PROCEEDINGS OF SCIENCE

A posteriori inclusion of PDFs in NLO QCD final-state calculations: The APPLGRID Project

---

Tancredi Carli

CERN, Department of Physics, Geneva, Switzerland

Dan Clements,

University of Glasgow, Glasgow, United Kingdom

Amanda Cooper-Sarkar, Claire Gwenlan,

University of Oxford, Oxford, United Kingdom

Gavin P. Salam,

LPTHE, Universities of Paris VI and VII and CNRS, 75005, Paris, France

Frank Siegert,

IPPP, University of Durham, Durham, United Kingdom

Pavel Starovoitov,

National Science and Education Center of Particle and High Energy Physics, Minsk, Belarus

Mark Sutton[*]
University of Sheffield, Sheffield, United Kingdom
E-mail: `sutt@cern.ch`

The calculation of cross-sections at Next-to-Leading order in QCD involves the integration over the final state phase space in order to cancel the infra-red divergences. For the calculation of cross sections for jet observables in deep-inelastic scattering or at hadron-hadron colliders this integration requires the Monte Carlo generation of a large number of event weights, and must be repeated for any calculation with a different choice of parton densities within the proton or different choice of factorisation or renormalisation scale. This makes the full calculation with many of the available parton density function error sets, or any iterative fit of the parton densities themselves, prohibitive in terms of the processing time required. A method for the *a posteriori* inclusion of the parton densities in the calculation is presented. In this method, the Monte Carlo weights from the integration over the hard-subprocess phase space are stored in a look-up table so that the full calculation need be performed only once, after which the cross section can be obtained with any parton density set by a fast convolution with the stored weights. A detailed example from inclusive jet production at the LHC is presented.

XVIII International Workshop on Deep-Inelastic Scattering and Related Subjects, DIS 2010
April 19-23, 2010
Firenze, Italy

---

[*] Speaker.

PoS(DIS 2010)051

1. Introduction

The calculation of cross-sections at Next-to-Leading order (NLO) in QCD involves numerical integration over the phase space of the final state partons in order to cancel the infra-red and collinear divergences. For the calculation of cross sections for jet observables in deep-inelastic scattering or at hadron-hadron colliders this integration is usually performed using the Monte Carlo generation of large numbers of event weights. Any cross section with a different choice for the proton parton density functions (PDFs) or a different choice of factorisation or renormalisation scale requires the complete calculation to be performed again. Many PDF sets are now available that include several different subsets to enable the study of the uncertainties arising from the PDF itself. To make use of these sets typically requires a calculation to be performed many times, once for each PDF subset within the set. These PDF sets are typically produced using an iterative fit over the available data, which requires the cross section to be calculated at each point in the parameter space. Including data from inclusive DIS processes in such a fit, is possible, since the cross section is often available in closed form, but the rigorous inclusion of jet data would require the full calculation which is prohibitive in terms of the required processing time.

One possible solution is to make use of the factorisation of the cross section into the parton distributions and hard subprocess to store the weights generated in the integration over the phase space in a lookup table referenced by momentum fraction x and momentum transfer Q. This means the full calculation need be performed only once to fill the lookup table. At any stage after this the convolution can be performed independently, using any choice of PDF in only a short time.

Techniques for this approach are reasonably mature [1, 2, 3]. This paper discusses recent developments in one such implementation - the open source APPLGRID project [4], which shares a similar approach to the well known fastNLO project [2] that was developed concurrently. In contrast to the fastNLO approach however, APPLGRID adopts a fully open source approach, enabling users to generate their own lookup tables for jet production and Electroweak boson production in hadron-hadron collisions, rather than limiting the user to previously calculated cross sections. In addition, arbitrary variation of the renormalisation and factorisation scales is possible within APPLGRID, whereas in fastNLO, the user is limited to a small number of combined renormalisation and factorisation scale choices implemented by the fast NLO authors when the grids were generated.

2. Overview of the method

In hadron-hadron collisions, the generic cross section can be written as a convolution of the large distance parton densities, f_1 and f_2, from the colliding hadrons, with the short distance process, $\hat{\sigma}$,

$$d\sigma = \int_{\Omega} \mathrm{d}\Omega \sum_{p} \alpha_s^{p}(Q_R^2)\ f_1(x_1, Q_F^2)\ f_2(x_2, Q_F^2)\ \hat{\sigma}_p(x_1, x_2, Q_R^2). \tag{2.1}$$

The integration over the phase space is generally achieved numerically by summation over a large number of hard subprocess event weights sampled from the phase space,

$$d\sigma = \sum_{p} \sum_{m=1}^{N} w_m^{(p)} \left(\frac{\alpha_s(Q_m^2)}{2\pi} \right)^{p} f_1(x_{1m}, Q_m^2)\ f_2(x_{2m}, Q_m^2). \tag{2.2}$$

Taking advantage of symmetries in the hard process it is possible to replace the summation over the 13×13 possible initial state parton products, $f_1(x_1)f_2(x_2)$ with a summation over a smaller number, M, of linear combinations of the $f_1(x_1)$ and $f_2(x_2)$ states,

$$\sum_{ij=q,\bar{q},g} w_{ij}\, f_{1i}(x_1)\, f_{2j}(x_2) = \sum_{k=1}^{M} w^{(k)} F^{(k)}(x_1,x_2). \tag{2.3}$$

Representing the parton distribution as a multidimensional Lagrange interpolation over n_x and n_Q nodes in x and Q^2,

$$f(x,Q^2) = \sum_{i=0}^{n_x}\sum_{j=0}^{n_Q} f(x_{(i)},Q^2_{(j)})\; I_i^{(n_x)}(x_{(i)},x)\; I_j^{(n_Q)}(Q^2_{(j)},Q^2),$$

and substituting into equation 2.3, by changing the order of the summations, it is possible to perform the summation over the weights instead using the product of the interpolating coefficients with the weights themselves so that the last summation over the M independent linear combinations of the parton densities can carried out after the summation over the full calculation of event weights,

$$\sum_{m=1}^{N} w_m^{(p)(k)} \left(\frac{\alpha_s(Q^2_m)}{2\pi}\right)^{p} F^{(k)}(x_{1m},x_{2m},Q^2_m) \tag{2.4}$$

$$\rightarrow \sum_{i_1 i_2 j} F^{(k)}(x_1^{(i_1)},x_2^{(i_2)},Q^2_{(j)}) \left(\frac{\alpha_s(Q^2_{(j)})}{2\pi}\right)^{p} \sum_{m}^{N} w_m^{(p)(k)} I_{i_1}(x_{1m}) I_{i_2}(x_{2m}) I_j(Q^2_m) \tag{2.5}$$

$$= \sum_{i_1 i_2 j} F^{(k)}(x_1^{(i_1)},x_2^{(i_2)}Q^2_{(j)}) \left(\frac{\alpha_s(Q^2_{(j)})}{2\pi}\right)^{p} W^{(p)(k)}_{i_1 i_2 j}. \tag{2.6}$$

Here, the W structure consists of a set of 3-Dimensional "grids" in the variable x_1, x_2 and Q^2, with one such grid for each of the M linear combinations of the parton densities at each order. For a differential cross section, this structure will be repeated for each bin in the observable.

It is also possible to isolate the variation of the renormalisation and factorisation scales, and the centre-of-mass energy of the colliding hadrons and allow each of them to also to be modified during this external convolution stage.

Different processes are handled using different linear combinations, F, of the PDFs to take account of the symmetries in the different sub-process matrix elements.

3. Implementation

The APPLGRID code itself is implemented as a library of C++ classes for creating and interacting with these grids. In addition a basic Fortran interface is also available which presents a subset of the full C++ functionality to the user. Modified versions of NLOjet++ [5] and MCFM [6] are also available to enable users to generate their own grids for jet production and Electroweak boson production respectively, although only the basic grid classes are required to simply read pre-existing grids.

The accuracy of the *a posteriori* convolution can be controlled by the number of nodes chosen in x and Q^2 and also the order of the interpolation. In addition, a variable transform is used to allow

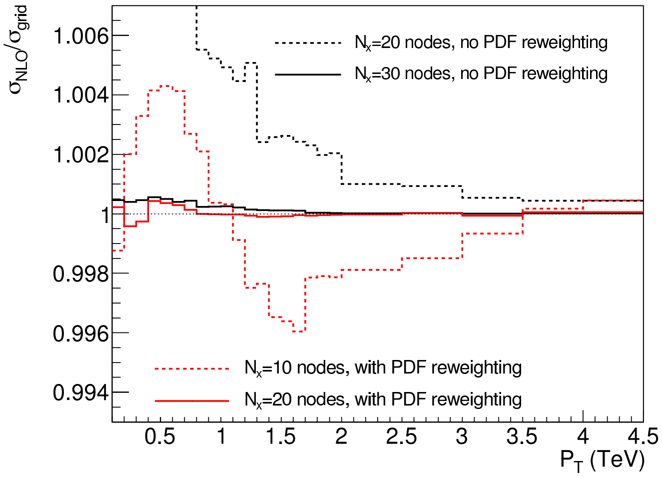

Figure 1: The effect of reweighting the PDF in the convolution for inclusive jet production at central rapidities. The black dashed and solid lines show the ratio for 20×20 nodes and 30×30 nodes in x respectively using a 5th order interpolation and no reweighting. The red dashed and solid lines show the results of a 20×20 and 10×10 grids respectively when including the PDF reweighting.

uniform node spacing without loss of precision, and a PDF reweighting factor can be applied to improve the accuracy of the interpolation when using fewer interpolation nodes. An example of the use of the PDF reweighting can be seen in figure 1. This shows the ratio between the full calculation and that resulting from the *a posteriori* convolution for the inclusive jet cross section at the LHC, with different scenarios of the number of grid nodes in each x variable. Here, the improvement in accuracy of the reproduction of the full calculation when including the PDF reweighting can be clearly seen, with the accuracy of the convolution performed with 20×20 nodes in each x variable, including the reweighting, being comparable or better than that with no reweighting, and 30×30 grid nodes in x. An additional feature of note from figure 1 is the observation of the general feature that at lower transverse energy, the cross section is described less well by the *a posteriori* convolution. This is due to the larger ranges of x contributing to the cross section in this region, since in this kinematic regime, both high and and low x partons can interact in the hard process.

The grids themselves make use of a custom sparse data structure to store the weights. This structure includes a zero suppression so that weights of zero need not be stored. This reduces the memory footprint for a typical grid by around 70%. The grids encode everything that is required to perform the convolution, with the exception of the PDFs themselves and the function required for the calculation of α_s, which must be provided by the user when performing the convolution. This makes the generation of the cross section from the grid very simple, with the reading of the grid from a file, and the convolution itself, which can return the results either as an STL vector or a histogram, possible with only two lines of code.

The convolution time scales linearly with the sum of the number of grid nodes in x_1 and x_2 rather than as their product, since the slowest contribution is the actual calculation of the PDFs on the grid nodes rather than the convolution itself. Knowledge of which elements are zero allows a further 40% reduction in the convolution time since weights which are known to be zero need not be interrogated.

4. Examples

A more detailed case study of the use of APPLGRID in the calculation of the inclusive jet production cross section, differential in jet transverse momentum p_T, is presented here. The cross

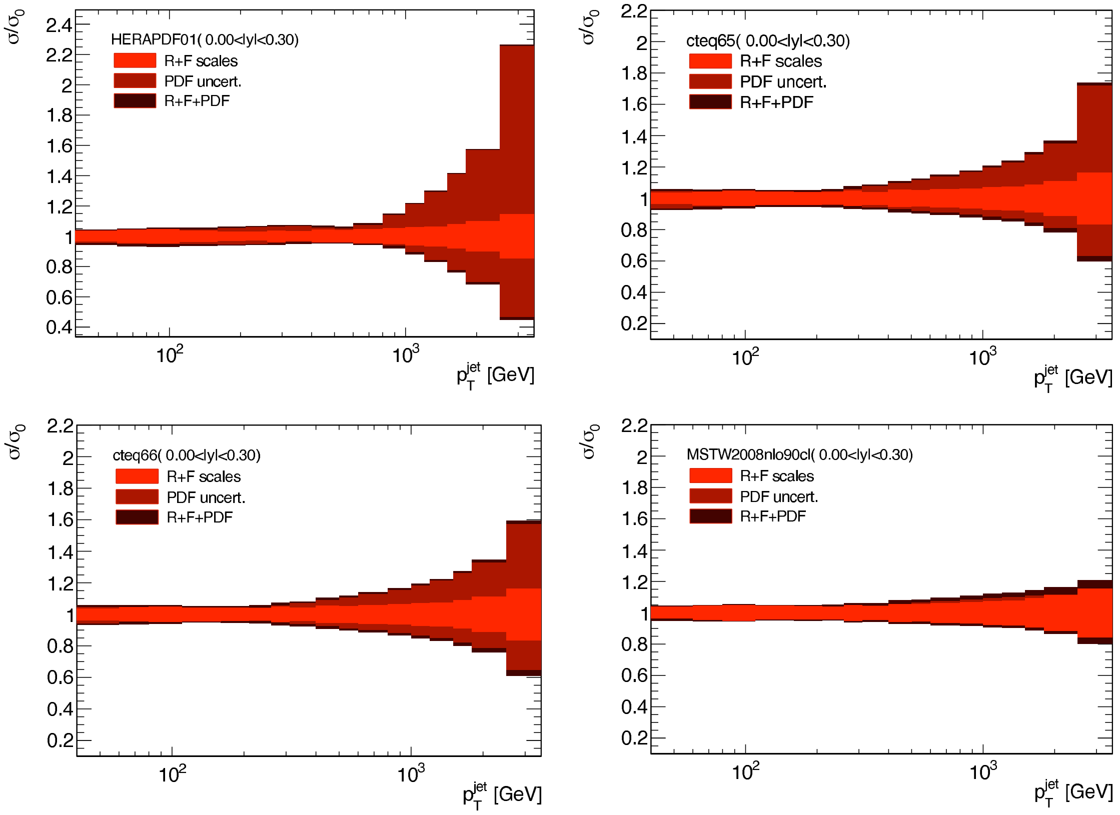

Figure 2: The uncertainty arising from the PDF set and the variation of the renormalisation and factorisation scales for the inclusive jet cross section at central rapidity at the LHC.

section defined has 17 bins in p_T, in five different regions of the jet rapitity, y. For the initial calculation around 10 million events were generated with a total processing time of between 10-15 hours. Using a grid with 30 nodes in x_1, x_2 and Q^2 and a 5th order interpolation, the size of the grid on disk is around 3 Mbytes. Using the custom sparse structure, this reduces to around 1 Mbyte in memory so that the full calculation can be reproduced to within 0.1% accuracy, and takes around 43 ms per rapidity region.

Figure 2 shows the scale and pdf uncertainties for 4 different PDF error sets - the CTEQ6.5 [7], CTEQ6.6 [8], and MSTW2008 [9] sets together with a recent fit to the combined HERA data [3]. Here each of the PDF sets in the full error sets have been used, and the convolution performed separately for each, together with the simultaneous variation of the renormalisation and factorisation scales. Both the CTEQ6.5 and MSTW2008 contains 41 subsets, the central sets plus 40 eigenvector sets that encapsulate the uncertainty on the PDF, one set for each of the up and down excursions for each of 20 eigenvectors from the fit. The CTEQ6.6 set contains 45 sets in total.

Figure 3 shows an example of rescaling the centre-of-mass energy for the cross sections for p_T(jet)>40 and p_T(jet)>60 GeV for each of the five rapidity regions, shown as a function of 50 values of the cms energy from 7 TeV to 14 TeV.

Figure 4 shows a detailed study of the effects of independently varying the renormalisation and factorisation scales with a high granularity, for 20 different values in both scales for central jet

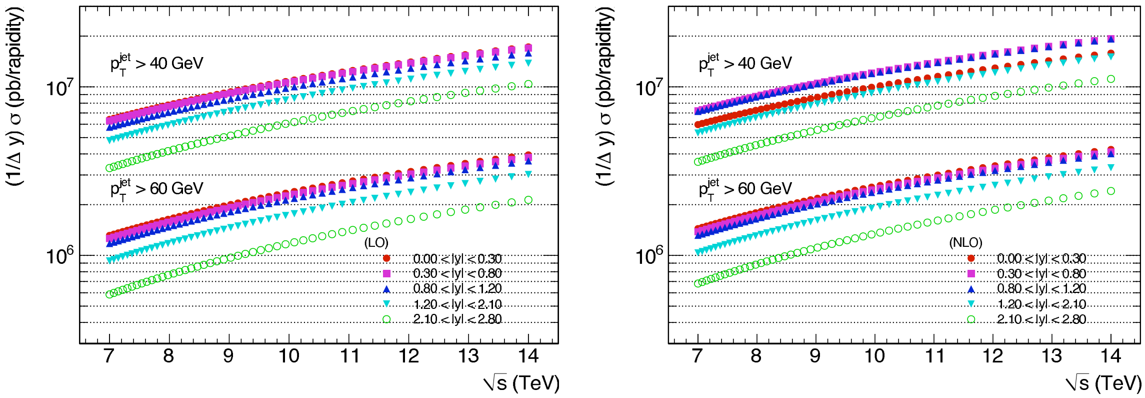

Figure 3: The cross section for inclusive jet production in five regions of jet rapidity as a function of the centre-of-mass of the colliding hadrons. On the left is the Leading order cross section, on the right, the cross section at NLO.

production at low, and intermediate transverse momenta. This represents 400 separate evaluations of the cross section, one for each scale combination, so for each rapidity region the total time for the convolution is less than 20 seconds for the entire plane. This compares favourably to a time greater than six months that would be required for running the full calculation separately for each point in the scale plane on a single machine, or around one week if making use of the ability of NLOjet++ to calculate the cross section with multiple scales simultaneously.

4.1 Interface with fastNLO

Built into the APPLGRID code is an interface capable of directly reading the fastNLO grids. This enables the APPLGRID code to perform the convolution for the fastNLO hadron-hadron and DIS jet cross sections and provide results which are numerically exact with respect to performing the same convolution with fastNLO itself. This for example, allows both Electroweak boson production from APPLGRID together with the inclusive jet production at the LHC including the threshold resummation of Kidonakis and Owens [10] from fastNLO to be calculated within a single framework.

In addition, the interface allows addition functionality over and above that provided by fastNLO since it allows the beam energy rescaling, and also the arbitrary variation of renormalisation and and factorisation scales discussed earlier. As a caveat here it should be noted that the fastNLO grids typically use only 12 interpolation nodes in x, and two in Q^2 for cross sections at the LHC, and only one in Q^2 for the Tevatron jet cross sections, so accuracy may be an issue when using beam energy scaling with the fastNLO grids.

5. Summary and Outlook

The APPLGRID project is an open source project for the generation and fast convolution of grids for storing weights from NLO cross section calculations of both jet production and Electroweak boson production at hadron-hadron colliders. Accuracy of better than 0.1% can be achieved with reasonably small grids and a typical convolution time of a few tens of milliseconds.

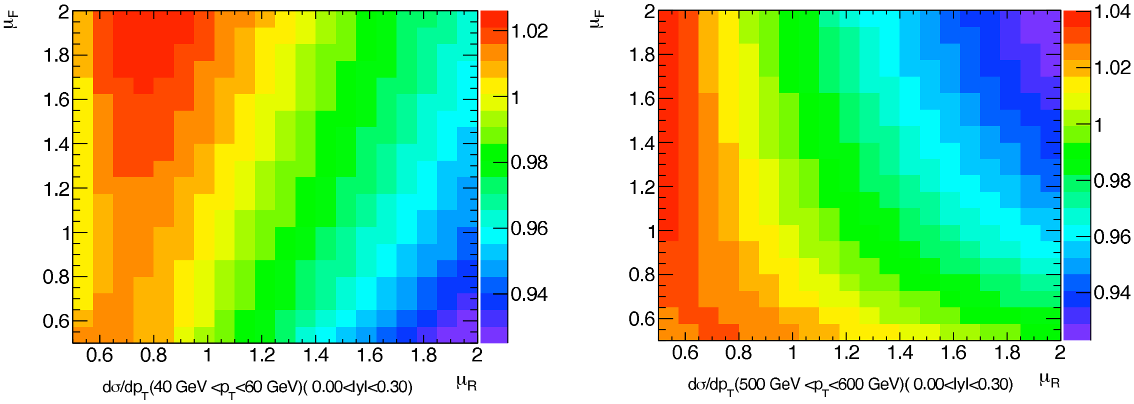

Figure 4: The ratio of the cross section for central jet production at the LHC, calculated with an independent fractional variation of the renormalisation and factorisation scales, to the cross section when each scale is set to the hard scale in the calculation. The left figure shows the cross section for at low transverse momenta, the right figure, shows that for intermediate transverse momenta.

Use of a custom data structure allows both the memory occupancy of the grid and the time required by the convolution to be reduced. The lookup tables provide a powerful tool for the study of the PDF and scale uncertainties and for the prediction of the cross section at different beam energies. Being a completely open source project, the full code is available, and can be downloaded from hepforge [11].

The first collisions at 7 TeV have recently been achieved by the LHC. The APPLGRID project is already being used by the ATLAS Collaboration as a standard tool for the calculation of the NLO cross section as part of the measurement of the jet production cross section with this first data.

References

[1] T. Carli, G. P. Salam and F. Siegert, Proceedings *HERA and the LHC: A Workshop on the Implications of HERA for LHC Physics*, Geneva (Switzerland) hep-ph/0510324 (2005).

[2] T. Kluge, K. Rabbertz and M. Wobisch, Proceedings of the *DIS 2006 conference*, Tsukuba (Japan), hep-ph/0609285 (2006).

[3] H1 and ZEUS Collaborations, JHEP 1001:109 (2010).

[4] T. Carli *et al*. Eur. Phys. J. **C66** (2010) 503-524.

[5] Z. Nagy, Phys. Rev. **D68** (2003) 094002; Z. Nagy, Phys. Rev. Lett. **88** (2002) 122003; Z. Nagy and Z. Trocsanyi, Phys. Rev. Lett. **87** (2001) 082001.

[6] J. M. Campbell and R. K. Ellis, Phys. Rev. **D60** (1999) 113006; J. M. Campbell and R. K. Ellis, Phys. Rev. **D62** (2000) 114012.

[7] W. K. Tung, *et al* JHEP 0702:053 (2007).

[8] P. M. Nadolsky *et al*, Phys. Rev. **D78** (2008) 013004.

[9] A. D. Martin *et al*, Eur. Phys. J. **C63** (2009) 189-285.

[10] N. Kidonakis and J. F. Owens, Phys. Rev. **D63** (2001) 054019.

[11] Hepforge, `http://projects.hepforge.org`.