The Bernstein polynomial basis: a centennial retrospective

a “sociological study” in the
evolution of mathematical ideas

Rida T. Farouki

Department of Mechanical & Aerospace Engineering,
University of California, Davis

— synopsis —

- 1912: Sergei Natanovich Bernstein — constructive proof of Weierstrass approximation theorem
- 1960s: Paul de Faget de Casteljau, Pierre Étienne Bézier and the origins of computer-aided geometric design
- elucidation of Bernstein basis properties and algorithms
- 1980s: intrinsic numerical stability of the Bernstein form
- algorithms & representations for computer-aided design
- diversification of applications in scientific computing

Weierstrass approximation theorem

Given any continuous function $f(x)$ on an interval $[\,a,b\,]$ and a tolerance $\epsilon > 0$, a polynomial $p_n(x)$ of sufficiently high degree n exists, such that

$$|f(x) - p_n(x)| \,\le\, \epsilon \qquad \text{for} \;\; x \in [\,a,b\,]\,.$$

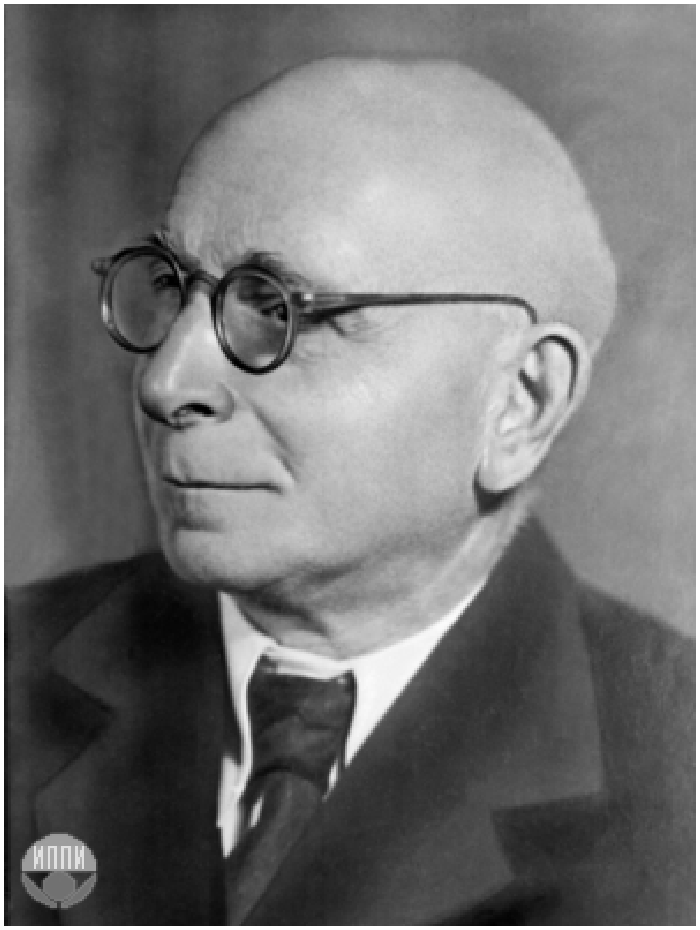
Polynomials can *uniformly approximate* any continuous $f(x)$, $x \in [\,a,b\,]$.

Original (1885) proof by Weierstrass is “existential” in nature — begins by expressing $f(x)$ as a convolution

$$f(x) \,=\, \lim_{k\to 0} \frac{1}{\sqrt{\pi}k} \int_{-\infty}^{+\infty} f(t)\, \exp\!\left[-\frac{(t-x)^2}{k^2}\right] \mathrm{d}t$$

with a Dirac delta function, and relies heavily on analytic limit arguments.

Sergei Natanovich Bernstein (1880-1968)



(photo: Russian Academy of Sciences)

academic career of S. N. Bernstein

- 1904: Sorbonne PhD thesis, on analytic nature of PDE solutions (worked with Hilbert at Göttingen during 1902-03 academic year)
- 1913: Kharkov PhD thesis (polynomial approximation of functions)
- 1912: *Comm. Math. Soc. Kharkov* paper (2 pages): constructive proof of Weierstrass theorem — introduction of Bernstein basis
- 1920-32: Professor in Kharkov & Director of Mathematical Institute political purge: moved to USSR Academy of Sciences (Leningrad)
- 1941-44: escapes to Kazakhstan during the siege of Leningrad
- 1944-57: Steklov Math. Institute, Russian Academy of Sciences, Moscow — edited complete works of Chebyshev (died 1968)
- collected works of Bernstein published in 4 volumes, 1952-64

Bernstein's proof of Weierstrass theorem

Russian school of approximation theory, founded by Chebyshev, favors *constructive approximation methods* over "existential" proofs

given $f(t)$ continuous on $t \in [\,0,1\,]$ define

$$p_n(t) \;=\; \sum_{k=0}^{n} f(k/n)\, b_k^n(t)\,, \quad b_k^n(t) \;=\; \binom{n}{k}(1-t)^{n-k}t^k$$

$p_n(t) =$ *convex combination* of sampled values $f(0), f(\frac{1}{n}), \ldots, f(1)$

$$|\, f(t) - p_n(t)\,| \;=\; O(\tfrac{1}{n}) \qquad \text{for } \; t \in [\,0,1\,]$$

$\Longrightarrow$ $p_n(t)$ *converges uniformly* to $f(t)$ as $n \to \infty$

derivatives of $p_n(t)$ also converge to those of $f(t)$ as $n \to \infty$

Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités.

Je me propose d'indiquer une démonstration fort simple du théorème suivant de Weierstrass:

Si $F(x)$ est une fonction continue quelconque dans l'intervalle 01, il est toujours possible, quel que petit que soit ε, de déterminer un polynome $E_n(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n$ de degré n assez élevé, tel qu' on ait

$$|F(x) - E_n(x)| < \varepsilon$$

en tout point de l'intervalle considéré.

A cet effet, je considère un évenement A, dont la probabilité est égale à x. Supposons qu'on effectue n expériences et que l'on convienne de payer à un joueur la somme $F\left(\frac{m}{n}\right)$, si l'évenement A se produit m fois. Dans ces conditions, l'espérance mathématique E_n du joueur aura pour valeur

$$E_n = \sum_{m=0}^{m=n} F\left(\frac{m}{n}\right) . C_n^m x^m (1-x)^{n-m}. \qquad (1)$$

Or, il résulte de la continuité de la fonction $F(x)$ qu'il est possible de fixer un nombre δ, tel que l'inégalité

$$|x - x_0| \leqq \delta$$

entraine

$$\left|F(x) - F(x_0)\right| < \frac{\varepsilon}{2};$$

de sorte que, si $\overline{F}(x)$ désigne le maximum et $\underline{F}(x)$ le minimum de $F(x)$ dans l'intervalle $(x - \delta,\ x + \delta)$, on a

$$\overline{F}(x) - F(x) < \frac{\varepsilon}{2},\quad F(x) - \underline{F}(x) < \frac{\varepsilon}{2}. \qquad (2)$$

S. N. Bernstein, Comm. Kharkov Math. Soc. (1912)

connection with probability theory

basis function $b_k^n(t) = \binom{n}{k}(1-t)^{n-k}t^k$

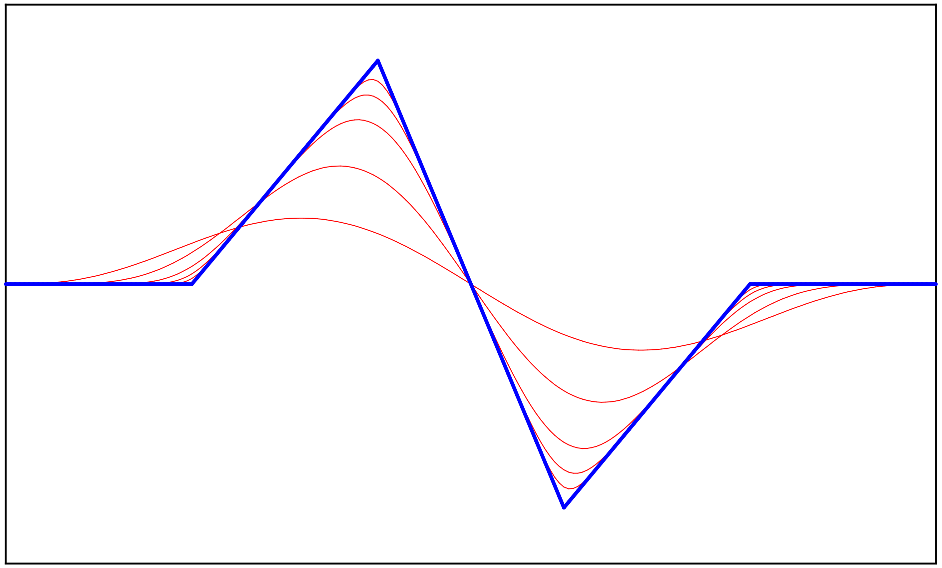
probability of k successes in n trials of random process

with individual probability of success t in each trial

$\longrightarrow$ binomial probability distribution

non-negativity & partition-of-unity properties of $b_k^n(t)$

slow convergence of Bernstein approximations



Bernstein polynomial approximations of degree
$n = 10, 30, 100, 300, 1000$ to a "triangular wave"

This fact seems to have precluded any numerical application of Bernstein polynomials from having been made. Perhaps they will find application when the properties of the approximant in the large are of more importance than the closeness of the approximation.

Philip J. Davis, *Interpolation and Approximation* (1963)

Paul de Casteljau & Pierre Bézier (1960s)

an emerging application — *computer-aided design*

- Paul de Faget de Casteljau — theory of “courbes et surfaces à pôles” developed at André Citroën, SA in the early 1960s

- de Casteljau’s work unpublised (regarded as proprietary by Citroën) — revealed to outside world by Wolfgang Böhm in mid-1980s

- Pierre Étienne Bézier — implemented methods for computer-aided design and manufacturing at Renault during 1960s and 1970s

- Bézier published numerous articles and books describing his ideas

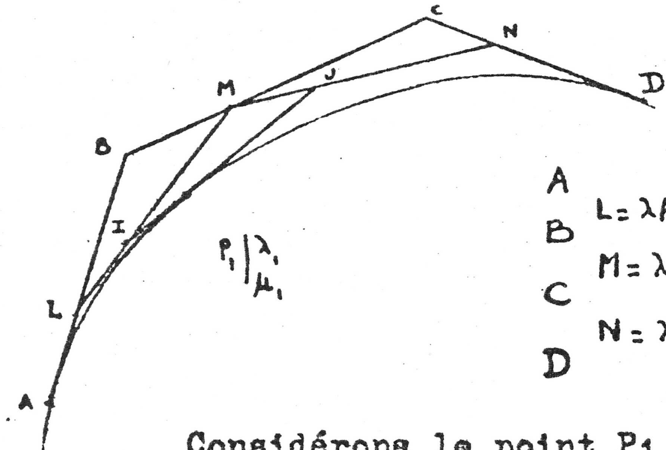
- basic problem: provide intuitive & interactive means for design and manipulation of “free–form” curves and surfaces by computer, in the automotive, aerospace, and related industries

- identification of de Casteljau’s and Bézier’s ideas with Bernstein form of polynomials came later, through work of Forrest, Riesenfeld, et al.

de Casteljau “Courbes et surfaces à pôles” (Société Anonyme André Citroën, 1963)

1.5.- Sous-Pôles d'une courbe

1.5.1.- Définition des sous-pôles. Considérons une cubique des pôles A, B, C, D. Nous avons vu précédemment que la construction par la méthode des barycentres donnent les différents points L, M, N, I, J, P.



A

B

C

D

$L = \lambda A + \mu B$

$M = \lambda B + \mu C$

$N = \lambda C + \mu D$

$I = \lambda^2 A + 2\lambda\mu B + \mu^2 C$

$J = \lambda^2 B + 2\lambda\mu C + \mu^2 D$

$P = \lambda^3 A + 3\lambda^2\mu B + 3\lambda\mu^2 C + \mu^3 D$

Considérons le point P_1 de paramètres λ_1 et μ_1 (avec $\lambda_1 + \mu_1 = 1$) (λ_1 varie de 0 à 1 et μ_1 1 à 0, lorsque P va en D).

Cherchons les pôles de la cubique P_1 D. Cette courbe dérive de la cubique initiale AD par changement des paramètres.

pôles = “pilot points” (inter*pol*ation of *pol*ynomials with *pol*ar forms)

de Casteljau — barycentric coordinates

de Casteljau's λ and μ = interval *barycentric coordinates*, with $\lambda + \mu = 1$

example — for $t \in [\,0,1\,]$ take $\lambda = 1 - t$ and $\mu = t$, and expand $(\lambda + \mu)^n$

$$1 \;=\; [\,(1-t)+t\,]^n \;=\; \sum_{k=0}^{n} \binom{n}{k} (1-t)^{n-k} t^k \;=\; \sum_{k=0}^{n} b_k^n(t)$$

$\Rightarrow$ Bernstein basis $\{b_k^n(t)\}$ is *non–negative* and forms a *partition of unity*

de Casteljau also considers extension to barycentric coordinates and multivariate polynomial bases on *triangular and simplex domains*

computer-aided design in the early 60s

. . . the designers were astonished and scandalized. Was it some kind of joke? It was considered nonsense to represent a car body mathematically. It was enough to please the eye, the word accuracy had no meaning . . .

reaction at Citroën to de Casteljau's ideas

Citroën's first attempts at digital shape representation used a Burroughs E101 computer featuring 128 program steps, a 220-word memory, and a 5 kW power consumption!

De Casteljau's "insane" persistence led to an increased adoption of computer-aided design methods in Citroën from 1963 onwards.

My stay at Citroën was not only an adventure for me,
but also an adventure for Citroën! . . . . P. de Casteljau

correspondence with de Casteljau (1991)

à Vélizy le 6 Septembre 1991

Cher Monsieur Farouki

L'Optique géométrique fournit, à peu près, les seuls exemples de Géométrie métrique, grace au principe de Fermat, et c'est pourquoi je l'aime : Aussi je vous remercie de votre envoi, auquel j'ai accordé le maximum d'attention. Il est regrettable qu'entre nous, il y ait cette barrière de langue et je suis infiniment désolé de ne pas pouvoir correspondre en anglais. J'espère aussi avoir donné satisfaction à votre collègue, J.C Chastang, bien que je sois tout le contraire d'un rat de bibliothèque.

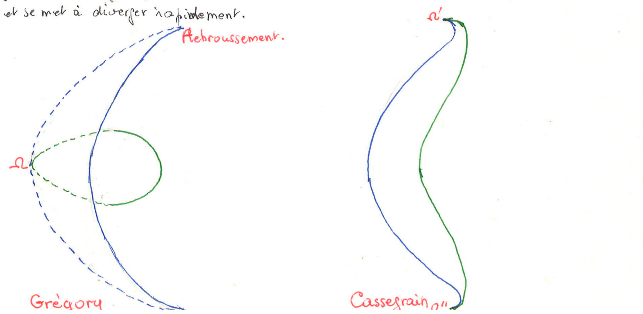
Dans sa lettre, votre collègue Mr Chastang semble seulement s'intéresser aux problèmes limités à un seul point source. Personnellement j'aurais une préférence pour le doublet ou mieux un petit cercle axial.

Voici un magnifique exemple, que j'appele le Ricochet de ce que l'on peut faire avec deux doublets, ou mieux deux cercles d'Airy. On obtient ainsi une solution d'équations aux différences finies, incroyablement précise : Par renvois successifs, on génère deux surfaces conjuguées d'un système aplanétique. Je serais prêt à croire que cette solution "approchée" est meilleure que la solution "exacte" dans le cas limite de la tache de diffraction d'Airy, puisque précisément calculée pour cela. (Voir à ce sujet, aussi, la Remarque de la page 118 de mon livre "Lissage")

Je possède tout un dossier sur la question : Réflexion, Réfraction, Courbe algébrique d'interpolation entre deux points, et encore propriété de l'intersection des tangentes en A_n et B_n des courbes conjuguées (points correspondants) qui joue un rôle vis à vis des centres de courbure.

On peut opposer cette récurrence à celle de la théorie des fractales : il faut en effet rechercher la stabilité des longueurs vectorielles $\overrightarrow{B_{2n-1}B_{2n+1}}$ ou $\overrightarrow{A_{2n}A_{2n+2}}$, en évitant toute convergence (ou à l'envers la divergence). Comme contre exemple, le principe d'Herschell appliqué à des segments axiaux M'M" et P'P" conduit à une rapide divergence.

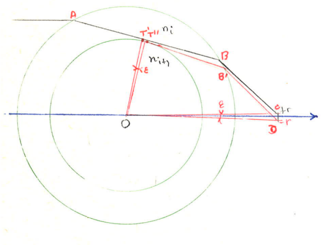
La forme mathématique est plus étendue que la partie exploitable physiquement. Les combinaisons de type Grégory, divergentes donnent des formes infiniment plus "jolies" que leur équivalent Cassegrain, convergentes. On termine dans les deux cas sur une singularité de degré élevé (Ω). On peut poursuivre la récurrence au delà de Ω, mais elle ne signifie plus rien et se met à diverger rapidement.



Dans mon livre "Lissage" je donne d'autres exemples, non tirés de l'Optique, de la Génération par différences finies, mais ils restent rares.

Il faudrait aussi montrer comment un cercle de grand rayon, connu que sur une petite portion doit s'exprimer sous la présentation surface d'onde locale, qui fournit un calcul "approché" bien plus rigoureux que la solution "exacte." Le passage du point à l'infini s'effectue alors en douceur. Je ne pense pas vous apprendre quelque chose.

Il existe un autre problème, tiré de l'optique qui utilise ce genre de principe : On suppose une lentille sphérique, réalisée en couches successives à la manière des peaux d'un oignon. On impose une tache focale de rayon donnée r, assez petite vue du centre O de la boule sous un angle ε



Dès que le rayon atteint la valeur +r, une nouvelle couche juste sous incidence rasante va le renvoyer en -r, ce qui impose :

$$\frac{n_{i+1}}{n_i} = \frac{1}{\cos\frac{\varepsilon}{2}} = C^{te}$$

puisque la nouvelle couche va être abordée au niveau de l'angle limite.

Ceci impose une progression géométrique des indices

$$n_i = n_0 \left(\frac{1}{\cos\frac{\varepsilon}{2}}\right)^i$$

La suite n'est plus qu'une question de calculs, pour déterminer de proche en proche les rayons successifs des couches

Là encore il suffit de limiter r, ou ε à la tache d'Airy.

Evidemment le système obeït à la loi de Bouguer $n r \sin i = C^{te}$ le long d'un rayon. à ce propos avez vous remarqué que $\sin i = 1$ $nr = C^{te}$. Si cette condition est réalisée dans un milieu à symétrie cylindrique le rayon devient circulaire : les surfaces d'ondes sont des plans passant par l'axe ! Cela semble en contradiction avec le principe de Fermat

s'il existe un gradient d'indice radial tel que $nr = C^{te}$, le rayon qui se propage à $r = C^{t}$ donc $n = C^{t}$, dans un milieu d'indice constant est circulaire !!!

Si vous effectuiez des développements sur l'une de ces questions, cela m'intéresserait d'être tenu au courant.

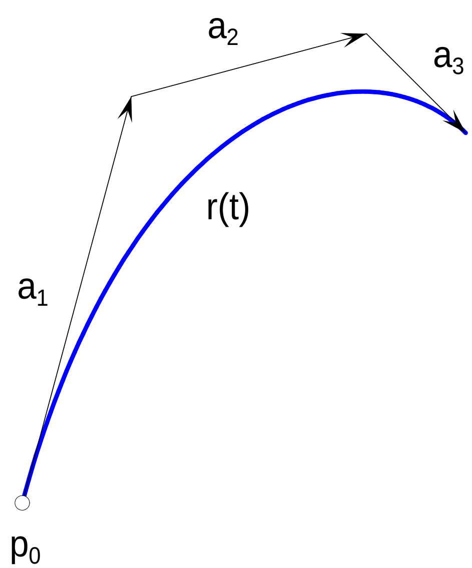
Je vous prie d'agréer, cher Monsieur Farouki l'expression de mes meilleurs sentiments

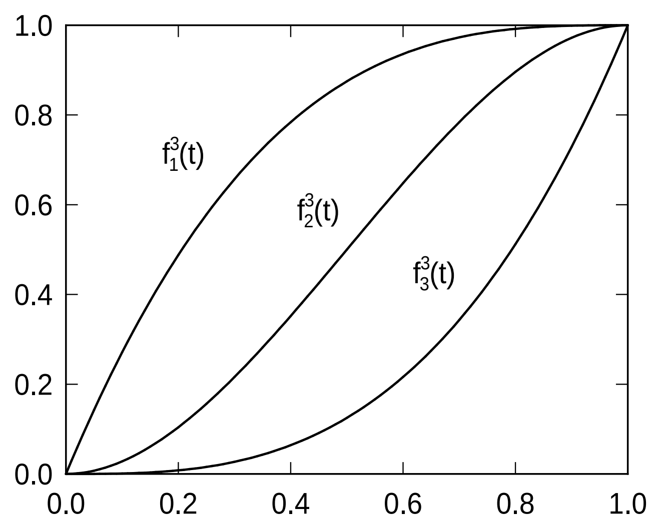
[signature]

Bézier's "point-vector" form of a polynomial curve

specify degree-n curve by initial point $\mathbf{p}_0$ and n vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$

$$\mathbf{r}(t) \;=\; \mathbf{p}_0 \,+\, \sum_{k=1}^{n} \mathbf{a}_k f_k^n(t)\,, \qquad f_k^n(t) \;=\; \frac{(-1)^k}{(k-1)!}\, t^k \, \frac{\mathrm{d}^{k-1}}{\mathrm{d}t^{k-1}} \, \frac{(1-t)^n - 1}{t}$$



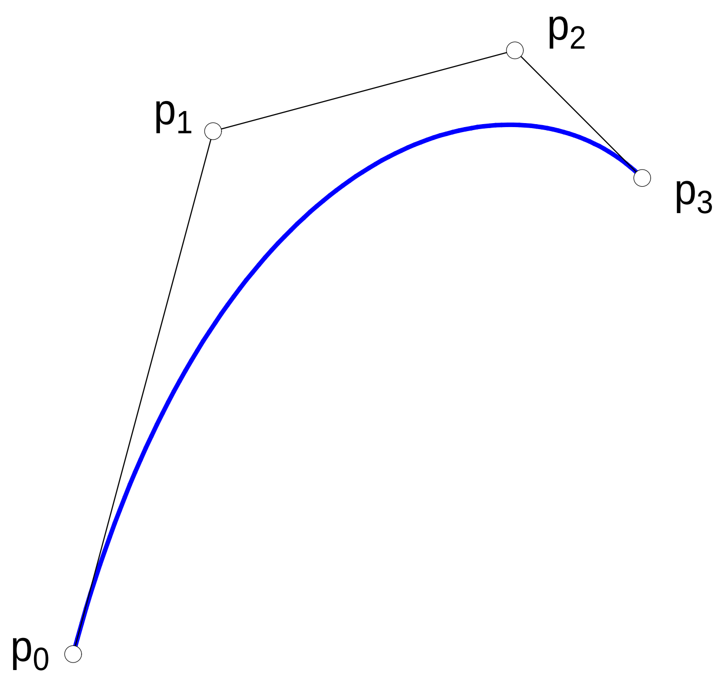


Left: Bézier point-vector specification of a cubic curve. Right: cubic basis functions $f_1^3(t)$, $f_2^3(t)$, $f_3^3(t)$ associated with the vectors $\mathbf{a}_1$ $\mathbf{a}_2$, $\mathbf{a}_3$.

mischievous Bézier — $f_1^n(t), \ldots, f_n^n(t) =$ basis of Onésime Durand !

control-point form of a Bézier curve

Forrest (1972) : $f_i^n(t) \;=\; \sum_{k=i}^{n} b_k^n(t)\,, \quad b_k^n(t) \;=\; \binom{n}{k}(1-t)^{n-k}t^k$

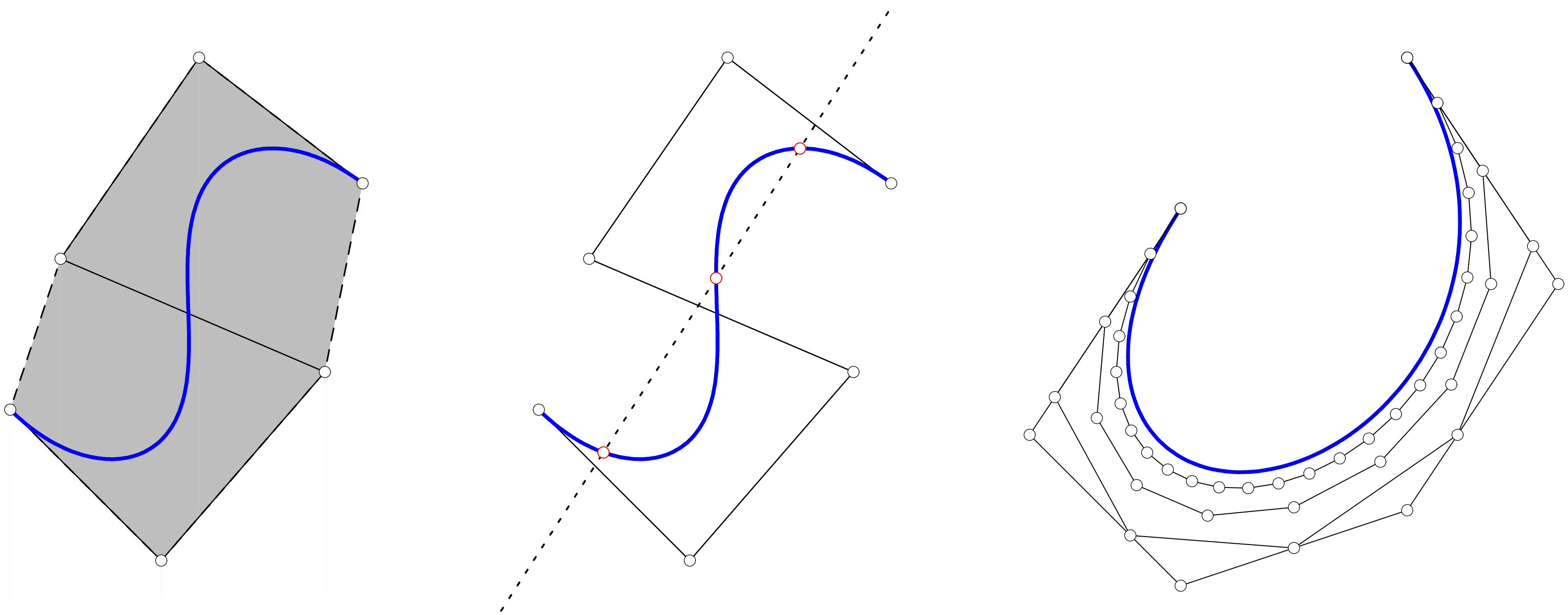


re-write as $\mathbf{r}(t) \;=\; \sum_{k=0}^{n} \mathbf{p}_k\, b_k^n(t)\,, \quad \mathbf{p}_k = \mathbf{p}_{k-1} + \mathbf{a}_k$

manipulate cure shape by moving control points $\mathbf{p}_0, \ldots, \mathbf{p}_n$

convex-hull, variation-diminishing, degree-elevation properties of the Bézier form

$$\mathbf{r}(t) \;=\; \sum_{k=0}^{n} \mathbf{p}_k \, b_k^n(t)\,, \quad \text{control points} \quad \mathbf{p}_0, \ldots, \mathbf{p}_n$$



de Casteljau algorithm — evaluates & subdivides $\mathbf{r}(t)$

initialize — set $t = \tau$ and $\mathbf{p}_k^0 = \mathbf{p}_k$ for $k = 0, \ldots, n$

```
for r = 1,...,n
    for j = r,...,n
```

$$\{ \; \mathbf{p}_j^r \; = \; (1 - \tau)\, \mathbf{p}_{j-1}^{r-1} \; + \; \tau\, \mathbf{p}_j^{r-1} \; \}$$

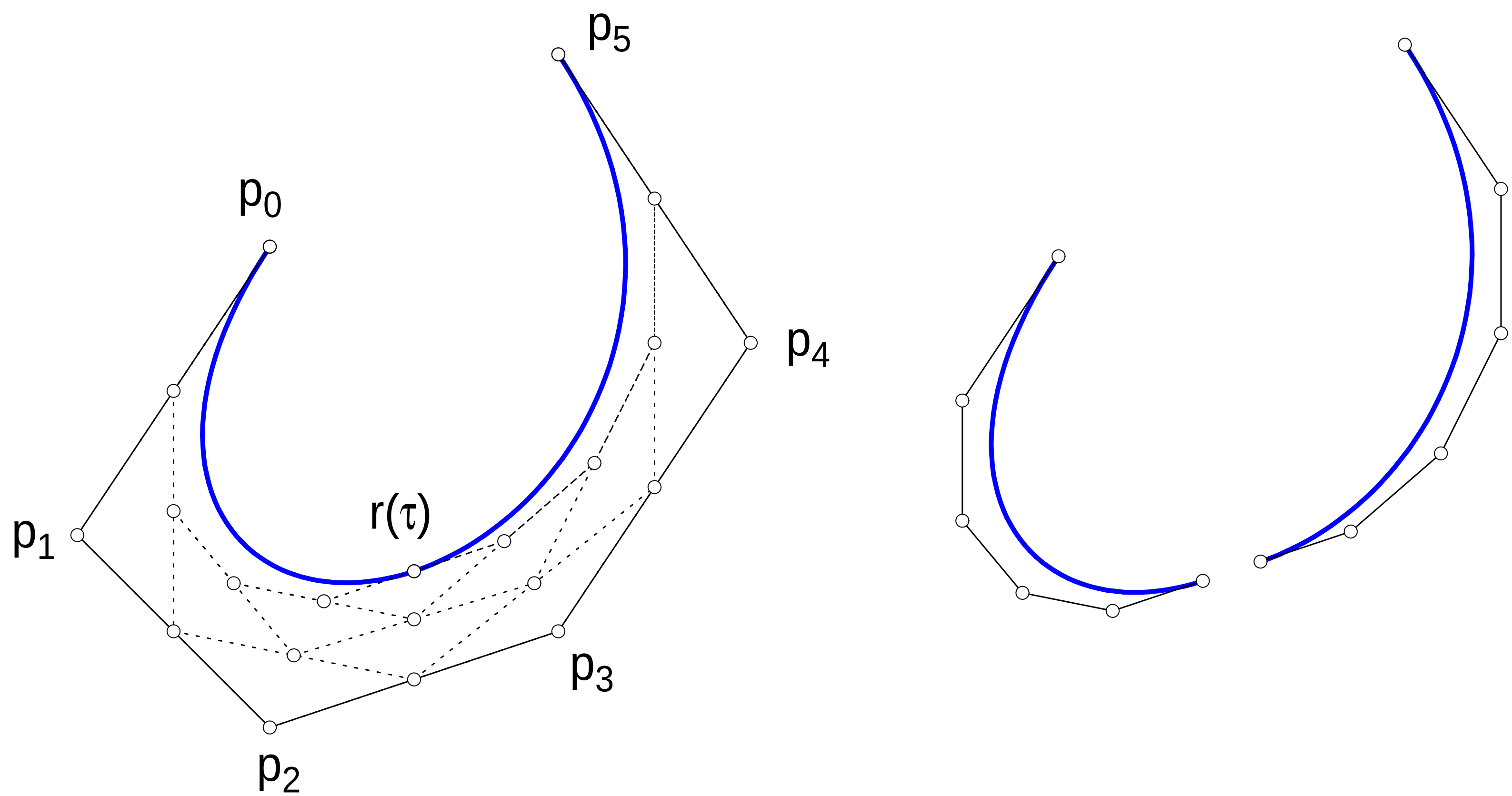
generates a triangular array of points $\{ \, \mathbf{p}_j^r \, \}$

$$\begin{array}{ccccccccccc}
\mathbf{p}_0^0 & & \mathbf{p}_1^0 & & \mathbf{p}_2^0 & & \cdot & & \cdot & & \mathbf{p}_n^0 \\
& \mathbf{p}_1^1 & & \mathbf{p}_2^1 & & \cdot & & \cdot & & \mathbf{p}_n^1 & \\
& & \mathbf{p}_2^2 & & \cdot & & \cdot & & \mathbf{p}_n^2 & & \\
& & & \cdot & & \cdot & & \cdot & & & \\
& & & & & \mathbf{p}_n^n & & & & &
\end{array}$$

$$\mathbf{p}_n^n \;=\; \text{evaluated point } \mathbf{r}(\tau) \text{ on curve}$$

$$\mathbf{p}_0^0, \mathbf{p}_1^1, \ldots, \mathbf{p}_{n-1}^{n-1}, \mathbf{p}_n^n \;=\; \text{control points for subsegment } t \in [\,0, \tau\,] \text{ of } \mathbf{r}(t)$$

$$\mathbf{p}_n^n, \mathbf{p}_n^{n-1}, \ldots, \mathbf{p}_n^1, \mathbf{p}_n^0 \;=\; \text{control points for subsegment } t \in [\,\tau, 1\,] \text{ of } \mathbf{r}(t)$$



interlude . . . "lost in translation"

warning sign on bathroom door in Beijing hotel

“English on vacation”

in a Bucharest hotel lobby —
The elevator is being fixed for the next day.
During that time we regret that you will be unbearable.

in a Paris hotel elevator —
Please leave your values at the front desk.

in a Zurich hotel —
Because of the impropriety of entertaining guests of the opposite sex in your bedroom, it is suggested that the lobby be used for this purpose.

in an Acapulco restaurant —
The manager has personally passed all the water served here.

in Germany’s Schwarzwald —
It is strictly forbidden on our Black Forest camping site that people of different sex — for instance, men and women — live together in one tent unless they are married with each other for that purpose.

in an Athens hotel —
Guests are expected to complain at the office between 9 and 11 am daily.

instructions for AC in Japanese hotel —
If you want just condition of warm in your room, please control yourself.

in a Yugoslav hotel —
The flattening of underwear with pleasure is the job of the chambermaid.

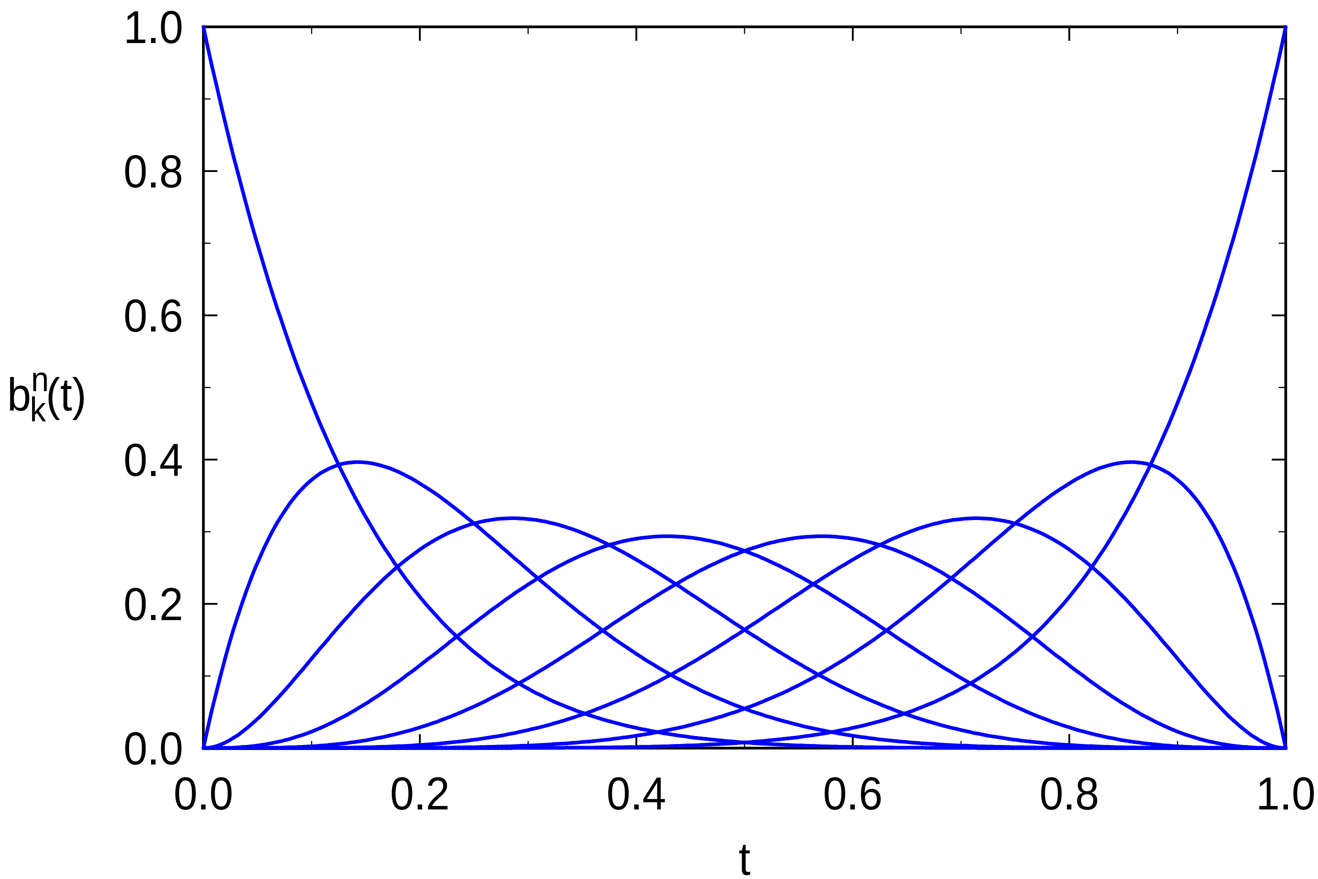
in a Japanese hotel —
You are invited to take advantage of the chambermaid.

on the menu of a Swiss restaurant —
Our wines leave you nothing to hope for.

in a Bangkok dry cleaners —
Drop your trousers here for best results.

Japanese rental car instructions —
When passenger of foot heave in sight, tootle the horn.
Trumpet him melodiously at first, but if he still obstacles
your passage, then tootle him with vigor.

Bernstein basis functions



$$b_k^n(t) = \binom{n}{k}(1-t)^{n-k}t^k$$

roots of multiplicity k and $n-k$ at $t=0$ and $t=1$

properties of the Bernstein basis

$$b_k^n(t) = \binom{n}{k}(1-t)^{n-k}t^k\,, \quad k = 0,\ldots,n$$

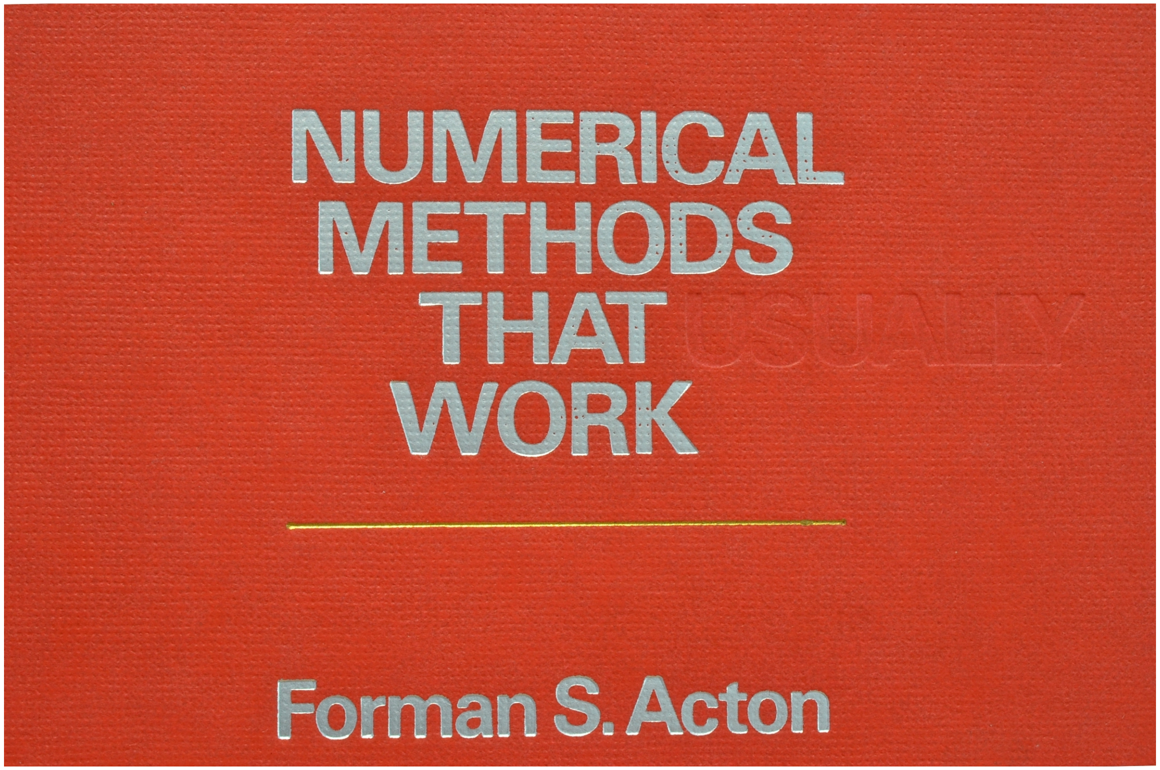
- partition of unity : $\sum_{k=0}^{n} b_k^n(t) \equiv 1$
- non-negativity : $b_k^n(t) \geq 0$ for $t \in [\,0,1\,]$
- symmetry : $b_k^n(t) = b_{n-k}^n(1-t)$
- recursion : $b_k^{n+1}(t) = t\,b_{k-1}^n(t) + (1-t)\,b_k^n(t)$
- unimodality : $b_k^n(t)$ has maximum at $t = k/n$

properties of Bernstein form, $p(t) = \sum_{k=0}^{n} c_k \, b_k^n(t)$

- end-point values : $p(0) = c_0$ and $p(1) = c_n$
- lower & upper bounds : $\min_k c_k \le p(t) \le \max_k c_k$
- variation diminishing : $\#$ roots $=$ signvar$(c_0, \ldots, c_n) - 2m$
- derivatives & integrals : coefficients of $p'(t)$ & $\int p(t)\,\mathrm{d}t$ $=$ differences & partial sums of $c_0, \ldots, c_n$
- recursive algorithms for subdivision, degree elevation, arithmetic operations, composition, resultants, etc.
- root isolation (subdivision & variation-diminishing property)

the plague of numerical instability

. . . or, the temptation to “*kick the computer*”



Do you ever want to kick the computer? Does it iterate endlessly on your newest algorithm that should have converged in three iterations? And does it finally come to a crashing halt with the insulting message that you divided by zero? These minor trauma are, in fact, the ways the computer manages to kick you and, unfortunately, you almost always deserve it! For it is a sad fact that most of us can more readily compute than think . . .

numerical stability of polynomials

$p(t)$ has coefficients $c_0, \ldots, c_n$ in basis $\Phi = \{\phi_0(t), \ldots, \phi_n(t)\}$

$$p(t) \;=\; \sum_{k=0}^{n} c_k \phi_k(t)$$

how sensitive is a value or root of $p(t)$ to perturbations of maximum relative magnitude ϵ in the coefficients $c_0, \ldots, c_n$?

condition number for value of $p(t)$:

$$|\delta p(t)| \;\le\; C_\Phi(p(t))\,\epsilon\,, \quad C_\Phi(p(t)) \;=\; \sum_{k=0}^{n} |c_k \phi_k(t)|$$

condition number for root τ of $p(t)$:

$$|\delta\tau| \;\le\; C_\Phi(\tau)\,\epsilon\,, \quad C_\Phi(\tau) \;=\; \frac{1}{|p'(\tau)|} \sum_{k=0}^{n} |c_k \phi_k(t)|$$

condition numbers for power and Bernstein forms

$$p(t) = \sum_{k=0}^{n} a_k t^k = \sum_{k=0}^{n} c_k b_k^n(t)$$

$$c_j = \sum_{k=0}^{j} \frac{\binom{j}{k}}{\binom{n}{k}} a_k, \qquad t^k = \sum_{j=k}^{n} \frac{\binom{j}{k}}{\binom{n}{k}} b_j^n(t)$$

Theorem. $C_B(p(t)) \leq C_P(p(t))$ for any polynomial $p(t)$ and all $t \in [\,0, 1\,]$.

Proof (triangle inequality).

$$\begin{aligned} C_B(p(t)) \;&=\; \sum_{j=0}^{n} |c_j b_j^n(t)| = \sum_{j=0}^{n} \left| \sum_{k=0}^{j} \frac{\binom{j}{k}}{\binom{n}{k}} a_k \right| b_j^n(t) \\ &\leq\; \sum_{k=0}^{n} |a_k| \sum_{j=k}^{n} \frac{\binom{j}{k}}{\binom{n}{k}} b_j^n(t) \;=\; \sum_{k=0}^{n} |a_k t^k| \;=\; C_P(p(t))\,. \end{aligned}$$

Wilkinson's "perfidious" polynomial

problem: compute the roots of the degree 20 polynomial

$$p(t) = (t-1)(t-2)\cdots(t-20) = \sum_{k=0}^{20} a_k t^k$$

using (software) floating-point arithmetic

J. H. Wilkinson (1959), The evaluation of the zeros of ill-conditioned polynomials, Parts I & II, *Numerische Mathematik* **1**, 150-166 & 167-180.

"*The cosy relationship that mathematicians enjoyed with polynomials suffered a severe setback in the early fifties when electronic computers came into general use. Speaking for myself, I regard it as the most traumatic experience in my career as a numerical analyst.*"

J. H. Wilkinson, The Perfidious Polynomial,
in *Studies in Numerical Analysis* (1984)

root condition numbers for Wilkinson polynomial

root	power basis	Bernstein basis
0.05	2.10×10^{1}	3.41×10^{0}
0.10	4.39×10^{3}	1.45×10^{2}
0.15	3.03×10^{5}	2.34×10^{3}
0.20	1.03×10^{7}	2.03×10^{4}
0.25	2.06×10^{8}	1.11×10^{5}
0.30	2.68×10^{9}	4.15×10^{5}
0.35	2.41×10^{10}	1.12×10^{6}
0.40	1.57×10^{11}	2.22×10^{6}
0.45	7.57×10^{11}	3.32×10^{6}
0.50	2.78×10^{12}	3.80×10^{6}
0.55	7.82×10^{12}	3.32×10^{6}
0.60	1.71×10^{13}	2.22×10^{6}
0.65	2.89×10^{13}	1.12×10^{6}
0.70	3.78×10^{13}	4.15×10^{5}
0.75	3.78×10^{13}	1.11×10^{5}
0.80	2.83×10^{13}	2.03×10^{4}
0.85	1.54×10^{13}	2.34×10^{3}
0.90	5.74×10^{12}	1.45×10^{2}
0.95	1.31×10^{12}	3.41×10^{0}
1.00	1.38×10^{11}	0.00×10^{0}

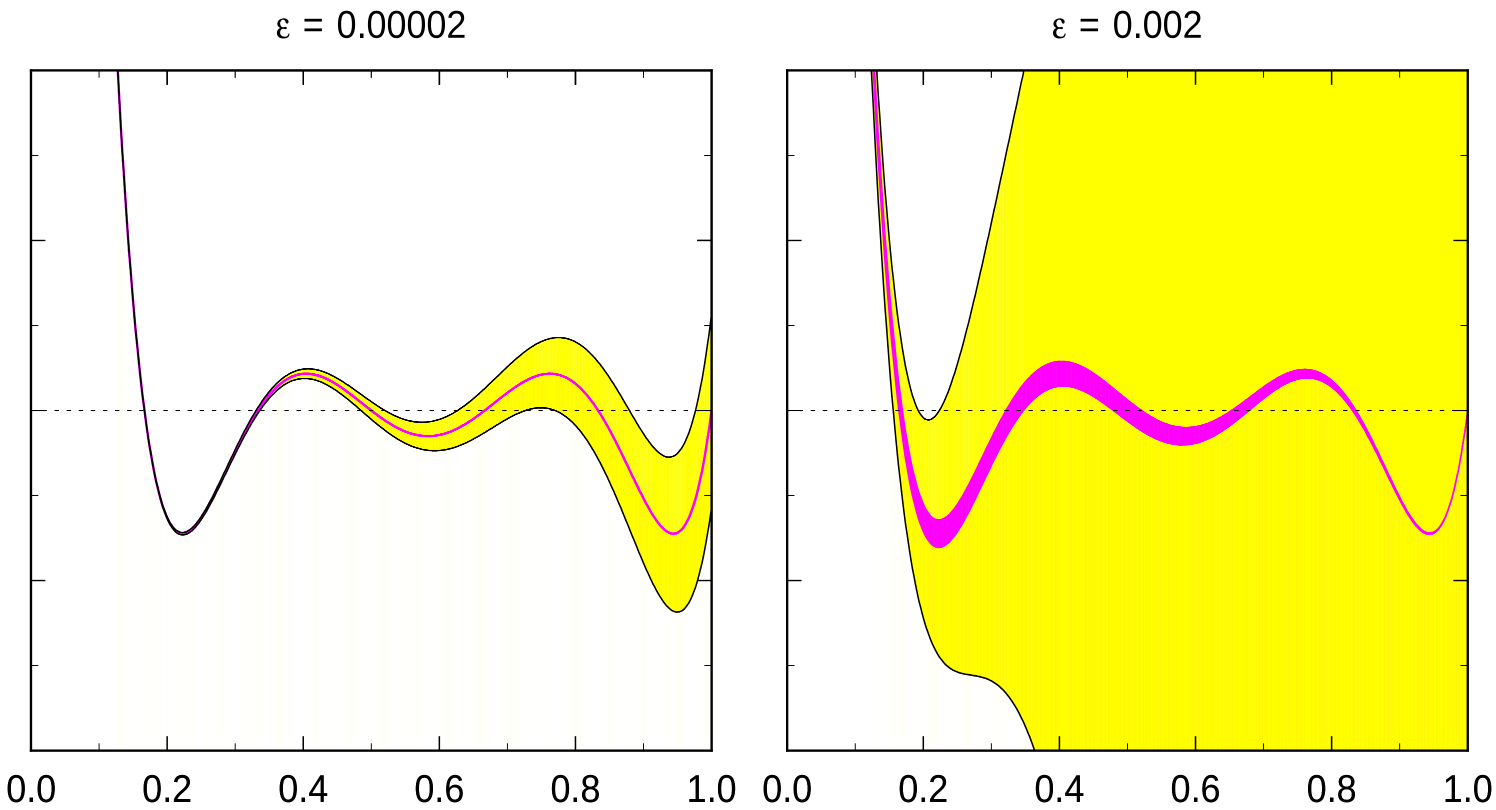
perturbed roots of Wilkinson polynomial — $\epsilon = 5 \times 10^{-10}$

root	power basis	Bernstein basis
0.05	0.05000000	0.0500000000
0.10	0.10000000	0.1000000000
0.15	0.15000000	0.1500000000
0.20	0.20000000	0.2000000000
0.25	0.25000000	0.2500000000
0.30	0.30000035	0.3000000000
0.35	0.34998486	0.3500000000
0.40	0.40036338	0.4000000000
0.45	0.44586251	0.4500000000
0.50	$0.50476331\pm$	0.5000000000
0.55	0.03217504 i	0.5499999997
0.60	$0.58968169\pm$	0.6000000010
0.65	0.08261649 i	0.6499999972
0.70	$0.69961791\pm$	0.7000000053
0.75	0.12594150 i	0.7499999930
0.80	$0.83653687\pm$	0.8000000063
0.85	0.14063124 i	0.8499999962
0.90	$0.97512197\pm$	0.9000000013
0.95	0.09701652 i	0.9499999998
1.00	1.04234541	1.0000000000

evaluating Wilkinson's polynomial @ $t = 0.525$

$$
\begin{aligned}
a_0 &= +0.000000023201961595 \\
a_1\,t &= -0.000000876483482227 \\
a_2\,t^2 &= +0.000014513630989446 \\
a_3\,t^3 &= -0.000142094724489860 \\
a_4\,t^4 &= +0.000931740809130569 \\
a_5\,t^5 &= -0.004381740078100366 \\
a_6\,t^6 &= +0.015421137443693244 \\
a_7\,t^7 &= -0.041778345191908158 \\
a_8\,t^8 &= +0.088811127150105239 \\
a_9\,t^9 &= -0.150051459849195639 \\
a_{10}\,t^{10} &= +0.203117060946715796 \\
a_{11}\,t^{11} &= -0.221153902712311843 \\
a_{12}\,t^{12} &= +0.193706822311568532 \\
a_{13}\,t^{13} &= -0.135971108107894016 \\
a_{14}\,t^{14} &= +0.075852737479877575 \\
a_{15}\,t^{15} &= -0.033154980855819210 \\
a_{16}\,t^{16} &= +0.011101552789116296 \\
a_{17}\,t^{17} &= -0.002747271750190952 \\
a_{18}\,t^{18} &= +0.000473141245866219 \\
a_{19}\,t^{19} &= -0.000050607637503518 \\
a_{20}\,t^{20} &= +0.000002530381875176 \\
\hline
p(t) &= 0.000000000000003899
\end{aligned}
$$

perturbation regions for $p(t) = (t - \frac{1}{6}) \cdots (t - 1)$



perturbed Bernstein form

perturbed power form

optimal stability of Bernstein basis

$\Psi = \{\psi_0(t), \ldots, \psi_n(t)\}$ and $\Phi = \{\phi_0(t), \ldots, \phi_n(t)\}$ non–negative on $[\,a, b\,]$

Theorem.

$$\text{If} \quad \psi_j(t) \,=\, \sum_{k=0}^{n} M_{jk}\phi_k(t) \quad \text{with} \quad M_{jk} \geq 0\,,$$

then the condition numbers for the value of *any* degree n polynomial $p(t)$ at *any* point $t \in [\,a, b\,]$ in the bases Φ and Ψ satisfy

$$C_\Phi(p(t)) \,\leq\, C_\Psi(p(t))\,.$$

We say that the Φ basis is *systematically more stable* than the Ψ basis.

Example: $\Phi = \{b_0^n(t), \ldots, b_n^n(t)\}$ and $\Psi = \{1, t, \ldots, t^n\}$ — in fact, the Bernstein basis is *optimally stable* (it is impossible to construct a basis on $[\,0, 1\,]$ that is systematically more stable).

optimal stability (sketch)

$\mathcal{P}_n =$ set of all *non–negative bases* for degree-n polynomials on $[\,a, b\,]$.

For $\Phi, \Psi \in \mathcal{P}_n$ we write $\Phi \prec \Psi$ if $\Psi \;=\; \mathbf{M}\,\Phi$ for a *non–negative matrix* $\mathbf{M}$.

The relation $\prec$ is a *partial ordering* of the set of non-negative bases $\mathcal{P}_n$.

Theorem. $\Phi \prec \Psi \iff C_\Phi(p(t)) \leq C_\Psi(p(t))$ for all $p(t) \in \mathcal{P}_n$ and $t \in [\,a, b\,]$.

Definition. Φ is a *minimal basis* in $\mathcal{P}_n$ if no Ψ exists, such that $\Psi \prec \Phi$.

A minimal basis in $\mathcal{P}_n$ is *optimally stable* — it is impossible to construct a non–negative basis on $[\,a, b\,]$ that is systematically more stable.

Theorem. The *Bernstein basis* is minimal in $\mathcal{P}_n$, and is optimally stable. It is the only minimal basis whose basis functions have no roots in (a, b).

MATHEMATICS OF COMPUTATION
Volume 65, Number 216
October 1996, Pages 1553–1566

ON THE OPTIMAL STABILITY OF THE BERNSTEIN BASIS

R. T. FAROUKI AND T. N. T. GOODMAN

ABSTRACT. We show that the Bernstein polynomial basis on a given interval is "optimally stable," in the sense that no other nonnegative basis yields systematically smaller condition numbers for the values or roots of arbitrary polynomials on that interval. This result follows from a partial ordering of the set of all nonnegative bases that is induced by nonnegative basis transformations. We further show, by means of some low–degree examples, that the Bernstein form is not uniquely optimal in this respect. However, it is the only optimally stable basis whose elements have no roots on the interior of the chosen interval. These ideas are illustrated by comparing the stability properties of the power, Bernstein, and generalized Ball bases.

1. INTRODUCTION

To represent a polynomial p in a digital computer, we store in memory its coefficients $c_0, \dots, c_n$ in a suitable basis. These coefficients, together with a value t of the independent variable, serve as input to an evaluation algorithm that furnishes the polynomial value $p(t)$ as output.

less stable than the power form. Incidentally, it is interesting to note that the Chebyshev basis on $t \in [0,1]$ also gives a very unstable representation of this polynomial; see Example 4.2′ in [6]. Some of the root condition numbers are as large as 10^{55}! (that's an exclamation mark, not a factorial — 10^{55} is surely a sufficiently impressive number in its own right).

condition numbers can be "very large" !

least-squares polynomial approximation

minimize $\displaystyle\int_0^1 [\, f(t) - p_n(t)\,]^2 \,\mathrm{d}t\,,\qquad p_n(t) \,=\, \sum_{k=0}^{n} a_k\phi_k(t)$

orthogonal basis $\displaystyle\int_0^1 \phi_j(t)\,\phi_k(t)\,\mathrm{d}t \,=\, \begin{cases} \beta_k & j=k \\ 0 & j\neq k \end{cases}$

$$\Longrightarrow \quad a_k \,=\, \frac{1}{\beta_k}\int_0^1 f(t)\,\phi_k(t)\,\mathrm{d}t$$

permanence of coefficients: $a_0,\ldots,a_n$ unchanged when $n \to n+1$

orthogonality impossible for non–negative bases,

but Bernstein basis is intimately related to Legendre basis

Legendre and Bernstein bases on $t \in [\,0,1\,]$

recurrence relation $L_0(t) = 1,\; L_1(t) = 2t - 1$

$$(k+1)L_{k+1}(t) \;=\; (2k+1)(2t-1)\,L_k(t) \;-\; k\,L_{k-1}(t)$$

Rodrigues' formula $L_k(t) \;=\; \dfrac{(-1)^k}{k!}\,\dfrac{\mathrm{d}^k}{\mathrm{d}t^k}\,[\,(1-t)t\,]^k$

Bernstein form $L_k(t) \;=\; \displaystyle\sum_{i=0}^{k} (-1)^{k+i}\binom{k}{i}\,b_i^k(t)$

$$\begin{array}{ccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & -1 & & 1 & & & & & & \\
 & & & & & 1 & & -2 & & 1 & & & & & \\
 & & & & -1 & & 3 & & -3 & & 1 & & & & \\
 & & & 1 & & -4 & & 6 & & -4 & & 1 & & & \\
 & & -1 & & 5 & & -10 & & 10 & & -5 & & 1 & & \\
 & 1 & & -6 & & 15 & & -20 & & 15 & & -6 & & 1 & \\
-1 & & 7 & & -21 & & 35 & & -35 & & 21 & & -7 & & 1
\end{array}$$

Pascal's triangle with alternating signs !!

Bernstein form of the Legendre polynomials

$$L_0(t) = 1\, b_0^0(t)\,,$$
$$L_1(t) = -\,1\, b_0^1(t) + 1\, b_1^1(t)\,,$$
$$L_2(t) = 1\, b_0^2(t) - 2\, b_1^2(t) + 1\, b_2^2(t)\,,$$
$$L_3(t) = -\,1\, b_0^3(t) + 3\, b_1^3(t) - 3\, b_2^3(t) + 1\, b_3^3(t)\,,$$
$$L_4(t) = 1\, b_0^4(t) - 4\, b_1^4(t) + 6\, b_2^4(t) - 4\, b_3^4(t) + 1\, b_4^4(t)\,,$$
$$L_5(t) = -\,1\, b_0^5(t) + 5\, b_1^5(t) - 10\, b_2^5(t) + 10\, b_3^5(t) - 5\, b_4^5(t) + 1\, b_5^5(t)\,,$$

Bernstein form of Legendre polynomial derivatives — e.g., $L_4(t)$

$$\begin{aligned}
L_4(t) &= 1\, b_0^4(t) - 4\, b_1^4(t) + 6\, b_2^4(t) - 4\, b_3^4(t) + 1\, b_4^4(t)\,,\\
\tfrac{1}{2}\, L_4'(t) &= 5\, b_0^3(t) - 10\, b_1^3(t) + 10\, b_2^3(t) - 5\, b_3^3(t)\,,\\
\tfrac{1}{3}\, L_4''(t) &= 15\, b_0^2(t) - 20\, b_1^2(t) + 15\, b_2^2(t)\,,\\
\tfrac{1}{3}\, L_4'''(t) &= 35\, b_0^1(t) - 35\, b_1^1(t)\,,\\
\tfrac{1}{5}\, L_4''''(t) &= 70\, b_0^0(t)\,,
\end{aligned}$$

Legendre–Bernstein basis transformations

$$p(t) \;=\; \sum_{k=0}^{n} a_k L_k(t) \;=\; \sum_{k=0}^{n} c_k b_k^n(t)$$

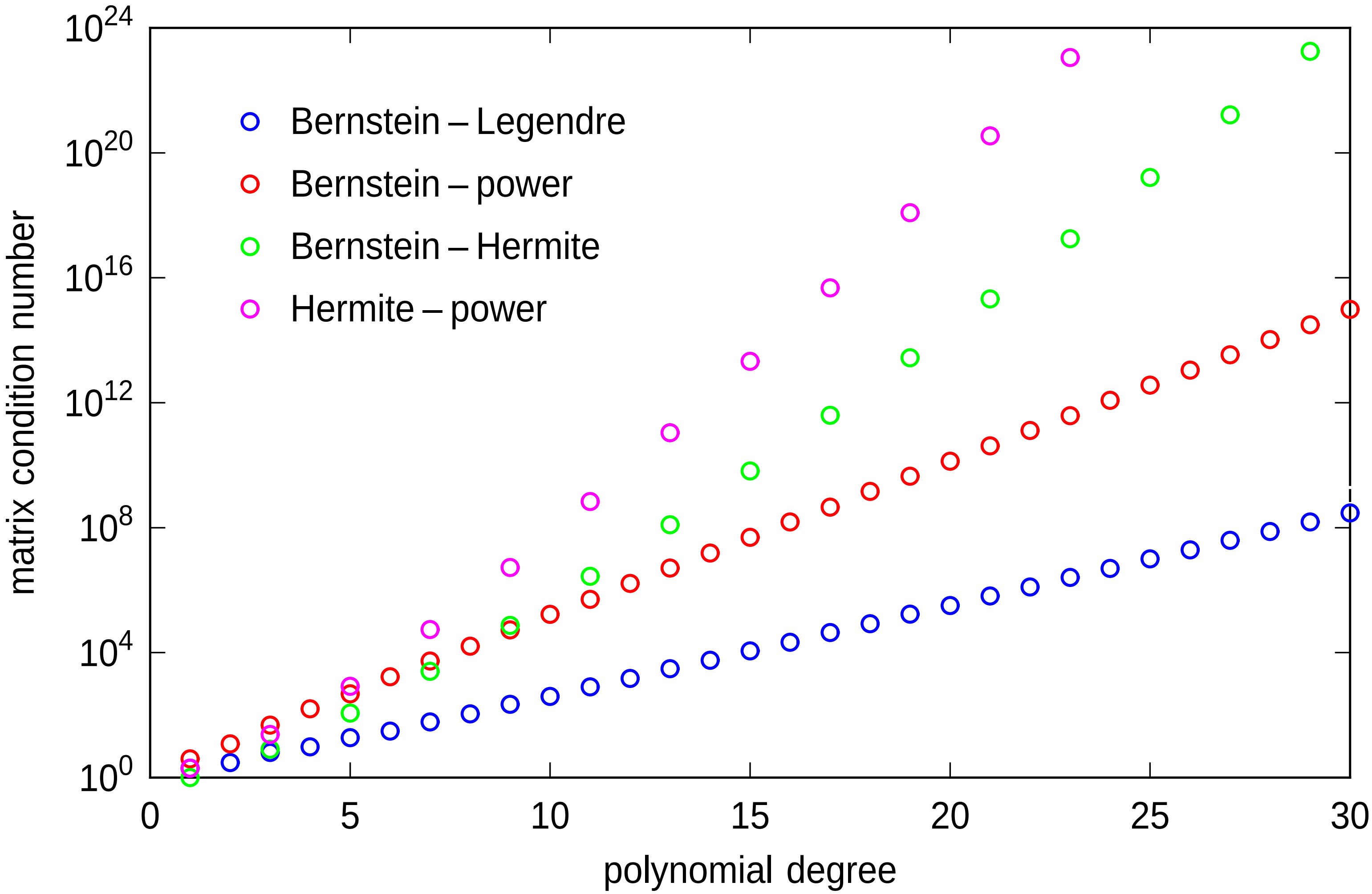
$$c_j \;=\; \sum_{k=0}^{n} M_{jk} a_k \,, \quad a_j \;=\; \sum_{k=0}^{n} M_{jk}^{-1} c_k$$

$$M_{jk} \;=\; \frac{1}{\binom{n}{k}} \sum_{i=\max(0,j+k-n)}^{\min(j,k)} (-1)^{k+i} \binom{j}{i} \binom{k}{i} \binom{n-k}{j-i}$$

$$M_{jk}^{-1} \;=\; \frac{2j+1}{n+j+1} \binom{n}{k} \sum_{i=0}^{j} (-1)^{j+i} \frac{\binom{j}{i}\binom{j}{i}}{\binom{n+j}{k+i}}$$

condition number $C_p(\mathbf{M}) = \|\mathbf{M}\|_p \|\mathbf{M}^{-1}\|_p \,, \quad C_1(\mathbf{M}) = 2^n > C_\infty(\mathbf{M})$

condition numbers for basis transformations



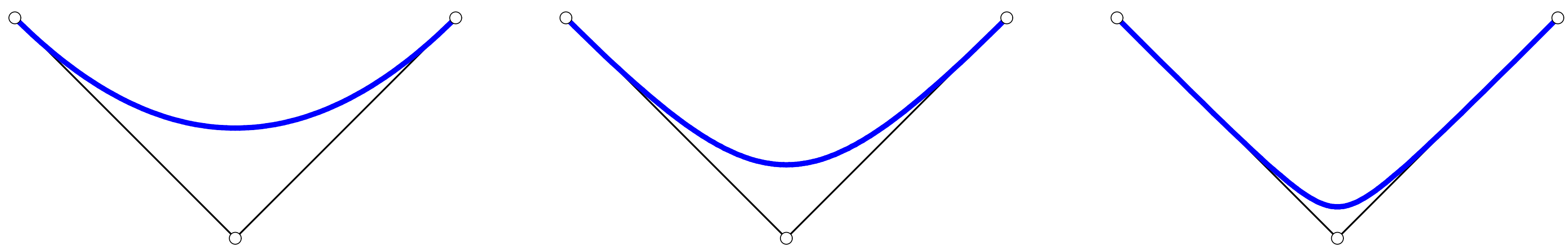
extension to rational forms

rational Bézier curve $$\mathbf{r}(t) = \frac{\sum_{k=0}^{n} w_k \mathbf{p}_k b_k^n(t)}{\sum_{k=0}^{n} w_k b_k^n(t)}$$

defined by control points $\mathbf{p}_0, \ldots, \mathbf{p}_n$ and scalar weights $w_0, \ldots, w_n$

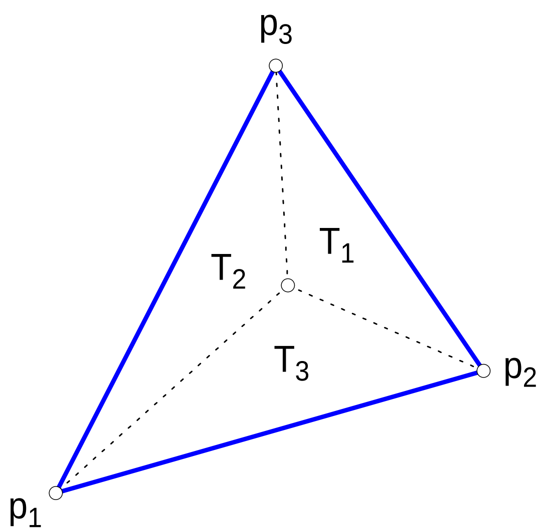
set of rational curves is closed under projective transformations

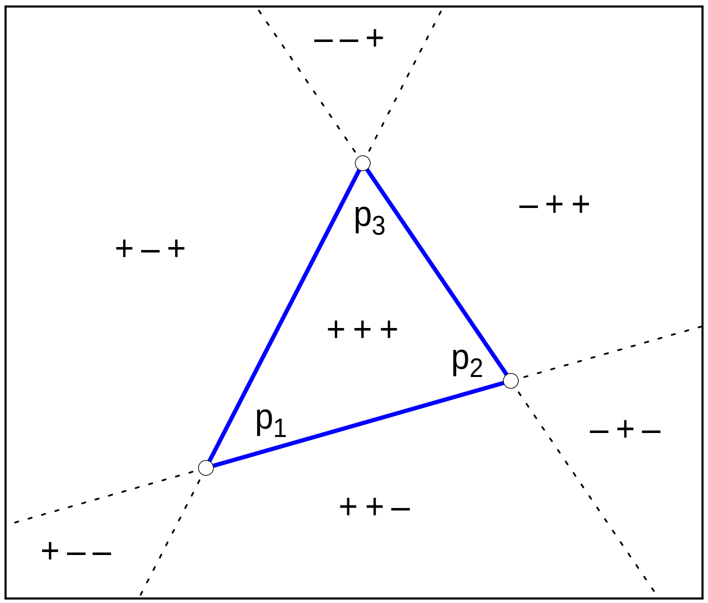
conic segments as rational quadratic Bézier curves $(w_0 = w_2 = 1)$



$w_1 < 1$ (ellipse) $w_1 = 1$ (parabola) $w_1 > 1$ (hyperbola)

bivariate & multivariate generalizations





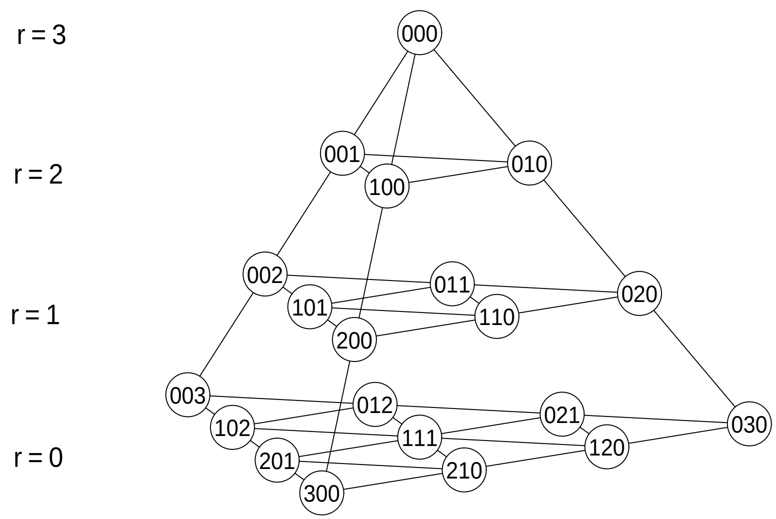
barycentric coordinates: $(u, v, w) = \frac{(\text{area}(T_1), \text{area}(T_2), \text{area}(T_3))}{\text{area}(T)}$

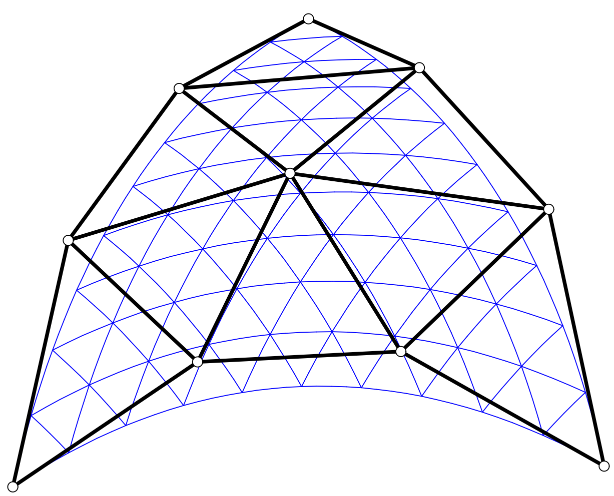
$$1 = (u + v + w)^n = \sum_{i+j+k=n} b^n_{ijk}(u, v, w)\,, \quad b^n_{ijk}(u, v, w) = \frac{n!}{i!j!k!} u^i v^j w^k$$

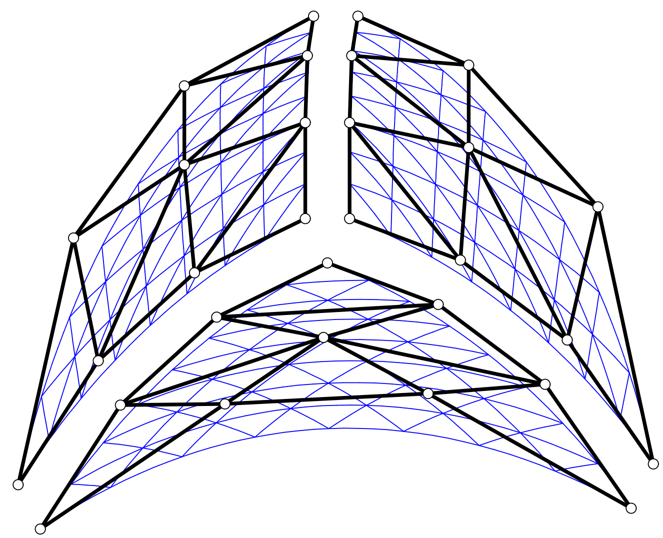
triangular surface patch: $\mathbf{r}(u, v, w) = \sum_{i+j+k=n} \mathbf{p}_{ijk} b^n_{ijk}(u, v, w)$

bivariate de Casteljau algorithm

generates tetrahedral array — *evaluates* and *subdivides* $\mathbf{r}(u,v,w)$

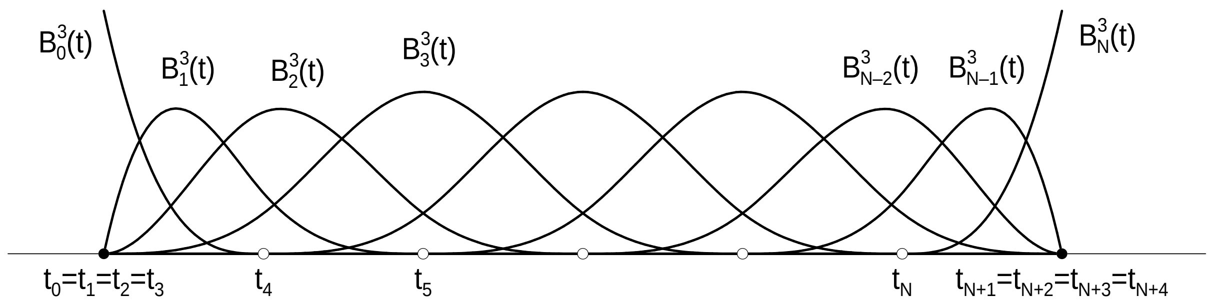




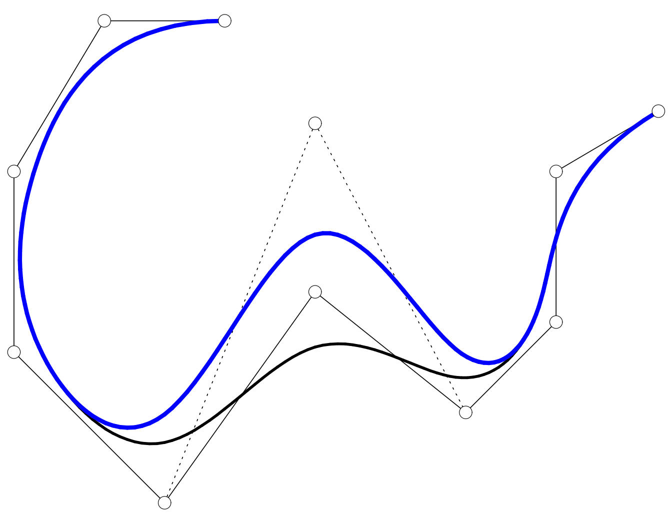


generalization to B-spline basis

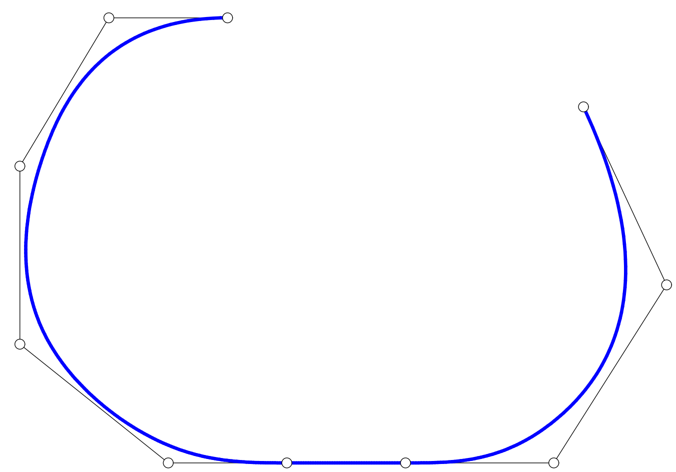
continuous domain $[0,1] \rightarrow$ partitioned domain $[t_0, t_1, \ldots, t_{N+n+1}]$



retain partition-of-unity, non-negativity, recursion properties
+ compact support & control over continuity (multiple knots)



local shape modification



C^2 embedded linear segment

scientific computing applications

- real solutions of systems of algebraic equations; identifying extrema or bounds on constrained or unconstrained polynomial functions in one or several variables (optimization) using Bernstein basis properties

- robust stability of dynamic systems with uncertain physical parameters (Kharitonov generalization of Routh-Hurwitz criterion)

- definition of barycentric coordinates and “partition-of-unity” polynomial basis functions over general polygon or polytope domains for use in the finite-element and meshless analysis methods

- modelling of inter-molecular potential energy surfaces; design of filters for signal processing applications; inputs to neurofuzzy networks modelling non-linear dynamical systems; reconstruction of 3D models and calibration of optical range sensors

closure

- 100 years have elapsed since introduction of Bernstein basis
- Bernstein form was limited to *theory*, rather than *practice*,* of polynomial approximation for ~ 50 years after its introduction
- applications in *design*, rather than *approximation*, pioneered ~ 50 years ago by de Casteljau and Bézier
- now universally adopted as a fundamental representation for computer-aided geometric design applications
- "optimally stable" basis for polynomials defined over finite domains
- Bernstein basis intimately related to Legendre orthogonal basis
- increasing adoption in diverse scientific computing applications

---

"In theory, there is no difference between theory and practice. In practice, there is." . . . Yogi Berra