2.1 Parametrisation, sum rules and positivity

The functional form, what constrains it before any data, and the FK-table pipeline that turns it into predictions.

---

1. The inverse problem

Part 1 ended with a forward statement: given the parton distribution functions (PDFs) at one scale Q_0, QCD predicts every hard cross-section at every other scale. A fit runs this backwards. This chapter establishes what has to be assumed before the backward question has an answer, and how the assumption is turned into numbers that can be compared with data.

Write the data as a vector D_I, $I = 1 \dots N_{\text{dat}}$, and the prediction for point I as a functional $T_I[f]$ of the distributions $f_j(x, Q_0^2)$, where j labels the flavour combination. The measurement model is

$$D_I = T_I[f] + \varepsilon_I, \qquad \varepsilon \sim \mathcal{N}(0, C), \tag{2.1.1}$$

with C the covariance matrix of the experimental errors (chapter 2.2). The data index is written I in this chapter, because i, j and k are needed for polynomial order, flavour and grid node; the following chapters write D_i. The unknown is a set of functions on $0 < x < 1$. The data are a finite list of numbers.

For deep inelastic scattering the functional is linear. Once the distributions are stored on a grid of n_x nodes x_k (section 7) it is a matrix, and any change δf that the matrix annihilates leaves every prediction unchanged:

$$\begin{aligned} T_I[f] &= \sum_{j,k} \mathrm{FK}_{I,jk}\, f_j(x_k, Q_0^2), \\ T[f + \delta f] &= T[f] \quad \text{if} \quad \sum_{j,k} \mathrm{FK}_{I,jk}\, \delta f_j(x_k) = 0 \ \text{ for all } I. \end{aligned} \tag{2.1.2}$$

Three facts make the inversion ill-posed.

- **There are more unknowns than the data can fix.** Even on the grid, the matrix FK has N_{dat} rows and $14\, n_x$ columns. Its null space contains every δf the data cannot see: flavour combinations no measured process couples to, and any shape at values of x the experiments do not reach.
- **The map smooths.** Each prediction is an integral of f over a range of x. Rapid oscillations average out, so inverting the map amplifies noise.
- **Exact constraints are not data.** The sum rules of chapter 1.3 hold for the true distributions, but nothing in Eq. (2.1.1) enforces them.

A fit therefore needs information beyond the data. Every group supplies it in the same way: restrict f to a family $f(x; \theta)$ with a finite number p of parameters θ, and infer θ [1].

> **A functional form is a prior.** Choosing a family $f(x;\theta)$ assigns probability zero to every function outside it. This is true of a 52-parameter polynomial form and of a neural network alike; they differ in the size of the family and in how its members behave where there are no data. The uncertainty a fit quotes is the uncertainty *within the family* (section 8).

---

2. The generic form and where its powers come from

Almost every fixed functional form used since the 1980s has the same skeleton, written for the momentum density xf at the input scale:

$$xf(x,Q_0^2) = A\,x^{\delta}\,(1-x)^{\eta}\,\mathcal{P}(x). \tag{2.1.3}$$

A is a normalisation, δ controls the behaviour as $x \to 0$, η controls the approach to $x = 1$, and $\mathcal{P}$ is a smooth function that interpolates between the ends. Moore gives the 1994 form of Martin, Roberts and Stirling as an early example, with $\mathcal{P} = 1 + \epsilon\sqrt{x} + \gamma x$ [1]. Each power has a physical argument behind it, and each argument has limits.

Small x: Regge behaviour

At small x the photon–proton system has a large invariant mass, $W^2 \simeq Q^2/x$, so deep inelastic scattering at fixed Q^2 and $x \to 0$ is the high-energy limit of the total cross-section for a virtual photon on a proton. Regge theory describes high-energy total cross-sections as the exchange of families of particles ("trajectories") in the crossed channel. It gives a power law in the energy, $\sigma_{\text{tot}} \propto (W^2)^{\alpha(0)-1}$, where $\alpha(0)$ is the *intercept* of the trajectory. Since $F_2 \propto Q^2\sigma_{\text{tot}}$ at small x (derivation D1),

$$F_2(x,Q^2) \;\propto\; x^{\,1-\alpha(0)} \quad (x \to 0), \qquad \text{so} \qquad \delta = 1-\alpha(0). \tag{2.1.4}$$

Two kinds of trajectory matter. The Pomeron carries the quantum numbers of the vacuum and can couple to the gluon and the flavour-singlet sea. Its intercept, fitted to hadronic total cross-sections, is about 1.08 [5], which gives $\delta \simeq -0.08$: a momentum density that rises slowly as $x \to 0$. The leading meson trajectories carry flavour and couple to valence combinations. Their intercept is about 0.5, which gives $\delta \simeq 0.5$: a valence density that vanishes at small x.

Large x: counting rules

Every distribution vanishes at $x = 1$, since no parton can carry more than the proton's whole momentum [1]. How fast is the subject of the quark counting rules, quoted here without derivation. For one parton to carry almost all the momentum, each of the other n_s constituents (the *spectators*) must be left with almost none, and at lowest order every spectator costs a hard gluon exchange. Counting powers gives

$$f(x) \;\propto\; (1-x)^{2n_s-1} \quad (x \to 1). \tag{2.1.5}$$

A valence quark in a three-quark proton has $n_s = 2$, so $\eta = 3$. A gluon has three spectators, $\eta = 5$. A sea quark has four, $\eta = 7$. The Snowmass report quotes the rule as $\eta \gtrsim 3$ for baryons [4].

The limits of both arguments

Neither argument fixes a number that can be put into a fit.

The exponents depend on the scale. Both arguments refer to some low, unspecified scale, and evolution changes the powers. At large x the change follows from chapter 1.4 (derivation D6):

$$\eta(\mu^2) = \eta(\mu_0^2) + \frac{C_F}{\pi b}\,\ln\frac{\alpha_S(\mu_0^2)}{\alpha_S(\mu^2)}. \tag{2.1.6}$$

For $n_f = 4$ the coefficient is 0.64, so halving the coupling adds 0.44 to η. At small x the gluon and the sea grow steeper as the scale rises, Eq. (1.4.32).

The powers are reached only at the end-points. What data constrain is an *effective* exponent at finite x,

$$\delta_{\text{eff}}(x) = \frac{\partial \ln(xf)}{\partial \ln x}, \qquad \eta_{\text{eff}}(x) = \frac{\partial \ln(xf)}{\partial \ln(1-x)}, \tag{2.1.7}$$

which tends to δ and η only as $x \to 0$ and $x \to 1$. For the up-valence distribution of section 4, with $\delta = 0.344$ and $\eta = 3.749$, the check script finds $\delta_{\text{eff}} = 0.007$ at $x = 10^{-2}$ and 0.365 at $x = 10^{-4}$; and $\eta_{\text{eff}} = 2.50$ at $x = 0.5$ and 3.35 at $x = 0.9$. Where data exist, the factor $\mathcal{P}$ matters as much as the nominal powers. The Snowmass report makes the same point: different forms can mimic one another, so the effective exponent is the more informative quantity [4].

In practice δ and η are free parameters; the Snowmass report lists as an open question how far such expectations should shape fitted distributions [4]. In the parameter set our fits start from (section 4), $\eta_{u_V} = 3.75$, $\eta_{d_V} = 3.98$, $\eta_g = 5.25$ and $\eta_S = 7.04$, each within one unit of the counting-rule value, and $\delta_S = -0.010$ against $\delta_{u_V} = 0.344$, in the order Regge theory suggests.

---

3. The interpolating factor: Chebyshev polynomials in $\sqrt{x}$

This section establishes why the MSHT20 form writes $\mathcal{P}$ as a series of Chebyshev polynomials in the variable $y = 1 - 2\sqrt{x}$.

The polynomials

The Chebyshev polynomial of the first kind, T_n, is the polynomial of degree n that turns $\cos\theta$ into $\cos n\theta$:

$$T_n(\cos\theta) = \cos n\theta, \qquad T_0 = 1, \quad T_1(y) = y, \quad T_{n+1}(y) = 2y\,T_n(y) - T_{n-1}(y). \tag{2.1.8}$$

The recurrence is the identity $\cos(n+1)\theta + \cos(n-1)\theta = 2\cos\theta\cos n\theta$. It gives $T_2 = 2y^2 - 1$, $T_3 = 4y^3 - 3y$, and so on. Two properties make them a good basis on $-1 \le y \le 1$:

- **They are bounded.** $|T_n(y)| \le 1$, so a coefficient c_n changes the function by at most $|c_n|$, at every order. In the basis $1, y, y^2, \ldots$ the high powers are nearly indistinguishable on the interval, and fitted coefficients become large with alternating signs.
- **They are orthogonal**, with weight $1/\sqrt{1-y^2}$, because $\int_0^\pi \cos m\theta \cos n\theta\, d\theta = 0$ for $m \ne n$. Adding one more term changes the lower coefficients little.

At the ends, $T_n(1) = 1$ and $T_n(-1) = (-1)^n$.

The variable

The zeros of T_n are spread over the whole interval in y. The map from x to y decides where in x that resolution is spent. MSHT20 uses

$$y = 1 - 2\sqrt{x}, \qquad x = 0 \mapsto y = 1, \qquad x = 1 \mapsto y = -1. \tag{2.1.9}$$

The data span several decades in x. With the linear choice $y = 1 - 2x$, the two decades $10^{-4} < x < 10^{-2}$ map to $0.98 < y < 0.9998$: 1% of the interval, where no polynomial of low degree can vary. With the square root they map to $0.80 < y < 0.98$, which is 9% (Figure 1). The square root also keeps every sum-rule integral in closed form, because powers of y are powers of $x^{1/2}$ (section 5).

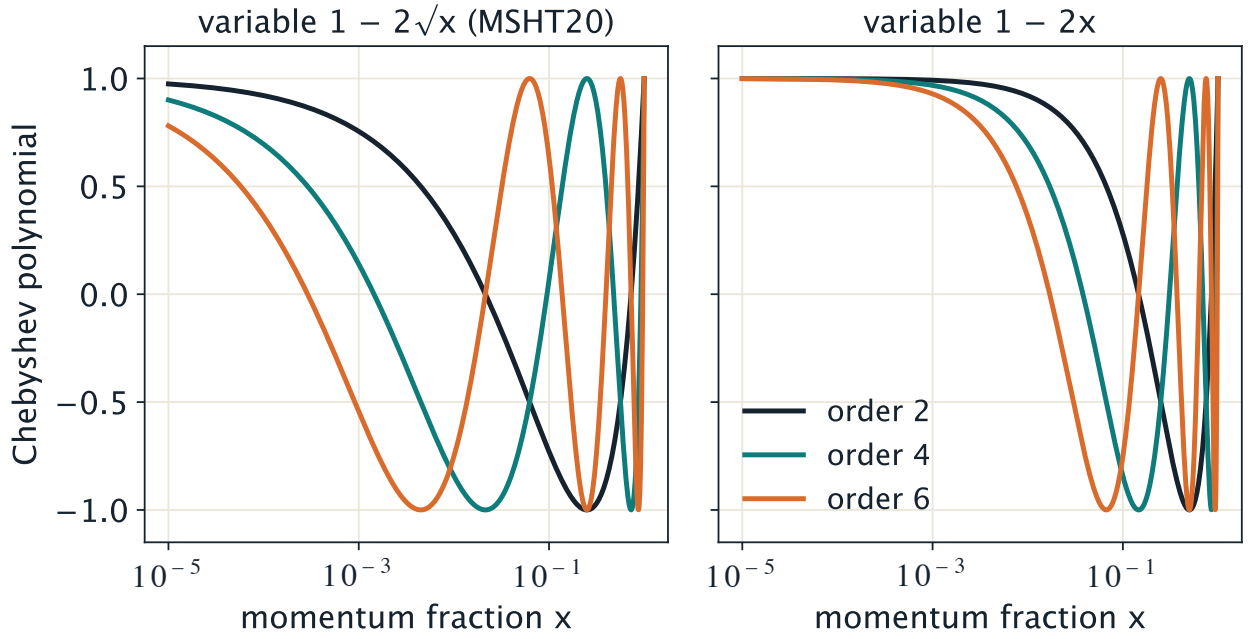

Figure 1. The same three Chebyshev polynomials as functions of x. Left: with $y = 1 - 2\sqrt{x}$ the zeros are spread over three decades. Right: with $y = 1 - 2x$ every polynomial is flat below $x \approx 0.05$ and all the structure is crowded into large x.

As $x \to 0$ every $T_n \to 1$, so the series tends to the constant $1 + \sum_i c_i$ and the small-x power remains δ. The approach is in powers of $\sqrt{x}$, which is slow: this is why δ_{eff} in section 2 is still far from δ at $x = 10^{-2}$.

---

4. The MSHT20 form as used in this project

The form is that of the MSHT20 analysis [6]. Our implementation is a JAX port of the FPPDF code's version of it, held in the project as `msht20_model`; the statements below were read from that code.

Seven distributions

The form parametrises seven combinations at Q_0: the two valence distributions $u_V = u - \bar{u}$ and $d_V = d - \bar{d}$; the sea S; the total strangeness $s_+ = s + \bar{s}$; the strangeness asymmetry $s_- = s - \bar{s}$; the ratio $\rho = \bar{d}/\bar{u}$; and the gluon. In the convention of the code the sea is the whole of it,

$$S = 2(\bar{u} + \bar{d}) + s + \bar{s},$$

so that $u_V + d_V + S$ is the sum of all quarks and antiquarks. The MSHT20 paper defines S in the same way [6]. That paper takes $Q_0^2 = 1\ \text{GeV}^2$; our fits use the same form at $Q_0 = 1.65$ GeV, so fitted parameter values are not directly

comparable with the published ones.

Four of the seven share one template, with six Chebyshev coefficients:

$$xf(x,Q_0^2) = A_f\, x^{\delta_f}\,(1-x)^{\eta_f}\Big[1+\sum_{i=1}^{6} c_{f,i}\,T_i(y)\Big], \qquad f \in \{u_V,\ d_V,\ S,\ s_+\}. \tag{2.1.10}$$

The gluon has four Chebyshev coefficients and a second term with its own powers:

$$xg(x,Q_0^2) = A_g\, x^{\delta_g}(1-x)^{\eta_g}\Big[1+\sum_{i=1}^{4} c_{g,i}\,T_i(y)\Big] + A_{g-}\, x^{\delta_{g-}}(1-x)^{\eta_{g-}}. \tag{2.1.11}$$

One power law cannot change sign. The second term, with negative A_{g-} and δ_{g-}, lets the input gluon turn over and become negative at small x. The remaining two are

$$\begin{aligned} xs_-(x,Q_0^2) &= A_-\, x^{\delta_-}(1-x)^{\eta_-}\Big(1-\frac{x}{x_0}\Big), \\ \rho(x,Q_0^2) &= A_\rho\,(1-x)^{\eta_\rho}\Big[1+\sum_{i=1}^{6} c_{\rho,i}\,T_i(y)\Big]. \end{aligned} \tag{2.1.12}$$

The asymmetry s_- must integrate to zero, so it must change sign; the factor $(1-x/x_0)$ puts the sign change at x_0. The ratio ρ has no power of x: it tends to the constant $A_\rho(1+\sum_i c_{\rho,i})$ as $x \to 0$.

The individual antiquarks and the strange quarks follow by algebra:

$$\begin{aligned} x(\bar u+\bar d) &= \tfrac{1}{2}\big(xS - xs_+\big), & \bar u &= \frac{\bar u+\bar d}{1+\rho}, \\ \bar d &= \frac{\rho\,(\bar u+\bar d)}{1+\rho}, & s,\bar s &= \tfrac{1}{2}\big(s_+ \pm s_-\big). \end{aligned} \tag{2.1.13}$$

Then $u = u_V + \bar u$ and $d = d_V + \bar d$. Charm is set to zero at Q_0 in our port; above Q_0 it is generated by evolution. The theory tables we use are built in a scheme that allows a fitted charm distribution, and the port does not use that freedom. The ledger records this as a deliberate mismatch whose effect has not been measured [7].

Counting to 52

The code stores the parameters in a vector of 73 slots. Not all are free.

DISTRIBUTION	FREE PARAMETERS	NUMBER	NOT FREE, AND WHY
u_V	$\delta, \eta, c_1 \dots c_6$	8	A: number sum rule
d_V	$\delta, \eta, c_1 \dots c_6$	8	A: number sum rule
S	$A, \delta, \eta, c_1 \dots c_6$	9	none
s_+	$A, \eta, c_1 \dots c_6$	8	δ: tied to δ_S
g	$\delta, \eta, c_1 \dots c_4, A_{g-}, \delta_{g-}, \eta_{g-}$	9	A_g: momentum sum rule
s_-	A, η	2	x_0: strangeness sum rule; δ and six Chebyshev slots held fixed
ρ	$A, \eta, c_1 \dots c_6$	8	none
charm	none	0	nine slots, all zero

The free parameters number $8 + 8 + 9 + 8 + 9 + 2 + 8 = 52$. From the other side: 73 slots, less 9 for charm, 4 fixed by sum rules, 1 tie, and 7 held fixed in s_-, is again 52.

Figure 2 shows the main shapes at the parameter point that ships with the code as its “publication” input file, which our fits use as template and starting point. It is computed by the check script from that file, with the sum rules of section 5 imposed.

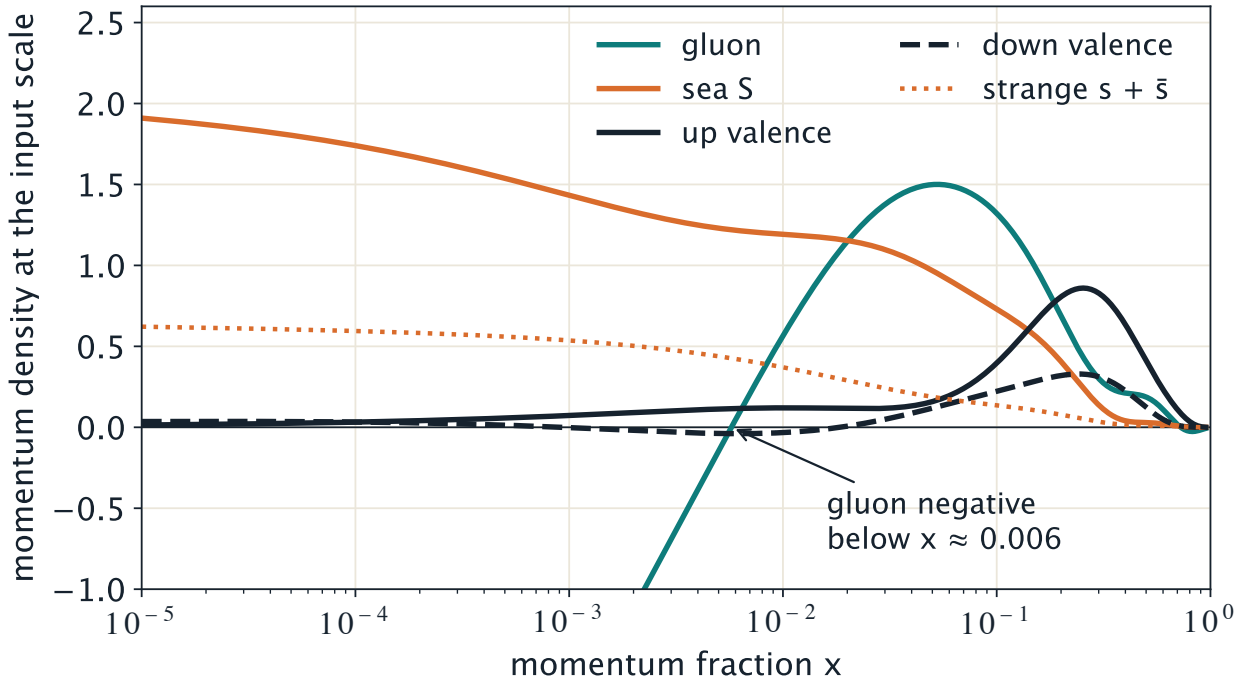

Figure 2. The input distributions of the MSHT20 form at the starting parameter point of the project’s port, with the four constrained parameters computed from the sum rules. The two-term gluon of Eq. (2.1.11) is negative below $x \approx 0.006$ at this point.

5. Sum rules imposed analytically

Four parameters of the form are not fitted. They are computed from the other 52 each time the model is evaluated, so that four exact constraints hold identically.

The constraints

Chapter 1.3 derived the number and momentum sum rules from conserved currents [1, 3]. For the proton, in the variables of section 4,

$$\int_0^1 dx\, u_V = 2, \qquad \int_0^1 dx\, d_V = 1, \qquad \int_0^1 dx\, s_- = 0,$$
$$\int_0^1 dx\; x\left[u_V + d_V + S + g\right] = 1. \tag{2.1.14}$$

They hold at every scale, and evolution preserves them (chapter 1.4), so imposing them at Q_0 is enough.

A constraint can be imposed by a penalty in the loss function, in which case it holds approximately and the result depends on the penalty's weight. Or it can be solved for one parameter and substituted. The second way is exact and removes a dimension from the parameter space, but it needs the integrals in closed form. For the form of section 4 they are.

One integral

Every term of Eqs. (2.1.10)–(2.1.12) is a power of x times a power of $1-x$ times a polynomial in $\sqrt{x}$. The only integral needed is Euler's Beta function (derivation D2):

$$I(a,b) \equiv \int_0^1 dx\, x^a (1-x)^b = \frac{\Gamma(a+1)\,\Gamma(b+1)}{\Gamma(a+b+2)}, \qquad a > -1,\; b > -1. \tag{2.1.15}$$

A Chebyshev polynomial in $y = 1 - 2\sqrt{x}$ is a polynomial in $\sqrt{x}$ with coefficients $\tau_{n,k}$ that follow from a two-term recurrence (derivation D3):

$$T_n\big(1-2\sqrt{x}\big) = \sum_{k=0}^{n} \tau_{n,k}\, x^{k/2}, \qquad \tau_{n,0} = 1, \qquad \tau_{n,k+1} = \frac{k^2 - n^2}{(k+\tfrac{1}{2})(k+1)}\,\tau_{n,k}. \tag{2.1.16}$$

For example $T_2 = 1 - 8\sqrt{x} + 8x$. Each term shifts the power of x by a half-integer, so

$$\begin{aligned} \int_0^1 dx\, x^a (1-x)^b\, T_n(y) &= \sum_{k=0}^{n} \tau_{n,k}\, I\big(a+\tfrac{k}{2},\, b\big), \\ J(a,b;c) &\equiv \int_0^1 dx\, x^a (1-x)^b \Big[1 + \sum_i c_i\, T_i(y)\Big] \\ &= I(a,b) + \sum_i c_i \sum_{k=0}^{i} \tau_{i,k}\, I\big(a+\tfrac{k}{2},\, b\big). \end{aligned} \tag{2.1.17}$$

The project's code reaches the same result by expanding $(1-2\sqrt{x})^k$ binomially; the check script confirms that the two routes agree.

The four solutions

The parametrisation gives xf, so a number integral divides by x and lowers the power by one, while a momentum integral uses the power as it stands (derivation D4).

Valence normalisations. From the first two constraints,

$$A_{u_V} = \frac{2}{J(\delta_u - 1, \eta_u; c_u)}, \qquad A_{d_V} = \frac{1}{J(\delta_d - 1, \eta_d; c_d)}. \tag{2.1.18}$$

The strangeness crossing. The third constraint fixes where s_- changes sign, and the Gamma functions reduce to a ratio of the exponents:

$$x_0 = \frac{I(\delta_-, \eta_-)}{I(\delta_- - 1, \eta_-)} = \frac{\delta_-}{\delta_- + \eta_- + 1}. \tag{2.1.19}$$

The amplitude A_- remains free: the sum rule decides where $s - \bar{s}$ changes sign, not how large it is.

The gluon normalisation. Write the momentum fraction carried by each quark combination as $\langle x \rangle_f = A_f \, J(\delta_f, \eta_f; c_f)$. Momentum conservation then gives the normalisation of the first gluon term:

$$A_g = \frac{1 - \langle x \rangle_{uv} - \langle x \rangle_{dV} - \langle x \rangle_S - A_{g-} \, I(\delta_{g-}, \eta_{g-})}{J(\delta_g, \eta_g; c_g)}. \tag{2.1.20}$$

The order matters: $\langle x \rangle_{u_V}$ and $\langle x \rangle_{d_V}$ contain the valence normalisations, so Eq. (2.1.18) is evaluated first.

Numbers at the starting point

At the starting parameter point of Figure 2 the check script finds

QUANTITY	VALUE
A_{uv}, A_{dV}	8.363, 0.690
x_0	1.762×10^{-4}
A_g	69.08
Momentum in u_V, d_V	0.350, 0.127
Momentum in the sea S	0.196, of which s_+ carries 0.045
Momentum in the gluon	$0.326 = 0.588 - 0.262$

and direct numerical integration of the assembled shapes returns 2, 1, 0 and 1 for the four sum rules to better than one part in 10^8. Two things can be read from the table.

- **Normalisations are not sizes.** The Chebyshev terms cancel most of the bare integral: $I(\delta_u - 1, \eta_u) = 1.557$ becomes $J = 0.239$ with the series, so A_{uv} is 8.36 where a pure power law would give 1.28.
- **The gluon is a difference.** The second gluon term carries momentum -0.262, so the first must carry 0.588. Changes in δ_{g-} or η_{g-} are compensated through A_g, which correlates the gluon parameters strongly.

The sum rules thus induce exact correlations between shape and normalisation, which reappear in the Hessian of chapter 2.3.

---

6. Integrability and positivity

Sum rules fix integrals. Two further requirements restrict the shape: the integrals must exist, and physical cross-sections must not be negative.

Integrability

Equation (2.1.15) converges only for $a > -1$. Applied to the sum rules of the form, this gives conditions on the small-x powers:

$$\begin{aligned} \delta_{u_V},\ \delta_{d_V},\ \delta_{-} &> 0 \qquad \text{(number integrals)},\\ \delta_{S},\ \delta_{g},\ \delta_{g_-} &> -1 \quad \text{(momentum integrals)}. \end{aligned} \tag{2.1.21}$$

Near the boundary the integral diverges as $I(\delta - 1, \eta) \simeq 1/\delta$. In a fixed form, integrability is a statement about the allowed range of a few exponents, enforced by the bounds on those parameters.

Other combinations are not covered by a sum rule but are still expected to have finite first moments, for example the triplet $T_3 = u + \bar{u} - d - \bar{d}$. The form does not guarantee this. At the starting point $\rho \to 0.358$ as $x \to 0$, so $x(\bar{d} - \bar{u})$ behaves as x^{δ_S} with $\delta_S = -0.010$ and does not vanish.

Positivity

Chapter 1.3 set out the status of positivity. At leading order a distribution is a density and cannot be negative. Beyond leading order the $\overline{\text{MS}}$ distribution is a subtracted quantity, and what must be positive is every physical cross-section; whether the distributions themselves are positive has been argued and disputed [1].

The form of section 4 guarantees neither kind of positivity. A power law times $1 + \sum c_i T_i$ changes sign wherever the bracket does, and the two-term gluon is built so that it can. At the starting point the check script finds that xg is negative below $x = 0.0057$ and above $x = 0.73$, that $x\,d_V$ is negative between $x \approx 0.001$ and $x \approx 0.018$, and that $x(\bar{u} + \bar{d})$ is negative above $x = 0.70$.

How NNPDF imposes both

A neural network has no exponents to bound, so NNPDF imposes these requirements through the loss function. In the method of Ubiali's thesis, positivity is required of observables: the longitudinal structure function F_L, which constrains the gluon and singlet at small x, and the dimuon cross-section, which constrains strangeness. Any negative prediction for these is penalised with a multiplier of order 10^{10} [2]. The constraint is imposed at a low scale because evolution preserves it.

In the later form quoted by Moore, the penalty is a smooth function and is applied to the distributions as well [1]:

$$\begin{aligned} \chi^2 \;&\to\; \chi^2 + \sum_{q} \Lambda_q \sum_{k=1}^{N_{\text{grid}}} \mathrm{ELU}_{\alpha}\big(-f_q(x_k,\,5\ \text{GeV}^2)\big), \\ \mathrm{ELU}_{\alpha}(t) &= \begin{cases} t, & t \geq 0, \\ \alpha\,(e^{t} - 1), & t < 0. \end{cases} \end{aligned} \tag{2.1.22}$$

A negative value $f_q = -|f_q|$ costs $\Lambda_q |f_q|$, linear in the violation. A positive value contributes between $-\alpha\Lambda_q$ and zero, negligible for small α. The multipliers Λ_q are hyperparameters. NNPDF adds the same kind of term for a set of pseudo-observables, because positive distributions do not guarantee positive cross-sections, nor the reverse [1].

Integrability is imposed in the same spirit. In Colibri's implementation, which our fits use, the penalty is the squared momentum density of chosen evolution-basis combinations at chosen small-x nodes:

$$\chi^2 \;\to\; \chi^2 + \Lambda_{\text{int}} \sum_{j \in \text{chosen}} \sum_{k \in \text{small } x} \left[x_k f_j(x_k, Q_0^2) \right]^2 . \tag{2.1.23}$$

It pushes $x f_j \to 0$ at small x, the condition for a finite first moment.

These penalties belong to the loss function (chapter 2.2), and they are part of the prior. Our Phase B fits use NNPDF's full set, recorded in the ledger as 17 positivity and 2 integrability constraints. Section 10 reports what happened without them.

---

7. From parameters to predictions

This section follows one evaluation of the model, from the 52 numbers θ to the N_{dat} predictions $T_I(\theta)$. The model enters only through its values on a grid. Everything else is fixed linear algebra.

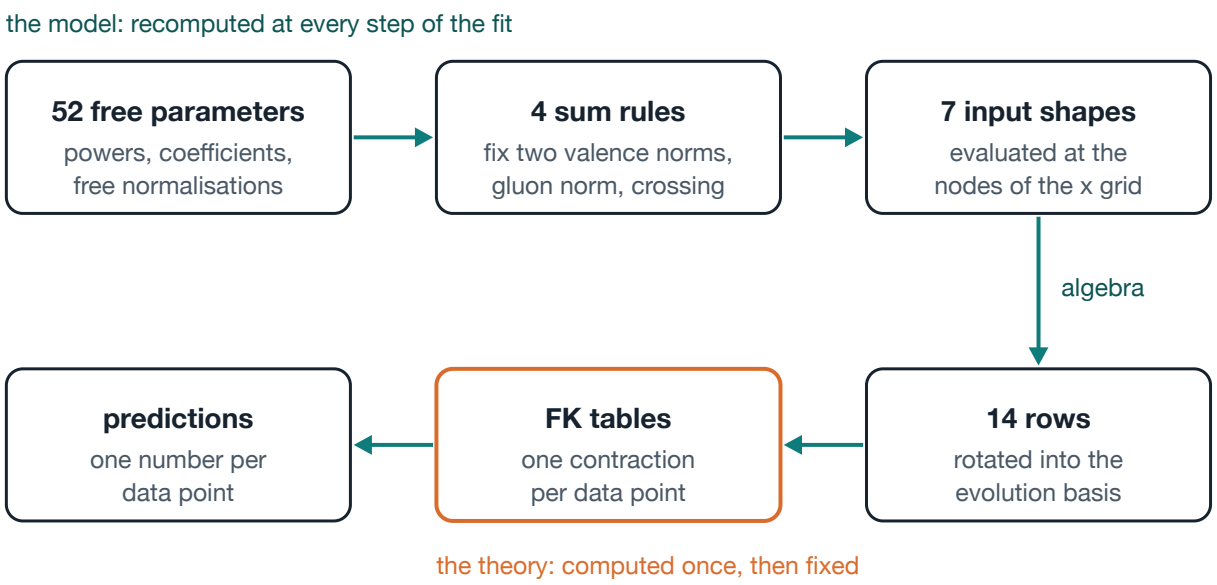

Figure 3. One evaluation of the model. The upper row and the rotation depend on θ. The FK tables (orange) contain the evolution from Q_0 and the hard cross-sections; they do not depend on θ and are loaded once.

The grid

Chapter 1.5 showed that a distribution stored at nodes $x_1 < \cdots < x_{n_x}$ and reconstructed by interpolation turns evolution into a matrix, and that folding the matrix with the hard cross-section gives the FK table. The interpolation is done inside the table, when it is built. At fit time the model is asked only for its values at the nodes. In Colibri the fit grid is the sorted union of the grids of all the tables in the fit, and the model is one function that maps θ to an array of shape $14 \times n_x$.

The evolution basis

FK tables take their input in the evolution basis of chapter 1.5. The model fills 14 rows, in this order:

$$\big(\gamma,\ \Sigma,\ g,\ V,\ V_3,\ V_8,\ V_{15},\ V_{24},\ V_{35},\ T_3,\ T_8,\ T_{15},\ T_{24},\ T_{35}\big). \tag{2.1.24}$$

With $q^{\pm} = q \pm \bar{q}$, the light-flavour rows are $\Sigma = u^+ + d^+ + s^+$, $V = u^- + d^- + s^-$, $V_3 = u^- - d^-$, $V_8 = u^- + d^- - 2s^-$, $T_3 = u^+ - d^+$ and $T_8 = u^+ + d^+ - 2s^+$. In terms of the fitted shapes, $\Sigma = u_V + d_V + S$, $V = u_V + d_V + s_-$, $T_8 = \Sigma - 3s_+$ and $V_8 = V - 3s_-$. With no heavy quarks at Q_0 the remaining rows carry no new information: $V_{15} = V_{24} = V_{35} = V$ and $T_{15} = T_{24} = T_{35} = \Sigma$. The photon row is zero.

A row left at zero is a physical statement, and the contraction will multiply it into the data's sensitivity without complaint.

The contraction

For one incoming hadron the prediction is linear in the array. For two it is quadratic:

$$\begin{aligned} T_I^{\mathrm{DIS}}(\theta) &= \sum_{j,k} \mathrm{FK}_{I,jk}\ f_j(x_k;\theta), \\ T_I^{\mathrm{had}}(\theta) &= \sum_{j,k}\sum_{l,m} \mathrm{FK}_{I,jk,lm}\ f_j(x_k;\theta)\, f_l(x_m;\theta). \end{aligned} \tag{2.1.25}$$

Here $f_j(x_k;\theta)$ is $x_k f_j(x_k, Q_0^2)$ for row j. In Colibri these are two `einsum` calls. The check script builds both kinds of table for toy kernels on an illustrative grid of 50 nodes, with linear interpolation in $\ln x$. The contraction reproduces the convolution of the interpolated distribution to eight digits or better, and the convolution of the exact shape to 1% and 2%, which is the interpolation error of so crude a scheme.

Why this is fast, and what it implies

- **Cost.** With $n_x = 50$, a deep inelastic point costs $14 \times 50 = 700$ multiplications and a hadronic point at most $700^2 = 490{,}000$. No integral, no splitting function and no value of α_S appears at fit time.
- **Derivatives are cheap.** The tables are constants, so the derivative of a prediction is the same contraction applied to the derivative of the model:

$$\frac{\partial T_I^{\mathrm{DIS}}}{\partial \theta_a} = \sum_{j,k} \mathrm{FK}_{I,jk}\, \frac{\partial f_j(x_k;\theta)}{\partial \theta_a}. \tag{2.1.26}$$

JAX differentiates the model automatically, including the Gamma functions of the sum rules. Gradient-based minimisation, the Hessian and Hamiltonian Monte Carlo all rest on this.

- **Linear in f, not in θ.** Deep inelastic predictions are linear in the distributions, but the distributions are not linear in the parameters: δ and η sit in exponents, and the sum rules divide by integrals. The posterior in θ is therefore not Gaussian, even for Gaussian data errors (chapter 2.3).

---

8. Parametrisation bias and the price of flexibility

A family that is too small cannot reach the truth. A family that is too large fits the noise. This section makes both statements quantitative for a model that is linear in its parameters, and illustrates them with a toy.

A linear model

Let the family be $xf(x;\theta) = \sum_{a=1}^{p} \theta_a \, \phi_a(x)$ with fixed basis functions ϕ_a, and let the data measure xf directly at points x_I with independent errors σ_I. In units of the errors, the design matrix, the true values and the least-squares fit are

$$\begin{aligned} J_{Ia} &= \frac{\phi_a(x_I)}{\sigma_I}, \qquad b_I = \frac{xf_{\text{true}}(x_I)}{\sigma_I}, \\ \hat{\theta} &= (J^T J)^{-1} J^T d, \qquad d = b + n, \quad n \sim \mathcal{N}(0,1). \end{aligned} \tag{2.1.27}$$

This J is the Jacobian $\partial T_I / \partial \theta_a$ of chapter 2.3 divided by the errors, so $J^T J$ is the Fisher matrix F of that chapter and $(J^T J)^{-1}$ the parameter covariance V. The matrix $P = J (J^T J)^{-1} J^T$ projects onto the p-dimensional subspace of data vectors the family can produce. The expected minimum of χ^2 is then (derivation D5)

$$\left\langle \chi^2_{\min} \right\rangle = N_{\text{dat}} - p + b^T (1 - P)\, b. \tag{2.1.28}$$

The last term is the squared distance, in units of the errors, between the truth and the nearest member of the family. It is the *parametrisation bias* as the data see it. If the truth lies in the family it vanishes. A χ^2 per point well above one signals a family too rigid for the data.

The uncertainty band the fit reports at any x, and its average over the data points, are

$$\sigma_f^2(x) = \phi(x)^T (J^T J)^{-1} \phi(x), \qquad \frac{1}{N_{\text{dat}}} \sum_I \frac{\sigma_f^2(x_I)}{\sigma_I^2} = \frac{p}{N_{\text{dat}}}. \tag{2.1.29}$$

Two consequences follow.

- **Variance grows with the number of parameters.** Averaged over the data, the fitted curve is known N_{dat}/p times better than a single data point, whatever the basis.
- **The band contains no bias.** $\sigma_f(x)$ depends on the basis functions and the errors, not on f_{true}. Where there are no data, it is set by how the basis functions continue outside the data region.

A toy

The check script applies this to 60 pseudo-data points with 3% errors, spread logarithmically over $10^{-3} < x < 0.75$. The truth is $xf = 2.2\, x^{0.5}(1-x)^3(1+3x)$. The family has fixed powers that differ from the truth's, $\phi_a = x^{0.34}(1-x)^{3.75}\, T_{a-1}(y)$, so the truth is outside it for every p.

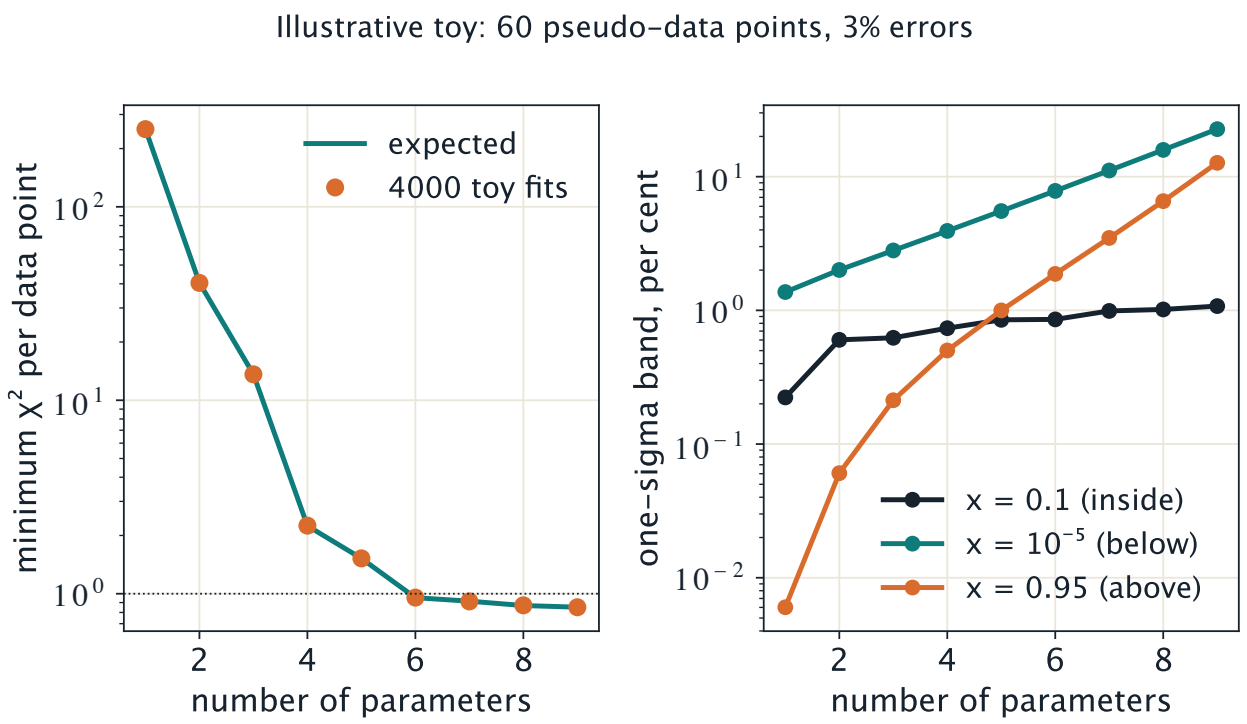

Figure 4. An illustrative toy, not a fit to data. Left: the expected minimum χ^2 per point from Eq. (2.1.28) and the mean of 4000 simulated fits. Right: the one-sigma band of Eq. (2.1.29) at one point inside the data region and at two points outside it.

With one parameter χ^2 per point is 252. With six it is 0.955 (expected 0.954), and the fit is statistically acceptable. At that point:

WHERE	BAND	ACTUAL ERROR OF THE MEAN FIT
$x = 0.1$, inside the data	0.9%	$+0.3\%$
$x = 10^{-5}$, below the data	7.8%	$+47\%$
$x = 0.95$, above the data	1.9%	-57%

Inside the data region the band is honest. Outside it the fit is wrong by six and by thirty times its quoted uncertainty, with an acceptable χ^2. The data cannot reveal the problem, because the fixed powers $x^{0.34}$ and $(1-x)^{3.75}$ are extrapolated with no freedom. Adding parameters widens the bands outside the data (to 23% and 13% at $p = 9$) without removing the error (still $+43\%$ and -45%).

What fitting groups do about it

The Snowmass report states the requirement as a form flexible "just enough" to describe the data without overfitting [4]. It lists the responses: MSHT20 moved to a more flexible form, which changed its down-valence distribution; the CT18 analysis tested more than 250 alternative forms to estimate the uncertainty due to the choice; and a cross-validation test has been proposed to choose the number of parameters.

Freeing the exponents, as the MSHT20 form does, removes the specific failure of the toy. The general point stands: outside the data, a fixed form answers with its own functional shape.

9. Neural networks

NNPDF takes the opposite approach to the trade-off of section 8: make the family so large that bias is negligible, and control the variance by other means [2, 4].

The form

Each independent distribution at Q_0 is a neural network times a power-law prefactor:

$$f_j(x, Q_0^2) = A_j \, x^{-n_j} \, (1-x)^{m_j} \, \mathrm{NN}_j(x). \tag{2.1.30}$$

The network is a multilayer feed-forward perceptron. Each neuron forms a weighted sum of the outputs of the previous layer, subtracts a threshold, and applies a sigmoid $g(h) = 1/(1+e^{-h})$. The inputs are x and $\ln x$. A network with $n(1), \ldots, n(L)$ neurons per layer has

$$N_{\mathrm{par}} = \sum_{l=1}^{L-1} n(l+1)\,\big[1 + n(l)\big] \tag{2.1.31}$$

weights and thresholds. The architecture in Ubiali's thesis is 2–5–3–1, which gives $5 \cdot 3 + 3 \cdot 6 + 1 \cdot 4 = 37$ parameters per distribution. With seven independent distributions (Σ, g, V, T_3, $\bar{d} - \bar{u}$, s_+ and s_-) the total is 259, against fewer than 30 in the fixed-form fits of the time [2]. The project ledger records NNPDF4.0 as a 763-parameter network.

A network of this kind can approximate any continuous function, given enough neurons [2]. Ubiali's thesis calls the parametrisation *redundant*, and intends it: redundancy removes the bias of a functional form where the data do not constrain the result.

Preprocessing

The prefactor $x^{-n}(1-x)^{m}$ is the pair of power laws of Eq. (2.1.3) in a different role. It divides out the dominant end-point behaviour so that a small network has only the deviations to learn, and it is not fitted. The exponents are drawn at random for each replica from a fixed range; for the singlet in the analysis described in the thesis, $m \in [2.55, 3.45]$ and $n \in [1.05, 1.35]$. The χ^2 of the replicas is almost uncorrelated with the exponents, with correlation coefficients all below 0.14 in magnitude, which is the check that the choice does not bias the result [2].

The Snowmass report adds that the prefactor is not strictly necessary, and that it speeds up convergence in the extrapolation regions "by providing a sensible prior" [4]. Where there are no data, a neural-network fit also extrapolates with the shape its prior gives it.

Sum rules and overfitting

The network's integrals have no closed form, so the sum rules are imposed by numerical integration. The momentum sum rule fixes the gluon normalisation as [2]

$$A_g = \frac{1 - \int_0^1 dx\; x\,\Sigma(x)}{\int_0^1 dx\; x\,(1-x)^{m_g} x^{-n_g}\,\mathrm{NN}_g(x)}, \tag{2.1.32}$$

which is Eq. (2.1.20) with quadrature in place of Gamma functions. Positivity and integrability enter through the penalties of section 6.

With hundreds of parameters, the minimum of χ^2 is not the answer: by Eq. (2.1.29) a family this large would follow the noise. NNPDF stops the minimisation early, using a validation subset of the data (chapter 2.4), and represents the uncertainty by an ensemble of fits to fluctuated data (chapter 2.3). Whether the resulting uncertainties are faithful is tested by closure tests on pseudo-data generated from a known truth [4]. The fixed form controls overfitting by having few parameters and risks bias from the form; the network removes that bias and depends instead on the training procedure.

---

10. In this project

Every number in this section is quoted from the project ledgers [7], with its result ID.

The port. The MSHT20 form was ported to JAX as a Colibri model with "52 free params, 4 analytic sum rules" (PPDF-10). It was tested against the reference implementation with 96 checks at four levels (integrals, shapes, sum-rule machinery, flavour assembly), at the starting point and at five random $\pm 10\%$ variations of it; the worst deviation was $\lesssim 10^{-9}$. The random variations caught an error the starting point hides: a factor in the gluon integral of Eq. (2.1.20) taken from the slot that holds A_g, which contains the placeholder 1 at the starting point.

Flexibility was the difference on real data. On the same 648 deep inelastic points, a rigid 13-parameter form reached a minimum $\chi^2/N_{\mathrm{dat}} = 5.60$ (PPDF-7) and the MSHT20 form reached 1.002 (PPDF-12). This is the bias term of Eq. (2.1.28) in a real fit.

The form carries the global dataset. On 82 datasets and 4,615 points the minimum is $\chi^2/N_{\mathrm{dat}} = 1.2368$; the ledger sets this beside the published NNPDF4.0 value of 1.17 (PPDF-31). A repeat of the same fit on different hardware gave 1.2380 (PPDF-35), the value quoted in chapter 2.2.

The penalties of section 6 are needed. The same fit without the positivity and integrability penalties gave $\chi^2/N_{\mathrm{dat}} = 16.35$, and a non-finite χ^2 on the first rung of 19 datasets and 3,089 points (PPDF-31). The ledger's reading is that the penalties act as a regulariser: without them the polynomial form reaches regions where the likelihood is not finite. The two penalties were switched off together, so the failure cannot be attributed to positivity alone. The ledger also notes that the starting point "was never once tested against positivity or integrability" before Phase B, which agrees with the sign changes found in section 6.

Not all 52 directions are constrained. The Fisher analysis of Phase B counts the flat directions among the 52 as data are added: $12 \to 9 \to 8 \to 8$ along the four dataset rungs (PPDF-32, as corrected in the ledger; the first count, from a forward-difference Hessian, read $14 \to 12 \to 10 \to 10$). A form flexible enough to remove the bias has directions the data do not determine. Chapters 2.3 and 3.4 take this up.

Derivations

D1. The small-x power from Regge behaviour

Derives (2.1.4)

The cross-section for a virtual photon. Chapter 1.1 expresses the structure functions through the cross-sections σ_T and σ_L for absorbing a transverse or longitudinal virtual photon. Adding the two relations given there, and using $F_2 = 2xF_1 + F_L$ and $2x\nu = Q^2$,

$$\sigma_T + \sigma_L = \frac{4\pi^2\alpha}{\nu(1-x)}\left[F_1 + \frac{F_L}{2x}\right] = \frac{4\pi^2\alpha}{Q^2(1-x)}\,F_2 \xrightarrow{x\to 0} \frac{4\pi^2\alpha}{Q^2}\,F_2(x,Q^2).$$

So at fixed Q^2, F_2 at small x is the total photon–proton cross-section up to a constant.

The energy. The invariant mass of the photon–proton system is $W^2 = (P+q)^2 = M^2 + Q^2(1-x)/x$. For $x \to 0$ at fixed Q^2, $W^2 \simeq Q^2/x \to \infty$.

The Regge power. In Regge theory the forward elastic amplitude at high energy is a sum over exchanged trajectories, $\mathcal{A}(s, t=0) \simeq \sum_i \beta_i \, s^{\alpha_i(0)}$, with $s = W^2$. The optical theorem relates the total cross-section to its imaginary part, $\sigma_{\text{tot}} = \operatorname{Im}\mathcal{A}(s,0)/s$ at high energy. Hence

$$\sigma_{\text{tot}}(W^2) \simeq \sum_i \beta_i' \, (W^2)^{\alpha_i(0)-1}.$$

The result. Substituting $W^2 \simeq Q^2/x$, each trajectory contributes $F_2 \propto x^{1-\alpha_i(0)}$ at fixed Q^2. Since F_2 is x times a sum of quark distributions, the momentum densities xq contributing to that term behave as $x^{1-\alpha_i(0)}$. Comparing with Eq. (2.1.3) gives $\delta = 1 - \alpha(0)$, Eq. (2.1.4).

What is assumed. The argument takes $W^2 \to \infty$ at fixed Q^2 and applies a description built for hadronic cross-sections to a virtual photon. It says nothing about how the power changes with Q^2.

D2. The Beta integral

Derives (2.1.15)

Product of two Gamma functions. Start from $\Gamma(a+1)\,\Gamma(b+1) = \int_0^\infty du \int_0^\infty dv \; u^a v^b e^{-u-v}$, which converges for $a, b > -1$.

Change of variables. Put $u = zt$ and $v = z(1-t)$, with $0 < z < \infty$ and $0 < t < 1$. Then $u + v = z$, and the Jacobian is $\partial(u,v)/\partial(z,t) = z$, so $du\,dv = z\,dz\,dt$:

$$\Gamma(a+1)\,\Gamma(b+1) = \int_0^\infty dz \; z^{a+b+1} e^{-z} \int_0^1 dt \; t^a (1-t)^b = \Gamma(a+b+2)\; I(a,b).$$

This is Eq. (2.1.15). **Near the boundary**, $a = \delta - 1$ with $\delta \to 0$: $\Gamma(\delta) \simeq 1/\delta$, so $I(\delta - 1, \eta) \simeq 1/\delta$.

D3. Chebyshev polynomials in $\sqrt{x}$ and their moments

Derives (2.1.16), (2.1.17)

The differential equation. Differentiate $T_n(\cos\theta) = \cos n\theta$ twice with respect to θ. With $y = \cos\theta$, $d/d\theta = -\sin\theta\, d/dy$, and the result is the Chebyshev equation

$$(1-y^2)\,T_n''(y) - y\,T_n'(y) + n^2\,T_n(y) = 0.$$

Change of variable. Put $y = 1 - 2s$, where $s = \sqrt{x}$. Then $d/dy = -\tfrac{1}{2}\,d/ds$ and $1 - y^2 = 4s(1-s)$. For $F(s) = T_n(1-2s)$ the equation becomes

$$s(1-s)\,F''(s) + \left(\tfrac{1}{2} - s\right)F'(s) + n^2 F(s) = 0.$$

Series solution. Substitute $F = \sum_k \tau_k \, s^k$ and collect the coefficient of s^k:

$$(k+1)k\,\tau_{k+1} - k(k-1)\,\tau_k + \tfrac{1}{2}(k+1)\,\tau_{k+1} - k\,\tau_k + n^2\tau_k = 0 \quad\Longrightarrow\quad \tau_{k+1} = \frac{k^2-n^2}{(k+1)(k+\frac{1}{2})}\,\tau_k.$$

The starting value is $\tau_0 = F(0) = T_n(1) = 1$. The numerator vanishes at $k = n$, so the series stops at s^n and is the polynomial itself. This is Eq. (2.1.16). The first steps give $\tau_{n,1} = -2n^2$, which reproduces $T_1 = 1 - 2s$, and for $n = 2$: $\tau_1 = -8$, $\tau_2 = -8 \cdot (1-4)/(2 \cdot \frac{3}{2}) = 8$.

The moments. Since $s^k = x^{k/2}$, multiplying by $x^a(1-x)^b$ and integrating term by term with Eq. (2.1.15) gives Eq. (2.1.17). The check script compares the series with the recurrence (2.1.8) for $n = 1 \ldots 8$ and the moments with numerical quadrature; they agree to better than one part in 10^8.

D4. The four constrained parameters

Derives (2.1.18), (2.1.19), (2.1.20)

Throughout, $J(a, b; c)$ is the integral of Eq. (2.1.17).

Valence number. Equation (2.1.10) gives xu_V, so $u_V = A_{u_V} x^{\delta_u - 1}(1-x)^{\eta_u}[1 + \ldots]$ and

$$2 = \int_0^1 dx\, u_V = A_{u_V}\, J(\delta_u - 1, \eta_u; c_u).$$

Solving for A_{u_V} gives Eq. (2.1.18). The same step with 1 in place of 2 gives A_{d_V}. The integral exists for $\delta_u > 0$.

Strangeness. From Eq. (2.1.12),

$$0 = \int_0^1 dx\, s_- = A_-\Big[I(\delta_- - 1, \eta_-) - \frac{1}{x_0}\, I(\delta_-, \eta_-)\Big] \quad\Longrightarrow\quad x_0 = \frac{I(\delta_-, \eta_-)}{I(\delta_- - 1, \eta_-)}.$$

Insert Eq. (2.1.15) and use $\Gamma(z+1) = z\,\Gamma(z)$ twice:

$$x_0 = \frac{\Gamma(\delta_- + 1)}{\Gamma(\delta_-)} \cdot \frac{\Gamma(\delta_- + \eta_- + 1)}{\Gamma(\delta_- + \eta_- + 2)} = \frac{\delta_-}{\delta_- + \eta_- + 1}.$$

This is Eq. (2.1.19).

Momentum. The momentum integral of each quark-like shape is $\int_0^1 dx\, xf = A_f\, J(\delta_f, \eta_f; c_f) = \langle x \rangle_f$, with the power unshifted. By Eq. (2.1.13) and the definition of S, the sum of all quarks and antiquarks is

$$\Sigma = u_V + d_V + 2(\bar{u} + \bar{d}) + s + \bar{s} = u_V + d_V + S,$$

so s_+, s_- and ρ do not appear in the momentum sum: they redistribute momentum within the sea. The two gluon terms integrate to $A_g\, J(\delta_g, \eta_g; c_g)$ and $A_{g-}\, I(\delta_{g-}, \eta_{g-})$. The last constraint of Eq. (2.1.14) is then

$$\langle x \rangle_{u_V} + \langle x \rangle_{d_V} + \langle x \rangle_S + A_g\, J(\delta_g, \eta_g; c_g) + A_{g-}\, I(\delta_{g-}, \eta_{g-}) = 1,$$

and solving for A_g gives Eq. (2.1.20).

Order of evaluation. A_g depends on the valence normalisations through the momentum fractions, so it is computed last, as in the project's `sumrules` function.

D5. Expected χ^2 and the band of a linear fit

Derives (2.1.28), (2.1.29)

Set-up. In the units of Eq. (2.1.27) the data vector is $d = b + n$ with $\langle n \rangle = 0$ and $\langle n n^T \rangle = 1$, and $\chi^2(\theta) = |d - J\theta|^2$.

The minimum. Setting the gradient $-2J^T(d - J\theta)$ to zero gives $\hat{\theta} = (J^T J)^{-1} J^T d$. The fitted values are $J\hat{\theta} = Pd$ with $P = J(J^T J)^{-1} J^T$. This matrix satisfies $P^2 = P$ and $P^T = P$, so it is an orthogonal projector, and $\mathrm{tr}\, P = \mathrm{tr}\left[(J^T J)^{-1} J^T J\right] = p$.

The expected minimum. The residual is $(1 - P)d$, and $1 - P$ is also a projector, so

$$\chi^2_{\min} = d^T (1 - P)\, d = b^T (1 - P) b + 2\, b^T (1 - P) n + n^T (1 - P) n.$$

The middle term averages to zero. The last averages to $\mathrm{tr}(1 - P) = N_{\mathrm{dat}} - p$. This is Eq. (2.1.28). If the truth is in the family, $b = J\theta_{\mathrm{true}}$ and $(1 - P)J = 0$, so the first term vanishes.

The band. $\hat{\theta} - \langle \hat{\theta} \rangle = (J^T J)^{-1} J^T n$, so the covariance of the fitted parameters is $(J^T J)^{-1} J^T \langle n n^T \rangle J (J^T J)^{-1} = (J^T J)^{-1}$. The fitted function at any x is $\phi(x)^T \hat{\theta}$, with variance $\phi(x)^T (J^T J)^{-1} \phi(x)$: the first part of Eq. (2.1.29). At a data point, $\phi(x_I) = \sigma_I J_{I\cdot}$, so $\sigma_f^2(x_I)/\sigma_I^2 = P_{II}$, and the average over the data is $\mathrm{tr}\, P / N_{\mathrm{dat}} = p / N_{\mathrm{dat}}$.

The bias of the fit. The mean fitted function is $\phi(x)^T (J^T J)^{-1} J^T b$. Its difference from $x f_{\mathrm{true}}(x)$ is the third column of the table in section 8. It does not appear in σ_f. The check script simulates 4000 fits for each p from 1 to 9 and confirms Eqs. (2.1.28) and (2.1.29).

D6. The large-x power under evolution

Derives (2.1.6)

Large x is large N. By Eq. (2.1.15) the moment of a pure power is $\int_0^1 dx\, x^{N-1}(1 - x)^\eta = \Gamma(N)\Gamma(\eta + 1)/\Gamma(N + \eta + 1)$. For $N \gg \eta$, $\Gamma(N)/\Gamma(N + \eta + 1) \simeq N^{-\eta - 1}$, so

$$q(x) \propto (1 - x)^\eta \ \text{ as } x \to 1 \qquad \Longleftrightarrow \qquad q(N) \propto N^{-\eta - 1} \ \text{ as } N \to \infty.$$

The anomalous dimension at large N. In Eq. (1.4.26), $\sum_{k=1}^{N} 1/k \simeq \ln N$, so $\gamma_{qq}(N) \simeq -2 C_F \ln N$.

Evolution. By Eq. (1.4.28) a non-singlet moment is multiplied by $[\alpha_S(\mu_0^2)/\alpha_S(\mu^2)]^{\gamma_{qq}/2\pi b}$. With $\zeta = \ln[\alpha_S(\mu_0^2)/\alpha_S(\mu^2)]$,

$$q(N, \mu^2) \propto N^{-\eta - 1} \exp\left[-\frac{C_F}{\pi b} \zeta \ln N \right] = N^{-\eta - 1 - C_F \zeta / \pi b}.$$

Reading the exponent back with the first step gives Eq. (2.1.6). The result is for a non-singlet (valence) distribution at leading order. For $n_f = 4$, $b = 25/12\pi$ and $C_F/\pi b = 16/25 = 0.64$.

References

1. J. M. Moore, *Parton Distributions in Beyond the Standard Model Theories*, PhD thesis, University of Cambridge, sections 1.2.2–1.2.4, 1.3.1 and 1.3.2 (the positivity penalty is its Eqs. 1.104 and 1.105).

2. M. Ubiali, *A new approach in the determination of Parton Distribution Functions. Application to processes with heavy quarks in the initial states*, PhD thesis, Université catholique de Louvain and University of Edinburgh (2010), sections 3.1.2 (its Eqs. 3.10 and 3.11, Tables 3.2 and 3.3) and 3.1.6 (its Eqs. 3.28–3.33).
3. R. K. Ellis, W. J. Stirling and B. R. Webber, *QCD and Collider Physics*, Cambridge University Press (1996), chapter 4. Open access under CC BY-NC 4.0, doi:10.1017/CBO9780511628788. Background for the sum rules and for derivation D6.
4. S. Amoroso et al., "Snowmass 2021 whitepaper: Proton structure at the precision frontier", arXiv:2203.13923.
5. A. Donnachie and P. V. Landshoff, "Total cross sections", Phys. Lett. B296 (1992) 227.
6. S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin and R. S. Thorne, "Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs", Eur. Phys. J. C 81 (2021) 341, arXiv:2012.04684.
7. Project ledgers `context/RESULTS_LEDGER.md` (PPDF-7, PPDF-10, PPDF-12) and `context/PHASE_B_LEDGER.md` (fixed setup and known seams, PPDF-31, PPDF-32, PPDF-35), and the project's port `analysis/msht20/msht20_model`.