One flavor mass reweighting: foundations

Björn Leder
with Jacob Finkenrath and Francesco Knechtli

Bergische Universität Wuppertal

July 29, 2013

LATTICE 13

JULY 29 – AUGUST 03 2013
MAINZ, GERMANY

Why (mass) reweighting?

Physics

- Tuning of quark masses,
  e.g., m_s in a 2+1 simulation, isospin splitting
- Quark mass dependence
- QED effects, finite chemical potential, ...

Algorithms

- Stabilization of HMC algorithm (next talk J. Finkenrath)
- Alternative update algorithms ([Comp.Phys.Comm. 2013])

Applied Maths

- Stochastic determination of determinants

Why (mass) reweighting?

Physics

- Tuning of quark masses,
  e.g., m_s in a 2+1 simulation, isospin splitting
- Quark mass dependence
- QED effects, finite chemical potential, ...

Algorithms

- Stabilization of HMC algorithm (next talk J. Finkenrath)
- Alternative update algorithms ([Comp.Phys.Comm. 2013])

Applied Maths

- Stochastic determination of determinants

Reweighting

Observable in lattice QCD

$$\langle \mathcal{O} \rangle_a = \frac{1}{Z_a} \int \mathrm{D}[U]\; P_a(U)\; \mathcal{O}_a(U)$$

- with weight $P_a(U) = \mathrm{e}^{-S_g(\beta,U)} \prod_{i=1}^{n_f} \det(D(U) + m_i)$
- bare parameter set $a = \{\beta, m_1, m_2, \ldots, m_{n_f}\}$
- and normalization $\int \mathrm{D}[U] P(U)/Z \equiv 1$

Observable at $b = \{\beta', m_1', m_2', \ldots, m_{n_f}'\}$

$$\langle \mathcal{O} \rangle_b = \frac{\langle \mathcal{O} W_{a,b} \rangle_a}{\langle W_{a,b} \rangle_a}, \quad \text{reweighting factor} \quad W_{a,b} = \frac{P_b}{P_a}$$

In practice Monte Carlo:

- generate $\{U_i\}$ according to $P_a(U)$
- use $\mathcal{O}(U)$ and $W_{a,b}(U)$ to compute $\langle \mathcal{O} \rangle_b$

Reweighting

Reweighting factors

- beta shift: $W_{\beta,\beta'} = \mathrm{e}^{-(S_g(\beta',U)-S_g(\beta,U))}$
- one flavor: $W_{m_i,m_i'} = \frac{\det(D(U)+m_i')}{\det(D(U)+m_i)}$
- all other are products of these two
- two flavor: $W_{m_i,m_i'} W_{m_j,m_j'}$
- two degenerate: $W_{m,m'}^2$, isospin: $W_{m,m-\Delta m} W_{m,m+\Delta m}$

In [arXiv:1306.3962]:

1. proof of integral representation of the complex determinant of a complex matrix
2. unbiased estimator with controlled stochastic error
3. expansions of stochastic error and ensemble fluctuations
4. based on 3: detailed scaling analysis and optimized reweighting strategies

Reweighting

Reweighting factors

- beta shift: $W_{\beta,\beta'} = \mathrm{e}^{-(S_g(\beta',U)-S_g(\beta,U))}$
- one flavor: $W_{m_i,m_i'} = \frac{\det(D(U)+m_i')}{\det(D(U)+m_i)}$
- all other are products of these two
- two flavor: $W_{m_i,m_i'}W_{m_j,m_j'}$
- two degenerate: $W^2_{m,m'}$, isospin: $W_{m,m-\Delta m}W_{m,m+\Delta m}$

In [arXiv:1306.3962]:

1. proof of integral representation of the complex determinant of a complex matrix
2. unbiased estimator with controlled stochastic error
3. expansions of stochastic error and ensemble fluctuations
4. based on 3: detailed scaling analysis and optimized reweighting strategies

Integral representation

Let $A \in \mathbb{C}^{n\times n}$ and $\eta \in \mathbb{C}^n$, then

$$\frac{1}{\det A} = \int \mathrm{D}[\eta]\, \mathrm{e}^{-\eta^\dagger A \eta} \quad \text{if} \quad \mathrm{Re}(\lambda) > 0\,,\ \forall \lambda \in \sigma(A)$$

- with $\mathrm{D}[\eta] = \prod_{i=1}^{n} \frac{d\mathrm{Re}(\eta_i)\, d\mathrm{Im}(\eta_i)}{\pi}$
- if A Hermitian, then $\mathrm{Re}(\lambda) > 0$ is equiv. to A positive definite and therefore $A = LL^\dagger$ exists and the determinant is the Jacobian of $\eta \to L^{\dagger^{-1}}\eta$
- for non-Hermitian A the proof uses the Schur decomposition $A = QUQ^{-1}$

Numerical evaluation: Monte Carlo

Let $\{\eta^{(k)},\, k=1,\dots,N_\eta\}$ be distributed as $p(\eta)$, then

$$W = \frac{1}{\det A} = \left\langle \frac{\mathrm{e}^{-\eta^\dagger A \eta}}{p(\eta)} \right\rangle_{p(\eta)} = \frac{1}{N_\eta} \sum_{k=1}^{N_\eta} \frac{\mathrm{e}^{-{\eta^{(k)}}^\dagger A \eta^{(k)}}}{p(\eta^{(k)})} + \mathrm{O}(1/\sqrt{N_\eta})$$

- where $\langle \mathcal{O} \rangle_{p(\eta)} = \int \mathrm{D}[\eta]\, p(\eta) \mathcal{O}(\eta)$
- convenient to choose a Gaussian $p(\eta) = \exp(-\eta^\dagger \eta)$
- variance: $\sigma_\eta^2 = \frac{1}{\det(A+A^\dagger-1)} - \frac{1}{\det(AA^\dagger)}$
- convergence of integral repr. of variance if $\lambda(A+A^\dagger-1) > 0 \rightarrow \mathrm{Re}(\lambda(A)) > 0.5$
- monitor variance to guarantee convergence of mean
- if $A = 1 + \epsilon B$ with $\epsilon ||B|| \ll 1$, can approximate

$$\frac{\sigma_\eta^2}{|W|^2} = \det\left(1 + \epsilon^2 \frac{BB^\dagger}{1+\epsilon(B+B^\dagger)}\right) - 1 = \epsilon^2 \mathrm{Tr}(BB^\dagger) + \mathrm{O}(\epsilon^3)$$

One flavor reweighting factor

Expansion in Δm

$$W(U) = \frac{1}{\det M(U)}, \quad M = \frac{D_m}{D_m - \Delta m} = 1 + \frac{\Delta m}{D_m} + \mathrm{O}(\Delta m^2)$$

with $D_m = D(U) + m$ and $\Delta m = m - m'$

Stochastic noise:

- $\frac{\sigma_\eta^2}{|W|^2} = \Delta m^2 \mathrm{Tr}((D_m D_m^\dagger)^{-1}) + \mathrm{O}(\Delta m^3)$
- reduction: factorization of determinant
- mass interpolation [Hasenbusch (2001), Hasenfratz,Hoffmann,Schaefer(2008)]
- Nth-root [Hasenbusch (2001)]

One flavor reweighting factor

Expansion in Δm

$$W(U) = \frac{1}{\det M(U)}\,, \quad M = \frac{D_m}{D_m - \Delta m} = 1 + \frac{\Delta m}{D_m} + \mathrm{O}(\Delta m^2)$$

with $D_m = D(U) + m$ and $\Delta m = m - m'$

Stochastic noise:

- $\frac{\sigma_\eta^2}{|W|^2} = \Delta m^2 \mathrm{Tr}((D_m D_m^\dagger)^{-1}) + \mathrm{O}(\Delta m^3)$
- reduction: factorization of determinant
- mass interpolation [Hasenbusch (2001), Hasenfratz,Hoffmann,Schaefer(2008)]
- Nth-root [Hasenbusch (2001)]

Mass interpolation

Factorization

$$W = \prod_{j=0}^{N-1} \frac{1}{\det M_l}\,, \quad M_l = \frac{D_l}{D_l - \delta m} = I + \frac{\delta m}{D_l} + \mathrm{O}(\delta m^2)$$

with $D_l = D_m - l\delta m$ and $\delta m = \Delta m/N = (m - m')/N$

Relative stochastic error

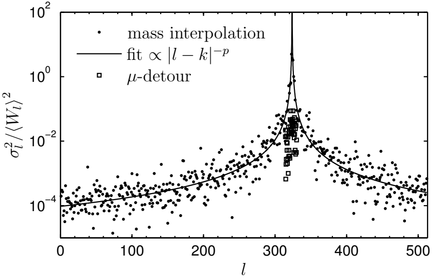
$$W = \prod_{l=0}^{N-1} W_l\,, \quad W_l = \frac{1}{N_\eta} \sum_{k=1}^{N_\eta} \mathrm{e}^{-\eta^{(k,l)\dagger}(M_l - 1)\eta^{(k,l)}}$$

$$\delta_\eta^2 = \frac{N}{N_\eta}\left[\delta m^2 \mathrm{Tr}((D_m D_m^\dagger)^{-1}) + \mathrm{O}(\delta m^2 \Delta m)\right] \approx k_\eta \frac{\Delta m^2 V}{N N_\eta}$$

higher order terms negligible for $N \gtrsim 8$ [PoS Lattice2012 190], i.e., for $\delta m ||D_m^{-1}|| = \Delta m/(\overline{m} N) \lesssim 1/16$

Zero crossings

$N_f = 2$ O(a) impr. Wilson fermions at $\beta = 5.3$ ($a = 0.066\,\mathrm{fm}$)



- $m_{PS} = 440\,\mathrm{MeV}$ or $\overline{m} \approx 33\,\mathrm{MeV} \approx \overline{m}_s/3$
- reweighting to $\overline{m} \approx \overline{m}_s/6$ (D5→D6 CLS)
- peak: real ev crosses zero
- fit: $k \approx 322$, $p \approx 1.8$

μ-detour

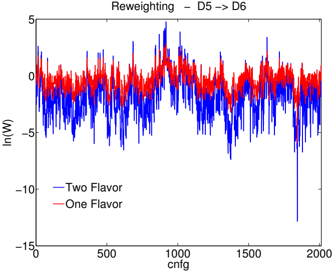
- singular values of $D_{m,\mu} = D_m + i\mu\gamma_5$ are $\geq |\mu|$
- half circle arround $l = 322$, penetration depth $\mu/\overline{m} =?$
- determinant ratio complex for $\mu \neq 0$
- imaginary parts add up to the correct phase -1

Ensemble fluctuations

Variance of reweighting factor

$$\frac{\sigma_{1f}^2}{\langle W\rangle^2} = \frac{\langle W^2\rangle}{\langle W\rangle^2} - 1\,, \quad W = \exp\left[\mathrm{Tr}\,\ln\left(1 - \frac{\Delta m}{D_m}\right)\right]$$

$$\frac{\sigma_{1f}^2}{\langle W\rangle^2} = \Delta m^2\left[\left\langle(\mathrm{Tr}(D_m^{-1})^2\right\rangle - \left\langle\mathrm{Tr}(D_m^{-1})\right\rangle^2\right] + \mathrm{O}(\Delta m^3)$$



- $\sigma_{1f}^2/\langle W\rangle^2 \approx k_{1f}\Delta m^2 V$
- stochastic error: $\delta_\eta^2 \approx k_\eta \frac{\Delta m^2 V}{N N_\eta}$
- ratio: $\frac{\langle\delta_\eta^2\rangle}{\sigma_{1f}^2/\langle W\rangle^2} \approx \frac{\langle k_\eta\rangle}{k_{1f} N N_\eta}$ with $\langle k_\eta\rangle / k_{1f} = 2 - 3$ for $0 \le \Delta m \le m/2$ and $\overline{m}_s/6 \le \overline{m} \le 4/3 \cdot \overline{m}_s$

(Anti-)Correlations

Idea

Reduce ensemble fluctuations by combining two or more reweighting factors that are anti-correlated.

beta-shift

Fix $\beta' = \beta + \Delta\beta$ by minimizing the variance of

$$\ln(W_{\beta,\beta'} W_{m,m'}) = -\Delta\beta S_g(U) + \mathrm{Tr}\,\ln\left(1 - \frac{\Delta m}{D_m}\right)$$

Combining two flavors

$$W_{2\mathrm{f}}^{(\gamma)} = W_{m_r, m_r - \gamma\Delta m} W_{m_s, m_s + \Delta m}$$

Two flavor reweighting $W_{2f}^{(\gamma)} = W_{m_r, m_r - \gamma\Delta m} W_{m_s, m_s + \Delta m}$

Combining two flavors

Without loss of generality $m_r \leq m_s$, define $m_\pm = (m_s \pm m_r)/2$

$$W_{2f}^{(\gamma)} = \det \left[I - \Delta m \frac{(\gamma - 1) D_{m_+} + (\gamma + 1) m_- + \gamma \Delta m}{D_{m_+}^2 - m_-^2} \right]$$

for deg. masses $m_r = m_s$ special cases $\gamma = -1, 0, 1$:
$W_{2f}^{(0)} = W_{m,m'}$, $W_{2f}^{(-1)} = W_{m,m'}^2$, isospin $W_{2f}^{(1)} = W_\pm$

Non-degenerate $m_r \neq m_s$

- $\gamma = 1$ means keeping sum of bare masses constant
- fix γ by minimizing the the ensemble fluctuations:

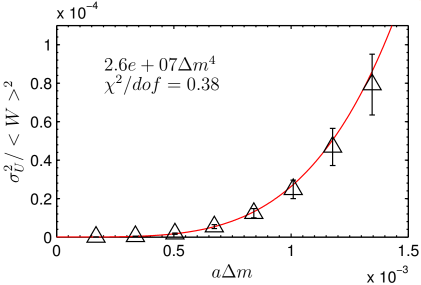
$$\gamma^* \approx 1 - 2m_-^2 \frac{k_\pm}{k_{1f}} + \mathrm{O}(\Delta m, m_-^3)$$

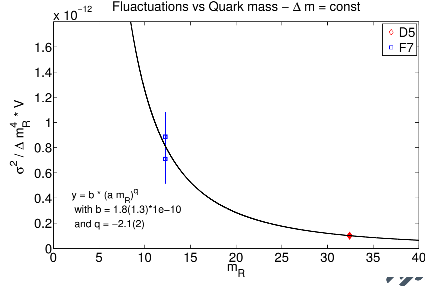
Isospin reweighting

Statistical error and ensemble fluctuations $\propto \Delta m^4$ since:

$$W_\pm = W_{m,m+\Delta m} W_{m,m-\Delta m} = \det\left[1 - \left(\frac{\Delta m}{D_m}\right)^2\right]$$

$$\frac{\sigma_\pm^2}{\langle W_\pm \rangle^2} = \Delta m^4 \left[\left\langle (\mathrm{Tr}(D_m^{-2}))^2 \right\rangle - \left\langle \mathrm{Tr}(D_m^{-2}) \right\rangle^2\right] + \mathrm{O}(\Delta m^6)$$





$$\sigma_\pm^2 / \langle W_\pm \rangle^2 \approx k_\pm \frac{\Delta m^4 V}{\overline{m}^2}$$

Recap

[arXiv:1306.3962]

1. proof of integral representation of the complex determinant of a complex matrix ($\mathrm{Re}(\lambda) > 0$)
2. unbiased estimator with controlled stochastic error based on mass interpolation and optional μ-detour (zero-crossings, correct complex phase)
3. expansions of stochastic error and ensemble fluctuations
4. based on 3: detailed scaling analysis and optimized reweighting strategies (isospin rew., strange mass rew.)

Fluctuations of isospin reweighting factor:

$$\sigma_{\pm}^2 / \langle W_{\pm} \rangle^2 \approx k_{\pm} \frac{\Delta m^4 V}{\overline{m}^2}$$

- ~ 0.001 at $m_{PS} = 270\,\text{MeV}$, $V = 96 \times 48^3$ (F7 CLS)
- $\rightarrow \sim 0.05$ at $m_{PS} = 135\,\text{MeV}$, $V = 128 \times 64^3$!!

Proof using $A = Q(D+K)Q^{-1}$

$$\int \mathrm{D}[\eta]\ \mathrm{e}^{-\eta^\dagger A \eta} \stackrel{(1)}{=} \int \mathrm{D}[\eta]\ \mathrm{e}^{-\eta^\dagger (D+K)\eta} \stackrel{(2)}{=} \int \mathrm{D}[\eta]\ \mathrm{e}^{-r^T D r - s^T D s}$$

$$\stackrel{(3)}{=} \int \left(\prod_i^n \frac{dr_i\, ds_i}{\pi}\right)\ \mathrm{e}^{-r^T D r - s^T D s} \stackrel{(4)}{=} \prod_i^n \frac{1}{\sqrt{\lambda_i}\sqrt{\lambda_i}} = \frac{1}{\det A}$$

(1) change of variables $\eta \to Q\eta$, $\det(Q) = 1$

(2) rewrite exponent using
$r = \frac{1}{2}[\eta + (I + D^{-1}K^T)\eta^*]$, $s = -\frac{i}{2}[\eta - (I + D^{-1}K^T)\eta^*]$

(3) change of variables $(\mathrm{Re}(\eta), \mathrm{Im}(\eta)) \to (r, s)$, $\det(M) = 1$

$$\begin{pmatrix} r \\ s \end{pmatrix} = M \begin{pmatrix} \mathrm{Re}(\eta) \\ \mathrm{Im}(\eta) \end{pmatrix},$$

$$M = \begin{pmatrix} I + \frac{1}{2}D^{-1}K^T & -\frac{i}{2}D^{-1}K^T \\ \frac{i}{2}D^{-1}K^T & I + \frac{1}{2}D^{-1}K^T \end{pmatrix}$$

(4) Gaussian integration $\mathrm{Re}(\lambda) > 0$

Domain decomposition

Factorization [Comp.Phys.Comm. 2013]

$$\det(D+m) = \det(\hat{D}_m)\,\det(D_{bb}+m)\,\det(D_{ww}+m)$$

- reweighting factor: $W = \hat{W}\, W_{bb}\, W_{ww}$
- exact: W_{bb}, W_{ww} $\rightarrow$ determinants of $(12 \cdot N_D)^2$ matrices
- special case domain size $N_D = 1^4$: even-odd
- exact treatment of UV-modes
- stochastic: $\hat{W} = \prod_{l=0}^{N-1} \frac{1}{\det(\hat{M}_l)}$
- combination of DD and mass interpolation

Number of inversions?

(i) mass interpolation steps for controlled stochastic variance:

$$N \gtrsim 16\, \delta m ||D_m^{-1}|| = 16\, \frac{\Delta m}{\overline{m}}$$

(ii) noise sources/factor to monitor stochastic variance:

$$N_\eta \gtrsim 6$$

(iii) stochastic error much smaller than ensemble fluctuations:

$$N N_\eta = \frac{\sigma_{1\mathrm{f}}^2}{\left\langle \delta_\eta^2 \right\rangle} \frac{k_\eta}{k_U} \gtrsim 25 - 55$$

beta-shift

Fix $\beta' = \beta + \Delta\beta$ by minimizing the variance of

$$\ln(W_{\beta,\beta'}W_{m,m'}) = -\Delta\beta S_g(U) + \mathrm{Tr}\,\ln\left(1 - \frac{\Delta m}{D_m}\right)$$

We find $\Delta\beta/n_f \simeq -3\times 10^{-4}$ (thus we keep c_{SW} constant) and $k_U(\Delta\beta)/k_U(0) \simeq 0.4$