NEURAL NETWORKS, PROBABILITY DISTRIBUTIONS, AND STRUCTURE FUNCTIONS

STEFANO FORTE
I.N.F.N. ROMA III

ANTWERPEN, SEPTEMBER 16, 2002

Credits

- **Neural Network Parametrization of Structure Functions**

  S. F., Lluís Garrido, José I. Latorre and Andrea Piccione, *JHEP* **205**, 62 (2002)

- **Truncated Moments of Parton Distributions**

  S. F. and Lorenzo Magnea, *Phys. Lett.* **B448**, 295 (1999); S. F., Lorenzo Magnea, Giovanni Ridolfi and Andrea Piccione, *Nucl. Phys.* **B594**, 46 (2001); Andrea Piccione, *Phys. Lett.* **B518**, 207 (2001)

- **Unbiased Determination of α_s**

  S. F., José I. Latorre, Lorenzo Magnea and Andrea Piccione, *Nucl. Phys. B, in press*

THE NAME OF THE GAME

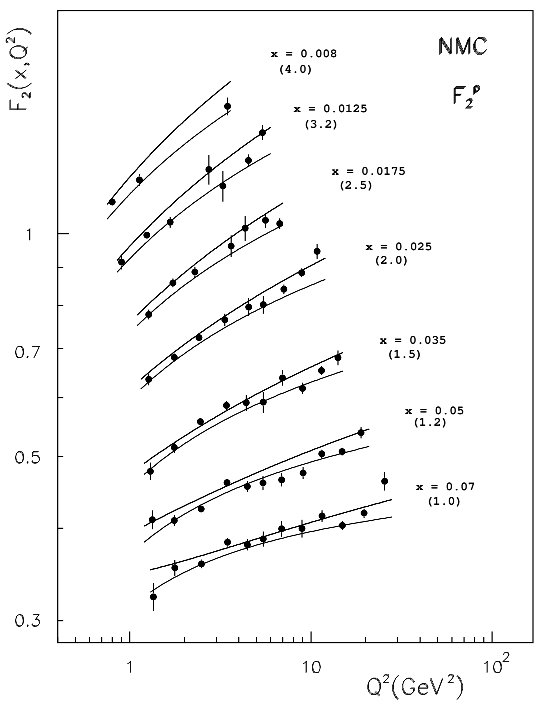
DIS DATA → STRUCTURE FUNCTIONS (FORM FACTORS, DEP. ON KIN. VARIABLES x, Q^2)

STRUCTURE FUNCTION=HARD COEFF.⊗PARTON DISTN.

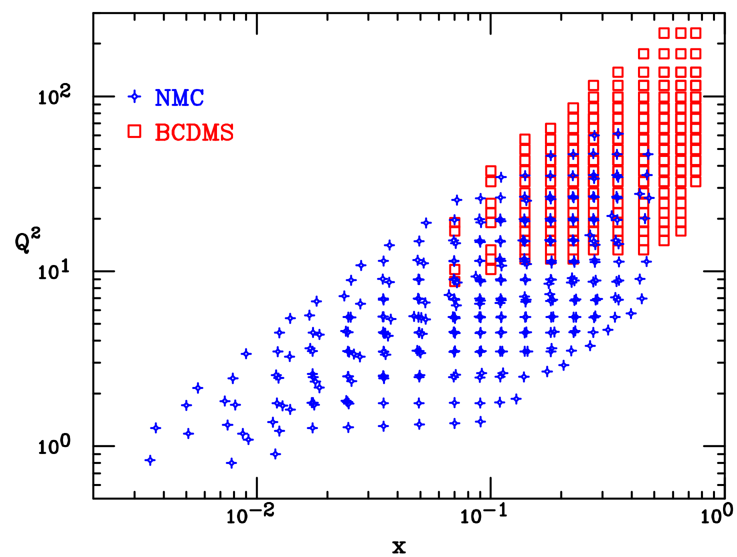
$$F_2^{NC}(x,Q^2) = x\sum_{\text{flav. } i} e_i^2(q_i+\bar{q}_i) + \alpha_s\left[C_i[\alpha_s]\otimes(q_i+\bar{q}_i) + C_g[\alpha_s]\otimes g\right]$$

- TRIVIAL COMPLICATIONS: DISENTANGLE INDIVIDUAL QUARK & GLUON CONTRIBUTION TO STRUCTURE FUNCTION; EVOLVE TO COMMON SCALE; DECONVOLUTE see below: truncated moms.
- SERIOUS COMPLICATION: DETERMINE ERROR ON FUNCTIONS $f(x)$, $f = q_i, \bar{q}_i, g$

A (MARGINALLY) SIMPLER PROBLEM: DETERMINE THE STRUCTURE FUNCTION



GIVEN A BUNCH OF EXPERIMENTAL DATA $F_2(x,Q^2)$ AT POINTS (x_i,Q_i^2), WITH STAT. ERRORS (fig.→bars) AND CORRELATED SYST. ERRORS (fig.→bands)
DETERMINE THE STRCTURE FUNCTION AND ASSOCIATE ERROR



WHAT'S THE PROBLEM? D. Kosower, 1999

- FOR A SINGLE QUANTITY, WE QUOTE 1 SIGMA ERRORS: VALUE$\pm$ ERROR
- FOR A PAIR OF NUMBERS, WE QUOTE A 1 SIGMA ELLIPSE
- FOR A FUNCTION, WE NEED AN "ERROR BAR" IN A SPACE OF FUNCTIONS

MUST DETERMINE THE PROBABILITY DENSITY (MEASURE) $\mathcal{P}[F_2]$ IN THE SPACE OF FUNCTIONS $F_2(x,Q^2)$

$\Rightarrow$ EXPECTATION VALUE OF AN OBSERVABLE $\mathcal{F}\left[F_2(x,Q^2)\right]$:

$$\left\langle \mathcal{F}\left[F_2(x,Q^2)\right]\right\rangle = \int \mathcal{D}F_2\,\mathcal{F}\left[F_2(x,Q^2)\right]\,\mathcal{P}[F_2],$$

PROBLEM: MUST DETERMINE AN INFINITE–DIMENSIONAL OBJECT FROM A FINITE SET OF DATA POINTS

SOLUTIONS. . .

- CHOOSE A FIXED FUNCTIONAL FORM, E.G. (SMC, 1998)

  $$F_2(x,Q^2) = x^{a_1}\, f(x,Q^2)$$
  $$f(x,Q^2) = A(x)\left[\frac{\log Q^2/\Lambda^2}{\log Q_0^2/\Lambda^2}\right]^{B(x)}\left[1+\frac{C(x)}{Q^2}\right]$$
  $$A(x) = (1-x)^{a_2}\left[a_3 + a_4\,(1-x) + a_5\,(1-x)^2 + a_6\,(1-x)^3 + a_7\,(1-x)^4\right]$$
  $$B(x) = b_1 + b_2\,x + \frac{b_3}{x+b_4}$$
  $$C(x) = c_1\,x + c_2\,x^2 + c_3\,x^3 + c_4\,x^4$$

  PROBLEM PROJECTED ONTO THE FINITE–DIMENSIONAL SPACE OF PARAMETERS

  WHAT IS THE BIAS (THEOR. ERROR) DUE TO THE CHOICE OF FUNCTIONAL FORM?

- EXPAND OVER A FINITE SET OF BASIS FUNCTIONS, E.G. ORTHOGONAL POLYNOMIALS (Yndurain 1975, Parisi, Sourlas 1976, Furmański, Petronzio, 1982)

  PROBLEM PROJECTED ONTO THE FINITE–DIMENSIONAL SPACE OF EXPANSION COEFFICIENTS

  WHAT IS THE BIAS (THEOR. ERROR) DUE TO THE CHOICE OF TRUNCATION?

  E.g. assume a periodic f. is expanded over a basis of ortho. polynomials, or a non-periodic f is Fourier-expanded . . .

- GENERATE A MONTE-CARLO SAMPLE OF FCTS. W. "REASONABLE" PRIOR DISTN., AND UPDATE FROM DATA USING BAYESIAN INFERENCE (Giele, Kosower, Keller 2001)

  PROBLEM IS MADE FINITE-DIMENSIONAL BY THE CHOICE OF PRIOR, BUT RESULT DO NOT DEPEND ON THE CHOICE IF SUFFICIENTLY GENERAL

  HARD TO HANDLE "FLAT DIRECTIONS" (Monte Carlo replicas which lead to same agreement with data); COMPUTATIONALLY VERY INTENSIVE

THE NEURAL MONTE CARLO APPROACH

BASIC IDEA: USE NEURAL NETWORKS AS UNIVERSAL UNBIASED INTERPOLANTS

- GENERATE A SET OF MONTE CARLO REPLICAS $F_2^{(k)}(x_i, Q^2)$ OF THE ORIGINAL DATASET $F_2^{(\mathrm{data})}(x_i, Q^2)$ WHICH IS LARGE ENOUGH TO REPRODUCE CENTRAL VALUES (AS AVERAGES), ERRORS (AS VARIANCES) AND CORRELATIONS (AS COVARIANCES)

  $\Rightarrow$ REPRESENTATION OF $\mathcal{P}[F_2]$ AT DISCRETE SET OF POINTS (x_i, Q_i^2)

- TRAIN A NEURAL NET ON EACH REPLICA, THUS OBTAINING A NEURAL REPRESENTATION OF THE FUNCTION $F_2^{(net)(k)}(x, Q)$

- THE SET OF NEURAL NETS IS A REPRESENTATION OF THE PROBABILITY DENSITY:

$$\left\langle \mathcal{F}\left[F_2(x, Q^2)\right] \right\rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \mathcal{F}\left[F_2{}^{(net)(k)}(x, Q^2)\right]$$

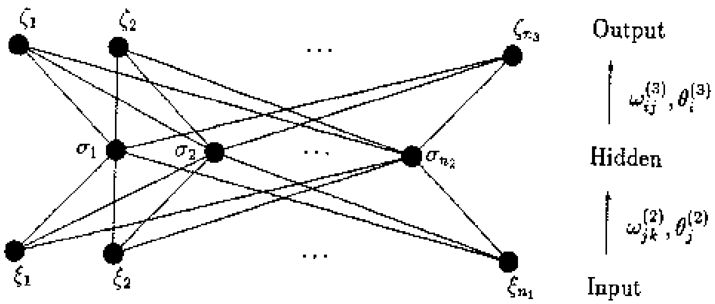
  EXAMPLE: MELLIN MOMENT

$$\left\langle \int_0^1 dx\, x^{N-1} F_2(x, Q^2) \right\rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \int_0^1 dx\, x^{N-1} F_2{}^{(net)(k)}(x, Q^2)$$

- CHECK GOODNESS OF FIT THROUGH STATISTICAL INDICATORS (χ^2, CORRELATION,...)

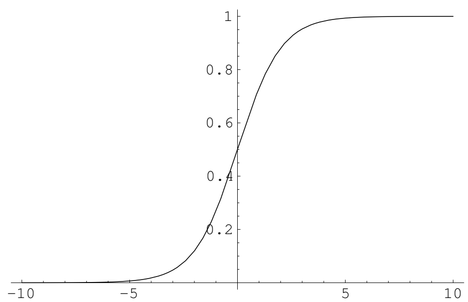
NEURAL NETWORKS

STRUCTURE



MULTILAYER FEED-FORWARD NETWORKS

- Each neuron receives input from neurons in preceding layer and feeds output to neurons in subsequent layer
- Activation determined by weights and thresholds
  $\xi_i = g\left(\sum_j \omega_{ij}\xi_j - \theta_i\right)$
- Sigmoid activation function
  $g(x) = \frac{1}{1+e^{-\beta x}}$



- WEIGHTS & THRESHOLDS CAN BE ADJUSTED SO THAT SIGMOIDS ARE IN CROSSOVER NONLINEAR REGION
- THANKS TO NONLINEAR BEHAVIOUR, ANY FUNCTION CAN BE EXPANDED OVER BASIS OF $g(x)$, $g(g(x))$, $g(g(g(x)))$ $\ldots$
- CAN CHOOSE REDUNDANT ARCHITECTURE (NO. OF LAYERS & NODES) TO MAKE SURE NO SMOOTHING BIAS IS INTRODUCED

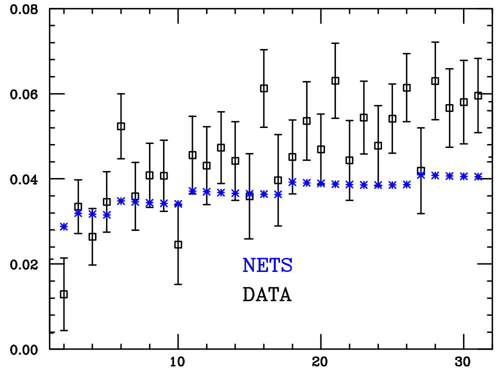
NEURAL NETWORKS

TRAINING

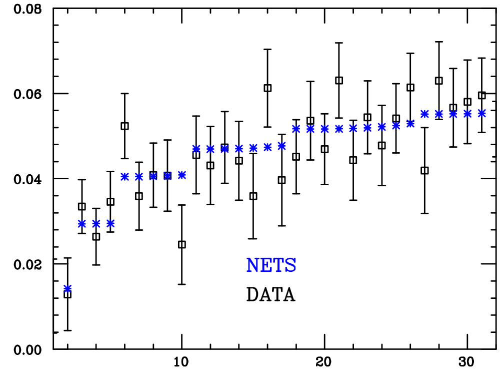
TRAINING BY BACK-PROPAGATION

- START WITH RANDOM NETWORK & COMPUTE OUTPUT FOR GIVEN INPUT (F_2 FOR GIVEN (x, Q^2))
- COMPARE COMPUTED OUTPUT TO DESIRED OUTPUT BY MEANS OF ENERGY FUNCTION ($e.g.$ χ^2)
- VARY WEIGHTS AND THRESHOLDS ALONG DIRECTION OF STEEPEST DESCENT OF ENERGY FUNCTION $\Rightarrow$ CAN BE DONE BY BACK-PROPAGATION
- ITERATE

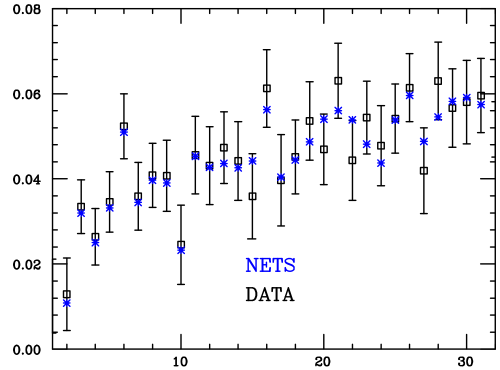
UNDERLEARNING



"PROPER" LEARNING



OVERLEARNING



WHEN SHOULD TRAINING STOP?

WHICH IS THE APPROPRIATE ENERGY FUNCTION?

OPTIMAL TRAINING

WITH LONG ENOUGH TRAINING & BIG ENOUGH NETWORK,

PREDICTION GOES THROUGH ALL POINTS

any error function proportional to (data-nets) will do: vanishes at minimum.

Q: DO WE REALLY WANT THIS?

NAIVE A: SURE! Then when averaging over MC sample, at (x, Q^2) of datapoints averaging over nets is *identical* to averaging over data

OBJECTION: WHAT IF WE HAVE TWO MEASUREMENTS AT THE SAME (x, Q^2)?

PERFORM WEIGHTED AVERAGE $\frac{F_2^{(1)}/\sigma_1 + F_2^{(2)}/\sigma_2}{1/\sigma_1 + 1/\sigma_2}$ BEFORE DATA GENERATION.

BUT WHAT IF WE HAVE TWO MEASUREMENTS AT (x_i, Q_i^2) WHICH ARE VERY CLOSE?

F_2 IS NOT A FRACTAL!

CLEVER A:

- ERROR FUNCTION → USUAL LOG-LIKELIHOOD

$$E^{(k)}[\omega, \theta] = \sum_{i=1}^{N_{dat}} \frac{\left(F_i^{(art)(k)} - F_i^{(net)(k)}\right)^2}{\sigma_{i,s}^{(exp)\,2}}$$

- ESTABLISH FIXED TRAINING LENGTH SUCH THAT $\frac{E^{(k)}[\omega,\theta]}{N_{dat}} \approx 1$

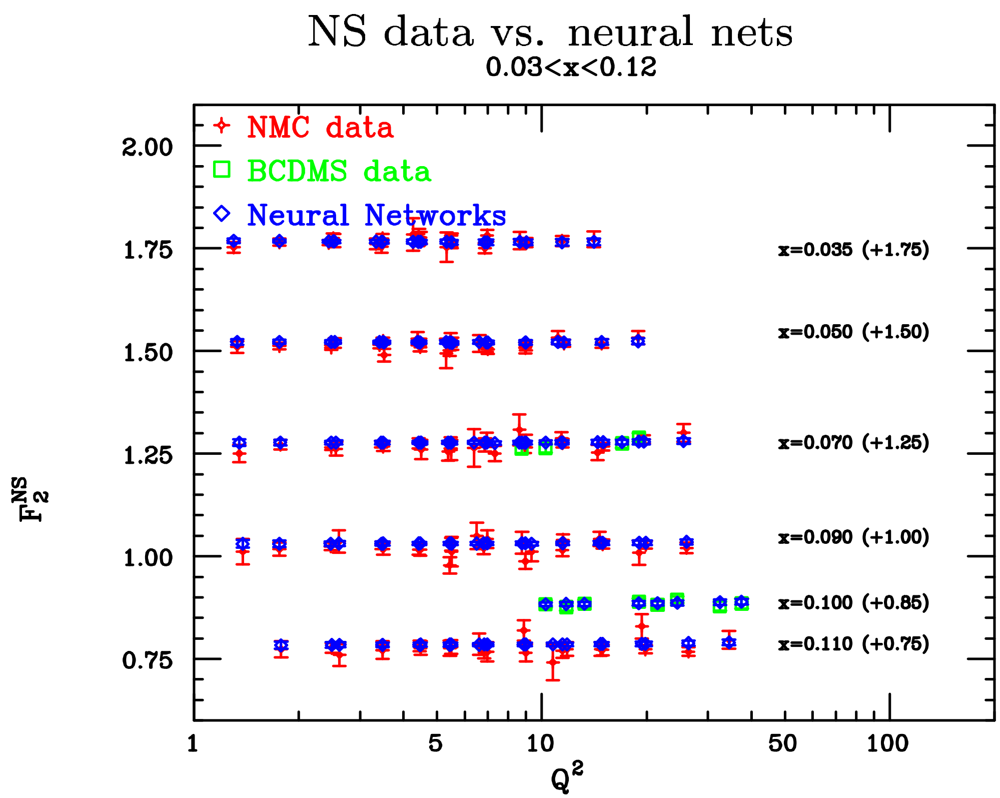
WHAT ABOUT SYST. ERRORS? TAKEN CARE OF BY MC DATA GENERATION!

$F_i^{(net)}$ provide best fit of $F_i^{(sys)(k)} \equiv F_i^{(exp)} + \sum_{p=1}^{N_{sys}} r_{i,p}^{(k)} \sigma_{i,p}$.

Including systematics in likelihood not practical (nonlocal back-propagation).

⇒ TRAIN 1000 PROTON, 1000 DEUTERON & 1000 NONSINGLET NETS

Combining data



In nonsinglet case,
average variance of nets $<<$ stat. error of data (factor 3–4)
is it due to smoothing bias?
or is it due to combining data?
recall error on weighted average

$$\sigma = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2} < \sigma_i$$

Can construct a statistical indicator to tell!

Average error $\langle E \rangle = \frac{1}{N_{rep}} \sum_{n=1}^{N_{rep}} \sum_{i=1}^{N_{dat}} \frac{\left(F_i^{(art)(n)} - F_i^{(net)(n)}\right)^2}{\sigma_{i,s}^{(exp)\,2}}$ ($n \to$ replica; $i \to$ datapoint)

"Central" error $\langle \tilde{E} \rangle = \frac{1}{N_{rep}} \sum_{n=1}^{N_{rep}} \sum_{i=1}^{N_{dat}} \frac{\left(F_i^{(exp)} - F_i^{(net)(n)}\right)^2}{\sigma_{i,s}^{(exp)\,2}}$

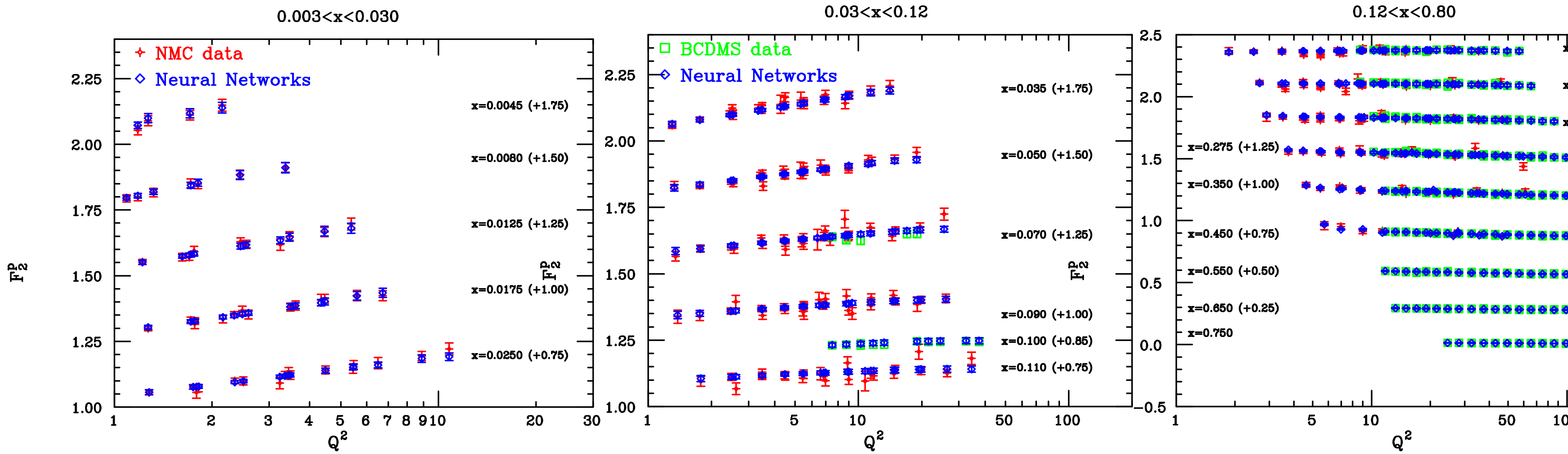
Bias indicator $\mathcal{R} \equiv \langle \tilde{E} \rangle / \langle E \rangle$: if $\sigma_{net} << \sigma_{exp}$ then

$\mathcal{R} \approx 1 \Rightarrow$ BIAS; $\mathcal{R} \approx 1/2 \Rightarrow$ ERROR REDUCTION HERE $\mathcal{R} = 0.58$ (0.53 NMC only)

- INHOMOGENEOUS ERRORS $\Rightarrow$ WEIGHTED TRAINING
- INCOMPATIBLE DATA $\Rightarrow$ DISCARDED BY NETS

RESULTS

NEURAL FIT TO PROTON F_2 DATA



- FULL NEURAL FIT TO F_2 FOR PROTON, DEUTERON & NONSINGLET AVAILABLE
- ERRORS AND CORRELATIONS FAITHFULLY REPRODUCED, BUT STAT. UNCERTAINTIES OPTIMALLY COMBINED
  $\Rightarrow$ FIT CAN BE USED IN LIEU OF DATA, BUT BETTER THAN THEM
- SOURCE CODE, DRIVER PROGRAM & GRAPHIC WEB INTERFACE FOR F_2 PLOTS & NUMERICAL COMPUTATION AVAILABLE @

`http://sophia.ecm.ub.es/f2neural`

An unbiased analysis method: truncated moments

x-space distn.: measurable, but evolution given by integro–differential eqn

N-space moments: evolution given by linear differential eqn, but not measurable

Truncated moments:

$$F_{2,\,N}^{NS}(x_0,\mu^2) \equiv \int_{x_0}^{1} dx\; x^{n-1} F_2^{NS}(x,\mu^2)$$

- measurable
- to any finite accuracy, satisfy coupled linear evolution equations with upper triangular anomalous dimension matrix:

$$\mu^2 \frac{d}{d\mu^2}\begin{pmatrix} F_{2,\,1}^{NS}(x_0,\mu^2) \\ F_{2,\,2}^{NS}(x_0,\mu^2) \\ F_{2,\,3}^{NS}(x_0,\mu^2) \\ \ldots \end{pmatrix} = \begin{pmatrix} \gamma_{11}^{M}(x_0,\alpha_s(\mu^2)) & \gamma_{12}^{M}(x_0,\alpha_s(\mu^2)) & \gamma_{13}^{M}(x_0,\alpha_s(\mu^2)) & \ldots \\ 0 & \gamma_{22}^{M}(x_0,\alpha_s(\mu^2)) & \gamma_{23}^{M}(x_0,\alpha_s(\mu^2)) & \ldots \\ 0 & 0 & \gamma_{33}^{M}(x_0,\alpha_s(\mu^2)) & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix} \begin{pmatrix} F_{2,\,1}^{NS}(x_0, \\ F_{2,\,2}^{NS}(x_0, \\ F_{2,\,3}^{NS}(x_0, \\ \ldots \end{pmatrix}$$

  M is order of truncation: only M moments coupled.

  As $M \to \infty$, accuracy becomes arbitrarily high.

- Can truncate to finite triangular anomalous dimension matrix
- Rapid convergence: for $x_0 \lesssim 0.1$, $M \approx 10$ ensures percent accuracy on evolution of all moments with $N \geq 2$. Same accuracy on first moment also possible with improved solution (non-triangular matrix).

DETERMINATION OF α_s

- MOMENTS CAN BE COMPUTED AT ANY SCALE IN TERMS OF MOMS. AT REF. SCALE Q_0^2

through evolution matrix $M(x_0; Q_0^2, Q_i^2; \alpha_s)$ determined by an. dim. and α_s:

$$q_n^{th}(x_0, Q_i^2) \equiv \sum_{p=n_{min}}^{M} M_{np}(x_0; Q_0^2, Q_i^2; \alpha_s)\; q_p(x_0, Q_0^2)$$

- CAN DETERMINE α_s BY MINIMIZING χ^2 with covariance matrix V^{-1} from neur. nets

$$\chi^2 = \sum_{n,i} \sum_{m,j} \left[q_n^{exp}(x_0, Q_i^2) - q_n^{th}(x_0, Q_i^2)\right] V^{-1}_{ni;mj} \left[q_m^{exp}(x_0, Q_j^2) - q_m^{th}(x_0, Q_j^2)\right]$$

MOMENTS AND CORRELATIONS

IN PRINCIPLE FIT α_s & ALL MOMENTS AT REF. SCALE

IN PRACTICE NEIGHBOURING MOMENTS HIGHLY CORRELATED;

OFF-DIAGONAL ANOMALOUS DIMS. SMALL $\Rightarrow$ FIT ONLY A SUBSET OF MOMENTS

(NMC + BCDMS)

SINGLE MOMENT

n	α_s
2	0.085 ± 0.070
3	0.106 ± 0.030
4	0.115 ± 0.019
5	0.123 ± 0.015
6	0.127 ± 0.014
7	0.129 ± 0.014
8	0.129 ± 0.016
9	0.129 ± 0.018

purple: minimal error

MORE MOMENTS

FITTED MOMENTS	α_s
2+3+4	0.126 ± 0.010
2+4+6	0.140 ± 0.008
3+5+7	0.138 ± 0.009
2+4+6+8	0.142 ± 0.009
3+5+7+9	0.124 ± 0.007
2+4+5+7	0.141 ± 0.009
3+4+5+6+7	0.1256 ± 0.0049
3+4+5+6+8	0.1247 ± 0.0050
2+4+5+6+8	0.1242 ± 0.0042
2+4+5+7+8	0.1254 ± 0.0044

red: optimal fit

OPTIMAL FIT

AS THE NUMBER OF FITTED MOMENTS IS INCREASED

ERROR DECREASES,

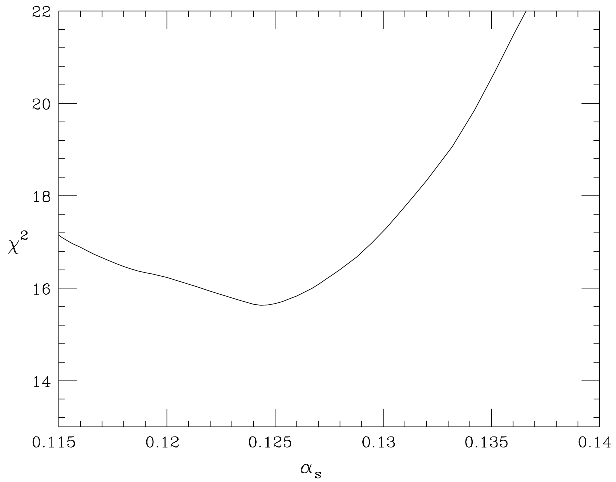
STABILITY OF CENTRAL VALUES IMPROVES

BUT IF CORRELATIONS LARGE, FIT UNSTABLE

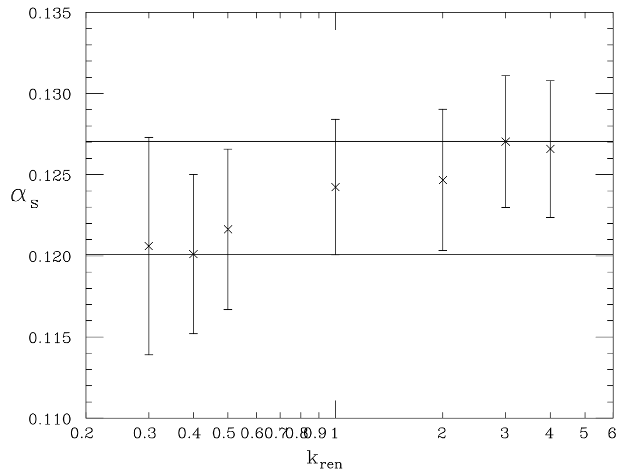
- $20 \leq Q^2 \leq 70$ GEV^2, THREE SCALES correlns. larger if Q^2 values closer
- $x_0 = 0.03$ correlns. larger if x_0 larger
- 2+4+5+6+8 higher moments less reliable and more correlated

UNCERTAINTIES

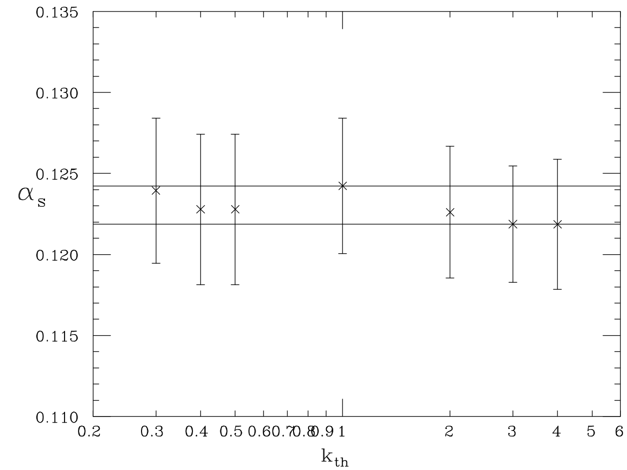
STAT. ERROR



REN. SCALE



HEAVY QUARK THR.



- ASYMMETRIC χ^2: $\sigma(\text{STAT.}) = {}^{+0.004}_{-0.007}$
- HIGHER ORDER CORRNS FROM $\mu^2_{ren} = k_{ren} Q^2$, $0.3 \leq k_{ren} \leq 4$: $\sigma(\text{REN.}) = {}^{+0.003}_{-0.004}$
- POSITION OF HQ THRESH. $Q^2_{th} = k_{th} M^2_q$, $0.3 \leq k_{th} \leq 4$ $\sigma(\text{THRESH.}) = {}^{+0.000}_{-0.002}$
- POWER CORRNS. VARY Q^2_{min} FROM 20 TO 30 GEV^2 $\sigma(\text{HT}) < 0.001$

$$\alpha_s(M_Z) = 0.124 \ {}^{+0.004}_{-0.007} \ (\text{EXP.}) \ {}^{+0.003}_{-0.004} \ (\text{TH.}) = 0.124 \ {}^{+0.005}_{-0.008} \ (\text{TOTAL})$$

ERROR: DOMINATED BY EXP. ERROR, TH. BIAS & UNCERTAINTY MINIMIZED

CENTRAL VALUE: CONSISTENT WITH WORLD AVERAGE BUT HIGH

EVIDENCE FOR SUDAKOV? High moments dominate the fit, $Q^2_{\text{eff}} = Q^2/N$;

α_s from a single moment increases with N

OUTLOOK

- SUCCESFUL IMPLEMENTATION OF NEURAL FITTING

  ⇒ NEURAL PARTON DISTRIBUTIONS!

- WORKING EVOLUTION CODE FOR TRUNCATED MOMENTS

  ⇒ GLUON SPIN FRACTION (AND MORE...)

...A WHOLE NEW SET OF TOOLS IN THE BOX!