PROCEEDINGS OF SCIENCE

CT14 Monte-Carlo parton distributions with positivity and asymmetric uncertainties

Tie-Jiun Hou[*a], **Jun Gao**[b], **Joey Huston**[c], **Pavel M. Nadolsky**[a], **Carl Schmidt**[c], **Daniel Stump**[c], **Bo-Ting Wang**[a], **Ke Ping Xie**[a], **Sayipjamal Dulat**[c,d], **Jon Pumplin**[c], **C.-P. Yuan**[c]

[a] *Department of Physics, Southern Methodist University, Dallas, TX 75275-0181 U.S.A.*
[b] *High Energy Physics Division, Argonne National Laboratory, Argonne, Illinois, 60439 U.S.A.*
[c] *Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824 U.S.A.*
[d] *Center for Theoretical Physics, Xinjiang University, Urumqi, Xinjiang 830046 China*
E-mail: `tiejiunh@mail.smu.edu`

In the context of the CT14 analysis of parton distribution functions, we propose a new technique for generation of Monte-Carlo replica sets from Hessian eigenvector sets. The technique preserves non-negative behavior and asymmetry properties of Hessian eigenvector sets. The generated ensemble of Monte-Carlo replicas is suitable for various applications, such as Bayesian reweighting and combination of PDF ensembles.

XXIV International Workshop on Deep-Inelastic Scattering and Related Subjects
11-15 April, 2016
DESY Hamburg, Germany

*Speaker.

PoS(DIS2016)034

Modern parton distribution functions (PDFs) [1, 2, 3, 4, 5, 6] with trustworthy estimates of uncertainties serve as main inputs for numerous theoretical predictions and measurements at hadron colliders. Cumulative uncertainties on the PDFs can be estimated using several techniques, including the Hessian [7], Monte Carlo(MC) [8, 9], Lagrange multiplier [10] and offset methods [11]. While the Hessian method is very efficient for propagating PDF uncertainties assuming the Gaussian approximation, the Monte-Carlo replica method is very general and can be used for powerful statistical applications, such as inclusion of new experimental data sets into the PDF fits by Bayesian reweighting [8, 13, 14] and combination of PDF ensembles obtained in different approaches [15, 16]. Sometimes it is necessary to convert the PDF error sets obtained via the Hessian eigenvector analysis, such as CT14 or MMHT14 error PDFs, into a large ensemble of MC replica PDFs. Watt and Thorne [12] proposed a procedure for doing this, sufficient for reproducing symmetric uncertainties of the input PDFs in well-constrained $\{x,Q\}$ regions. To retain more features of the Hessian ensemble, the procedure needs an extension beyond the linear approximation adopted by the Watt-Thorne approach. For instance, the CTEQ parametrizations are constructed to be non-negative in order to guarantee positive physical cross sections. In Ref. [17], we outlined a general method for constructing an MC replica ensemble that reproduces the asymmetric uncertainties and positivity constraints of the input Hessian ensemble. We will now summarize this method.

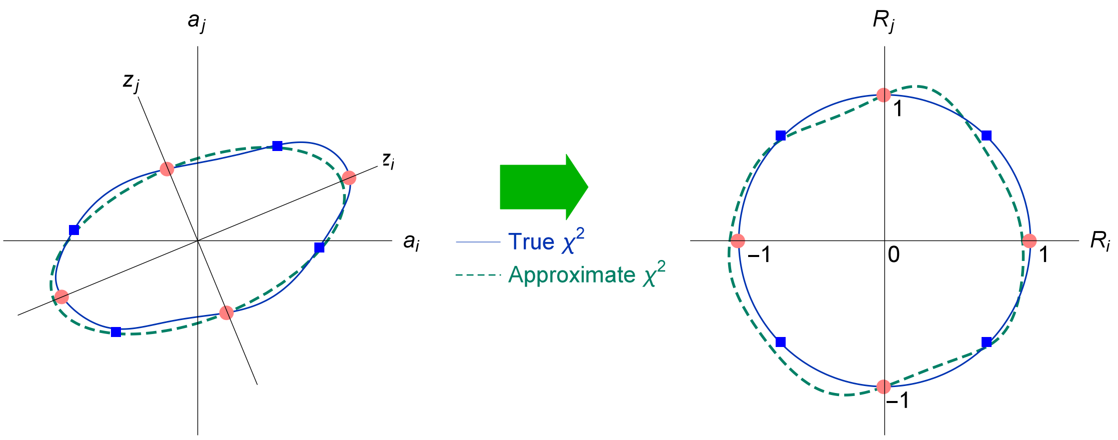

Figure 1: Contours of the constant χ^2 in planes of physical (a_i) and orthonormal (R_i) PDF parameters. The solid and dashed contours are for the exact χ^2 and its Gaussian approximation, respectively. The red (round) and blue (square) dots indicate extreme displacements along the eigenvector directions and diagonal directions.

Recall that the uncertainty estimation in the Hessian method with D PDF parameters is done by introducing an ensemble of $2D$ eigenvector sets that delineate the boundary of the uncertainty region around the best-fit combination of the PDF parameters, at a given confidence level (c.l.). These eigenvector sets allow one to calculate the PDF uncertainty on a QCD observable using one of the available master formulas [7, 18, 19]. If we notice that the probability distribution in a typical PDF fit is close to a Gaussian one near the minimum of the log-likelihood, one can write

$$Probability(\{R\}) \sim e^{-\sum_{i=1}^{D} R_i^2/2}, \tag{1}$$

where we introduced independent PDF parameters R_i ($i = 1,...,D$) such that each R_i varies along the i-th eigenvector direction and takes the values of zero at the minimum χ_0^2 of χ^2, and ± 1 at the boundary of the confidence region. The log-likelihood χ^2 is an approximately quadratic function of R_i in this representation:

$$\chi^2 \approx \chi_0^2 + \sum_{i=1}^{D} R_i^2. \tag{2}$$

The transformation from the physical PDF parameters a_i to the parameters R_i is illustrated in Fig. 1. We can parametrize the exact ("true") χ^2 returned by the global fit so that the confidence region boundary, indicated by the solid contour, is essentially a sphere of a unit radius in space of the R parameters. This representation holds even if the true χ^2 is not perfectly symmetric, e.g., if it deviates from the symmetric Gaussian approximation for χ^2, corresponding to the dashed contour of the constant χ^2 in the figure.

The PDF uncertainty on a QCD function X, such as a PDF or a cross section, is usually estimated from the maximal variation of $X(R)$ within the 68% (or 90%) c.l. region for the true χ^2. We can expand X in a Taylor series around its value $X(\{0\}) \equiv X_0$ at the minimum of χ^2,

$$X(\{R\}) = X(\{0\}) + \sum_{i=1}^{D} \frac{\partial X}{\partial R_i} R_i + \frac{1}{2} \sum_{i=1}^{D} \frac{\partial^2 X}{\partial R_i \partial R_j} R_i R_j + ..., \tag{3}$$

and approximate the derivatives of X by finite-difference formulas [17] in terms of the X values on the boundary, *i.e.*, on the solid contour in Fig. 1. To compute the first-order derivatives $\partial X/\partial R_i$ and second-order *diagonal* derivatives $\partial^2 X/\partial R_i^2$, it suffices to know X for R at the red circles, that is, to perform $2D$ evaluations of $X_{\pm i} \equiv X(0,0,...,R_i = \pm 1,...0)$. The non-diagonal derivatives $\partial^2 X/(\partial R_i \partial R_j)$ require $2D(D-1)$ additional evaluations of X (or PDFs) for $\{R_i, R_j\} = \{\pm 2^{-1/2}, \pm 2^{-1/2}\}$ at the blue squares, and these are usually not provided by the Hessian error PDFs. We can nevertheless get an idea about the asymmetry of $X(R)$ by including the diagonal derivatives $\partial^2 X/\partial R_i^2$ into the PDF uncertainty formulas.

This brings us to the asymmetric PDF uncertainties, given in terms of predictions for the Hessian eigenvector sets by [19]

$$\delta_{68}^{H,>} X = \sqrt{\sum_i \left(max[X_{+i} - X_0, X_{-i} - X_0, 0]\right)^2}, \; \delta_{68}^{H,<} X = \sqrt{\sum_i \left(max[X_0 - X_{+i}, X_0 - X_{-i}, 0]\right)^2}. \tag{4}$$

To construct N_{rep} MC replicas $X^{(k)}$ that reproduce these asymmetric uncertainty bands, we generate them by

$$X^{(k)} = X(\{0\}) + d^{(k)} - \Delta, \tag{5}$$

$$d^{(k)} = \sum_{i=1}^{D} \frac{X_{+i} - X_{-i}}{2} R_i^{(k)} + \frac{1}{2} \sum_{i=1}^{D} (X_{+i} + X_{-i} - 2X_0)(R_i^{(k)})^2, \tag{6}$$

using random numbers $R_i^{(k)}$ sampled from a standard normal distribution. By setting $X = f$ or $X = \ln(f)$, we produce two ensembles of PDF replicas from the CT14 Hessian PDFs, designated as CT14 MC1 and MC2. The second one satisfies positivity, as the Hessian CT14 PDFs do. All the

PoS(DIS2016)034

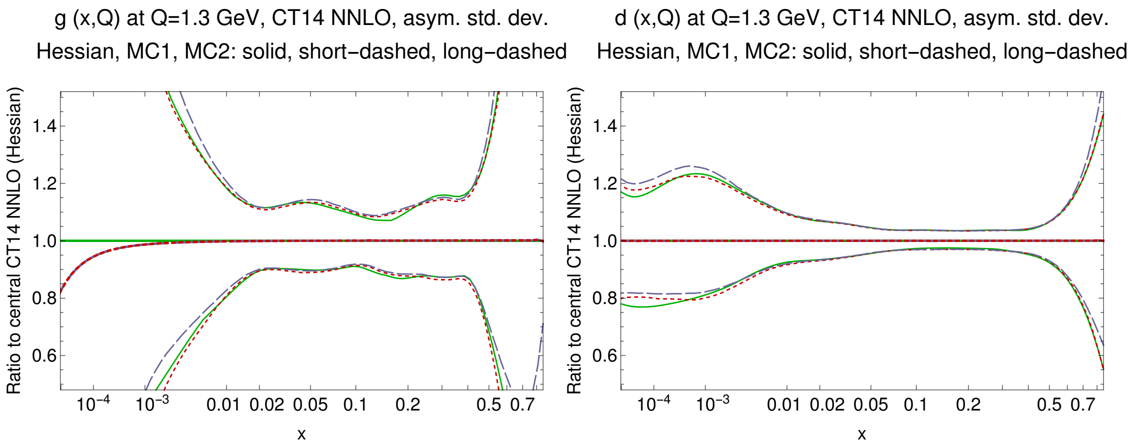

Figure 2: The mean values and asymmetric standard deviations of the CT14 NNLO MC1 (short-dashed) and MC2 (long-dashed) PDFs, compared to the mean and 68% c.l. uncertainty (solid) of the CT14 NNLO Hessian PDF. The PDFs are shown as ratios to the central CT14 fit.

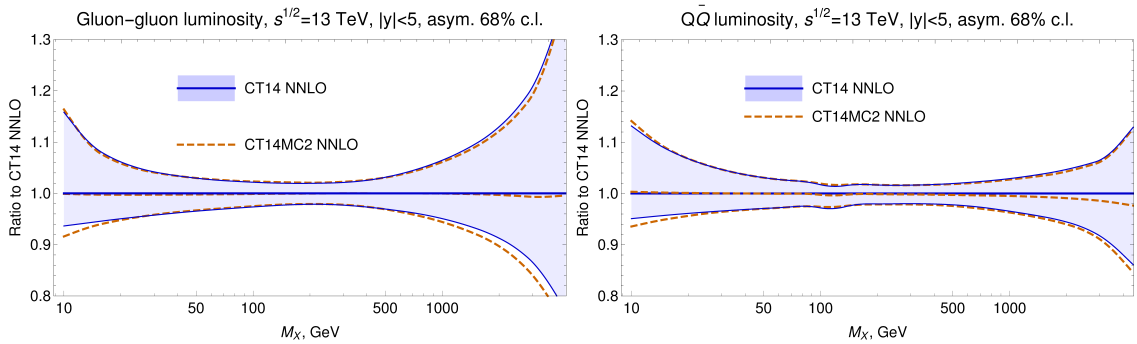

Figure 3: 68% c.l. asymmetric uncertainties of the gg and $q\bar{q}$ luminosities for the CT14 NNLO Hessian (solid) and MC2 (dashed) ensembles, computed at $\sqrt{s} = 13$ TeV with a constraint $|y| < 5$ on the rapidity of the heavy state.

replica sets are shifted by a constant amount Δ in order to set the average $\langle X \rangle = \left(\sum_{i=1}^{N_{rep}} X^{(k)}\right)/N_{rep}$ of the replicas to coincide with the best-fit value X_0 of the Hessian ensemble, regardless of N_{rep}.

As the distribution of the replicas is no longer symmetric with respect to X_0, the uncertainties are given by the standard deviations that are constructed separately from the positive and negative displacements of $X^{(k)}$ from $\langle X \rangle$:

$$\delta_{68}^{MC,>} X = \sqrt{\langle (X - \langle X \rangle)^2 \rangle_{X > \langle X \rangle}}, \; \delta_{68}^{MC,<} X = \sqrt{\langle (X - \langle X \rangle)^2 \rangle_{X < \langle X \rangle}}. \tag{7}$$

Fig. 2 illustrates on the example of the $g(x,Q)$ and $d(x,Q)$ distributions that the asymmetric standard deviations (7) of both the CT14 MC1 and MC2 NNLO ensembles reproduce well the 68% c.l. asymmetric uncertainties (4) of the CT14 Hessian ensemble. In the same spirit, the CT14 MC parton luminosities obtained are very close to the Hessian ones across the LHC kinematic range, cf. Fig. 3.

In the presented approach, the MC ensemble samples the probability distribution in the $D-$dimensional parameter space that is reconstructed under the Gaussian assumption from the distribution of the Hessian error PDFs on the $(D-1)$-dimensional boundary of a confidence region

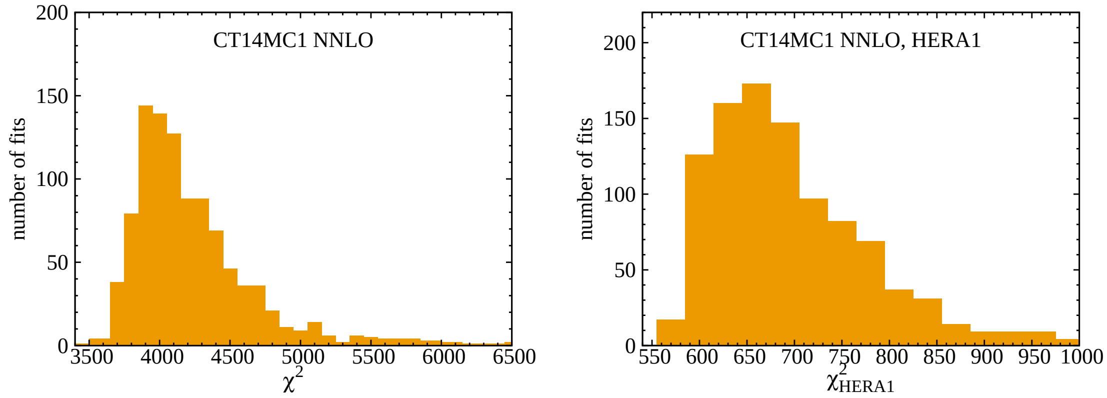

Figure 4: The χ^2 distribution of the 1000 CT14 NNLO MC replicas for the total χ^2 (left) and χ^2 of the combined HERA-I data (right).

enclosing the best fit. When the PDF uncertainties are small, we observe good agreement between the CT14 Hessian, MC1, and MC2 error bands. In the extrapolation regions, where the linear approximations cease to be adequate, differences between the Hessian, MC1 and MC2 error bands are more pronounced and reveal intrinsic ambiguities in the replica generation methods. The positivity constraint on CT14 MC2 yields more physical behavior in poorly constrained x intervals. A C++ code MCGEN to generate MC replicas with or without positivity constraints is published on HepForge. Finally, we notice that individual replicas tend to be poor fits to the hadronic data used in the global analysis; however, their averages and standard deviations provide excellent approximations for the Hessian central PDF set and 68% c.l. uncertainties. This is demonstrated in Fig. 4, showing histograms of χ^2 values for the global data (3174 data points; left panel) and for the combined HERA-1 data (579 data points [22]; right panel), for the 1000 replicas in the CT14 NNLO MC1 and MC2 ensembles. The vast majority of replicas yield very large χ^2 values for the global data, and even for the single experiment. The random fluctuations of the individual replicas, which result in large χ^2 values for any single replica, will largely cancel in the ensemble averages.

Acknowledgement. This research was supported in part by the National Science Foundation under Grants No. PHY-1410972 and PHY11-25915; by the U.S. Department of Energy under Contract No. DE-AC02-06CH11357 and Grants DE-SC0013681 and DE-SC0010129; and by the National Natural Science Foundation of China under Grant No. 11465018.

References

[1] S. Dulat *et al.*, Phys. Rev. D **93**, no. 3, 033006 (2016) [arXiv:1506.07443 [hep-ph]].

[2] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, Eur. Phys. J. C **75**, no. 5, 204 (2015) [arXiv:1412.3989 [hep-ph]].

[3] R. D. Ball *et al.* [NNPDF Collaboration], JHEP **1504**, 040 (2015) [arXiv:1410.8849 [hep-ph]].

[4] H. Abramowicz *et al.* [H1 and ZEUS Collaborations], Eur. Phys. J. C **75**, no. 12, 580 (2015) [arXiv:1506.06042 [hep-ex]].

[5] S. Alekhin, J. Blumlein and S. Moch, Phys. Rev. D **89**, no. 5, 054028 (2014) [arXiv:1310.3059 [hep-ph]].

[6] A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens and N. Sato, Phys. Rev. D **93**, no. 11, 114017 (2016) [arXiv:1602.03154 [hep-ph]].

[7] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai and W. K. Tung, Phys. Rev. D **65**, 014013 (2001) [hep-ph/0101032].

[8] W. T. Giele and S. Keller, Phys. Rev. D **58**, 094023 (1998) [hep-ph/9803393].

[9] W. T. Giele, S. A. Keller and D. A. Kosower, hep-ph/0104052.

[10] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai and W. K. Tung, Phys. Rev. D **65**, 014012 (2001) [hep-ph/0101051].

[11] A. M. Cooper-Sarkar, J. Phys. G **28**, 2669 (2002) [hep-ph/0205153].

[12] G. Watt and R. S. Thorne, JHEP **1208**, 052 (2012) [arXiv:1205.4024 [hep-ph]].

[13] R. D. Ball *et al.* [NNPDF Collaboration], Nucl. Phys. B **849**, 112 (2011) Erratum: [Nucl. Phys. B **854**, 926 (2012)] Erratum: [Nucl. Phys. B **855**, 927 (2012)] doi:10.1016/j.nuclphysb.2011.03.017, 10.1016/j.nuclphysb.2011.10.024, 10.1016/j.nuclphysb.2011.09.011 [arXiv:1012.0836 [hep-ph]].

[14] R. D. Ball *et al.*, Nucl. Phys. B **855**, 608 (2012) doi:10.1016/j.nuclphysb.2011.10.018 [arXiv:1108.1758 [hep-ph]].

[15] J. Gao and P. Nadolsky, JHEP **1407**, 035 (2014) doi:10.1007/JHEP07(2014)035 [arXiv:1401.0013 [hep-ph]].

[16] S. Carrazza, J. I. Latorre, J. Rojo and G. Watt, Eur. Phys. J. C **75**, 474 (2015) doi:10.1140/epjc/s10052-015-3703-3 [arXiv:1504.06469 [hep-ph]].

[17] T. J. Hou *et al.*, arXiv:1607.06066 [hep-ph].

[18] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky and W. K. Tung, JHEP **0207**, 012 (2002) [hep-ph/0201195].

[19] P. M. Nadolsky and Z. Sullivan, eConf C **010630**, P510 (2001) [hep-ph/0110378].

[20] S. Forte, L. Garrido, J. I. Latorre and A. Piccione, JHEP **0205**, 062 (2002) [hep-ph/0204232].

[21] J. Butterworth *et al.*, J. Phys. G **43**, 023001 (2016) [arXiv:1510.03865 [hep-ph]].

[22] F. D. Aaron *et al.* [H1 and ZEUS Collaborations], JHEP **1001**, 109 (2010) [arXiv:0911.0884 [hep-ex]].

PoS(DIS2016)034