Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Plug-and-Play ADMM for Distributed Multifractal Analysis of Large-Scale Time Series

Sawon Pratiher

sawon1234@gmail.com

Howrah https://orcid.org/0000-0001-8471-8148

Research Article

Keywords: Multifractal analysis, scaling exponents, distributed estimation, decentralized optimization, ADMM, plug-and-play priors, wavelet leaders, detrended fluctuation analysis.

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11875332/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Plug-and-Play ADMM for Distributed Multifractal Analysis of Large-Scale Time Series

Sawon Pratiher

***Abstract*—Multifractal analysis describes how the local regularity of a signal fluctuates in time, but its estimators assume that one machine holds the whole record. When a long record is split into shards stored on different machines or sensors, the common practice of analyzing each shard and averaging the results is biased: logarithms of local moments are averaged instead of logarithms of global ones, the bias does not vanish as shards of fixed length are added, and the result depends on the storage layout. We develop a distributed framework that reproduces the centralized estimate exactly while each node exchanges only a small parameter vector with its neighbors. First, we show that detrended fluctuation, moving-average, wavelet-coefficient, wavelet-leader, continuous-wavelet and cross-correlation estimators share one property: their partition functions are sums of terms that depend on bounded windows of the signal, so a one-time halo exchange makes their decomposition across nodes exact. Second, we recast scaling-exponent regression as a separable convex quasi-likelihood problem whose minimizer depends only on pooled moments. Third, we solve it with a decentralized alternating direction method of multipliers in which the multifractal estimator, the prior on the scaling function and the local solver are interchangeable modules. We prove convergence for every convex prior and linear convergence for smooth priors, with explicit network dependence, and derive the costs. Experiments with up to 33 million samples, six network topologies and 7.7 million intraday exchange-rate and stock-index returns confirm exactness, the predicted rates and the layout dependence of shard averaging.**

***Index Terms*—Multifractal analysis, scaling exponents, distributed estimation, decentralized optimization, ADMM, plug-and-play priors, wavelet leaders, detrended fluctuation analysis.**

I. Introduction

MULTIFRACTAL analysis characterizes a signal by the way its local regularity fluctuates in time [1]–[3]. In practice it measures how q-th order moments of multiscale quantities, such as detrended fluctuations [4], moving-average residuals [5], wavelet coefficients [6], [7] or wavelet leaders [3], [8], scale with the analysis scale s. The exponents $\zeta(q)$ of these power laws and their Legendre transform, the multifractal spectrum, are used in physiology [9], [10], finance [11], [12], turbulence [13] and texture analysis [14]–[16]. Toolboxes collect many variants of these estimators [10], [17].

All of these estimators assume that one machine holds the whole record. That assumption fails more and more often. Intraday financial archives, long-term physiological monitoring and sensor networks produce records with 10^7–10^9 samples that are stored as shards (one file per day or month, one segment per sensor or storage node), and the shards may not be moved because of bandwidth, cost or privacy [18], [19]. The usual workaround is *divide and conquer*: analyze every shard on its own and average the shard estimates of $\zeta(q)$ or of the spectrum. This workaround is biased. A scaling exponent is a slope in a log-log plot, and the logarithm of a local mean moment is a biased estimate of the logarithm of the global mean moment. The bias is inversely proportional to the number of terms in a shard, so it does not vanish as more shards of a fixed length are added (Proposition 2). At large $|q|$ the moments are dominated by rare events, and each shard sees fewer of them than the full record does, so estimates linearize in q earlier [20]–[22]. The result therefore depends on how the record happens to be stored. For a monofractal signal, shard averaging produces spurious multifractality that grows with the number of shards (Section VII-C).

This paper asks whether a network of nodes, each holding one shard and talking only to its neighbors, can compute the estimate that a single machine would compute on the full record, at a small communication cost, and with the freedom to choose both the multifractal estimator and the prior on the scaling function. We answer yes, through three steps (Fig. 1).

Exact decomposition. We isolate a property that is shared by the estimators listed in [17]: their partition functions are sums of nonnegative terms, and each term depends only on a bounded window of the profile. We call such estimators *s-local* (Assumption 1). With a one-time *halo* exchange of at most a few maximal scales of samples between time-adjacent nodes, the global partition function is exactly the sum of the node partition functions (Proposition 1). Without the halo, the error is controlled by the fraction of lost terms and by a moment-dominance ratio that grows quickly with $|q|$.

Separable convex formulation. The centralized regression of $\log m(q,s)$ on $\log s$ is not separable over nodes, because the logarithm acts on a sum. We replace it by a quasi-likelihood (QL) criterion [23], [24] with a log link. The QL criterion is a sum of node terms, it is convex, its minimizer depends on the data only through pooled moments, and it differs from the log-log least-squares estimate only by terms that are quadratic in the regression residuals (Proposition 3). Physical constraints, such as concavity of $\zeta(q)$, enter as a convex prior.

Plug-and-play decentralized ADMM. We solve the resulting consensus problem with a decentralized alternating direction method of multipliers (ADMM) [25]–[27]. Three components are *plug-and-play*: the multifractal front-end, the prior on the scaling function (through its proximal operator, or a denoiser as in [28], [29]), and the local solver (an exact Newton step or a single linearized gradient step). Every iteration exchanges $2Q$ numbers per neighbor, where Q is the number of moment orders, independently of the length of the series and of the number of scales.

Contributions. (i) We show that the six main estimator families in common toolboxes (MFDFA, MFDMA, MODWT structure functions, wavelet leaders, CWT modulus maxima and MF-DXA) admit an exact distributed decomposition, and

Submitted to a journal for review. Content may be updated post-review. Subject to journal policies regarding preprint hosting. (E-mail: sawon1234@gmail.com.). Supporting code and data are available from https://github.com/Sawon1234/ADMM-for-Distributed-Multifractal-Analysis.

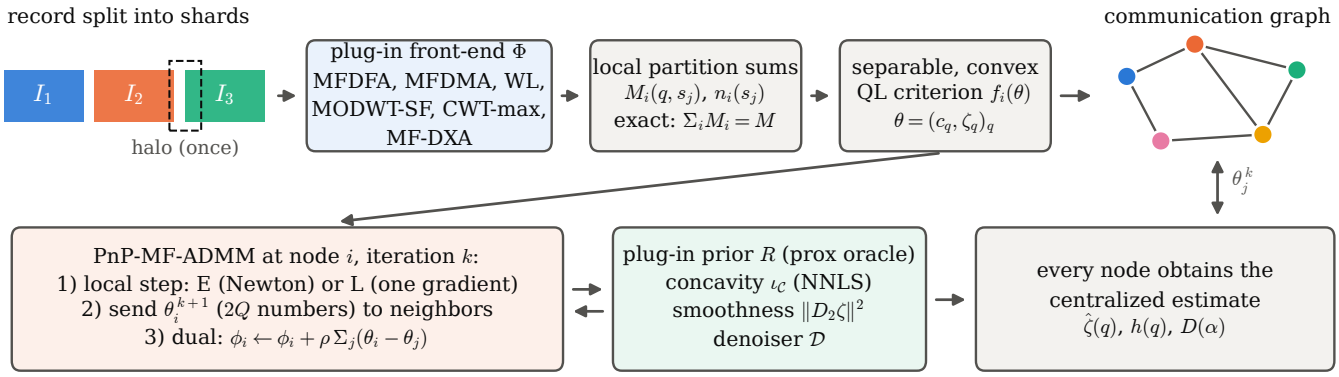


Figure 1: PnP-MF-ADMM. Each node holds one shard and, after a one-time halo exchange, computes exact local partition sums with any s-local front-end. The nodes then solve a separable quasi-likelihood problem by decentralized ADMM, exchanging $2Q$ numbers per neighbor and iteration; the prior on the scaling function enters only through a proximal oracle. Every node converges to the estimate of a single machine that holds the whole record.

we quantify what is lost without a halo. (ii) We give the first analysis, to our knowledge, of the bias of shard averaging in multifractal estimation: to leading order it is K times the bias of the exact estimator for K equal shards, and it makes the estimate depend on the storage layout. (iii) We propose a separable QL formulation and the PnP-MF-ADMM algorithm. (iv) We prove global convergence for any proper closed convex prior and linear convergence for smooth priors, with explicit constants in the network's algebraic connectivity, and we give an error decomposition, iteration complexity and computation and communication costs. (v) We validate the framework on synthetic series of up to 2^{25} samples, on six network families with up to 256 nodes, against four baselines, and on 5.6 million EUR/USD and 2.1 million S&P 500 intraday returns.

Organization. Section II reviews related work. Section III defines s-local front-ends and proves the exact decomposition and the bias of shard averaging. Section IV gives the QL formulation and the priors. Section V presents the algorithm and its cost, and Section VI its convergence. Section VII reports experiments, and Section VIII concludes. Proofs, implementation details and additional experiments are in the supplementary material; items numbered with the prefix S are found there.

Notation. $\|\cdot\|$ is the Euclidean norm, and $\|v\|_G^2 = v^\top G v$ for $G \succeq 0$. $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ are extreme eigenvalues. For a graph with adjacency matrix A and degree matrix D, $L_- = D - A$ is the Laplacian and $L_+ = D + A$ the signless Laplacian; λ_2 is the algebraic connectivity, the smallest nonzero eigenvalue of L_-. Stacked node variables are written in bold, $\boldsymbol{\theta} = [\theta_1; \ldots; \theta_K]$, and $\otimes$ is the Kronecker product.

II. Related Work

Multifractal estimators. The partition-function formalism [1] was adapted to signals through wavelet modulus maxima (WTMM) [2], [6], detrended fluctuation analysis (MFDFA) [4], detrending moving averages (MFDMA) [5], and cross-correlation variants (DCCA and MF-DXA) [30], [31]. Wavelet leaders [3], [8], [13], [32] and p-leaders [33], [34] are now the reference for negative moments and for the theory of the formalism. The linearization effect [20], [21] and the statistics of cascade estimators [22] show that the range of reliably estimated q grows with the sample size. That fact is why shard averaging fails. Fast single-machine algorithms [35], [36] reduce the cost of DFA and DMA but still need the whole record on one machine.

Optimization for scaling analysis. Convex formulations have been used to regularize *local* regularity estimates in images, through total variation and proximal algorithms [16], [37]. Bayesian models have been used for the multifractality parameter [15]. These works operate on one machine. We instead regularize the *global* scaling function across q and distribute the computation.

Decentralized optimization. Consensus-based estimation over networks [38], [39] and decentralized ADMM [26], [40], with its linearized variants DLM [27] and IC-ADMM [41], are mature. EXTRA, PG-EXTRA [42], [43], decentralized gradient descent [44] and proximal methods with a common nonsmooth term [45] are first-order alternatives. Average consensus [46] can pool sufficient statistics. ADMM has also been used for nonconvex distributed problems such as network dissensus [47]. Plug-and-play (PnP) priors replace a proximal operator with a denoiser [28], [29], [48], [49]. Our contribution is to connect these tools to multifractal estimation through an exact decomposition and a separable convex criterion, with guarantees tailored to its structure.

III. Multifractal Front-Ends as Additive Partition Functions

A. Scaling Function and Spectrum

Let $x[n]$, $n = 0, \ldots, N-1$, be the record, $\bar{x}$ its mean and $Y[n] = \sum_{u \le n}(x[u] - \bar{x})$ its profile. A front-end Φ maps Y to nonnegative multiscale *terms* $T_v(s)$, attached to *anchors* v, at scales $s_1 < \cdots < s_J$. For moment orders $q \in \mathcal{Q} = \{q_1, \ldots, q_Q\}$, $q \neq 0$, it forms the partition function and the counts

$$M(q, s) = \sum_{v \in \mathcal{A}(s)} T_v(s)^{pq}, \qquad n(s) = |\mathcal{A}(s)|, \tag{1}$$

where $\mathcal{A}(s)$ is the set of anchors that carry a term and $p \in \{1/2, 1\}$ depends on whether T is a squared fluctuation or a magnitude. The mean moment $m(q, s) = M(q, s)/n(s)$ is assumed to follow the power law $m(q, s) \simeq C_q s^{\zeta(q)}$ over the scaling range. For MFDFA, $\zeta(q) = qh(q)$ with $h(q)$ the generalized Hurst exponent [4]. With the conventions used

Table I: s-Local Front-Ends: Exponent p, Anchor Spacing $\delta(s)$, Windows and Negative-q Support. $L_j = (2^j-1)(2L_f-1)+1$ Is the Level-j Filter Length for Daubechies Filters With L_f Vanishing Moments, and $c = \lfloor (s-1)\vartheta \rfloor$

Front-end	p	$\delta(s)$	$\lambda^-(s)$	$\lambda^+(s)$	$q<0$
MFDFAm [4]	$1/2$	s	0	$s-1$	yes
MFDMA(ϑ) [5]	$1/2$	s	$s-1-c$	$s-1+c$	yes
MODWT-SF [7]	1	2^j	L_j-1	0	no
WL [3]	1	2^j	$L_j-1+3\cdot 2^{j-1}$	$3\cdot 2^{j-1}$	yes
CWT-max [6]	1	1 (maxima)	$\lceil 5a \rceil + 1 + \lceil a \rceil$	same	yes
MF-DXAm [31]	$1/2$	s	0	$s-1$	yes

here (all front-ends act on the profile, and wavelet coefficients are L^1-normalized), the same $\zeta(q)$ is estimated by every front-end. The multifractal spectrum follows from the Legendre transform $\alpha(q) = \zeta'(q)$, $D(\alpha(q)) = 1 + q\alpha(q) - \zeta(q)$ [1], [3]. Its width $\Delta\alpha = \alpha(q_{\min}) - \alpha(q_{\max})$ is the usual summary of multifractality. The function ζ is concave and $\zeta(0) = 0$, so $h(q) = \zeta(q)/q$ is nonincreasing and $\alpha(q)$ is monotone.

B. s-Local Front-Ends

Assumption 1 (s-locality)**.** *For every scale s there are a candidate anchor grid $\mathcal{G}_N(s) \subset \{0,\dots,N-1\}$ that depends only on (N,s), and window lengths $\lambda^-(s), \lambda^+(s) \geq 0$. For every $v \in \mathcal{G}_N(s)$, the membership indicator $\chi_v(s) = \mathbb{1}[v \in \mathcal{A}(s)]$ and the term $T_v(s)$ are functions of Y restricted to $[v-\lambda^-(s), v+\lambda^+(s)]$ only. They are invariant to adding a constant to Y.*

Table I lists the estimators covered in this paper, their windows and their anchor grids; exact definitions are in Section S-II. For MFDFA of order m, the terms are the squared residuals $F_v^2(s)$ of an order-m polynomial fit over the segment $[v, v+s)$. The anchors are $v = ks$ (forward segments) and $v = (N \bmod s) + ks$ (backward segments, as in [4]). For MFDMA [5], the terms are mean squared residuals of the profile around a moving average with position parameter ϑ. The wavelet front-ends use the non-circular MODWT [7] of Y with Daubechies filters [50]: MODWT-SF uses coefficient magnitudes and WL uses leaders, i.e., suprema over finer scales and a time window of length $3\cdot 2^j$ [3]. CWT-max uses local maxima of a Mexican-hat transform with the same supremum, which is a WTMM-type partition function [6]. MF-DXA [31] uses $|\langle r_v^x, r_v^y \rangle|/s$ for detrended residuals of two profiles. Invariance to constants holds because the detrending removes them, because moving averages commute with constants, and because the wavelets have vanishing moments. For the CWT we subtract the mean of the discretized kernel so that this holds exactly.

C. Exact Decomposition Across Nodes

The record is split into contiguous shards $I_i = [a_i, b_i)$, $i \in \mathcal{V} = \{1,\dots,K\}$, with lengths $N_i = b_i - a_i$. Node i *owns* the anchors in I_i and computes

$$M_i(q,s) = \sum_{v \in \mathcal{A}(s) \cap I_i} T_v(s)^{pq}, \qquad n_i(s) = |\mathcal{A}(s) \cap I_i|. \quad (2)$$

Proposition 1 (Exact decomposition)**.** *Let Assumption 1 hold and let $\lambda^\pm = \max_j \lambda^\pm(s_j)$. Suppose node i knows N and $\bar{x}$, and holds x on the extended interval $[a_i - \lambda^-, b_i + \lambda^+) \cap [0, N)$, i.e., its shard plus a halo received once from its time-neighbors. Then $\sum_i M_i(q,s) = M(q,s)$ and $\sum_i n_i(s) = n(s)$ for all q and s. If instead node i holds only its shard and keeps only the anchors whose window lies inside it, let $\epsilon(s)$ be the fraction of lost terms and $R_q(s) = \max_v T_v(s)^{pq}/m(q,s)$ the moment-dominance ratio. The pooled mean moment $\tilde{m}(q,s)$ then satisfies*

$$\left| \frac{\tilde{m}(q,s)}{m(q,s)} - 1 \right| \leq \frac{\epsilon(s)\,(1+R_q(s))}{1-\epsilon(s)}, \quad (3)$$

with $\epsilon(s) \leq (K-1)\nu(s)/n(s)$, where $\nu(s)$ is the largest number of candidate anchors in $\lambda^-(s)+\lambda^+(s)+1$ consecutive samples (e.g., $\nu(s) \leq 2$ for the two MFDFA grids).

The proof is in Section S-I. The halo has $\lambda^- + \lambda^+$ samples per shard boundary, i.e., one maximal scale for MFDFA and a few for wavelet leaders (Table II). It is exchanged once, before any iteration. The global mean is one scalar average consensus on $(\sum_{n \in I_i} x[n], N_i)$. Because of Assumption 1, node profiles may differ from the global profile by a constant, so no prefix sums are needed. In (3), $R_q(s)$ grows exponentially with $|q|$ for intermittent signals: dropping even a few boundary terms can move high-order moments substantially, which Table II confirms.

D. Why Averaging Shard Estimates Is Biased

Let $\hat{m}_i(q,s) = M_i(q,s)/n_i(s)$ and let $\hat{\zeta}_i(q) = \sum_j a_j \log_2 \hat{m}_i(q,s_j)$ be the local log-log least-squares slope, with weights satisfying $\sum_j a_j = 0$ and $\sum_j a_j \log_2 s_j = 1$. The *naive* estimate is $\hat{\zeta}_{\mathrm{nv}}(q) = K^{-1}\sum_i \hat{\zeta}_i(q)$. The *exact* estimate $\hat{\zeta}_{\mathrm{ex}}(q)$ applies the same slope weights to the pooled $\log_2 \hat{m}(q,s_j)$.

Proposition 2 (Bias of shard averaging)**.** *Assume that the terms at scale s have mean $\mu(q,s)$ and that local mean moments satisfy $\mathrm{Var}[\hat{m}_i(q,s)]/\mu(q,s)^2 = \kappa_q(s)/n_i(s) + o(1/n_i(s))$, with third absolute central moments of $\hat{m}_i/\mu$ of order $o(1/n_i)$. Here $\kappa_q(s)$ is an effective normalized variance that absorbs the dependence between terms. Then*

$$\mathbb{E}\,\hat{\zeta}_{\mathrm{nv}}(q) - \zeta(q) = -\frac{1}{2\ln 2}\sum_j a_j \kappa_q(s_j)\,\frac{1}{K}\sum_i \frac{1}{n_i(s_j)} + o(\cdot), \quad (4)$$

$$\mathbb{E}\,\hat{\zeta}_{\mathrm{ex}}(q) - \zeta(q) = -\frac{1}{2\ln 2}\sum_j a_j \kappa_q(s_j)\,\frac{1}{n(s_j)} + o(\cdot). \quad (5)$$

For equal shards, $n_i = n/K$, the leading bias of $\hat{\zeta}_{\mathrm{nv}}$ is K times that of $\hat{\zeta}_{\mathrm{ex}}$. If the shard length is fixed and $N \to \infty$, the leading bias of $\hat{\zeta}_{\mathrm{ex}}$ vanishes but that of $\hat{\zeta}_{\mathrm{nv}}$ does not.

Proposition 2 is a second-order delta-method statement (proof in Section S-I). It concerns moments with finite variance, i.e., q below the critical order of the shard. Beyond it, local sample moments are dominated by single extremes and linearize in q [20], [21]. Since the critical order grows with sample size, shards linearize at smaller $|q|$ than the full record, which further shifts the averaged spectrum (Fig. 2). The two propositions together say: to reproduce the centralized analysis, decompose the *moments*, not the *estimates*.

IV. Separable Quasi-Likelihood Formulation

A. Quasi-Likelihood Criterion

Pooling the moments with a consensus and then fitting is one exact option (the CSS baseline in Section VII). It requires

every message to carry the whole $Q \times J$ moment array, and it cannot use a prior during the iterations. We instead look for a criterion that is a *sum of node functions of the parameters*. Let $\theta = \{(c_q, \zeta_q)\}_{q\in\mathcal{Q}} \in \mathbb{R}^{Q\times 2}$ collect intercepts and exponents, let $\ell_j = \log_2 s_j - \bar{\ell}$ be centered log-scales with $\bar{\ell}$ their mean, and let

$$\eta_{qj}(\theta) = \ln 2\,(c_q + \zeta_q \ell_j) = x_j^\top \theta_q, \quad x_j = \ln 2\,[1,\ \ell_j]^\top. \quad (6)$$

Fix positive normalizers c_j (e.g., the deterministic global counts $n(s_j)$, which every node can compute from N) and weights $w_j > 0$. Define $a_{iqj} = \ln\big(M_i(q, s_j)/c_j\big)$ and $\pi_{ij} = n_i(s_j)/c_j$. The node criterion is

$$f_i(\theta) = \sum_{q\in\mathcal{Q}} \sum_{j=1}^{J} w_j \Big[\psi_B\big(a_{iqj} - \eta_{qj}(\theta)\big) + \pi_{ij}\,\eta_{qj}(\theta)\Big], \quad (7)$$

where $\psi_B(t) = e^t$ for $t \le B$ and $\psi_B(t) = e^B(1 + (t-B) + \frac{1}{2}(t-B)^2)$ for $t > B$. With $B = \infty$, (7) is the Gamma/log-link quasi-likelihood of the pooled mean moments [23], [24]. The cap B only makes the gradient globally Lipschitz (Lemma 1); it is inactive at the solution in all our experiments. Let $F = \sum_i f_i$ and $r_{qj}(\theta) = \ln m(q, s_j) - \eta_{qj}(\theta)$ be the log-residuals of the pooled power-law fit.

Proposition 3 (Properties of the QL criterion)**.** *Let $B = \infty$, $J \ge 2$, and $M(q, s_j) > 0$ for all q, j.*

(a) *F depends on the data only through the pooled $M(q, s_j)$ and $n(s_j)$. For each (q, j), the minimizer of F over an unconstrained η_{qj} is $\eta_{qj} = \ln m(q, s_j)$.*

(b) *F is strictly convex and coercive, so $\hat{\theta} = \arg\min F$ exists and is unique.*

(c) *Let $\tilde{\omega}_j = w_j n(s_j)/c_j$, $X = [x_1, \dots, x_J]^\top$ and $\Omega = \mathrm{diag}(\tilde{\omega}_j)$, and let $\hat{\theta}_q^{\mathrm{LS}}$ be the weighted least-squares fit of $\ln m(q, s_j)$ on x_j with weights $\tilde{\omega}_j$. Then*

$$\hat{\theta}_q - \hat{\theta}_q^{\mathrm{LS}} = (X^\top \Omega X)^{-1} \sum_j \tilde{\omega}_j\, \varphi(r_{qj})\, x_j, \quad (8)$$

where $\varphi(r) = e^r - 1 - r \in [0, \frac{1}{2} r^2 e^{|r|}]$ and $r_{qj} = r_{qj}(\hat{\theta})$.

With $c_j = n(s_j)$ and $w_j = 1$, $\hat{\theta}^{\mathrm{LS}}$ is the classical unweighted log-log regression of the centralized estimator.

So $\hat{\theta}$ equals the centralized regression up to $O(\max_j r_{qj}^2)$. Log-residuals of a good scaling range are small. On MRW and fGn records, the two estimates have the same accuracy at every q. For the binomial cascade, whose log-moments oscillate around the power law, they differ at $|q| \ge 3.5$ (Section VII-C). This difference is the price of separability. A criterion that is a sum of node terms and depends on the data only through pooled moments must be affine in (M_i, n_i) (for continuous criteria), so its estimating equations are linear in the pooled mean moments $m(q, s_j)$, whereas log-log regression is linear in $\ln m(q, s_j)$. The point of (7) is that it acts on moments, not on log-moments: each node contributes its own exponentials, and the sum reproduces the pooled moment. The weights $\tilde{\omega}_j$ give the usual quasi-likelihood efficiency argument [23]: a mean moment estimated from $n(s_j)$ terms has variance proportional to $\mu^2/n(s_j)$, the variance function of the Gamma family.

B. Priors on the Scaling Function

The prior $R : \mathbb{R}^{Q\times 2} \to \mathbb{R} \cup \{+\infty\}$ is proper, closed and convex, and the PnP-MF-ADMM algorithm accesses it only through a proximal oracle. We use three modules.

Concavity (physical admissibility). $R = \iota_{\mathcal{C}}$ is the indicator of $\mathcal{C} = \{\theta : \zeta \text{ is concave on } \mathcal{Q} \cup \{0\},\ \zeta(0) = 0\}$. Every element of $\mathcal{C}$ can be written as $\zeta(q) = bq - \sum_k \beta_k[(q-\kappa_k)_+ - (-\kappa_k)_+]$ with $b \in \mathbb{R}$, $\beta \ge 0$ and knots κ_k at the interior points of $\mathcal{Q} \cup \{0\}$. The (weighted) Euclidean projection onto $\mathcal{C}$ is therefore a nonnegative least-squares problem in Q variables, solved exactly by an active-set method [51]; compare concave regression [52]. The projection guarantees a monotone $h(q)$ and a well-defined Legendre spectrum.

Smoothness. $R(\theta) = \frac{\lambda}{2}\|D_2\zeta\|^2$, with D_2 the second-difference operator on the nonuniform grid $\mathcal{Q}\cup\{0\}$ (anchored at $\zeta(0) = 0$). This is a linear denoiser and is smooth, so it falls under Theorem 2.

Denoisers. Any map $\mathcal{D} : \mathbb{R}^{Q\times 2} \to \mathbb{R}^{Q\times 2}$ can be placed in the prior slot of the linearized variant, as in PnP-ADMM [28], [29]. Convergence then follows from fixed-point arguments if $\mathcal{D}$ is the proximal operator of a convex function (e.g., an averaged linear smoother), and from [49] for suitably contractive learned denoisers. We state guarantees only for the proximal case.

C. Decentralized Consensus Form

The nodes communicate over a connected undirected graph $\mathcal{N} = (\mathcal{V}, \mathcal{E})$, which need not coincide with the time-adjacency of shards. Assign the prior evenly, $g_i = R/K$, and introduce local copies θ_i:

$$\min_{\theta_1,\dots,\theta_K} \sum_{i=1}^{K} \big[f_i(\theta_i) + \tfrac{1}{K} R(\theta_i)\big] \quad \text{s.t.} \quad \theta_i = \theta_j,\ (i,j) \in \mathcal{E}. \quad (9)$$

Since $\mathcal{N}$ is connected, (9) is equivalent to $\min_\theta F(\theta) + R(\theta)$. Its solution, the *centralized estimator* $\hat{\theta}$, is what a single machine holding the whole record would compute (Propositions 1–3).

V. PnP-MF-ADMM

A. Algorithm

We follow the edge-based decentralized ADMM [26], [27]. Each constraint $\theta_i = \theta_j$ is split through an auxiliary edge variable, the auxiliary variables are eliminated, and the edge multipliers are aggregated into one dual variable ϕ_i per node, initialized at $\phi_i^0 = 0$. With penalty $\rho > 0$ and

$$b_i^k = \rho\Big(d_i \theta_i^k + \sum_{j\in\mathcal{N}_i} \theta_j^k\Big) - \phi_i^k, \quad (10)$$

one iteration of PnP-MF-ADMM consists of a local primal step, a broadcast of θ_i^{k+1} to the neighbors, and the dual step

$$\phi_i^{k+1} = \phi_i^k + \rho \sum_{j\in\mathcal{N}_i} \big(\theta_i^{k+1} - \theta_j^{k+1}\big). \quad (11)$$

Two interchangeable local steps are available.

Exact (E) step: for a small $\tau_e > 0$,

$$\theta_i^{k+1} = \arg\min_\theta\ f_i(\theta) + \tfrac{1}{K} R(\theta) + \rho d_i \|\theta\|^2 - \langle b_i^k, \theta\rangle + \tfrac{\tau_e}{2}\|\theta - \theta_i^k\|^2. \quad (12)$$

Algorithm 1 PnP-MF-ADMM (Node i)

Require: shard $x[a_i{:}b_i]$, front-end Φ, orders $\mathcal{Q}$, scales $\{s_j\}$, prior oracle R, penalty ρ

Setup (once): average consensus for $\bar{x}$; receive halo of $\lambda^{\pm}$ samples from time-neighbors; compute $M_i(q, s_j)$, $n_i(s_j)$ by (2)

$\theta_i^0 \leftarrow$ local log-log fit of $\hat{m}_i$; $\phi_i^0 \leftarrow 0$; send θ_i^0 to neighbors

for $k = 0, 1, 2, \ldots$ **do**

end

$b_i^k \leftarrow \rho(d_i\theta_i^k + \sum_{j\in\mathcal{N}_i}\theta_j^k) - \phi_i^k$

$\theta_i^{k+1} \leftarrow$ E-step (12) or L-step (14)

broadcast θ_i^{k+1} ($2Q$ numbers) to $j \in \mathcal{N}_i$; receive θ_j^{k+1}

$\phi_i^{k+1} \leftarrow \phi_i^k + \rho\sum_{j\in\mathcal{N}_i}(\theta_i^{k+1} - \theta_j^{k+1})$

return ζ_i, and $\alpha_i(q), D_i(\alpha)$ by Legendre transform

Linearized (L) step: with $\tau_i > 0$,

$$u_i^k = \frac{\tau_i\theta_i^k + b_i^k - \nabla f_i(\theta_i^k)}{2\rho d_i + \tau_i}, \tag{13}$$

$$\theta_i^{k+1} = \operatorname{prox}_{R/(K(2\rho d_i+\tau_i))}\big(u_i^k\big). \tag{14}$$

Algorithm 1 summarizes the method. The E-step is a strongly convex problem in $2Q$ variables. Because f_i is separable over q, its Hessian is block diagonal with 2×2 blocks

$$H_{iq}(\theta) = \sum_j w_j\,\psi_B''\big(a_{iqj} - \eta_{qj}(\theta)\big)\,x_j x_j^\top. \tag{15}$$

We solve (12) by a damped (projected) Newton method warm-started at θ_i^k. For $R = \iota_{\mathcal{C}}$, the intercepts are eliminated in closed form and the ζ-update is a projection onto $\mathcal{C}$ weighted by the per-q Schur complements of (15), i.e., a weighted nonnegative least-squares problem. In our runs the E-step takes 2 to 4 Newton iterations on average. The L-step needs one gradient and one call of the prior oracle, which makes it the natural slot for arbitrary denoisers.

Penalty and step parameters. Because the moments are normalized by c_j, the Hessian of F at an exact power law is $H_0 = \sum_j w_j x_j x_j^\top$ for every q, independently of the data. We therefore use the scale-free choice $\rho = \bar{\rho}\,\lambda_{\max}(H_0)/(K\bar{d})$, with $\bar{d}$ the mean degree and $\bar{\rho} = 1$ by default, and $\tau_e = 10^{-3}\rho$. For the L-step, τ_i must exceed half the curvature of f_i (Theorem 1). In practice we set $\tau_i = 0.6\lambda_{\max}(H_i(\theta_i^k))$ and increase it when the local curvature grows, in the spirit of Remark 1 of [47]; Lemma 1 provides a global bound if one is needed.

B. Computation and Communication

Proposition 4 (Costs)**.** *Let J scales, Q orders, shard length N_i and maximal window $\lambda = \lambda^- + \lambda^+$ be given. (i)* Setup: *the front-end costs $O(N_i J m)$ flops for MFDFAm and MF-DXAm, $O(N_i J)$ for MFDMA (cumulative sums), $O(N_i J L)$ for MODWT and WL with L-tap filters, and $O(N_i J \log N_i)$ for CWT by FFT. The moments (2) cost $O(Q\sum_j n_i(s_j))$ more; for MFDFA this is $O(QN_i\sum_j s_j^{-1})$. Memory is $O(N_i + \lambda)$, and the halo is λ samples per shard boundary, exchanged once. (ii)* Per iteration: *the L-step costs $O(QJ)$ flops plus one prior call, and the E-step ν times that (ν Newton steps). The concavity projection is a nonnegative least-squares problem with Q variables, which costs at most $O(Q^3)$. Each node sends $2Q$ numbers to each neighbor, i.e., $4Q|\mathcal{E}|$ numbers per iteration network-wide. (iii)* Comparison: *centralizing the raw data moves N samples (times the hop distance). Consensus on sufficient statistics (CSS) moves $(Q+1)J$ numbers per edge direction per iteration, i.e., $J(Q+1)/(2Q) \approx J/2$ times more per iteration than PnP-MF-ADMM.*

The message volume of PnP-MF-ADMM is $4Q|\mathcal{E}|$ per iteration, independent of N, whereas centralization moves $N\bar{h}$ numbers, with $\bar{h}$ the mean hop distance to a fusion node; Table V gives both for $N = 2^{25}$. Each node stores and processes only N/K samples plus its halo.

VI. Convergence Analysis

Write $f(\boldsymbol{\theta}) = \sum_i f_i(\theta_i)$, $g(\boldsymbol{\theta}) = \frac{1}{K}\sum_i R(\theta_i)$, $\mathbf{L}_\pm = L_\pm \otimes I_{2Q}$, and $\mathbf{M} = \mathbf{L}_-^{1/2}$. Let $P = e^B \operatorname{diag}_q(H_0)$, where diag_q denotes the block-diagonal matrix with Q copies of H_0.

Lemma 1 (Smoothness and inactive cap)**.** *Each f_i is convex and C^1 with $0 \preceq \nabla^2 f_i \preceq P$ almost everywhere, so ∇f_i is P-cocoercive: $\langle \nabla f_i(\theta) - \nabla f_i(\theta'), \theta - \theta'\rangle \geq \|\nabla f_i(\theta) - \nabla f_i(\theta')\|^2_{P^{-1}}$. If $J \geq 2$ and every node has a positive term at two or more scales for each q, then f_i is strictly convex, and it is strongly convex on every bounded set. Let $\hat{\theta}^\infty$ minimize $F + R$ with $B = \infty$. If $B > B^\star := \max_{q,j}\big[\ln(n(s_j)/c_j) + r_{qj}(\hat{\theta}^\infty)\big]$, then $\hat{\theta}^\infty$ also uniquely minimizes the capped problem.*

With $c_j = n(s_j)$, $B^\star$ is the largest log-residual of the fit, so the cap $B = 5$ can only be active at the solution if the power law misfits a pooled moment by a factor $e^5 \approx 148$.

Lemma 2 (Optimality)**.** *$\hat{\theta}$ minimizes $F + R$ if and only if $\boldsymbol{\theta}^\star = \mathbf{1} \otimes \hat{\theta}$ and some $\boldsymbol{\beta}^\star \in \operatorname{range}(\mathbf{M})$ satisfy $0 \in \nabla f(\boldsymbol{\theta}^\star) + \partial g(\boldsymbol{\theta}^\star) + \mathbf{M}\boldsymbol{\beta}^\star$ and $\mathbf{M}\boldsymbol{\theta}^\star = 0$. Such a $\boldsymbol{\beta}^\star$ exists and is unique in $\operatorname{range}(\mathbf{M})$.*

Writing $\boldsymbol{\phi}^k = \mathbf{M}\boldsymbol{\beta}^k$, Algorithm 1 is the primal–dual recursion

$$0 \in \nabla f(\boldsymbol{\xi}^k) + \partial g(\boldsymbol{\theta}^{k+1}) + \mathbf{M}\boldsymbol{\beta}^{k+1} + \mathbf{G}(\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k), \tag{16}$$

$$\boldsymbol{\beta}^{k+1} = \boldsymbol{\beta}^k + \rho\mathbf{M}\boldsymbol{\theta}^{k+1}, \tag{17}$$

with $\boldsymbol{\xi}^k = \boldsymbol{\theta}^{k+1}$ and $\mathbf{G} = \rho\mathbf{L}_+ + \tau_e I$ for the E-step, and $\boldsymbol{\xi}^k = \boldsymbol{\theta}^k$ and $\mathbf{G} = \rho\mathbf{L}_+ + \operatorname{diag}(\tau_i) \otimes I$ for the L-step (Section S-I). Define the Lyapunov function

$$V^k = \|\boldsymbol{\theta}^k - \boldsymbol{\theta}^\star\|^2_{\mathbf{G}} + \tfrac{1}{\rho}\|\boldsymbol{\beta}^k - \boldsymbol{\beta}^\star\|^2. \tag{18}$$

Theorem 1 (Global convergence)**.** *Let $\mathcal{N}$ be connected and let R be proper, closed and convex, with $\phi^0 = 0$. For the E-step with any $\tau_e > 0$, and for the L-step with $\tau_i > \lambda_{\max}(P)/2$, the iterates satisfy*

$$V^{k+1} \leq V^k - \|\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k\|^2_{\mathbf{G}-\gamma P} - \tfrac{1}{\rho}\|\boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^k\|^2, \tag{19}$$

with $\gamma = 0$ (E) or $\gamma = \frac{1}{2}$ (L, P acting blockwise). Consequently $\theta_i^k \to \hat{\theta}$ for every i, and ϕ^k converges. Moreover, the KKT residual $\mathcal{R}^{k+1} = \operatorname{dist}(0, \nabla f(\boldsymbol{\theta}^{k+1}) + \partial g(\boldsymbol{\theta}^{k+1}) + \mathbf{M}\boldsymbol{\beta}^{k+1}) + \|\mathbf{M}\boldsymbol{\theta}^{k+1}\|$ satisfies $\min_{t\leq k}(\mathcal{R}^{t+1})^2 = o(1/k)$.

Theorem 2 (Linear convergence for smooth priors)**.** *In addition, let R be convex with an L_R-Lipschitz gradient (e.g.,*

$R \equiv 0$ *or the smoothness prior), and fold* R/K *into* f_i*. Let* $S = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^\star\|_{\mathbf{G}}^2 \le V^0\}$*,* $\mu > 0$ *the strong-convexity modulus of the* f_i *on* S*,* ℓ *the Lipschitz constant of* ∇f *on* S*, and* λ_2 *the algebraic connectivity. For the E-step,*

$$V^{k+1} \le \frac{1}{1+\delta} V^k, \quad \textit{with} \\ \frac{1}{\delta} = \max\Big\{ \frac{\lambda_{\max}(\mathbf{G}) + 2\ell^2/(\rho\lambda_2)}{2\mu}, \; \frac{2\lambda_{\max}(\mathbf{G})}{\rho\lambda_2} \Big\}. \tag{20}$$

For the L-step with $\mathbf{G} - P \succeq \gamma_0 I \succ 0$*, the same holds with* $1/\delta = \max\{(2\lambda_{\max}(\mathbf{G}) + 2\ell^2/(\rho\lambda_2))/\mu,\ (2\lambda_{\max}(\mathbf{G}) + 2\lambda_{\max}(\mathbf{G})^2/(\rho\lambda_2))/\gamma_0\}$.

Corollary 1 (Error decomposition and iteration complexity)**.** *Under Theorem 2, for every node and every* k*,*

$$\|\theta_i^k - \theta^{\text{true}}\| \le \underbrace{\|\hat{\theta} - \theta^{\text{true}}\|}_{\textit{statistical}} + \underbrace{\sqrt{V^0/\lambda_{\min}(\mathbf{G})}\,(1+\delta)^{-k/2}}_{\textit{optimization}}. \tag{21}$$

An optimization error ε *is reached after* $k \ge \ln\big(V^0/(\lambda_{\min}(\mathbf{G})\varepsilon^2)\big)/\ln(1+\delta) = O(\delta^{-1}\ln\varepsilon^{-1})$ *iterations. For the E-step with* $\lambda_{\max}(\mathbf{G}) \approx 2\rho d_{\max}$*, choosing* $\rho \propto \ell/\sqrt{d_{\max}\lambda_2}$ *gives*

$$\delta^{-1} = O\Big(\frac{\ell}{\mu}\sqrt{\frac{d_{\max}}{\lambda_2}} + \frac{d_{\max}}{\lambda_2}\Big). \tag{22}$$

Remark 1 (Interpretation). The first term of δ^{-1} is the conditioning of the local criteria, which is governed by how unevenly the moments are shared among nodes: at large $|q|$ a few shards may carry most of $M(q,s)$. The second term is the graph's condition number $d_{\max}/\lambda_2$. On a ring, $d_{\max}/\lambda_2 \approx K^2/(2\pi^2)$; on Erdős–Rényi graphs with $p \propto \ln K/K$ it grows only slowly with K. Section VII-E confirms both dependences.

Remark 2 (Nonsmooth priors). For $R = \iota_{\mathcal{C}}$ (polyhedral), Theorem 1 guarantees convergence and an $o(1/k)$ best-iterate residual. Linear rates for decentralized proximal methods with a common nonsmooth term require a different splitting [45]. In all our experiments, PnP-MF-ADMM with $\iota_{\mathcal{C}}$ nevertheless converges linearly (Figs. 3 and 4). A plausible explanation is that the active set of the polyhedral $\mathcal{C}$ is identified after finitely many iterations, after which the problem is locally smooth.

Remark 3 (What is exchanged). Nodes exchange estimates of θ, never samples (apart from the one-time halo) and never their moment arrays. Asynchronous and quantized variants follow standard extensions of decentralized ADMM [41] and are left for future work.

VII. Numerical Experiments

A. Setup

Signals. We use three models with known scaling functions (Section S-III): a conservative random binomial cascade with multiplier 0.7, for which $\zeta(q) = 1 - \log_2(0.7^q + 0.3^q)$; a multifractal random walk (MRW) [53] with $\lambda^2 = 0.04$, for which $\zeta(q) = 0.54q - 0.02q^2$; and fractional Gaussian noise (fGn) [54] with $H = 0.7$, which is monofractal, $\zeta(q) = 0.7q$ and $\Delta\alpha = 0$. *Estimation.* We use $\mathcal{Q} = \{\pm 0.5, \pm 1, \dots, \pm 5\}$ ($Q = 20$), 14 log-spaced scales in $[16, 4096]$ for the DFA/DMA families, $j = 4, \dots, 12$ for MODWT-SF and WL (Daubechies, three vanishing moments), and 12 scales in $[8, 1024]$ for CWT-max. We set $w_j = 1$, c_j to the deterministic expected counts, $B = 5$, $\bar{\rho} = 1$ and $\tau_e = 10^{-3}\rho$. The default local step is E.

Table II: Exactness ($N = 2^{20}$ MRW). Halo Length; Largest Discrepancy of Pooled vs. Centralized $\log M$ With Halo; Lost Terms (%) and $\max_q |\Delta\hat{\zeta}|$ Without Halo; PnP-MF-ADMM Iterations to 10^{-10} on an RGG

		halo	lost %		$\lvert\Delta\hat{\zeta}\rvert$, no halo		ADMM it.	
Front-end	λ	$\lvert\Delta \log M\rvert$	16	128	16	128	16	128
MFDFA1	4096	1.9×10^{-13}	0.09	0.77	0.003	0.003	260	2430
MFDFA2	4096	1.7×10^{-13}	0.09	0.77	0.003	0.008	260	2270
MFDMA0	8190	1.2×10^{-9}	0.20	1.70	0.011	0.039	260	2240
MFDMA0.5	10237	1.1×10^{-8}	0.20	1.70	0.011	0.031	270	2100
MODWT-SF	20475	7.1×10^{-15}	0.52	3.98	0.025	—	310	1890
WL	32763	4.8×10^{-13}	0.83	6.01	0.053	—	290	>3000
CWT-max	12290	2.1×10^{-12}	0.58	4.90	0.021	—	450	1810
MFDXA1	4096	1.1×10^{-9}	0.09	0.77	0.004	0.166	460	2900

Networks. Ring, 2-D torus, Erdős–Rényi (ER, $p = 2\ln K/K$), random geometric (RGG, radius $1.2\sqrt{2\ln K/(\pi K)}$), star and complete graphs, all connected. *Metric.* $\max_i \|\theta_i^k - \hat{\theta}\|/\|\hat{\theta}\|$, where $\hat{\theta}$ is the centralized solution computed by a Newton or projected Newton method to machine precision. *Baselines.* CSS (Metropolis average consensus on the moment arrays, then a local fit) [46]; shard averaging ("naive"); projected DGD [44]; and PG-EXTRA [43]. For DGD and PG-EXTRA we report the best of four step sizes, because the exponential criterion has no small global Lipschitz constant. The network is simulated in one process on a two-core virtual machine, and per-node times are measured separately. All numbers come from the stored results of the companion notebook.

B. Exactness of the Decomposition

Table II applies the eight front-ends to one MRW record of $N = 2^{20}$ samples split over $K = 16$ and $K = 128$ shards. With the halo, the pooled partition functions match the centralized ones to round-off (10^{-15} to 10^{-8} in $\log M$, Section S-II), and the counts match exactly. Without the halo, 0.1–6% of the terms are lost. The resulting change of $\hat{\zeta}$ ranges from 0.003 to 0.17, i.e., up to the order of the statistical error. For the wavelet front-ends at $K = 128$, whole scales lose all their terms because the windows exceed the shard length. PnP-MF-ADMM on an RGG then reproduces the centralized estimate to 10^{-10} for every front-end, except WL at $K = 128$, which reaches 3×10^{-10} in the 3000-iteration budget.

C. Shard Averaging Changes the Estimate

Fig. 2 uses $N = 2^{22}$, $K \in \{1, 4, \dots, 1024\}$ shards (down to 4096 samples each), MFDFA1 with scales 16–1024, and 16 realizations per model. By Proposition 1, the exact distributed estimate does not depend on K. The shard average does, in a model-dependent way (Fig. 2a). For fGn it creates spurious multifractality: $\Delta\alpha$ grows from 0.014 at $K = 1$ to 0.102 at $K = 1024$. The additional bias of the shard average, relative to $K = 1$, grows roughly in proportion to K, as predicted by Proposition 2: at $q = -5$ it is -0.0023 for $K = 16$ and -0.093 for $K = 1024$, a factor of 40 for a 64-fold increase of K (Fig. 2b). For the cascade, the spectrum narrows by 0.09, consistent with an earlier linearization in short shards; for the MRW it widens by 0.03. The QL estimate and the pooled log-log regression coincide for fGn and MRW (their RMSEs differ by at most 0.01 at every q). For the cascade they differ at $|q| \ge 3.5$, where the log-residuals are large (Proposition 3c).

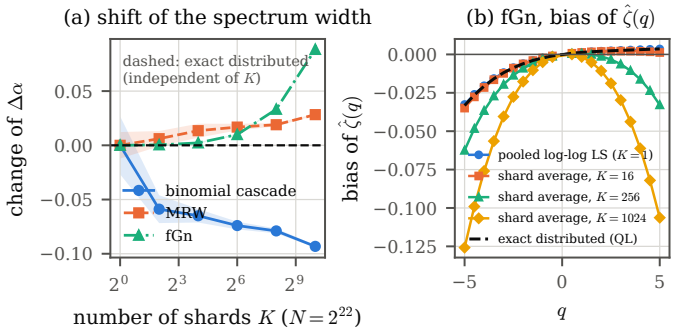


Figure 2: Shard averaging versus exact distributed estimation ($N = 2^{22}$, MFDFA1, 16 realizations). (a) Change of the spectrum width of the shard average relative to $K = 1$ (shaded: ± 1 standard error); the exact distributed estimate is independent of K. (b) Bias of $\hat{\zeta}(q)$ for monofractal fGn; any curvature of $\hat{\zeta}$ in q is spurious multifractality.

Two caveats complete the picture. First, the shard average is *not* uniformly worse in mean squared error. At the extreme orders of intermittent signals, the pooled moments are dominated by a few events, and a mean of logarithms has a much smaller variance. For example, at $q = -5$ for the cascade the RMSE is 0.233 for the pooled estimate and 0.061 for the shard average with $K = 16$ (Table S-II). Second, this variance reduction comes with an estimand that drifts with the storage layout, so results computed on differently sharded copies of the same record disagree, and for fGn at $|q| = 5$ the shard average has between 1.9 and 16 times the RMSE of the pooled estimate once $K \geq 256$ (Table S-II). The goal of this paper is to reproduce the centralized analysis; robust centralized estimators, e.g., log-cumulants (Section VIII), are a separate question.

D. Convergence and Communication

Fig. 3 compares the methods with $K = 64$, $N = 2^{22}$, MFDFA1 and the concavity prior. On the RGG with MRW data, PnP-MF-ADMM (E) reaches 10^{-8} in 530 iterations, with 1.2×10^7 numbers sent network-wide. CSS needs 563 iterations but 9.3×10^7 numbers, i.e., 7.8 times more, because each message carries the $(Q+1)J = 294$-entry moment array instead of $2Q = 40$ parameters. The linearized variant (L) needs 1545 iterations and remains cheaper per iteration (one gradient and one prox). PG-EXTRA and DGD stall above 10^{-3} even with tuned steps. The curvature of the exponential criterion varies by orders of magnitude between nodes and iterations, and a single step size cannot accommodate it. On the ring ($\lambda_2 = 0.0096$) every method slows down, as predicted by the $d_{\max}/\lambda_2$ term of Corollary 1. For the cascade, the moments at large $|q|$ are concentrated on a few shards, i.e., the local criteria are badly conditioned (Remark 1). The E-step handles this through its exact local solves and still reaches 10^{-6} with 2.5 times fewer numbers than CSS, whereas the L-step stalls near 10^{-2}. The E-step thus is the robust default, and the L-step is the economical choice when moments are evenly shared. Table S-I lists all final errors.

E. Network Size, Topology and Penalty

Fig. 4 reports the iterations to 10^{-6} for $K = 16$ to 256 on one MRW record of 2^{22} samples. Both terms of Corollary 1 are visible. On well-connected graphs ($\bar{d}/\lambda_2 \leq 6$: complete, ER, star), the count still grows with K, from about 143 at $K = 16$ to 553–1110 at $K = 256$. This is the conditioning term: shorter shards share the moments less evenly, so the local criteria become worse conditioned. On poorly connected graphs the graph term dominates. On the ring the count grows from 286 ($K = 16$) to 1784 ($K = 64$), and for $K \geq 128$ it exceeds the budget of 6000. In Fig. 4b, the points with large $\bar{d}/\lambda_2$ (ring, RGG) grow with a slope between $1/2$ and 1 in log-log coordinates, i.e., between the $\sqrt{d_{\max}/\lambda_2}$ and $d_{\max}/\lambda_2$ terms. The scale-free default $\bar{\rho} = 1$ is the best of the five tested values on the RGG and ER graphs, and within a factor of 1.1 of the best on the ring (Fig. 4c). The E-step used 2.4–3.4 Newton iterations on average.

Table III: Plug-and-Play Priors on Short Records (200 Realizations, $K = 4$, $q \in [-8, 8]$): Fraction of Runs Where $\alpha(q)$ Increases by More Than 0.01 Without a Prior, and RMSE of $\hat{\zeta}$ and of $\Delta\alpha$ With No Prior (N), Concavity (C) and Smoothness Plus Concavity (S+C)

			RMSE $\hat{\zeta}$			RMSE $\Delta\alpha$		
Model	N	%	N	C	S+C	N	C	S+C
cascade	2^{10}	46	1.189	1.177	1.182	0.49	0.51	0.51
cascade	2^{12}	26	0.578	0.574	0.576	0.22	0.23	0.24
cascade	2^{14}	13	0.352	0.352	0.352	0.17	0.18	0.18
MRW	2^{10}	42	0.654	0.620	0.620	0.28	0.33	0.33
MRW	2^{12}	18	0.382	0.379	0.379	0.24	0.25	0.26
MRW	2^{14}	6	0.236	0.235	0.236	0.21	0.21	0.21

F. Plug-and-Play Priors

The global estimate is rarely inadmissible when N is large, but short records, or analyses restricted to a few shards, often produce scaling functions that are not concave. The estimated $h(q)$ is then not monotone, and the Legendre transform folds back (Fig. 5). Table III reports 200 realizations for each of $N = 2^{10}, 2^{12}, 2^{14}$ ($K = 4$ shards, MFDFA1, $q \in [-8, 8]$). Without a prior, $\alpha(q)$ increases somewhere by more than 0.01 in 6–46% of the runs. The concavity module $\iota_{\mathcal{C}}$ removes these violations by construction and lowers the RMSE of $\hat{\zeta}$ by 0.2–5.3%. It is not a free lunch for the width: removing the folds also shortens the range of α. At these short lengths $\Delta\alpha$ is already underestimated, so its RMSE increases by up to 16% (at $N = 2^{10}$) and by at most 2% at $N = 2^{14}$. Adding the smoothness module ($\lambda = 0.05$) changes little at these settings. In the distributed algorithm, each module is a single call in the local step, so every node holds an admissible spectrum at every iteration.

G. Plug-and-Play Front-Ends

Because the decomposition is exact for every s-local front-end, switching estimators changes only the module that produces $M_i(q, s)$. Table IV compares seven univariate front-ends inside the same pipeline (QL criterion, concavity prior, E-step). It uses $N = 2^{20}$, $K = 8$ and 10 realizations per model; by Table II, each distributed estimate equals its centralized counterpart. MFDFA1, MFDFA2, WL and MODWT-SF (positive q only) have similar accuracy on the three models. Both MFDMA variants are less accurate on the cascade, the backward one markedly so. We used the same scaling range (16–4096) for all front-ends and did not tune it per

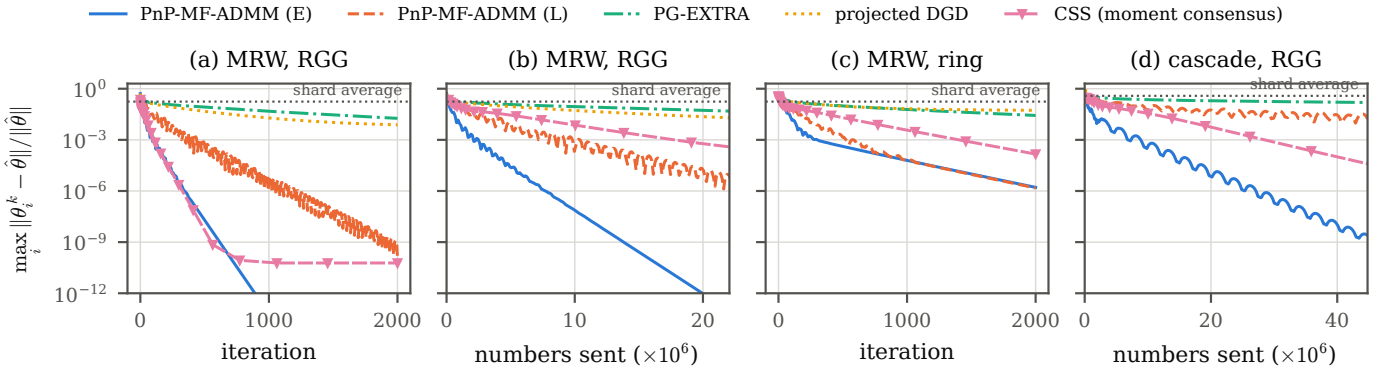


Figure 3: Convergence to the centralized estimator ($K = 64$, $N = 2^{22}$, MFDFA1, concavity prior). (a), (c) Versus iterations. (b), (d) Versus the total number of scalars sent over the network. The dotted line is the error of the shard average.

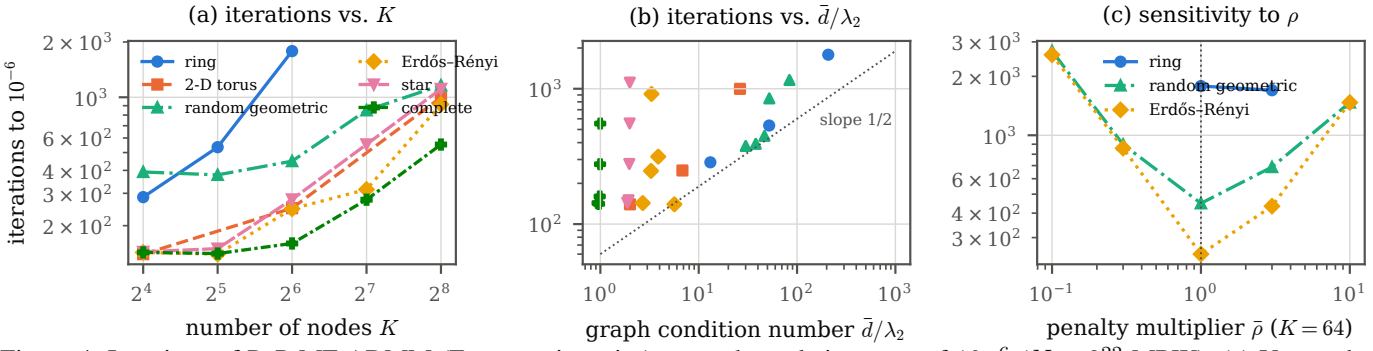


Figure 4: Iterations of PnP-MF-ADMM (E, concavity prior) to reach a relative error of 10^{-6} ($N = 2^{22}$ MRW). (a) Versus the number of nodes. (b) Versus the graph condition number $\bar{d}/\lambda_2$. (c) Versus the penalty multiplier $\bar{\rho}$ for $K = 64$; missing points did not converge within 3000 iterations.

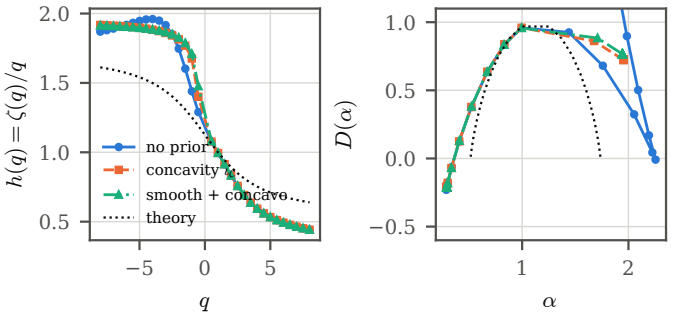


Figure 5: Effect of the concavity prior on a short cascade record ($N = 2^{10}$). Without a prior, $h(q)$ is not monotone and the Legendre spectrum folds back; with $\iota_{\mathcal{C}}$ (alone or after smoothing) it is admissible.

Table IV: Plug-and-Play Front-Ends in the Distributed Pipeline ($N = 2^{20}$, $K = 8$, 10 Realizations): RMSE of $\hat{\zeta}(q)$ Over $|q| \le 4$, and Front-End Time per Node

Front-end	MRW	cascade	fGn	ms/node
MFDFA1	0.024	0.052	0.012	40
MFDFA2	0.021	0.050	0.014	39
MFDMA0	0.042	0.834	0.031	41
MFDMA0.5	0.027	0.163	0.012	40
MODWT-SF	0.025	0.037	0.016	1069
WL	0.024	0.057	0.014	1174
CWT-max	0.204	0.089	0.193	110

MODWT-SF: $q > 0$ only (structure functions of coefficients).

estimator; [5] reports a different ranking of window positions for binomial measures with other settings. Our light CWT-max substitute for WTMM is the least accurate for MRW and fGn because of its negative-q branch; it is included to show that maxima-based partition functions also fit the framework. The per-node cost of DFA-type front-ends is tens of milliseconds for 1.3×10^5 samples, and the wavelet pyramids cost about one second in our pure-NumPy implementation.

H. Large-Scale Run

Table V reports one MRW record of $N = 2^{25} \approx 3.4 \times 10^7$ samples on $K = 256$ nodes of an RGG (mean hop distance $\bar{h} = 5.2$), with 14 scales up to 8192 (MFDFA1) and levels 4–13 (WL). Each node processes 2^{17} samples plus its halo. The front-end takes 0.04 s on average per node for MFDFA1 and 2.9 s for WL; the same code on one machine takes K times longer (the serial total). PnP-MF-ADMM reaches a relative error of 10^{-4} in 750–1262 iterations and 10^{-8} in 3116–4331 iterations. The message volume does not depend on N. Up to 10^{-4}, it amounts to 59–100% of what centralizing the raw record over the same graph would move ($N\bar{h}$ numbers). Up to 10^{-8}, on this poorly connected graph, it is 2.5–3.4 times that amount; with the same iteration count it would fall below it for records longer than 1.1×10^8 samples, and fewer iterations are needed on better-connected graphs (Fig. 4). The one-time halos add 6% (MFDFA1) and 50% (WL) of N; the wavelet halo is larger because its windows grow with the filter length and the largest scale. The estimated $\hat{\zeta}$ is within 0.003 of the theoretical MRW scaling function for $|q| \le 2$.

Table V: Large-Scale Run ($N = 2^{25}$ MRW, $K = 256$, RGG): Front-End Time per Node (Mean, Max) and Serial Total; ADMM Iterations to 10^{-4} and 10^{-8} and Time; Scalars Sent by ADMM to 10^{-4} and 10^{-8}, and by Centralization ($N\bar{h}$)

	front-end (s)			ADMM			scalars sent		
	mean	max	serial	10^{-4}	10^{-8}	s	10^{-4}	10^{-8}	central
MFDFA1	0.04	0.10	10	1262	4331	160	1.7×10^8	6.0×10^8	1.7×10^8
WL	2.88	3.55	737	750	3116	108	1.0×10^8	4.3×10^8	1.7×10^8

Table VI: Intraday Data: $\hat{h}(2)$, Spectrum Width (Exact Distributed, Monthly Shard Average, Within-Shard Shuffled), Distance of PnP-MF-ADMM to the Single-Machine $\hat{\zeta}$, and Iterations to 10^{-10}

			$\Delta\alpha$				
Data	Front-end	$\hat{h}(2)$	exact	avg.	shuf.	$\lvert\Delta\hat{\zeta}\rvert$	it.
EUR/USD	MFDFA1	0.486	0.964	0.427	0.504	2.0×10^{-10}	2640
EUR/USD	WL	0.480	2.327	0.364	0.390	6.3×10^{-7}	>3000
S&P 500	MFDFA1	0.494	0.203	0.294	0.318	3.5×10^{-11}	630
S&P 500	WL	0.496	0.229	0.289	0.271	2.0×10^{-11}	605

I. *Intraday Exchange Rate and Stock Index*

We analyze two public intraday archives stored as one file per month.[1] *EUR/USD*: OANDA bars from January 2005 to May 2020, giving $N = 5{,}583{,}560$ bar-to-bar log-returns in $K = 185$ monthly shards (13,037–33,467 returns each). *S&P 500*: one-minute index bars from November 2010 to December 2018, giving $N = 2{,}117{,}666$ returns in $K = 98$ monthly shards. Each shard is one node of an RGG. Returns are divided by an intraday profile of mean absolute returns in 15-minute bins [55]. The profile is a ratio of sums over nodes, i.e., one average consensus on a 96-entry vector. The first return of a shard uses the last price of the previous shard, which is a one-sample halo. EUR/USD has 11% zero returns (stale quotes), so some segments are exactly polynomial; the resolution floor of Section S-III discards them identically on every node. Table VI and Fig. 6 summarize the results.

The pooled partition functions agree with a single-machine computation to 2×10^{-8} or better. PnP-MF-ADMM reaches the centralized $\hat{\zeta}$ to 10^{-10} in 605–2640 iterations, except for WL on EUR/USD, which stops at 6×10^{-7} after 3000 iterations because the moments are very unevenly shared among months. The exact and the monthly-averaged analyses disagree substantially. For EUR/USD, the MFDFA1 width is 0.96 for the record as a whole and 0.43 for the average of monthly analyses. The difference comes from $q < 0$, where the global moments are dominated by the calmest months, which a monthly average dilutes. For the S&P 500, the shard average overstates the width (0.29 versus 0.20), and MFDFA1 and WL agree on the exact width (0.20 and 0.23). Shuffling the returns within each shard, which each node can do on its own, gives the surrogate widths in Table VI. For EUR/USD the surrogate width is about half the original, so temporal correlations contribute to the multifractality. For the S&P 500 it is not smaller than the original, so its multifractality at these scales cannot be attributed to correlations. For EUR/USD the WL estimate at $q < 0$ jumps to $h(q) \approx 2$. Wavelet leaders nearly vanish over stale-quote stretches, which illustrates that the choice of front-end matters on quantized data. The distributed computation reproduces this centralized behavior exactly.

VIII. Conclusion

We have shown that multifractal analysis of a sharded record can be carried out by a network of nodes without moving the data and without changing the answer. The key observation is that the partition functions of the common estimators are sums of terms with bounded windows. A one-time halo exchange therefore makes their decomposition across nodes exact, whereas averaging shard estimates changes the estimand in a way that depends on the storage layout. A separable quasi-likelihood criterion turns scaling regression into a consensus problem whose messages have $2Q$ entries. PnP-MF-ADMM solves it with interchangeable front-ends, priors and local solvers, and it provably converges, linearly for smooth priors, at a rate governed by the moment shares and the graph condition number.

Three limitations point to future work. First, the QL estimate differs from log-log regression by a term quadratic in the log-residuals. This is negligible for random walks and noises but visible at extreme orders of strongly oscillating cascades. When the log-domain fit is required, consensus on the $Q \times J$ moment arrays (CSS), which our decomposition also makes exact, is the alternative. Second, at extreme orders of intermittent signals the variance of moment-based estimators, centralized or distributed, is large. Estimators built on cumulants of log-leaders [8], [14] use pooled sums of powers of $\ln T_v(s)$ and therefore satisfy Proposition 1 as well; a distributed log-cumulant analysis is a natural next step. Third, asynchronous, quantized and privacy-preserving communication [18], [41], bootstrap confidence intervals [8] and images or other d-dimensional fields, where halos become boundary layers, follow the same template and are left for future work.

[1] https://github.com/FutureSharks/financial-data; we use the folders `currencies/oanda/EUR_USD` and `stocks/histdata/SPXUSD`.

Appendix: Proofs

A. *Proof of Proposition prop:exact*

Exactness. Fix s and a candidate anchor $v \in \mathcal{G}_N(s)$ with $v \in I_i$. Its window $W_v = [v - \lambda^-(s), v + \lambda^+(s)] \cap [0, N)$ is contained in the extended interval $E_i = [a_i - \lambda^-, b_i + \lambda^+) \cap [0, N)$ held by node i, because $\lambda^{\pm}(s) \le \lambda^{\pm}$. Node i forms its profile as $Y_i[n] = \sum_{u=\min E_i}^{n} (x[u] - \bar{x})$ for $n \in E_i$, which equals the global profile minus the constant $Y[\min E_i - 1]$ (zero if $\min E_i = 0$). By Assumption ass:local, $\chi_v(s)$ and $T_v(s)$ computed from Y_i coincide with their centralized values. Every candidate anchor belongs to exactly one I_i, because the shards partition $\{0, \dots, N-1\}$. Summing over i gives $\sum_i M_i = M$ and $\sum_i n_i = n$.

No halo. Without the halo, node i can evaluate exactly the anchors $v \in I_i$ with $W_v \subset I_i$. A lost anchor has a window that contains a shard boundary b_i, $i < K$. For each boundary, such anchors lie in an interval of length $\lambda^-(s) + \lambda^+(s) + 1$, which contains at most $\nu(s)$ candidate anchors by definition (for a grid with spacing $\delta(s)$, at most $\lceil (\lambda^- + \lambda^+)/\delta \rceil + 1$ per grid). Hence $n_L \le (K-1)\nu(s)$ and $\epsilon(s) = n_L/n(s)$. Let M_L and n_L be the lost partition sum and count, so that $\tilde{m} = (M - M_L)/(n - n_L)$. Then $\tilde{m} - m = (m n_L - M_L)/(n - n_L)$. Since $0 \le M_L \le n_L R_q(s) m$, we get $|\tilde{m} - m| \le m\, n_L (1 + R_q)/(n - n_L)$. Dividing by m gives (eq:nohalo). ∎

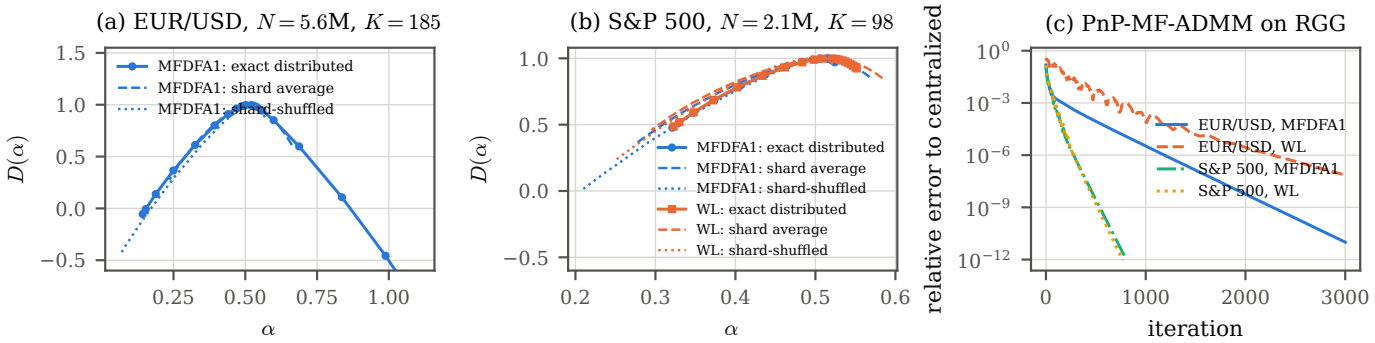


Figure 6: Intraday data. (a), (b) Multifractal spectra from the exact distributed estimate, the average of the monthly analyses, and within-shard shuffled surrogates. (c) Convergence of PnP-MF-ADMM (E-step, concavity prior) to the single-machine estimate.

B. Proof of Proposition prop:naive

Write $U_i = \hat{m}_i(q,s)/\mu(q,s) - 1$, so $\mathbb{E}U_i = 0$ and $\mathbb{E}U_i^2 = \kappa_q(s)/n_i + o(1/n_i)$. Assume in addition the uniform integrability condition $\mathbb{E}[|\ln(1+U_i)|\mathbb{1}\{|U_i| > 1/2\}] = o(1/n_i)$, which is standard for the delta method. On $\{|U_i| \le 1/2\}$, $\ln(1+U) = U - U^2/2 + \varrho(U)$ with $|\varrho(U)| \le \frac{8}{3}|U|^3$. Hence $\mathbb{E}\ln\hat{m}_i = \ln\mu - \kappa_q(s)/(2n_i) + o(1/n_i)$. The local slope is linear in $\log_2 \hat{m}_i$, so

$$\mathbb{E}\hat{\zeta}_i - \zeta = \frac{1}{\ln 2}\sum_j a_j\Big(-\frac{\kappa_q(s_j)}{2n_i(s_j)}\Big) + o(\cdot), \tag{23}$$

where the true log-moments drop out because $\sum_j a_j = 0$ and $\sum_j a_j \log_2 s_j = 1$. Averaging over i gives (eq:bias-nv). The pooled $\hat{m} = \sum_i (n_i/n)\hat{m}_i$ is a mean over n terms, so the same expansion holds with n_i replaced by n and gives (eq:bias-ex). With $n_i = n/K$, $K^{-1}\sum_i 1/n_i = K/n$. With a fixed shard length, $n_i(s)$ is fixed while $n(s) = Kn_i(s) \to \infty$. ∎

Remark. For fixed s the two estimators have the same first-order variance when the shards are weakly dependent (an average of K slopes, each with variance proportional to K/n, versus one slope with variance proportional to $1/n$). The mean squared error gap is therefore dominated by the bias, which is what Fig. fig:naive shows.

C. Proof of Proposition prop:ql

(a) With $B = \infty$, $\sum_i f_i(\theta) = \sum_{q,j} w_j[(M(q,s_j)/c_j)e^{-\eta_{qj}} + (n(s_j)/c_j)\eta_{qj}]$. For fixed (q,j), the scalar function $t \mapsto Ae^{-t} + Bt$ with $A, B > 0$ is minimized at $t = \ln(A/B) = \ln m(q,s_j)$.

(b) For each q, $\eta_q \mapsto \sum_j w_j[A_j e^{-\eta_j} + B_j\eta_j]$ is strictly convex and coercive on $\mathbb{R}^J$. The map $\theta_q \mapsto X\theta_q$ is injective because $J \ge 2$ distinct scales make X full column rank. Composition preserves strict convexity and coercivity, and F is separable over q.

(c) Setting the gradient to zero gives $\sum_j w_j(n_j/c_j)\big(1 - m_{qj}e^{-\eta_{qj}}\big)x_j = 0$, i.e., $\sum_j \tilde{\omega}_j(e^{r_{qj}} - 1)x_j = 0$. The weighted least-squares fit satisfies $\sum_j \tilde{\omega}_j(y_{qj} - x_j^\top \hat{\theta}_q^{\mathrm{LS}})x_j = 0$ with $y_{qj} = \ln m_{qj}$. Subtracting, and using $y_{qj} - x_j^\top\hat{\theta}_q = r_{qj}$, gives $X^\top\Omega X(\hat{\theta}_q^{\mathrm{LS}} - \hat{\theta}_q) = \sum_j \tilde{\omega}_j(r_{qj} - e^{r_{qj}} + 1)x_j$. The bound $0 \le \varphi(r) \le \frac{1}{2}r^2 e^{|r|}$ follows from Taylor's theorem with Lagrange remainder. ∎

D. Proof of Lemma lem:cap

ψ_B is C^2, with $\psi_B''(t) = e^{\min(t,B)} \in (0, e^B]$. By the chain rule, $\nabla^2 f_i = \mathrm{diag}_q\big(\sum_j w_j\psi_B''(\cdot)x_j x_j^\top\big)$, which lies between 0 and P. For cocoercivity, let $h(u) = f_i(P^{-1/2}u)$. Then $\nabla^2 h \preceq I$, so ∇h is 1-cocoercive by the Baillon–Haddad theorem [56]. Substituting $\nabla h(u) = P^{-1/2}\nabla f_i(P^{-1/2}u)$ gives the claim. Strict convexity on $\mathbb{R}^{2Q}$ follows as in Proposition prop:ql(b), because $\psi_B'' > 0$ and every q has at least two active scales. Strong convexity on a bounded set follows from the continuity and positivity of the smallest eigenvalue of $\nabla^2 f_i$ on its closure.

For the cap, let $t_{iqj}^\star = a_{iqj} - \eta_{qj}(\hat{\theta}^\infty)$. Because $M_i \le M$, $e^{t_{iqj}^\star} \le (M(q,s_j)/c_j)e^{-\eta_{qj}} = (n(s_j)/c_j)e^{r_{qj}}$, so $t_{iqj}^\star \le B^\star < B$. By continuity, the capped criterion F_B coincides with F on a neighborhood U of $\hat{\theta}^\infty$. Then $\hat{\theta}^\infty$ minimizes $F_B + R$ on U, hence globally by convexity, and uniquely by strict convexity. ∎

E. Proof of Lemma lem:kkt

Since $\mathcal{N}$ is connected, $\ker\mathbf{L}_- = \ker\mathbf{M} = \{\mathbf{1}\otimes\theta\}$ and $\mathrm{range}(\mathbf{M}) = \ker(\mathbf{1}^\top \otimes I)$. ($\Rightarrow$) If $\hat{\theta}$ minimizes $F + R$, Fermat's rule and the sum rule (with F finite and differentiable everywhere) give $u \in \partial R(\hat{\theta})$ with $\nabla F(\hat{\theta}) + u = 0$. Set $\boldsymbol{\theta}^\star = \mathbf{1}\otimes\hat{\theta}$ and $\mathbf{v} = [\nabla f_i(\hat{\theta}) + u/K]_i$. Then $(\mathbf{1}^\top\otimes I)\mathbf{v} = 0$, so $\mathbf{v} \in \mathrm{range}(\mathbf{M})$, and there is a unique $\boldsymbol{\beta}^\star \in \mathrm{range}(\mathbf{M})$ with $\mathbf{M}\boldsymbol{\beta}^\star = -\mathbf{v}$. ($\Leftarrow$) $\mathbf{M}\boldsymbol{\theta}^\star = 0$ forces consensus. Multiplying the inclusion by $\mathbf{1}^\top \otimes I$ annihilates $\mathbf{M}\boldsymbol{\beta}^\star$ and gives $0 \in \nabla F(\hat{\theta}) + \sum_i \frac{1}{K}\partial R(\hat{\theta}) \subseteq \nabla F(\hat{\theta}) + \partial R(\hat{\theta})$, by convexity of $\partial R(\hat{\theta})$. ∎

F. Derivation of the Recursion (eq:incl)

Use $2D = L_+ + L_-$. The optimality condition of the E-step at node i is $0 \in \nabla f_i(\theta_i^{k+1}) + \frac{1}{K}\partial R(\theta_i^{k+1}) + 2\rho d_i\theta_i^{k+1} - b_i^k + \tau_e(\theta_i^{k+1} - \theta_i^k)$. Stacked, with $\mathbf{b}^k = \rho\mathbf{L}_+\boldsymbol{\theta}^k - \boldsymbol{\phi}^k$, it becomes

$$0 \in \nabla f(\boldsymbol{\theta}^{k+1}) + \partial g(\boldsymbol{\theta}^{k+1}) + \boldsymbol{\phi}^k + \rho\mathbf{L}_-\boldsymbol{\theta}^{k+1} + (\rho\mathbf{L}_+ + \tau_e I)(\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k).$$

By (eq:dual), $\boldsymbol{\phi}^k + \rho\mathbf{L}_-\boldsymbol{\theta}^{k+1} = \boldsymbol{\phi}^{k+1}$. Since $\boldsymbol{\phi}^0 = 0$, induction gives $\boldsymbol{\phi}^k \in \mathrm{range}(\mathbf{M})$, so $\boldsymbol{\phi}^k = \mathbf{M}\boldsymbol{\beta}^k$ with $\boldsymbol{\beta}^{k+1} = \boldsymbol{\beta}^k + \rho\mathbf{M}\boldsymbol{\theta}^{k+1}$. The L-step is the prox optimality condition $0 \in \frac{1}{K}\partial R(\theta_i^{k+1}) + (2\rho d_i + \tau_i)(\theta_i^{k+1} - u_i^k)$, which yields the same form with $\nabla f(\boldsymbol{\theta}^k)$ and τ_e replaced by $\mathrm{diag}(\tau_i)$.

G. Proof of Theorem thm:conv

Let $\mathbf{e}^k = \boldsymbol{\theta}^k - \boldsymbol{\theta}^\star$, $\Delta\boldsymbol{\theta} = \boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k$ and $\Delta\boldsymbol{\beta} = \boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^k$. Subtract the KKT inclusion of Lemma lem:kkt from (eq:incl) and take the inner product with $\mathbf{e}^{k+1}$. Monotonicity of ∂g removes the subgradient terms. Since $\mathbf{M}\boldsymbol{\theta}^\star = 0$, $\langle \boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^\star, \mathbf{M}\boldsymbol{\theta}^{k+1}\rangle = \frac{1}{\rho}\langle \boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^\star, \Delta\boldsymbol{\beta}\rangle$. With $2\langle a-b, a-c\rangle = \|a-b\|^2 + \|a-c\|^2 - \|b-c\|^2$ we obtain

$$\langle \nabla f(\boldsymbol{\xi}^k) - \nabla f(\boldsymbol{\theta}^\star), \mathbf{e}^{k+1}\rangle + \tfrac{1}{2}\big[V^{k+1} - V^k + \|\Delta\boldsymbol{\theta}\|_{\mathbf{G}}^2 + \tfrac{1}{\rho}\|\Delta\boldsymbol{\beta}\|^2\big] \le 0. \quad (24)$$

E-step: $\boldsymbol{\xi}^k = \boldsymbol{\theta}^{k+1}$, and the first term is nonnegative by convexity. *L-step:* with $\Delta\mathbf{g} = \nabla f(\boldsymbol{\theta}^k) - \nabla f(\boldsymbol{\theta}^\star)$, cocoercivity (Lemma lem:cap) and Young's inequality give $\langle \Delta\mathbf{g}, \mathbf{e}^k\rangle + \langle \Delta\mathbf{g}, \Delta\boldsymbol{\theta}\rangle \ge \|\Delta\mathbf{g}\|_{P^{-1}}^2 - \|\Delta\mathbf{g}\|_{P^{-1}}^2 - \frac{1}{4}\|\Delta\boldsymbol{\theta}\|_P^2$. Either way, (eq:descent) follows, and $\mathbf{G} - \gamma P \succ 0$ under the stated conditions because $\mathbf{L}_+ \succeq 0$.

Summing (eq:descent) shows that $\sum_k (\|\Delta\boldsymbol{\theta}\|^2 + \|\Delta\boldsymbol{\beta}\|^2) \le \infty$ and that $(\boldsymbol{\theta}^k, \boldsymbol{\beta}^k)$ is bounded. Take a cluster point $(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\beta}})$ along a subsequence k_t. Since $\Delta\boldsymbol{\theta}, \Delta\boldsymbol{\beta} \to 0$, passing to the limit in (eq:incl) uses continuity of ∇f and closedness of the graph of the maximally monotone ∂g [56]. It gives $0 \in \nabla f(\bar{\boldsymbol{\theta}}) + \partial g(\bar{\boldsymbol{\theta}}) + \mathbf{M}\bar{\boldsymbol{\beta}}$, and $\mathbf{M}\bar{\boldsymbol{\theta}} = \lim \Delta\boldsymbol{\beta}/\rho = 0$, so $(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\beta}})$ is a KKT pair, with $\bar{\boldsymbol{\beta}} \in \mathrm{range}(\mathbf{M})$ since that subspace is closed. Applying (eq:descent) with this pair in place of $(\boldsymbol{\theta}^\star, \boldsymbol{\beta}^\star)$ shows that V^k, now computed w.r.t. $(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\beta}})$, is nonincreasing and tends to 0 along k_t. Hence the whole sequence converges, and $\bar{\boldsymbol{\theta}} = \mathbf{1} \otimes \hat{\theta}$ by Lemma lem:kkt and uniqueness.

For the rate, (eq:incl) gives $\nabla f(\boldsymbol{\theta}^{k+1}) + \mathbf{u}^{k+1} + \mathbf{M}\boldsymbol{\beta}^{k+1} = \nabla f(\boldsymbol{\theta}^{k+1}) - \nabla f(\boldsymbol{\xi}^k) - \mathbf{G}\Delta\boldsymbol{\theta}$ for some $\mathbf{u}^{k+1} \in \partial g(\boldsymbol{\theta}^{k+1})$. Also $\mathbf{M}\boldsymbol{\theta}^{k+1} = \Delta\boldsymbol{\beta}/\rho$. Therefore $(\mathcal{R}^{k+1})^2 \le c_1\|\Delta\boldsymbol{\theta}\|^2 + c_2\|\Delta\boldsymbol{\beta}\|^2$, with $c_1 = 2(\gamma'\lambda_{\max}(P) + \lambda_{\max}(\mathbf{G}))^2$ ($\gamma' = 0$ for E, 1 for L) and $c_2 = 2/\rho^2$. The right-hand side is summable, and for a nonnegative summable sequence a_t, $k \min_{t \le k} a_t \le 2\sum_{t \ge k/2} a_t \to 0$. ∎

H. Proof of Theorem thm:linear

By (eq:descent), $V^k \le V^0$, so all iterates lie in the convex set S, which also contains $\boldsymbol{\theta}^\star$. *E-step.* Strong convexity on S gives $\langle \nabla f(\boldsymbol{\theta}^{k+1}) - \nabla f(\boldsymbol{\theta}^\star), \mathbf{e}^{k+1}\rangle \ge \mu\|\mathbf{e}^{k+1}\|^2$, and (24) gives

$$V^k - V^{k+1} \ge 2\mu\|\mathbf{e}^{k+1}\|^2 + \|\Delta\boldsymbol{\theta}\|_{\mathbf{G}}^2 + \tfrac{1}{\rho}\|\Delta\boldsymbol{\beta}\|^2. \quad (25)$$

Since $\boldsymbol{\beta}^0 = 0$ and $\boldsymbol{\beta}^\star \in \mathrm{range}(\mathbf{M})$, $\boldsymbol{\beta}^k - \boldsymbol{\beta}^\star \in \mathrm{range}(\mathbf{M})$, where $\|\mathbf{M}\mathbf{z}\|^2 \ge \lambda_2\|\mathbf{z}\|^2$. From (eq:incl) with g absorbed, $\mathbf{M}(\boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^\star) = -(\nabla f(\boldsymbol{\theta}^{k+1}) - \nabla f(\boldsymbol{\theta}^\star)) - \mathbf{G}\Delta\boldsymbol{\theta}$. Using $\|\mathbf{G}\mathbf{z}\|^2 \le \lambda_{\max}(\mathbf{G})\|\mathbf{z}\|_{\mathbf{G}}^2$,

$$\|\boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^\star\|^2 \le \tfrac{2}{\lambda_2}\big(\ell^2\|\mathbf{e}^{k+1}\|^2 + \lambda_{\max}(\mathbf{G})\|\Delta\boldsymbol{\theta}\|_{\mathbf{G}}^2\big).$$

Hence

$$V^{k+1} \le \Big(\lambda_{\max}(\mathbf{G}) + \tfrac{2\ell^2}{\rho\lambda_2}\Big)\|\mathbf{e}^{k+1}\|^2 + \tfrac{2\lambda_{\max}(\mathbf{G})}{\rho\lambda_2}\|\Delta\boldsymbol{\theta}\|_{\mathbf{G}}^2 \le \frac{1}{\delta}\,(V^k - V^{k+1}),$$

with δ as in (eq:rate), so $V^{k+1} \le V^k/(1+\delta)$.

L-step. Halving strong monotonicity on S and cocoercivity gives $\langle \Delta\mathbf{g}, \mathbf{e}^k\rangle \ge \frac{\mu}{2}\|\mathbf{e}^k\|^2 + \frac{1}{2}\|\Delta\mathbf{g}\|_{P^{-1}}^2$. Young's inequality gives $\langle \Delta\mathbf{g}, \Delta\boldsymbol{\theta}\rangle \ge -\frac{1}{2}\|\Delta\mathbf{g}\|_{P^{-1}}^2 - \frac{1}{2}\|\Delta\boldsymbol{\theta}\|_P^2$. Then (24) yields $V^k - V^{k+1} \ge \mu\|\mathbf{e}^k\|^2 + \|\Delta\boldsymbol{\theta}\|_{\mathbf{G}-P}^2 + \frac{1}{\rho}\|\Delta\boldsymbol{\beta}\|^2$. Bounding $\|\mathbf{e}^{k+1}\|_{\mathbf{G}}^2 \le 2\lambda_{\max}(\mathbf{G})(\|\mathbf{e}^k\|^2 + \|\Delta\boldsymbol{\theta}\|^2)$ and $\|\boldsymbol{\beta}^{k+1} - \boldsymbol{\beta}^\star\|^2 \le \frac{2}{\lambda_2}(\ell^2\|\mathbf{e}^k\|^2 + \lambda_{\max}(\mathbf{G})^2\|\Delta\boldsymbol{\theta}\|^2)$, and using $\mathbf{G} - P \succeq \gamma_0 I$, gives the stated constant. ∎

I. Proof of Corollary cor:err

By the triangle inequality, $\|\theta_i^k - \theta^{\mathrm{true}}\| \le \|\hat{\theta} - \theta^{\mathrm{true}}\| + \|\mathbf{e}^k\|$, and $\lambda_{\min}(\mathbf{G})\|\mathbf{e}^k\|^2 \le V^k \le (1+\delta)^{-k}V^0$. Solving for k gives the iteration count. With $\lambda_{\max}(\mathbf{G}) \le 2\rho d_{\max} + \tau_e$ and $\tau_e \to 0$, $\delta^{-1} \le \max\{(\rho d_{\max} + \ell^2/(\rho\lambda_2))/\mu,\ 4d_{\max}/\lambda_2\}$. The first term is minimized at $\rho = \ell/\sqrt{d_{\max}\lambda_2}$, where it equals $2(\ell/\mu)\sqrt{d_{\max}/\lambda_2}$. ∎

IX. Front-End Definitions and Implementation

All front-ends act on the profile Y and are implemented in `pnpmf/frontends.py`. Each returns, for the anchors owned by a node, the positive terms $T_v(s)$; zero terms are discarded and are not counted. Anchors live on the global index grid. A node with extended interval $[e_0, e_1)$ keeps an owned anchor v only if its window lies in $[e_0, e_1)$. This rule implements both the halo and the no-halo settings of Proposition prop:exact.

MFDFAm. Let $Q_s \in \mathbb{R}^{s\times m}$ be an orthonormal basis of the polynomials of degrees $1, \dots, m$ on s equispaced points, orthogonalized against constants. For a segment $y \in \mathbb{R}^s$ with mean $\bar{y}$, $F^2 = (\|y - \bar{y}\|^2 - \|Q_s^\top(y - \bar{y})\|^2)/s$. This costs $O(sm)$ per segment and $O(Nm)$ per scale, and avoids $s \times s$ projectors. The anchors are $\{ks\}$ and $\{(N \bmod s) + ks\}$, $k < \lfloor N/s \rfloor$, as in the two-pass scheme of [4], and $p = 1/2$.

MFDMA(ϑ). Let $c = \lfloor (s-1)\vartheta \rfloor$ and $\tilde{Y}(t) = \frac{1}{s}\sum_{u=t-(s-1-c)}^{t+c} Y(u)$, computed by cumulative sums. The residual $\varepsilon(t) = Y(t) - \tilde{Y}(t)$ is defined for $t \in [s-1-c, N-1-c]$. It is cut into $\lfloor (N-s+1)/s \rfloor$ segments of length s starting at $s-1-c$, and $T_v = \frac{1}{s}\sum \varepsilon^2$ over each segment [5]. Here $\vartheta = 0$ is the backward and $\vartheta = 1/2$ the centered moving average.

MODWT-SF and WL. Daubechies scaling filters with L_f vanishing moments are obtained by spectral factorization, so no external wavelet package is used. The non-circular MODWT pyramid [7] uses the filters $h/\sqrt{2}$ and $g/\sqrt{2}$, upsampled by 2^{j-1} at level j, with "valid" convolutions. Coefficient $W_j(t)$ therefore depends on $Y[t - L_j + 1..t]$ only, and is defined exactly on any interval that holds this window. MODWT-SF uses $T_v = |W_j(v)|$ at $v = k2^j$. WL uses $T_v = \max_{|t-v| \le 3\cdot 2^{j-1}} \max_{j' \le j} |W_{j'}(t)|$, the supremum over finer scales and a time neighborhood of length $3 \cdot 2^j$ [3], [8], again at $v = k2^j$. We use $L_f = 3$ and $s_j = 2^j$.

CWT-max. The Mexican-hat kernel $\psi(u) = (1-u^2)e^{-u^2/2}$ is sampled on $|u| \le \lceil 5a \rceil$, scaled by $1/a$ (L^1 normalization), and its mean is subtracted so that it annihilates constants exactly. At scale a, the anchors are the local maxima of $|W_a(t)|$ in t (strict on the left). The term is $\max_{|t'-t| \le \lceil a \rceil} \max_{a' \le a} |W_{a'}(t')|$, i.e., a supremum along the finer scales. This is a computationally light substitute for chaining maxima lines in WTMM [6]. Maxima detection is a deterministic local rule, so Assumption ass:local holds with a data-dependent χ_v. Agreement with the centralized computation holds up to FFT round-off, which can only flip a comparison between two numerically equal moduli.

MF-DXAm. The terms are $|\langle r_v^x, r_v^y\rangle|/s$, where r_v^x and r_v^y are the order-m detrending residuals of the two profiles on segment v, with $p = 1/2$ [31].

Numerical exactness. For MFDFA, MODWT-SF, WL and CWT-max, the pooled log-moments agree with the centralized ones to 10^{-12} (Table tab:exact). For MFDMA and MF-DXA the agreement is 10^{-8} to 10^{-9}, and the agreement is exact in exact arithmetic. The moving averages of MFDMA are differences of cumulative sums, whose magnitude depends on where a node's profile starts. The cross-products of MF-DXA can nearly cancel, and negative orders amplify the relative round-off of such small terms.

Appendix: Experimental Details

Synthetic signals. fGn is generated exactly by circulant embedding [54]. The MRW increments are $\varepsilon(t)e^{\omega(t)}$, where ω is Gaussian with $\mathrm{Cov}(\omega(t),\omega(t+\tau)) = \lambda^2 \ln(T/(|\tau|+1))$ for $|\tau| < T$ and mean $-\mathrm{Var}\,\omega$; it is generated by circulant embedding with integral scale $T = N$ [53]. The walk has $\zeta(q) = (\frac{1}{2}+\lambda^2)q - \frac{\lambda^2}{2}q^2$ at scales much smaller than T. The conservative binomial cascade assigns the weights $(a, 1-a)$ to the two children in random order at each of $\log_2 N$ levels, so $\zeta(q) = 1 - \log_2(a^q + (1-a)^q)$ for the MFDFA convention [4]. The theoretical spectra in the figures are Legendre transforms of these ζ.

Estimation settings. Orders: $\mathcal{Q} = \{\pm 0.5, \ldots, \pm 5\}$ (Section sec:exp), and $\{\pm 0.5, \ldots, \pm 8\}$ for the prior study. DFA, DMA and DXA scales: 14 integer log-spaced values in $[16, 4096]$, or 12 in $[16, 1024]$ when shards can be as short as 4096 samples (Fig. fig:naive); for the prior study, 8 values in $[16, \max(64, N/16)]$. Wavelets: Daubechies with $L_f = 3$, levels 4–12 (4–13 in Table tab:large). CWT: 12 geometric scales in $[8, 1024]$. QL: $w_j = 1$, c_j the expected global count, $B = 5$, and a resolution floor that discards terms with $T^p \le 10^{-6}\hat{\sigma}$, where $\hat{\sigma}$ is the global standard deviation (one scalar consensus). The floor only matters for quantized market data, where runs of identical prices make some segments exactly polynomial. ADMM: $\bar{\rho} = 1$, $\tau_e = 10^{-3}\rho$; E-step damped Newton with tolerance 10^{-12}; L-step $\tau_i = 0.6\lambda_{\max}(H_i)$, increased to 1.5 times the current local curvature whenever it is exceeded. Initialization: local log-log fits, $\phi^0 = 0$.

Baselines. CSS: Metropolis–Hastings weights [46] on the vector $[M_i/c,\ n_i/c]$ with $(Q+1)J$ entries; at each checkpoint every node solves the centralized problem on its running averages. DGD and PG-EXTRA [43], [44]: Metropolis weights, local copies of the prior R/K, and step $\alpha_0 s$ with $s \in \{1, 0.3, 0.1, 0.03\}$. Here $\alpha_0 = 1/L$ for DGD and $\alpha_0 = \lambda_{\min}(\tilde{W})/L$ for PG-EXTRA, where L is the largest node curvature at the solution and at the initial point; the best final error is reported.

Graphs and seeds. The graph seeds equal the experiment seeds, and connectivity is enforced by resampling. Every experiment is seeded, and `python -m pnpmf.experiments <name>` recreates the stored JSON.

X. Additional Results

Variance of shard averaging. Table VIII complements Fig. fig:naive with RMSEs. At $|q| \le 2$ all estimators are comparable. At $|q| = 5$ the pooled estimators of the cascade and the MRW have large variances, because a few extreme segments dominate the pooled moments. The shard average, a mean of logarithms, is less variable there, but its bias grows with K; for fGn it is the less accurate estimator once $K \ge 256$. The QL and pooled least-squares estimates differ only for the cascade at $|q| = 5$ (Proposition prop:qlc).

Fig. 7 shows the bias of $\hat{\zeta}(q)$ for the cascade and the MRW. The shard-average curves move away from the pooled one as K grows, at both ends of the q range.

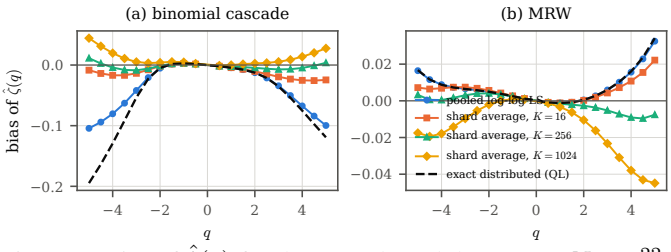


Figure 7: Bias of $\hat{\zeta}(q)$ for the cascade and the MRW ($N = 2^{22}$, MFDFA1, 16 realizations); compare Fig. fig:naiveb for fGn.

Front-ends per order. Fig. 8 resolves Table tab:fe by q. The RMSE of every front-end is smallest near $q = 0$ and grows towards the ends of the range. CWT-max deteriorates at negative q for MRW and fGn, and the MFDMA variants are inaccurate for the cascade at negative q.

Network scaling. Table IX lists the iteration counts of Fig. fig:scalinga with the algebraic connectivity of each graph.

A non-proximal plug-in denoiser. Fig. 9 runs the L-step on an MRW record ($N = 2^{20}$, $K = 32$, RGG) with the denoiser $\mathcal{D} = P_{\mathcal{C}} \circ S_\lambda$, a Tikhonov smoother across q followed by the concavity projection. This composition is not a proximal operator, so Theorem thm:conv does not apply. Nevertheless, the consensus residual decreases to 2.1×10^{-5} after 1500 iterations, and the fixed point differs from the concavity-only estimate by 2.3×10^{-3} in relative norm. With the proximal module $\iota_{\mathcal{C}}$, the same run reaches the centralized estimate to 5.0×10^{-9}.

Intraday data: generalized Hurst exponents. Fig. 10 shows $h(q)$ behind Fig. fig:real. For the S&P 500, MFDFA1 and WL agree, and $h(q)$ is nearly constant for $q < 0$. For EUR/USD, WL jumps to $h \approx 2$ for $q \le -1$. Over stale-quote stretches the profile is nearly linear, so the wavelet coefficients, which have three vanishing moments, and hence the leaders become very small and dominate the negative moments. MFDFA1 is affected much less. The monthly shard averages (dashed) dilute these effects, which is the layout dependence discussed in Section sec:exp-naive.

XI. Reproducibility

The package `pnpmf` (Python, NumPy/SciPy only) and the notebook `PnP_MF_ADMM.ipynb` are at https://github.com/Sawon1234/ADMM-for-Distributed-Multifractal-Analysis. Running the notebook in Google Colab with the default flags regenerates every figure and table from the stored JSON results in a few minutes. It then compares the regenerated tables with those included in this manuscript and runs the unit tests. Setting a flag recomputes the corresponding experiment; recomputing all of them takes about 2.5 hours on a Colab CPU. The real-data experiment fetches about 420 MB of public intraday data by a sparse `git` checkout.

References

[1] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, "Fractal measures and their singularities: The characterization of strange sets," *Phys. Rev. A*, vol. 33, no. 2, pp. 1141–1151, 1986.

Table VII: Final Relative Error After 2000 Iterations and Total Scalars Sent Until 10^{-6} ($K = 64$, $N = 2^{22}$; "—": not reached; "div.": diverged for every step size)

	MRW, ring		MRW, RGG		cascade, ring		cascade, RGG	
Method	final	sent	final	sent	final	sent	final	sent
PnP-MF-ADMM (E)	1.6×10^{-6}	—	9.4×10^{-14}	7.8×10^{6}	1.8×10^{-4}	—	2.5×10^{-9}	2.7×10^{7}
PnP-MF-ADMM (L)	1.5×10^{-6}	—	1.3×10^{-10}	2.3×10^{7}	4.2×10^{-2}	—	1.7×10^{-2}	—
PG-EXTRA	2.7×10^{-2}	—	1.8×10^{-2}	—	1.6×10^{-1}	—	1.6×10^{-1}	—
DGD	5.4×10^{-2}	—	7.6×10^{-3}	—	div.	—	div.	—
CSS	1.5×10^{-4}	—	6.0×10^{-11}	6.8×10^{7}	1.9×10^{-4}	—	3.7×10^{-11}	6.8×10^{7}

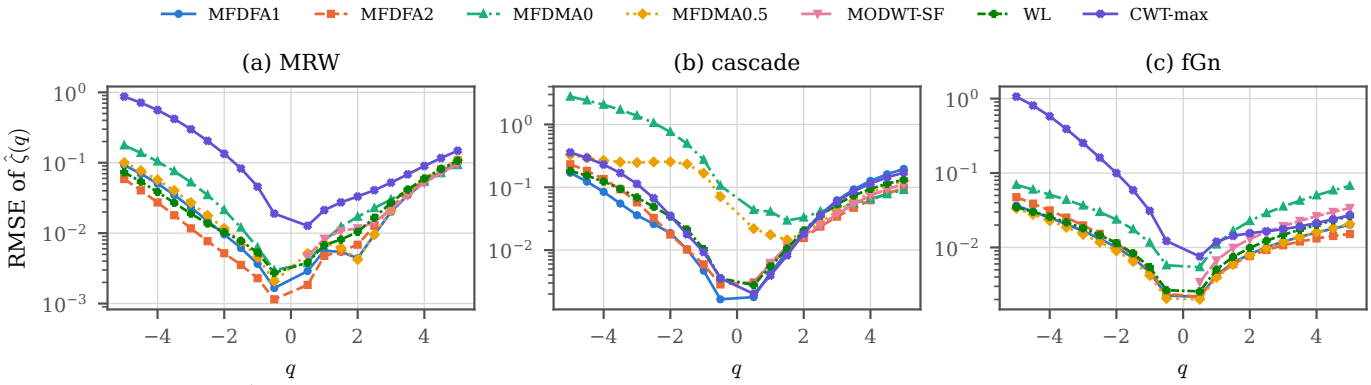


Figure 8: RMSE of $\hat{\zeta}(q)$ per order for the seven front-ends in the distributed pipeline ($N = 2^{20}$, $K = 8$, 10 realizations).

Table VIII: RMSE of $\hat{\zeta}(q)$ ($N = 2^{22}$, MFDFA1, 16 Realizations)

Model	Estimator	$q = -5$	$q = -2$	$q = 2$	$q = 5$
cascade	QL (exact)	0.456	0.021	0.017	0.223
	pooled LS	0.233	0.020	0.017	0.190
	shard avg., $K = 16$	0.061	0.008	0.012	0.049
	shard avg., $K = 256$	0.027	0.005	0.006	0.016
	shard avg., $K = 1024$	0.046	0.005	0.004	0.029
MRW	QL (exact)	0.046	0.006	0.005	0.116
	pooled LS	0.046	0.006	0.005	0.106
	shard avg., $K = 16$	0.031	0.006	0.004	0.039
	shard avg., $K = 256$	0.011	0.005	0.004	0.013
	shard avg., $K = 1024$	0.019	0.003	0.011	0.045
fGn	QL (exact)	0.034	0.006	0.003	0.007
	pooled LS	0.033	0.006	0.003	0.006
	shard avg., $K = 16$	0.036	0.007	0.002	0.005
	shard avg., $K = 256$	0.063	0.009	0.002	0.033
	shard avg., $K = 1024$	0.126	0.018	0.011	0.106

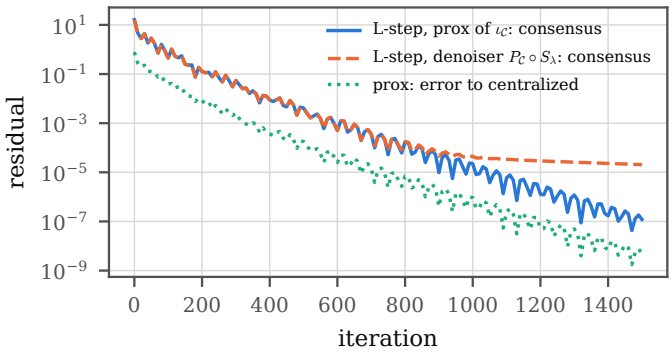


Figure 9: L-step with a proximal prior and with a plug-in denoiser ($N = 2^{20}$ MRW, $K = 32$, RGG). Consensus residual $\max_i \|\theta_i - \bar{\theta}\|$ and, for the proximal prior, the relative error to the centralized estimate.

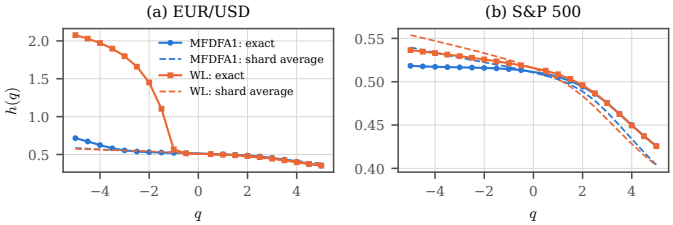


Figure 10: Generalized Hurst exponents of the intraday data: exact distributed estimates (solid) and monthly shard averages (dashed).

[2] J. F. Muzy, E. Bacry, and A. Arneodo, "Wavelets and multifractal formalism for singular signals: Application to turbulence data," *Phys. Rev. Lett.*, vol. 67, no. 25, pp. 3515–3518, 1991.

[3] S. Jaffard, "Wavelet techniques in multifractal analysis," in *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot*, ser. Proc. Symp. Pure Math. Providence, RI: Amer. Math. Soc., 2004, vol. 72, pp. 91–151.

[4] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, "Multifractal detrended fluctuation analysis of nonstationary time series," *Physica A*, vol. 316, no. 1–4, pp. 87–114, 2002.

[5] G.-F. Gu and W.-X. Zhou, "Detrending moving average algorithm for multifractals," *Phys. Rev. E*, vol. 82, p. 011136, 2010.

[6] J. F. Muzy, E. Bacry, and A. Arneodo, "Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method," *Phys. Rev. E*, vol. 47, no. 2, pp. 875–884, 1993.

[7] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[8] H. Wendt, P. Abry, and S. Jaffard, "Bootstrap for empirical multifractal analysis," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 38–48, 2007.

[9] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, "Multifractality in human heartbeat dynamics," *Nature*, vol. 399, pp. 461–465, 1999.

[10] E. A. F. Ihlen, "Introduction to multifractal detrended fluctuation analysis in Matlab," *Front. Physiol.*, vol. 3, p. 141, 2012.

[11] L. Calvet and A. Fisher, "Multifractality in asset returns: Theory and evidence," *Rev. Econ. Stat.*, vol. 84, no. 3, pp. 381–406, 2002.

[12] Z.-Q. Jiang, W.-J. Xie, W.-X. Zhou, and D. Sornette, "Multifractal analysis of financial markets: A review," *Rep. Prog. Phys.*, vol. 82, no. 12, p. 125901, 2019.

[13] B. Lashermes, S. G. Roux, P. Abry, and S. Jaffard, "Comprehensive multifractal analysis of turbulent velocity using the wavelet leaders," *Eur. Phys. J. B*, vol. 61, no. 2, pp. 201–215, 2008.

[14] H. Wendt, S. G. Roux, S. Jaffard, and P. Abry, "Wavelet leaders and

Table IX: Iterations of PnP-MF-ADMM (E) to 10^{-6} and Algebraic Connectivity λ_2 (in Parentheses); $N = 2^{22}$ MRW; Tori Need a Square K

Graph	$K = 16$	32	64	128	256
ring	286 (0.15)	535 (0.038)	1784 (0.0096)	>6000 (0.0024)	>6000 (0.0006)
grid	140 (2)	n/a	249 (0.59)	n/a	998 (0.15)
er	143 (2.4)	140 (1.2)	247 (2.4)	315 (2.5)	915 (3.4)
rgg	392 (0.12)	378 (0.25)	449 (0.19)	849 (0.21)	1161 (0.17)
star	144 (1)	150 (1)	278 (1)	554 (1)	1110 (1)
complete	143 (16)	141 (32)	160 (64)	277 (1.3e+02)	553 (2.6e+02)

bootstrap for multifractal analysis of images," *Signal Process.*, vol. 89, no. 6, pp. 1100–1114, 2009.

[15] S. Combrexelle, H. Wendt, N. Dobigeon, J.-Y. Tourneret, S. McLaughlin, and P. Abry, "Bayesian estimation of the multifractality parameter for image texture using a Whittle approximation," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2540–2551, 2015.

[16] B. Pascal, N. Pustelnik, and P. Abry, "Strongly convex optimization for joint fractal feature estimation and texture segmentation," *Appl. Comput. Harmon. Anal.*, vol. 54, pp. 303–322, 2021.

[17] E. A. F. Ihlen and B. Vereijken, "Interaction-dominant dynamics in human cognition: Beyond $1/f^{\alpha}$ fluctuation," *J. Exp. Psychol. Gen.*, vol. 139, no. 3, pp. 436–463, 2010.

[18] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, 2020.

[19] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[20] B. Lashermes, P. Abry, and P. Chainais, "New insights into the estimation of scaling exponents," *Int. J. Wavelets Multiresolut. Inf. Process.*, vol. 2, no. 4, pp. 497–523, 2004.

[21] F. Angeletti, M. Mézard, E. Bertin, and P. Abry, "Linearization effect in multifractal analysis: Insights from the random energy model," *Physica D*, vol. 240, no. 15, pp. 1245–1253, 2011.

[22] M. Ossiander and E. C. Waymire, "Statistical estimation for multiplicative cascades," *Ann. Statist.*, vol. 28, no. 6, pp. 1533–1560, 2000.

[23] R. W. M. Wedderburn, "Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method," *Biometrika*, vol. 61, no. 3, pp. 439–447, 1974.

[24] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, 2nd ed. London, U.K.: Chapman & Hall, 1989.

[25] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[26] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, "On the linear convergence of the ADMM in decentralized consensus optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1750–1761, 2014.

[27] Q. Ling, W. Shi, G. Wu, and A. Ribeiro, "DLM: Decentralized linearized alternating direction method of multipliers," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 4051–4064, 2015.

[28] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, 2013, pp. 945–948.

[29] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 84–98, 2017.

[30] B. Podobnik and H. E. Stanley, "Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series," *Phys. Rev. Lett.*, vol. 100, p. 084102, 2008.

[31] W.-X. Zhou, "Multifractal detrended cross-correlation analysis for two nonstationary signals," *Phys. Rev. E*, vol. 77, p. 066211, 2008.

[32] H. Wendt and P. Abry, "Multifractality tests using bootstrapped wavelet leaders," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4811–4820, 2007.

[33] R. Leonarduzzi, H. Wendt, P. Abry, S. Jaffard, C. Melot, S. G. Roux, and M. E. Torres, "p-exponent and p-leaders, Part II: Multifractal analysis. Relations to detrended fluctuation analysis," *Physica A*, vol. 448, pp. 319–339, 2016.

[34] R. Leonarduzzi, H. Wendt, P. Abry, S. Jaffard, and C. Melot, "Finite-resolution effects in p-leader multifractal analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3359–3368, 2017.

[35] P. Castiglioni and A. Faini, "A fast DFA algorithm for multifractal multiscale analysis of physiological time series," *Front. Physiol.*, vol. 10, p. 115, 2019.

[36] Y. Tsujimoto, Y. Miki, S. Shimatani, and K. Kiyono, "Fast algorithm for scaling analysis with higher-order detrending moving average method," *Phys. Rev. E*, vol. 93, p. 053304, 2016.

[37] N. Pustelnik, H. Wendt, P. Abry, and N. Dobigeon, "Combining local regularity estimation and total variation optimization for scale-free texture segmentation," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 468–479, 2016.

[38] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc WSNs with noisy links—Part I: Distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, 2008.

[39] G. Mateos, J. A. Bazerque, and G. B. Giannakis, "Distributed sparse linear regression," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5262–5276, 2010.

[40] A. Makhdoumi and A. Ozdaglar, "Convergence rate of distributed ADMM over networks," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 5082–5095, 2017.

[41] T.-H. Chang, M. Hong, and X. Wang, "Multi-agent distributed optimization via inexact consensus ADMM," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 482–497, 2015.

[42] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: An exact first-order algorithm for decentralized consensus optimization," *SIAM J. Optim.*, vol. 25, no. 2, pp. 944–966, 2015.

[43] ——, "A proximal gradient algorithm for decentralized composite optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6013–6023, 2015.

[44] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, 2009.

[45] S. A. Alghunaim, E. K. Ryu, K. Yuan, and A. H. Sayed, "Decentralized proximal gradient algorithms with linear convergence rates," *IEEE Trans. Autom. Control*, vol. 66, no. 6, pp. 2787–2794, 2021.

[46] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, 2004.

[47] C. Kumar and K. Rajawat, "Network dissensus via distributed ADMM," *IEEE Trans. Signal Process.*, vol. 68, pp. 2287–2301, 2020.

[48] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons, and C. A. Bouman, "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 408–423, 2016.

[49] E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 5546–5557.

[50] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[51] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

[52] C. Hildreth, "Point estimates of ordinates of concave functions," *J. Amer. Statist. Assoc.*, vol. 49, no. 267, pp. 598–619, 1954.

[53] E. Bacry, J. Delour, and J. F. Muzy, "Multifractal random walk," *Phys. Rev. E*, vol. 64, p. 026103, 2001.

[54] R. B. Davies and D. S. Harte, "Tests for Hurst effect," *Biometrika*, vol. 74, no. 1, pp. 95–101, 1987.

[55] T. G. Andersen and T. Bollerslev, "Intraday periodicity and volatility persistence in financial markets," *J. Empir. Finance*, vol. 4, no. 2–3, pp. 115–158, 1997.

[56] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. Cham, Switzerland: Springer, 2017.