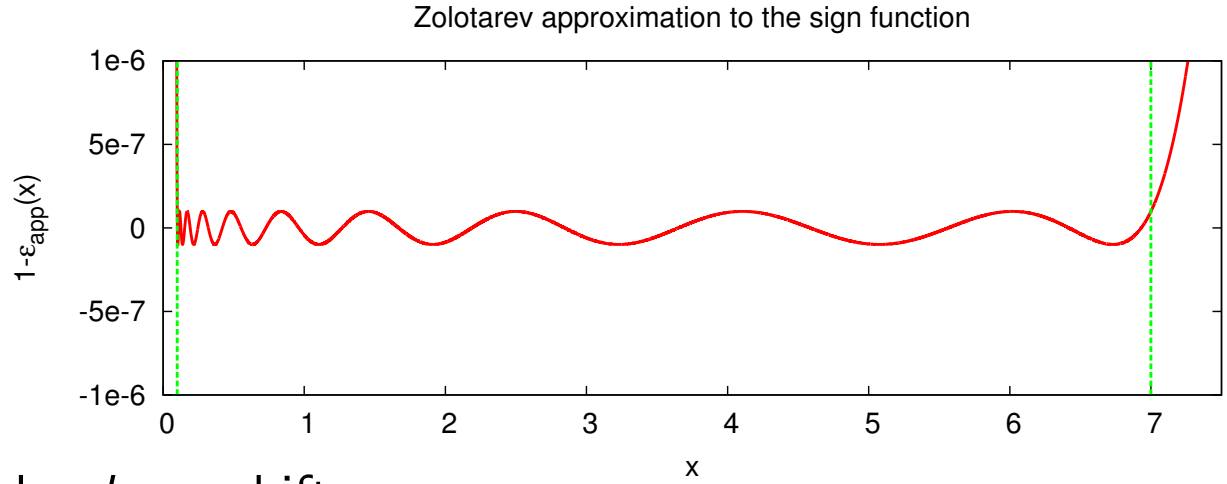


The sign function: Rational approximations

$$\epsilon(h) = h \sum_{i=1}^{M} \frac{b_i}{h^2 + c_i}$$

- coefficients determined by accuracy and range of the approximation
- need to solve systems $(h^2 + c_i)\, x \;=\; v$
- multi-mass
- interplay between accuracy of the solver & precision of the scalar approximation
- need several hundred h on vector multiplications for on D_{ov}.

Example: Zolotarev approximation



n	residue b_n	shift c_n
0	0.037	0.0008
1	0.050	0.009
2	0.078	0.037
3	0.132	0.124
4	0.229	0.395
5	0.406	1.240
6	0.741	3.941
7	1.472	13.244
8	3.841	54.340
9	32.04	601.173

$$\epsilon_{app}(x) = x \sum_n \frac{b_n}{x^2 + c_n}$$

- 10 shifts, $x \in [0.1, 7] \Rightarrow 1 \cdot 10^{-7}$ accuracy
- cost determined by smallest shift
- smallest shift decreases with larger range of approx.

Chebychev Polynomials

Approximate $1/\sqrt{h^2}$ by a sum over Chebychev polynomials

$$\epsilon(h) \approx h \sum_{n=0}^{N} c_n T_n(h^2)$$

- sum is computed with the (recursive) Clenshaw algorithm: numerically very stable
- need several hundred terms
- one parameter determines the accuracy of the approximation

Summary I: Contruction of the overlap operator

$$D_{\mathrm{ov}}(m) = (R_0 - \frac{m}{2})\left[1 + \gamma_5 \epsilon(h(-R_0))\right] + m$$

- expensive part is the application of the sign function on a vector
- project low modes of the kernel operator $h(-R_0)$
- use rational or polynomial approximation for the rest of the spectrum
- several hundred times more expensive than applying Wilson operator
- for larger volume, there are more modes to project

Solving systems of linear equations

$$D_{\mathrm{ov}}^{\dagger}(m) D_{\mathrm{ov}}(m) x = b$$

- use an iterative solver: many tried, best depends on physical situation
- $D_{\mathrm{ov}}^{\dagger} D_{\mathrm{ov}}$ has strict upper bound for conditioning number $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{R_0^2}{m^2}$
- total cost $n_{iter} \times n_{construct}$
- application of $D^{\dagger} D$ as expensive as application of D

Improvements I: adaptive / relaxed precision

Strategy for improvement:

- Krylov methods build solution from $\{(D^\dagger D)^n v; n = 0, ..., N\}$

$$x = \sum_{n=0}^{N} c_i (D^\dagger D)^n v$$

- identify less important contributions and compute $\epsilon(h)w$ with reduced accuracy
  - lower order polynomial
  - less accurate rational approximation / reduced inner solver precision

Recent publications:
Cundy et al, CPC 165:221-242,2005
Chiarappa et al, hep-lat/0609023

Example RelCG

Taken from Cundy et al, CPC 165:221-242,2005

compute x such that $\|D_{\mathrm{ov}}^{\dagger} D_{\mathrm{ov}} x - b\| \leq \epsilon \cdot \|b\|$

$x = 0$; $r = b$; $p = r$;

$\gamma_{old} = \gamma = r^{\dagger} \cdot r$; $\zeta = 1/\gamma$;

while $\sqrt{\gamma} > \epsilon \cdot \|b\|$ **do**

compute q with $\|D_{\mathrm{ov}}^{\dagger} D_{\mathrm{ov}} p - q\| \leq \epsilon \cdot \|b\| \cdot \|p\| \cdot \sqrt{\zeta}$;

$\beta = q^{\dagger} \cdot p$;

$\alpha = \gamma/\beta$;

$x = x + \alpha \cdot p$;

$r = r - \alpha \cdot q$;

$\gamma = r^{\dagger} \cdot r$;

$\zeta = \zeta + 1/\gamma$;

$p = r + \gamma/\gamma_{old} \cdot p$;

$\gamma_{old} = \gamma$;

end while

- Many algorithms can be modified in this way
- key is to find a strategy to relax accuracy and keep convergence under control
- gain depends on the situation and the solver: up to factor 5 reported
- Cundy et al, CPC 165:221-242,2005
- Chiarappa et al, hep-lat/0609023 (computation of inverse 30-120 times more expensive than TM)

Improvements II: Low mode preconditioning

Same trick as used for the construction of the sign function

$$(D_{\mathrm{ov}}^{\dagger}(m)D_{\mathrm{ov}}(m))^{-1}v = \sum_{i=1}^{N} \frac{1}{\lambda_i}\psi_{\lambda_i}\psi_{\lambda_i}^{\dagger}v + (1-P)(D_{\mathrm{ov}}^{\dagger}D_{\mathrm{ov}}(m))^{-1}v$$

- Can be made exact by Schur complement technique (Chiarappa et al, hep-lat/0609023)
- even relatively poor eigenvectors can lead to significant speed-up
- best suited if system is solved for many sources

Conclusions

- chiral symmetry simplifies many lattice computations
- chiral fermions are expensive: need to construct sign function
- state of the art: low-mode projection + rational (Zolotarev) or polynomial (Chebychev) approximation
- solver can be greatly improved my relaxing the accuracy of the sign function
- many solvers have been tested; best setup depends on physical situation
- total cost 30–100 times of twisted mass fermions — in quenched
- dynamical simultations: coarse lattice should be possible