Representation of complex probabilities and complex Gibbs sampling

L.L. Salcedo[1,2]

[1]Departamento de Física Atómica Molecular y Nuclear

[2]Instituto Carlos I de Física Teórica y Computacional
Universidad de Granada

Lattice 2017

Outline

Complex probabilities: sign problem

- Many physical problems reduce to computing weighted averages

$$\langle A \rangle_P = \frac{\int d^n x \, P(x) \, A(x)}{\int d^n x \, P(x)} \qquad x \in \mathbb{R}^n \qquad P(x) = e^{-S(x)}$$

  For a large number of (entangled) degrees of freedom, **Monte Carlo** is often the only suitable approach

$$x \sim P$$ (**Importance sampling** with respect to P)

- However **complex probabilities** may be needed. Outstanding example: lattice QCD at finite density.
- Straightforward importance sampling is meaningless when $P(x)$ is complex: this is the well known **sign problem**

Existing approaches

Several proposals have been put forward to deal with complex weights within Monte Carlo

- Particular treatments for special problems, e.g., worm algorithms
- Reweighting
- Complex Langevin equation
- Lefschetz thimbles and variants
- Reweighting the CLE
- (Your favorite method here)

The goal

- The sign problem is hard. No existing method can be considered as the complete solution. New alternative approaches are worth pursuing.*
- Our goal is to explore a complex version of the **heat bath** method. The idea is to do the updates by means of a **representation** of the conditional **complex** probability.
- This requires studying **representations** of complex probabilities by themselves (i.e., quite independently of the CLE construction).

*At least until the coming of quantum computers/simulators.

Direct representations. I

- The CLE is based on **representing** $P(x) = e^{-S(x)}$ by an ordinary probability distribution $\rho(z)$ defined on the **complexified manifold**, so that ideally

$$\langle A \rangle_P = \langle A \rangle_\rho \equiv \frac{\int d^{2n}z\, \rho(z)\, A(z)}{\int d^{2n}z\, \rho(z)} \qquad \rho(z) \geq 0$$

where $A(z)$ is the **analytical extension** of $A(x)$.

- Abstracting the idea, one define a **representation** $\rho(z)$ of a complex distribution $P(x)$ on $\mathbb{R}^n$, as a distribution on $\mathbb{C}^n$ such that

$$\int d^{2n}z\, \rho(z)\, A(z) = \int d^n x\, P(x)\, A(x) \qquad \text{for generic } A(x)$$

(No residual reweighting.)

Direct representations. II

Of interest for Monte Carlo are the **positive representations**, that is, with $\rho(z) \geq 0$

- Positive representations exists very generally (several constructions are available)
- $P(x)$ needs not admit an analytical extension from $\mathbb{R}^n$ to $\mathbb{C}^n$. Any (normalizable) complex distribution has positive representations, and with the same symmetries.
- The holomorphic function expectations do not fully fix $\rho(z)$. Representations of a given $P(x)$ are by no means unique, e.g.

$$\rho' = e^{t\nabla^2}\rho \quad (t > 0)$$

Direct representations. III

- Nevertheless, the sign problem is not solved in practice since the available constructions are not viable for large systems
- Moreover, poor representations may not be better than reweighting: **Smaller widths yield smaller variances**
- As a rule, **the more complex is** $P(x)$**, the wider is** $\rho(z)$.
  The **width** (size of the support in the imaginary direction) of any **positive** $\rho(z)$ is **bounded from below**. For instance,

$$|\langle A(x)\rangle_P| = |\langle A(z)\rangle_\rho| \leq \max_{z\in\rho}\{|A(z)|\} \quad \text{for arbitrary } A$$

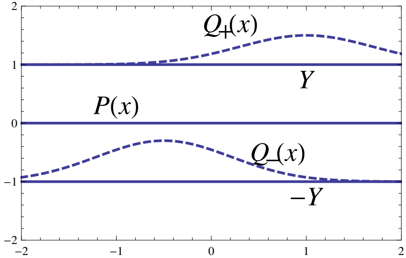
- For local actions, good quality positive representations, with bounded width as n increases, should eventually outperform reweighting: CLE would be an example (when it works).

Two-branch representations: one-dimensional case. I

Let us consider **two-branch representations**: representations with support on two horizontal lines, $\mathbb{R} \pm iY$:

$$\rho(z) = Q_+(x)\,\delta(y - Y) + Q_-(x)\,\delta(y + Y)$$

$$Q_\pm(x) \geq 0$$



$$P(x) = Q_+(x - iY) + Q_-(x + iY)$$

Then $\langle A(x)\rangle_P = \sum_{\sigma=\pm} N_\sigma \langle A(x + i\sigma Y)\rangle_{Q_\sigma}, \quad N_\sigma \equiv \int dx\, Q_\sigma(x)$

Two-branch representations: one-dimensional case. II

- Certainly **real** functions $Q_\pm(x)$ exist (and they are unique):

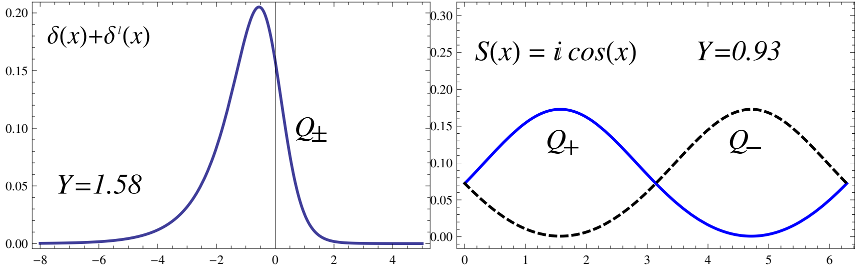
$$\tilde{Q}_\pm(k) = \pm\frac{e^{\pm Y}\tilde{P}(k) - e^{\mp Y}\tilde{P}(-k)^*}{2\sinh(2kY)} \quad (k \neq 0)$$

with $P(x) = \sum_k e^{ikx}\tilde{P}(k)$

- For Y above some critical value, $Q_\sigma(x)$ become **positive** (more e^Y in denominator and eventually the positive constant mode dominates)

Two-branch representations: one-dimensional case. III

Two examples of two-branch representations:



Two-branch representations: higher-dimensional case. I

- For $\mathbb{R}^n$ $(n > 1)$ one can try a two-branch scheme

$$P(\vec{x}) = \sum_{\sigma=\pm} Q_\sigma(\vec{x} - i\sigma\vec{Y}), \qquad Q_\sigma(\vec{x}) \geq 0$$

$$Q_\sigma(\vec{x}) = N_\sigma + 2\mathrm{Re}\sum_{\vec{k}\neq 0} \frac{e^{i\vec{k}\cdot(\vec{x}-i\sigma\vec{Y})}\tilde{P}(k)}{2\sinh(2\sigma\vec{k}\cdot\vec{Y})}$$

$$\equiv N_\sigma + \mathrm{Re}\left(C(\vec{x};\sigma\vec{Y}) * P(\vec{x})\right)$$

- This is a **formal** solution but not a **practical** one.
- No good choices for $\vec{Y}$: the natural choice $\vec{Y} = (Y, \ldots, Y)$ meets $\vec{k}\cdot\vec{Y} = 0$ for some Fourier modes; these components are not moved into the complex manifold.

Two-branch representations: higher-dimensional case. II

- Asymmetric $\vec{Y}$ are unnatural and also tend to require some Y_i larger than necessary.
- An elegant solution is to use 2^n branches: For each Fourier mode $\vec{k}$ and variable x_i, either $+Y$ of $-Y$ is selected so that $k_i Y_i \geq 0$:

$$P(x) = \sum_{\vec{\sigma}} Q_{\vec{\sigma}}(\vec{x} - i\vec{\sigma}Y) \qquad \vec{\sigma} = (\pm, \ldots, \pm)$$

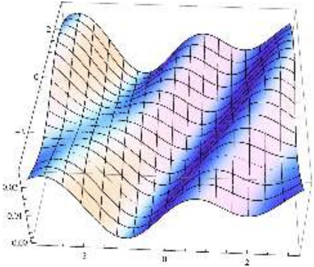
- Having more branches is compensated by i) simplicity, ii) restoration of symmetry, iii) smaller values of Y, and iv) mandatory in the non periodic case.

Two-branch representations: higher-dimensional case. III

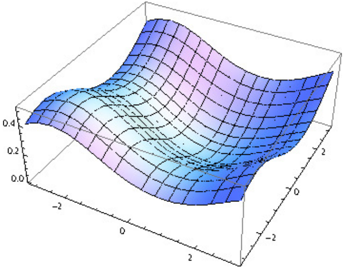
Example of two-branch representations in two dimensions:

$$P(x_1, x_2) = (1 + \beta\cos(x_1))(1 + \beta\cos(x_2))(1 + \beta\cos(x_1 - x_2))$$

$\beta = 2i$



$Q_{\pm}(\vec{x}), \quad \vec{Y} = (1.48, 4.43)$



$Q_{++}(\vec{x}), \quad \vec{Y} = (2.4, 2.4)$

Representation of complex probabilities on groups. I

- $P(g)$ complex probability defined on $G = \{e^{-i\vec{a}\cdot\vec{T}},\ \vec{a} \in \mathbb{R}^m\}$
- The observables can be analytically extended to the **complexified group**

$$\tilde{G} = \{e^{-i\vec{a}\cdot\vec{T}},\ \vec{a} \in \mathbb{C}^m\}$$

Within a **two-branch scheme**, letting $G_I \equiv \{e^{-i\vec{a}\cdot\vec{T}},\ \vec{a} \in i\mathbb{R}^m\}$

$$P(g) = Q_+(gh) + Q_-(gh^{-1}), \quad g \in G, \quad h \in G_I$$

$$\langle A \rangle_P = N_+ \langle A(gh^{-1}) \rangle_{Q_+} + N_- \langle A(gh) \rangle_{Q_-}$$

- $A(gh^{-1}), A(gh)$ refer to the analytical extension in $\tilde{G}$
- The Monte Carlo sampling is carried out with the real and positive weights $Q_\pm(g)$ on G

Representation of complex probabilities on groups. II

- To do the matching, P and $Q_\pm$ are expanded as a linear combination of **group representations** (like the Fourier modes of $\mathrm{U}(1)$)
- For instance in $\mathrm{U}(N)$, for $h = \mathrm{diag}(\omega_1, \dots, \omega_N) \quad (\omega_i > 0)$

$$P(g) = \sum_n a^{j_1 \cdots j_n}_{i_1 \cdots i_n} g^{i_1}{}_{j_1} \cdots g^{i_n}{}_{j_n}$$

$$Q_\pm(g) = 2\,\mathrm{Re} \sum_n a^{j_1 \cdots j_n}_{i_1 \cdots i_n} g^{i_1}{}_{j_1} \cdots g^{i_n}{}_{j_n}\, c(\Omega^{\pm 1}_{j_1,\dots,j_n})$$

with $\quad \Omega_{j_1,\dots,j_n} \equiv \omega_{j_1} \cdots \omega_{j_n} \quad$ and

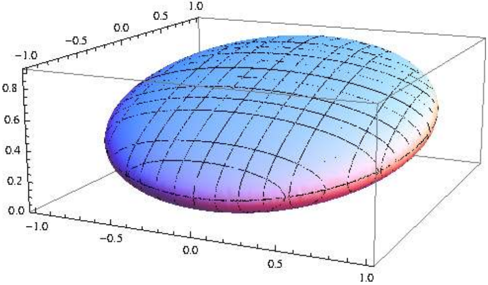
$$c(\Omega) \equiv \frac{\Omega}{\Omega^2 - \Omega^{-2}} \qquad \left(\quad = \frac{e^Y}{2\sinh(2Y)} \quad \right)$$

- For sufficiently large h the singlet dominates and $Q_\sigma > 0$

Representation of complex probabilities on groups. III

A two-branch representation on $\mathrm{SU}(2)$: $P(g) = 1 + \beta\,\mathrm{tr}(\sigma_3\, g)$

$$Q_\pm(g) = \frac{1}{2} + \mathrm{tr}\left[\left(\pm\frac{\beta_R}{\Omega - \Omega^{-1}} + i\sigma_3\frac{\beta_I}{\Omega + \Omega^{-1}}\right) g\right]$$



$h = \mathrm{diag}(\Omega, \Omega^{-1})$

$g = a_0 - i\vec{a}\cdot\vec{\sigma}, \quad a \in \mathrm{S}^3$

Plot of $Q_+(g)$ on the plane (a_1, a_3) with $a_2 = 0$ for

$\beta = \frac{1}{2}(1+i), \quad \Omega = e^{2.5}$

Complex heat bath. I

- Each variable is updated in turn using a **representation** of its **conditional** complex probability:

$$(z_1,\dots,z_i,\dots,z_n)\to(z_1,\dots,z_i',\dots,z_n),\quad z_i'\sim\rho_{\rm rep}(z_i'|\{z_{j\neq i}\})$$

where $\rho_{\rm rep}(z_i'|\{z_{j\neq i}\})$ is a representation (with respect to z_i') of

$$P(z_i'|\{z_{j\neq i}\})=\frac{P(z_1,\dots,z_i',\dots,z_n)}{P(z_1,\dots,\widehat{z}_i,\dots,z_n)}$$

- Only one-variable representations are needed
- Relies on the analytical extension of $P(\vec{x})$, just like CLE
- Convergence to a representation of $P(x)$ is not guaranteed
- Potential problem from zeroes of the **marginal** probabilities $P(z_1,\dots,\widehat{z}_i,\dots,z_n)$

Complex heat bath. II

To see the performance of the method we study a concrete model:

- Complex action in a d-dimensional periodic lattice

$$S[\phi] = \sum_x \left(\lambda \phi_x^4 + \phi_x^2 + \beta\, \phi_x \sum_{\mu=1}^{d} \phi_{x+\hat{\mu}} \right), \quad \beta \in \mathbb{C} \quad \lambda = 0, 1$$

$$P[\phi_x | \{\phi_{x' \neq x}\}] \propto e^{-\lambda \phi_x^4 - \phi_x^2 - \beta\, \phi_x \sum_{\mu=1}^{d} (\phi_{x+\hat{\mu}} + \phi_{x-\hat{\mu}})}$$

- In the **free** case ($\lambda = 0$), the conditional probability is a displaced real Gaussian. The marginal probability has no zeros. The method works alright for a bounded region of the β plane.

Complex heat bath: Interacting case. I

In the **interacting** case ($\lambda = 1$), we apply the complex Gibbs sampling for imaginary β and compute four observables $O_{1,2}$ and $I_{1,2}$.†

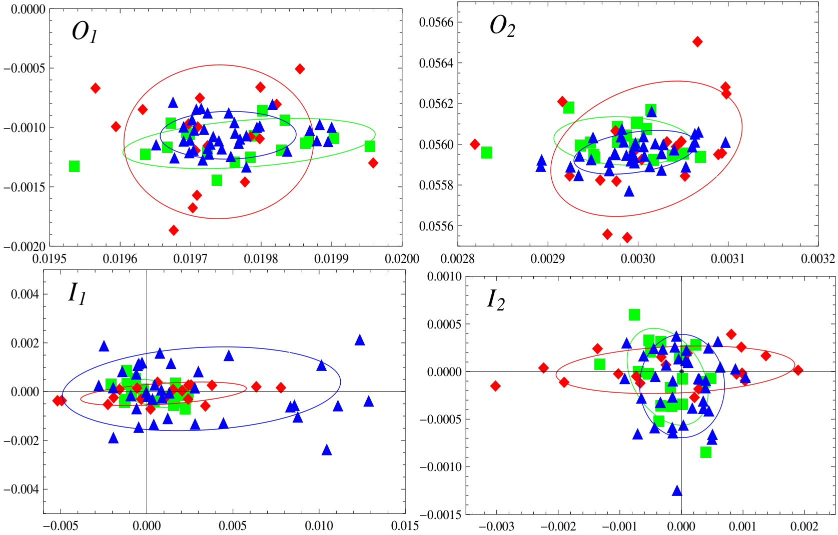
- Checks from T-matrix calculations (for $d = 1$ only). The complex Gibbs method works up to $|\beta| \approx 1$.
- Checks from reweighting and complex Langevin
- Checks from virial (or Schwinger-Dyson) relations

$$\left\langle \frac{\partial A}{\partial \phi_x} \right\rangle = \left\langle A \frac{\partial S}{\partial \phi_x} \right\rangle$$

The choices $A = \phi_x, \phi_{x+\hat{\mu}}$, yield $O_{1,2}$, $I_{1,2}$ with

$$\langle I_1 \rangle = \langle I_2 \rangle = 0$$

†Thanks to CSIRC (Granada University) for providing computing time in the cluster Alhambra.



3^3 lattice, $\beta = 0.25i$, RW squares, CLE rhombuses, CHB triangles.

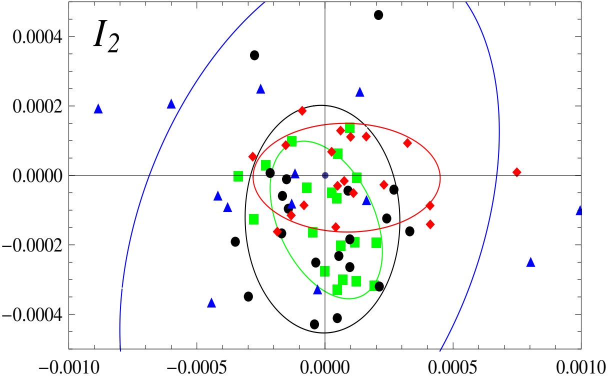
Complex heat bath: Interacting case. II

β	N^d	$10^3 \times \mathrm{Re}\langle O_1 \rangle$	$10^3 \times \mathrm{Im}\langle O_2 \rangle$	$10^3 \times \mathrm{Re}\langle I_1 \rangle$	$10^3 \times \mathrm{Im}\langle I_2 \rangle$	Method‡
0.25i	3^3	19.783 (28)	56.017 (19)	0.49 (33)	− 0.05 (8)	RW
0.25i	3^3	19.740 (21)	55.978 (51)	0.97 (73)	0.02 (4)	CLE
0.25i	3^3	19.745 (14)	55.967 (16)	3.39 (115)	− 0.14 (9)	CHB
0.25i	8^3	19.985 (137)	55.958 (44)	0.59 (63)	− 0.67 (42)	RW
0.25i	8^3	19.746 (5)	55.977 (12)	0.29 (24)	0.01 (1)	CLE
0.25i	8^3	19.749 (4)	55.969 (5)	4.04 (28)	− 0.13 (2)	CHB
0.5i	3^3	71.642 (116)	99.851 (47)	− 0.39 (40)	− 0.15 (21)	RW
0.5i	3^3	71.628 (76)	100.073 (98)	0.16 (75)	0.03 (9)	CLE
0.5i	3^3	71.578 (41)	99.882 (32)	4.70 (125)	0.21 (29)	CHB
0.5i	8^3	71.332 (17)	99.579 (18)	0.24 (18)	− 0.01 (3)	CLE
0.5i	8^3	71.305 (8)	99.545 (9)	2.44 (85)	− 0.01 (5)	CHB
0.5i	8^3	71.305 (8)	99.544 (9)	0.14 (20)	0.03 (3)	CHB**
0.5i	8^3	71.352 (11)	99.570 (8)	6.95 (212)	− 0.16 (13)	CHB*
0.5i	16^3	71.303 (7)	99.569 (7)	0.11 (7)	0.03 (1)	CLE
0.5i	16^3	71.328 (5)	99.567 (4)	5.15 (9)	− 0.19 (3)	CHB
0.5i	8^4	90.652 (9)	94.365 (9)	0.06 (9)	0.03 (1)	CLE
0.5i	8^4	90.670 (5)	94.369 (5)	3.19 (20)	0.14 (3)	CHB

‡ * No cutoff, ** Filtered

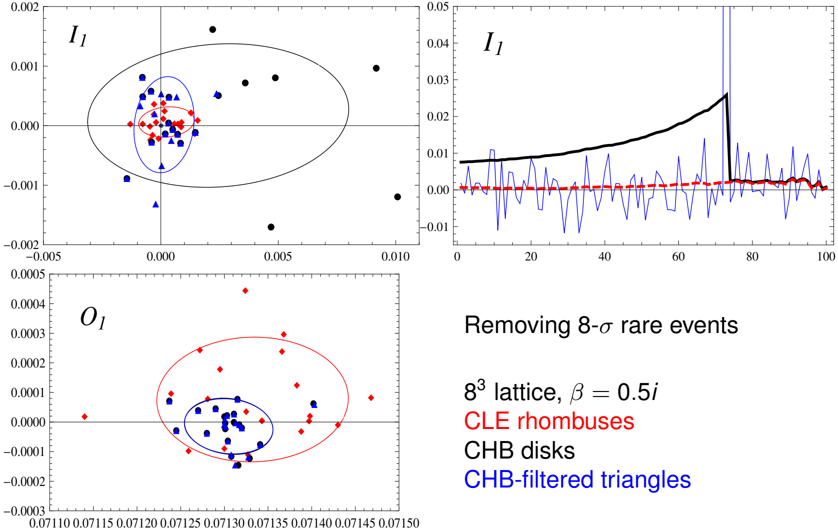
Effect of the cutoff

Updates with Y above some Y_s (cutoff) are skipped



8^3, $\beta = 0.5i$
CLE rhombuses
CHB-2 squares
CHB-5 disks
CHB-∞ triangles

Filtering out Markovian jolts



Removing 8-σ rare events

8^3 lattice, $\beta = 0.5i$
CLE rhombuses
CHB disks
CHB-filtered triangles

Bias from zeroes of the marginal probability

- Gibbs update: $(z_1, z_2) \to (z_1', z_2)$, $\qquad \langle A \rangle_\rho \to \langle A \rangle_{\rho'}$

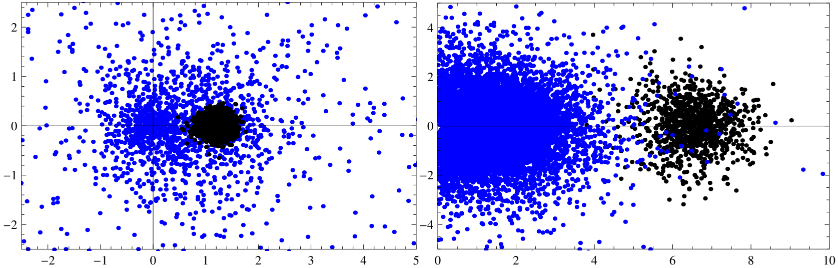
$$\begin{aligned}\langle A \rangle_{\rho'} &= \int d^2 z_2\, d^2 z_1'\, \rho(z_2)\, \rho_{\rm rep}(z_1'|z_2)\, A(z_1', z_2) \\ &= \int d^2 z_2\, dx_1'\, \rho(z_2) \frac{P(x_1', z_2)}{P(z_2)} A(x_1', z_2) \\ &= \int dx_2\, dx_1'\, P(x_2) \frac{P(x_1', x_2)}{P(x_2)} A(x_1', x_2) + \textbf{Residues}\end{aligned}$$

- The idea is verified for two-branch representations of

$$P(x_1, x_2) = (1 + \beta\cos(x_1))(1 + \beta\cos(x_2))(1 + \beta\cos(x_1 - x_2))$$

A **bias** is found whenever $[-Y, Y]$ **contains** the zero of $P(z_2) \sim 1 + \frac{1}{2}\beta^2\cos(z_2)$. This cannot be prevented for sufficiently large β (Y increases and the zero in z_2 decreases).

That the **marginal probabilities** stay away from zero can be **monitored** on-the-flight during the Markovian chain:



$\lambda\phi^4$ model, $\beta = 0.5i, i$

XY model, 8^3 lattice, $(\beta, \mu) = (1, 1), (0.4, 0.2)$

Summary

- Positive representations exists quite generally with variable degrees of quality, related to their **extension** on the complexified manifold. Good quality positive representations are expected to exist for local actions.
- Unfortunately, there is no known practical way to obtain them, beyond low dimensions (barring CLE, when it works).
- The complex heat bath method works in some cases. Its validity can be assessed by monitoring the marginal probabilities on-the-flight.

Outlook/speculations

- May be the bias introduced by zeros of the marginal probability can be removed by extracting the poles in the construction of the representations of the conditional probabilities
- Also the bias problem could perhaps be alleviated by considering larger **clusters** of variables to be updated at each step.
- Since positive representations exists, this means that there are **real** actions on the complexified manifold that correctly reproduce, say QCD at finite density. Is there a **local** real action in the same universality class?

Some references

- Does the complex Langevin method give unbiased results?
  Phys. Rev. D94 (2016) 114505
- Gibbs sampling of complex valued distributions
  Phys. Rev. D94 (2016) 074503
- Existence of positive representations for complex weights
  J. Phys. A40 (2007) 9399
- Representation of complex probabilities
  J. Math. Phys. 38 (1997) 1710-1722
- Spurious solutions of the complex Langevin equation
  Phys. Lett. B305 (1993) 125-130

Backup

Reweighting

For $P_0(x) > 0$

$$\langle A \rangle_P = \frac{\langle wA \rangle_{P_0}}{\langle w \rangle_{P_0}}, \qquad w(x) = \frac{P(x)}{P_0(x)}$$

Often $P_0(x) = |P(x)|$ (phase quenched probability)

- Simple and sound, mathematically rigorous
- Afflicted by the **overlap problem**: $x \sim P_0$ not P
- As the volume of the system increases, the signal-to-noise ratio quenches exponentially: The Monte Carlo estimate becomes useless

Complex Langevin. I

- It relies on a driven random walk on $\mathbb{C}^n$

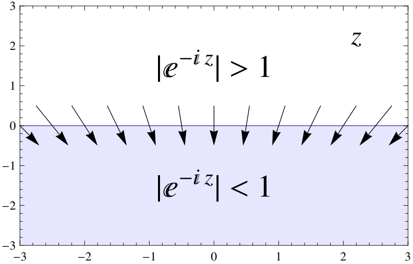
$$dz = -\nabla S(z)\,dt + \sqrt{2dt}\,\eta\,,$$

taking the average of $A(z)$, the analytical extension of $A(x)$.

- The CLE method is elegant and worthy but **not reliable**

$$\text{One-dimensional example:}\quad S(x) = \frac{1}{8}x^4 + 2ix$$

By construction, the CLE walker ends up in $\mathrm{Im}\, z < 0$



hence $\quad |\langle e^{-iz}\rangle_{\mathrm{CLE}}| \le 1$

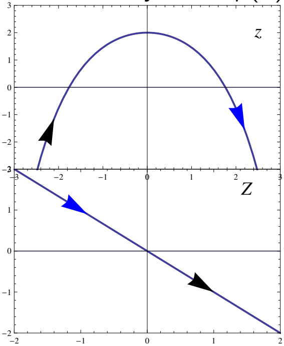
yet $\quad \langle e^{-ix}\rangle_P = -4.98$

Complex Langevin. II

The Fokker-Planck eq. for $\rho(z)$ implies a F-P eq. for $P(x,t)$:

$$-\partial_t P = -\nabla^2 P + \nabla(vP) + \mathcal{A}, \qquad P = e^{-S} \Rightarrow \mathcal{A} = 0$$

The **anomaly** $\mathcal{A}$ is a boundary term from integration by parts. It vanishes only when $\rho(z)$ is sufficiently convergent at ∞



$$Z = \int_\infty^z \frac{dz'}{v(z')} \qquad \text{“time to reach } \infty\text{”}$$

In the new variable ∞ is a regular point.
A finite density at $Z = 0$ is tied to $\mathcal{A} \neq 0$

Direct representations. III

A concrete construction:

$$\text{Let} \quad \int d^n x\, P(x) = \int d^n x\, P_0(x), \qquad P_0(x) > 0$$

$$\begin{aligned} P(x) &= P_0(x) + \nabla(P_0(x)H(x)) \\ &= \int d^n x'\, P_0(x')\left(\delta(x'-x) - H(x')\nabla\delta(x'-x)\right) \end{aligned}$$

Then, if $q(z)$ is a positive representation of $\delta(x) + \delta'(x)$

(for instance: $q(z) = \frac{1}{8\pi}\left|1 - \frac{z}{2}\right|^2 e^{-|z|^2/4}$)

$$\rho(z) = \int d^n x'\, P_0(x') \int d^2 z'\, q(z')\, \delta(z - x' - z' H(x'))$$

is a positive representation of $P(x)$. That is,

$$x' \sim P_0, \quad z' \sim q \quad \Rightarrow \quad z = x' + z' H(x') \sim \rho \cong P$$

Complex representations

Abstracting the procedure of construction of representations: we look for a **complex function** $\hat{\rho}$ such that (having extracted the constant mode of P)

$$P = K\hat{\rho}, \qquad 0 = K\hat{\rho}^*, \qquad \rho = 2\mathrm{Re}\hat{\rho}$$

where K is the analytical projection $\rho \to P$. Equivalently

$$P = K\hat{\rho}, \qquad 0 = \bar{K}\hat{\rho}$$

($\bar{K}$ being the anti-analytical projection). This is a **linear system**. So for instance, if $P^{(i)} \cong \hat{\rho}^{(i)}$,

$$P = \sum_i a_i P^{(i)} \;\Rightarrow\; \rho = 2\mathrm{Re}\sum_i a_i \hat{\rho}^{(i)}, \qquad a_i \in \mathbb{C}$$

Unlike ρ, the complex representation $\hat{\rho}$ contains information on the **phase** of P.

Representation of complex probabilities on groups. III

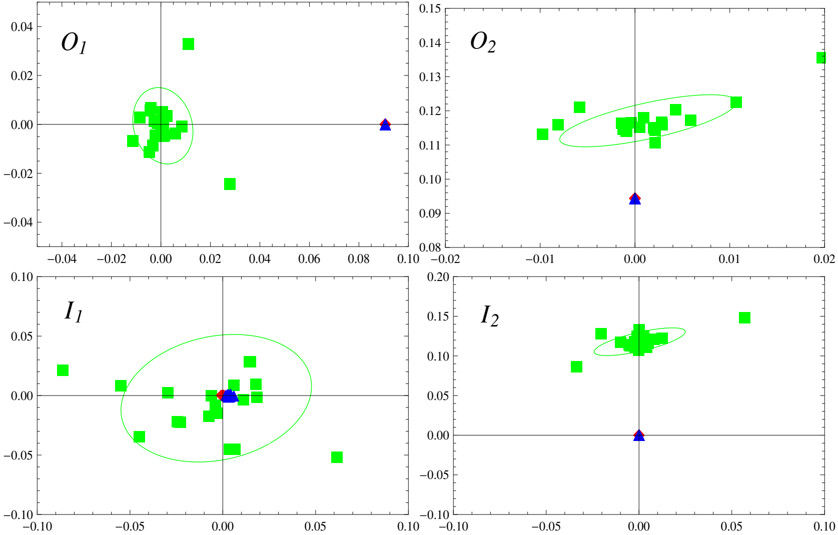
- **Complication** in the non-Abelian case: regardless of the choice of h, certain components (**singlet with respect to** h) will remain unchanged ($\Omega = 1$). The relation

$$P(g) = Q_+(gh) + Q_-(gh^{-1})$$

can only be fulfilled if P is real along those components.

- There, a different element $h' \in G_l$ has to be applied, fulfilling to the condition

$$\langle \text{singlet } h \,|\, \text{singlet } h' \rangle = 0$$



8^4 lattice, $\beta = 0.5i$, RW squares, CLE rhombuses, CHB triangles.