The Matrix Element Method in the LHC era

XIIth Quark Confinement and the Hadron Spectrum – Thessaloniki, Greece

Sébastien Wertz

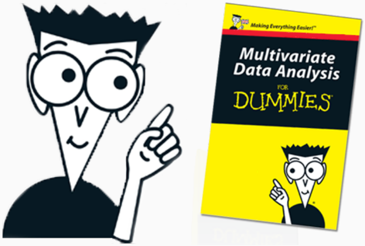
September 2nd, 2016

Université catholique de Louvain
Centre for Cosmology, Particle Physics and Phenomenology

Introduction

Motivation

- Use of Machine Learning MVA techniques in HEP ever more pervasive
- "MVA for dummies": easy to train & run BDT, ANN
- Usually satisfying results with "default" parameters!
- Important effort to make advanced techniques accessible (e.g. PyMVA)
- Even with tweaking, can be limited e.g. by **size of training sample**
  ...Keep independent test samples! Not always followed...
  Is there a way around?



Motivation

Aim:

Most powerful discriminator between simple hypotheses α (“signal”) and β (“background”)

Motivation

Aim:

Most powerful discriminator between simple hypotheses α (“signal”) and β (“background”)

Answer: Neyman–Pearson lemma

The most powerful discriminator to reject β in favor of α for event x is the likelihood ratio:

$$\Lambda(\mathsf{x}) = \frac{L(\mathsf{x}|\alpha)}{L(\mathsf{x}|\beta)}$$

Neyman, Pearson (1933)

...or any “nice” function of it

Motivation

Aim:

Most powerful discriminator between simple hypotheses α ("signal") and β ("background")

Answer: Neyman–Pearson lemma

The most powerful discriminator to reject β in favor of α for event $\mathbf{x}$ is the likelihood ratio:

$$\Lambda(\mathbf{x}) = \frac{L(\mathbf{x}|\alpha)}{L(\mathbf{x}|\beta)}$$

Neyman, Pearson (1933)

...or any "nice" function of it

ML algorithms try to approximate $\Lambda(\mathbf{x})$ given limited knowledge of the PDFs for α and β in the phase space (over-training)

Motivation

Aim:

Most powerful discriminator between simple hypotheses α ("signal") and β ("background")

Answer: Neyman–Pearson lemma

The most powerful discriminator to reject β in favor of α for event $\mathbf{x}$ is the likelihood ratio:

$$\Lambda(\mathbf{x}) = \frac{L(\mathbf{x}|\alpha)}{L(\mathbf{x}|\beta)}$$

Neyman, Pearson (1933)

...or any "nice" function of it

ML algorithms try to approximate $\Lambda(\mathbf{x})$ given limited knowledge of the PDFs for α and β in the phase space (over-training)

What about computing $\Lambda(\mathbf{x})$ from first principles?

MEM in a nutshell

Matrix Element Method in a nutshell

Hadron collider → likelihood for event **x** under hypothesis α:

Kondo (1988)

$$L(\mathbf{x}|\alpha) = \frac{1}{A_\alpha \sigma_\alpha} \int \mathrm{d}\Phi(y) \frac{\mathrm{d}x_1 \mathrm{d}x_2}{x_1 x_2 s} f(x_1) f(x_2) |\mathcal{M}_\alpha(y, x_1, x_2)|^2 W_\alpha(\mathbf{x}|y) \epsilon_\alpha(y)$$

Matrix Element Method in a nutshell

Hadron collider $\rightarrow$ likelihood for event $\mathbf{x}$ under hypothesis α:

Kondo (1988)

$$L(\mathbf{x}|\alpha) = \frac{1}{A_\alpha \sigma_\alpha} \int \mathrm{d}\Phi(y) \frac{\mathrm{d}x_1 \mathrm{d}x_2}{x_1 x_2 s} f(x_1) f(x_2) |\mathcal{M}_\alpha(y, x_1, x_2)|^2 W_\alpha(\mathbf{x}|y) \epsilon_\alpha(y)$$

Matrix element

- Typically LO
- May be simplified (e.g. no spin correlations)
- Analytical expression/exported from generator (e.g.: MadGraph)
- Implied: sum (average) over initial (final) state permutations (spin, colour, flavour, ...)

Matrix Element Method in a nutshell

Hadron collider $\rightarrow$ likelihood for event $\mathbf{x}$ under hypothesis α:

Kondo (1988)

$$L(\mathbf{x}|\alpha) = \frac{1}{A_\alpha \sigma_\alpha} \int \mathrm{d}\Phi(y) \frac{\mathrm{d}x_1 \mathrm{d}x_2}{x_1 x_2 s} f(x_1) f(x_2) |\mathcal{M}_\alpha(y, x_1, x_2)|^2 W_\alpha(\mathbf{x}|y) \epsilon_\alpha(y)$$

Transfer function

- Detector resolution, parton shower & hadronisation
- Normalised as $\int_\Omega \mathrm{d}\mathbf{x}\, W(\mathbf{x}|y) = 1$, where Ω is exp. acceptance
- In practice, one assumes factorisation & perfect angular resolution:

$$W(\mathbf{x}|y) = \prod_i W_i(E_i^{rec}|E_i^{gen})\, \delta(\theta_i^{rec} - \theta_i^{gen})\, \delta(\phi_i^{rec} - \phi_i^{gen})$$

- Fitted: single/double Gaussian, σ, $\mu \sim A + B \cdot \sqrt{E_{gen}} + C \cdot E_{gen}$
- Tabulated: 2D histogram of E^{rec} vs. E^{gen}

Matrix Element Method in a nutshell

Hadron collider $\rightarrow$ likelihood for event $\mathbf{x}$ under hypothesis α:

Kondo (1988)

$$L(\mathbf{x}|\alpha) = \frac{1}{A_\alpha \sigma_\alpha} \int \mathrm{d}\Phi(y) \frac{\mathrm{d}x_1 \mathrm{d}x_2}{x_1 x_2 s} f(x_1) f(x_2) |\mathcal{M}_\alpha(y, x_1, x_2)|^2 W_\alpha(\mathbf{x}|y) \epsilon_\alpha(y)$$

Efficiency

Probability for partonic configuration y to end up as a selected event
i.e., end up in Ω

Matrix Element Method in a nutshell

Hadron collider → likelihood for event **x** under hypothesis α:

Kondo (1988)

$$L(\mathbf{x}|\alpha) = \frac{1}{A_\alpha \sigma_\alpha} \int d\Phi(y) \frac{dx_1 dx_2}{x_1 x_2 s} f(x_1) f(x_2) |\mathcal{M}_\alpha(y, x_1, x_2)|^2 W_\alpha(\mathbf{x}|y) \epsilon_\alpha(y)$$

And all the rest of it:

- σ_α: cross-section for hypothesis α
- A_α: acceptance and efficiency, $A_\alpha = \langle \epsilon_\alpha(y) \rangle_\alpha$
- $f(x)$: parton distribution functions & Björken x
- s: hadronic centre-of-mass energy
- $d\Phi(y)$: phase-space density

With these definitions, $L(\mathbf{x}|\alpha)$ properly normalised
...Not crucial for some use cases

Volobouev, 1101.2259

MEM use cases

- Now that we can compute $L(\mathsf{x}|\alpha)$ and $L(\mathsf{x}|\beta)$, getting $\Lambda(\mathsf{x})$ is trivial
- Straightforward generalisation if several backgrounds β_i (relative expected yields f_i):

$$\Lambda(\mathsf{x}) = \frac{L(\mathsf{x}|\alpha)}{\sum_i f_i L(\mathsf{x}|\beta_i)}$$

- Instead of event-by-event discrimination, build sample likelihood:

$$L^{\alpha}_{\text{sample}} = \prod_{i \in \text{ sample}} L(\mathsf{x}_i|\alpha)$$

  and either:
  - Maximise w.r.t parameters (M.L. estimation)
    $\rightsquigarrow$ calibrate, calibrate, calibrate …
  - Test hypotheses using $L^{\alpha}_{\text{sample}}/L^{\beta}_{\text{sample}}$

MEM at the LHC

$t\bar{t}h$ searches

- $t\bar{t}h$: allows for *direct* measurement of Top-Higgs Yukawa coupling
- Small cross section (130/500fb @ 8/13 TeV), huge backgrounds, combinatorics
- Difficult S/B discrimination

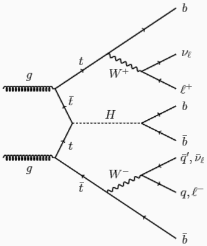
Many final states, several analyses by ATLAS & CMS. Those using MEM:

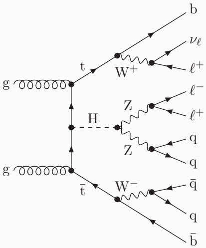
ATLAS *ATLAS (2015)*

$h \to b\bar{b}$, $t\bar{t} \to$ 1/2 leptons, 8 TeV:
Combination of MEM & ANN

CMS *CMS (2015, 2016)*

1. $h \to b\bar{b}$, $t\bar{t} \to$ 1/2 leptons, 8 TeV: MEM discriminant + b-tag info.
2. Multi-lepton: $h \to VV, \tau\tau$, 13 TeV: MEM used in $\geq$ 3-lepton category ($t\bar{t}V$ backgrounds), combination with BDT

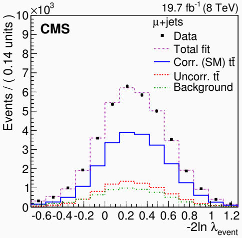


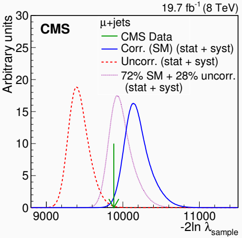


Dedicated implementations of the MEM used

Spin correlations in $t\bar{t}$

- $t\bar{t}$ pairs produced spin-correlated in SM
- Top quark short lifetime → correlation imprinted on decay products
- Measurement of correlation: test of SM





Analysis by CMS with MEM, μ+jets channel: *CMS (2015)*

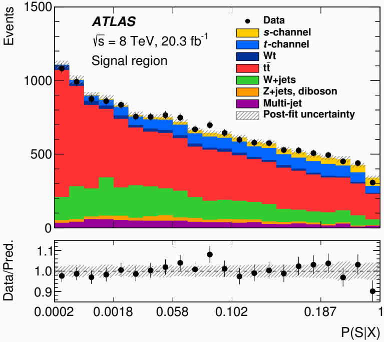
- Event likelihood: $\lambda(\mathbf{x}) = L(\mathbf{x}|\mathrm{uncorr})/L(\mathbf{x}|\mathrm{corr})$
  → calibrated template fit of $\lambda(\mathbf{x})$
  → $f_{\mathrm{corr}} = 0.72 \pm 0.09(\mathrm{stat})^{+0.15}_{-0.13}(\mathrm{syst})$
- Sample likelihood $\lambda_{\mathrm{sample}} = \prod_i \lambda(\mathbf{x}_i)$
  → Agreement with SM hypothesis: 2.2σ
  → Agreement with "no correlations": 2.9σ
- Most precise measurement in ℓ+jets channel
- Likelihoods computed with MadWeight

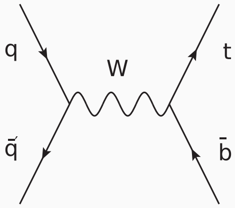
s-channel single-Top search

- Rare Top quark production mode (LHC8: $\sim$ 5pb $\sim 2\% \cdot t\bar{t}$)
- First evidence @ Tevatron (DØ, CDF): MEM & ML techniques
- More challenging @ LHC – exp./obs.: ATLAS $1.4/1.3\sigma$, CMS $0.9/0.7\sigma$
- ATLAS: re-analysis, use of MEM instead of BDT $\rightarrow 3.9/3.3\sigma$
  - 2 signal hypotheses: 2 & 3 partons
  - 6 bkg. hypotheses: t-channel, $t\bar{t}$ $1\ell/2\ell$, $W + jj$, $W + cj$, $W + bb$
  - In-house MEM implementation: MEMTK (CUBA, LHAPDF5, MCFM)

ATLAS (2014)
CMS (2013)

ATLAS (2015)





MEM implementation

Implementation: difficulties

On the market:

- A few process-dedicated codes (interfaced with analyses)
- MadWeight: (almost) any user-chosen hypothesis

Combining ME, PDFs, transfer functions, integration not very difficult...
Difficult part: $\mathrm{d}\Phi(y) \propto \prod_i \frac{\mathrm{d}^3 p_i}{2E_i} \delta^4(P_{in} - P_{out})$:

- Adaptive MC integration (Vegas) doesn't like peaked integrands
- Need to align peaks with integration coordinates *Lepage (1978)*
- Delta (momentum conservation) $\rightarrow$ need *some* thought beforehand
- Need to align propagators without spoiling transfer functions:

$$\mathrm{d}^3 p = |E|^2 \, \sin\theta \, \mathrm{d}E \, \mathrm{d}\theta \, \mathrm{d}\phi \, W(E^{rec}|E) \, \delta(\theta^{rec} - \theta) \, \delta(\phi^{rec} - \phi)$$

- "Standard" phase-space reduction techniques don't work

MadWeight: smart phase-space choices

MadWeight: *automatic* algorithm identifying most efficient parametrisations
$\rightarrow$ combine: *Artoisenet et al. (2010)*

"Main block" : remove δ^4, align some propagators

"Secondary blocks" : if needed, align other propagators

Alignment always (ie., within possibilities) such that removed d.o.f. is:

- Invisible (flat transfer function) – e.g. neutrinos
- As fallback, not too much constrained (e.g. jet, bad resolution)

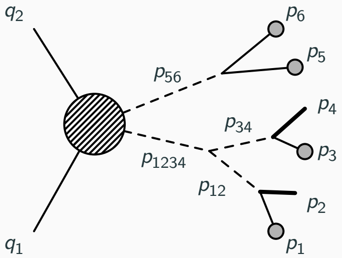
Simple example

$p_1 \rightarrow p_2 p_3 \Longleftrightarrow \mathrm{d}E_2\, \mathrm{d}\theta_2\, \mathrm{d}\phi_2\, \mathrm{d}E_3\, \mathrm{d}\theta_3\, \mathrm{d}\phi_3$, visible final state
$\Rightarrow$ Change of variables to $\mathrm{d}s_1\, \mathrm{d}E_2\, \mathrm{d}\theta_2\, \mathrm{d}\phi_2\, \mathrm{d}\theta_3\, \mathrm{d}\phi_3$
$\Rightarrow$ Propagator for p_1 aligned with grid

+ Aligned propagators can be removed (analytically invert Breit-Wigner)

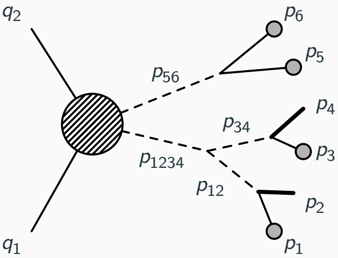
MadWeight phase-space: a more convoluted example

Di-Higgs production, $h(b\bar{b})h(W(\ell\nu)W(\ell\nu))$:



MadWeight phase-space: a more convoluted example

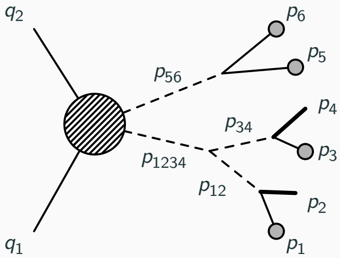
Di-Higgs production, $h(b\bar{b})h(W(\ell\nu)W(\ell\nu))$:



- Align p_{56}: “Secondary block D”
  - Trade one jet T.F. for h propagator
  - Other jet T.F. aligned
  - All angles: fixed

MadWeight phase-space: a more convoluted example

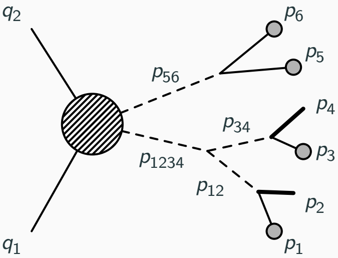
Di-Higgs production, $h(b\bar{b})h(W(\ell\nu)W(\ell\nu))$:



- Align p_{56}: “Secondary block D”
  - Trade one jet T.F. for h propagator
  - Other jet T.F. aligned
  - All angles: fixed
- Align p_{34}: “Secondary block C”
  - Assumes that W on-shell
  - Lepton p_3 T.F. aligned, angles fixed
  - Integrate p_4’s angles (flat)

MadWeight phase-space: a more convoluted example

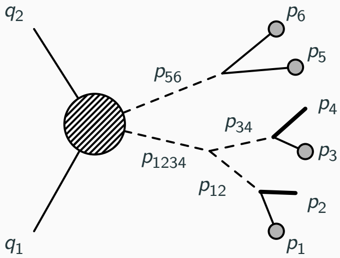
Di-Higgs production, $h(b\bar{b})h(W(\ell\nu)W(\ell\nu))$:



- Align p_{56}: "Secondary block D"
  - Trade one jet T.F. for h propagator
  - Other jet T.F. aligned
  - All angles: fixed
- Align p_{34}: "Secondary block C"
  - Assumes that W on-shell
  - Lepton p_3 T.F. aligned, angles fixed
  - Integrate p_4's angles (flat)
- Align p_{1234}, remove δ^4: "Main block B"
  - Lepton p_1 T.F. aligned, angles fixed
  - Trade q_1, q_2, p_2 for s_{1234}

MadWeight phase-space: a more convoluted example

Di-Higgs production, $h(b\bar{b})h(W(\ell\nu)W(\ell\nu))$:



- Align p_{56}: "Secondary block D"
  - Trade one jet T.F. for h propagator
  - Other jet T.F. aligned
  - All angles: fixed
- Align p_{34}: "Secondary block C"
  - Assumes that W on-shell
  - Lepton p_3 T.F. aligned, angles fixed
  - Integrate p_4's angles (flat)
- Align p_{1234}, remove δ^4: "Main block B"
  - Lepton p_1 T.F. aligned, angles fixed
  - Trade q_1, q_2, p_2 for s_{1234}

In all, $2 \times 2 = 4$ possibilities $\rightarrow$ multi-channel needed!

MoMEMta: smart *and* modular

MW: efficient parametrisations but: rigid structure, no API (I/O: text files)
...not in development/supported anymore!

New project: MoMEMta ▸ Website

- Written in C++ (more common knowledge than `Fortran`)
- Choice: not automatic → more freedom to the user
- Modular structure:
  - Module ⇔ C++ class: retrieve inputs, produce output
  - Structure of integrand steered through `Lua` script files
  - Packaged: modules for T.F.s, MadWeight's blocks, ...
- New C++ matrix element exporter (plugin for MG5_aMC@NLO) *Alwall et al. (2014)*
- Based on Root, LHAPDF6, Cuba *Buckley et al. (2014); Hahn (2005)*

Several topologies available, more to come...!

Team: Sébastien Brochet, Alessia Saggio, Miguel Vidal, SW

MEM on the GPU

Same function evaluated $\mathcal{O}(10^5)$ times: $\int f(x)\,\mathrm{d}x \cong \frac{1}{\sum w_i} \sum w_i f(x_i)$

$\rightarrow$ Embarrassingly parallel! $\rightarrow$ Use GPUs or co-processors?

Schouten, DeAbreu, Stelzer (2014)

- $t\bar{t}h(b\bar{b})$, $t\bar{t}b\bar{b}$ hypotheses
- CUDA, OPENCL technologies
  $\rightarrow$ parallelisation on CPU/GPU
- $t_{\mathrm{CPUC}^*}/t_{\mathrm{GPU}} \sim \mathcal{O}(50)$

Grasseau et al. (2015)

- $t\bar{t}h(\tau^+\tau^-)$
- MPI, OPENCL code
  $\rightarrow$ CPU, Xeon Phi, GPU
- $t_{\mathrm{CPUC}^*}/t_{\mathrm{GPU}} \sim \mathcal{O}(50 - 100)$

Impressive gains? but...

- GPUs underused (limited by data transfers; double/single precision)
- Usually many more CPUCs* than GPUs available ($\mathcal{O}(100 - 1000)$ more!)
  $\rightarrow$ simply parallelise events (small memory use!)
- Overall cost/gain ratio not good enough
- No general implementation (yet?)

* = CPU core

Conclusion

Conclusions & prospects

- MEM powerful technique for:
  - S/B discrimination
  - Hypothesis testing
  - Measurements
- Specific uses at the LHC
- Obstacles against wider use:
  Implementation difficult, (still) computationally heavy
- Automatic & efficient: MadWeight ...but hard to use in practice
- New project: MoMEMta $\rightarrow$ efficient & modular
- Recent developments in computing (GPGPU) ...to be followed!

Thank you!