Sparse modeling approach to obtaining the shear viscosity from smeared correlation functions

Etsuko Itou
(Keio U., Kochi U., RCNP Osaka U.)
collaboration with Yuki Nagai (JAEA, RIKEN AIP)

references:
E.I. and Y. Nagai, JHEP 07 (2020) 007
E.I. and S.Aoki, PoS INPC2016 (2017) 342
Asakawa, Hatsuda, E.I., Kitazawa, Suzuki (FlowQCD coll.) Phys.Rev. D90 (2014) 1, 011501

APLAT 2020, (August 6th, 2020)
Summary of this work

Three difficulties to obtain the shear viscosity

(i) How to define the renormalized EMT on the lattice
(ii) How to improve a bad signal-to-noise ratio of the correlation function of EMT
(iii) How to estimate $\rho(\omega)$ from the limited number of the data $C(\tau)$

As for (i) (ii), the gradient flow method looks promising.

Asakawa, Hatsuda, E.I., Kitazawa, Suzuki, Phys.Rev. D90 (2014) 1, 011501

Today, we propose the sparse modeling method for (iii).
Good tool to extract an essential data from sparse data

$\rho(\omega)$ from $C(\tau)$ is proposed in condensed matter theory

H. Shinaoka, J. Otsuki, M. Ohzeki, K. Yoshimi, Phys. Rev. B 96 (2017) 035147 [arXiv:1702.03054].
J. Otsuki, M. Ohzeki, H. Shinaoka, K. Yoshimi, Phys. Rev. E 95 (2017) 061302 [arXiv:1702.03056].
Sample code of SpM

You can find numerical codes of SpM on arXiv page:

https://arxiv.org/src/2004.02426v2/anc
written in FORTRAN, fortran2003, C++, julia

(1) Give a data of \( C(\tau) \) w/ or w/o statistical error in "samplectau.in"

(2) Write "NT=N_\tau" in "kinput.in"

(3) julia calc_kernel-SVD.jl (make the SVD data, you can use LAPACK etc)

(4) execute “main_v2.f”, then obtain \( \rho(\omega) \) and the output \( C(\tau) \) constructed by obtained \( \rho(\omega) \).

(Note that maximal number of iteration is reduced in the original code. Please change “itemax = 1000” to the large value in real calculation)
Introduction
shear viscosity in lattice calculation

Shear viscosity is given by the spectral function

$$\eta(T) = \pi \frac{d\rho(\omega)}{d\omega} \bigg|_{\omega=0}$$

$\rho(\omega)$ is defined from Euclidean correlation function of the renormalized spatial EMT

$$C(\tau) = \frac{1}{T^5} \int d\vec{x} \langle T_{12}^R(0, \vec{0}) T_{12}^R(\tau, \vec{x}) \rangle = \int_{-\infty}^{+\infty} d\omega K(\tau, \omega) \rho(\omega)$$

In lattice calculation,

1. $C(\tau)$ is measured by generated configurations
2. $\rho(\omega)$ is estimated.

Here, $K(\tau, \omega) = \frac{\cosh \left( \omega \left( \frac{1}{2T} - \tau \right) \right)}{\sinh \left( \frac{\omega}{2T} \right)}$. It is independent of the Monte Carlo data.
Usage of gradient flow method

(i) How to define the renormalized EMT
(ii) How to improve a bad signal-to-noise ratio of the correlation function of EMT
(iii) How to estimate $\rho(\omega)$ from the limited number of the data $C(\tau)$

Renormalized EMT in gradient flow method

$Luescher and Weisz, JHEP 1102, 051(2011)$
$Suzuki, PTEP 2013, no8, 083B03$

$T_{\mu\nu}^R(x) = \lim_{t \to 0} \lim_{\alpha \to 0} \left( \frac{1}{\alpha_U(t)} U_{\mu\nu}(t, x) + \frac{\delta_{\mu\nu}}{4\alpha_E(t)} [E(t, x) - \langle E(t, x) \rangle_0] \right)$

It is supported by the UV finiteness for boson composite operators.
Gradient flow is a continuous stout smearing, so that it reduces the statistical errors.

**Fiducial window of flow-time**

$t$ should be longer than lattice spacing
we want to avoid an over-smeared regime

Theoretically, $2a < \sqrt{8t} < N_\tau a/2$
 Actually, the data show a plateau.

$Asakawa, Hatsuda, E.I., Kitazawa, Suzuki Phys.Rev. D90 (2014) 1, 011501$
Two-point fn. of EMT using the gradient flow method

\[ C(\tau) = \frac{1}{T^5} \int d\vec{x} \langle T_{12}^R(0, \vec{0}) T_{12}^R(\tau, \vec{x}) \rangle \]
Improvement of signal-to-noise ratio

In these statistics, 2000-5000 configurations, $C(\tau)$ are highly fluctuated

Some non-smeared data take negative value because of the large fluctuation.
But smeared $C(t, \tau)$ correctly take positive value with small statistical errors.
Flow-time dependence of $C(t, \tau)$

Lattice size: $64^3 \times 16$, parameter: $\beta = 6/g_0^2 = 6.93$

$\langle T_{12}(0, \vec{x}) T_{12}(\tau, \vec{x}') \rangle$

- Errors get smaller in whole $\tau$-regime
- Slope is changed in shorter $\tau$-regime.

- The data $\tau \lesssim \sqrt{8t}$ is over-smeared, since the smeared regime of $T_{12}(\tau, \vec{x}')$ overlaps $T_{12}(0, \vec{0})$ in $\langle T_{12}^R(0, \vec{0}) T_{12}^R(\tau, \vec{x}') \rangle$ measurement.
- We would like to eliminate them from analysis to estimate $\rho(\omega)$. 
Usage of gradient flow method

(i) How to define the renormalized EMT
(ii) How to improve a bad signal-to-noise ratio of the correlation function of EMT
(iii) How to estimate $\rho(\omega)$ from the limited number of the data $C(\tau)$

As for (i) (ii), the gradient flow method looks promising.
On the other hand, the gradient flow method makes harder (iii).

Essential difficulty of (iii) comes from the smallness of the data points $C(\tau)$.

integral equation

$$C(\tau) = \int_{-\omega_{cut}}^{+\omega_{cut}} d\omega K(\tau, \omega)\rho(\omega)$$

a set of linear eqs.

$$\begin{pmatrix}
C(\tau_1) \\
C(\tau_2) \\
\vdots \\
C(\tau_{N_t})
\end{pmatrix} =
\begin{pmatrix}
K_{11} & K_{12} & \cdots & K_{1N_\omega} \\
K_{11} & K_{12} & \cdots & K_{1N_\omega} \\
\vdots & \vdots & \ddots & \vdots \\
K_{N_t1} & K_{N_t2} & \cdots & K_{N_tN_\omega}
\end{pmatrix}
\begin{pmatrix}
\rho_1 \\
\rho_2 \\
\vdots \\
\rho_{N_\omega}
\end{pmatrix}$$

If $N_t < N_\omega$, then there are several possible solutions.

As explained, the flowed $C(\tau)$ is deformed by over-smearing.

If we eliminate the over-smereared data, then the situation gets worse.
Solve the inverse problem

We need a powerful tool to find $\rho(\omega)$ from very limited data of $C(\tau)$

Until now, several estimation methods, $\rho(\omega)$ from $C(\tau)$, have been proposed.

- fitting the data using some functional form of $\rho(\omega)$
  
  e.g., Breit-Wigner ansatz
  
  $$\frac{\rho(\omega)}{\omega} = \frac{F}{1 + b^2(\omega - \omega_0)^2} + \frac{F}{1 + b^2(\omega + \omega_0)^2}$$

- find a likely function based on Bayesian statistics
  
  e.g., maximum entropy method

Sparse-modeling (SpM) method

1) satisfy the Bayesian theorem
2) perform the SVD of kernel matrix

$$K_{ij} = U_{ik} S_{kl} V_{lj}^\dagger$$

$$\begin{cases} S : N_\tau \times N_\omega \text{ diagonal matrix} \\ U : N_\tau \times N_\tau \text{ unitary matrix} \\ V : N_\omega \times N_\omega \text{ unitary matrix} \end{cases}$$

3) transform vectors $\vec{C}$ and $\vec{\rho}$ into the IR (SVD) basis using unitary matrices, then the rank of $\vec{C}'$ becomes the same with $\vec{\rho}'' (= N_\omega)$

4) cut the modes with sufficiently small singular values. It stabilizes numerically

5) add $L_1$ regularization term to the optimization problem to be consistent with a reduction of modes
SpM

Find a “best” solution which is described by a small-component vector $\vec{\rho}'$ to be consistent with the cutoff of modes.

It is possible to do by introducing $L_1$ regularization into the optimization problem:

\[
F(\vec{\rho}') = \frac{1}{2} \| \vec{C}' - S\vec{\rho}' \|_2^2 + \lambda \| \vec{\rho}' \|_1
\]

$L_1$ term: $\| \vec{\rho}' \|_1 \equiv \sum_l |\rho'_l|$.

Find minimum of the sum

\[
\frac{1}{2} \| \vec{C}' - S\vec{\rho}' \|_2^2 = \text{const.} \quad \text{and} \quad \| \vec{\rho}' \|_1 = \text{const.}
\]

by tuning $\lambda$ (Lagrange multiplier).

*L1* term favors the solution with a small number of components. The tendency will give a consistent solution with a cutoff $s_l$. 
The spectral function becomes featureless in the under-fitting, where the $L_1$ regularization term is too strong and the number of components $\vec{\rho}'$ is too reduced.

Artificial spikes appear in the over-fitting, where the $L_1$ term is too weak and the vector $\vec{\rho}'$ has redundancy.

---

**Role of $L_1$ term**

Test in a statistical model

J. Otsuki et al, Phys. Rev. E 95 (2017) 061302
Standard cost function

The optimization problem with $L_1$ and/or $L_2$ regularization is called the LASSO (Least Absolute Shrinkage and Selection Operators) problem.

It allows obtaining the global minimum regardless of initial conditions.

We can also add the non-negativity $\rho(\omega) \geq 0$ and/or the sum rule $\sum_j \rho_j = 1$.

The standard cost function in this work is:

$$\tilde{F}(\bar{\rho}', \tilde{z}', \tilde{z}) = \frac{1}{2\lambda} \| \bar{C}' - S \bar{\rho}' \|_2^2 - \nu(\langle V \bar{\rho}' \rangle - 1) + \| \tilde{z}' \|_1 + \lim_{\gamma \to \infty} \sum_j \Theta(-z_j)$$

where $\tilde{z}, \tilde{z}'$ are auxiliary vectors, and minimize the cost fn. to be $\tilde{z}' = \bar{\rho}'$, $\tilde{z} = V \bar{x}'$.

The ADMM algorithm:

The numerical algorithm is called the alternating direction method of multipliers (ADMM) algorithm, where some auxiliary vectors are introduced to satisfy the conditions.

S. Boyd et al., Foundations and Trends R in Machine Learning 3, 1 (2011).
See also Appendix A in our paper.
Results of test calculation

J. Otsuki, M. Ohzeki, H. Shinaoka, K. Yoshimi, Phys. Rev. E 95 (2017) 061302 [arXiv:1702.03056].

SpM can be applied for non-smeared $C(\tau)$ without the statistical error.

Preparations:

$\rho(\omega)^{\text{exact}}$ is given by hand (it has 3 Gaussian peaks)

$G(\tau)^{\text{exact}}$ is constructed by $\rho(\omega)^{\text{exact}}$

Test(1): Estimate $\rho(\omega)^{\text{SpM}}$ from $G(\tau)^{\text{exact}}$

Test(2): Make $G(\tau)^{\text{input}} = G(\tau)^{\text{exact}} + \eta$ and estimate $\rho(\omega)^{\text{SpM}}$

We find SpM works well and it looks very stable around $\omega \approx 0$.

(shear viscosity: $\eta \propto d\rho(\omega)/d\omega |_{\omega=0}$)
Simulation results for quenched QCD
simulation setup

* Lattice action: Wilson plaquette gauge action
* Lattice size: $64^3 \times 16$
* parameter: $\beta = 6/g_0^2 = 6.93$
* # of configurations: 2,000

  cf.) Nakamura-Sakai(2005): 800,000 conf.
  Borsanyi et al.(2018): 6 million conf.

* Temperature, $T = 1.65T_c$

  ALPHA collaboration NPB535 (1998)389,
  G.Boyd et al., NPB469(1996)419

* the thermal entropy @ $T = 1.65T_c$: $s/T^3 = 4.98(24)$ in continuum limit

  Asakawa, Hatsuda, E.l., Kitazawa, Suzuki, PRD90 (2014) 1, 011501
Results of spectral function

\[
\tilde{\rho}(t, a\omega)
\]

sum of d.o.f. \( N = \int_{-a\omega_{\text{cut}}}^{a\omega_{\text{cut}}} \tilde{\rho}(t, a\omega) d(a\omega) \) gets smaller in large flow-time.

The gradient flow can be interpreted as a renormalization group flow. Renormalization group decreases the d.o.f. of the system. The higher frequency modes are gradually suppressed by the gradient flow. The results support this intuitive property of the gradient flow. We also see the statistical error in long flow time is small as expected.
Comparison with input and output $C(\tau)$

$C_{\text{output}}(t, \tau/a)$ is constructed by the obtained $\tilde{\rho}(\omega)$

The output $C(\tau)$ almost reproduce the input data.

We consider the SpM analysis works well.
Current status on shear viscosity

We have not taken $a \to 0$ limit and then $t \to 0$ limit yet, therefore, it may not be fair to compare our data with the other works….

A poor statistics of our data (only 2,000 conf.) can give a result of the viscosity by combination of the gradient flow and SpM methods.
Toward precise determination of viscosity

(i) How to define the renormalized EMT
(ii) How to improve a bad signal-to-noise ratio of the correlation function of EMT
(iii) How to estimate $\rho(\omega)$ from the limited number of the data $C(\tau)$

$$C(\tau) = \frac{1}{T^5} \int d\vec{x} \langle T_{12}^R(0, \vec{0}) T_{12}^R(\tau, \vec{x}) \rangle = \int_{-\infty}^{+\infty} d\omega K(\tau, \omega) \rho(\omega)$$

$C(\tau)$ is measured using the gradient flow for (i) and (ii)

$\rho(\omega)$ is estimated using the sparse modeling method for (iii)

Both theoretically and technically, these methods look promising. It means that these methods reduce the noise and give a stable results.

The sparse modeling is a general framework to estimate $\rho(\omega)$ from $C(\tau)$. I hope that it will be a powerful tool for various subjects in lattice QCD calculations.