Terascale ISTP 2010

Hands–on Z' reconstruction with ROOT/C++

Ivor Fleck • Marcus Rammes

1 Preface

Any analysis of LHC data is a very complex procedure and the search for new physics is a difficult challenge in particular. There is a vast variety of the most different theories, models and scenarios on the market that predict very similar physical signatures. In addition to the problem how to distinguish those signatures, we mostly are concerned with very small cross sections; we can extract only a few events, overlaid by a large amount of irreducible background.
Most of the models and scenarios will be excluded at a certain confidence level, indicating upper bounds for the cross section.

In this exercise, we only can cover very few elements of the whole sequence of reconstruction and analysis of Z' physics in order to demonstrate an unsubtle strategy along general lines.

2 Some theory in a nutshell

2.1 What is a Z'?

Basically, a Z' is an electrically neutral resonance more massive then the Standard model (SM) Z^0. In the SM gauge group extension $SU(2)_{EW} \times U(1)_Y \rightarrow SU(2)_{EW} \times U(1)_Y \times U'(1)_{Z'}$, the Z' is a gauge boson that mixes with the normal Standard Model Z^0. If it exists, its mass is supposed to be $M_{Z'} \gtrsim 500\,\mathrm{GeV}$.
Due to its large mass, it will decay most preferably into a pair of top quarks ($Z' \rightarrow t\bar{t}$). In Z' gauge boson models, the cross section is about a factor 30-200 smaller than for SM $t\bar{t}$ production. For that reason, the standard $t\bar{t}$ production processes will dominate the irreducible background.
This very brief explanation should be sufficient for now, if you want to know more about the Z' you should refer to [1].

2.2 The top quark

The top quark is the 6th flavor in the quark sector of the SM. It decays almost completely into a W boson and a b quark:

$$t \rightarrow W^+ b \qquad \bar{t} \rightarrow W^- \bar{b}$$

The top quark is by far the heaviest one of the six quarks in the SM ($M_t \approx 172.5\,\mathrm{GeV}$). For that reason, the top quark has some interesting features with respect to physics at the terascale:

- In NLO calculations for heavy SUSY particles $\tilde{X}$ coupling to a quark flavor q_i, the coupling strength is proportional to $\ln\left(M_{q_i}{}^2/M_{\tilde{X}}^2\right)$. For light quarks, you will have no chance to detect any anomalies in signatures compared to SM predictions. The top quark is heavy enough to significantly exhibit such differences.
- The top quark has a very short lifetime, it decays before it can hadronize. For that reason, spin correlations between t and $\bar{t}$ and W polarization studies can be performed.
- In resonant decays of some heavy particle X, $X \to t\bar{t}$ is the only decay channel that can be observed with a feasible background (the invariant mass of X has then to be at least twice the top mass, of course).

2.3 $t\bar{t}$ decay modes

Most of the top quarks at the LHC are produced as $t\bar{t}$ pairs. Assuming that they almost always decay via $t \to Wb$, $t\bar{t}$ events can be devided into three classes, depending on how the W bosons are decaying afterwards:

- **full–hadronic decay:** Both W bosons decay into a light quark and a anti–quark pair ($\to$ 4 light jets) ($W^+W^- \to j_1j_2j_3j_4$)
- **semi–leptonic decay:** One W boson decays into a light quark and a anti–quark pair ($\to$ 2 light jets), the other one into a charged lepton and a neutrino ($W^+W^- \to j_1j_2\ell\nu$)
- **di–leptonic decay:** Both W bosons decay into a charged lepton and a neutrino ($W^+W^- \to 2\ell2\nu$)

The full–hadronic mode has the largest branching ratio (BR), but is contaminated by a big amount of QCD background, furthermore it is difficult to trigger on this dacay mode. The di–leptonic channel is the cleanest one, but has only a BR of $\approx 18\%$.
The semi–leptonic mode has a moderate background combined with a feasible BR. Therefore, this mode is also called the "golden channel".
For Z' discovery, the semi–leptonic decay mode is chosen.

Besides, the decay mode $W \to \tau\nu_\tau$ is most often omitted since the τ is difficult to reconstruct. For the semi–leptonic decay mode with $\ell = e, \mu$, a BR of $\approx 31\,\%$ remains.

3 Analysis templates and data samples

3.1 The analysis software package

A basic framework for Z' reconstruction has been prepared for this exercise. We have been trying to keep the program structure as simple as possible, so we

renounced using complicated classes with multiple inheritance, function/class templates etc.
The most important physical objects are grouped in structs such as jets, muons, electrons, W bosons etc. Topic-related `C`-functions are organized in different files that just have to be included to your main program file. All functions are documented.

We prepared 10 exercises for this tutorial, each one stored in a corresponding include file `exercise1.cc` to `exercise10.cc`. They simply can be invoked by `exercise1()` etc. from the main method.
Although all exercises are completely available, you are strongly advised to try to do at least some of the exercises on your own and just refer to the solution when you're stuck.

3.1.1 `typedefs.cc`

Here you find the definitions of the data types and structs used for the analysis. The members of the structs are named according to the naming conventions in the ROOT NTuple files.

`struct Jet`

- `E` : jet energy
- `Et` : jet transverse energy
- `p` : norm of jet 3-momentum
- `pt`: jet p_T
- `px,py,pz` : vector components of jet 3-momentum
- `double eta,phi`: jet η, φ
- `tag_weight` : statistical weight of b tagging
- `bjet_prob` : b tag weight converted to a probability

`struct Electron`

- `E` : electron energy
- `Et` : electron transverse energy
- `p` : norm of electron 3-momentum
- `pt`: electron p_T
- `px,py,pz` : vector components of electron 3-momentum
- `double eta,phi`: electron η, φ
- `charge` : charge of the electron
- `etCone20` : energy deposition in a $\Delta R = 0.2$ cone around the electron track

struct Muon

- double E : muon energy
- double Et : muon transverse energy
- double p : norm of muon 3-momentum
- double pt: muon p_T
- double px,py,pz : vector components of muon 3-momentum
- double eta,phi: muon η, φ
- double tag_weight : statistical weight of b tagging
- double charge : charge of the muon
- double etCone20 : energy deposition in a $\Delta R = 0.2$ cone around the muon track

struct Event

- std::vector<Jet>: all jets of the event
- std::vector<Electron>: all electrons of the event
- std::vector<Muon>: all muons of the event
- double missingET: missing transverse energy $\not\!\!E_T$
- double missingET_phi: azimuthal direction of $\not\!\!E_T$ (angle in x, y–plane)
- weight: the event weight (only important when scaling samples up to a mixed data sample, is 1.0 in the prefabricated samples)

struct HadronicWBoson

- Jet jet1,jet2: light jet constituents of the hadronic W boson
- double mass: reconstructed invariant W mass

struct LeptonicWBoson

- double El,pxl,pyl,pzl: energy and 3-momentum components of the charged lepton constituent of the W boson
- double ptnu: transverse momentum of the recovered neutrino
- double pxnu,pynu: x and y 3-momentum components of the recovered neutrino
- std::pair<double,double> pznu: the two soluintions of the p_z^ν recovery (see 4.4)
- std::pair<double,double> Enu: the two energies calculated from the p_z^ν solutions of the recovered neutrino

- `std::pair<double,double> mass`: the two invariant W masses calculated from the p_z^ν solutions
- `double Enu_final`: E_ν for the best W fit
- `double pznu_final`: p_z^ν for the best W fit
- `double mass_final`: invariant W mass for the best W fit

`struct HadronicTop`

- `Jet bJet`: b–tagged jet closest in ΔR to the previously reconstructed hadronic W boson
- `HadronicWBoson wBoson`: the hadronic W boson
- `double mass`: reconstructed invariant hadronic top mass
- `double deltaR`: ΔR between the b jet and the hadronic W

`struct LeptonicTop`

- `Jet bJet`: b–tagged jet closest in ΔR to the previously reconstructed leptonic W boson
- `LeptonicWBoson wBoson`: the leptonic W boson
- `double mass`: reconstructed invariant leptonic top mass (best agreement of the two leptonic W solutions with hadronic top mass)
- `double deltaR`: ΔR between the b jet and the leptonic W

`struct RecoObject`

- `Event event`: the raw event associated to the reconstruction
- `HadronicTop hadTop`: the reconstructed hadronic top quark
- `LeptonicTop lepTop`: the reconstructed leptonic top quark
- `double ZprimeMass`: the reconstructed invariant Z' mass

The typedefs are just introduced for a better legibility, the jet operators to ease the comparison between jet objects.

3.1.2 `readNTuple.cc`

This file contains only one function. `Data* readNTuple(const char* file_name)` returns a pointer to a `std::vector<Event>`, containing all events available in the ROOT NTuple with the given file name.

3.1.3 reconstruction.cc

In reconstruction.cc you will find the standard methods for W boson, top quark and Z' mass reconstruction:

- HadronicWBoson getBestHadronicWFit(Event):
  This method reconstructs the hadronic W boson as described in 4.2
- HadronicTop getBestHadronicTopFit(Event):
  Returns the hadronic top quark reconstructed via the method explained in 4.3
- LeptonicWBoson getBestLeptonicWFit(Event, LeptonicWBoson):
  Reconstructs the leptonic W boson as described in 4.4
- LeptonicTop getBestLeptonicTopFit(Event, LeptonicWBoson, HadronicTop):
  Returns the leptonic top quark reconstructed as explained in 4.5
- RecoData* getReconstructedData(Data*):
  Returns a pointer to a std::vector<RecoObject>. The function applies all cuts, reconstructs the hadronic and leptonic W bosons and top quarks and calculates the invariant Z' mass.

3.1.4 cuts.cc

The most important function in cuts.cc is "bool isGoodEvent(Event)". The function checks if the event fulfils all cuts (see 4.1).

3.1.5 kinematics.cc

This file contains a set of various helper functions for kinematic reconstruction, such as the calculation of ΔR, η, invariant masses etc. Please refer to the commented source code for further information.

3.2 MC data samples

We have generated simple ROOT NTuples for signal and the most important background processes containing only the most relevant information you need for this exercise. They have been produced from official ATLAS MonteCarlo production data sets. The samples are provided in the directory ./data/.

file name	contained data	σ [pb]	$\int \mathcal{L}\, dt$ [pb^{-1}]
zee_jets.root	$Z \to e^+e^- + n$ jets	28.81	1000
zmumu_jets.root	$Z \to \mu^+\mu^- + n$ jets	29.41	1000
wenu_jets.root	$W \to e\nu + n$ jets	160.57	200
wmunu_jets.root	$W \to \mu\nu + n$ jets	90.77	200
wbb_jets.root	$Wb\bar{b} + n$ jets	14.64	1000
ttbar_sample.root	$t\bar{t}$, semi–lep	220.66	250
zprime_sample.root	$Z' \to t\bar{t}$, $M_{Z'} = 1000$ GeV	0.634	30000
data_sample.root	mixed $Z'/t\bar{t}$ sample	—	300

The luminosities of the samples were chosen in a way to end up with similar number of events in each sample.

3.3 Compiling/running the analysis

Everytime you make changes to the source code, you have to recompile the program by just typing `make`. Then you can run the analysis with `./main`.

Many people prefer the ROOT built–in C interpreter CINT instead of compiling source code.
If you want to use CINT and the ROOT system, you might write the histograms created by this analysis into a `TFile`, then you can open them in a `TBrowser` or in CINT afterwards to apply your styles, resize/move legends, try out fit functions etc.

4 The project: Finding the Z'

We are now going through the whole process of Z' reconstruction, step by step. The analysis is intended to be based on a data sample with an integrated luminosity of $3\,\text{fb}^{-1}$. This is a trade–off between having a minimum number of Z' events available and a still moderate event weight of the other samples that have a much higher cross section.
Therefore, always be aware to scale your histograms by the correct factor when normalizing your plots to a certain luminosity! Also be warned that **all energies and momenta are stored in units of MeV, not GeV!!!**

The analysis presented in this exercise is mainly based upon the methods described in the ATLAS top quark CSC notes 2008 [2].
Of course this analysis cannot cover a complete investigation of all uncertainties. Especially systematic errors are completely neglected, the methods of calculation of statistical errors are only estimations and maybe also not very accurate. In a real analysis for your thesis, this would be a central an time-consuming part of your work!

4.1 Event selection

In order to suppress background and to improve the purity of the reconstructed top quarks/Z's, several cuts have to be applied onto the data.

- At least 4 jets with $|\eta| < 2.5$ and $p_T > 40\,\text{GeV}$
- Exactly one isolated lepton with $p_T > 20,\text{GeV}$ for the electrons and $p_T > 25\,\text{GeV}$ for the muons; excluded crack region $1.37 < |\eta| < 1.52$ [1].
  The isolation cuts are $E_T^{\text{cone20}} < C1 + C2 \cdot E_T$ ($C1 = 4.0\,\text{GeV}$, $C2 = 0.023$) for the electrons and $E_T^{\text{cone20}}/p_T < 0.1$ for the muons
- Exactly two b–tagged jets (tag weight > 6.0)
- Missing transverse energy $\not{E}_T > 20\,\text{GeV}$

[1] The crack region is the region in the detector where the barrel detectors end and the end cap detectors begin. In this region, the tracking efficiency for leptons strongly drops. To avoid additional systematic uncertainties, this crack region is excluded in most analyses.

Exercise 1

Make two plots for the jet multiplicities of the $t\bar{t}$ and the Z' sample out of two overlaid histograms: One histogram filled with the jet mutliplicity before and one after applying the cuts. Normalize the histograms to their integrals.

Calculate the event selection efficiencies $\epsilon_{t\bar{t}}^{\mathrm{C1}}$ and $\epsilon_{Z'}^{\mathrm{C1}}$ for the $t\bar{t}$ and the Z' sample. From the efficiencies of all the background samples (except $t\bar{t}$) calculate an average background suppression factor $\langle b \rangle = 1 - \langle \epsilon \rangle$.

4.2 Reconstruction of the hadronic W boson

The hadronic W boson in a semi–leptonic top quark event is the W boson that decays into two light flavor (LF) jets. LF jets are all jets that have not been b–tagged (tag weight < 6.0). Among all LF jets, the two ones are used for the kinematic W reconstruction that build up the W mass fitting the best ($M_W = 80.4\,\mathrm{GeV}$).

Besides this method, which is the most intuitive one and was therefore chosen for this tutorial, there are two other ways for reconstructing the W:

- **The χ^2 method:** The two LF jets that minimize

$$\chi^2 = \frac{(M_{jj}(\alpha_{E_{j1}}, \alpha_{E_{j2}}) - M_W)^2}{{\Gamma_W}^2} + \frac{(E_{j1}(1-\alpha_{E_{j1}}))^2}{{\sigma_1}^2} + \frac{(E_{j2}(1-\alpha_{E_{j2}}))^2}{{\sigma_2}^2}$$

  with $\sigma = E \cdot \sqrt{(a^2/E) + b^2}$ ($a = 0.989\,\mathrm{GeV}^{1/2}$ and $b = 0.075$) and ${\Gamma_W}^2$ being the spectral linewidth of the W boson are selected.

- **The geometric method:** The LF jets that are closest in ΔR are used.

The most important background in this process is the so–called "combinatorical background": If you have n LF jets in your event, there are $n(n-1)$ possibilities to combine them two the W boson. For $n = 5$ LF jets, there are $5 \cdot 4 = 20$ combinations, only one of them is true.

The combinatorical background of the W reconstruction can be described by a 4th degree Tschebyschow polynomial, the correctly combined jets can be assumed to be Gaussian distributed around the W mass. The first four orders of Tschebyschow polynomials are:

$$T_0 = 1 \qquad T_1 = x \qquad T_2 = 2x^2 - 1$$

$$T_3 = 4x^3 - 3x \qquad T_4 = 8x^4 - 8x^2 + 1$$

Exercise 2

Combine the hadronic W mass spectra of the $t\bar{t}$ and the Z' samples in one plot. Normalize both histograms to their integrals.

Exercise 3

Fit the sum of a 4th degree Tschebyschow and a Gaussian distribution to your hadronic W mass spectrum of the $t\bar{t}$ sample and extract the W mass and width from the mean and the RMS of the Gaussian.

Hint: First define five `TF1`'s representing your Tschebyschoff polynomials $T_0 \ldots T_4$ and one `TF1` for the Gauss. In a sixth `TF1`, add the five polynomials, each one provided with a free fit parameter for scaling, and the Gaussian `TF1`. Use the standard `TH1::Fit(...)` method. Maybe you have to set start parameters for the Gauss to bring the fit to converge.

4.3 Reconstruction of the hadronic top quark

After the hadronic W is reconstructed, the b jet that is closest in ΔR to the hadronic W boson is chosen to reconstruct the hadronic top quark.

Exercise 4

Plot the hadronic top mass spectrum for the $t\bar{t}$ and the Z' sample (normalized to their integrals).
Perform the previous exercise with the Tschebyschoff fit to obtain the top mass and width of the $t\bar{t}$ sample.

4.4 Reconstruction of the leptonic W boson

The difficulty of reconstructing the W that decays into two leptons is the unknown neutrino momentum vector. The only observable that can be measured in a hadron collider is the missing transverse energy $\not{E}_T$; besides, the direction of $\not{E}_T$ in the x, y–plane can be reconstructed. In good approximation, $\not{E}_T$ can be considered to be the transverse energy of the neutrino.

Taking the W boson mass as a fixed value, one can obtain a quadratic equation in p_z^ν:

$$M_W{}^2 = m_\ell{}^2 - 2(p_x^\ell p_x^\nu + p_y^\ell p_y^\nu) + 2E_\ell \sqrt{\not{E}_T{}^2 + p_z^{\nu 2}} - 2(p_z^\ell p_z^\nu) \tag{1}$$

p_x^ν and p_y^ν can be calculated from $\not{E}_T$ and $\varphi(\not{E}_T)$:

$$p_x^\nu = \not{E}_T \cdot \cos(\varphi(\not{E}_T)) \qquad p_y^\nu = \not{E}_T \cdot \sin(\varphi(\not{E}_T))$$

(1) has either two solutions or none. If it has none, p_T^ν is iterativelly decreased until the equation has two solutions. Later on, the leptonic top mass will be reconstructed with the leptonic W from both solutions. The leptonic top mass that agrees best with the hadronic top mass will be kept.

Exercise 5

Combine the histograms of the leptonic W mass for both p_z^ν solutions in one plot. Make one plot for each the $t\bar{t}$ and the Z' sample.

4.5 Reconstruction of the leptonic top quark

After the reconstruction of the hadronic top quark, there will be one remaining b jet. Together with the neutrino solution for the section before that agrees best with the mass of the hadronic top quark is chosen to be the reconstructed leptonic top quark.

4.6 Z' reconstruction

Finally, the Z' invariant mass can be calculated from the reconstructed tops. Since the cross section for SM $t\bar{t}$ production is about a factor 370 higher than $\sigma(Z' \to t\bar{t})$, one has to think about a way to distinguish the production channels.

$t\bar{t}$ production from a Z' is a resonant decay, i. e. the top quarks will be more boosted in average. As consequence, the opening angle of the cone of the decay products from the top quarks will be more narrow. This should be seen as a clear signature in the ΔR distributions of the b jet and the corresponding W boson.

Exercise 6

Fill two histograms with the invariant Z' mass spectrum from the Z' and $t\bar{t}$ samples and combine them in one plot. Choose logarithmic y scaling.

Exercise 7

For each the hadronic and the leptonic top quark, fill two histograms with the ΔR spectrum of the Z' and $t\bar{t}$ samples. Make two plots out of them (e. g. "`hadronic_top_deltaR.eps`" and "`leptonic_top_deltaR.eps`"). Normalize the two histograms in each plot to their integrals.

Choose a cut in ΔR in a way to maximize purity on the on hand, but not cutting away too many events on the other hand. Apply an additional, appropriate cut on the invariant Z' mass; choose a quite naarow window around 1000 GeV

With these additional cuts, again fill two histograms with the invariant Z' mass, this time with linear y scaling. Normalize the histograms to $3\,\mathrm{fb}^{-1}$ now.
Calculate the efficiencies $\epsilon_{t\bar{t}}^{\mathrm{C2}}$ und $\epsilon_{Z'}^{\mathrm{C2}}$ for the additional cuts.

4.7 Discovery potential

Let's assume there was no Z' at all (the *null hypothesis*). Then, the number of expected $t\bar{t}$ is $N_{\text{exp}} = \epsilon_{t\bar{t}} \cdot N_{t\bar{t}}$. N_{exp} will fluctuate according to a Poisson distribution. For large numbers the uncertainty is given by:

$$\Delta N_{\text{exp}} = \sqrt{N_{\text{exp}}}$$

We can ask now the question how many events we may count additionally to an expected number N_{exp} of $t\bar{t}$ events after Z' reconstruction and still trust the null hypothesis. Obviously, the number of $t\bar{t}$ we really observe is given by

$$N_{\text{obs}} = N_{\text{exp}} + \eta \cdot \Delta N_{\text{exp}} = N_{\text{exp}} + \eta \cdot \sqrt{N_{\text{exp}}} \Rightarrow \eta = \frac{N_{\text{obs}} - N_{\text{exp}}}{\sqrt{N_{\text{exp}}}}$$

η is called the *discovery potential*, it tells you the number of standard deviations the observed number of events is lying above the expected number of events.

Exercise 8

Extract N_{obs} and N_{exp} from your analysis and calculate the discovery potential for a $1000\,\text{GeV}$ Z' at an integrated luminosity of $3\,\text{fb}^{-1}$.

Plot the function "discovery portential vs. integrated luminosity". Estimate your value for N_{exp} and N_{obs} from $\sigma_{t\bar{t}}$ and the reconstruction efficiency. Use a `TF1` with draw option `"AL"`.

Hint: The overall efficiencies for sequential cuts always factorize, i. e. $\epsilon = \epsilon^{\text{C1}} \cdot \epsilon^{\text{C2}}$.

4.8 Cross section limits

Generally, the number of events for a given process depends on both the cross section σ and the integrated luminosity L:

$$N = \sigma \cdot L$$

The number of observed events N_{obs}, i.e. after all cuts have been applied, depends on the reconstruction efficiency ϵ, so the cross section can be reconstructed via

$$\sigma = \frac{N_{\text{obs}}}{\epsilon \cdot L} \tag{2}$$

The number of observed evets will not only contain signal events but also events from background processes. Let's now assume we have counted N_{obs} events. This is a fixed number. This number will somehow split into

$$N_{\text{obs}} = N_{\text{obs},t\bar{t}} + N_{\text{obs},Z'} \tag{3}$$

where $N_{\text{obs},t\bar{t}}$ and $N_{\text{obs},Z'}$ are completely unknown. But we can initially set $N_{\text{obs}} \stackrel{!}{=} N_{\text{obs},t\bar{t}} = N_{\text{exp},t\bar{t}}$ and increase N_{obs} in a way that it is still compatible with the expected number of $t\bar{t}$ events up to a certain confidence level. So we can write:

$$N_{\text{obs}}(\mu, \text{CL}) = N_{\text{exp},t\bar{t}} + N_{\text{max},Z'} \tag{4}$$

$N_{\text{obs}}(\mu, \text{CL})$ is iterativelly increased until a value is found where the cumulative Poissonian PDF exceeds the CL with the given $\mu = N_{\text{exp},t\bar{t}}$ taken from your previous analysis. With $N_{\text{max},Z'} = N_{\text{obs}} - N_{\text{exp},t\bar{t}}$ you obtain the upper bound for $\sigma(Z')$ for your desired CL using (2).

Exercise 9

Plot a function "CL vs. upper bound cross section" in the range of CL $= [60 \ldots 99\,\%]$. Use a `TGraph` and add x, y–values in a `for` loop. Use the draw option `"AL"`.

4.9 Determination of the cross section

In the SM gauge boson extension model for the Z' that we considered in the exercises before, the cross section of $\sigma(Z' \to t\bar{t}) = 0.634\,\text{fb}$ is by far too low to detect any Z' peak above the $t\bar{t}$ background with our given luminosity of $3\,\text{fb}^{-1}$; we would have to increase the amount od available MC samples strongly to do that exercise. But instead, we could assume a different model where there is some other resonant decay at $M_{\text{inv}} \approx 1000\,\text{GeV}$ with a much higher cross section. Then we could see a peak and subtract the background to extract the signal with a few hundreds of pb^{-1}.

For this purpose, we prepared a "fake" data sample `data_sample.root` which consists of a $t\bar{t}$ and Z' events, randomly mixed. The sample contains $300\,\text{pb}^{-1}$ of data and was assembled with the standard $t\bar{t}$ cross section of 220.66 pb and an unknown fraction of Z' events.

To estimate the cross section, we have somehow to recover the signal from the sample. The recovery procedure might be as follows:

1. To the MC $t\bar{t}$ distribution of the invariant Z' mass a suitable function is fitted to obtain a parameterized description of the background (a Landau distribution in this case)
2. The fit function is scaled from the luminosity of the background sample to the luminosity of the data sample (in this case: $R = \frac{300\,\text{pb}^{-1}}{250\,\text{pb}^{-1}}$)
3. With the scaled Landau, the background is removed from the data sample bin by bin, i. e. the function value at the bin center is subtracted from the bin content and a new histogram is filled with those differences. If the difference is smaller then zero, the new bin content is set to zero.
4. The integral of the new histogram is taken as number of signal events. Then the cross section can be calculated by $\sigma = \frac{N_{\text{sig}}}{L \cdot \epsilon_{Z'}}$. $\epsilon_{Z'}$ had been calculated from the MC Z' sample in exercise 1.

To estimate the uncertainties, we assume both the Landau value $f(x)$ and the bin content $N(x)$ at position x to be Poissonian distributed. Then the uncertainty of the recovered Z' bin $N'(x)$ is $\Delta N'(x) = \sqrt{f(x) + N(x)}$. If we consider the bins to be uncorrelated, the uncertainty of the whole sum of bin contents Σ is (error propagation):

$$\Delta\Sigma = \sqrt{\sum_{i=1}^{\text{bins}} (\Delta N'(x_i))^2} = \sqrt{\sum_{i=1}^{\text{bins}} (f(x_i) + N(x_i))}$$

Exercise 10

Make a histogram of the invariant Z' mass of the $t\bar{t}$ sample. Fit a Landau distribution to the histogram.

Make a histogram of the invariant Z' mass spectrum of `data_sample.root`. Scale the Landau fit by the ratio of luminosities.

With the scaled Landau fit, go through the data histogram bin by bin and subtract the function value at the bin center from the bin content. Fill a new histogram with that difference bin by bin. Take care that you set equal bin numbers for all histograms!

Calculate the cross section of the "unknown" resonance and its uncertainty.

5 Going more into detail

If you are now at the point that you have completed all exercises we prepared for you, you might be willing to refine your analysis. Here are some suggestions how you might go on:

- Beside the standard CSC cuts, there are additonal cuts (C2 to C6) to improve W boson and top quark purity, initially proposed for top quark mass measurements (see [2]).
  Apply those cuts (or at least some of them) and study how your analysis performs.
- Compare the performance of the χ^2 and geometric W boson reconstruction method to the method of "best W mass reconstruction".
- For the estimation of the upper cross section bound of Z' production, a more sophisticated method had been proposed in [3]. Apply this method and compare the result to the simple method used in this exercise.
- Play around with ROOT colors, fill styles, legends, error bands etc. to embellish your plots.

And, last but not least, you are kindly invited to prepare a presentation for the students' report session tomorrow... :)

References

[1] Thomas G. Rizzo
Z' Phenomenology and the LHC
arXiv:hep-ph/0610104

[2] M. Bosman, P. Ferrari
Expected performance of the ATLAS experiment: detector, trigger and physics
arXiv:0901.0512

[3] E. Cogneras, D. Pallin
Generic $t\bar{t}$ resonance search with the ATLAS detector
ATL–PHYS–PUB–2006–033