Wigglyness is Diagonalisable

James Chok and Geoffrey M. Vasil

Tnks @ Orgnz.

Linear Regression

$$\min_{a_0, a_1 \in \mathbb{R}} \sum_{i=1}^{N} [y_i \ - \ (a_0 + a_1 x_i)]^2$$

$$X \ = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}, \quad a \ = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} \quad y \ = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

$$\min_{a \in \mathbb{R}^2} \ \|y - Xa\|^2$$

Polynomial Regression

$$\min_{a_0,\ldots,a_M\in\mathbb{R}} \sum_{i=1}^{N} [y_i \;-\; (a_0 + a_1\, x_i + \cdots + a_M\, x_i^M)]^2$$

$$X = \begin{pmatrix} 1 & x_1 & \cdots & x_1^M \\ 1 & x_2 & \cdots & x_2^M \\ \vdots & \vdots & & \\ 1 & x_N & \cdots & x_N^M \end{pmatrix}, \quad a = \begin{pmatrix} a_0 \\ \vdots \\ a_M \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

$$\min_{a\in\mathbb{R}^{M+1}} \|y - Xa\|^2$$

Polynomial Regression

$$\min_{a \in \mathbb{R}^{M+1}} \|y - Xa\|^2$$

$$\begin{aligned} \mathcal{L} \;=\; \|y - Xa\|^2 \;&=\; (y - Xa)^T (y - Xa) \\ &=\; a^T X^T X a \;-\; 2a^T X^T y \;+\; y^T y \end{aligned}$$

$$\frac{\partial \mathcal{L}}{\partial a} \;=\; 2a^T X^T X \;-\; 2X^T y \;=\; 0 \;\;\Longrightarrow\;\; a \;= (X^T X)^{-1} X^T y$$

Plot Wiggly Polynomial Regression

Polynomial Regression

$$X = \begin{pmatrix} 1 & x_1 & \cdots & x_1^M \\ 1 & x_2 & \cdots & x_2^M \\ \vdots & \vdots & & \\ 1 & x_N & \cdots & x_N^M \end{pmatrix}, \qquad a = (X^T X)^{-1} X^T y$$

Vandermonde Matrix is Numerically Bad

Polynomial Regression

$$X = \begin{pmatrix} 1 & x_1 & \cdots & x_1^M \\ 1 & x_2 & \cdots & x_2^M \\ \vdots & \vdots & & \\ 1 & x_N & \cdots & x_N^M \end{pmatrix}, \qquad a = (X^T X)^{-1} X^T y$$

$M = (X^T X)$ we hope $M^{-1} M = I$

Polynomial Regression (Preconditioning)

$$\sum_{i=0}^{M} a_i\, x^i \;=\; \sum_{i=0}^{M} c_i\, P_i(x)$$

$\{P_i(x)\}_{i=0}^{M}$ spans polynomials of degree M

Polynomial Regression

$$\min_{a_0,\ldots,a_M \in \mathbb{R}} \sum_{i=1}^{N} \left[y_i - \sum_{i=0}^{M} c_i P_i(x) \right]^2$$

$$X = \begin{pmatrix} P_0(x_1) & P_1(x_1) & \cdots & P_M(x_2) \\ P_0(x_2) & P_1(x_2) & \cdots & P_M(x_2) \\ \vdots & \vdots & & \\ P_0(x_N) & P_1(x_N) & \cdots & P_M(x_N) \end{pmatrix}, \quad c = \begin{pmatrix} c_0 \\ \vdots \\ c_M \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

$$\min_{a \in \mathbb{R}^{M+1}} \|y - Xc\|^2 \implies c = (X^T X)^{-1} X^T y$$

Wigglyness

$$\mathscr{W}(f) \;=\; \int (f''(x))^2 \, dx$$

Penalized Regression

$$\mathscr{W}(f) \;=\; \int (f''(x))^2\,dx$$

$$\min_f \|y - f(x)\|^2 \qquad \text{such that} \qquad \mathscr{W}(f) \leq \eta$$

$$\min_f \|y - f(x)\|^2 \qquad \text{such that} \qquad \mathscr{W}(f) \le \eta$$

Equivalent

$$\min_f \left[\|y - f(x)\|^2 \; + \; \lambda \mathscr{W}(f) \right]$$

$$\min_f \|y - f(x)\|^2 \qquad \text{such that} \qquad \mathscr{W}(f) \leq \eta$$

For small η, λ is big

$$\min_f \left[\|y - f(x)\|^2 \; + \; \lambda \mathscr{W}(f) \right]$$

Smoothing Penalty

$$\min_{a_0,\ldots,a_M \in \mathbb{R}} \sum_{i=1}^{N} \left[y_i - \sum_{k=0}^{M} a_k \, x_i^k \right]^2 + \lambda \int_0^1 \left(\frac{d^2}{dx^2} \sum_{k=0}^{M} a_k \, x^k \right)^2 dx$$

Wigglyness is Not Diagonalisable

L2 - Regularisation

$$\min_{a_0,\ldots,a_M \in \mathbb{R}} \sum_{i=1}^{N} \left[y_i - \sum_{k=0}^{M} a_k\, x_i^k \right]^2 + \lambda \sum_{i=0}^{M} a_k^2$$

$$\min_{a_0,\ldots,a_M \in \mathbb{R}} \sum_{i=1}^{N} \left[y_i - \sum_{k=0}^{M} a_k\, x_i^k \right]^2 \quad \text{such that} \quad \sum_{i=0}^{M} a_k^2 \le \eta$$

L2 - Regularisation

$$\min_{a_0,\ldots,a_M \in \mathbb{R}} \sum_{i=1}^{N} \left[y_i - \sum_{k=0}^{M} a_k \, x_i^k \right]^2 + \lambda \sum_{i=0}^{M} a_k^2$$

$$a = (X^T X + \lambda I)^{-1} X^T y$$

Show Plot of l2-regularization

Polynomial Regression

$$\min_{a_0,\ldots,a_M \in \mathbb{R}} \sum_{i=1}^{N} \left[y_i - \sum_{i=0}^{M} c_i P_i(x) \right]^2 + \lambda \int_0^1 \left(\frac{d^2}{dx^2} \sum_{i=0}^{M} c_i P_i(x) \right)^2 dx$$

$P_i(x)$ = i-th Legendre Polynomial

Show Plot of High/Low-pass filter

Classical Orthogonal Polynomials have Orthogonal Derivatives

Classical Orthogonal Polynomials

- Jacobi Polynomials
- Hermite Polynomials
- Laguerre Polynomials

Classical Orthogonal Polynomials

$$< P_n, P_m > = \int_\Omega P_n(x)\, P_m(x)\, d\mu_0 \;=\; c_{n,0}\, \delta_{n,m}$$

$$< P_n^{(k)}, P_m^{(k)} > = \int_\Omega P_n^{(k)}(x)\, P_m^{(k)}(x)\, d\mu_k \;=\; c_{n,k}\, \delta_{n,m}$$

Smoothing Penalty

$$\int_0^1 \left(\frac{d^k}{dx^k} \sum_{n=0}^{N} c_n P_n(x) \right)^2 d\mu_k = < \sum_{n=0}^{N} c_n P_n^{(k)}(x), \sum_{n=0}^{N} c_n P_n^{(k)}(x) >_{\mu_k}$$

$$= \sum_{n=0}^{N} c_n^2 \; < P_n^{(k)}(x), P_n^{(k)}(x) >_{\mu_k}$$

Smoothing Penalty

$$\sum_{k\geq 0}\int_0^1\left(\frac{d^k}{dx^k}\sum_{n=0}^{N} c_n P_n(x)\right)^2 d\mu_k \;=\; \sum_{k\geq 0}\sum_{n=0}^{N} c_n^2 \;< P_n^{(k)}(x), P_n^{(k)}(x) >_{\mu_k}$$

$$=\; \sum_{n=0}^{N} c_n^2 \;\sum_{k=0}^{n} < P_n^{(k)}(x), P_n^{(k)}(x) >_{\mu_k}$$

$$\mathscr{W}(f) \qquad \sim\; \sum_{n=0}^{N} c_n^2\; n^{2n}$$

Compare with polynomial wigglyness

Wigglyness is Diagonalisable

Smoothing Penalty

Sobolev-Jacobi Smoothing

$$\sum_{k\geq 0} \int_0^1 \left(\frac{d^k}{dx^k} \sum_{n=0}^{N} c_n P_n(x) \right)^2 d\mu_k \ \sim \ \sum_{n=0}^{N} c_n^2 \, n^{2n}$$

Classical Smoothing

$$\int_0^1 \left(\frac{d^2}{dx^2} \sum_{n=0}^{N} c_n P_n(x) \right)^2 d\mu_k \ \sim \ ??$$

Smoothing Penalty vs L2-Penalty

Sobolev-Jacobi Smoothing

$$\sum_{k\geq 0}\int_0^1 \left(\frac{d^k}{dx^k} \sum_{n=0}^{N} c_n P_n(x) \right)^2 d\mu_k \ \sim \ \sum_{n=0}^{N} c_n^2 \ n^{2n}$$

L2-Penalty

$$\sum_{n=0}^{N} c_n^2$$

Polynomial Regression

$$\min_{a_0,\ldots,a_M \in \mathbb{R}} \sum_{i=1}^{N} \left[y_i - \sum_{m=0}^{M} c_m P_m(x) \right]^2 + \lambda \sum_{m=0}^{M} c_m^2 m^{2m}$$

$$\min_{a_0,\ldots,a_M \in \mathbb{R}} \sum_{i=1}^{N} \left[y_i - \sum_{m=0}^{M} c_m P_m(x) \right]^2 \quad \text{such that} \quad \sum_{m=0}^{M} c_m^2 m^{2m} \leq \eta$$

Polynomial Regression

$$\min_{a_0,\ldots,a_M \in \mathbb{R}} \sum_{i=1}^{N} \left[y_i - \sum_{m=0}^{M} c_m P_m(x) \right]^2 \quad \text{such that} \quad \sum_{m=0}^{M} c_m^2 m^{2m} \le \eta$$

$$M = 2 \implies c_0^2 + c_1^2 + 16 c_2^2 \le \eta$$

Polynomial Regression

$$\min_{a_0,\ldots,a_M \in \mathbb{R}} \sum_{i=1}^{N} \left[y_i - \sum_{m=0}^{M} c_m P_m(x) \right]^2 \quad \text{such that} \quad \sum_{m=0}^{M} c_m^2 m^{2m} \leq \eta$$

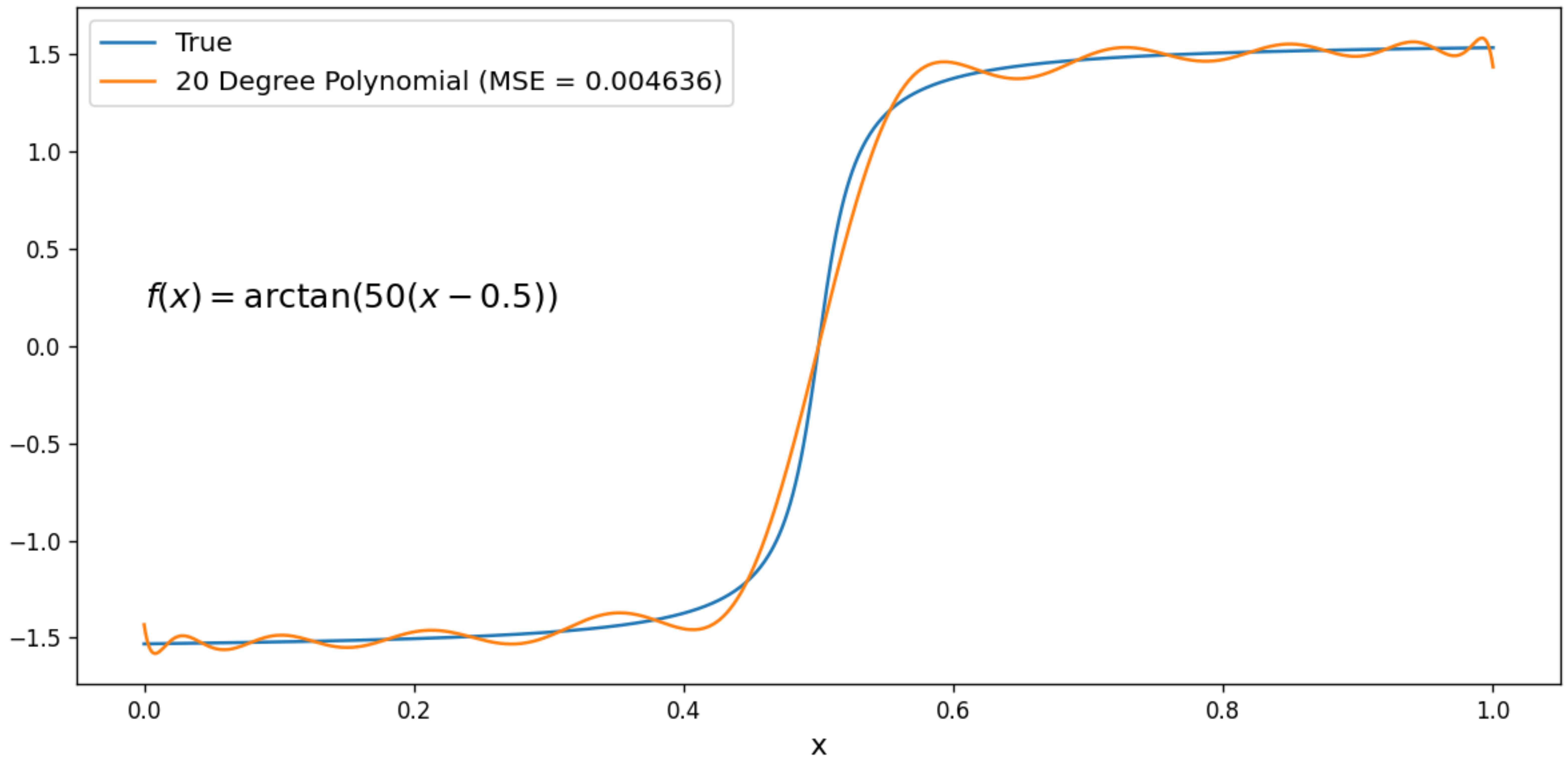
$$M = 2 \implies c_0^2 + c_1^2 + 16c_2^2 \leq \eta$$

$$c = (X^T X + \mathrm{Diag}(m^{2m}))^{-1} X^T y$$

Polynomial Approximation

$$f(x) \approx \sum_{n=0}^{N} a_n P_n(x) = \sum_{n=0}^{N} c_n x^n$$

(e.g. Legendre or Chebyshev)



Rational Approximation

$$f(x) \;\approx\; R_{N,M}(x) = \sum_{n=0}^{N} a_n P_n(x) \Bigg/ \sum_{m=0}^{M} b_m Q_m(x)$$

- Reduces Runge's Phenomena
- Faster convergence than ordinary polynomials

AAA Algorithm

Nakatsukasa, Sète, and Trefethen (2018)

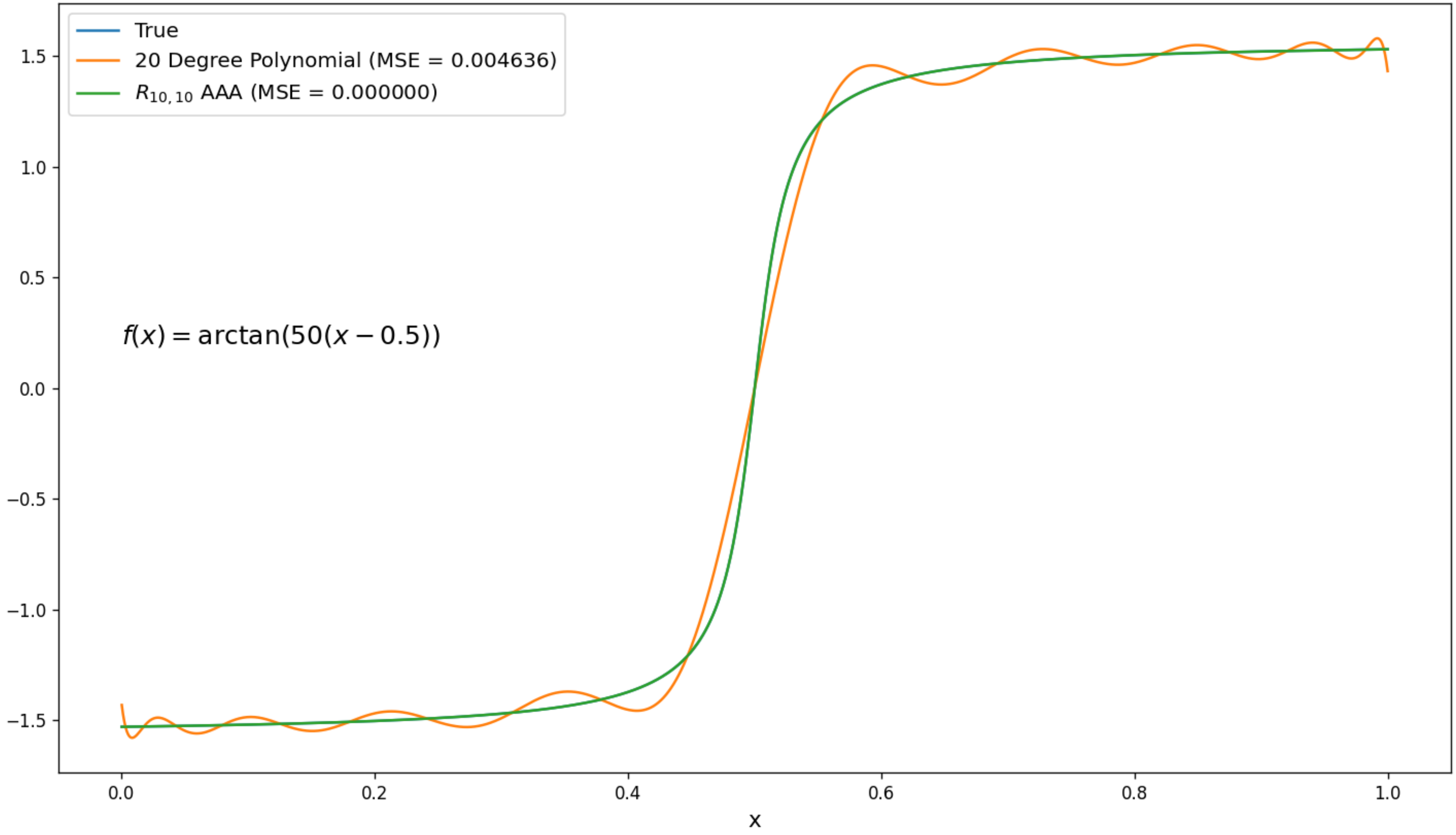
$$f(x) \approx R_{N,N}(x) = \sum_{n=0}^{N} \frac{w_n f_n}{x - x_n} \Bigg/ \sum_{n=0}^{N} \frac{w_n}{x - x_n}$$

$$f_n \;=\; f(x_n), \qquad w_n \;\neq\; 0$$

$0 < x_0 < x_1 < \ldots < x_N \leq 1$ partitions $[0,1]$. Non-zero denominator at x_i

$$\min_{w} \sum_{i} \left[f_i \left(\sum_{n=0}^{N} \frac{w_n}{x_i - x_n} \right) - \left(\sum_{n=0}^{N} \frac{w_n f_n}{x_i - x_n} \right) \right]^2$$

Normalizing Condition: $\|w\| = 1$



True
20 Degree Polynomial (MSE = 0.004636)
$R_{10,10}$ AAA (MSE = 0.000000)
$f(x) = \arctan(50(x - 0.5))$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.2
0.4
0.6
0.8
1.0
x

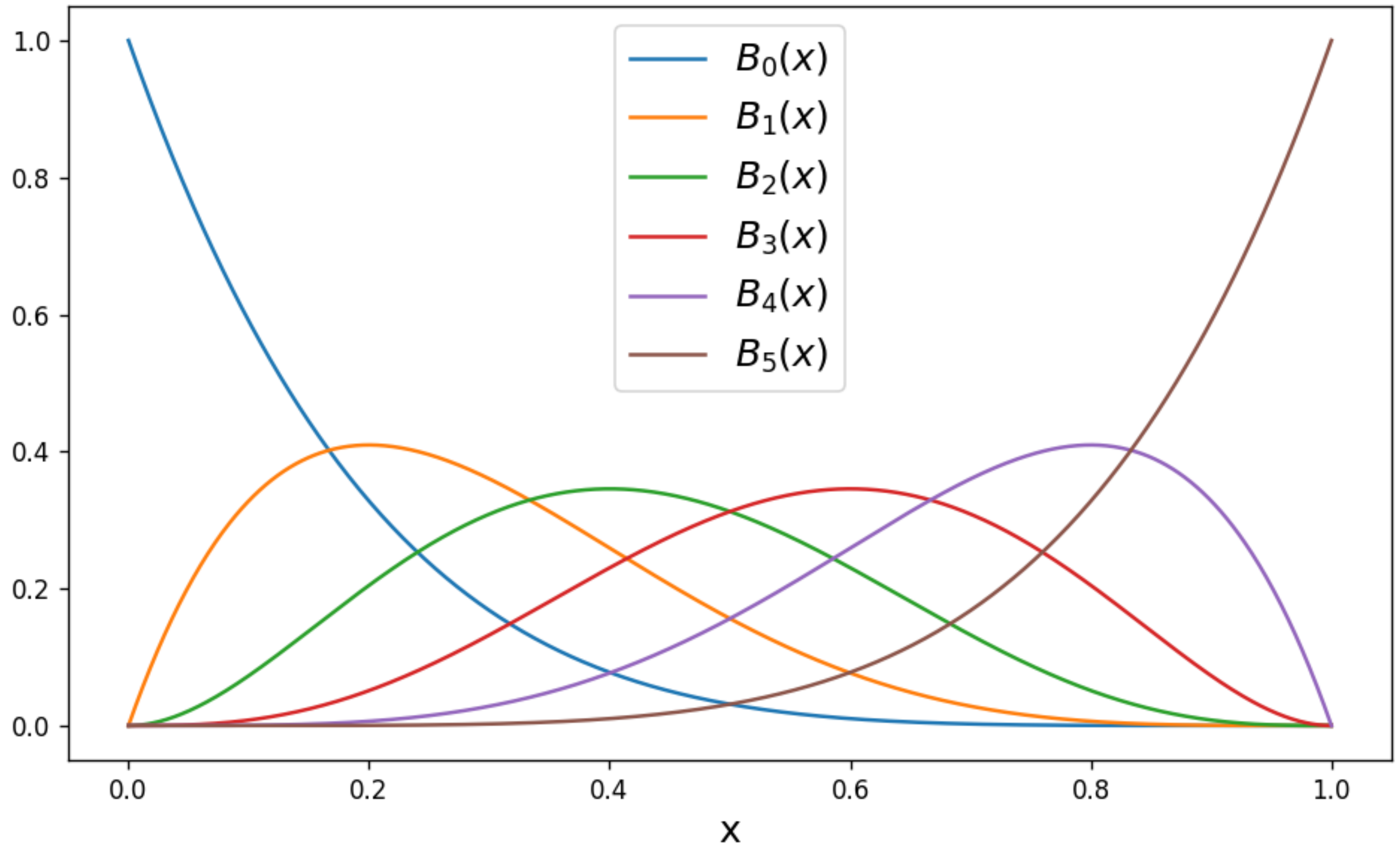
New Problem

$$f(x) \;\approx\; R_{N,M}(x) = \frac{P(x)}{Q(x)} \quad \text{with} \quad \underline{Q(x) > 0} \quad \text{for} \quad x \in [0,1]$$

Bernstein Polynomials

$$B_k^{(N)}(x) \;=\; \binom{N}{k} x^k (1-x)^{N-k}$$

$$B_k^{(N)}(x) \;>\; 0 \quad \text{for} \quad x \in (0,1)$$



Sergei Natanovich Bernstein

Our Proposal

$$f(x) \;\approx\; R_{N,M}(x) = \frac{P(x)}{Q(x)} \quad \text{with} \quad Q(x) > 0 \quad \text{for} \quad x \in [0,1]$$

$$Q(x) \;=\; \sum_{m=0}^{M} w_m B_m(x) \quad \text{where} \quad B_m(x) \;=\; \binom{M}{m} x^m (1-x)^{M-m}$$

Our Proposal

$$f(x) \;\approx\; R_{N,M}(x) = \frac{P(x)}{Q(x)} \quad \text{with} \quad Q(x) > 0 \quad \text{for} \quad x \in [0,1]$$

$$Q(x) \;=\; \sum_{m=0}^{M} w_m B_m(x) \quad \text{where} \quad B_m(x) \;=\; \binom{M}{m} x^m (1-x)^{M-m}$$

Positivity

$$w_m \;\geq\; 0$$

Normalization

$$\sum_m w_m \;=\; 1$$

Our Proposal

$$f(x) \;\approx\; R_{N,M}(x) = \frac{P(x)}{Q(x)} \quad \text{with} \quad Q(x) > 0 \quad \text{for} \quad x \in [0,1]$$

$$Q(x) \;=\; \sum_{m=0}^{M} w_m\, B_m(x) \quad \text{where} \quad B_m(x) \;=\; \binom{M}{m} x^m\, (1-x)^{M-m}$$

Positivity

$$w_m \;\geq\; 0$$

Normalization

$$\sum_m w_m \;=\; 1$$

$$f(x) \;\approx\; R_{N,M}(x) = \sum_{n=0}^{N} a_n\, P_n(x) \Big/ \sum_{m=0}^{M} w_m\, B_m(x)$$

For some $\{P_n(x)\}_n$

How to Solve

$$f(x) \;\approx\; R_{N,M}(x) = \sum_{n=0}^{N} a_n P_n(x) \Big/ \sum_{m=0}^{M} w_m B_m(x)$$

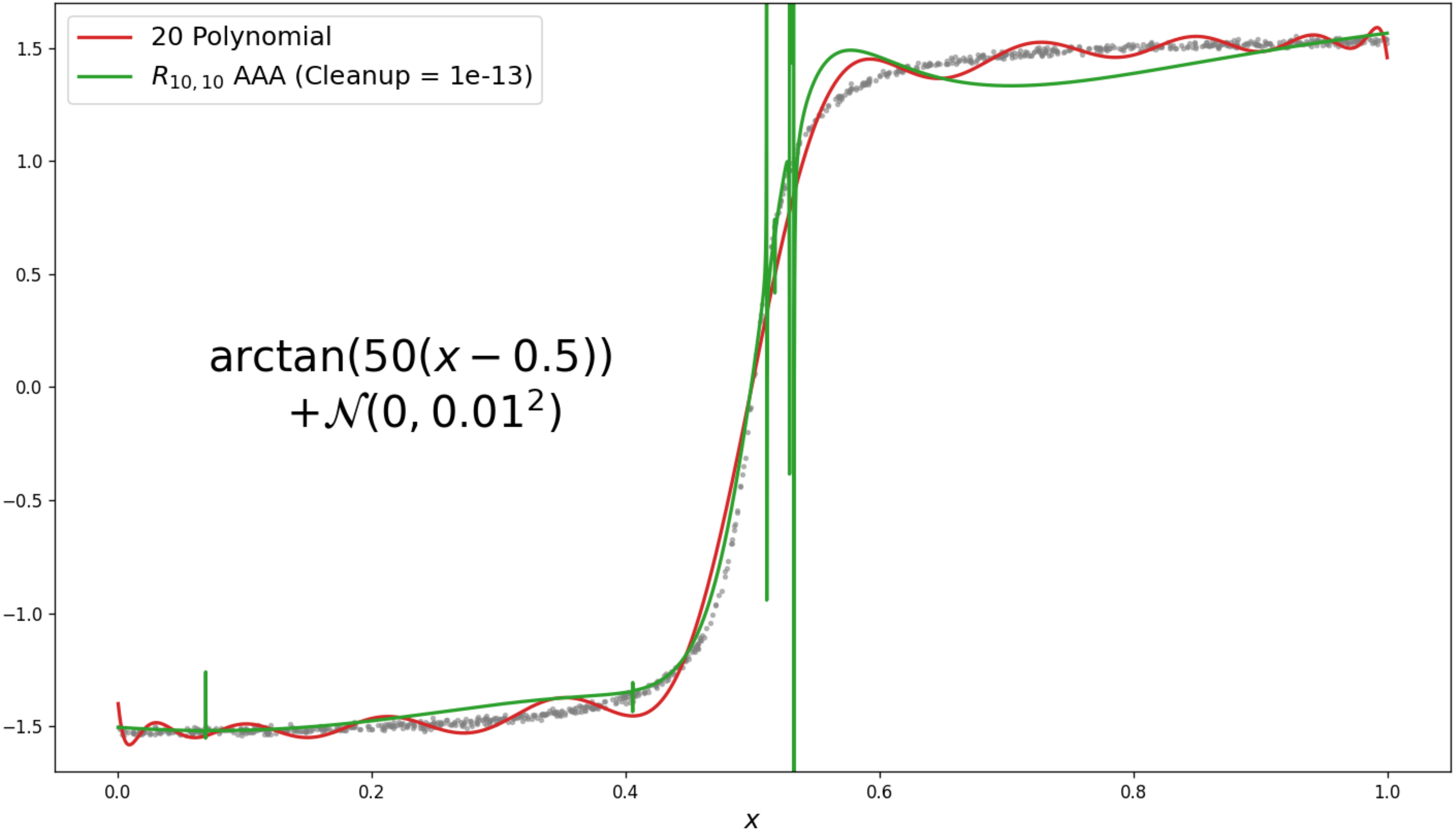
$$w \;\in\; \Delta^{M+1} \;=\; \left\{ w \in \mathbb{R}^{M+1} \mid w_m \;\geq\; 0 \text{ and } \sum_m w_m \;=\; 0 \right\}$$

Linearized Problem

$$\min_{\substack{a \in \mathbb{R}^{N+1} \\ w \in \Delta^{M+1}}} \left\| f(x) \sum_m w_m B_m(x) \;-\; \sum_n a_n P_n(x) \right\|$$

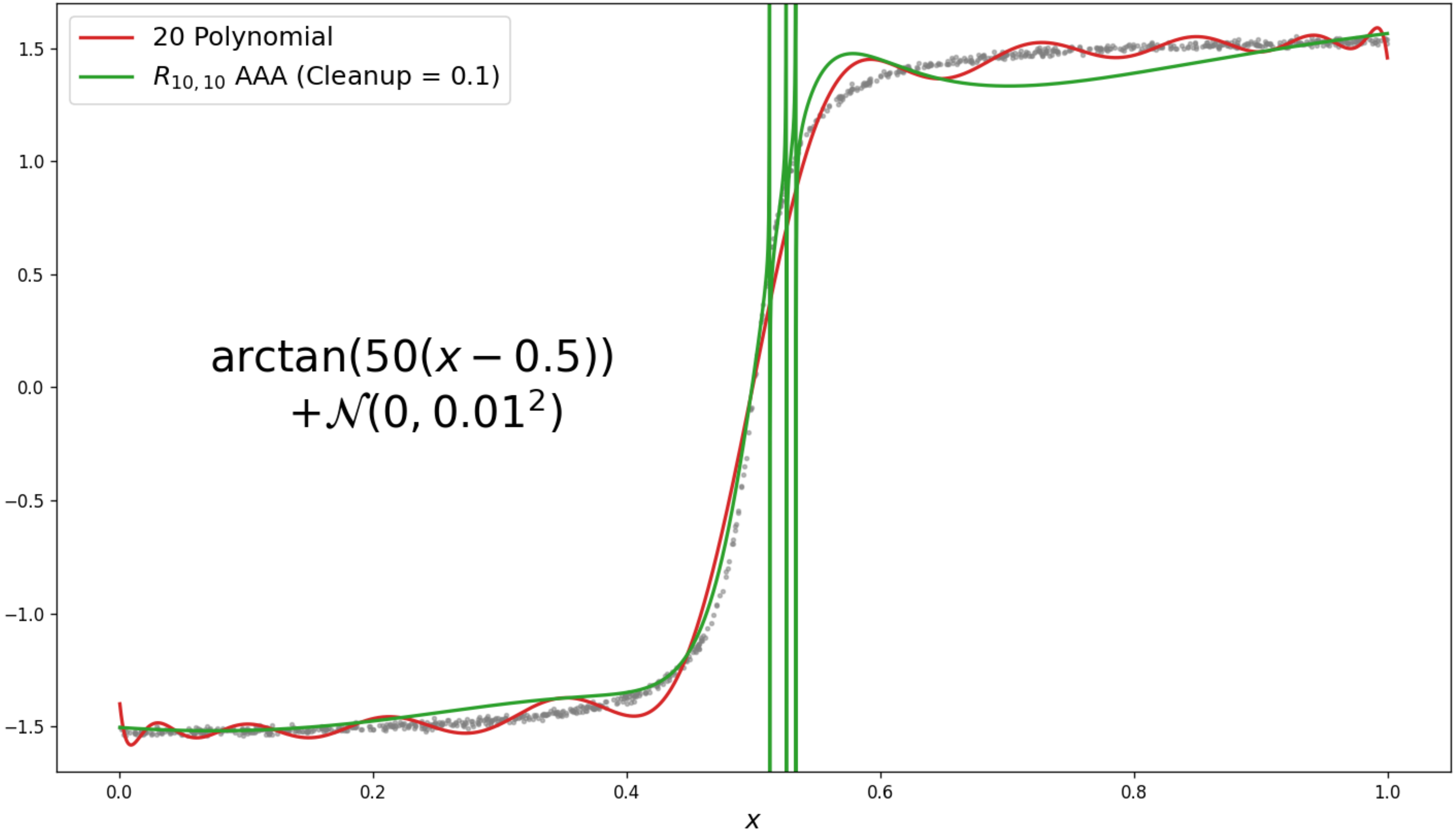
Very much like Lanczos (1938)

Noisy Data



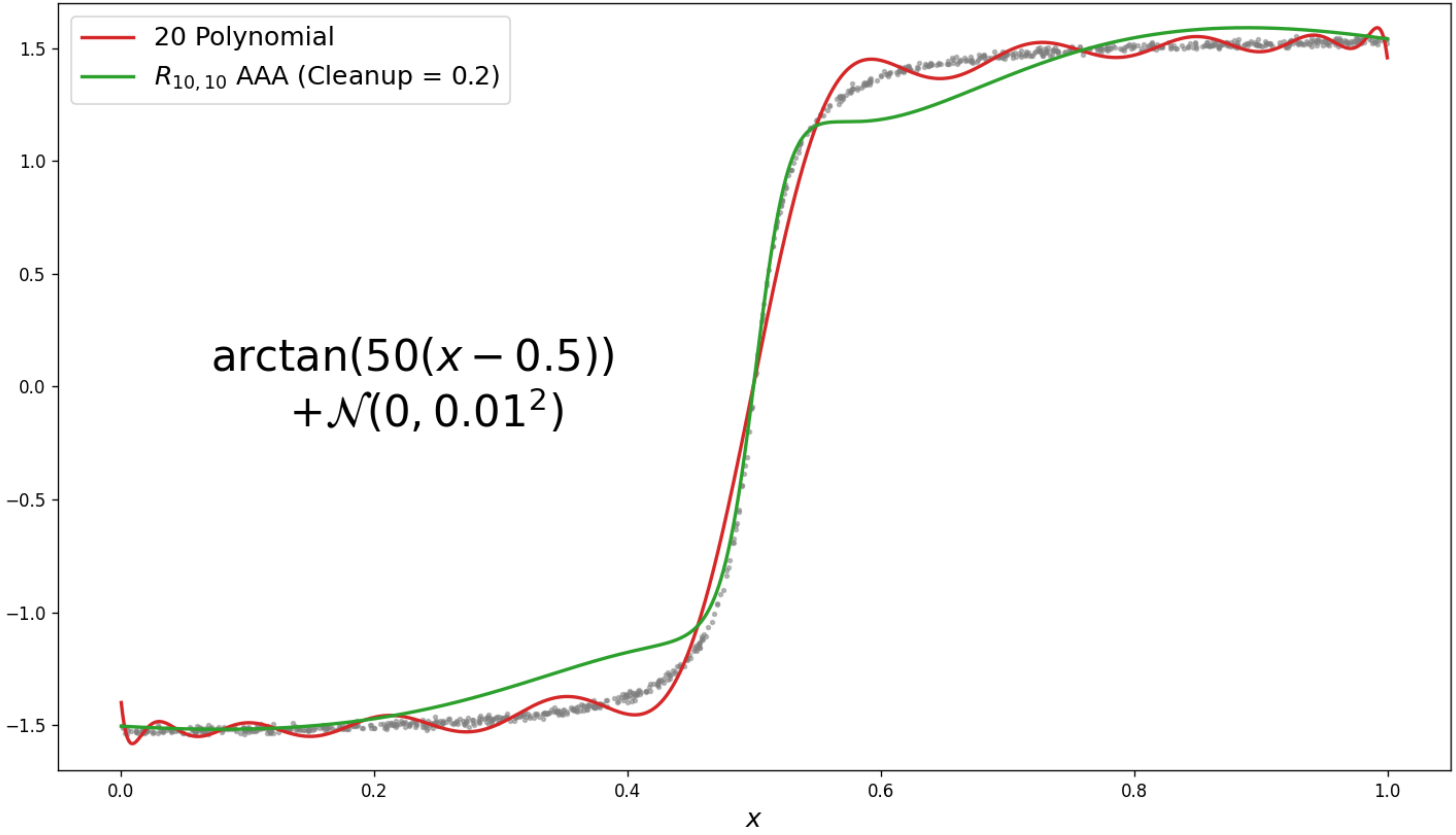
20 Polynomial
$R_{10,10}$ AAA (Cleanup = 1e-13)
$\arctan(50(x-0.5))$
$+\mathcal{N}(0, 0.01^2)$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.2
0.4
0.6
0.8
1.0
x

Noisy Data



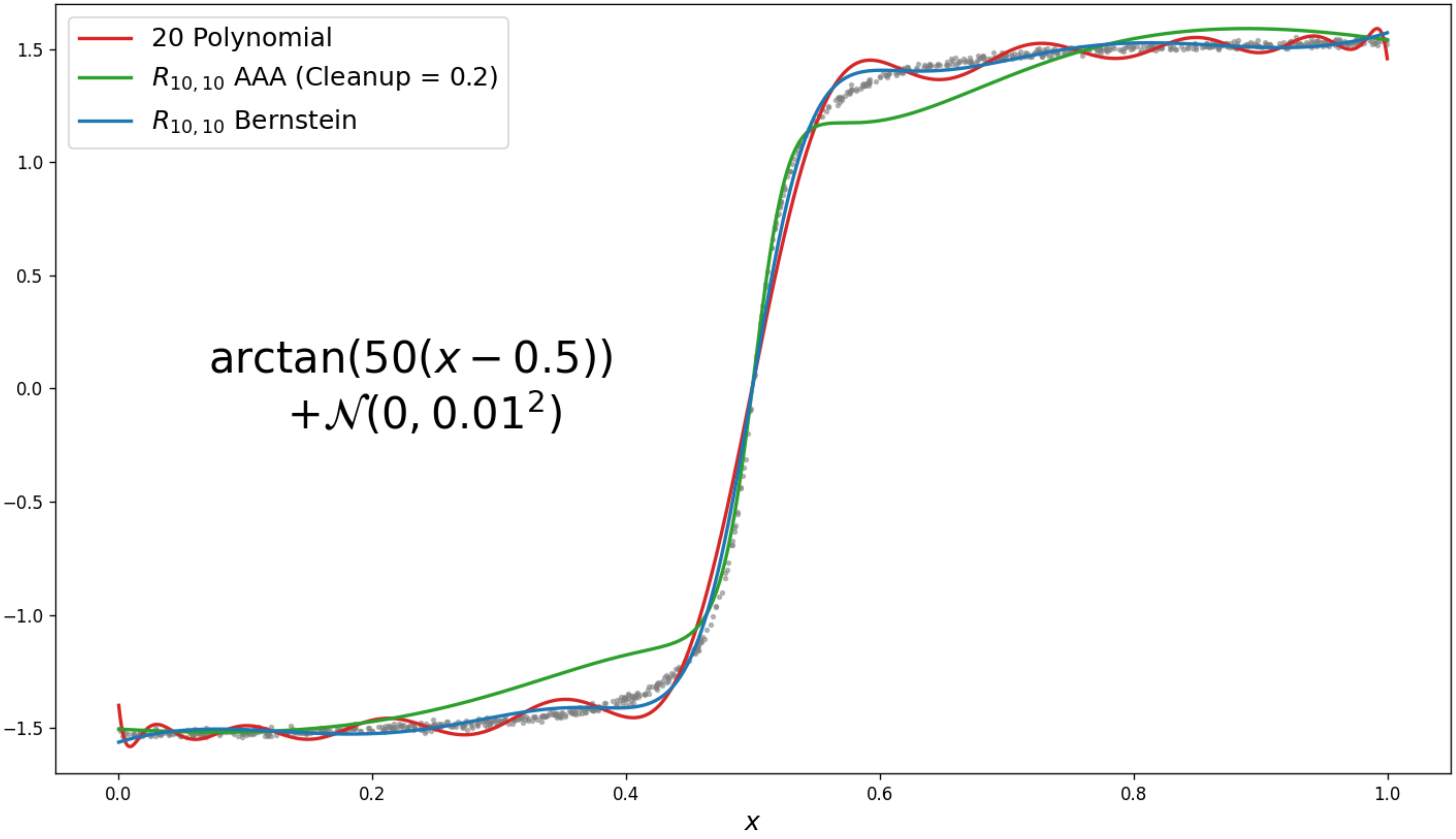
20 Polynomial
$R_{10,10}$ AAA (Cleanup = 0.1)
$\arctan(50(x-0.5))$
$+\mathcal{N}(0, 0.01^2)$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.2
0.4
0.6
0.8
1.0
x

Noisy Data



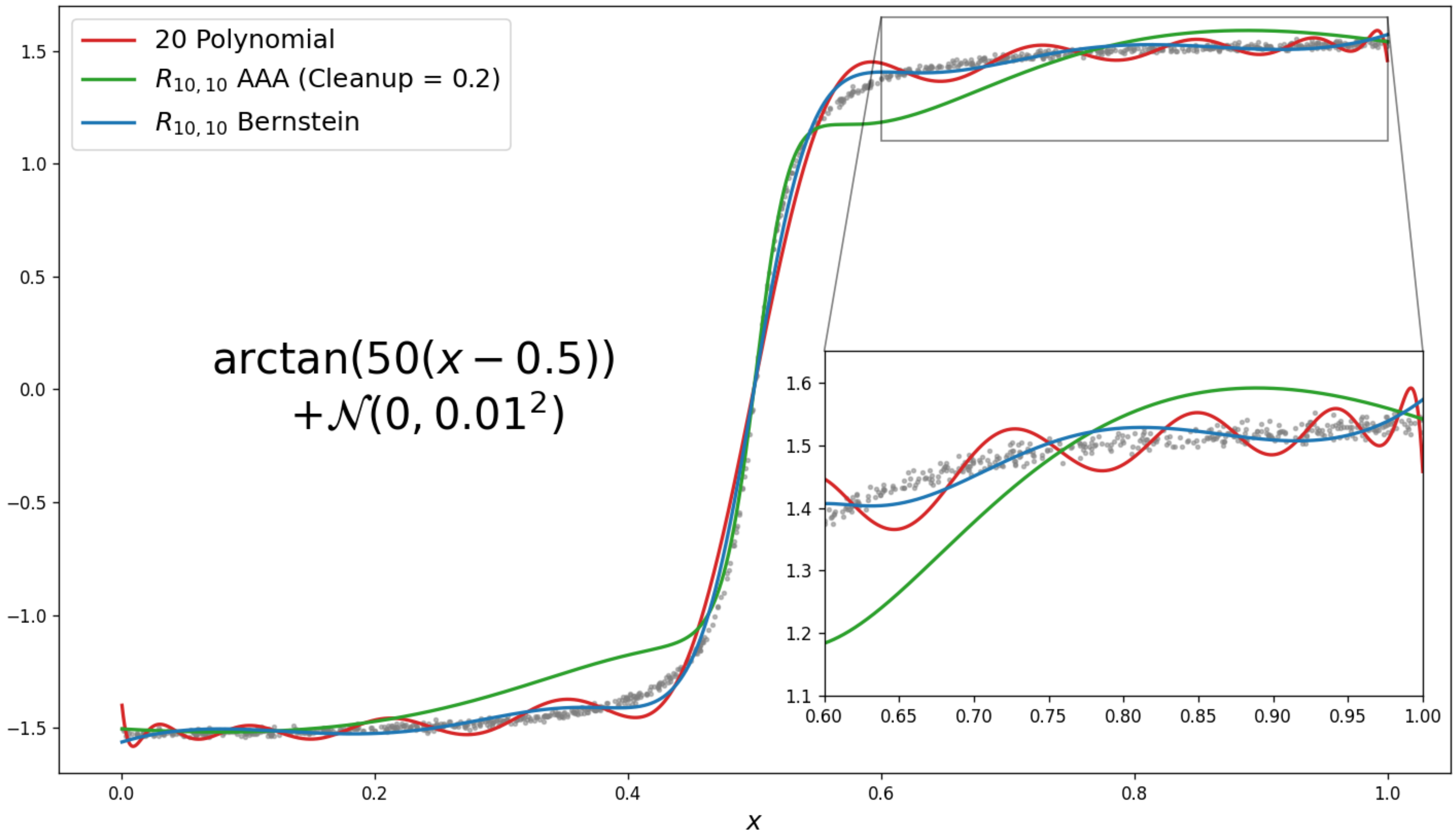
20 Polynomial
$R_{10,10}$ AAA (Cleanup = 0.2)
$\arctan(50(x-0.5))$
$+\mathcal{N}(0, 0.01^2)$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.2
0.4
0.6
0.8
1.0
x

Noisy Data



20 Polynomial
$R_{10,10}$ AAA (Cleanup = 0.2)
$R_{10,10}$ Bernstein
$\arctan(50(x-0.5))$
$+\mathcal{N}(0, 0.01^2)$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.2
0.4
0.6
0.8
1.0
x

Noisy Data



20 Polynomial
$R_{10,10}$ AAA (Cleanup = 0.2)
$R_{10,10}$ Bernstein
$\arctan(50(x-0.5))$
$+\mathcal{N}(0, 0.01^2)$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.2
0.4
0.6
0.8
1.0
x
1.6
1.5
1.4
1.3
1.2
1.1
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00

Penalization & smoothing

$$f(x) \;\approx\; R_{N,M}(x) \quad \text{and} \quad R_{N,M} \text{ is smooth}$$

$$\approx\; R_{N,M}(x) \quad \text{and} \quad \text{Numerator is smooth}$$

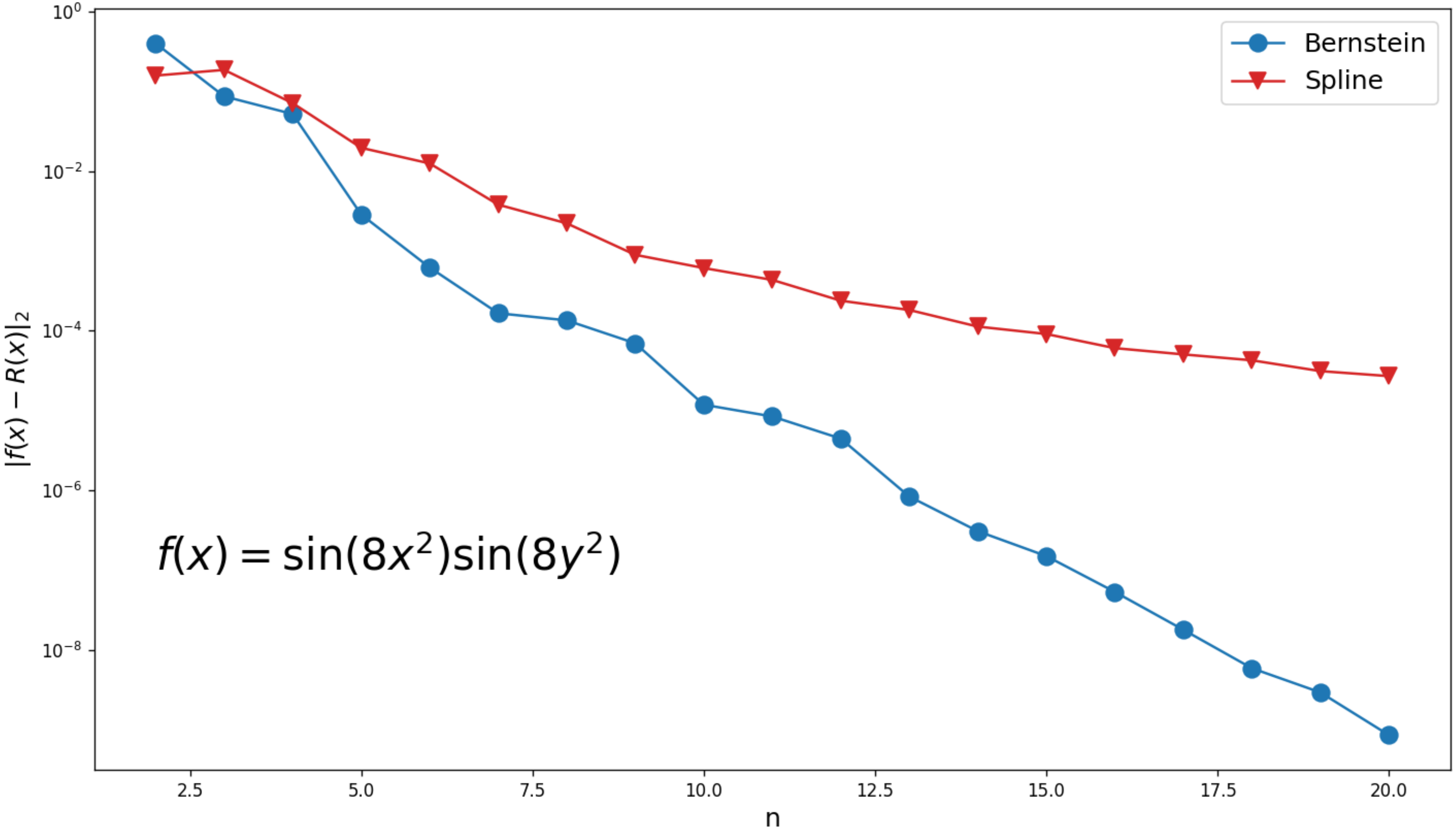
$$\rightarrow \quad \min_{g} \left\| f(x) \;-\; g(x) \right\| \;+\; \sum_{k \geq 0} \lambda_k \int (g^{(k)}(x))^2 \, d\mu_k \quad \text{for} \quad \lambda_k \;\geq\; 0$$

Bivariate

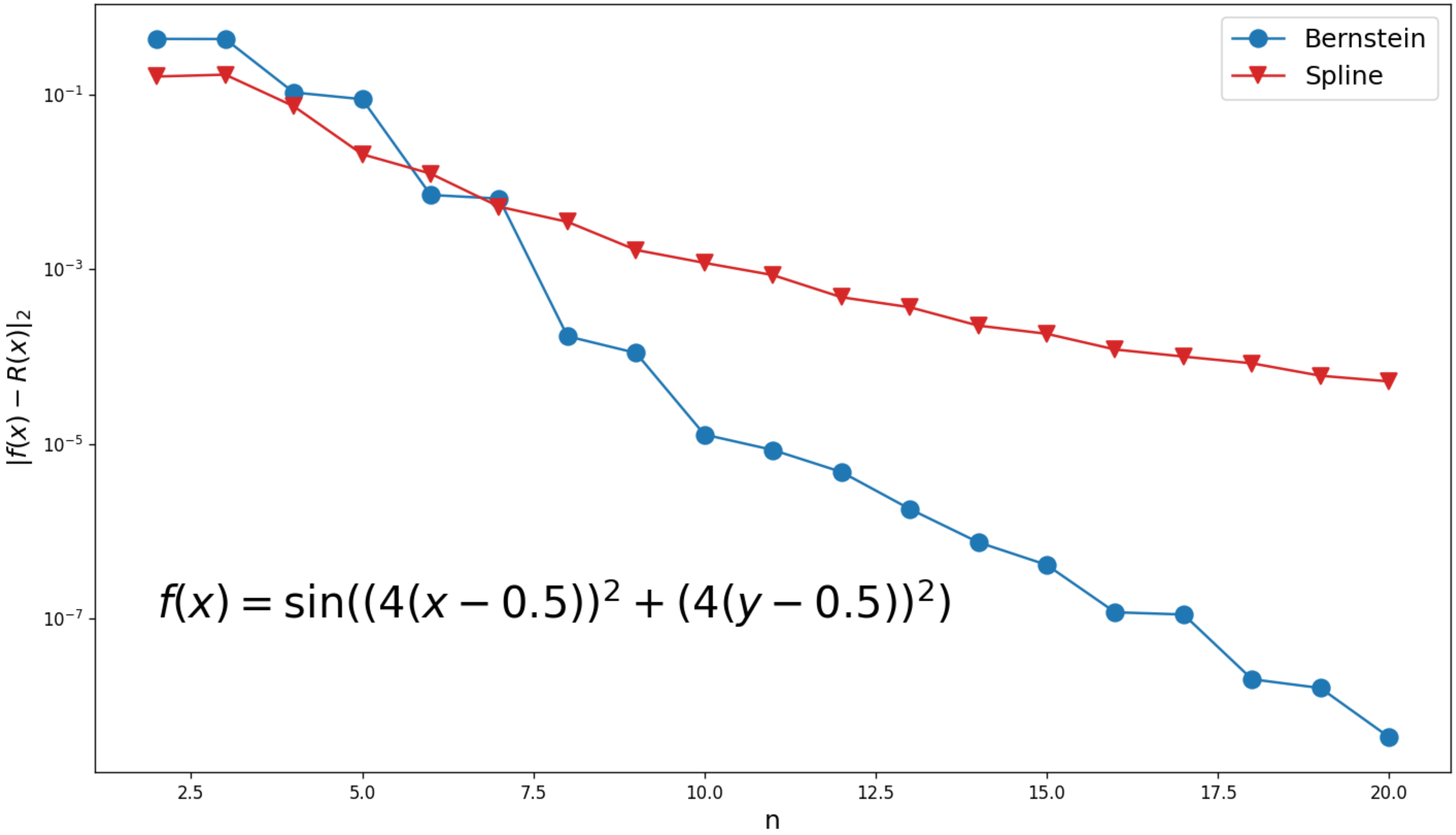
$$R(x,y) \;=\; \sum_{n,\,m} a_{n,\,m}\, P_n(x)\, P_m(y) \Bigg/ \sum_{j,\,k} w_{j,\,k}\, B_j(x)\, B_k(y)$$

$$w_{j,k} \;\geq\; 0 \quad \text{and} \quad \sum_{j,k} w_{j,k} \;=\; 1$$

Numerical Convergence

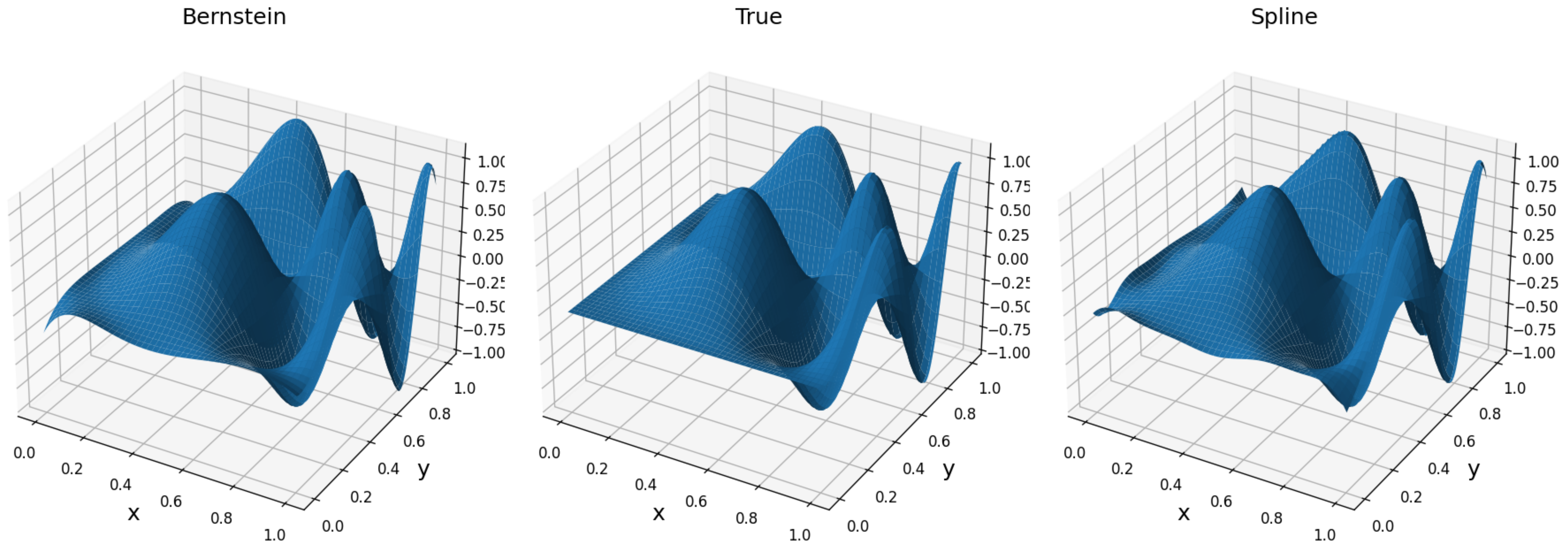


Numerical Convergence



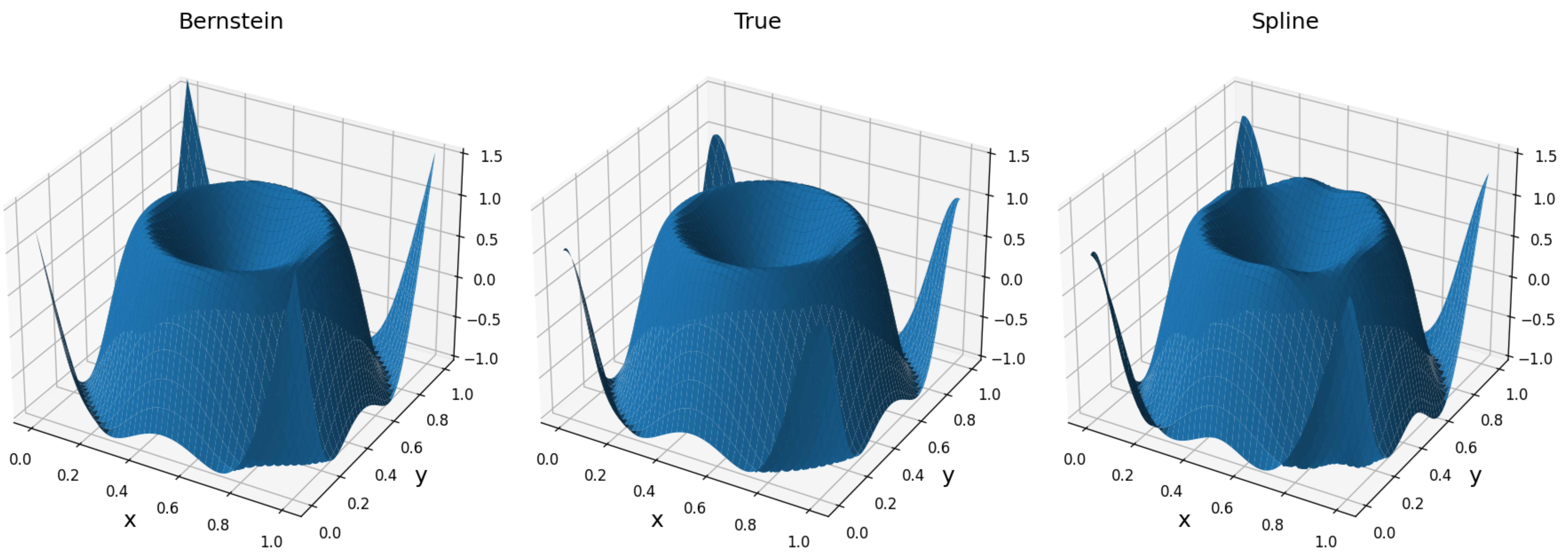
Bernstein vs Spline

$$\sin(8x^2 + 8y^2) + \mathcal{N}(0, 0.1^2)$$

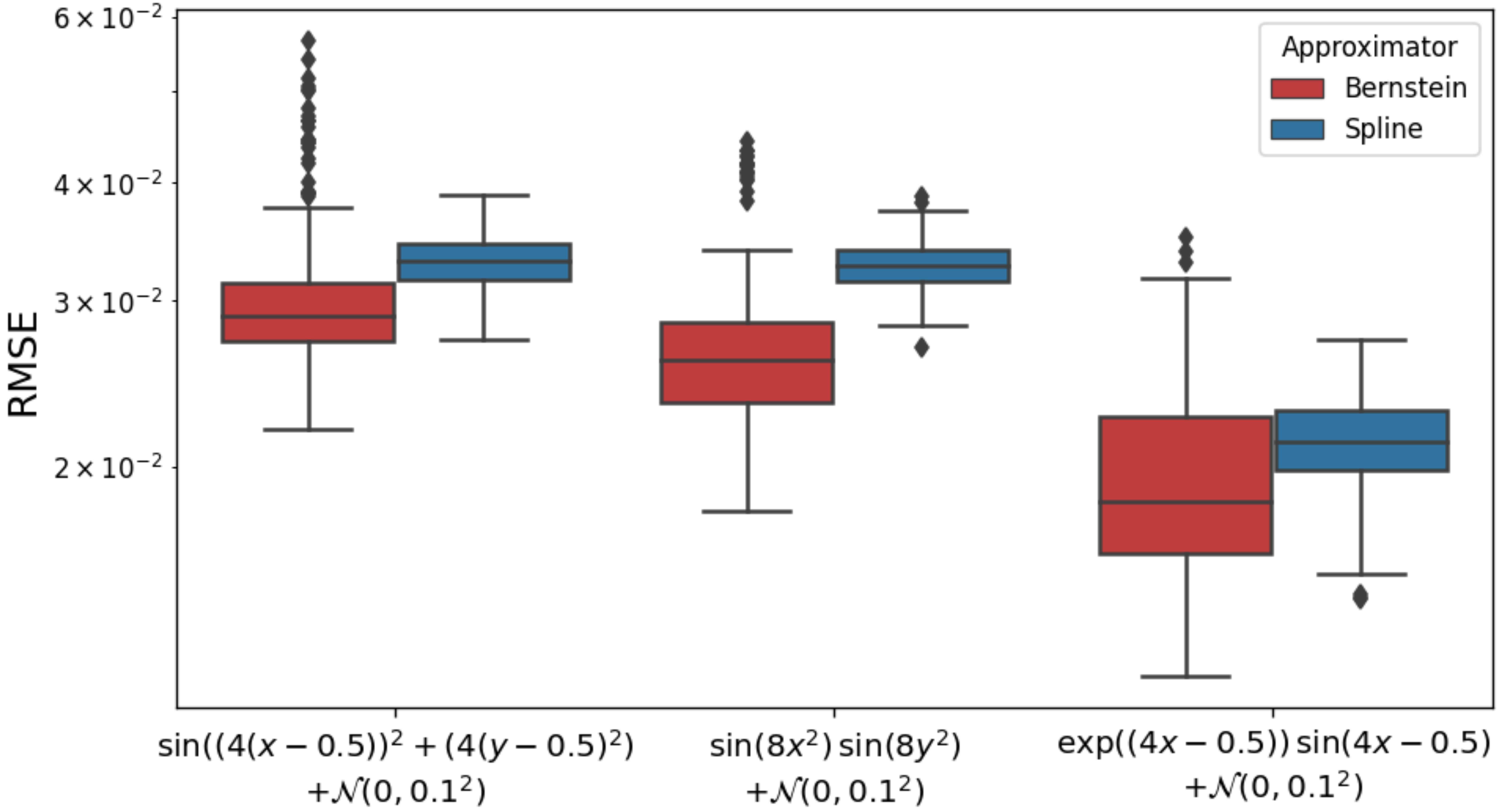


Bernstein vs Spline

$$\sin\left((4(x-0.5))^2 \;+\; (4(y-0.5))^2\right) \;+\; \mathcal{N}(0, 0.1^2)$$

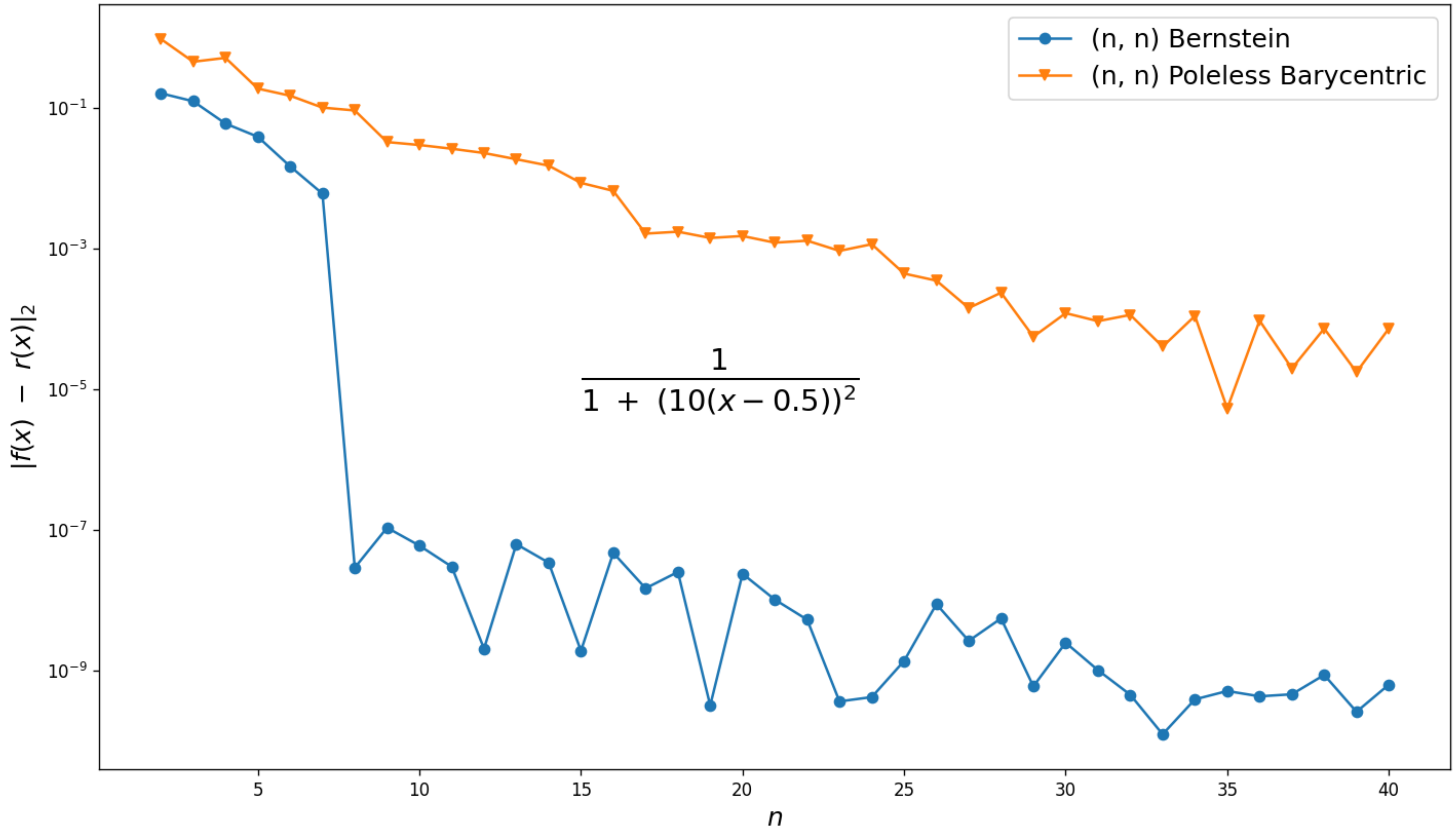


Bernstein vs Spline



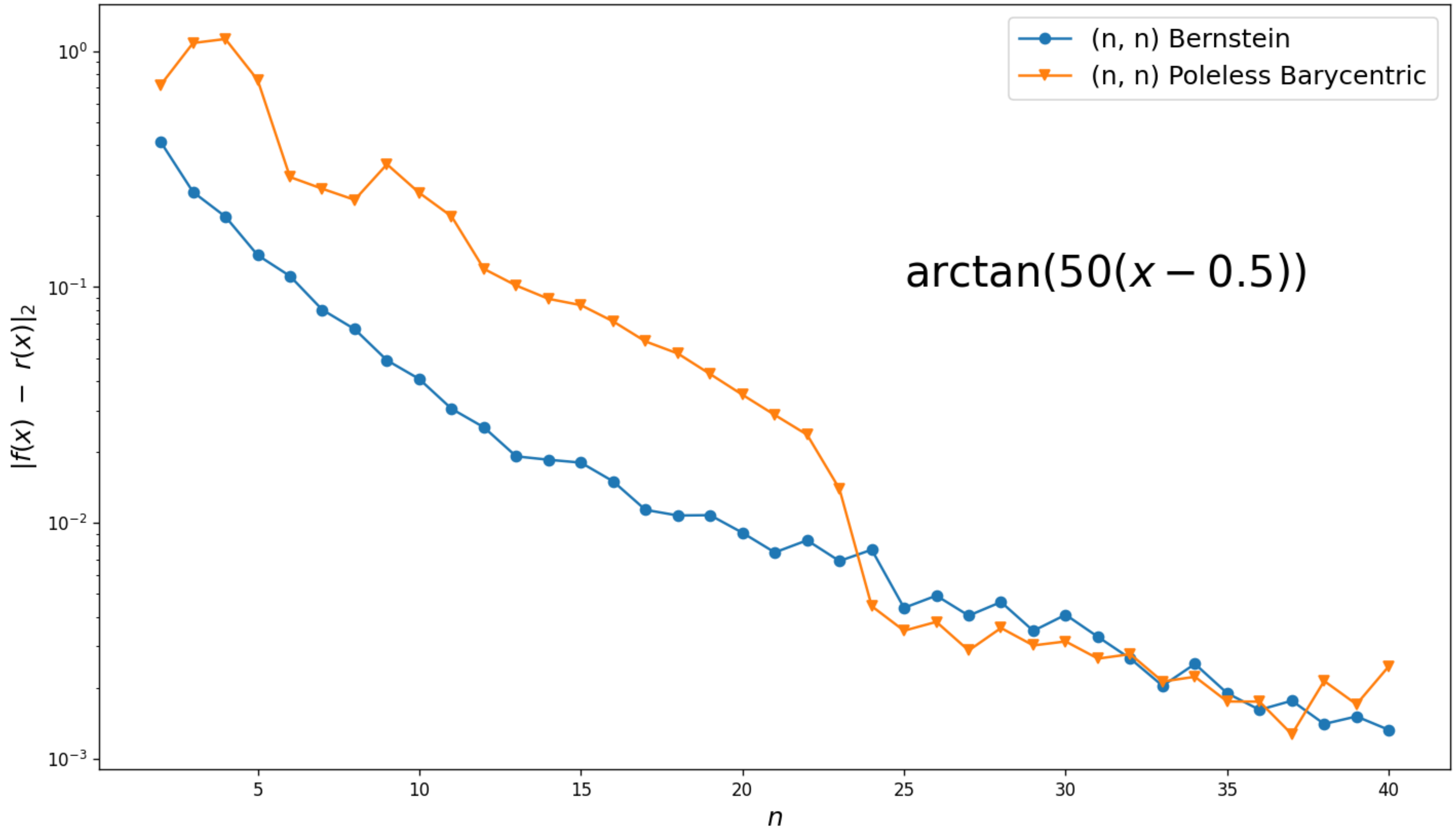
Poleless Barycentric

J. P. Berrut (1988) and M. S. Floater and K. Hormann (2007)



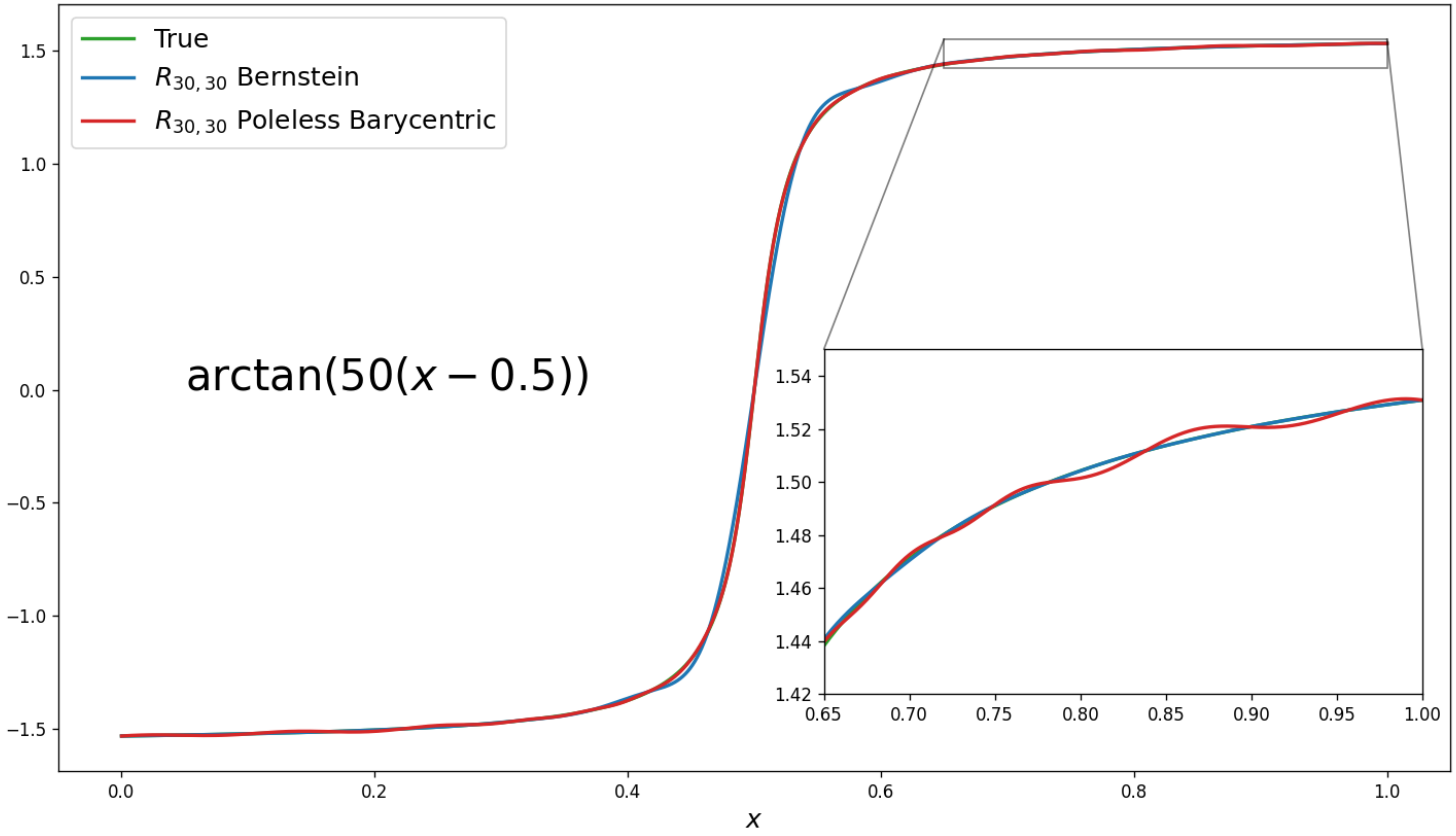
Poleless Barycentric

J. P. Berrut (1988) and M. S. Floater and K. Hormann (2007)



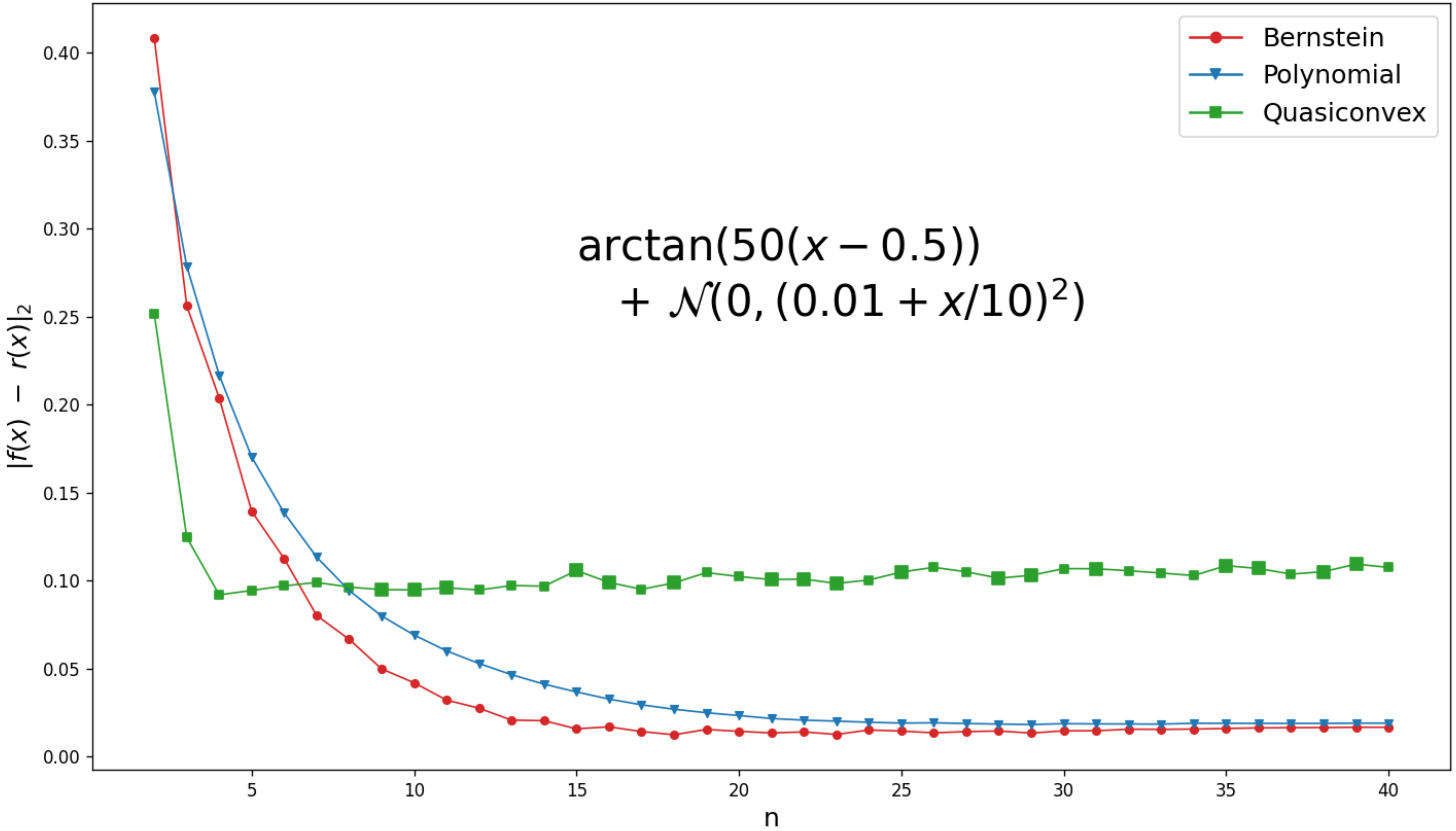
Poleless Barycentric

J. P. Berrut (1988) and M. S. Floater and K. Hormann (2007)



Quasiconvex

V. Peiris, N. Sharon, N. Sukhorukova, and J. Ugon (2021)



Quasiconvex

V. Peiris, N. Sharon, N. Sukhorukova, and J. Ugon (2021)