Linear Jet Tagging with the Energy Flow Basis

Machine Learning for Jet Physics Workshop

Berkeley, CA – December 12, 2017

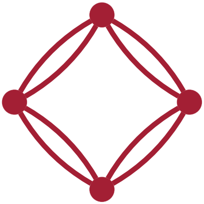
Patrick T. Komiske III

Center for Theoretical Physics, Massachusetts Institute of Technology

Based on work with Eric M. Metodiev and Jesse Thaler

Outline

Lightning Review of EFPs



Linear Classification with EFPs

Comparison with Modern Machine Learning

Fast Computation of EFPs

EFP Essentials

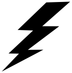
$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell} \tag{1}$$

$$z_i = \frac{p_{T,i}}{\sum_i p_{T,i}}, \quad \theta_{ij} = (\Delta y_{ij}^2 + \Delta\phi_{ij}^2)^{\beta/2}$$

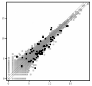
See Eric Metodiev's talk

- Jet substructure observables:
  - Fixed energy structure (necessary and sufficient for IRC safety)
  - Angular structure encoded by a multigraph G
  - Related to jet mass, ECFs, angularities, geometric moments, etc.
- Jet representation:
  - Complete, linear basis for IRC-safe information
  - Linearity of basis naturally suggests linear learning methods

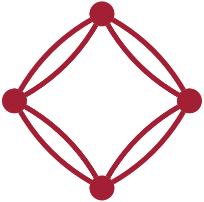
Outline



Lightning Review of EFPs



Linear Classification with EFPs



Comparison with Modern Machine Learning

Fast Computation of EFPs

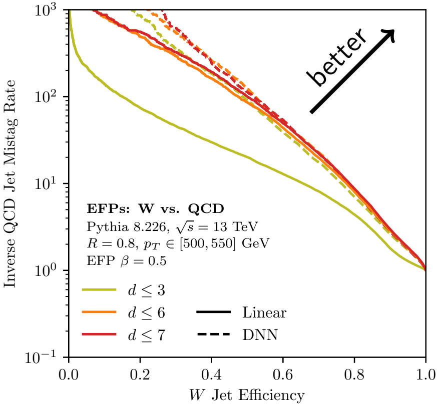
Linear Classification Overview

- Fit a decision plane, determined by a vector $\boldsymbol{w}$
  - Fisher's linear discriminant (LDA): closed-form solution
  - Logistic regression: Convex, iterative solution
- Decision threshold t is determined by distance from the plane
- $\mathcal{G}$ is finite set of graphs corresponding to the inputs
  - Organization by d is natural (equivalent to the order of the expansion)
  - Organization by N or χ also possible, (where is the information?)

$$\text{Classifier} = \left\{ \left(t + \sum_{G \in \mathcal{G}} w_G \text{EFP}_G \right) \begin{array}{ll} \geq 0, & \text{signal} \\ < 0, & \text{background} \end{array} \right.$$

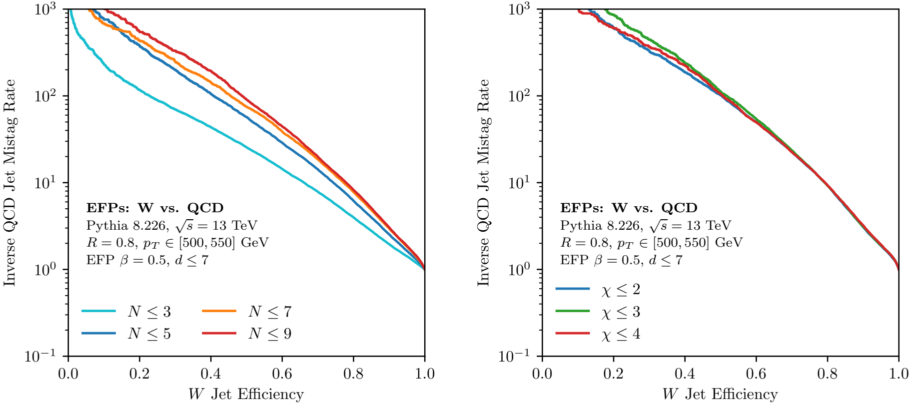
Linear Classification with EFPs

- W vs. QCD jet classification (quark/gluon and top tagging in backup)



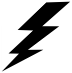
- 300k training samples
- Linear: Fisher's linear discriminant
  - num. params. = num. EFPs < 1000
  - 100k test samples
- DNN: Dense neural net
  - (100 node fully-connected layer) $\times$ 3
  - $\sim$ 120k parameters
  - 50k validation, 50k test samples

Which EFPs are Important?

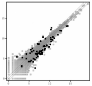


- High-N EFPs are important for classification performance
- Great classification performance with just $\chi = 2$ EFPs!

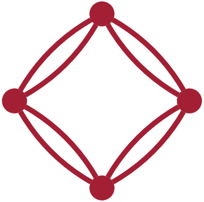
Outline

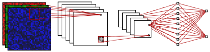


Lightning Review of EFPs



Linear Classification with EFPs

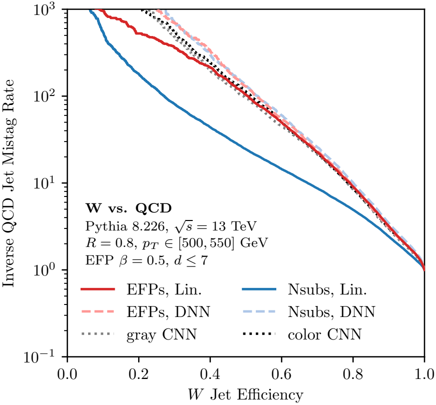




Comparison with Modern Machine Learning

Fast Computation of EFPs

Modern Machine Learning Comparison

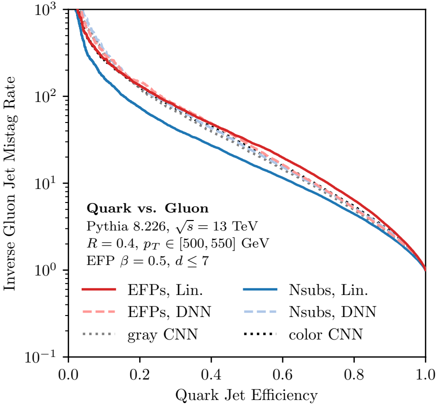


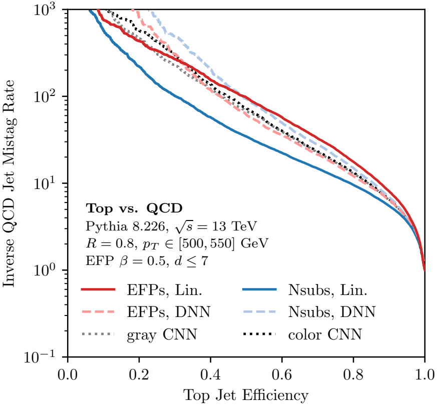
- Linear and DNN same as before
- CNN: Convolutional neural net
  - 33×33 jet images
  - (48 filters) $\times\, 3$
  - gray: p_T channel only
  - color: p_T and mult. channels
  - ~ 80k parameters

- (Linear classification with EFPs) $\sim$ (MML) for $\varepsilon_s \gtrsim 0.5$!

N-subjetiness: 1011.2268, N-subjetiness basis: 1704.08249, NN Review: 1709.04464

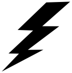
Modern Machine Learning Comparison



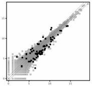


- (Linear classification with EFPs) $\gtrsim$ (MML) for $\varepsilon_s \gtrsim 0.5$

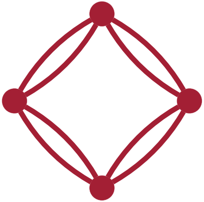
Outline

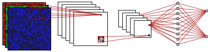


Lightning Review of EFPs



Linear Classification with EFPs





Comparison with Modern Machine Learning

Fast Computation of EFPs

Computational Complexity of ECF(G)s

$$\sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} E_{i_1} \cdots E_{i_N} \begin{cases} \prod_{i<j\in\{i_1,\ldots,i_N\}} \theta_{ij}^{\beta}, & \mathsf{ECF}_N^{\beta} \quad \text{1305.0007} \\ \prod_{m=1}^{v} \min_{(m)}\{\theta_{ij}^{\beta}\}_{i<j\in\{i_1,\ldots,i_N\}}, & {}_v\mathsf{ECFG}_N^{\beta} \quad \text{1609.07483} \end{cases}$$

- Implementation of ECF(G) formula runs in time $\mathcal{O}(M^N)$
- With $M \sim 100$, ECF(G)$_{N=4}$ $\sim$ one hundred million operations
- $N = 4$ is barely tractable, $N \geq 5$ is essentially inaccessible

Computational Complexity of ECF(G)s

$$\sum_{i_1=1}^{M}\cdots\sum_{i_N=1}^{M} E_{i_1}\cdots E_{i_N}\begin{cases}\displaystyle\prod_{i<j\in\{i_1,\ldots,i_N\}}\theta_{ij}^{\beta}, & \text{ECF}_N^{\beta} \quad 1305.0007\\ \displaystyle\prod_{m=1}^{v}\min_{(m)}\{\theta_{ij}^{\beta}\}_{i<j\in\{i_1,\ldots,i_N\}}, & {}_v\text{ECFG}_N^{\beta} \quad 1609.07483\end{cases}$$

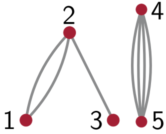
- Implementation of ECF(G) formula runs in time $\mathcal{O}(M^N)$
- With $M \sim 100$, ECF(G)$_{N=4}$ $\sim$ one hundred million operations
- $N = 4$ is barely tractable, $N \geq 5$ is essentially inaccessible

```
// if N > 5, then throw error
if (_N > 5) {
    throw Error("EnergyCorrelator is only hard coded for N = 0,1,2,3,4,5");
}
```

fjcontrib – EnergyCorrelator 1.2.0

Computational Complexity of EFPs

- Like other energy correlators, EFPs are naively $\mathcal{O}(M^N)$
- *Factorability* of summand in EFP formula can speed up computation

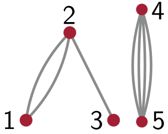


$$= \left(\sum_{i_1=1}^{M}\sum_{i_1=1}^{M}\sum_{i_3=1}^{M} z_{i_1} z_{i_2} z_{i_3} \theta_{i_1 i_2}^2 \theta_{i_2 i_3}\right)\left(\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_4} z_{i_5} \theta_{i_4 i_5}^4\right)$$

Composite EFPs are products of prime EFPs

Computational Complexity of EFPs

- Like other energy correlators, EFPs are naively $\mathcal{O}(M^N)$
- *Factorability* of summand in EFP formula can speed up computation



$$= \left(\sum_{i_1=1}^{M}\sum_{i_1=1}^{M}\sum_{i_3=1}^{M} z_{i_1} z_{i_2} z_{i_3} \theta_{i_1 i_2}^2 \theta_{i_2 i_3}\right)\left(\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_4} z_{i_5} \theta_{i_4 i_5}^4\right)$$

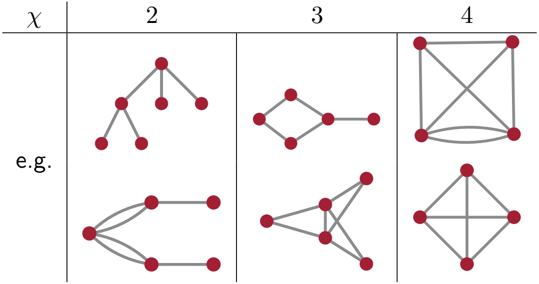
Composite EFPs are products of prime EFPs

$$= \underbrace{\sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\sum_{i_5=1}^{M}\sum_{i_6=1}^{M}\sum_{i_7=1}^{M}\sum_{i_8=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} z_{i_6} z_{i_7} z_{i_8} \theta_{i_1 i_2}\theta_{i_1 i_3}\theta_{i_1 i_4}\theta_{i_1 i_5}\theta_{i_1 i_6}\theta_{i_1 i_7}\theta_{i_1 i_8}}_{\mathcal{O}(M^8)}$$

$$= \underbrace{\sum_{i_1=1}^{M} z_{i_1}\left(\sum_{i_2=1}^{M} z_{i_2}\theta_{i_1 i_2}\right)^7}_{\mathcal{O}(M^2)}$$

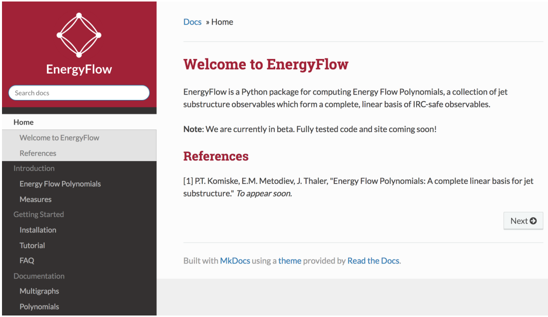
Variable Elimination (VE)

- Algorithm for finding optimal parentheses placement in EFP formula
- Reduces EFP computational complexity to $\mathcal{O}(M^{\chi})$:
  - Best case (NP-hard): $\chi = \texttt{treewidth}(G) + 1$
  - Heuristics can be used which work excellently for our small graphs
  - $\chi = 2$ for all tree graphs, $\chi = 3$ for single-cycle graphs, $\chi = N$ for K_N



EnergyFlow Python Package

- A convenient and simple package for efficient implementation of EFPs
- Currently written in pure Python using the NumPy library
  - Need a fast, arbitrary dimension multi-array
  - We've considered a C++ implementation (possible, but not simple)




```
pip install energyflow
```

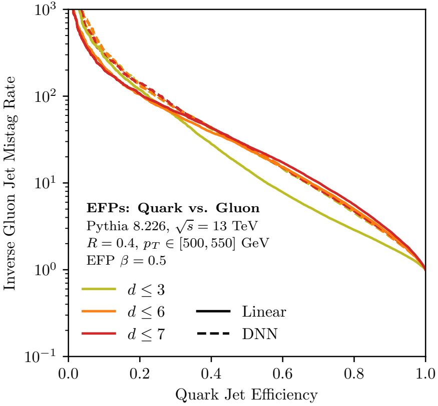
```python
import energyflow as ef
efps = ef.EFPSet('d<=7')
results = efps.compute(event)
```

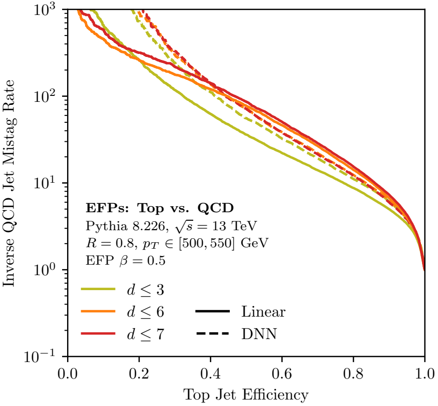
Conclusions

- Linear classification with EFPs very comparable to MML methods
- Linear methods $\implies$ very nice both theoretically and experimentally
  - EFP linear structure potentially allows for theoretical calculation
  - Fully differentiable model, uncertainty/error propagation simple
  - Convex, global minimum is guaranteed
  - No hyperparameters
  - Interesting methods made possible by linearity
    - Lasso regression for automatic feature selection
    - PCA, orthogonal subspaces, etc.
- Efficient computation of EFPs has been achieved
  - `EnergyFlow` Python package here, stay tuned for more
- EFPs potentially bridge MML performance & theory understanding

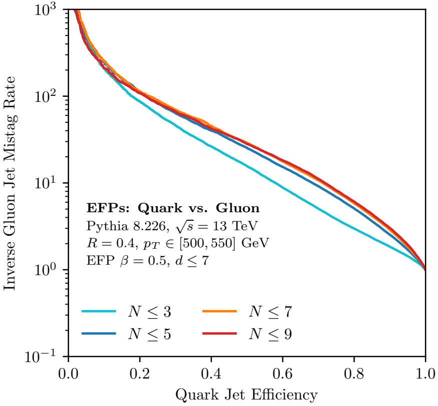
Additional Slides

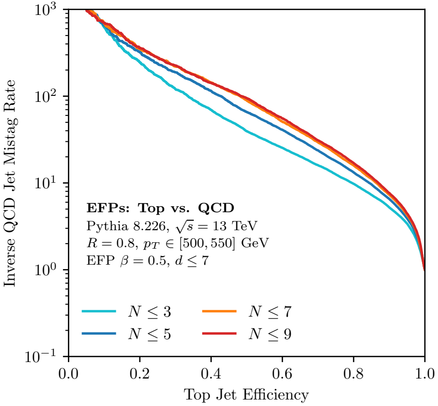
Quark/Gluon, Top Linear Classification with EFPs



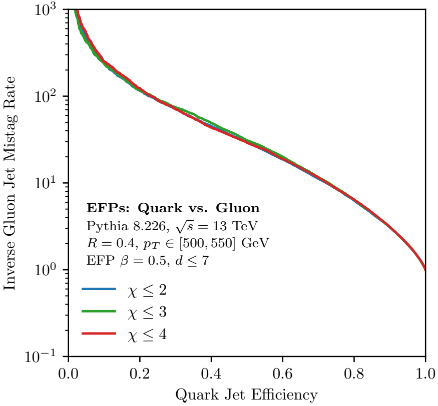


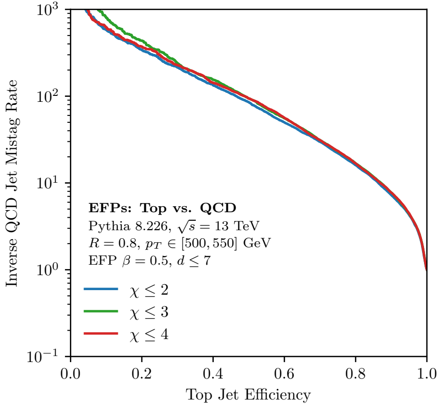
Quark/Gluon and Top Tagging N Sweep



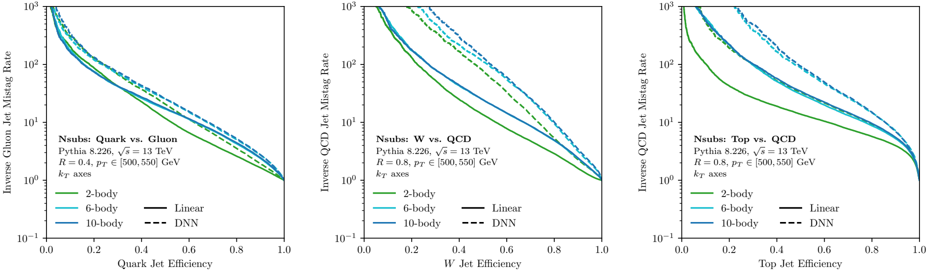


Quark/Gluon and Top Tagging χ Sweep



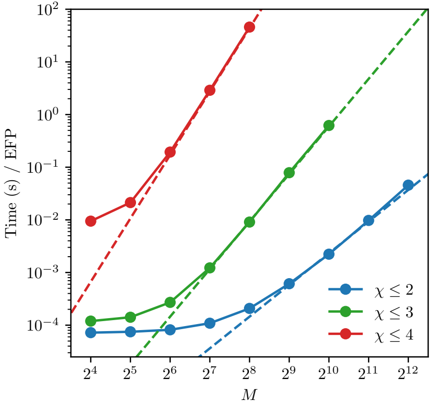


N-Subjettiness Linear/DNN Comparison



VE Timing

- Test our implementation of VE averaged over all EFPs with $d \leq 7$
- This includes prime EFPs up to $N = 8$! (Imagine $N = 8$ ECF, OMG)



N		2	3	4	5	6	7	8
χ	2	7	12	33	50	65	48	23
	3		11	42	82	80	33	
	4			2	1			

Prime EFPs with $d \leq 7$