Deep Finite Temperature Bootstrap

Costis Papageorgakis, Queen Mary University of London

Athens Christmas Workshop in Theoretical Physics 2025
18th of December 2025

Based on 2508.08560 with V. Niarchos, A. Stratoudakis and M. Woolley

The bootstrap philosophy

Consistency conditions should be enough to completely determine physics

E.g. old S-matrix programme using analyticity, unitarity and crossing to determine observables for strong interactions

A version of this idea used to great effect in the modern conformal bootstrap

⇒ CFTs are ubiquitous: UV/IR behaviour of QFTs, critical phenomena, quantum gravity via AdS/CFT correspondence…

But there are many other applications:

- Modern S-matrix bootstrap
- Cosmological bootstrap
- Matrix quantum mechanics bootstrap
- Modular bootstrap

Motivation

Conformal bootstrap programme opens the way for nonperturbative solution of CFTs and numerical bootstrap has yielded impressive results,

Relies on leveraging sum rules arising from consistency conditions:

$$\sum_J \sum_\Delta \vec{\mathscr{X}}_{\Delta,J} = 0$$

Infinite unknowns involved and non-convex: direct solution is hard

Workaround: The linear functional method maps to convex linear/semi-definite problem

⇒ Rigorous, high-precision results obtained through e.g. SDPB
[Rattazzi, Rychkov, Tonni, Vichi '08, Simmons-Duffin '15]

Revisit direct (constructive) solutions towards:

- Exploring the solution space and guiding rigorous approaches
- Developing more efficient numerics (e.g. scaling multi-correlator bootstrap)
- Applying bootstrap approach to scenarios without positivity (e.g. finite temperature, defects/boundaries, higher-point correlators, non-unitarity…)

However: How to attack a problem with infinite unknowns?

⇒ Hard truncations: drop all unknowns above a cutoff [Gliozzi '13]

⇒ Soft truncations: model contributions above a cutoff
(e.g. effective operators [Niarchos, CP, Richmond, Stapleton, Woolley '23],
Tauberian theorems [Marchetto, Miscioscia, Pomoni '23])

Can be effective for specific cases but in general hard to control in terms of systematic error…

Today

⇒ Introduce a novel framework for capturing all unknowns using dispersion relations and neural networks, without dropping any terms

⇒ Showcase a numerical application in the context of thermal CFTs dual to black holes in AdS

The main idea

Consider the following split of the sum rule:

$$\sum_J \sum_\Delta \vec{\mathcal{X}}_{\Delta,J} = 0$$

The main idea

Consider the following split of the sum rule:

$$\sum_{J} \sum_{\Delta} \vec{\mathcal{X}}_{\Delta,J} = 0$$

$$\sum_{J>J_*} \sum_{\Delta} \vec{\mathcal{X}}_{\Delta,J}$$

'High-spin' part
(approximated through
dispersion relations)

The main idea

Consider the following split of the sum rule:

$$\sum_J \sum_\Delta \vec{\mathcal{X}}_{\Delta,J} = 0$$

$$\sum_{J>J_*} \sum_\Delta \vec{\mathcal{X}}_{\Delta,J}$$

'High-spin' part (approximated through dispersion relations)

$$\sum_{J\leq J_*} \sum_\Delta \vec{\mathcal{X}}_{\Delta,J}$$

'Low-spin' part

The main idea

Consider the following split of the sum rule:

$$\sum_J \sum_\Delta \vec{\mathcal{X}}_{\Delta,J} = 0$$

$$\sum_{J>J_*} \sum_\Delta \vec{\mathcal{X}}_{\Delta,J}$$

'High-spin' part (approximated through dispersion relations)

$$\sum_{J\leq J_*} \sum_\Delta \vec{\mathcal{X}}_{\Delta,J}$$

'Low-spin' part

$$\sum_{J\leq J_*} \sum_{\Delta\leq\Delta_*(J)} \vec{\mathcal{X}}_{\Delta,J}$$

'Low-spin' & 'low-dimension' part (discrete data)

The main idea

Consider the following split of the sum rule:

$$\sum_J \sum_\Delta \vec{\mathcal{X}}_{\Delta,J} = 0$$

$$\sum_{J>J_*} \sum_\Delta \vec{\mathcal{X}}_{\Delta,J}$$

'High-spin' part (approximated through dispersion relations)

$$\sum_{J\leq J_*} \sum_\Delta \vec{\mathcal{X}}_{\Delta,J}$$

'Low-spin' part

$$\sum_{J\leq J_*} \sum_{\Delta\leq\Delta_*(J)} \vec{\mathcal{X}}_{\Delta,J}$$

'Low-spin' & 'low-dimension' part (discrete data)

$$\sum_{J\leq J_*} \sum_{\Delta>\Delta_*(J)} \vec{\mathcal{X}}_{\Delta,J}$$

'Low-spin' & 'high-dimension' part (NNs)

The main idea

Consider the following split of the sum rule:

$$\sum_{J}\sum_{\Delta}\vec{\mathcal{X}}_{\Delta,J}=0$$

$$\sum_{J>J_*}\sum_{\Delta}\vec{\mathcal{X}}_{\Delta,J}$$

'High-spin' part (approximated through dispersion relations)

$$\sum_{J\leq J_*}\sum_{\Delta}\vec{\mathcal{X}}_{\Delta,J}$$

'Low-spin' part

$$\sum_{J\leq J_*}\sum_{\Delta\leq\Delta_*(J)}\vec{\mathcal{X}}_{\Delta,J}$$

'Low-spin' & 'low-dimension' part (discrete data)

$$\sum_{J\leq J_*}\sum_{\Delta>\Delta_*(J)}\vec{\mathcal{X}}_{\Delta,J}$$

'Low-spin' & 'high-dimension' part (NNs)

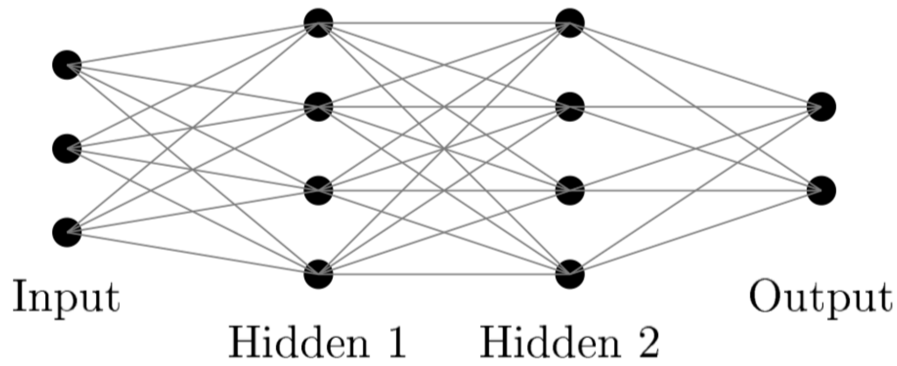
⇒ The J* and Δ* are arbitrary

Artificial Neural Networks

Multi-Layer Perceptron (MLP): $f_\theta : \mathbb{R}^{\text{in}} \to \mathbb{R}^{\text{out}}$

Each layer takes: $\vec{h}^{(\ell+1)} = \sigma\left(W^{(\ell)}\vec{h}^{(\ell)} + \vec{b}^{(\ell)} \right)$

The parameters $\vec{\theta} = \{W^{(\ell)}, \vec{b}^{(\ell)}\}$ are learnt during training, aiming to minimise the loss



Universal approximation theorem: for nonlinear σ the NN can approximate any continuous function

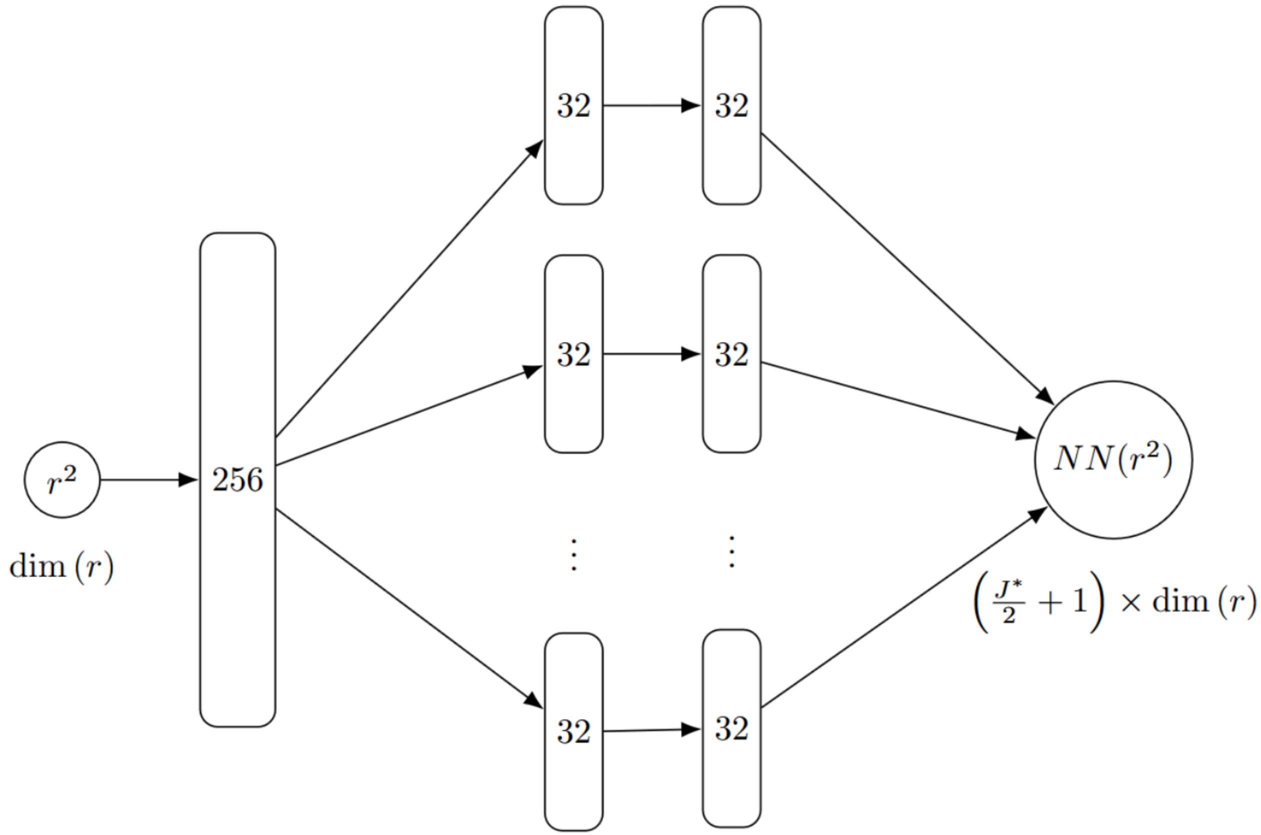
More explicitly for thermal CFTs:

$$\sum_{J \leq J_*} \sum_{\Delta > \Delta_*(J)} \vec{\mathcal{X}}_{\Delta,J}$$

↓

$$\sum_{J < J_*} A_J(r) C_J^{(d)}(\cos\theta)$$

and the $A_J(r)$ are 1D 'tail' functions approximated using PINNs



Summary so far

Try to find constructive solutions to sum rules for bootstrap problem

Capture 'high-spin' contributions through dispersion relations, and optimise for the 'low-spin' contributions (discrete data and NNs for 'tails')

Organise search in terms of a semi-positive definite loss function $\mathscr{L}(D, \vec{\theta})$ that we want to minimise (D : discrete data $\vec{\theta}$: NN parameters)

Choose for loss function:

- L_p norm of the crossing vector ('abs-loss' for p=1)
- A dot-product loss for the crossing vector ('dot-loss')

Application to Thermal Bootstrap

Consider CFT on $\mathbb{R}^{d-1} \times S^1_\beta$ and associated 2pt function:

$$g(\tau, |x|) = \langle \phi(x)\phi(0) \rangle_\beta \,, \quad |x| = \sqrt{\tau^2 + \vec{x}^2}$$

Choose $\vec{x} = (\sigma, 0, \ldots, 0)$ and $z := \tau + i\sigma = rw, \ \ \bar{z} := \tau - i\sigma = rw^{-1}$

The Kubo-Martin-Schwinger (KMS) condition dictates (for $\beta = 1, \ |x| = r$) [El-Showk, Papadodimas '11]

$$g(\tau, r) = g(1 - \tau, r)$$

$$g(1 - z, 1 - \bar{z}) = g(z, \bar{z})$$

Expand 2pt function using ‘thermal OPE’

$$g(rw, rw^{-1}) = \sum_{\mathcal{O}_{\Delta,J} \in \phi\times\phi} a_{\mathcal{O}_{\Delta,J}} C_J^{(\frac{d-2}{2})}\left(\frac{1}{2}(w + w^{-1})\right) r^{\Delta - 2\Delta_\phi}$$

where the thermal 1pt functions $b_{\mathcal{O}}$ are encoded through

$$a_{\mathcal{O}} := \frac{f_{\phi\phi\mathcal{O}} b_{\mathcal{O}}}{c_{\mathcal{O}}} \frac{J!}{2^J (\frac{d-2}{2})_J}$$

Bootstrap: Assuming knowledge of zero-temperature data ($f_{\phi\phi\mathcal{O}}$, Δ) KMS condition poses nontrivial constraints for thermal 1pt functions

Use our framework ('high/low-spin') to write the 2pt function as:

$$
\begin{aligned}
g(rw, rw^{-1}) \;=\;& \sum_{J=0}^{J_*} \sum_{\Delta \le \Delta_*(J)} a_{\Delta,J} C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi} \\
&+ \sum_{J=0}^{J_*} A_{\Delta_*(J),J}(r)\, C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) \\
&+2\left(\int_{-\infty}^{-r^{-1}} + \int_{r^{-1}}^{\infty}\right) dw'\, \mathcal{K}_{J_*}(w,w')\, \mathrm{Disc}\left[g(rw', rw'^{-1})\right]
\end{aligned}
$$

Discrete data

Tail functions

controlled approx.

Δ_*, J_* are freely tuneable. The discontinuity involves reduced theory-specific input.

Exact: No terms have been dropped!

Calibration: Generalised Free Fields

Analytic solution known. For a scalar with dimension Δ_ϕ:

double-twist operators

$$g(z,\bar{z}) = \sum_{m=-\infty}^{\infty} \frac{1}{\left[(m-z)(m-\bar{z})\right]^{\Delta_\phi}}$$

$$[\phi\phi]_{n,J}, \quad n = 0, 1, \ldots, \quad J = 2\ell \quad \ell = 0, 1, \ldots$$

$$\Delta_{n,J} = 2\Delta_\phi + 2n + 2\ell$$

$$g(rw, rw^{-1}) = r^{-2\Delta_\phi} + \sum_{n=0}^{\infty}\sum_{\ell=0}^{\infty} a_{n,2\ell}\, C_{2\ell}^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{2(n+\ell)}$$

$$a_{n,J} = 2\zeta(2\Delta_\phi + 2n + J)\frac{(J+\nu)(\Delta_\phi)_{J+n}(\Delta_\phi - \nu)_n}{n!(\nu)_{J+n+1}}$$

Given this knowledge, how good is our approximate KMS condition?

J_*	$\mathcal{L}_{\overline{\text{abs}}}$	$\mathcal{L}_{\text{dot}(0)}$
0	0.1177	0.0125
2	0.0300	1.3×10^{-5}
4	0.0067	1.6×10^{-7}
6	0.0013	3.3×10^{-9}
8	0.0002	0.8×10^{-10}

0 discrete data

J_*	$\mathcal{L}_{\text{dot}(1)}$	$a_{1,0} + 3\,a_{2,0}$ from dot loss
2	9.5×10^{-6}	15.16582
4	1.4×10^{-6}	15.07614
6	7.8×10^{-8}	15.06252
8	3.5×10^{-9}	15.06049
		Exact value: 15.06013

1 discrete datum

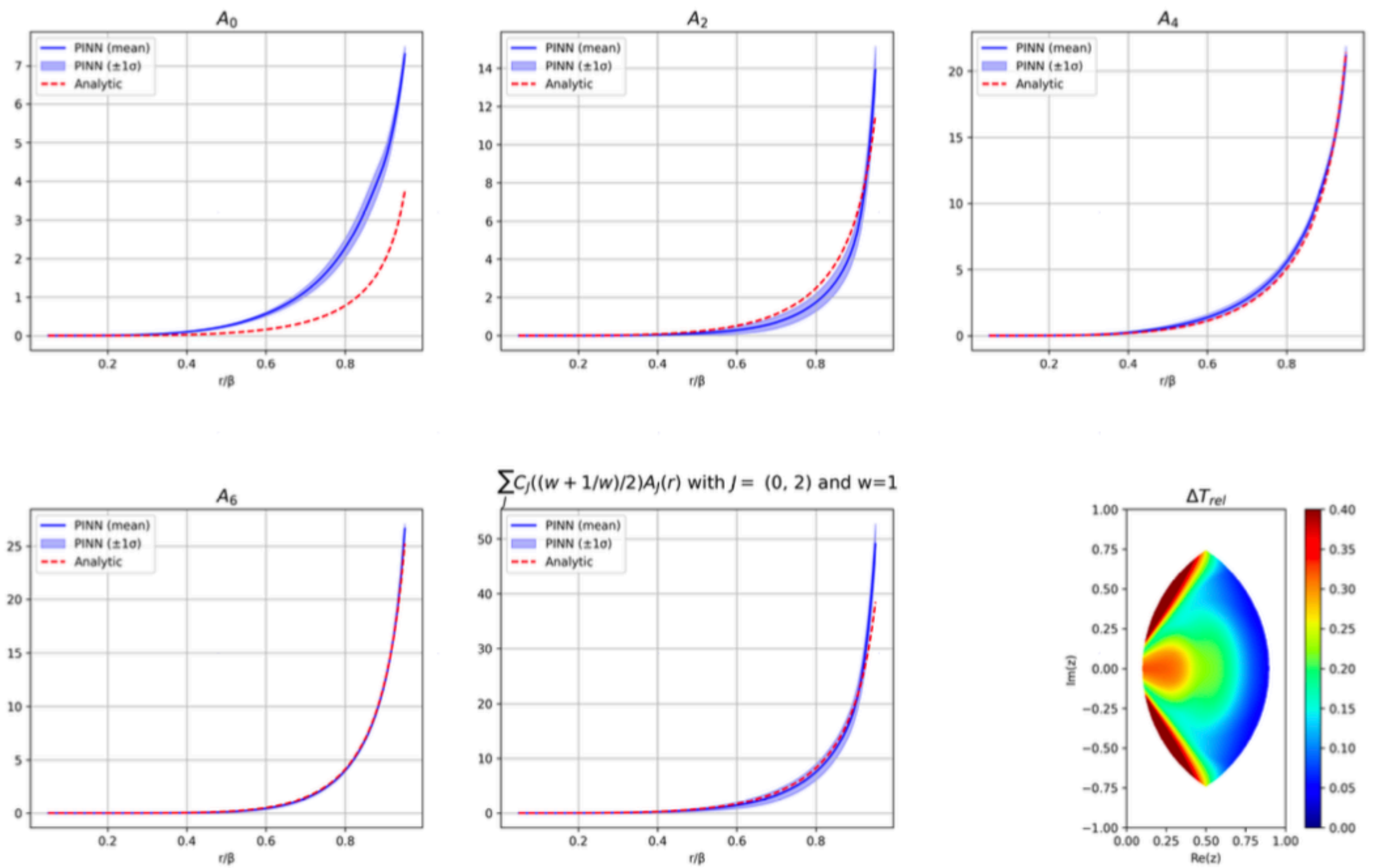
$\Rightarrow$ Approximation works well and improves systematically as J^* increases

Results: GFF with no analytic input

For d=4, $J^* = 6$, $\Delta_\phi = 1.68$ and one discrete datum we have the 4 tails and also:

$$a_{1,0} + 3a_{0,2} = 13.29 \pm 2.82$$

with analytic 15.06013

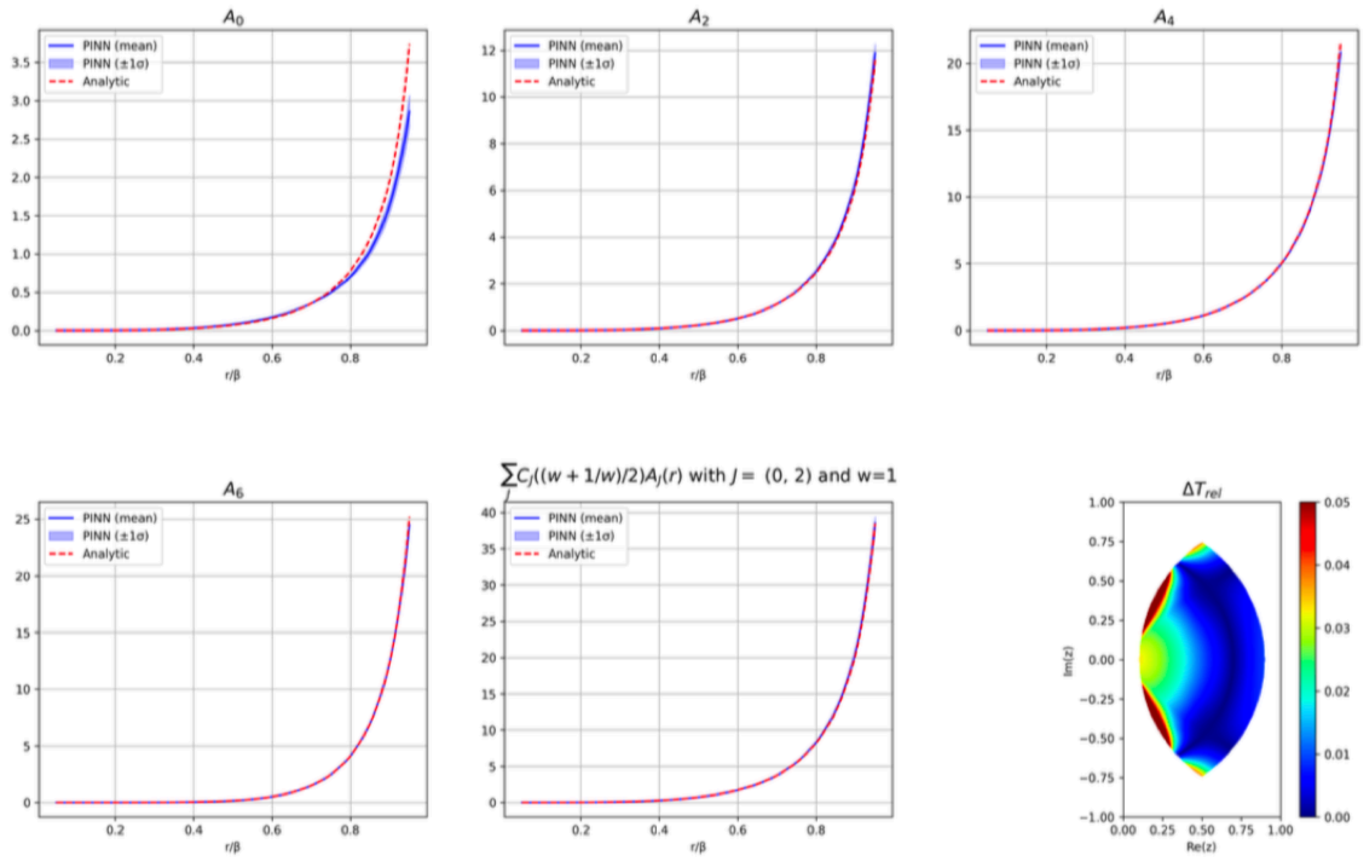


Results: GFF with analytic ‘anchor’ at $r = 0.7$

For d=4, $J^* = 6$, $\Delta_\phi = 1.68$ and one discrete datum we have the 4 tails and also:

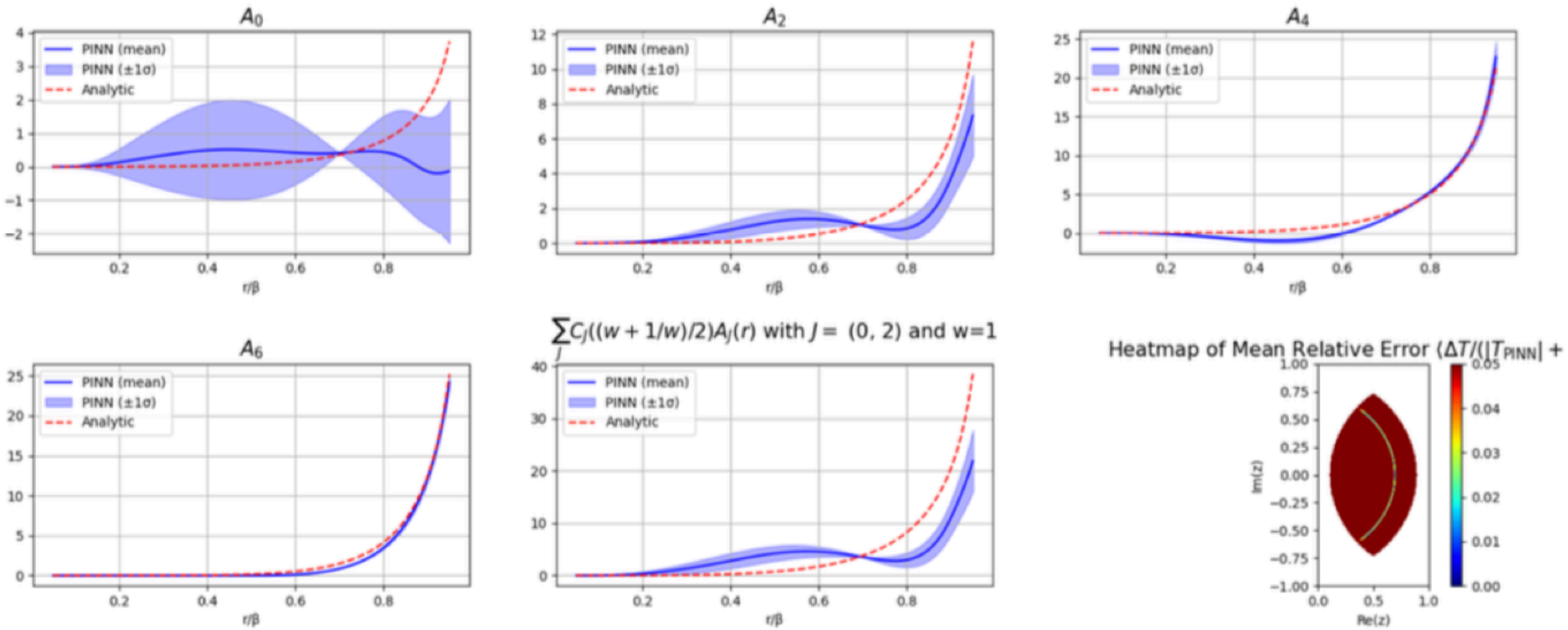
$$a_{1,0} + 3a_{0,2} = 15.0647 \pm 0.0291$$

with analytic 15.06013



GFF results summary

- Recovering the approximate correlator is not stable without additional input
- But once correct ‘anchor point’ is provided solution is determined
- If an incorrect anchor point is provided stability is destroyed



Holographic correlators

$\Rightarrow$ Find solution through stability argument in cases of interest:
4D holographic correlators in a large-c CFT with a gravity dual

The OPE involves:

Operator	Scaling dimension Δ	Spin J	$a_{\Delta,J}$
Identity	0	0	$a_1 = 1$
$T_{\mu\nu}$	d	2	a_T
$[\phi\phi]_{n,J}$	$2(\Delta_\phi + n + \ell)$	$0 \leq 2\ell$	$a_{n,J}$
$[T^k]_J$	$d\,k$	$0 \leq 2\ell \leq 2k$	$a_J^{(k)}$

Note: Same spectrum, infinite KMS solutions corresponding to arbitrary higher-derivative gravities!

Holographic computation: Solve free wave equation on black brane background

- Energy momentum data determined from asymptotics [Fitzpatrick, Huang '19]
- Lowest-twist energy-momentum multi-trace data have $a_{2k}^{(k)} \propto a_T^k$ (universality)
- Double-twist data harder, requiring full bulk solution

Questions for bootstrap:

- Can we detect multiple KMS solutions corresponding to different gravities?
- Can we detect universality relations?
- Can we recover the double-twist data?

Holographic computation: Solve free wave equation on black brane background

- Energy momentum data determined from asymptotics [Fitzpatrick, Huang '19]
- Lowest-twist energy-momentum multi-trace data have $a_{2k}^{(k)} \propto a_T^k$ (universality)
- Double-twist data harder, requiring full bulk solution

Questions for bootstrap:

- Can we detect multiple KMS solutions corresponding to different gravities?
- Can we detect universality relations?
- Can we recover the double-twist data? $\Leftarrow$ Go for this

Results 3: Holographic theories

Pick $J^* = 6$, $\Delta_\phi = 1.5$ and optimise for $a_{1,0} + 3a_{0,2}$. Include the following 7 holography-provided data in the discontinuity:

$$\text{identity} + T_{\mu\nu} + [T^2]_0 + [T^2]_2 + [T^2]_4 + [T^3]_4 + [T^3]_6 + [T^4]_8$$

$$a_{T,\mathrm{GR}} = 1.21761,\ a^{(2)}_{0,\mathrm{GR}} = -1.37668,\ a^{(2)}_{2,\mathrm{GR}} = 1.58848,\ a^{(2)}_{4,\mathrm{GR}} = -4.05945,$$
$$a^{(3)}_{4,\mathrm{GR}} = 1.77035,\ a^{(3)}_{6,\mathrm{GR}} = 8.52362,\ a^{(4)}_{8,\mathrm{GR}} = -15.9641.$$

To help the search, use GFF finite-r input as a starting point, use a hinge loss and look for stable solutions

Results 3: Holographic theories

The loss function is then $\mathscr{L}_{\text{KMS}} + \mathscr{L}_{\text{ReLU}}$ with

$$\mathcal{L}_{\text{ReLU}} = \frac{1}{\frac{J_*}{2}+1} \sum_J \text{ReLU}\left(\left| \mathcal{A}_J(r_i) - \mathcal{A}_J(r_i)|_{\text{GFF}} \right| - \text{p} \times \left| \mathcal{A}_J(r_i)|_{\text{GFF}} \right| \right)$$

$$\text{ReLU}(x) = \max(0, x)$$

and varying p until we hit a stable configuration. This happens around p = 0.20

$$\text{p} = 0.20 \;\; : \;\; a_{10} + 3a_{02} = 9.37 \pm 0.44$$

Compare to zero-spatial separation result (using 40 energy-momentum data): 7.686 [Burić, Gusaev, Parnachev '25]

Summary

- Described a new framework to directly bootstrap without truncations
- High operator-dimension tails can be bootstrapped using NNs
- Presented examples in the case of the thermal bootstrap but the method can be applied to any other bootstrap context, including where positivity constraints are not available
- Within finite-T: Unlike other approaches, we work at non-zero spatial separation throughout

Outlook

- Improve optimisation accuracy: incorporate more energy-momentum data and dynamical discontinuities to get better double-twist data predictions in holographic CFTs
- Better understanding of the island of stability criteria for algorithm convergence
- Extend applications to zero-T, theories with defects/boundaries, higher-point bootstrap…

Outlook

- Improve optimisation accuracy: incorporate more energy-momentum data and dynamical discontinuities to get better double-twist data predictions in holographic CFTs
- Better understanding of the island of stability criteria for algorithm convergence
- Extend applications to zero-T, theories with defects/boundaries, higher-point bootstrap…

Thank you

Note on loss functions

- L_p-norm (a rather obvious choice): If $\vec{\mathscr{F}}$ is the crossing vector we try to set to 0, then

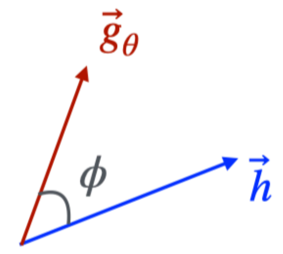
$$\mathscr{L}(\vec{\mathscr{F}}) := \left(\sum_{n=1}^{\dim \vec{\mathscr{F}}} \left| \mathscr{F}_n \right|^p \right)^{1/p} \qquad \text{for} \quad p \geq 1 \qquad\qquad (\textbf{abs-loss} \text{ for } p = 1)$$

- A (novel) alternative that works better: **dot-loss**

V1: D is empty (**no exposed CFT data**) and $\vec{\mathscr{F}} = \vec{g}_\theta - \vec{h}$ ($\vec{g}_\theta$ optimizable, $\vec{h}$ known)

$$\mathscr{L}(\vec{\mathscr{F}}(\theta)) = 1 - \frac{|\vec{g}_\theta \cdot \vec{h}|}{|\vec{g}_\theta||\vec{h}|} + \left| 1 - \frac{\vec{g}_\theta \cdot \vec{h}}{|\vec{h}|^2} \right|$$

$$0 \leq 1 - |\cos(\phi)| \leq 1$$



V2: D has one element (**1 exposed CFT datum a**) and

$$\overrightarrow{\mathcal{F}}(a,\vec{\theta}) = a\vec{f} + \vec{g}_\theta - \vec{h} \qquad (\vec{g}_\theta \text{ optimizable, } \vec{h} \text{ known, } \vec{f} \text{ known crossed block})$$

In this case:

$$\mathcal{L}(\overrightarrow{\mathcal{F}}(\theta)) = 1 - \frac{|(\vec{g}_\theta - \vec{h}) \cdot \vec{f}|}{|\vec{g}_\theta - \vec{h}||\vec{f}|}$$

you optimize only with respect to the unknown tail functions and you get the coefficient a for free at the end

$$a = -\frac{(\vec{g}_{\theta_*} - \vec{h}) \cdot \vec{f}}{|\vec{f}|^2}$$