PADÉ APPROXIMATION

Given: formal power series (=: FPS)

$$f(z) := \sum_{k=0}^{\infty} \phi_k z^k \qquad \phi_k : \text{moments, Markov param., } \dots$$

Task: for every pair $(m,n) \in \mathbb{N}^2$ find (m,n) Padé approximant (=: PA) $r = r_{m,n} \in \mathcal{R}_{m,n}$ for which

(1) $$\boxed{f - r = \underline{\mathcal{O}(z^{\varkappa})}}$$ with $\varkappa$ as large as possible.

Taylor series of r at 0

FPS starting at $z^{\varkappa}$

[Padé, thesis, 1892]

There is a unique solution; "normally" $\varkappa = m+n+1$, which is equal to the number of free parameters.

To find r, consider the linearized problem:

Task: for every pair $(m,n) \in \mathbb{N}^2$ find (m,n) Padé forms (=: PFs) $(p,q) \in \mathcal{P}_m \times \mathcal{P}_n$ for which

(2) $$\boxed{fq - p = \mathcal{O}(z^{m+n+1})}$$

(2) has always nontrivial solutions, and they yield the unique rational function $p/q = r_{m,n}$.

I call $r_{m,n}$ a true Padé interpolant if $\varkappa \geq m+n+1$ in (1); otherwise I call $r_{m,n}$ deficient.

THM: The general solution of (2) is

(3) $$\boxed{(p,q) = (z^{\sigma}\hat{p}\,w,\ z^{\sigma}\hat{q}\,w)}$$

where

$\hat{p}, \hat{q}$ are fixed, relatively prime $\left(\Rightarrow r_{m,n} = \frac{\hat{p}}{\hat{q}}\right)$

$\sigma := \max\{0,\ m+n+1 - \underbrace{\text{order of } (f\hat{q}-\hat{p})}_{=\varkappa}\}$ deficiency

$\leq \delta := \min\{m-\partial\hat{p},\ n-\partial\hat{q}\}$ defect of $r_{m,n}$ in $\mathcal{R}_{m,n}$

$w \in \mathcal{P}_{\delta-\sigma}$ arbitrary

$r_{m,n}$ is a true interpolant iff $\sigma = 0$. [Gragg, SIAM Rev. 14, 1972]

COR. (CHARACTERIZATION THEOREM):

$r \in \mathcal{R}_{m,n}$ is the (m,n) PA of f iff

(4) $$\boxed{f - r = \mathcal{O}(z^{m+n+1-\delta})},$$

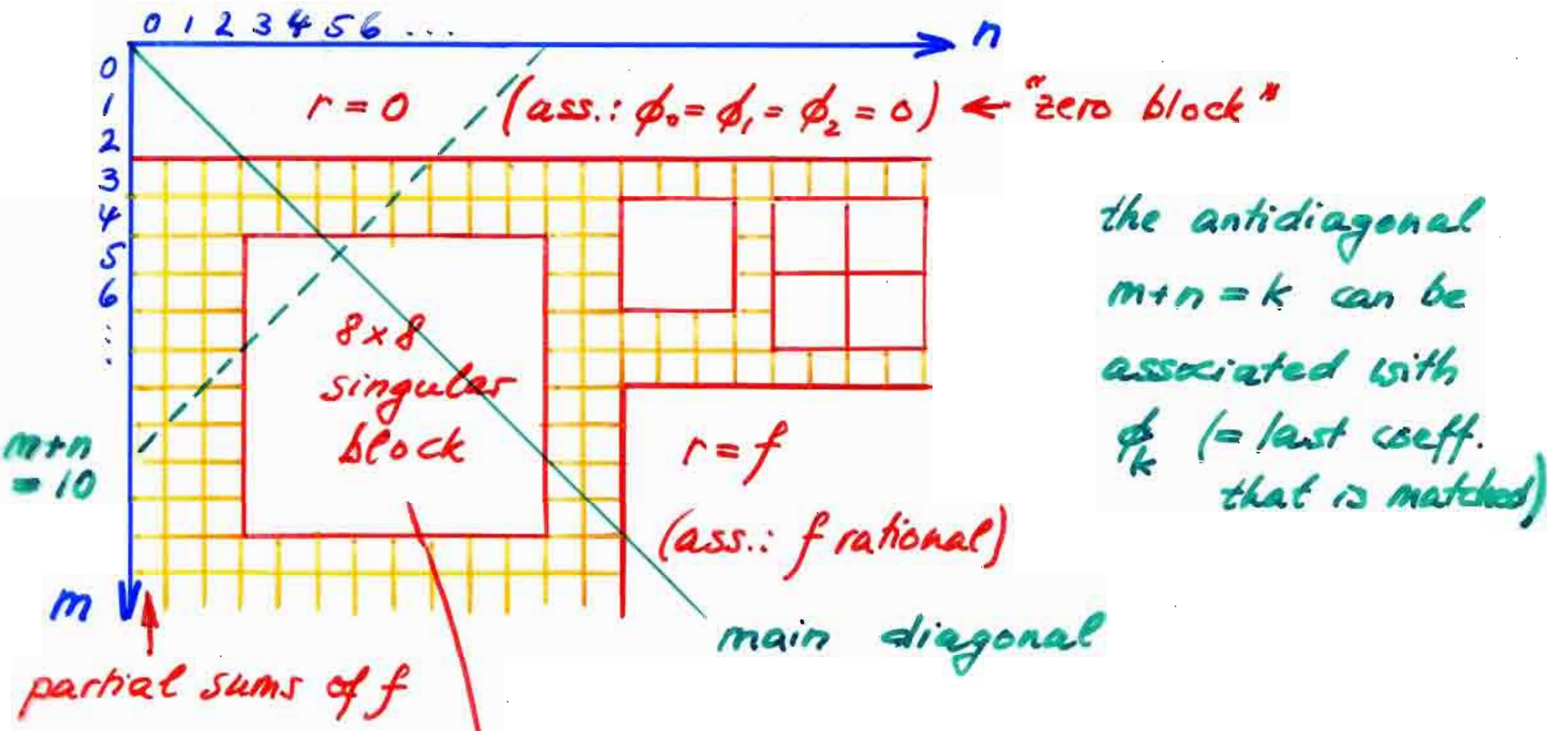
i.e., iff r "interpolates" $m+n+1-\delta$ Taylor coefficients of f.

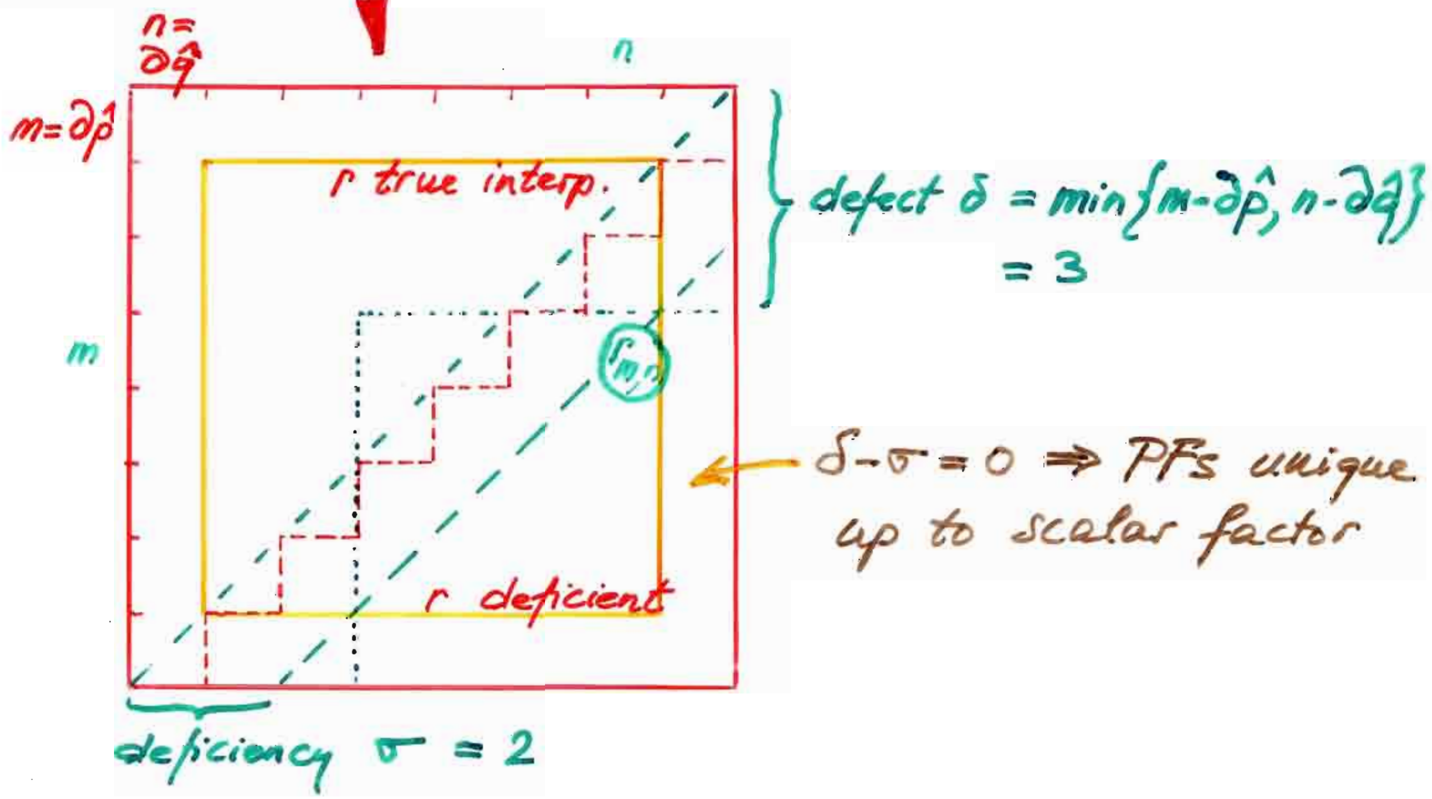
COR. (BLOCK STRUCTURE THEOREM)

The set of (m,n) pairs for which a particular nonzero rational function is a PA of f covers a (possibly empty) square in the (m,n)-plane.

If $r = f$, the square is infinite.

The Padé table: double-entry table whose (m,n) entry is the (m,n) Padé approximant $r_{m,n}$





The (m,n) PA is called **normal** if its block is 1×1.
The Padé table of f is called **normal** if all its entries are normal.

The condition (2), $fq - p = O(z^{m+n+1})$, translates into a linear system for the coefficients of p and q. It has a very special structure. Let

$$p(z) =: \sum_{k=0}^{m} \pi_k \, z^k , \qquad q(z) =: \sum_{k=0}^{n} \rho_k \, z^k ,$$

and, as before, $f(z) = \sum_{k=0}^{\infty} \phi_k z^k$. Compare coeffs of $z^{m+1}, \dots, z^{m+n}$:

$$\begin{bmatrix} \phi_{m+1} & \phi_m & \phi_{m-1} & \cdots & \phi_{m-n+1} \\ \phi_{m+2} & \phi_{m+1} & \phi_m & \cdots & \phi_{m-n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{m+n} & \phi_{m+n-1} & \phi_{m+n-2} & \cdots & \phi_m \end{bmatrix} \begin{bmatrix} \rho_0 \\ \rho_1 \\ \vdots \\ \rho_{n-1} \\ \rho_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$T_{m+1;n}$ $\qquad$ $T_{m;n}$

$n \times (n+1)$ Toeplitz system of rank $n-(\delta-\sigma)$

Unique solution with $\rho_{0,n} = 1$ (i.e., $q(0) = 1$) iff $T_{m;n}$ is nonsingular [in first row or column of block since $\delta = 0$ there, which implies $\sigma = 0$, $w \in \mathbb{P}_0$]

Unique solution with $\rho_{n,n} = 1$ (i.e., q monic) iff $T_{m+1;n}$ is nonsingular [in last row of block, since $\delta = \sigma$ and $\partial q = n$ there]

In the last column of a block there is a unique solution with q monic of degree $\partial \hat{q} + \sigma$, but this is of no use since $\partial \hat{q} + \sigma$ changes under perturbation of the problem (e.g., due to roundoff). ($\Rightarrow$ reformulation needed to make use of uniqueness in last column: solution with monic p is unique & stable)

If we compare *all* the coefficients in

(2) $$fq - p = O(z^{m+n+1}) =: z^{m+n+1} \underbrace{e}_{\text{"residual" (FPS)}}$$

$$\underset{f}{\begin{bmatrix} \phi_0 \\ \phi_1 & \phi_0 \\ \phi_2 & \phi_1 & \phi_0 \\ \vdots & \vdots & \vdots & \ddots \\ \phi_m & \phi_{m-1} & \phi_{m-2} & \cdots & \phi_{m-n} \\ \hline \phi_{m+1} & \phi_m & \phi_{m-1} & \cdots & \phi_{m-n+1} \\ \phi_{m+2} & \phi_{m+1} & \phi_m & \cdots & \phi_{m-n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{m+n} & \phi_{m+n-1} & \phi_{m+n-2} & \cdots & \phi_m \\ \hline \phi_{m+n+1} & \phi_{m+n} & \phi_{m+n-1} & \cdots & \phi_{m+1} \\ \phi_{m+n+2} & \phi_{m+n+1} & \phi_{m+n} & \cdots & \phi_{m+2} \\ \vdots & \vdots & \vdots & & \vdots \end{bmatrix}} \underset{q}{\begin{bmatrix} \rho_0 \\ \rho_1 \\ \rho_2 \\ \vdots \\ \rho_n \end{bmatrix}} = \begin{array}{c} \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \vdots \\ \pi_m \\ \hline 0 \\ 0 \\ \vdots \\ 0 \\ \hline \varepsilon_0 \\ \varepsilon_1 \\ \vdots \end{bmatrix} \end{array} \begin{array}{l} \leftarrow p \\ \\ \leftarrow e \end{array}$$

$T_{m;n}$

If m is fixed, but n varies, $T_{m;n}$ is the nth principal submatrix of the infinite Toeplitz matrix $T_{m;\infty}$ — rows of Padé table

If $m-n =: \ell$ if fixed, $M_{\ell+1;n} := J_n T_{m;n}$ with $J_n := \begin{bmatrix} 0 & & 1 \\ & ⋰ & \\ 1 & & 0 \end{bmatrix}$ — diagonal

is the nth principal submatrix of the infinite Hankel matrix $M_{\ell+1;\infty}$. ("moment matrices")

In both cases, a *nested* sequence of linear systems

Two nearly trivial generalizations

(1) Given formal Laurent series (=: FLS)

$$f(z) := \sum_{k=-\infty}^{\infty} \phi_k z^k$$

For $m \in \mathbb{Z}$, $n \in \mathbb{N}$ choose $L \leq m-n+1$ and split f up into

$$f(z) = z^{-L}\left(f^-(z) + f^+(z)\right) = \sum_{k=-\infty}^{L-1} \phi_k z^k + \sum_{k=L}^{\infty} \phi_k z^k$$

Let (p^+, q) be an $(m-L, n)$ PF of f^+, and $r^+ = p^+/q$ be the $(m-L, n)$ PA of f^+

DEF: $(p, q) := (z^{-L}[p^+ + qf^-], q)$ is a (one-sided) (m,n) Padé form of f.
$r := z^{-L} f^- + r^+$ is a (one-sided) (m,n) Padé approximant

It can be shown that (p,q) and r are independant of L $\Rightarrow$ we could always choose $L = m-n+1$, so that $m-L = n-1$.

As before holds

$$fq - p = O(z^{m+n+1}), \quad f - r = O(z^{\varkappa})$$

$\varkappa$ maximum

The Padé table of f covers now the right half-plane. The block structure is as before.

See Trefethen/G [1987], Bultheel [1987].

② Given: two formal power series

$$f(z) := \sum_{k=0}^{\infty} \phi_k z^k, \qquad g(z) := \sum_{k=0}^{\infty} \gamma_k z^k, \quad \gamma_0 \neq 0.$$

DEF: (m,n) Padé form of (g,f): $(p,q) \in \mathcal{P}_m \times \mathcal{P}_n$ satisfying

(2') $$\boxed{gp + fq = \mathcal{O}(z^{m+n+1})}$$

$$\Big\Updownarrow \quad h := -f/g \quad \text{FPS}$$

(2) $$hq - p = \mathcal{O}(z^{m+n+1})$$

Hence, (2') is equivalent to (2).
However, comparing coefficients in (2') yields another linear system, equivalent to the one for (2).

May write (2') as

$$[g \;\; f] \begin{bmatrix} p \\ q \end{bmatrix} = z^{m+n+1} \mathbf{e}$$

$\mathbf{e}$: residual (FPS)
not the same as before