The Matrix Element Method in the LHC era

XII\textsuperscript{th} Quark Confinement and the Hadron Spectrum – Thessaloniki, Greece

Sébastien Wertz
September 2\textsuperscript{nd}, 2016

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

MEM in a nutshell

MEM at the LHC

MEM implementation

Conclusion
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?
Aim:
Most powerful discriminator between simple hypotheses $\alpha$ ("signal") and $\beta$ ("background")
Motivation

Aim:
Most powerful discriminator between simple hypotheses $\alpha$ ("signal") and $\beta$ ("background")

Answer: Neyman–Pearson lemma
The most powerful discriminator to reject $\beta$ in favor of $\alpha$ for event $x$ is the likelihood ratio:

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

...or any “nice” function of it

Neyman, Pearson (1933)
**Motivation**

**Aim:**
Most powerful discriminator between simple hypotheses $\alpha$ ("signal") and $\beta$ ("background")

**Answer: Neyman–Pearson lemma**
The most powerful discriminator to reject $\beta$ in favor of $\alpha$ for event $x$ is the likelihood ratio:

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

...or any “nice” function of it

Neyman, Pearson (1933)

ML algorithms try to approximate $\Lambda(x)$ given limited knowledge of the PDFs for $\alpha$ and $\beta$ in the phase space (over-training)
Motivation

Aim:
Most powerful discriminator between simple hypotheses \( \alpha \) ("signal") and \( \beta \) ("background")

Answer: Neyman–Pearson lemma
The most powerful discriminator to reject \( \beta \) in favor of \( \alpha \) for event \( x \) is the likelihood ratio:

\[
\Lambda(x) = \frac{L(x|\alpha)}{L(x|\beta)}
\]

...or any "nice" function of it

Neyman, Pearson (1933)

ML algorithms try to approximate \( \Lambda(x) \) given limited knowledge of the PDFs for \( \alpha \) and \( \beta \) in the phase space (over-training)

What about computing \( \Lambda(x) \) from first principles?
MEM in a nutshell
Hadron collider $\rightarrow$ likelihood for event $x$ under hypothesis $\alpha$: 

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

Kondo (1988)
Matrix Element Method in a nutshell

Hadron collider → likelihood for event $\mathbf{x}$ under hypothesis $\alpha$:

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

**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, ...)

Kondo (1988)
Matrix Element Method in a nutshell

Hadron collider → likelihood for event \( x \) under hypothesis \( \alpha \):

\[
L(x|\alpha) = \frac{1}{A_\alpha \sigma_\alpha} \int d\Phi(y) \frac{dx_1dx_2}{x_1x_2} f(x_1)f(x_2) |M_\alpha(y, x_1, x_2)|^2 W_\alpha(x|y) \epsilon_\alpha(y)
\]

Transfer function

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

\[
W(x|y) = \prod_i W_i(E_{i}^\text{rec}|E_{i}^\text{gen}) \delta(\theta_{i}^\text{rec} - \theta_{i}^\text{gen}) \delta(\phi_{i}^\text{rec} - \phi_{i}^\text{gen})
\]

- Fitted: single/double Gaussian, \( \sigma, \mu \sim A + B \cdot \sqrt{E_{\text{gen}}} + C \cdot E_{\text{gen}} \)
- Tabulated: 2D histogram of \( E_{\text{rec}} \) vs. \( E_{\text{gen}} \)
Matrix Element Method in a nutshell

Hadron collider → likelihood for event $x$ under hypothesis $\alpha$:

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

Efficiency

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

Kondo (1988)
Matrix Element Method in a nutshell

Hadron collider → likelihood for event \( x \) under hypothesis \( \alpha \):

\[
L(x|\alpha) = \frac{1}{A_\alpha \sigma_\alpha} \int d\Phi(y) \frac{dx_1 dx_2}{x_1 x_2 s} f(x_1) f(x_2) |M_\alpha(y, x_1, x_2)|^2 W_\alpha(x|y) \epsilon_\alpha(y)
\]

And all the rest of it:

- \( \sigma_\alpha \): cross-section for hypothesis \( \alpha \)
- \( A_\alpha \): 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, \( L(x|\alpha) \) properly normalised

...Not crucial for some use cases
MEM use cases

- Now that we can compute $L(x|\alpha)$ and $L(x|\beta)$, getting $\Lambda(x)$ is trivial.

- Straightforward generalisation if several backgrounds $\beta_i$ (relative expected yields $f_i$):

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

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

$$L_{\text{sample}}^\alpha = \prod_{i \in \text{sample}} L(x_i|\alpha)$$

and either:

- Maximise w.r.t parameters (M.L. estimation)
  $\Rightarrow$ calibrate, calibrate, calibrate ...

- Test hypotheses using $L_{\text{sample}}^\alpha / L_{\text{sample}}^\beta$
MEM at the LHC
tth searches

- tth: 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:

<table>
<thead>
<tr>
<th>ATLAS</th>
<th>ATLAS (2015)</th>
</tr>
</thead>
<tbody>
<tr>
<td>h → b, t → 1/2 leptons, 8 TeV:</td>
<td>Combination of MEM &amp; ANN</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CMS</th>
<th>CMS (2015, 2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. h → b, t → 1/2 leptons, 8 TeV:</td>
<td>MEM discriminant + b-tag info.</td>
</tr>
<tr>
<td>2. Multi-lepton: h → VV, ττ, 13 TeV:</td>
<td>MEM used in ≥ 3-lepton category (tV backgrounds), combination with BDT</td>
</tr>
</tbody>
</table>

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 $\rightarrow$ correlation imprinted on decay products
- Measurement of correlation: test of SM

Analysis by CMS with MEM, $\mu$+jets channel: $CMS$ ($2015$)

- Event likelihood: $\lambda(x) = L(x|\text{uncorr})/L(x|\text{corr})$
  $\rightarrow$ calibrated template fit of $\lambda(x)$
  $\rightarrow f_{\text{corr}} = 0.72 \pm 0.09\text{(stat)}^{+0.15}_{-0.13}\text{(syst)}$

- Sample likelihood $\lambda_{\text{sample}} = \prod_i \lambda(x_i)$
  $\rightarrow$ Agreement with SM hypothesis: $2.2\sigma$
  $\rightarrow$ Agreement with “no correlations”: $2.9\sigma$

- Most precise measurement in $\ell$+jets channel
- Likelihoods computed with MADWEIGHT
s-channel single-Top search

- Rare Top quark production mode (LHC8: $\sim 5\text{pb} \sim 2\% \cdot \tilde{t}\tilde{t}$)
- 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, $t\tilde{t}$ $1\ell/2\ell$, $W + jj$, $W + cj$, $W + bb$
  - In-house MEM implementation: MEMtk (CUBA, LHAPDF5, MCFM)

![Graph depicting s-channel single-Top search](chart.png)

**Event distributions**

- **Data**
- **s-channel**
- **t-channel**
- **Wt**
- **tt**
- **W+jets**
- **Z+jets, diboson**
- **Multi-jet**
- **Post-fit uncertainty**

**Signal region**

- **$\sqrt{s} = 8\text{ TeV}, 20.3\text{ fb}^{-1}$**

**Event counts**

<table>
<thead>
<tr>
<th>Events</th>
<th>Data/Pred.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0$</td>
<td>$0.9$</td>
</tr>
<tr>
<td>$500$</td>
<td>$1.0$</td>
</tr>
<tr>
<td>$1000$</td>
<td>$1.1$</td>
</tr>
<tr>
<td>$1500$</td>
<td>$0.0002$</td>
</tr>
</tbody>
</table>

**P-values**

- **ATLAS**
- **CMS**
- **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: $d\Phi(y) \propto \prod_i \frac{d^3 p_i}{2E_i} \delta^4(P_{in} - P_{out})$:

- Adaptive MC integration (**VEGAS**) doesn’t like peaked integrands
- Need to align peaks with integration coordinates
- Delta (momentum conservation) → need some thought beforehand
- Need to align **propagators** without spoiling **transfer functions**:

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

- “Standard” phase-space reduction techniques don’t work
**MADWEIGHT: smart phase-space choices**

**MADWEIGHT: automatic** algorithm identifying most efficient parametrisations
→ combine:

“Main block” : remove $\delta^4$, 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**

$p_1 \rightarrow p_2 p_3 \iff dE_2 \, d\theta_2 \, d\phi_2 \, dE_3 \, d\theta_3 \, d\phi_3$, visible final state
⇒ Change of variables to $ds_1 \, dE_2 \, d\theta_2 \, d\phi_2 \, d\theta_3 \, d\phi_3$
⇒ Propagator for $p_1$ aligned with grid

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

Artoisenet et al. (2010)
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
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)
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 $\delta^4$: “Main block B”
  - Lepton $p_1$ T.F. aligned, angles fixed
  - Trade $q_1, q_2, p_2$ for $s_{1234}$
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 $\delta^4$: “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)
► Based on ROOT, LHAPDF6, CUBA

Alwall et al. (2014)
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 $O(10^5)$ times: $\int f(x) \, dx \approx \frac{1}{\sum w_i} \sum w_i f(x_i)$

→ Embarrassingly parallel! → Use GPUs or co-processors?

### Schouten, DeAbreu, Stelzer (2014)
- $t_{\bar{t}t\bar{b}b}$, $t\bar{t}b\bar{b}$ hypotheses
- CUDA, OPENCL technologies
  → parallelisation on CPU/GPU
- $t_{CPUC}^*/t_{GPU} \sim O(50)$

### Grasseau et al. (2015)
- $t_{\tau^+\tau^-}$
- MPI, OPENCL code
  → CPU, Xeon Phi, GPU
- $t_{CPUC}^*/t_{GPU} \sim O(50 \text{ to } 100)$

Impressive gains? but...
- GPUs underused (limited by data transfers; double/single precision)
- Usually many more CPUCs* than GPUs available ($O(100 \text{ to } 1000)$ more!)
  → 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 → efficient & modular

- Recent developments in computing (GPGPU) ...to be followed!
Thank you!