INTRODUCTION

Lifestyle [1] and pharmacotherapy interventions [2,3] have been shown in randomised trials to be effective in achieving weight loss among individuals with overweight or obesity with high cardiovascular (CVD) risk because of type 2 diabetes, cardiovascular diseases or additional risk factors [3,4] In such high risk patients, there is evidence that weight loss is also beneficial for cardiovascular risk factors [4]. However, there are mixed findings on the effect of weight loss on CVD, as some studies found no effect of weight loss on fatal and non-fatal CVD [1,5,6], while a recent meta-analysis of trials reported moderate lower risk of CVD following weight loss [4]. In any case, there are no randomised controlled trials (RCTs) of any weight loss intervention assessing the effectiveness for primary prevention of cardiovascular diseases (CVDs) in otherwise healthy individuals with overweight or obesity. The prospect of performing such trials is low because of the very high number of participants required for reliable estimation of a primary preventive effect on CVD events. Therefore, it is not known whether weight loss reduces the risk of incident cardiovascular disease among people in the general population with overweight or obesity. This is particularly important as higher body mass index is associated with the onset of cardiovascular diseases [7] and the prevalence of obesity, already high, is predicted to risk further [8].

In such settings emulating target pragmatic trials using causal inference methods in large scale observational data may play a role in distinguishing effects of weight change in people with normal weight, overweight and obesity [9-12]. Observational studies of weight (or BMI) change are conflicting, some reporting increased risk of CVD [13-16], no association [17-20] or lower risk, especially after bariatric surgery in people with severe obesity [21]; weight gain has been associated with increased CVD risk in some studies[13-16], but not others[21]. Emulation of weight loss trials has been carried out in a consented cohort, the Nurses' Health Study, which found no relationship between weight loss and CHD [17,18].

Short Biography of Author

Maria Rosaria Acocella is currently a researcher at the University of Salerno. She received her B.Sc. in 2001 and PhD in 2005 from the University of Salerno. She was a postdoctoral research fellow at the same university, studying new homogeneous and heterogeneous catalysts and their possible application in reactions of synthetic interest, with particular attention to the catalysts with low environmental impacts. Since 2012, her research has concerned green functionalization, characterization and the application of heterogeneous carbon-based catalytic systems. She has published over 58 peer-reviewed articles, including ACS Catalysis, ChemSusChem, ChemCatChem, Carbon and Chemistry of Materials.

Given that the introduction of a speed bump increases trading latency, it is expected to reduce HFT activity. Therefore, if our ML-generated HFT metrics accurately capture the dynamics of HFT activity, we should observe a reduction in the metrics on Amex post the speed bump implementation. To formally test this hypothesis, we employ the following stock-day regression:

$$HFT_{i,t}^{ML,D} = \alpha_i + \beta_t + \gamma_1 Post_{i,t} * Amex_{i,t} + \sum_{k=1}^{4} \delta_{i,t}^{k} C_{i,t}^{k} + \varepsilon_{i,t} \quad (4)$$

$$HFT_{i,t}^{ML,S} = \alpha_i + \beta_t + \gamma_2 Post_{i,t} * Amex_{i,t} + \sum_{k=1}^{4} \delta_{i,t}^{k} C_{i,t}^{k} + \varepsilon_{i,t} \quad (5)$$

where $HFT_{i,t}^{ML,D}$ and $HFT_{i,t}^{ML,S}$ correspond to ML-generated liquidity-demanding and -supplying HFT activity, respectively. We incorporate stock-specific fixed effects (α_i) and day fixed effects (β_t) to account for individual stock characteristics and daily variations, respectively. $Post_{i,t}$ is an indicator variable, taking the value of 1 on July 24, 2017, when the speed bump was implemented and thereafter, and 0 before, while $Amex_{i,t}$ corresponds to 1 for NYSE Amex-listed stocks and 0 for NYSE- and NASDAQ-listed firms. The standard errors are double clustered by firm and day.

$C_{i,t}^{k}$ includes a range of control variables, such as volatility ($Volatility_{i,t}$), relative quoted spread ($Spread_{i,t}$), inverse price ($InvPrice_{i,t}$), and trading volume in dollars ($Volume_{i,t}$). $Volatility_{i,t}$ is calculated as the daily (t) standard deviation of the transactional-level returns for stock i. $Spread_{i,t}$ is the daily average of transaction-level bid-ask spreads. The transaction-level bid-ask spread is calculated as the difference between ask and bid prices divided by the average of ask and bid prices for each transaction. All these variables are obtained from the TAQ database.

A few points of clarification are necessary with respect to the estimating of Models (4) and (5). First, our model does not explicitly include $Post_{i,t}$ and $Amex_{i,t}$ indicator variables, as their effects are already accounted for through the inclusion of time and stock fixed effects.

and $\delta E(\mathbf{h})$ is the energy difference between initial and final collinear particles, expanded to first nontrivial order in thermal masses and transverse momenta:

$$\delta E(\mathbf{h}) = \frac{h^2}{2pk(p-k)} + \frac{m_{\infty,c}^2}{2k} + \frac{m_{\infty,b}^2}{2(p-k)} - \frac{m_{\infty,a}^2}{2p}. \tag{3.13}$$

We encountered $m_{\infty,q}^2$ already in Eq. (3.4). For gluons, $m_{\infty,g}^2 = m_D^2/2$.

3.2 Next-to-Leading Order corrections

We now focus on the next-to-leading order corrections. Treating kinetic theory to NLO was pioneered in Ref. [33, 34] and applied to QCD shear viscosity in Ref. [15]. In this work, the authors observed that the analytically complex parts of the collision operator do not receive $\mathcal{O}(g)$ corrections, which instead only modify collinear exchange, incorporated by changing γ_{bc}^a, and soft exchange, which can be modeled as momentum-diffusion and identity-changing processes. We briefly review the extended treatment that captures these $O(g)$ corrections, and then we summarize it and present an explicit form of the collision operator at NLO.

3.2.1 Momentum diffusion

The small momentum exchange part of diagrams (A), (B), (C) can be well approximated by momentum diffusion for the particles on each line. Beyond leading order, there are additional small-momentum-exchange effects which are only suppressed by $\mathcal{O}(g)$ relative to LO, which must be incorporated. One such effect is the possibility for the gluon in diagrams (A), (B) to carry a soft momentum. Other effects involve the interference between multiple soft scatterings, soft gluon emission during scattering, and nonabelian interactions between scattering processes, such as those depicted in Figure 2.

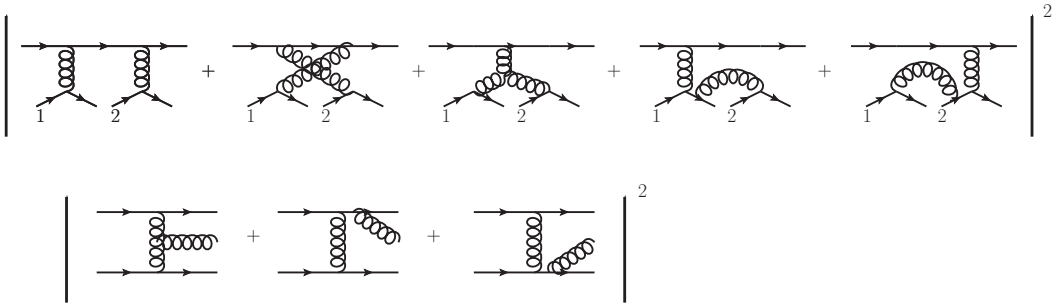

Figure 2. Some scattering effects which affect momentum diffusion at NLO.

Because the momentum exchanges involved in these processes are small (that is the reason they can occur at $\mathcal{O}(g)$ and not $\mathcal{O}(g^2)$), these terms can be adequately described in terms of Langevin dynamics, that is, in terms of momentum drag and longitudinal and transverse momentum diffusion. This is analyzed in the context of heavy quark dyamics in Ref. [35] and jet energy loss in Ref. [34], and applied to shear viscosity in Ref. [15]. As shown there, the effective collision term due to soft scatterings, cut off at a scale $\mu_\perp$, is

$$(C_{\text{diff}}^{2\leftrightarrow 2} f_1)^a(\boldsymbol{p}) = -\frac{1}{2}\frac{\partial}{\partial p^i}\,\hat{q}_a^{ij}\, f_0^a(p)(1\pm f_0^a(p))\frac{\partial f_1^a(\boldsymbol{p})}{\partial p^j} + \text{gain-terms}\,, \tag{3.14}$$

Federico Cacciamani, Andrea Celli, Marco Ciccone, and Nicola Gatti. Multi-agent coordination in adversarial environments through signal mediated strategies. *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, page 269–278, May 2021.

Murray Campbell, A Joseph Hoane Jr, and Feng-hsiung Hsu. Deep blue. *Artificial Intelligence*, 134(1-2):57–83, 2002.

Luca Carminati, Federico Cacciamani, Marco Ciccone, and Nicola Gatti. A marriage between adversarial team games and 2-player games: Enabling abstractions, no-regret learning, and subgame solving. *Proceedings of the 39th International Conference on Machine Learning*, 162: 2638–2657, July 2022.

Andrea Celli and Nicola Gatti. Computational results for extensive-form adversarial team games. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), Apr. 2018.

Gabriele Farina, Andrea Celli, Nicola Gatti, and Tuomas Sandholm. Ex ante coordination and collusion in zero-sum multi-player extensive-form games. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, page 9661–9671, 2018.

Gabriele Farina, Andrea Celli, Nicola Gatti, and Tuomas Sandholm. Connecting optimal ex-ante collusion in teams to extensive-form correlation: Faster algorithms and positive complexity results. In *Proceedings of the 38th International Conference on Machine Learning*, pages 3164–3173, 18–24 Jul 2021.

Kristoffer Arnsfelt Hansen, Thomas Dueholm Hansen, Peter Bro Miltersen, and Troels Bjerre Sørensen. Approximability and parameterized complexity of minmax values. In *Internet and Network Economics: 4th International Workshop*, pages 684–695, 2008.

Harold W Kuhn. Extensive games. *Proceedings of the National Academy of Sciences*, 36(10):570–576, 1950a.

Harold W Kuhn. A simplified two-person poker. *Contributions to the Theory of Games*, 1:97–103, 1950b.

Chanjuan Liu, Enqiang Zhu, Qiang Zhang, and Xiaopeng Wei. Exploring the effects of computational costs in extensive games via modeling and simulation. *Int J Intell Syst*, 36(8):4065–4087, 2021.

Wenjun Ma, Xudong Luo, and Yuncheng Jiang. Matrix games with missing, interval, and ambiguous lottery payoffs of pure strategy profiles and compound strategy profiles. *Int J Intell Syst*, 33(3):529–559, 2018.

Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisỳ, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337):508–513, 2017.

John Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.

Ashutosh Nayyar, Aditya Mahajan, and Demosthenis Teneketzis. Decentralized stochastic control with partial history sharing: A common information approach. *IEEE Transactions on Automatic Control*, 58(7):1644–1658, 2013.

Sheldon M Ross. Goofspiel—the game of pure strategy. *Journal of Applied Probability*, 8(3):621–625, 1971.

David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, 2017.

Finnegan Southey, Michael H. Bowling, Bryce Larson, Carmelo Piccione, Neil Burch, Darse Billings, and D. Chris Rayner. Bayes' bluff: Opponent modelling in poker. *Proceedings of the 21st Conference in Uncertainty in Artificial Intelligence*, pages 550–558, 2005.

Bernhard von Stengel and Daphne Koller. Team-maxmin equilibria. *Games and Economic Behavior*, 21(1-2):309–321, 1997.

Brian Hu Zhang and Tuomas Sandholm. Team correlated equilibria in zero-sum extensive-form games via tree decompositions. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(5):5252–5259, Jun. 2022.

Youzhi Zhang and Bo An. Computing team-maxmin equilibria in zero-sum multiplayer extensive-form games. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(02):2318–2325, Apr. 2020a.

Youzhi Zhang and Bo An. Converging to team-maxmin equilibria in zero-sum multiplayer games. *Proceedings of the 37th International Conference on Machine Learning*, pages 11033–11043, 2020b.

Youzhi Zhang, Bo An, and Jakub Černý. Computing ex ante coordinated team-maxmin equilibria in zero-sum multiplayer extensive-form games. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(6):5813–5821, May 2021.

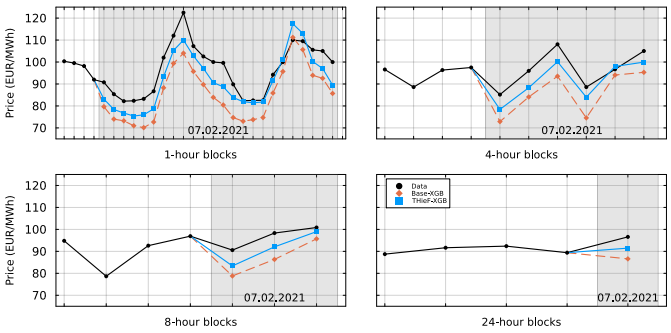

Fig. 1. Stylized example of the impact of THieF for the XBG model on one test day (07.02.2021) and four levels of the hierarchy (1H, 4H, 8H, 24H).

boosting, and a transformer. The first two are often used as benchmarks in EPF [15]–[19]. The third has been reported to perform well in numerous forecasting competitions, including M5 [20]. The fourth uses state-of-the-art transformer architecture, which is still rare in EPF [21]. It is AutoGluon's new tabular foundation model called *Mitra*, which excels on datasets with less than 5,000 samples and 100 features [22].

All four models compute the electricity price forecast $\hat{p}_{d,h}$ for day d and block h as a function of 20 features, as in [19]:

$$\hat{p}_{d,h} = f\big(p_{d-1,h}, ..., p_{d-7,h}, p^{min}_{d-1}, p^{max}_{d-1}, \hat{L}_{d,h}, \hat{W}_{d,h}, \text{API}_{d-2}, \text{TTF}_{d-2}, \text{D}^{(1)}_{d}, ..., \text{D}^{(7)}_{d}\big), \quad (2)$$

where $p_{d-i,h}$ are the lagged prices of the same block in the last seven days $i = 1, ..., 7$, p^{min}_{d-1} and p^{max}_{d-1} are the minimum and maximum hourly prices of the previous day, $\hat{L}_{d,h}$ and $\hat{W}_{d,h}$ are the day-ahead load and wind generation forecasts for the target block (source: ENTSO-E Transparency), API_{d-2} and TTF_{d-2} are the closing prices of the nearest to delivery monthly coal (API2) and yearly natural gas futures (TTF) from day $d-2$ (source: Investing.com), and $D^{(i)}_{d}$ are the weekday dummies. All models are trained independently for each block h using a 3-year window of past values; each day the window is rolled forward by one day. Overall, we consider 60 blocks: 24 × 1H (i.e., 24 one-hour blocks), 12 × 2H, 8 × 3H, 6 × 4H, 4 × 6H, 3 × 8H, 2 × 12H, and 1 × 24H (i.e., baseload).

For more robust parameter estimation, following [15], before training the models, we preprocess the inputs using the area hyperbolic sine transformation: $\text{asinh}\left((y - \hat{\mu}_y)/\hat{\sigma}_y\right)$, where $\hat{\mu}_y$ and $\hat{\sigma}_y$ are the sample mean and sample standard deviation of y estimated on the training set. More precisely, $p_{d-i,h}$ for all $i = 1, ..., 7$ and each selected block h are transformed with a common $\hat{\mu}_y$ and $\hat{\sigma}_y$ calculated for the vector $[p_{d-1092,h}, ..., p_{d-1,h}]$. The extremes p^{min}_{d-1} and p^{max}_{d-1} as well as the exogenous variables $\hat{L}_{d,h}$, $\hat{W}_{d,h}$, API_{d-2} and TTF_{d-2} are transformed independently, also using 3-year vectors of past values; weekday dummies are not transformed.

1) AutoRegression with eXogenous inputs (ARX): This expert – in the sense of [23] – model is estimated via ordinary least squares (OLS) and uses the same inputs as [19] to allow for direct comparisons; the dataset is identical.

2) Nonlinear ARX (NARX): The nonlinear counterpart of ARX approximates $f(\cdot)$ in Eq. (2) with a shallow feedforward neural network using the series-parallel architecture [24]. As in [16], the hidden layer consists of 5 neurons and uses hyperbolic tangent activation, with a linear function in the output layer. The weights are calculated in Matlab R2025a using the Levenberg-Marquadt algorithm with early stopping based on a 10% validation set. To mitigate the uncertainty of parameter estimation, the output is clipped to $[-3, 3]$ before applying the inverse transformation, i.e., the hyperbolic sine, and the final $\hat{p}_{d,h}$ is obtained by training the network 10 times for each d and h, and averaging the 10 price forecasts; no hyperparameter optimization is performed.

3) EXtreme Gradient Boosting (XGB): The third model is an ensemble of gradient-boosted decision trees (GBDT) [25]. We use eXtreme GBDT implemented in the XGBoost (v1.7) Python package [26], with MSE as the loss function for training and hyperparameter optimization. The latter is carried out 10 times at the beginning of each year (independently for each block h) using 3 years of past data and Bayesian sequential optimization [27] with early stopping for the number of trees ($\leq$1000).[1] $\hat{p}_{d,h}$ is obtained by averaging the forecasts generated using these 10 sets of hyperparameters.

4) Mitra: The final forecaster is Amazon's state-of-the-art tabular foundation model released with AutoGluon v1.4 in July 2025, based on a 12-layer 72 million-parameter transformer architecture [22]. Since it is pre-trained on purely synthetic data, evaluating its performance on historic time series does not pose data contamination issues. Although *Mitra* is not specifically tailored to time series, our forecasting task requires only one-step-ahead predictions, with Eq. (2) being equivalent to a tabular regression problem. We generate forecasts in a *zero-shot* mode, relying on a model pre-trained for regression tasks. At each timestep, the training set is fed as support examples for in-context learning. *Mitra* then generates a single prediction based on the input variables acting as a query.

IV. Empirical Results

To ensure a sound assessment of the THieF approach, we consider one of the largest power markets in Europe (EPEX-DE, Germany) and a dataset that spans 7 years (05.01.2018-31.12.2024); the same as in [19]. The first 1092 days (until 31.12.2020) are the initial training window; each day it is rolled forward by 24 hours. The remaining 4-year period (starting 01.01.2021) is a challenging test set that includes the COVID-19 pandemic (the first day of hard lockdown in Germany was 15.12.2020), the Russian invasion of Ukraine (24.02.2022) and the soaring natural gas prices (Q4 2021-Q4 2022), as well as the appearance of negative price spikes due to low demand and high renewable generation (e.g., on Sunday 02.07.2023 at 3 p.m. the price dropped to −500 EUR/MWh).

In Table I we report the mean absolute errors (MAE) and the root mean squared errors (RMSE) of the four models over the entire test period, as well as the respective gains from forecast reconciliation (in %). The latter are all significant at the 5%

[1]The search ranges are: *max_depth* $\in [2, 10]$, *learning_rate* $\in [0.0001, 1.0]$ (log scale), *subsample* $\in [0.5, 1.0]$, *min_child_weight* $\in [0, 10]$, $\gamma \in [0.0, 0.5]$, $\lambda \in [0.001, 10.0]$ (log scale), and $\alpha \in [0.001, 10.0]$ (log scale). The remaining parameters use default values.

7.1 Alignment of Accuracy with User Preferences

We first need to assess if BLEU scores and similar metrics are good proxies for human preferences. To do this we study the outcomes of our user preference survey.

When aggregating over the entire dataset, users prefer Nougat the most (with a frequency of 57.1%) followed by Marker (49.1%) and PyMuPDF (48.6%). Throughput does not necessarily translate to user preferences. PyMuPDF, for example, offered a 2133× higher throughput while experiencing a BLEU score difference of 0.5%. However, users are not indifferent to parsers, as indicated by the low win frequency of 2.1% for pypdf. As these frequencies are determined by a binary tournament of different pairings of the seven parsers present in the study, percentages do not sum to 100%. Therefore, we report normalized win rates instead.

Users are highly willing to make their preferences known, doing so 91.3% of the time and while picking "neither" only in the remaining 8.7%. Moreover, participants have a high agreement in the choices they make. Among the 405 triplets of page document and two parser output texts shown to multiple users, participants made the same choice 82.2% of the time. This high consensus rate—achieved despite scientists' diverse disciplinary backgrounds—suggests a degree of objectivity in participants' preferences, underscoring their usefulness for model refinement. Importantly, these preferences are collected only once and used offline to adapt the model's weights via DPO during post-training, so that no further human input is required when the model is deployed for parsing.

A key result of this study is that the BLEU score, while indicative of user preference, is hardly predictive. The BLEU is highly correlated with the win rate (correlation $\hat{\rho}$ = 0.47), which is statistically significant as $H_0 : \rho = 0$ is rejected with $p = 8.4^{-49}$. Yet the correlation is also far from 1, explaining only 47% of the variation in user choices. We view this result as justification that the BLEU score is a robust quality indicator of parser text output and a suitable target for LLM-finetuning, but also not completely predictive requiring the consideration of other measures of quality.

7.2 AdaParse Quality Assessment

Since AdaParse manages diverse parsers during its execution, it is important to probe the mechanism for parser selection and to gauge the improvement over using individual parsers. Because no single quality measure is completely predictive of user preferences, we consider a set of quality measures. The empirical investigation includes document coverage (as measured by the number of retrieved document pages), BLEU, ROUGE, and character-accuracy rate (CAR). It also includes two metrics we devised from the user preference study: *win rate* (WR) which measures how often a parser was selected over the others for a given document and *accepted tokens* (AT) that tracks the relative frequency of tokens that exceed a critical BLEU threshold.

To evaluate AdaParse, we run it on a held-out test set of 1000 digitally born PDFs that were not used during the training. We report three rounds of metrics: the first with no changes to any layer of the PDFs, the second with augmentations applied only to the image layer, and the third with alterations applied only to the text layer.

We show in Table 1 the default quality on the test set. Marker has the highest coverage rate, but does not have the highest quality according to any other metric. Nougat has the highest win rate between parsers by a slim margin. AdaParse even with the requirement to allocate no more than 5% of the documents to its high-quality parser (Nougat), produces the best BLEU and ROUGE scores, and the second-best CAR. Additionally, AdaParse has the highest percentage of accepted tokens based on the user preference data at 76.9%. AdaParse can achieve better performance than any of its constituent parsers by delegating to the method that is most suitable for an individual document. While a parser like Nougat may perform best on average, it is not the best parser for each document allowing AdaParse to exceed it if it accurately infers a better parser-document matching.

The performance of AdaParse is based on the capability in predicting the BLEU score of PyMuPDF and Nougat-parsed text, with an $R^2 = 40.0\%$ and $R^2 = 46.5\%$, respectively. This is largely based on parameter-efficient finetuning through low-rank adaptation (LoRA) (Hu et al., 2021) and DPO on its weights in decoder mode. DPO post-training has been shown to improve performance in related downstream prediction tasks using relatively little preference data, even in high-performance computing (HPC) applications (Dharuman et al., 2024).

Table 1. Accuracy on born-digital PDFs: Document- (coverage rate), word- (BLEU, ROUGE), and character-level (CAR) accuracies. CAR = Character accuracy rate. WR = Win rate. AT = Accepted tokens. All %.

Parser	Coverage	BLEU	ROUGE	CAR	WR	AT
Marker	**96.7**	47.5	64.2	59.6	26.6	73.3
Nougat	93.0	48.1	66.5	65.8	**27.9**	69.8
PyMuPDF	91.3	51.9	67.3	67.0	24.4	76.7
pypdf	92.0	43.6	58.7	32.3	2.4	72.4
GROBID	81.0	26.5	52.4	54.8	–	20.6
Tesseract	91.3	48.8	64.2	**67.8**	18.7	72.5
AdaParse	91.5	**52.1**	**67.6**	67.1	25.5	**76.9**

Additionally, we test parsing performance under simulated image degradation to mimic low-quality scans. Low-quality scans are common in older academic and book datasets. We emulate this quality degradation with random rotations,

Next we give a simple proof of [21, Theorem 2.4]. Indeed, we prove a more general fact, which gives a partial answer to Question 4.2. First, we give some concepts.

Recall that a space is called *collectionwise Hausdorff* if each closed discrete subset of X can be separated by a discrete family of open neighborhoods. A subset D of a space X is said to be *transfinite sequentially dense* if each point $x \in X$ is the unique accumulation point of a transfinite sequence of points in D.

Theorem 4.3. *Let G be a hereditarily collectionwise Hausdorff topological group with a (closed) suitable set S and H be a transfinite sequentially dense subgroup of G. Then H has a (closed) suitable set.*

Proof. Since G is hereditarily collectionwise Hausdorff, it follows that $G \setminus \{e\}$ is collectionwise Hausdorff, hence, for each $x \in S$, there exists an open neighborhood $U_x \subset G \setminus \{e\}$ of x such that the family $\{U_x : x \in S\}$ is discrete in $G \setminus \{e\}$. Now put $S_x = \{x\}$ whenever $x \in H \cap S$. If $x \in S \setminus H$, let S_x be a transfinite sequence of $U_x \cap H$ such that x is the unique accumulation point of S_x in G. Put $S^* = \bigcup_{x \in S} S_x$. Then S^* is closed and discrete in $G \setminus \{e\}$ since the family $\{S_x : x \in S\}$ is a locally finite family and each S_x is discrete and closed in $G \setminus \{e\}$. It is obvious that $S \subset \overline{\langle S^* \rangle}$ and $\langle S \rangle$ is dense in G, it follows that $\langle S^* \rangle$ is dense in H. Therefore, S^* is a suitable set for H.

Clearly, if S is closed in G, then S^* is closed in H. □

Proposition 4.4. Each dense subgroup of a linearly orderable topological group G is a transfinite sequentially dense subgroup.

Proof. Let H be a dense subgroup of G. For each $x \in G$, define a transfinite sequence in H as follows. If $x \in H$, then put $S_x = \{x\}$; otherwise pick a transfinite sequence $S_x = \{z(x, \alpha) : \alpha < \tau\}$, where each $z(x, \alpha) \in H \cap xU_\alpha$ with $\alpha < \tau$. Clearly, x is the unique accumulation point of S_x in H. Therefore, H is a transfinite sequentially dense subgroup. □

Corollary 4.5. [21, Theorem 2.4] *Each dense subgroup of a linearly orderable topological group G with a (closed) suitable set has a (closed) suitable set.*

Proof. Since G is hereditarily paracompact by [20, Theorem 8], it follows G is hereditarily collectionwise Hausdorff. Now the conclusion is immediate by applying Theorem 4.3 and Proposition 4.4. □

Moreover, we can prove the following theorem.

Theorem 4.6. *Let G be a linearly orderable topological group. Then each subgroup of G has a suitable set if and only if each closed linearly orderable subgroup of G has a suitable set.*

Proof. The sufficiency is obvious, hence is suffices to prove the necessity. Let H be a subgroup of G. If $\overline{H}$ is metrizable, then $\overline{H}$ and H all have suitable sets by [4, Theorem 6.6]. If $\overline{H}$ is not metrizable, then G is not metrizable, hence it follows from [20] that G contains a decreasing chain of open subgroups $\{U_\alpha : \alpha \in I\}$ which forms a local base at the identity. Then $\{U_\alpha \cap \overline{H} : \alpha \in I\}$ is a decreasing chain of open subgroups in $\overline{H}$ which form a local base at the identity. Hence $\overline{H}$ is a linearly orderable topological subgroup. By our assumption, $\overline{H}$ has a suitable set, then H has a suitable set by [21, Theorem 2.4]. □

Lemma 2.4 (Exercise 1.5.H in [2])**.** *In the Niemytzki plane every closed set is a G_δ-set.*

Solution. Let F be closed in N and divide it onto F_1, its intersection with the x-axis, and $F_2 = F \setminus F_1$.

The set F_2 is closed in the open subspace $\{\langle x, y\rangle : y > 0\}$, which carries its Euclidean topology. Hence F_2 is a G_δ-set in that subspace and hence in N.

Let $A = \{x : \langle x, 0\rangle \in F\}$; then $F_1 = \bigcap_{k\in\mathbb{N}} \bigcup_{x\in A} U(x, k)$.

So F is the union of two G_δ-sets. □

Proposition 2.5. *In the space $B \cup \{\infty\}$ every open set is an F_σ-set.*

Proof. Let O be open in $B\cup\{\infty\}$, then $\pi^{\leftarrow}[O]$ is open in $(N \times \mathbb{N}) \cup \{\infty\}$ and hence an F_σ-set: By Lemma 2.4 its intersection with every sheet $N \times \{n\}$ is an F_σ-set, and, if $\infty \in \pi^{\leftarrow}[O]$ then $\{\infty\}$ is a contributing closed set as well.

Say $\pi^{\leftarrow}[O] = \bigcup_{k\in\mathbb{N}} F_k$, then $O = \bigcup_{k\in\mathbb{N}} \pi[F_k]$ is an F_σ-set by Lemma 2.3. □

Remark 2.6. This example is not quite the one that we want because although $\{\infty\}$ is a G_δ-set it is not a zero-set. For assume $f : B \cup \{\infty\} \to [0, 1]$ is continuous and such that $f(\infty) = 0$ and $f\big(\pi(x, 0, 0)\big) > 0$ for all $x \in \mathbb{R}$.

Then by the method proof of Lemma 1.4 we find, natural numbers k and m, an open interval J and a dense subset A of J such that for all $a \in A$ we have $f\big(\pi(a, 0, 0)\big) \geqslant 2^{-k}$ and $f\big(\pi(x, y, 0)\big) > \frac{1}{2} f\big(\pi(a, 0, 0)\big)$ for all $\langle x, y\rangle \in U'(a, m)$. As before this implies that $f\big(\pi(x, 0, 0)\big) \geqslant 2^{-k-1}$ for all $x \in J$ and that there is a single G_δ-set G that is dense in J and such that $f\big(\pi(x, 0, n)\big) \geqslant 2^{-k-1}$ for all $x \in G$ and all $n \in \mathbb{N}$. This would contradict the continuity of f at ∞.

3. A whole Library

We redo our construction.

We start with the product $N \times \mathbb{N} \times \mathbb{N}$ and add a point ∞. The basic neighbourhoods of ∞ are the sets

$$V_k = \bigcup_{m\geqslant k} \bigcup_{n\geqslant k} N \times \{\langle m, n\rangle\}$$

We apply the Jones construction to every column in this matrix, so that we get a book B_m for every m.

The resulting quotient space, which we call L is the union of these books and the point ∞ (yes, L for Library). We let π denote the quotient map.

Proposition 3.1. *The space L is regular, but not completely regular.*

Proof. The space L is completely regular at all points except possibly ∞.

The space is regular at ∞ by the same argument as for B: for every k the closed neighbourhood V_{k+1} is contained in the interior of V_k.

To see that L is not completely regular at ∞ let $f : L \to [0, 1]$ be continuous and such that $f(x, 0, m, 0) = 0$ for all $x \in \mathbb{R}$ and all $m \in \mathbb{N}$. By Remark 2.2 applied to every column we obtain a single dense G_δ-set G in $\mathbb{R}$ such that for all $x \in G$ and all $m, n \in \mathbb{N}$ we have $f(\pi(x, 0, m, n)) = 0$. By continuity this implies that $f(\infty) = 0$. □

Proposition 3.2. *Every one-point set in L is a zero-set.*

Proof. We only need to check this for $\{\infty\}$.

But this is straightforward: let $f : L \to [0, 1]$ take on the value 2^{-m} on book B_m, and let $f(\infty) = 0$. □

To summarize in a theorem.

Any vertex in G is either in S_m, or L. This means that the only bad edges that are not counted by T_m must be incident to a vertex in S_m and a vertex in L. The edges that are incident to a vertex in S_m and $G \setminus S_m = L$ are entirely described by E_m. Thus, the number of bad edges that exist between a vertex in S_m and a vertex in L is given by $\sum E_m$, which is $O\left(\varepsilon^2 m^2\right)$. This is dominated by $T_m = O\left(\varepsilon n^2\right) + T_0$, so the asymptotic growth rate of the number of bad edges in the graph is given solely by that of T_m.

It remains to show that T_0 cannot grow faster than $O\left(\varepsilon^2 n^2 |S_0|\right) \geq \Theta(\varepsilon n^2)$. By construction, every vertex in S_0 must begin incident to at least one bad edge, so the starting number of bad edges is at least $T_0/2 = \Theta(T_0)$. By Lemma 6, we start with $O\left(\varepsilon^{-2}\right)$ bad edges, so $T_0 = O\left(\varepsilon^{-2}\right)$. This is also $O(\varepsilon n^2)$ because $\varepsilon = \omega\left(n^{-1/2}\right) \geq \Theta\left(n^{-2/3}\right)$. □

Lemma 11. *The quantity $\sum E_0$ grows on the order of $O\left(m|S_0|\right)$.*

Proof. Since there are only n vertices in the graph, every vertex in S_0 can have at most n edges. Recall that $\sum E_0$ counts the number of edges incident to a vertex in S_0 and a vertex in $G \setminus S_0$. By Lemma 8, we have

$$\sum E_0 = O\left(n|S_0|\right)$$
$$\sum E_0 = O\left(m|S_0|\right).$$

□

We will now consider the time reversed process of playing moves on the E_k.

Definition 12. *A **reversed move** on a sequence E_k consists of the following steps:*

- *Select a list of positive elements $\{e_1, e_2, \ldots, e_\alpha\}$ from E_k*
- *Subtract one from each of the e_j for $1 \leq j \leq \alpha$*
- *Calculate $z = \lceil \alpha/(1+\epsilon) \rceil$*
- *Create E_{k-1} by inserting z into E_k as a new element*

Let $\sum E_k^2$ denote the sum of the squares of the elements of E_k. In order to investigate the asymptotic growth rate of $\sum E_m$, we wish to find the largest starting value of $\sum E_m$ for which it is possible to play m reversed moves to get to a desired value of $\sum E_0$.

Our plan is to bound the behavior of $\sum E_k$ and $\sum E_k^2$ as reversed moves are played. Intuitively, playing a reversed move on many elements of E_k would result in a large decrease in from $\sum E_k$ to $\sum E_{k-1}$. However, playing such a move would also result in the creation of a new, very large element in E_{k-1}. Since we can only ever subtract one from each element per move, having the sum $\sum E_{k-1}$ concentrated in a few large elements will make decreasing $\sum E$ more difficult. We consider $\sum E_k^2$ as a way to measure how "spread out" the sequence E_k is, and this particular quantity is preferable over other measures because of how the change from $\sum E_k^2$ to $\sum E_{k-1}^2$ can be cleanly bounded by an expression in terms of $\sum E_k$.

Consider some arbitrary sequence E_k. Let α denote the number of elements of E that we play the current reversed game move on, and let these elements be denoted $e_1, e_2, \ldots, e_\alpha$. By the rules of the reversed game, playing such a move will create a new element of size $z = \lceil \alpha/(1+\varepsilon) \rceil$. It is easy to see that the effect of such a move on $\sum E_k$ is

$$\sum E_{k-1} = \sum E_k - \alpha + z.$$

Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59:1551–1580.

Kristensen, D. and Rahbek, A. (2010) Likelihood-based inference for cointegration with nonlinear error-correction. *Journal of Econometrics,* 158:78–94.

Lee, S. and Hansen, B. (1994) Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator. *Econometric Theory,* 10:29–52.

Li, Y., Mykland, P.A., Renault, E., Zhang, L., and Zheng, X. (2014) Realized volatility when sampling times are possibly endogenous. *Econometric Theory*, 30:580–605.

Pacurar, M. (2008) Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. *Journal of Economic Surveys*, 22:711—751.

Phillips, P.C.B. (1991) Optimal inference in cointegrated systems, *Econometrica*, 59:283–306.

Sin, C. (2014) QMLE of a standard exponential ACD model: Asymptotic distribution and residual correlation. *Annals of Financial Economics*, Vol. 09: Issue 2.

Sweeting, (1980) Uniform asymptotic normality of the maximum likelihood estimator. *Annals of Statistics*, 8:1375–1381.

Zolotarev, V.M. (1986) *One-Dimensional Stable Distributions.* American Mathematical Society, Providence, Rhode Island.

Appendix

A Proofs

Proof of Lemma 2.1

Observe that the ACD equation (1.2) can be formulated as a stochastic recurrence equation (SRE):

$$x_i = A_i\, x_{i-1} + B_i, \qquad i \in \mathbb{Z}\,, \tag{A.1}$$

with a sequence $(A_i, B_i) = (\omega\,, \alpha)\varepsilon_i$, $i \in \mathbb{Z}$, of random vectors with i.i.d. positive $\{\varepsilon_i\}$. Using the SRE representation, it follows that $\{x_i\}$ is strictly stationary geometrically ergodic if and only if $\mathbb{E}[\ln(\alpha\varepsilon)] < 0$ by Theorem 2.1.3 and Proposition 2.2.4 in Buraczewski, Damek and Mikosch (2016), BDM henceforth. The power-law tail behavior $\mathbb{P}(x > z) \sim c_\kappa\, z^{-\kappa}$ follows from Theorem 2.4.4 in BDM, and it holds that c_κ is given by

$$c_\kappa = \frac{\mathbb{E}[(\omega + (\alpha\,\varepsilon)\,x)^\kappa - ((\alpha\,\varepsilon)x)^\kappa]}{\kappa\,\mathbb{E}[(\alpha\varepsilon)^\kappa\,\ln(\alpha\varepsilon)]}\,. \tag{A.2}$$

□

Proof of Lemma 2.2

The results hold by Lemma 2.1, noting that for the exponential case,

$$1 = \mathbb{E}[(\alpha\varepsilon)^\kappa] = \alpha^\kappa \int_0^\infty x^\kappa \exp(-x)\, dx = \alpha^\kappa\, \Gamma(\kappa + 1)\,. \tag{A.3}$$

□

compactification of a uncountable discrete space. Hence $A(X)$ contains a non-trivial convergent sequence. However, since X contains no non-trivial convergent sequences and $A(X)$ is snf-countable, it follows from [12, Theorem 3.1] that $A(X)$ contains no non-trivial convergent sequences, which is a contradiction. Therefore, $A(X)$ does not have a suitable set. □

It follows from [12, Theorem 3.3] that, for a topological group G, the free Abelian topological group $A(G)$ is snf-countable if and only if $A(G)$ contains no non-trivial convergent sequences. Therefore, by [12, Theorems 3.1 and 3.3], the following corollary gives a partial answer to Problem 1.1.

Corollary 3.3. *Let X be a non-separable k_ω-space without non-trivial convergent sequences. If X is homeomorphicto atopological group, then $A(X)$ does not have a suitable set.*

However, the following question is unknown for us.

Question 3.4. *Let G be a non-separable k_ω-topological group without non-trivial convergent sequences. Does $A(G)^n$ have a suitable set for some $n \geq 2$?*

Recall that a subset Y of a space X is called *C^*-embedded* in X if any bounded continuous real-valued function on Y extends to a continuous real-valued function on X. If X is a compact space in which every countable discrete subset is C^*-embedded, then it follows from [22, Theorem 2.1.4] that X does not contain any non-trivial convergent sequences. Therefore, the following result is a partial answer to Question 3.4.

Theorem 3.5. *Let X be a non-separable compact space in which every countable discrete subset is C^*-embedded. Then $A(X)^n$ does not have a suitable set for any $n \in \mathbb{N}$.*

In Theorem 3.2, the assumption of "non-separable" is necessary since each free topological group over a separable space has a closed suitable set, see [4]. If X is a Lindelöf space and $F(X)$ or $A(X)$ has a closed suitable set, does X is separable? The following Theorem 3.7 gives a partial answer to this question. First, we give a lemma.

Lemma 3.6. *Let G be a Lindelöf group. Then*

(a) If $G \in \mathscr{S}_c$, then G is separable.

(b) If G is separable and non-compact then $G \in \mathscr{S}_c$.

(c) If $G \in \mathscr{S}_{cg}$ if and only if G is countable.

Proof. (a) Assume S is a closed suitable set for G. Then S is Lindelöf since S is closed in Lindelöf group G. Hence S is countable, which shows that $\langle S \rangle$ is countable. Thus G is separable.

(b) If G is separable and non-compact, then G is not pseudocompact since a regular Lindelöf pseudocompact space is compact. Now it follows from [5, Corollary 3.9] that $G \in \mathscr{S}_c$.

(c) By the proof of (a), if $G \in \mathscr{S}_{cg}$ then G is countable. The converse follows from [4, Theorem 2.2]. □

Theorem 3.7. *Let X be a regular space such that X^n is Lindelöf for each $n \in \mathbb{N}$. Then the following conditions are equivalent:*

Theorem 3.3. *The space L is regular, but not completely regular. Every one-point set is a zero-set of L, and every closed set is a G_δ-set.*

Proof. The only thing that still needs to be verified is the last statement.

But that is done almost verbatim as for $B\cup\{\infty\}$. Adapt the proofs of Lemma 2.3 and Proposition 2.5 to the present situation. □

Remark 3.4. This example also has an application in lattice theory. The lattice $\mathrm{O}(X)$ of open sets in a topological space is readily seen to be complete: the union and the interior of the intersection serve as the supremum and infimum, respectively, of a family of open sets. An important sublattice of $\mathrm{O}(X)$ is that of the cozero-sets: $\mathrm{Coz}(X)$.

It turns out that $\mathrm{O}(X)$ is the Dedekind-MacNeille completion of $\mathrm{Coz}(X)$ if and only if X is completely regular and each singleton set is a zero-set.

As shown by the present example this result is sharp in that complete regularity cannot be replaced by regularity.

Thanks to Guram Bezhanishvili for supplying this information.

References

[1] P. Alexandroff and H. Hopf, *Topologie. I*, Die Grundlehren der mathematischen Wissenschaften, vol. Band 45, Springer-Verlag, Berlin-New York, 1974. Berichtigter Reprint. MR0345087

[2] Ryszard Engelking, *General topology*, 2nd ed., Sigma Series in Pure Mathematics, vol. 6, Heldermann Verlag, Berlin, 1989. Translated from the Polish by the author. MR1039321

[3] F. Burton Jones, *Hereditarily separable, non-completely regular spaces*, Topology Conference. Blacksburg, VA, USA, 1973 (Dickman Jr Raymond F. and Peter Fletcher, eds.), Lecture Notes in Math., vol. 375, Springer, Berlin-New York, 1974, pp. 149–152. MR0413044

[4] A. Mysior, *A regular space which is not completely regular*, Proc. Amer. Math. Soc. **81** (1981), no. 4, 652–653, DOI 10.2307/2044178. MR0601748

[5] Lynn Arthur Steen and J. Arthur Seebach Jr., *Counterexamples in topology*, 2nd ed., Springer-Verlag, New York-Heidelberg, 1978. MR507446

[6] A. Tychonoff, *Über die topologische Erweiterung von Räumen*, Math. Ann. **102** (1930), no. 1, 544–561, DOI 10.1007/BF01782364 (German). MR1512595

Faculty EEMCS, TU Delft, Postbus 5031, 2600 GA Delft, the Netherlands
Email address: k.p.hart@tudelft.nl
URL: https://fa.ewi.tudelft.nl/~hart

language modeling with pathways. *J. Mach. Learn. Res.*, April 2022.

Cohan, A., Feldman, S., Beltagy, I., Downey, D., and Weld, D. S. Specter: Document-level representation learning using citation-informed transformers. *arXiv preprint arXiv:2004.07180*, 2020.

Corrêa, A. S. and Zander, P.-O. Unleashing tabular content to open data: A survey on PDF table extraction methods and tools. In *18th Annual International Conference on Digital Government Research*, pp. 54–63, 2017.

Coulon, R., Toro, F. G., and Michotte, C. Machine-readable data and metadata of international key comparisons in radionuclide metrology. *Measurement Science and Technology*, 34(7):074009, 2023.

Dharuman, G., Hippe, K., Brace, A., Foreman, S., Hatanpää, V., Sastry, V. K., Zheng, H., Ward, L., Muralidharan, S., Vasan, A., et al. Mprot-dpo: Breaking the exaflops barrier for multimodal protein design workflows with direct preference optimization. In *SC24: International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–13. IEEE, 2024.

Du, Y., Li, C., Guo, R., Yin, X., Liu, W., Zhou, J., Bai, Y., Yu, Z., Yang, Y., Dang, Q., et al. PP-OCR: A practical ultra lightweight OCR system. *arXiv preprint arXiv:2009.09941*, 2020.

Elazar, Y., Bhagia, A., Magnusson, I., Ravichander, A., Schwenk, D., Suhr, A., Walsh, P., Groeneveld, D., Soldaini, L., Singh, S., et al. What's in my big data? *arXiv preprint arXiv:2310.20707*, 2023.

Fenniak, M., Stamy, M., pubpub zz, Thoma, M., Peveler, M., exiledkingcc, and pypdf Contributors. The pypdf library, 2024. URL https://pypi.org/project/pypdf/.

Freund, Y., Schapire, R. E., et al. Experiments with a new boosting algorithm. In *icml*, volume 96, pp. 148–156. Citeseer, 1996.

Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N., Presser, S., and Leahy, C. The Pile: An 800GB dataset of diverse text for language modeling, December 2020. URL http://arxiv.org/abs/2101.00027. arXiv:2101.00027 [cs].

Graham, Y. Re-evaluating automatic summarization with bleu and 192 shades of rouge. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pp. 128–137, 2015.

Groleau, A., Chee, K. W., Larson, S., Maini, S., and Boarman, J. Augraphy: A data augmentation library for document images. In *International Conference on Document Analysis and Recognition*, pp. 384–401. Springer, 2023.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

Jaume, G., Ekenel, H. K., and Thiran, J.-P. FUNSD: A dataset for form understanding in noisy scanned documents. In *International Conference on Document Analysis and Recognition Workshops*, volume 2, pp. 1–6. IEEE, 2019.

Jimeno Yepes, A., Zhong, P., and Burdick, D. Icdar 2021 competition on scientific literature parsing. In *16th International Conference on Document Analysis and Recognition*, pp. 605–617. Springer, 2021.

Kim, G., Hong, T., Yim, M., Nam, J., Park, J., Yim, J., Hwang, W., Yun, S., Han, D., and Park, S. OCR-free document understanding transformer. In *European Conference on Computer Vision*, pp. 498–517. Springer, 2022.

Kuribayashi, M. and Wong, K. StealthPDF: Data hiding method for PDF file with no visual degradation. *Journal of Information Security and Applications*, 61:102875, 2021.

Lab, D. Marker - data lab. https://www.datalab.to/marker, 2024. Accessed: 2024-10-29.

Levenshtein, V. Binary codes capable of correcting deletions, insertions, and reversals. *Proceedings of the Soviet physics doklady*, 1966.

Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z., and Wei, F. TrOCR: Transformer-based optical character recognition with pre-trained models. In *AAAI Conference on Artificial Intelligence*, volume 37, pp. 13094–13102, 2023.

Li, S., Huang, J., Zhuang, J., Shi, Y., Cai, X., Xu, M., Wang, X., Zhang, L., Ke, G., and Cai, H. Scilitllm: How to adapt llms for scientific literature understanding. *arXiv preprint arXiv:2408.15545*, 2024.

Lin, C.-Y. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pp. 74–81, 2004.

Lo, K., Wang, L. L., Neumann, M., Kinney, R., and Weld, D. S. S2ORC: The Semantic Scholar open research corpus. *arXiv preprint arXiv:1911.02782*, 2019.

to alter the overlap.

We should also note that we should really separate the moduli of the original genus-2 surface defining the half genus-2 state in two: the modulus of the handle and the modulus of the cylinder. Given a fixed modulus of the cylinder, there is a unique choice of β for which this overlap gives one. In turn, this uniquely fixes $\tilde{\beta}$. On the other hand, the modulus of the handles can be tuned at will here, without affecting the choice of β or the leading order value of the overlap. This is the CFT incarnation of what we found in the bulk: there are many different choices of wormhole states that all have large overlap with BTZ.

4.2 Comments on the $U(1)$ symmetry.

There is one aspect of the wormhole state that is worth discussing in more detail and is related to symmetry. The thermofield double state is rotationally invariant, this means that it is annihilated by the sum of the spin operators[9]

$$(J_L + J_R)\,|\text{TFD}\rangle = 0\,. \tag{4.11}$$

This is also manifest in the preparation of the state, which is done on a cylinder that is rotationally invariant. This symmetry extends into the bulk, giving a killing vector of the spacetime. For the wormhole state, there is no such symmetry in its preparation. This means we expect

$$(J_L + J_R)\,|\psi_{\text{WH}}\rangle \neq 0\,. \tag{4.12}$$

When we stare at the definition of the wormhole state, we see from (4.5) that the CPT nature of the left-right pairing guarantees that at the level of the primaries, the operator $(J_L + J_R)$ annihilates the state. But this happens only at the level of the primaries. For the descendants, we will get non-zero contributions confirming indeed that $(J_L + J_R)$ does not annihilate the state.

There is a version of the this happening in the bulk: while the preparation of the state does not preserve a $U(1)$ rotation symmetry, the initial data actually does preserve the symmetry. This suggests that the expectation value of $(J_L + J_R)$ vanishes in the semi-classical limit, and that its variance is small, and only visible at the quantum level. This is perfectly compatible with the state being annihilated by $(J_L + J_R)$ at the level of primaries, but not at the level of descendants. It would be interesting to see if there were a deeper connection between initial data, symmetries, and primary vs descendants in the dual CFT.

4.3 A Lorentzian version of the factorization puzzle

Before concluding, we will discuss a Lorentzian incarnation of the factorization puzzle [1]. The factorization puzzle was originally formulated in the Euclidean context: the

[9]Note the sign difference between the sum of spin operators and the operator that annihilates the thermofield double state $H_L - H_R$. This is due to the CPT operation that takes place between the left and right states.

3.1 $D = 5$ Spinning Particle

In the description of the spinning particle in curved space-time, we can make the minimal coupling $\eta_{AB} \to g_{AB}(x)$ in the action related to the Lagrangian, equation (2.1), with the Chern-Simons term. Then, the action for a $N = 2$ spinning particle is given by:

$$\begin{aligned} S = \int \Bigg[& \frac{E^{-1}}{2} g_{AB} \dot{x}^A \dot{x}^B - \frac{E}{2} M_0^2 - \frac{i}{2} g_{AB} \psi_k^A \frac{D\psi_k^B}{D\tau} - \frac{i}{2} \psi_k^5 \dot{\psi}_k^5 \\ & - \frac{i}{2} \lambda^k (E^{-1} g_{AB} \psi_k^A \dot{x}^B + M_0 \psi_k^5) - \frac{1}{8} E^{-1} g_{AB} (\lambda_k \psi_k^A)(\lambda_k \psi_k^B) \\ & + f \left(-i \frac{1}{2} \epsilon_{kl} (g_{AB} \psi_k^A \psi_l^B + \psi_k^5 \psi_l^5) - \left(q - \frac{1}{2} D \right) \right) \Bigg] d\tau, \end{aligned} \tag{3.1}$$

where

$$\frac{D\psi_k^A}{D\tau} = \dot{\psi}_k^A + \Gamma_{BC}^A \psi_k^B \dot{x}^C. \tag{3.2}$$

is the covariant derivative of the Grassmann variable ψ_k^A [26] and g_{AB} is, initially, a general metric in a 5-dimensional spacetime. Then, we obtain from equation (3.1) the following equations of motion:

$$\frac{\delta S}{\delta E} \to g_{AB} \left[\dot{x}^A \dot{x}^B - i \lambda_k \dot{x}^A \psi^B - \frac{1}{4} (\lambda_k \psi_k^A)(\lambda_k \psi_k^B) \right] + E^2 M_0^2 = 0, \tag{3.3}$$

$$\frac{\delta S}{\delta \lambda_k} \to g_{AB} \left[\dot{x}^A \psi^B - \frac{i}{2} (\lambda_k \psi_k^A) \psi_k^B \right] + E M_0 \psi_k^5 = 0, \tag{3.4}$$

$$\frac{\delta S}{\delta f} \to i \frac{1}{2} \epsilon_{kl} (g_{AB} \psi_k^A \psi_l^B + \psi_k^5 \psi_l^5) + \left(q - \frac{1}{2} D \right) = 0 \tag{3.5}$$

$$\frac{\delta S}{\delta \psi_k^B} \to \frac{D\psi_k^B}{D\tau} - \frac{1}{2} E^{-1} \lambda_k \dot{x}^B + \frac{i}{4} E^{-1} \lambda_k (\lambda_k \psi_k^B) + f \epsilon_{lk} \psi_l^B, \tag{3.6}$$

$$\frac{\delta S}{\delta \psi_k^5} \to \dot{\psi}_k^5 - 2 \psi_l^5 V_{kl} - \frac{1}{2} \lambda_k M_0 = 0, \tag{3.7}$$

$$\begin{aligned} \frac{\delta S}{\delta x^A} \to & \frac{D}{D\tau} \left[E^{-1} g_{AF} \dot{x}^F \right] - \frac{i}{2} \frac{d}{d\tau} \left[\lambda_k E^{-1} g_{AF} \psi_k^F \right] \\ & + \frac{i}{2} R_{ASQR} \psi_k^Q \psi_k^R \dot{x}^S = 0. \end{aligned} \tag{3.8}$$

We can note that E, λ_k and f have no dynamics. Therefore, they can be gauged away in the theory. In opposition, the equations (3.6)-(3.7) and (3.8) contain the dynamics of the spinning particle. From now on, we will fix the gauges and study the confinement of the spinning particle.

Let us consider the metric g_{AB} in the form of a generic braneworld Randall-Sundrum background with the metric given by

$$ds^2 = e^{2A(y)} \eta_{\mu\nu} dx^\mu dx^\nu + dy^2, \tag{3.9}$$

where, $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$ and $A(y)$ is the warp factor. If we choose the gauge conditions on equations (3.3)-(3.5) as $E = e^{2A(y)}$, $\lambda_k = 0$ and $f = 0$, respectively, we obtain the

Acknowledgements

The author is supported in part by NSF grant PHY-2310588.

References

[1] V. Braun, Y. H. He and B. A. Ovrut, "Yukawa couplings in heterotic standard models," JHEP **0604**, 019 (2006) [hep-th/0601204].

[2] V. Bouchard, M. Cvetic and R. Donagi, "Tri-linear couplings in an heterotic minimal supersymmetric standard model," Nucl. Phys. B **745**, 62 (2006) [arXiv:hep-th/0602096].

[3] L. B. Anderson, J. Gray and B. Ovrut, "Yukawa Textures From Heterotic Stability Walls," JHEP **1005**, 086 (2010) [arXiv:1001.2317 [hep-th]].

[4] E. I. Buchbinder, A. Constantin and A. Lukas, "Non-generic Couplings in Supersymmetric Standard Models," Phys. Lett. B **748**, 251 (2015) [arXiv:1409.2412 [hep-th]].

[5] S. Blesneag, E. I. Buchbinder, P. Candelas and A. Lukas, "Holomorphic Yukawa Couplings in Heterotic String Theory," JHEP **01**, 152 (2016) [arXiv:1512.05322 [hep-th]].

[6] S. Blesneag, E. I. Buchbinder and A. Lukas, "Holomorphic Yukawa Couplings for Complete Intersection Calabi-Yau Manifolds," JHEP **01**, 119 (2017) [arXiv:1607.03461 [hep-th]].

[7] J. Gray and J. Wang, "Jumping Spectra and Vanishing Couplings in Heterotic Line Bundle Standard Models," JHEP **11**, 073 (2019) [arXiv:1906.09373 [hep-th]].

[8] L. B. Anderson, J. Gray, M. Larfors, M. Magill and R. Schneider, "Generalized Vanishing Theorems for Yukawa Couplings in Heterotic Compactifications," JHEP **05**, 085 (2021) [arXiv:2103.10454 [hep-th]].

[9] L. B. Anderson, J. Gray, M. Larfors and M. Magill, "Vanishing Yukawa Couplings and the Geometry of String Theory Models," [arXiv:2201.10357 [hep-th]].

[10] J. Gray, "Vanishing conditions for higher order couplings in heterotic theories," Phys. Rev. D **110**, no.8, 086007 (2024) [arXiv:2406.19191 [hep-th]].

[11] T. Kobayashi, S. L. Parameswaran, S. Ramos-Sanchez and I. Zavala, "Revisiting Coupling Selection Rules in Heterotic Orbifold Models," JHEP **05**, 008 (2012) [erratum: JHEP **12**, 049 (2012)] [arXiv:1107.2137 [hep-th]].

[12] N. G. Cabo Bizet, T. Kobayashi, D. K. Mayorga Pena, S. L. Parameswaran, M. Schmitz and I. Zavala, "R-charge Conservation and More in Factorizable and Non-Factorizable Orbifolds," JHEP **05**, 076 (2013) [arXiv:1301.2322 [hep-th]].

[13] S. L. Parameswaran and I. Zavala, "Worldsheet instantons and coupling selection rules in heterotic orbifolds," Phys. Part. Nucl. Lett. **11**, no.7, 981-983 (2014) [arXiv:1401.6162 [hep-th]].

2. Tensor Harmonic Spectrum on S^3

One can expand TT–tensors in tensor spherical harmonics $h_{ab}^{(\ell)}$ labeled by $\ell = 0, 1, 2, \ldots$, which satisfy

$$\Delta_L \, h_{ab}^{(\ell)} = \mu_\ell \, h_{ab}^{(\ell)}, \tag{C7}$$

$$\mu_\ell = \frac{\ell(\ell+2) - 2}{R^2}. \tag{C8}$$

Inserting into the linearized equation gives the eigenvalue condition

$$\begin{aligned} \mu_\ell = 2nK \implies \ell(\ell+2) - 2 = 6, \\ \implies \ell = 2 \quad (\ell = -4 \text{ discarded}). \end{aligned} \tag{C9}$$

Hence *the only* nontrivial solution of $(\Delta_L - 2nK)h_{ab} = 0$ in TT gauge on S^3 is the quadrupole mode $\ell = 2$.

3. Exclusion of $\ell = 0, 1$ Modes

- $\ell = 0$ (monopole): A constant rescaling

$$h_{ab} \propto g_{ab} \tag{C10}$$

changes the volume rather than shape; in TT gauge $h^a{}_a = 0$ forbids such a trace mode.

- $\ell = 1$ (dipole): These correspond to infinitesimal diffeomorphisms (Killing vectors) on S^3,

$$h_{ab} = \nabla_a \xi_b + \nabla_b \xi_a, \tag{C11}$$

which can be entirely removed by a coordinate redefinition. In TT gauge, one requires $\nabla^a h_{ab} = 0$, and one finds no non-gauge $\ell = 1$ TT tensors.

Therefore, when restricting to *physical*, source-free metric perturbations of the round S^3, only the $\ell = 2$ harmonics survive, justifying the truncation to the quadrupole sector in the main text.

References

[1] R. R. Metsaev and A. A. Tseytlin, "Type IIB superstring action in AdS(5) x S(5) background," Nucl. Phys. B **533**, 109 (1998) [arXiv:hep-th/9805028]. "Superstring action in AdS(5) x S(5): kappa-symmetry light cone gauge," Phys. Rev. D **63**, 046002 (2001) [arXiv:hep-th/0007036].

[2] N. Berkovits, "Super-Poincare covariant quantization of the superstring," JHEP **0004**, 018 (2000) [arXiv:hep-th/0001035].

[3] K. Pohlmeyer, "Integrable Hamiltonian Systems And Interactions Through Quadratic Constraints," Commun. Math. Phys. **46**, 207 (1976).

[4] M. Luscher and K. Pohlmeyer, "Scattering Of Massless Lumps And Nonlocal Charges In The Two-Dimensional Classical Nonlinear Sigma Model," Nucl. Phys. B **137**, 46 (1978).

[5] E. Brezin, C. Itzykson, J. Zinn-Justin and J. B. Zuber, "Remarks About The Existence Of Nonlocal Charges In Two-Dimensional Models," Phys. Lett. B **82** (1979) 442.

[6] E. Witten, talk given at California Institute of Technology, Nov. 3-4, 2001, in a meeting to celebrate 60 years of J. Schwarz, http://quark.caltech.edu/jhs60/witten/

[7] J. M. Evans, M. Hassan, N. J. MacKay and A. J. Mountain, "Local conserved charges in principal chiral models," Nucl. Phys. B **561**, 385 (1999) [arXiv:hep-th/9902008].

[8] A. Giveon, D. Kutasov and N. Seiberg, "Comments on string theory on AdS(3)," Adv. Theor. Math. Phys. **2**, 733 (1998) [arXiv:hep-th/9806194].

[9] S. S. Gubser, I. R. Klebanov, and A. M. Polyakov, *"A Semiclassical Limit of the Gauge/String Correspondence"*, hep-th/0204051.

[10] S. Frolov and A. A. Tseytlin, *"Semiclassical Quantization of Rotating Superstring in $AdS_5 \times S^5$"*, hep-th/0204226.

[11] S. Mukhi, M. Rangamani, E. Verlinde, *"Strings from Quivers, Membranes from Moose"*, hep-th/0204147, JHEP **0205**, 023 (2002).

[12] M. Blau, J. Figueroa-O'Farrill, C. Hull and G. Papadopoulos, "A new maximally supersymmetric background of IIB superstring theory," JHEP **0201**, 047 (2002) [arXiv:hep-th/0110242].

[13] M. Blau, J. Figueroa-O'Farrill, C. Hull and G. Papadopoulos, "Penrose limits and maximal supersymmetry," [arXiv:hep-th/0201081].

[14] M. Blau, J. Figueroa-O'Farrill and G. Papadopoulos, "Penrose limits, supergravity and brane dynamics," [arXiv:hep-th/0201081].

[15] R. R. Metsaev, "Type IIB Green-Schwarz superstring in plane wave Ramond-Ramond background," Nucl. Phys. B **625**, 70 (2002) [arXiv:hep-th/0112044].

[16] R. R. Metsaev and A. A. Tseytlin, *"Exactly Solvable Model of Superstring in Ramond-Ramond Plane Wave Background",* hep-th/0202109.

[17] D. Berenstein, J. M. Maldacena and H. Nastase, *"Strings in Flat Space and PP Waves from N=4 Superyang-Mills",* JHEP **0204**, 013 (2002), hep-th/0202021.

[18] D. Berenstein, E. Gava, J. M. Maldacena, H. Nastase and K. S. Narain, *"Open Strings on Plane Waves and Their Yang-Mills Duals",* hep-th/0203249.

asymptotics and taking into account a symmetry coefficient, we obtain

$$\dot{\mathbb{H}}_0^{\mathrm{sc}} \prod_{n=1}^{+\infty} \Gamma_{0,2n+2}^{\alpha_n} = \Lambda^2 \mathrm{V}_{2\hat{\alpha}} \vartheta(\dot{\alpha}) \frac{(-4)^{\dot{\alpha}} \dot{\alpha}!}{2 \prod_{n \geqslant 1} \alpha_n!}, \tag{7}$$

where the auxiliary function $\vartheta(\cdot)$ is defined by the equality

$$\vartheta(n) = \underbrace{AG_0^1 \star \ldots \star AG_0^1}_{\text{n times}}(0)$$

for all $n \geqslant 1$. The operation $\star$ denotes the convolution. Next, using representation (3) and a combination of convolution with the Fourier transform, the formula is rewritten as

$$\vartheta(n) = \int_{\mathbb{R}^2} \frac{\mathrm{d}^2 k}{4\pi^2} \, \rho^n(|k|).$$

Thus, after substituting (7) into (6), we get

$$-\frac{\Lambda^2}{2} \sum_{\alpha \colon \dot{\alpha} > 0} \left((-4)^{\dot{\alpha}} \dot{\alpha}! \mathrm{V}_{2\hat{\alpha}} \vartheta(\dot{\alpha}) \prod_{n=1}^{+\infty} \frac{\gamma^{2n\alpha_n} 2^{-\alpha_n}}{[(2n+2)!]^{\alpha_n} \alpha_n! \alpha_n!} \right).$$

Next, note that we can proceed to summation by components of the multiindex. To do this, we first use the definition for the vertex from (4) and the representation for the factorial

$$n! = \int_{\mathbb{R}_+} \mathrm{d}s \, s^n e^{-s}.$$

Then, by adding and subtracting the summand corresponding to $\dot{\alpha} = 0$, we obtain a factorization of the sums, in each of which we can use the representation for the Bessel function of the first kind, see Section 8.441 in [44],

$$\sum_{k=0}^{+\infty} \frac{(-s)^k}{k!k!} = J_0(2\sqrt{s}).$$

As a result, we obtain for Ψ the following formula

$$-\frac{\Lambda^2}{2} \mathrm{Tr}_{x,k} \Big[\mathrm{Kh}\Big(\phi(x)\gamma, 2\sqrt{\rho(|k|)} \Big) - \mathbf{1} \Big], \tag{8}$$

where "$\mathbf{1}$" denotes the identity matrix, the integral operator $\mathrm{Tr}_{x,k}$ is given by the formula

$$\mathrm{Tr}_{x,k} = \int_{\mathbb{R}^2} \mathrm{d}^2 x \int_{\mathbb{R}^2} \frac{\mathrm{d}^2 k}{4\pi^2} \, \mathrm{tr},$$

and the auxiliary function $\mathrm{Kh}(\cdot\,,\cdot)$ is defined by the equalities

$$\mathrm{Kh}(u, v) = \lim_{j \to +\infty} \mathrm{Kh}_j(u, v), \tag{9}$$

$$\mathrm{Kh}_j(u, v) = \int_{\mathbb{R}_+} \mathrm{d}s \, e^{-s} \left[\prod_{n=1}^{j} J_0\Big(\frac{u^n v \sqrt{s}}{\sqrt{n+1}} \Big) \right]. \tag{10}$$

Note that formula (8) can be reduced to the sum over the roots of characteristic polynomial of the matrix $\phi(x)$. Indeed, note that the matrix $\phi(x)$ is real and skew-symmetric. Therefore, see Chapter VI in [45], each root of its characteristic polynomial for each fixed "x" is either purely imaginary or zero. Moreover, imaginary ones come in pairs (with their conjugates). Notate the roots as $i\lambda^a(x)$, where $a \in \{1, \ldots, n^2-1\}$ and $\lambda^a(x) \in \mathbb{R}$. Then formula (8) is rewritten as

$$\Psi = -\frac{\Lambda^2}{2} \sum_{a=1}^{n^2-1} \mathrm{Tr}_{x,k} \Big[\mathrm{Kh}\Big(i\lambda^a(x)\gamma, 2\sqrt{\rho(|k|)} \Big) - \mathbf{1} \Big]. \tag{11}$$