Use of Schwinger-Dyson equation in constructing an approximate trivializing map

Nobuyuki Matsumoto (RBRC)

LATTICE2022 8/8/2022 @Bonn
work in collaboration with
P. Boyle, T. Izubuchi, L. Jin, C. Jung, C. Lehner, and A. Tomiya BNL/Edinburgh BNL/RBRC UConn/RBRC BNL/RBRC BNL/Regensburg/RBRC IPUT Osaka
(work in progress)

## Introduction (1/3)

QCD lattice calculation has been taking an important and crucial role in the precision test of the standard model.
e.g., HVP contribution in muon g-2

RBC/UKQCD update of
17:10 disconnected S.Bacchio
the HVP contribution to the muon g-2"
17:30 EM coupling M.T.San Jose Perez
17:50 EM coupling H.Wittig, S.J.Perez
borrowed from Christoph's slide;
overview plot of tentative lattice results
and the experimental R -ratio result

The required precision is getting in a magnificent order, and calculation on fine lattices has become more demanding.

However, as we reach the continuum limit, we face the critical slowing down, which is often characterized by very long autocorrelation of topological charge.

## Goal of this study

Develop the idea of trivializing map Nicolai 1980, Lüscher 0907.5491 map between finite $\beta$ theory and $\beta=0$ theory
and utilize it to generate fine lattice configurations with reduced numerical cost.

## Introduction (2/3)

There are many promising approaches for solving the critical slowing down/topological freezing:

```
Trivializing/normalizing map
    Mon HS7 14:00 Flow-based density of states for complex actions Julian Urban
            14:20 Stochastic normalizing flows for lattice field theory Elia Cellini
            15:00 Gauge-equivariant flow models for sampling in lattice field theories with pseudofermions Fernando Romero-Lopez
            16:30 Status update on flow models for gauge field generation Phiala Shanahan
            17:50 Machine Learning Trivializing Maps Joe Marsh Rossney
            18:10 Learning trivializing flows David Albandea
            Wed HS717:50 Generative models for scalar field theories: how to deal with poor scaling? Javad Komijani
Master field analysis
    Mon HS5 16:40 Translating topological benefits in very cold master-field simulations Anthony Francis
Metadynamics
    Mon HS7 17:30 Topology changing update algorithms for SU(3) gauge theory Timo Eichhorn
    Poster B Metadynamics Surfing on Topology Barriers in the Schwinger Model Philip Rouenhoff
Fourier acceleration
    Poster B Fourier acceleration in strongly-interacting linear sigma models Joshua Swaim
Parallel tempering
    Tue HS5 14:20 Towards glueball masses of large-N SU(N) Yang-Mills theories without topological freezing via parallel tempering on boundary conditions Claudio Bonanno
            HS1 15:20 Parallel tempering algorithm applied for the deconfinement transition of quenched QCD Ruben Kara
Multicanonical approach
    Tue HS5 15:20 Topological susceptibility in high temperature full QCD via staggered spectral projectors Francesco D'Angelo
Skewed detailed balance
    Thu HS7 10:40 Towards the Application of Skewed Detailed Balance in Lattice Gauge Theories Joao C. Pinto Barros
Fermion algorithms
    Sat HS2 11:00 Review on Algorithms for dynamical fermions Jacob Finkenrath
```

Please let me know if I missed your work.
Also around BNL there are intense studies on

- Gauged-fixed Fourier acceleration Sheta, Zhao, Christ 2108.05486
- Riemannian manifold HMC Nguyen, Boyle, Christ, Jang, Jung 2112.04556
- L2HMC Foreman, X.Y.Jin, Osborn 2201.01582
- Machine learned trivializing map X.Y.Jin 2201.01862


## Introduction (3/3)

- Alternatively to Luscher's trivializing flow (flow kernel obtained as a $t$-expansion), Lüscher 0907.5491 flow time
we design the approximate trivializing flow with the Schwinger-Dyson (SD) equation.
Gonzalez-Arroyo, Okawa 1987, de Forcrand, Perez, Hashimoto, Hioki, Matsufuru, Miyamura,
Nakamura, Stamatescu, Tagoi, Takaishij, Umeda hep-lat/9806008
- By extending L. Jin's code of field-transformed HMC to include generic flow kernels, Jin LATTICE 2021
we perform the HMC with the resulting exact effective action after this flow.


## Advantages of this method

- the basis for the flow kernel can be chosen by hand
- can be applied to the general action of interest
- the coefficients in the kernel are determined by lattice estimates of the observables; no need for analytic calculation such as $t$-expansion
- truncation effects and goodness of the flow can be measured by the force norm
- We apply our method to Wilson and DBW2 actions and show that:
- With the SD method, we can have a better control of the effective action than $t$-expansion
- In particular cases, faster decorrelation (in MC step unit) is observed for long-ranged observables by adding rectangle and chair to the flow
- However, we have large algorithmic overhead, and need to check the scaling with larger statistics to confirm the actual benefits at large $\beta$

Trivializing map

## Trivializing map $(1 / 2) \quad$ Lüscher 0907.5491

- For a field transformation $\mathcal{F}_{t}: V \mapsto U=\mathcal{F}_{t}(V)$, we have the effective action $S_{\text {eff }, t}(V)$ :

$$
\begin{aligned}
\int d U e^{-S(U)} \mathcal{O}(U) & =\int d V \operatorname{det} \mathcal{F}_{t *}(V) e^{-S\left(\mathcal{F}_{t}(V)\right)} \mathcal{O}\left(\mathcal{F}_{t}(V)\right) \\
& \equiv \int d V e^{-S_{\text {eff }, t}(V)} \mathcal{O}\left(\mathcal{F}_{t}(V)\right)
\end{aligned}
$$

i.e.,

$$
S_{\mathrm{eff}, t}(V) \equiv S\left(\mathcal{F}_{t}(V)\right)-\ln \operatorname{det} \mathcal{F}_{t *}(V)
$$

$$
\mathcal{F}_{t *}(V)=\left(\mathcal{F}_{t *}^{A B}(V)\right): \text { Jacobian matrix }
$$

$$
\begin{aligned}
d \theta_{(U)}^{A}=\mathcal{F}_{t *}^{A B}(V) d \theta_{(V)}^{B} \quad \text { where } \quad & U_{x, \mu}^{\prime}=e^{\theta_{x, \mu}^{a} T^{a}} U_{x, \mu} \\
& \text { or } \partial_{x, \mu}^{a} U_{x, \mu}=T^{a} U_{x, \mu}
\end{aligned}
$$

- Lüscher proposed to use the trivializing map $\mathcal{F}_{t}$ in HMC
that makes the finite $\beta$ theory mapped to
Duane, Kennedy, Pendleton, Roweth 1987 the strong coupling limit $\beta=0$ at $t=1$ :

$$
S_{\mathrm{eff}, t=1}(V)=\text { const. }
$$

$V$ will be decorrelated faster under the trivial action, and so does the configurations $U=\mathcal{F}_{t}(V)$

$$
\begin{aligned}
& T^{a}: \operatorname{su}(3) \text { generators, } \operatorname{tr}\left(T^{a} T^{b}\right)=-\frac{1}{2} \delta^{a b} \\
& \text { Haar measure: }(d U) \propto \prod_{A} d \theta^{A} \\
& A=(x, \mu, a) \text { labels DOF }
\end{aligned}
$$

$$
\left[\partial_{\theta_{x, \mu}^{a}} \equiv \partial_{x, \mu}^{a}\right]
$$



- Gradient flow ansatz:

$$
\dot{\mathcal{F}}_{t, \epsilon}(U)_{x, \mu}=-T^{a} \partial_{x, \mu}^{a} \tilde{S}_{t}(U) \cdot U_{x, \mu}
$$

- Require that $\mathcal{F}_{t}$ trivializes the theory at $t=1$ :

$$
\begin{aligned}
S_{\mathrm{eff}, t}(V) & \stackrel{\Delta}{=} S\left(\mathcal{F}_{t}(V)\right)-\ln \operatorname{det} \mathcal{F}_{t *}(V) \\
& * \text { requirement } \\
& =(1-t) S\left(\mathcal{F}_{t}(V)\right) \\
& \longmapsto d / d t
\end{aligned}
$$

equation for the kernel:

$$
\partial_{x, \mu}^{a} \sim T^{a} \text { insertion }
$$

$$
\begin{aligned}
& \left(\partial^{a}\right)^{2} \overparen{C}=-\frac{4}{3} \overparen{\square} \\
& \operatorname{tr}\left[\left(T^{a}\right)^{2} \ldots\right]=-\frac{4}{3} \operatorname{tr}[\ldots]
\end{aligned}
$$

$$
\left\{\begin{array}{c}
\frac{2}{2 a}=-\frac{1}{2}(\sqrt{2}+\cdots \\
\operatorname{tr}\left[T^{a} A\right] \operatorname{tr}\left[T^{a} B\right]=-\frac{1}{2}\left(\operatorname{tr}[A B]-\frac{1}{3} \operatorname{tr} A \cdot \operatorname{tr} B\right)
\end{array}\right.
$$

$$
-\left(\partial^{A}\right)^{2} \tilde{S}_{t}+t \partial^{A} S \partial^{A} \tilde{S}_{t}=S \quad \text { (up to irrelevant const) }
$$

from Jacobian from action

- Solution (for Wilson action $S=S_{W}$ ):

$$
\begin{aligned}
\tilde{S}_{t} & =-\frac{\beta}{32} W_{0} \leftarrow \text { LO: plaquette } \\
& +t \frac{\beta^{2}}{192}\left(-\frac{4}{33} W_{1}+\frac{12}{119} W_{2}+\frac{1}{33} W_{3}-\frac{5}{119} W_{4}+\frac{3}{10} W_{5}-\frac{1}{5} W_{6}+\frac{1}{9} W_{7}\right) \\
& +O\left(t^{2}\right) \quad \text { NLO: "fectangle, chair, twisted rectangle } \ldots \\
& \text { "footprint } 2 \text { shapes" }
\end{aligned}
$$

- Note that the $t$-expansion is performed around $t=0$; this corresponds to expanding around the trivialized theory.

In fact, the expansion parameter $t$ appears in the combination $\beta t$ $\therefore$ approximation is better for small $\beta$.

However, our primary target is the large $\beta$ theory; $\therefore$ we rather want to approximate the map based on the information at large $\beta$.

- In the Schwinger-Dyson (SD) method below, we sequentially determine the flow from the finite $\beta$ theory.
- At each intermediate step, we determine the effective couplings by the SD equation.
- We then design the flow to decrease the couplings from the lattice estimates of Wilson loops.

This determination of the flow can be seen as iteratively reexpanding the flow kernel at each step.

direction of
construction
in the SD method


## Schwinger-Dyson equation

- Suppose that

$$
S_{\mathrm{eff}}(V)=\sum_{j} \beta_{j} W_{j} \quad \begin{aligned}
& W_{j}: \text { Wilson loops } \\
& \text { we omit the subscript } t \text { momentarily }
\end{aligned}
$$

- The couplings $\beta_{j}$ obey the linear equation:

$$
\sum_{j} \beta_{j}\left\langle\partial^{A} W_{j} \partial^{A} W_{i}\right\rangle_{S_{\text {eff }}}=\left\langle\left(\partial^{A}\right)^{2} W_{i}\right\rangle_{S_{\text {eff }}} \quad\langle\cdot\rangle_{S_{\text {eff }}}: \text { expectation value with respect to } S_{\text {eff }}
$$

- Consider a variation using $W_{i}$ as the flow kernel:

$$
\delta V=-\epsilon T^{A} \partial^{A} W_{i} \cdot V
$$

- The path integral is invariant under this variation (Schwinger-Dyson equation):

$$
0=\delta \int(d V) e^{-S_{\text {eff }}(V)}=\int(d V) e^{-S_{\text {eff }}(V)} \epsilon \frac{\left[-\left(\partial^{A}\right)^{2} W_{i}\right.}{\frac{f^{\prime}}{\text { from Jacobian }}}+\frac{\left.\partial^{A} S_{\text {eff }} \partial^{A} W_{i}\right]}{\text { from action }}
$$

- Combining this formula with the expansion of $S_{\text {eff }}(V)$ :

$$
\sum_{j} \beta_{j}\left\langle\partial^{A} W_{j} \partial^{A} W_{i}\right\rangle_{S_{\mathrm{eff}}}=\left\langle\left(\partial^{A}\right)^{2} W_{i}\right\rangle_{S_{\mathrm{eff}}}
$$

- Generically, we need infinite number of couplings to exactly parametrize $S_{\text {eff }}(V)=\sum_{j} \beta_{j} W_{j}$.
$\therefore \quad \sum_{j} \beta_{j}\left\langle\partial^{A} W_{j} \partial^{A} W_{i}\right\rangle_{S_{\text {eff }}}=\left\langle\left(\partial^{A}\right)^{2} W_{i}\right\rangle_{S_{\text {eff }}}$ : infinite-dimensional matrix
- Instead, we can try to mimic $S_{\text {eff }}(V)$ with a finite basis:

$$
S_{\text {eff }}^{\prime}(V)=\sum_{j}^{\prime} \beta_{j}^{\prime} W_{j} . \quad \longleftarrow \quad \begin{aligned}
& \text { prime symbols indicate truncation } \\
& (j \text { runs a finite range })
\end{aligned}
$$

We determine $\beta_{j}^{\prime}$ by:

$$
\Sigma_{j}^{\prime} \beta_{j}^{\prime}\left\langle\partial^{A} W_{j} \partial^{A} W_{i}\right\rangle_{S_{\text {eff }}}=\left\langle\left(\partial^{A}\right)^{2} W_{i}\right\rangle_{S_{\text {eff }}} \quad: \begin{aligned}
& \text { finite-dimensional matrix } \\
& (i \text { is also restricted to the finite range })
\end{aligned}
$$

Such $\beta_{j}^{\prime}$ give the best approximation of $S_{\text {eff }}(V)$ in the sence that it minimizes the norm

$$
\left\|S_{\text {eff }}-S_{\text {eff }}^{\prime}\right\|_{S_{\text {eff }}}, \quad \text { where } \frac{\|S\|_{S_{\text {eff }}}^{2} \equiv\left\langle\left(\partial^{A} S\right)^{2}\right\rangle_{S_{\text {eff }}}}{L 2 \text { norm of the force }}
$$

$\therefore$ The truncation error is calculable.

This Schwinger-Dyson method
gives us a way to parametrize effective actions.
$\because$ Subtracted equation

$$
\Sigma_{j}^{\prime}\left(\beta_{j}-\beta_{j}^{\prime}\right)\left\langle\partial^{A} W_{j} \partial^{A} W_{i}\right\rangle_{S_{\text {eff }}}=0
$$

is the stationary condition:

$$
\begin{aligned}
& \frac{\partial}{\partial \beta_{i}^{I}}\left\|S_{\text {eff }}-S_{\text {eff }}^{\prime}\right\| S_{\text {eff }}^{2} \\
& =\frac{\partial}{\partial \beta_{i}^{\prime}}\left\langle\left[\partial^{A}\left(S_{\text {eff }}-S_{\text {eff }}^{\prime}\right)\right]^{2}\right\rangle_{S_{\text {eff }}} \\
& =-2 \sum_{j}^{\prime}\left(\beta_{j}-\beta_{j}^{\prime}\right)\left\langle\partial^{A} W_{j} \partial^{A} W_{i}\right\rangle_{S_{\text {eff }}} \\
& \equiv 0
\end{aligned}
$$

- We parametrize $\tilde{S}_{t}$ in the same truncated space:

$$
\tilde{S}_{t}(V)=\sum_{k}^{\prime} \gamma_{k, t} W_{k}
$$

- Differentiating the equation for $\beta_{j, t}^{\prime}$ :

$$
\begin{aligned}
& \left\langle-\left(\partial^{A}\right)^{2} W_{i}+\sum_{j}^{\prime} \beta_{j, t}^{\prime} \partial^{A} W_{j} \partial^{A} W_{i}\right\rangle_{S_{\text {eff }, t}}=0 \\
& \quad-d / d t \\
& \sum_{k}^{\prime} \gamma_{k, t}\left\langle\partial^{B} W_{k} \partial^{B}\left[-\left(\partial^{A}\right) W_{i}+\partial^{A} S_{\text {eff }, t}^{\prime} \partial^{A} W_{i}\right]\right\rangle_{S_{\text {eff }, t}}=-\sum_{j}^{\prime} \dot{\beta}_{j, t}^{\prime}\left\langle\partial^{A} W_{j} \partial^{A} W_{i}\right\rangle_{S_{\text {eff }, t}} \\
& \quad(d / d t) \text { acting on the Boltzmann weight acting on } \beta_{j, t}^{\prime}
\end{aligned}
$$

This equation gives the coefficients $\gamma_{k, t}$ for a given $\dot{\beta}_{j, t}^{\prime}$ (thus a trajectory of $S_{\text {eff }, t}^{\prime}$ )

- We here require $\dot{\beta}_{j, t}^{\prime}=-\frac{\beta_{j, t}^{\prime}}{1-t} \quad$ so that $\quad \beta_{j, t}^{\prime}=(1-t) \beta_{j, t=0}^{\prime}=(1-t) \beta_{j, t=0} \quad$ (linear decrease)

Linear equation for $\gamma_{k, t}$ :

$$
\sum_{k}^{\prime} \gamma_{k, t}\left\langle\partial^{B} W_{k} \partial^{B}\left[-\left(\partial^{A}\right) W_{i}+\partial^{A} S_{\mathrm{eff}, t}^{\prime} \partial^{A} W_{i}\right]\right\rangle_{S_{\mathrm{eff}, t}}=\frac{1}{1-t}\left\langle\partial^{A} S_{\mathrm{eff}, t}^{\prime} \partial^{A} W_{i}\right\rangle_{S_{\mathrm{eff}, t}}
$$

In practice, we use the numerical derivative with the five-point formula to calculate this matrix

## Remarks (1/1)

- Some of the basis functions are not linearly independent ("Mandelstam constraints") Mandelstam 1979

Relevant example:


$$
\because(\operatorname{tr} U)^{2}=\operatorname{tr}\left(U^{2}\right)+2 \operatorname{tr} U^{\dagger} \quad[U \in S U(3)]
$$

We need to choose linearly independent basis in the inversion.

- We use the HMC algorithm with the exact transformed action $S_{\text {eff }}(V)$ developed by Luscher.

Luscher 0907.5491

## Advantages of designing the trivializing map with the Schwinger-Dyson equation

- the basis for the flow kernel can be chosen arbitrarily by hand
- can be applied to the general action of interest
- the coefficients in the kernel, $\gamma_{k, t}$, are determined by lattice estimates of the observables; no need for analytic calculation such as $t$-expansion
- truncation effects and goodness of the flow can be measured by the force norm
- Fully parallelized code based on qlat software (C++ codebase)
https://github.com/waterret/Qlattice
We extended L. Jin's code of field-transformed HMC
Jin LATTICE 2021
to include generic flow kernels including all the footprint 2 loops.
- Most costly part is the multiplication of a matrix including the Hessian $\partial^{A} \partial^{B} \tilde{S}_{t}$ in propagating the force.
E.g., plaquette

rectangle (short edge)

(long edge)

$\partial^{B}: 6$ patterns each
- By dividing the directions of the flowed links and appropriately coloring/masking the lattice for each type of loops, we can run the multiplication in parallel. Lüscher 0907.5491, Boyda,

Kanwar. Pacaniere, Rezende, Albergo, Cranmer, Hackett, Shanahan 2008.05456


Volume scaling is basically linear $O$ (vol) for local Wilson loop bases

> Results

## Computation resources (1/1)

- RIKEN HOKUSAI

HOKUSAI
bIG WATERFALL

- Univ of Tokyo Oakforest-PACS (retired)
- USQCD facility at BNL (KNL) funded by US DOE

Brookhaven
National Laboratory

picture taken from HP of CCS

We are grateful for these resources.

## Evolution of effective action [Wilson action] (1/1)

$8^{4}, \beta=6.13$ Wilson $\left(a^{-1}=2.56 \mathrm{GeV}\right)$
Ce-Consonni-Engel-Giusti 1506.06052
Determined $\gamma_{0, t}$ (plaqutte coefficient)
Difference from the target trajectory of effective action



With the SD method, we can have better control of the effective action

## Evolution of effective action [Wilson action] (1/1)

$8^{4}, \beta=6.13$ Wilson $\left(a^{-1}=2.56 \mathrm{GeV}\right)$
Ce-Consonni-Engel-Giusti 1506.06052
Determined $\gamma_{0, t}$ (plaqutte coefficient)
Difference from the target trajectory of effective action



Naively adding the rectangle term to the LO $t$-expansion makes the deviation more significant

## Autocorrelation [DBW2 action] (1/3)

$$
\begin{array}{ll}
8^{3} \times 16, \beta=0.89 & \left(a^{-1}=1.49 \mathrm{GeV}\right) \\
& \text { DBW2 } \\
\left(c_{1}=-1.4008\right) & \text { Necco hep-lat/0309017 }
\end{array}
$$

## Difference from the target trajectory of effective action



Boyle, Izubuchi, L. Jin, Jung, NM, Lehner, Tomiya work in progress

Computational cost (1 step flow)


The increase is due to the increase of the nonzero matrix elements in $\partial^{A} \partial^{B} \tilde{S}_{t}$.

It seems essential to design the flow with a few types of loops to circumvent large algorithm overhead.

## Autocorrelation [DBW2 action] $(2 / 3)$

$8^{3} \times 16, \beta=0.89$ DBW2 $\quad\left(a^{-1}=1.49 \mathrm{GeV}\right)$
Normalized autocorrelation function $\rho(n)$ for the smeared energy density ( $t_{w}=30 t_{0}$ )


Faster decorrelation (in MC steps) by including extended loops

$$
\begin{aligned}
& t_{W}: \text { Wilson smearing flow time } \\
& \left.t_{W}^{2}\langle E\rangle\right|_{t_{W}=t_{0}}=0.3 \quad \text { Luscher } 1006.4518
\end{aligned}
$$



## Autocorrelation [DBW2 action] (3/3)

Faster decorrelation is observed by the extended loops, but the autocorrelation is not controlled completely: (and statistics should be increased for definite conclusions)


- Integrated autocorrelation time of the topological charge increases exponentially with $\beta$ ( $:$ with a power of $a^{-1}$ ).

If not complete trivialization, we can have large benefits by changing the exponent.

## Possible strategies?

- SD method itself has a large room for optimization choose a different path of $\beta_{t}^{\prime}$ (e.g., exponential decrease, decrease $c_{1}$ ), $t$-dependent step sizes, change basis at each step ...

- On the other hand, the relatively small improvements show that the truncated large loops contribute to increasing the autocorrelation.

In fact, we need an infinite number of loops to obtain the exact trivializing flow, Lüscher 0907.5491 with which we expect to decrease the autocorrelation of all modes of the system;
since we need to deal with finite number of loops, we can be more specific to particular slow observables.
e.g., change instanton potential to stimulate tunneling
cf. Smith, Teper hep-lat/9801008
DeGrand, Hasenfratz, Kovacs hep-lat/9801037
De Forcrand, Garcia Perez, Hetrick, Stamatescu hep-lat/9802017 Hasenfratz-Nieter hep-lat/9806026

## Summary and Outlook

## Summary

- We proposed a way to design an approximate trivializing map with the Schwinger-Dyson equation

Advantages of this method

- the basis for the flow kernel can be chosen arbitrarily by hand
- can be applied to the general action of interest
- the coefficients in the kernel are determined by lattice estimates of the observables; no need for analytic calculation such as $t$-expansion
- truncation effects and goodness of the flow can be measured by the force norm
- We showed that
- With the SD method, we can have a better control of the effective action
- We have positive effects in autocorrelation of long-ranged objects by adding rectangle and chair to the flow


## Outlook

- Improve statistics for definite conclusions \& confirm the scaling
- Develop more efficient strategies
E.g., - Choose different trajectory of $\beta_{t}^{\prime}$ (exponential decrease, decrease $c_{1}, \ldots$ )
- Be specific to particular slow modes (change instanton potential, ...)
- Include fermion (Do we need more extended Wilson loops in the flow? )
cf. Grid: https://github.com/paboyle/Grid
GPT: https://github.com/lehner/gpt

Thank you.

## Integrated autocorrelation time [DBW2 action] $(1 / 1)$



