1 Die Bernsteinschen Grundpolynome

Die *Bernstein-Polynome* sind benannt nach dem bedeutenden russischen Mathematiker S. N. Bernstein (1880-1968), der 1912 eine sehr kurze elegante Herleitung des Weierstraßschen Approximationssatzes aus dem Gesetz der großen Zahl publizierte: Bei n unabhängigen Wiederholungen eines Experiments, bei dem mit Wahrscheinlichkeit p ein „Erfolg“ (= ein bestimmtes interessierendes Ereignis) eintritt, ist die Wahrscheinlichkeit von genau k Erfolgen bekanntlich gegeben durch $\binom{n}{k}p^k(1-p)^{n-k}$. Mit wachsendem n wird es immer wahrscheinlicher, dass $\frac{k}{n} \approx p$, weshalb für eine stetige Funktion $f : [0,1] \to \mathbb{R}$ gilt:

$$\boldsymbol{B_n f(p)} := \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} p^k (1-p)^{n-k} \approx f(p)$$

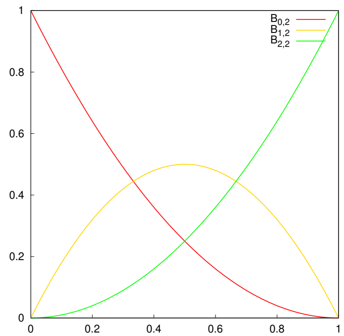
Wir werden dies weiter unten *direkt*, ohne Rückgriff aufs Gesetz der großen Zahl beweisen; dieses Gesetz wird umgekehrt *mit*bewiesen.

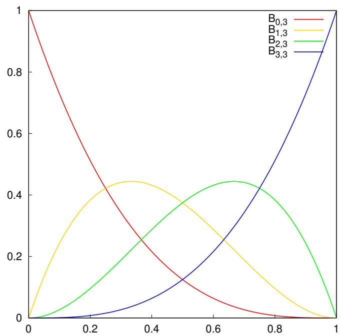
Wir führen folgende Bezeichnung ein für die sogenannten Bernsteinschen *Grundpolynome*:

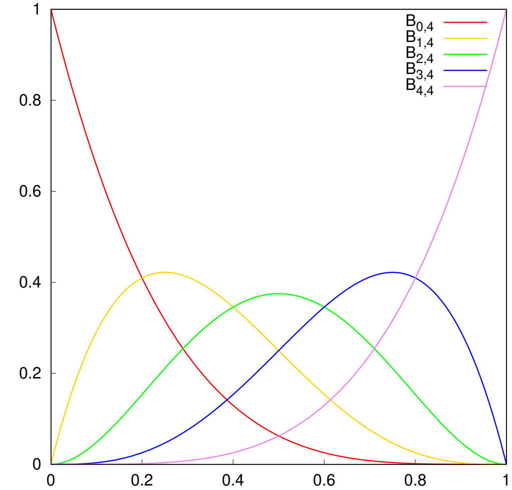
$$\boxed{\boldsymbol{B_{k,n}(t)} := \begin{cases} \binom{n}{k} t^k (1-t)^{n-k}, & 0 \le k \le n, \\ 0, & \text{sonst} \end{cases}} \tag{1}$$

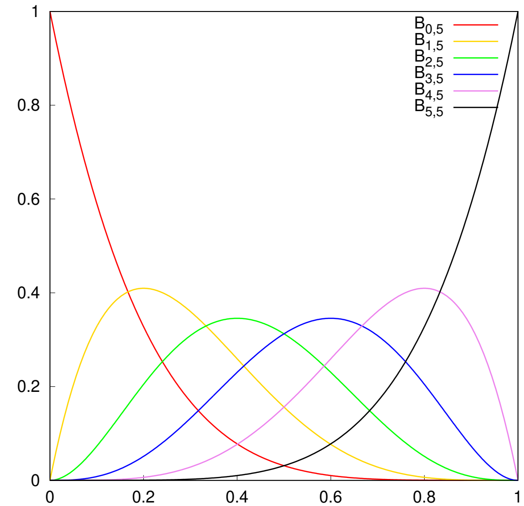
mit $0 \le t \le 1$.

Zunächst die ersten $B_{k,n}$ als Plots.









Die wichtigsten Eigenschaften der Grundpolynome:

a) Positivität

$$B_{k,n}(t) \geq 0 \quad (0 \leq t \leq 1) \tag{2}$$

b) Zerlegung der Eins

$$\sum_{k=0}^{n} B_{k,n}(t) = 1 \quad (0 \leq t \leq 1, n \geq 0) \tag{3}$$

c) Das **Maximum** in $[0,1]$ nimmt $B_{k,n}(t)$ an der Stelle $t = k/n$ an und wächst (fällt) streng monoton links (rechts) davon; hier $n \geq 1,\ 0 \leq k \leq n$; es gilt $B_{k+1,n}(\frac{k+1}{n})/B_{k,n}(\frac{k}{n}) = (1+\frac{1}{k})^k/(1+\frac{1}{n-k-1})^{n-k-1} \leq 1 \Leftrightarrow k \leq \frac{n-1}{2}$.

d) Rekursionsformeln

$$\begin{aligned} &B_{k,n+1}(t) = (1-t)B_{k,n}(t) + tB_{k-1,n}(t) \quad (0 \leq t \leq 1,\ 0 \leq k \leq n) \\ &B_{k,n+1}(t) = \frac{(n+1)(1-t)}{n+1-k}B_{k,n}(t), \quad B_{k+1,n+1}(t) = \frac{(n+1)t}{k+1}B_{k,n}(t) \end{aligned} \tag{4}$$

e) Symmetrie

$$B_{k,n}(t) = B_{n-k,n}(1-t) \quad (0 \leq t \leq 1) \tag{5}$$

f) Produkt-Identität

$$B_{k,n}(st) = \sum_{i=k}^{n} B_{k,i}(s)B_{i,n}(t) \left(\underset{(1)}{=} \sum_{i=0}^{n} B_{k,i}(s)B_{i,n}(t) \right) \quad (0 \leq s,t \leq 1) \tag{6}$$

g) Erste Ableitung

$$\begin{aligned} B'_{k,n}(t) &= n\left(B_{k-1,n-1}(t) - B_{k,n-1}(t)\right) \quad (0 \leq t \leq 1,\ 0 \leq k \leq n) \\ &= \frac{k-nt}{t(1-t)}B_{k,n}(t) = \frac{n(k-nt)}{(n-k)t}B_{k,n-1}(t) \quad (t \in (0,1),\ 0 \leq k < n) \end{aligned} \tag{7}$$

h) Zweite Ableitung

$$\begin{aligned} B''_{k,n} &= n(n-1)\left(B_{k-2,n-2}(t) - 2B_{k-1,n-2}(t) + B_{k,n-2}(t)\right) \ (t \in [0,1],\ 0 \leq k \leq n) \\ &= \frac{k(k-1) - 2k(n-1)t + n(n-1)t^2}{t^2(1-t)^2}B_{k,n}(t) \ (t \in (0,1),\ 0 \leq k \leq n) \end{aligned} \tag{8}$$

Beweise:
a) ist trivial, b) folgt unmittelbar aus $\sum_{k=0}^{n} B_{k,n}(t) = (t + (1-t))^n$.

Zu c): $B'_{k,n}(t) = \binom{n}{k}\big(k(1-t) - (n-k)t\big)t^{k-1}(1-t)^{n-k-1}$ für $1 < k < n$. Und $k(1-t) - (n-k)t = k - nt$ ist positiv für $t < k/n$, negativ für $t > k/n$. Die zweite Aussage folgt durch einfaches Umformen; $(1+\frac{1}{n})^n$ wächst streng mit n.

Zu d): $B_{k,n+1}(t) = \binom{n+1}{k}(1-t)^{n+1-k}t^k =$ (wegen $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$) $\binom{n}{k}(1-t)(1-t)^{n-k}t^k + \binom{n}{k-1}(1-t)^{n-(k-1)}t\,t^{k-1} = (1-t)B_{k,n}(t) + tB_{k-1,n}(t)$ im Falle $k \geq 1$. Falls $k = 0$, fällt der zweite Summand weg, und die Formel bleibt wegen (1) trotzdem richtig. Mit $\binom{n+1}{k} = \frac{n+1}{n+1-k}\binom{n}{k}$ folgt unmittelbar die zweite Rekursionsformel, ebenfalls für $0 \leq t \leq 1,\ 0 \leq k \leq n$. Da $\binom{n+1}{k+1} = \frac{n+1}{k+1}\binom{n}{k}$, ergibt sich die dritte Formel, wieder für $0 \leq t \leq 1,\ 0 \leq k \leq n$.

Zu e): Dies ist klar wegen $\binom{n}{k} = \binom{n}{n-k}$ für $0 \leq k \leq n$ und ist wegen (1) auch sonst richtig.

Zu f): Es gilt $\binom{n}{k}\binom{n-k}{i} = \frac{n!}{k!(n-k)!}\frac{(n-k)!}{i!(n-k-i)!} = \frac{n!}{i!(n-i)!}\frac{(n-i)!}{k!(n-i-k)!} = \binom{n}{i}\binom{n-i}{k}$ $(\bullet)$.

$$\begin{aligned} B_{k,n}(st) &= \binom{n}{k}(1-st)^{n-k}(st)^k = \binom{n}{k}(1-t+t(1-s))^{n-k}s^k t^k \\ &= \sum_{i=0}^{n-k}\binom{n}{k}\binom{n-k}{i}(1-t)^i t^{n-i}(1-s)^{n-k-i}s^k \\ &\underset{(\bullet)}{=} \sum_{i=0}^{n-k} B_{i,n}(1-t)B_{k,n-i}(s) \underset{\text{e)}}{=} \sum_{i=0}^{n-k} B_{k,n-i}(s)B_{n-i,n}(t) \end{aligned}$$

Zu g): Die erste der Identitäten ist mehr oder weniger trivial und gilt auch für $k = 0,\ n = 0$ wegen der Festlegung (1). Die Identität $B'_{k,n}(t) = \frac{k-nt}{t(1-t)}B_{k,n}(t)$ $(0 < t < 1,\ 0 \leq k \leq n)$ ergibt sich unmittelbar aus der Rechnung bei c). Die dritte Darstellung folgt mit $\binom{n}{k} = \frac{n}{n-k}\binom{n-1}{k}$.

Zu h): Die erste Darstellung folgt durch nochmalige Ableitung und Anwendung der ersten Formel aus g). Grenzfälle sind dabei wieder durch (1) einbezogen. Die zweite Darstellung ergibt sich durch Ableitung und Anwendung der zweiten Formel aus g). ∎

Die Produkt-Identität f) ist ein *Matrizenprodukt*: Mit $M_n(x) := \big(B_{i,k}(x)\big)_{\substack{0 \leq i \leq n \\ 0 \leq k \leq n}}$ gilt $M_n(s)M_n(t) = M_n(st)$. Gemäß (1) geht es dabei um obere Dreiecksmatrizen.

2 Approximation durch Bernstein-Polynome

Wir diskutieren nun detailliert die Approximation stetiger Funktionen durch Bernstein-Polynome – ein faszinierendes Thema.

a) *Ist f auf $[0,1]$ stetig, strebt $B_n f(x) = \sum_{k=0}^n f\big(\frac{k}{n}\big) B_{k,n}(x)$ gleichmäßig gegen $f(x)$; d.h.:*
Zu jedem $\delta > 0$ gibt's ein n_0 mit $|f(x) - B_n f(x)| < \delta$ $(x \in [0,1])$ für $n \geq n_0$.

Dazu betrachten wir mit $M := \max_{0\leq x\leq 1} |f(x)|$

$$|f(x) - B_n f(x)| = \left|\sum_{k=0}^n \Big(f(x) - f\big(\frac{k}{n}\big)\Big) B_{k,n}(x)\right|$$
$$\leq \sum_{|x-k/n|<\varepsilon} \left|f(x) - f\big(\frac{k}{n}\big)\right| B_{k,n}(x) + 2M \sum_{|x-k/n|\geq\varepsilon} B_{k,n}(x)$$

Die Summe derjenigen Binomialgewichte, für die der Abtastpunkt k/n um mindestens ε von x entfernt ist, wird mit wachsendem n immer kleiner:

$$\begin{aligned}\sum_{|x-k/n|\geq\varepsilon} B_{k,n}(x) &\leq \sum_{k=0}^n \frac{(nx-k)^2}{n^2\varepsilon^2} B_{k,n}(x) \quad (!!)\\ &= \frac{1}{n^2\varepsilon^2} \sum_{k=0}^n \left(n^2x^2 - 2nxk + k^2\right) B_{k,n}(x)\\ &= \frac{x^2}{\varepsilon^2} - \frac{2n^2x^2}{n^2\varepsilon^2} \sum_{k=1}^n B_{k-1,n-1}(x) + \frac{1}{n^2\varepsilon^2} \sum_{k=1}^n (k(k-1)+k) B_{k,n}(x)\\ &= \frac{-x^2}{\varepsilon^2} + \frac{x}{n\varepsilon^2} + \frac{x}{n\varepsilon^2} \sum_{k=2}^n (k-1) B_{k-1,n-1}(x) = \frac{x(1-x)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}\end{aligned}$$

Dabei wurde mehrfach (3) und (4), insbesondere $kB_{k,n}(x) = nxB_{k-1,n-1}(x)$ und $(k-1)B_{k-1,n-1}(x) = (n-1)xB_{k-2,n-2}(x)$ benutzt.

Ist nun ε so klein, dass $|f(x_1) - f(x_2)| < \frac{\delta}{2}$ für $x_1, x_2 \in [0,1]$ und $|x_1 - x_2| < \varepsilon$, ferner n so groß, dass $\frac{M}{2n\varepsilon^2} < \frac{\delta}{2}$, folgt $|f(x) - B_n f(x)| < \delta$ $(0 \leq x \leq 1)$. ∎

Den Bezug zum Gesetz der großen Zahl, der dieser Polynom-Approximation zugrunde liegt, hat knapp und prägnant S. N. Bernstein in seiner sehr kurzen Originalarbeit von 1912 formuliert (in französischer Sprache abgefasst). Umgekehrt folgt aus der hier dargestellten Abschätzung auch leicht die Tschebyscheffsche Ungleichung und damit Jakob Bernoullis *Gesetz der großen Zahl*; denn bei n unabhängigen Wiederholungen eines Zufallsexperiments ist die Wahrscheinlichkeit, dass die relative Häufigkeit eines Ereignisses E um mindestens ε von seiner Eintretenswahrscheinlichkeit $\mathrm{P}(E) = p$ abweicht, gerade $\sum_{|p-k/n|\geq\varepsilon} B_{k,n}(p) \leq \frac{p(1-p)}{n\varepsilon^2}$.

b) *Ist f auf $[0,1]$ (streng) monoton wachsend, so auch $B_n f$.*

Denn mit $f(0) \leq f(\frac{1}{n}) \leq f(\frac{2}{n}) \leq \cdots \leq f(1)$, $\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$ und $\binom{n}{k} = \binom{n}{n-k}$ folgt (man beachte, dass $n \geq 1$)

$$\begin{aligned}B_n' f(x) &= \sum_{k=0}^n f\big(\frac{k}{n}\big) \left(\binom{n}{k} x^k (1-x)^{n-k}\right)'\\ &= n \sum_{k=1}^n f\big(\frac{k}{n}\big) \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k} - n \sum_{k=0}^{n-1} f\big(\frac{k}{n}\big) \binom{n-1}{n-k-1} x^k (1-x)^{n-k-1}\\ &= n \sum_{k=0}^{n-1} \left(f\big(\frac{k+1}{n}\big) - f\big(\frac{k}{n}\big)\right) \binom{n-1}{k} x^k (1-x)^{n-1-k} \geq 0\end{aligned}$$

Im Falle *strenger* Monotonie von f folgt $B_n' f(x) > 0$. ∎

c) *Ist f auf $[0,1]$ (strikt) konvex, so auch $B_n f$.* (Exkurs Konvexität am Schluss.)

Ist f konvex, sind alle Funktionen $f_h(x) := \frac{f(x+h)-f(x)}{h}$ monoton wachsend, im Falle strikter Konvexität streng wachsend. Mit $f(x)$ ist offenbar auch jede Funktion $f_n(x) := \frac{n}{n-1} f\big(\frac{n-1}{n} x\big)$ (strikt) konvex. Die Rechnung bei b) zeigte, dass $B_n' f = B_{n-1} \varphi_n$ mit $\varphi_n(x) := \frac{f_n(x+\frac{1}{n-1}) - f_n(x)}{1/(n-1)}$, so dass B_n' nach b) bei (strikt) konvexem f (streng) wächst. ∎

d) *Ist f stetig differenzierbar, strebt $B_n' f$ gleichmäßig gegen f'.*

Gemäß Mittelwertsatz gibt es Zahlen $\xi_{k,n} \in [\frac{k}{n}, \frac{k+1}{n}]$ $(0 \leq k \leq n)$, so dass $B_n' f(x) = \sum_{k=0}^{n-1} f'(\xi_{k,n}) \binom{n-1}{k} x^k (1-x)^{n-1-k}$ nach b), also für $n > 1$

$$|B_n' f(x) - B_{n-1} f'(x)| \leq \sum_{k=0}^{n-1} \left|f'(\xi_{k,n}) - f'\big(\frac{k}{n-1}\big)\right| \binom{n-1}{k} x^k (1-x)^{n-1-k}$$

Da f' gleichmäßig stetig ist, gilt $|f'(\xi_{k,n}) - f'\big(\frac{k}{n-1}\big)| < \varepsilon$ $(0 \leq k < n)$ für alle hinreichend großen n und damit $|B_n' f(x) - B_{n-1} f'(x)| < \varepsilon$. Da $B_{n-1} f'$ nach a) gleichmäßig gegen f' strebt, folgt dasselbe für $B_n' f$. ∎

e) *Ist f zweimal stetig differenzierbar, strebt $B_n'' f$ gleichmäßig gegen f''.*

Mit der Differenzen-Notation $\Delta_h f(x) := f(x+h) - f(x)$ gilt nach b) $B_n' f(x) = n \sum_{k=0}^{n-1} \Delta_{1/n} f(k/n) \binom{n-1}{k} x^k (1-x)^{n-1-k}$. Durch nochmalige Ableitung und völlig analoge Umformungsschritte wie bei b) ergibt sich

$$B_n''f(x) = n(n-1)\sum_{k=0}^{n-2}\Delta^2_{1/n}f(k/n)\binom{n-2}{k}x^k(1-x)^{n-2-k}$$

mit $\Delta_h^2 f(x) = \Delta_h\Delta_h f(x) = f(x+2h) - 2f(x+h) + f(x)$.
Da mit dem Mittelwertsatz

$$\Delta_{1/n}\Delta_{1/n}f(x) = \frac{1}{n}\Delta'_{1/n}f(\xi_1) = \frac{1}{n}\left(f'(\xi_1+1/n) - f'(\xi_1)\right) = \frac{1}{n^2}f''(\xi_1+\xi_2)$$

mit $\xi_1 \in (x, x+1/n),\ \xi_2 \in (0, 1/n)$, folgt

$$B_n''f(x) = \left(1-\frac{1}{n}\right)\sum_{k=0}^{n-2}f''(\xi_{n,k})\binom{n-2}{k}x^k(1-x)^{n-2-k}$$

mit $\xi_{n,k} \in (k/n, (k+2)/n)$. Analog zu d) ergibt sich die Behauptung. ∎

f) *Die Variation von B_nf im Intervall $[0,1]$ ist nicht größer als die von f.*

Mit $\mathcal{Z}_{[a,b]} := \{(x_0, x_1, \ldots, x_n) \mid n \in \mathbb{N}, a = x_0 < x_1 < \cdots < x_n = b\}$ ist die *Variation* oder *Schwankung* einer Funktion im Intervall $[a,b]$ definiert durch

$$\operatorname{Var}_{[a,b]} f := \sup\left\{\sum_{k=1}^{n}|f(x_k) - f(x_{k-1})| \;\middle|\; (x_0, x_1, \ldots, x_n) \in \mathcal{Z}_{[a,b]}\right\} \tag{9}$$

Ist g auf $[a,b]$ differenzierbar und g' integrierbar, folgt mit dem Mittelwertsatz, dass jede Schwankungssumme $\sum|g(x_k)-g(x_{k-1})|$ eine Riemann-Summe zum Integral $\int_a^b|g'|$ ist, also zwischen Ober- und Untersumme zur gegebenen Zerlegung des Intervalls $[a,b]$ liegt; daher gilt

$$\operatorname{Var}_{[a,b]} g = \int_a^b |g'(x)|\mathrm{d}x$$

Der Beweis von b) zeigt folglich

$$\operatorname{Var}_{[0,1]} B_nf \le n\sum_{k=0}^{n-1}\left|f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right|\binom{n-1}{k}\int_0^1 x^k(1-x)^{n-1-k}\mathrm{d}x$$

Partielle Integration ergibt für $0 \le k < n-1$

$$\begin{aligned}
&\binom{n-1}{k}\int_0^1 \underbrace{x^k}_{\uparrow}\underbrace{(1-x)^{n-1-k}}_{\downarrow}\mathrm{d}x\\
&= \underbrace{\frac{x^{k+1}}{k+1}(1-x)^{n-1-k}\Big|_{x=0}^{1}}_{=0} + \frac{n-1-k}{k+1}\binom{n-1}{k}\int_0^1 x^{k+1}(1-x)^{n-1-k-1}\mathrm{d}x\\
&= \binom{n-1}{k+1}\int_0^1 x^{k+1}(1-x)^{n-1-k-1}\mathrm{d}x
\end{aligned}$$

Also sind die Integrale $\int_0^1 B_{k,n-1}(x)\mathrm{d}x$ $(0 \le k \le n-1)$ alle gleich und damit nach (3) alle gleich $1/n$. Insbesondere folgt

$$\operatorname{Var}_{[0,1]} B_nf \le \sum_{k=0}^{n-1}\left|f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right| \le \operatorname{Var}_{[0,1]} f$$

und damit die Behauptung. ∎

Die weitere bemerkenswerte Eigenschaft der Bernstein-Gewichte, die hier eine Rolle spielte, sei noch explizit als Formel hervorgehoben:

$$\int_{\mathbf{0}}^{\mathbf{1}} \boldsymbol{B_{k,n}(x)\mathrm{d}x} = \binom{n}{k}\int_0^1 x^k(1-x)^{n-k}\mathrm{d}x = \frac{\mathbf{1}}{\boldsymbol{n+1}} \quad (0 \le k \le n) \tag{10}$$

g) *Ist f konvex, folgt $B_nf(x) \ge B_{n+1}f(x)$ $(0 \le x \le 1)$ für alle $n \in \mathbb{N}$.*

$$\begin{aligned}
&B_nf(x) - B_{n+1}f(x) =\\
&= \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}\cdot(x+(1-x))\\
&\quad - \sum_{k=1}^{n} f\left(\frac{k}{n+1}\right)\binom{n+1}{k}x^k(1-x)^{n+1-k} - f(0)(1-x)^{n+1} - f(1)x^{n+1}\\
&= \sum_{k=0}^{n-1} f\left(\frac{k}{n}\right)\binom{n}{k}x^{k+1}(1-x)^{n+1-(k+1)} + \sum_{k=1}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n+1-k}\\
&\quad - \sum_{k=1}^{n} f\left(\frac{k}{n+1}\right)\binom{n+1}{k}x^k(1-x)^{n+1-k}\\
&= \sum_{k=1}^{n}\left\{f\left(\frac{k-1}{n}\right)\frac{k}{n+1} + f\left(\frac{k}{n}\right)\frac{n-k+1}{n+1} - f\left(\frac{k}{n+1}\right)\right\}\binom{n+1}{k}x^k(1-x)^{n+1-k}
\end{aligned}$$

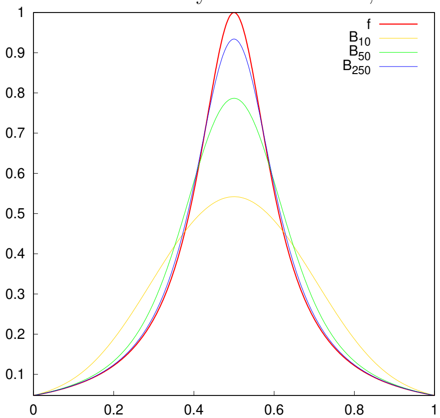
wegen $\binom{n}{k-1} = \frac{k}{n+1}\binom{n+1}{k}$ und $\binom{n}{k} = \frac{n-k+1}{n+1}\binom{n}{k}$. Da $\frac{k}{n+1}\cdot\frac{k-1}{n} + \frac{n-k+1}{n+1}\cdot\frac{k}{n} = \frac{k}{n+1}$, sind alle $\{\cdots\}$-Faktoren der Summe ≥ 0 wegen Konvexität von f. ■

Der Beweis zeigt, dass sogar $B_n f(x) > B_{n+1} f(x)$ $(0 < x < 1)$, wenn f in irgendeinem der Teilintervalle $[\frac{k-1}{n}, \frac{k}{n}]$ *nicht* linear ist.

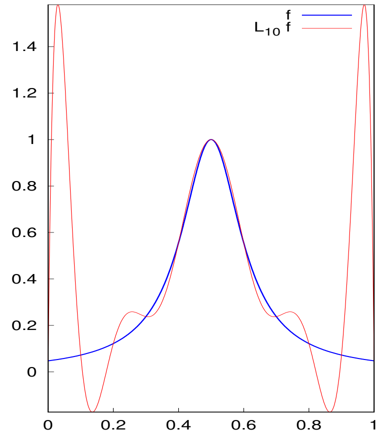
Als Beispiel für die Approximation durch Bernstein-Polynome wählen wir eine „Runge-Funktion“ (nach Carl Runge, berühmt unter anderem für das Runge/Kutta-Verfahren zur numerischen Lösung von Differentialgleichungen und den Approximationssatz von Runge in der komplexen Funktionentheorie)

$$f(x) := \frac{1}{1 + 5(4x-2)^2}\,.$$

Solche Funktionen, obwohl prinzipiell einfach gebaut, sind berüchtigt dafür, sehr schlecht durch klassische Interpolationspolynome angenähert zu werden. Die Funktion und ihre Bernstein-Polynome für n = 10, n = 50, n = 250:



Die Bernstein-Polynome nähern sich nur langsam an die Funktion an, haben aber von Anfang an die richtige Gestalt, da sie ja neben den Funktionswerten auch die Steigung und die Krümmung gleichmäßig approximieren. Ganz anders verhält sich das Interpolationspolynom, das an den Abtast-Stellen k/n $(0 \leq k \leq n)$ mit der Funktion übereinstimmt. Es reicht, den Fall $n = 10$ zu betrachten:



Der Zwang, an vorgegebenen äquidistanten Stellen exakt mit der Funktion übereinzustimmen, lässt das Polynom im Randbereich erheblich „ausschlagen“. Interpolationspolynome zu äquidistanten Stützstellen reagieren stark auf kleine Veränderungen der Stützwerte, sind bei nur mäßig großem n schon sehr schlecht konditioniert. – Nach einem berühmten Satz von L. Fejér sieht es anders aus, wenn man die auf $[0,1]$ umskalierten Tschebyscheff-Nullstellen als Stützstellen wählt und zusätzlich an diesen Stellen die Ableitung 0 vorschreibt (Hermite-Interpolation). Die Fejérschen „Treppenparabeln“ (so nannte er diese speziellen Interpolationspolynome) konvergieren gleichmäßig gegen eine stetige Funktion. Im Gegensatz zu den Bernstein-Polynomen konvergieren aber dabei die Ableitungen *nicht* gegen die der Grenzfunktion.

Wie *langsam* die Bernstein-Polynome – bei aller geometrischen Qualität der Approximation – selbst sehr gutartige Funktionen annähern, zeigt das Beispiel $\tilde{f}(x) := x(1-x)$; denn es ergibt sich (kleine Nebenrechnung)

$$B_n\tilde{f}(x) = \Big(1 - \frac{1}{n}\Big)x(1-x) \quad (n \geq 2).$$

Übrigens $B_n f = f$ $(n \geq 1)$ für $f \in \mathcal{P}_1$ und $f \in \mathcal{P}_k \;\Rightarrow\; B_n f \in \mathcal{P}_k$ $(n \geq k)$, wie leicht aus (19) weiter unten gefolgert werden kann. Wir zeigen noch (E. W. Woronowskaja, 1932):

h) *Ist f beschränkt auf $[0,1]$, differenzierbar in einer Umgebung von $x \in (0,1)$ und existiert $f''(x)$, so folgt*

$$\lim_{n\to\infty} n\Big(B_n f(x) - f(x)\Big) = \frac{x(1-x)}{2} f''(x)\,. \tag{11}$$

Man kann sogar zeigen (siehe Lorentz [4], S. 102ff., Beweis lang und *schwierig*):

$$|f(x) - B_n f(x)| \le \frac{Mx(1-x)}{2n} \;\Leftrightarrow\; f \text{ differenzierbar, } |f'(x) - f'(y)| \le M|x-y|$$

Insbesondere: $\max_{x\in[0,1]} n|f(x) - B_n f(x)|/\big(x(1-x)\big) \to 0\,(n \to \infty) \;\Leftrightarrow\; f$ linear.

Beweis von (11):
Nach Voraussetzung gilt für das Taylor-Polynom $T_2 f$ zweiten Grades zu f an der Stelle x, dass $f(t) - T_2 f(t) = \mathrm{o}((t-x)^2)$ $(t \to x)$. D.h.

$$f\big(t\big) = f(x) + f'(x)(t-x) + \Big(\frac{f''(x)}{2} + \delta(t)\Big)(t-x)^2 \qquad (\bullet)$$

mit einer beschränkten Funktion $\delta(t)$ mit $\delta(t) \to 0 = \delta(x)$ $(t \to x)$. Also:

$$\begin{aligned} B_n f(x) - f(x) &= \sum_{k=0}^n \Big(f\big(\frac{k}{n}\big) - f(x)\Big) B_{k,n}(x) \\ &\underset{(\bullet)}{=} \sum_{k=0}^n \Big(f'(x)\big(\frac{k}{n} - x\big) + \Big(\frac{f''(x)}{2} + \delta\big(\frac{k}{n}\big)\Big)\big(\frac{k}{n} - x\big)^2\Big) B_{k,n}(x) \\ &= \frac{f''(x)}{2n^2} \sum_{k=0}^n (k-nx)^2 B_{k,n}(x) + \frac{1}{n^2} \sum_{k=0}^n \delta\big(\frac{k}{n}\big)(k-nx)^2 B_{k,n}(x) \end{aligned}$$

Dabei wurde benutzt, dass lineare Funktionen $f(x) = ax + b$ mit ihrem Bernstein-Polynom übereinstimmen. Nun nutzen wir die im Beweis von a) hergeleitete Identität bzw. Abschätzung

$$\sum_{k=0}^n (k-nx)^2 B_{k,n}(x) = nx(1-x), \qquad \sum_{|x-k/n|\ge\varepsilon} B_{k,n}(x) \le \frac{x(1-x)}{n\varepsilon^2}. \qquad (12)$$

Wir fügen eine analoge, aber in gewissem Sinne *schärfere* Abschätzung hinzu. Zu dem Zweck betrachten wir die Summenausdrücke

$$\boldsymbol{S_{i,n}(x)} := \sum_{k=0}^n (k-nx)^i B_{k,n}(x), \text{ also } S_{0,n}(x) \equiv 1,\ S_{1,n}(x) \equiv 0,\ S_{2,n}(x) = nx(1-x)$$

Die Herleitung der Identität $S_{2,n}(x) = nx(1-x)$ benutzte die Rekursion $kB_{k,n}(x) = nxB_{k-1,n-1}(x)$. Wir leiten nun eine allgemeine Rekursionsbeziehung für die $S_{i,n}(x)$ her mittels $B'_{k,n}(x) = \frac{k-nx}{x(1-x)} B_{k,n}(x)$ (siehe (7)). Demnach ergibt sich $S'_{i,n}(x) = -i\,n\,S_{i-1,n}(x) + \frac{1}{x(1-x)} S_{i+1,n}(x)$ und damit

$$S_{i+1,n}(x) = x(1-x)\,S'_{i,n}(x) + i\,n\,x(1-x)\,S_{i-1,n}(x) \qquad (13)$$

Also $S_{3,n}(x) = x(1-x)\,n(1-2x) + 2nx(1-x)\cdot 0$ und nach kurzer Nebenrechnung

$$S_{4,n}(x) = \sum_{k=0}^n (k-nx)^4 B_{k,n}(x) = nx(1-x) + 3n(n-2)x^2(1-x)^2 \qquad (14)$$

und daher für $\varepsilon \in (0,1)$

$$\sum_{|x-k/n|\ge\varepsilon} (k-nx)^2 B_{k,n}(x) \le \sum_{k=0}^n \frac{(k-nx)^4}{n^2\varepsilon^2} B_{k,n}(x) = \frac{S_{4,n}(x)}{n^2\varepsilon^2} \qquad (15)$$

Folglich (beachte (12), (14), (15) und $\frac{1}{n} + \big(3 - \frac{6}{n}\big)x(1-x) \le 1$)

$$\frac{1}{n}\sum_{k=0}^n \Big|\delta\big(\frac{k}{n}\big)\Big|(k-nx)^2 B_{k,n}(x) \le x(1-x)\left(\sup_{|x-k/n|<\varepsilon} \Big|\delta\big(\frac{k}{n}\big)\Big| + \frac{\sup_{0\le t\le 1}|\delta(t)|}{n\varepsilon^2}\right)$$

und damit (11). ∎

Ergänzende Anmerkungen: (i) Per Induktion folgt, dass die $S_{i,n}(x)$ abwechselnd Polynome der Gestalt $P\big(x(1-x)\big)$ bzw. $Q\big(x(1-x)\big)(1-2x)$ sind, da $\big(x(1-x)\big)' = 1-2x$ und $(1-2x)^2 = 1 - 4x(1-x)$. Ebenso sieht man, dass $n^{\lfloor i/2 \rfloor}$ die maximale in $S_{i,n}$ vorkommende n-Potenz ist. Damit

$$\sum_{|x-k/n|\ge\varepsilon} B_{k,n}(x) \le \frac{S_{2i,n}(x)}{n^{2i}\varepsilon^{2i}} \le \frac{c_i}{n^i\varepsilon^{2i}}. \qquad (16)$$

(ii) Alternativer Rechenweg zur Herleitung von (14):

$$\begin{aligned} S_{3,n}(x) &= -nx\sum_{k=0}^n (k-nx)^2 B_{k,n}(x) + \sum_{k=1}^n (k-1-(n-1)x + (1-x))^2 k B_{k,n}(x) \\ &= -nxS_{2,n}(x) + nx\sum_{k=1}^n \Big(\big((k-1)-(n-1)x\big)^2 B_{k-1,n-1}(x) + \\ &\quad 2nx(1-x)\sum_{k=1}^n \big((k-1)-(n-1)x\big)B_{k-1,n-1}(x) + nx(1-x)^2 \sum_{k=1}^n B_{k-1,n-1}(x) \\ &= -nxS_{2,n}(x) + nxS_{2,n-1}(x) + 2nx(1-x)S_{1,n-1}(x) + nx(1-x)^2 S_{0,n-1}(x) \\ &= -n^2x^2(1-x) + n(n-1)x^2(1-x) + 0 + nx(1-x)^2 = nx(1-x)(1-2x) \end{aligned}$$

Analog

$$\begin{aligned} S_{4,n}(x) &= \sum_{k=0}^n (k-nx)(k-nx)^3 B_{k,n}(x) = -nxS_{3,n}(x) + \sum_{k=1}^n (k-nx)^3 k B_{k,n}(x) \\ &= -nxS_{3,n}(x) + nxS_{3,n-1}(x) + 3nx(1-x)S_{2,n-1}(x) + nx(1-x)^3 \\ &= nx(1-x) + 3n(n-2)\Big(x(1-x)\Big)^2 \end{aligned}$$

Es ist klar, dass man auch so rekursiv alle $S_{i,n}(x)$ bestimmen kann.

Beide Rechenwege nutzen im Umgang mit Bernstein-Gewichten nützliche Beziehungen.

(iii) Nun ein *dritter* Umformungsweg, umständlich, aber doch interessant.

$$\sum_{k=0}^{n} k(k-1)(k-2)B_{k,n}(x) = nx\sum_{k=1}^{n}(k-1)(k-2)B_{k-1,n-1}(x)$$
$$= n(n-1)x^2\sum_{k=2}^{n}(k-2)B_{k-2,n-2}(x) = n(n-1)(n-2)x^3\sum_{k=3}^{n}B_{k-3,n-3}(x),$$

also allgemein

$$\sum_{k=0}^{n} k(k-1)\cdots(k-i)B_{k,n}(x) = n(n-1)\cdots(n-i)x^{i+1} \quad (n \geq i+1) \quad (17)$$

$$\text{Rekursion:} \begin{cases} k^i = \displaystyle\sum_{j=0}^{i-1} c_j \prod_{l=0}^{j}(k-l) \text{ mit } c_0 = c_{i-1} = 1 \quad \Rightarrow \\ k^{i+1} = k + \displaystyle\sum_{j=1}^{i-1}(c_{j-1}+(j+1)c_j)\prod_{l=0}^{j}(k-l) + \prod_{l=0}^{i}(k-l) \end{cases}$$

Insbesondere: $\quad k^2 = k + k(k-1), \quad k^3 = k + 3k(k-1) + k(k-1)(k-2),$
$\quad k^4 = k + 7k(k-1) + 6k(k-1)(k-2) + k(k-1)(k-2)(k-3)\,.$

Daraus ergibt sich auch $\sum\limits_{k=0}^{n}(k-nx)^4B_{k,n}(x)$. Nebenbei: $\sum\limits_{k=0}^{n}\binom{k}{i+1}B_{k,n}(x) = \binom{n}{i+1}x^{i+1}$.

3 Bernstein-Basis und Monom-Basis.

Die Ableitungsberechnungen im Abschnitt 2 ergeben per Induktion offensichtlich die allgemeine Formel

$$B_n^{(k)}f(x) = n(n-1)\cdots(n-k+1)\sum_{i=0}^{n-k}\Delta_{\frac{1}{n}}^k f\left(\frac{i}{n}\right)\binom{n-k}{i}x^i(1-x)^{n-k-i} \quad (18)$$

Insbesondere folgt $\frac{1}{k!}B_n^{(k)}f(0) = \binom{n}{k}\Delta_{\frac{1}{n}}^k f(0)$ $(0 \leq k \leq n)$ und damit (Taylor!)

$$\boxed{B_n f(x) = \sum_{k=0}^{n} x^k\binom{n}{k}\Delta_{\frac{1}{n}}^k f(0)} \quad (19)$$

Die Koeffizienten dieser Monomdarstellung $B_nf(x) = c_0 + c_1x + c_2x^2 + \cdots + c_nx^n$ zeigen: $f \in \mathcal{P}_m \;\Rightarrow B_nf \in \mathcal{P}_m$ $(n \geq m)$; denn $\Delta_h\, x^n \in \mathcal{P}_{n-1}$ $(n \in \mathbb{N})$.

Und wegen $n^k\Delta_{\frac{1}{n}}^k f(0) \to f^{(k)}(0)$ folgt hieraus mit einfachen Abschätzungen, dass die Bernstein-Polynome zu *ganzen Funktionen* f in beschränkten $\mathbb{C}$-Teilmengen gleichmäßig gegen $f(z)$ konvergieren. (Siehe auch W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. II*, ^{2}1971, p. 222ff. hinsichtlich weiterer bemerkenswerter Folgerungen aus (19).)

Zum anderen können wir die Formeln allgemeiner formulieren, ohne Bezug zur Darstellung einer Funktion, indem wir von einer beliebigen Wertefolge $b_0, b_1, \ldots, b_n$ ausgehen und die Differenzen $\Delta b_k = b_{k+1} - b_k$ $(0 \leq k \leq n-1)$ und $\Delta^m b_k$ $(0 \leq m \leq n-k)$ betrachten:

$$\sum_{k=0}^{n} b_kB_{k,n}(x) = \sum_{k=0}^{n} b_k\binom{n}{k}x^k(1-x)^{n-k} = \sum_{k=0}^{n}\binom{n}{k}\Delta^kb_0\,x^k \quad (20)$$

So wird eine *beliebige* Linearkombination der Bernstein-Gewichte als Linearkombination der Monome dargestellt.

Wir überlegen nun, wie man umgekehrt eine völlig beliebige Linearkombination $c_0 + c_1x + \cdots + c_nx^n$ der Monome als Linearkombination der Bernstein-Gewichte ausdrücken kann.

Aus der Rekursionsformel $kB_{k,n}(x) = nxB_{k-1,n-1}(x)$ ergibt sich dies unmittelbar per Induktion. Denn es folgt $k(k-1)B_{k,n}(x) = n(n-1)x^2B_{k-2,,n-2}(x)$, $k(k-1)(k-2)B_{k,n}(x) = n(n-1)(n-2)x^3B_{k-3,n-3}(x)$, usw. Indem man diese Indentitäten von $k = 1$ bis n, von $k = 2$ bis n, von $k = 3$ bis n, usw.

aufsummiert, erhält man

$$\begin{aligned}
1 &= \sum_{k=0}^{n} B_{k,n}(x), \\
x &= \sum_{k=1}^{n} \frac{k}{n} B_{k,n}(x), \\
x^2 &= \sum_{k=2}^{n} \frac{k(k-1)}{n(n-1)} B_{k,n}(x), \\
&\vdots \\
x^n &= \sum_{k=n}^{n} \frac{k(k-1)\cdots(k-(n-1))}{n(n-1)\cdots(n-(n-1))} B_{k,n}(x)
\end{aligned}$$

Kompakter lässt sich dies schreiben als

$$x^i = \sum_{k=i}^{n} \frac{\binom{k}{i}}{\binom{n}{i}} B_{k,n}(x) \quad (0 \le i \le n)\,, \tag{21}$$

und es folgt

$$\sum_{i=0}^{n} c_i x^i = \sum_{i=0}^{n} c_i \sum_{k=i}^{n} \frac{\binom{k}{i}}{\binom{n}{i}} B_{k,n}(x) = \sum_{k=0}^{n} B_{k,n}(x) \sum_{i=0}^{k} c_i \frac{\binom{k}{i}}{\binom{n}{i}}\,. \tag{22}$$

Insbesondere ist gezeigt:

Die Bernstein-Gewichte $B_{0,n}(x), B_{1,n}(x), \ldots, B_{n,n}(x)$ sind eine Basis des Vektorraumes $\mathcal{P}_n$ der Polynome vom Grade $\le n$.

Es ist leicht, das einzelne Bernstein-Gewicht explizit als Linearkombination der Monome darzustellen, also die Umkehrung von (21) anzugeben:

$$\begin{aligned}
B_{i,n}(x) &= \binom{n}{i} x^i (1-x)^{n-i} = \binom{n}{i} x^i \sum_{k=0}^{n-i} \binom{n-i}{k} (-1)^k x^k \\
&= \binom{n}{i} \sum_{k=i}^{n} \binom{n-i}{k-i} (-1)^{k-i} x^k,
\end{aligned}$$

also

$$B_{i,n}(x) = \sum_{k=i}^{n} \binom{n}{k}\binom{k}{i} (-1)^{k-i} x^k \quad (0 \le i \le n) \tag{23}$$

wegen $\binom{n}{i}\binom{n-i}{k-i} = \binom{n}{k}\binom{k}{i}$. Damit ergibt sich

$$\begin{aligned}
\sum_{i=0}^{n} b_i B_{i,n}(x) &= \sum_{i=0}^{n} \sum_{k=i}^{n} b_i \binom{n}{k}\binom{k}{i} (-1)^{k-i} x^k \\
&= \sum_{k=0}^{n} \binom{n}{k} x^k \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} b_i
\end{aligned} \tag{20'}$$

Vergleich mit (20) zeigt:

$$\Delta^k b_0 = \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} b_i$$

Diese explizite Darstellung der Differenzen Δ^k kann man auch durch eine einfache algebraische „Operator"-Überlegung begründen: Mit $T\,a_k := a_{k+1}$, $I\,a_k := a_k$ folgt $\Delta a_k = a_{k+1} - a_k = (T-I)a_k$ und damit wegen $TI = IT$

$$\Delta^m a_k = (T-I)^m a_k = \sum_{i=0}^{m} \binom{m}{i} T^i (-I)^{m-i} a_k = \sum_{i=0}^{m} \binom{m}{i} (-1)^{m-i} a_{k+i}$$

Die Beziehung zwischen den Koeffizienten der Monom- und der Bernstein-Darstellung ein und desselben Polynoms

$$p(x) = b_0 B_{0,n}(x) + b_1 B_{1,n}(x) + \cdots + b_n B_{n,n}(x) = c_0 + c_1 x + \cdots + c_n x^n$$

lässt sich auch übersichtlich in Matrixform ausdrücken.

Nach (20) bzw. (20') gilt (platzsparende Notation: $(-)^n$ statt $(-1)^n$ usw.)

$$\begin{pmatrix} c_0/\binom{n}{0} \\ c_1/\binom{n}{1} \\ c_2/\binom{n}{2} \\ \vdots \\ c_n/\binom{n}{n} \end{pmatrix} = \underbrace{\begin{pmatrix} \binom{0}{0} & 0 & \cdots & \cdots & 0 \\ -\binom{1}{0} & \binom{1}{1} & \ddots & & \vdots \\ \binom{2}{0} & -\binom{2}{1} & \binom{2}{2} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ (-)^n\binom{n}{0} & (-)^{n-1}\binom{n}{1} & (-)^{n-2}\binom{n}{2} & \cdots & \binom{n}{n} \end{pmatrix}}_{=:A} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{24}$$

Ebenso ergibt (22)

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \underbrace{\begin{pmatrix} \binom{0}{0} & 0 & \cdots & \cdots & 0 \\ \binom{1}{0} & \binom{1}{1} & \ddots & & \vdots \\ \binom{2}{0} & \binom{2}{1} & \binom{2}{2} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \binom{n}{0} & \binom{n}{1} & \binom{n}{2} & \cdots & \binom{n}{n} \end{pmatrix}}_{=:B} \begin{pmatrix} c_0/\binom{n}{0} \\ c_1/\binom{n}{1} \\ c_2/\binom{n}{2} \\ \vdots \\ c_n/\binom{n}{n} \end{pmatrix} \tag{25}$$

Die beiden Matrizen A und B sind also zueinander *invers*, $A^{-1} = B$. Nebenbei sei bemerkt, dass man mit der Vandermonde-Identität für Binomialkoeffizienten zeigen kann: $BB^T = \left(\binom{i+k}{k}\right)_{\substack{0\le i\le n\\0\le k\le n}}$ („Pascal-Matrix", interessantes Beispiel einer *positiv definiten* Matrix).

Fazit: Einem beliebigen Polynom $p(x) = c_0 + c_1x + \cdots + c_nx^n$ entspricht umkehrbar eindeutig eine „Bernstein-Darstellung" $p(x) = \sum_{k=0}^n b_kB_{k,n}(x)$, und es gilt

$$\boxed{\begin{aligned} b_k &= \sum_{i=0}^{k} c_i\binom{k}{i}\Big/\binom{n}{i}, \\ c_k &= \binom{n}{k}\sum_{i=0}^{k} b_i(-1)^{k-i}\binom{k}{i} \end{aligned} \qquad (0\le k\le n)} \tag{26}$$

Die Formeln lassen sich ohne jede Änderung auf den mehrdimensionalen Fall übertragen, den der sogenannten *Bézier-Kurven*: Sind $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_n$ die *Kontrollpunkte*, die Eckpunkte des *Kontrollpolygons* einer Bézier-Kurve, kann man mittels der zweiten Formel (26) die entsprechenden – wenn man so will – Kontrollpunkte $\mathbf{c}_k$ bzgl. der Monom-Basis berechnen, die aber keine vergleichbare geometrische Bedeutung haben.

4 Exkurs: Konvexität

Eine Funktion $f : [a, b] \to \mathbb{R}$ heißt *konvex* (anschaulich: „überall durchhängend"), wenn

$$f(\lambda x_1+(1-\lambda)x_2) \underset{(*)}{\le} \lambda f(x_1)+(1-\lambda)f(x_2)(0 < \lambda < 1,\ a \le x_1 < x_2 \le b) \tag{27}$$

und *strikt* konvex, wenn bei $(*)$ stets „<" gilt.

Eine Funktion ist (strikt) konvex genau dann, wenn die Differenzenquotienten $\frac{f(x_2)-f(x_1)}{x_2-x_1}$ $(a \le x_1 < x_2 \le b)$ sowohl bzgl. x_1 als auch bzgl. x_2 (streng) monoton wachsen. Monotonie bzgl. x_1 (oder bzgl. x_2) *allein* ist gleichbedeutend mit Monotonie bzgl. *beiden*.

Dies ist leicht zu beweisen, wenn man die folgenden beiden Fakten beachtet:

$$\begin{aligned} &x = \frac{x_2-x}{x_2-x_1}x_1 + \frac{x-x_1}{x_2-x_1}x_2 \quad (a \le x_1 < x < x_2 \le b), \\ &\frac{a}{c} < \frac{a+b}{c+d} \Leftrightarrow \frac{a}{c} < \frac{b}{d} \Leftrightarrow \frac{a+b}{c+d} < \frac{b}{d} \quad (a,b,c,d \in \mathbb{R},\ c,d > 0) \end{aligned} \tag{28}$$

Eine auf $[a, b]$ konvexe Funktion ist überall stetig, besitzt sogar rechtsseitige und linksseitige Ableitungen – mit eventueller Ausnahme der Randpunkte a, b.

Denn: Sei $x \in (a, b)$ und $\varphi(\varepsilon) := (f(x+\varepsilon) - f(x))/\varepsilon$ $(\varepsilon \ne 0)$. Da φ monoton wächst, existieren $f'_-(x) := \lim_{\varepsilon\nearrow 0}\varphi(\varepsilon) \le \lim_{\varepsilon\searrow 0}\varphi(\varepsilon) =: f'_+(x)$. ■

Ist f differenzierbar und f' *streng monoton wachsend*, so ist f *strikt konvex*.

Denn für $x_1 < x_2 < x_3$ gilt gemäß Mittelwertsatz

$$\frac{f(x_2)-f(x_1)}{x_2-x_1} < \frac{f(x_3)-f(x_2)}{x_3-x_2},$$

so dass nach (28) die strenge Monotonie der Differenzenquotienten folgt.

Nun zu *Mengen*. Eine Menge heißt konvex genau dann, wenn sie mit je zwei Punkten auch deren Verbindungsstrecke umfasst; d.h.

$$\mathcal{M} \subseteq \mathbb{R}^n \text{ konvex} \quad :\Leftrightarrow \big(\mathbf{a}, \mathbf{b} \in \mathcal{M} \Rightarrow \{\lambda\mathbf{a}+(1-\lambda)\mathbf{b} \mid 0 \le \lambda \le 1\} \subseteq \mathcal{M}\big) \tag{29}$$

Ist $\mathcal{M}$ eine konvexe Menge, gilt $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{M}$, $\lambda_1, \ldots, \lambda_n \ge 0$ sowie $\lambda_1 + \cdots + \lambda_n = 1$, so folgt $\lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n \in \mathcal{M}$.

Dies ergibt sich leicht per Induktion aus (29). Hier nur der Fall $n = 3$:

Seien $\lambda_i > 0$ $(i = 1, 2, 3)$, gelte $\lambda_1 + \lambda_2 + \lambda_3 = 1$, ferner $\mathbf{x}_i \in \mathcal{M}$ $(i = 1, 2, 3)$. Dann folgt $\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + \lambda_3\mathbf{x}_3 = (\lambda_1 + \lambda_2)\big(\frac{\lambda_1}{\lambda_1+\lambda_2}\mathbf{x}_1 + \frac{\lambda_2}{\lambda_1+\lambda_2}\mathbf{x}_2\big) + \lambda_3\mathbf{x}_3 \in \mathcal{M}$ durch zweimalige Anwendung von (29). ∎

Man nennt jede Linearkombination $\lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n$ mit $\lambda_1 + \cdots + \lambda_n = 1$ und $\lambda_i \geq 0$ $(1 \leq i \leq n)$ eine *Konvex-Kombination* der Punkte $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

Trivial: Der Durchschnitt zweier (oder beliebig vieler) konvexer Mengen ist ebenfalls konvex. (Man beachte: Gemäß (29) ist $\emptyset$ konvex.) Dadurch lässt sich die *konvexe Hülle* $\mathrm{Co}(\mathcal{M})$ einer Menge $\mathcal{M} \subseteq \mathbb{R}^n$, die *kleinste konvexe Obermenge* der Menge, ganz einfach definieren als Durchschnitt aller konvexen Obermengen von $\mathcal{M}$.

Es gilt $\mathrm{Co}\big(\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}\big) = \{\lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n \mid \lambda_1 + \cdots + \lambda_n = 1, \lambda_i \in [0, 1]\,(1 \leq i \leq n)\}$ und allgemeiner $\mathrm{Co}(\mathcal{M}) = \big\{\sum_{i=1}^m \lambda_i\mathbf{x}_i \mid m \in \mathbb{N}, \mathbf{x}_i \in \mathcal{M}, \lambda_i \geq 0, \sum \lambda_i = 1\big\}$.

Denn erstens muss jede Konvex-Kombination von Punkten aus $\mathcal{M}$ in der konvexen Hülle vorkommen. Zweitens *ist* die rechts stehende Menge aller Konvex-Kombinationen konvex; denn die Konvex-Kombination zweier Konvex-Kombinationen von Punkten ist wieder eine: $\lambda(\lambda_1\mathbf{x}_1 + \cdots + \lambda_m\mathbf{x}_m) + (1-\lambda)(\lambda_1^*\mathbf{x}_1^* + \cdots + \lambda_{m^*}^*\mathbf{x}_{m^*}^*) = \sum_{i=1}^{m+m^*}\mu_i\mathbf{y}_i, \mu_i = \lambda\lambda_i\,(i \leq m)$, etc.

Exemplarisch seien abschließend zwei grundlegende Aussagen über konvexe Funktionen bzw. Mengen vorgestellt.

Jensensche Ungleichung: *Ist $f : [a, b] \to \mathbb{R}$ strikt konvex, gilt*

$$f(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n) < \lambda_1 f(x_1) + \lambda_2 f(x_2) + \cdots + \lambda_n f(x_n)$$

für $\lambda_1, \lambda_2, \ldots, \lambda_n > 0$, $\sum \lambda_i = 1$ und nicht sämtlich übereinstimmende $x_i \in [a, b]$.

Beweis:
Man kann dies ganz leicht per Induktion über n beweisen. Stattdessen hier eine mehr *geometrische* Begründung.

Mit der Formulierung des Satzes entsprechenden λ_i und x_i folgt $x_0 := \lambda_1 x_1 + \cdots + \lambda_n x_n \in (a, b)$, und $T(x) := f(x_0) + m(x - x_0)$ mit $f'_-(x_0) \leq m \leq f'_+(x_0)$ ist eine *Stützgerade* zu f (eventuell – im Fall $f'_-(x_0) = f'_+(x_0)$ – sogar eine *Tangente*), die strikt unterhalb von f verläuft, mit Ausnahme des Berührpunktes x_0; d.h. $T(x_i) < f(x_i)$ für $x_i \neq x_0$.

Also folgt $f(x_0) = T(x_0) = \sum\limits_{i=1}^n \lambda_i T(x_i) < \sum\limits_{i=1}^n \lambda_i f(x_i)$. ∎

Viele klassische Ungleichungen sind nichts als Spezialfälle der Jensenschen:

a) Die Exponentialfunktion $f(x) = \mathrm{e}^x$ ist strikt konvex, da $f''(x) = f(x) > 0$ für alle x. Sind also $x_1, \ldots . x_n > 0$, nicht alle gleich, ferner $\lambda_1, \ldots, \lambda_n > 0$, $\lambda_1 + \cdots + \lambda_n = 1$, so folgt gemäß Jensen

$$f(\lambda_1 \ln x_1 + \cdots + \lambda_n \ln x_n) < \lambda_1 f(\ln x_1) + \cdots + \lambda_n f(\ln x_n)$$

und damit

$$x_1^{\lambda_1} \cdots x_n^{\lambda_n} < \lambda_1 x_1 + \cdots + \lambda_n x_n, \tag{30}$$

die verallgemeinerte *Ungleichung zwischen arithmetischem und geometrischem Mittel*; $\lambda_i = n^{-1}$ $(1 \leq i \leq n)$ ist der klassische Fall.

b) Die Funktion $f(x) = x^p$ $(x > 0)$ mit $p > 1$ ist strikt konvex, da $f'(x) = px^{p-1}$ offenbar streng wächst. Mit $x_1, \ldots, x_n > 0$, nicht alle gleich, und $\lambda_1, \ldots, \lambda_n > 0$, $\sum \lambda_i = 1$ gilt also

$$(\lambda_1 x_1 + \cdots + \lambda_n x_n)^p < \lambda_1 x_1^p + \cdots + \lambda_n x_n^p. \tag{$*$}$$

Setzen wir $\frac{1}{p} + \frac{1}{q} := 1$, so dass $pq = p + q$, $q = \frac{p}{p-1}$, $p = \frac{q}{q-1}$, $p - \frac{p}{q} = q - \frac{q}{p} = (p-1)(q-1) = 1$, ergibt sich mit *positiven* a_i und b_i sowie $x_i = \frac{a_i}{b_i^{q-1}}$, $\lambda_i = \frac{b_i^q}{b_1^q + \cdots + b_n^q}$

$$(a_1 b_1 + \cdots + a_n b_n)^p < \big(b_1^q + \cdots + b_n^q\big)^{p-1}\big(a_1^p + \cdots + a_n^p\big) \tag{31}$$

wegen $b_i^{q-p(q-1)} = b_i^0$; da $(p-1)/p = 1/q$, ist dies die strikte *Höldersche Ungleichung* $a_1 b_1 + \cdots + a_n b_n < \big(a_1^p + \cdots + a_n^p\big)^{1/p}\big(b_1^q + \cdots + b_n^q\big)^{1/q}$ für *nicht proportionale* positive Vektoren $(a_1^p, \ldots, a_n^p)$ und $(b_1^q, \ldots, b_n^q)$ $(q - 1 = q/p)$.

Setzen wir stattdessen $x_k = \frac{a_{ik} w_i^{1/p}}{\|\mathbf{a}_k\|_p}$ $(1 \leq k \leq n)$, wobei $\mathbf{a}_k = (a_{1k}, \ldots, a_{mk})$, $a_{ik} \geq 0$ und $\|\mathbf{a}_k\|_p := \Big(\sum_{i=1}^m a_{ik}^p w_i\Big)^{1/p}$ mit $w_i > 0$ in $(*)$, ergibt sich

$$\Big(\lambda_1 \frac{a_{i1}}{\|\mathbf{a}_1\|_p} + \cdots + \lambda_n \frac{a_{in}}{\|\mathbf{a}_n\|_p}\Big)^p w_i < \lambda_1 \frac{a_{i1}^p w_i}{\|\mathbf{a}_1\|_p^p} + \cdots + \lambda_n \frac{a_{in}^p w_i}{\|\mathbf{a}_n\|_p^p}.$$

Summation $i = 1, \ldots, m$: $\sum_{i=1}^m \Big(\lambda_1 \frac{a_{i1}}{\|\mathbf{a}_1\|_p} + \cdots + \lambda_n \frac{a_{in}}{\|\mathbf{a}_n\|_p}\Big)^p w_i < 1$.
Setzt man nun noch $\lambda_k = \|\mathbf{a}_k\|_p / \sum_{j=1}^n \|\mathbf{a}_j\|_p$, erhält man

$$\|\mathbf{a}_1 + \cdots + \mathbf{a}_n\|_p < \|\mathbf{a}_1\|_p + \cdots + \|\mathbf{a}_n\|_p, \tag{32}$$

die *verallgemeinerte Minkowskische Ungleichung*.
(Diese *Anwendung* ist um einiges raffinierter als die Jensensche Ungleichung selbst.

Die von F. Riesz stammende Rückführung der Minkowskischen auf die Höldersche Ungleichung ist weit eleganter und kürzer; die Schlüssel-Umformung dabei ist $(a_i+b_i)^p w_i = a_i w_i^{1/p}(a_i+b_i)^{p-1}w_i^{1/q} + b_i w_i^{1/p}(a_i+b_i)^{p-1}w_i^{1/q}, \dots$)

Satz von C. Carathéodory: *Sei* $\mathcal{M} \subseteq \mathbb{R}^n$ *und* $x_0 \in \mathcal{M}$. *Dann besteht die konvexe Hülle* $\mathrm{Co}(\mathcal{M})$ *aus der Gesamtheit aller Konvex-Kombinationen von* x_0 *mit beliebigen weiteren höchstens* n *Punkten aus* $\mathcal{M}$*:*

$$\mathrm{Co}(\mathcal{M}) = \left\{ \lambda_0 \mathbf{x}_0 + \sum_{i=1}^{n} \lambda_i \mathbf{x}_i \;\middle|\; \mathbf{x}_i \in \mathcal{M}\,(1 \le i \le n),\ \lambda_i \ge 0\,(0 \le i \le n),\ \sum_{i=0}^{n} \lambda_i = 1 \right\}$$

Beweis:
Es ist nur zu zeigen: Jede Konvex-Kombination von mindestens $n{+}2$ Punkten aus $\mathcal{M}$ kann durch eine Konvex-Kombination aus *weniger* Punkten ersetzt werden.

Sei eine Konvex-Kombination $\mathbf{x} = \lambda_0\mathbf{x}_0 + \lambda_1\mathbf{x}_1 + \cdots + \lambda_{n+1}\mathbf{x}_{n+1}$ mit $\mathbf{x}_i \in \mathcal{M}\,(1 \le i \le n{+}1)$ gegeben. Da die Punkte $\mathbf{x}_i - \mathbf{x}_0\ (1 \le i \le n{+}1)$ insgesamt linear abhängig sind, gibt es eine nichttriviale (nicht alle $\mu_i = 0$) verschwindende Linearkombination $\sum_{i=1}^{n+1} \mu_i(\mathbf{x}_i - \mathbf{x}_0) = \mathbf{0}$, so dass also

$$\mu_0\mathbf{x}_0 + \mu_1\mathbf{x}_1 + \cdots + \mu_{n+1}\mathbf{x}_{n+1} = \mathbf{0}$$

mit $\mu_0 := -\mu_1 - \cdots - \mu_{n+1},\ \sum_{i=0}^{n+1} \mu_i = 0$, nicht alle $\mu_i = 0$.

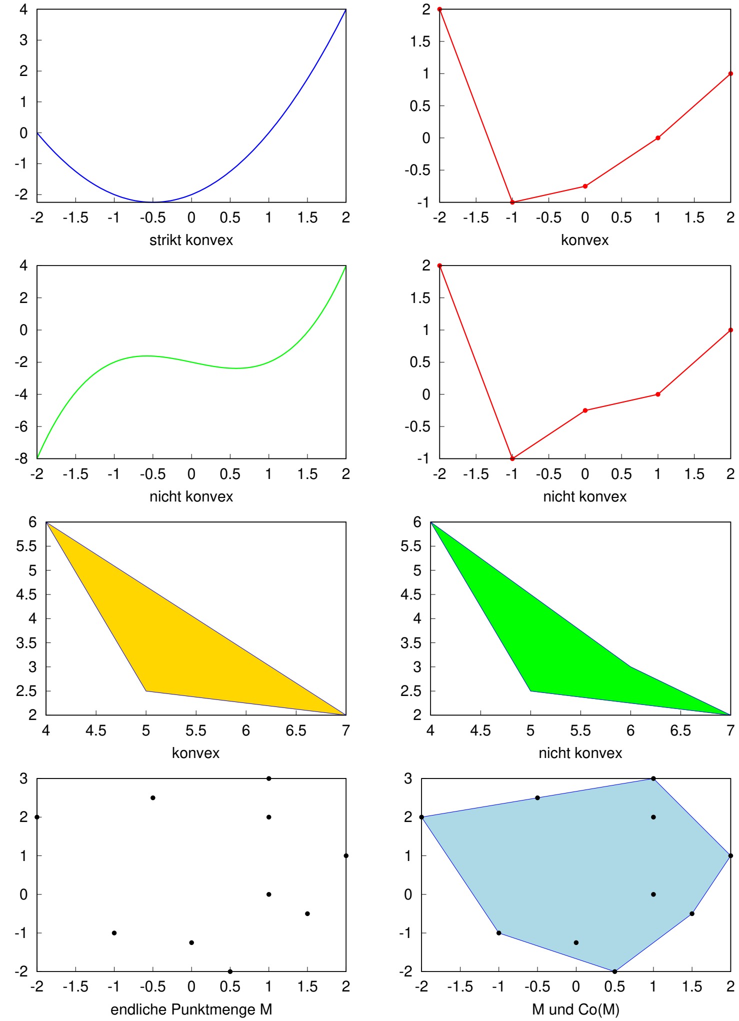
O.B.d.A. gelte $\lambda_1, \lambda_2, \dots, \lambda_{n+1} > 0$ (anderenfalls bleibt ja nichts zu zeigen). Dann gilt für kleine $|t|$ auch noch $\lambda_i + t\mu_i > 0\,(1 \le i \le n+1)$.

Gilt $\mu_0 > 0$, ist mindestens ein anderes μ_i negativ. Wir wählen das kleinste $t > 0$, für das mindestens ein Ausdruck $\lambda_i + t\mu_i$ mit $i \ge 1$ verschwindet. Dann ist

$$\mathbf{x} = (\lambda_0 + t\mu_0)\mathbf{x}_0 + \sum_{i=1}^{n+1} (\lambda_i + t\mu_i)\mathbf{x}_i$$

eine Konvex-Kombination der behaupteten Art, da ja mindestens ein Summand wegfällt. Im Falle $\mu_0 < 0$ macht man es analog, aber mit negativem t. Im Fall $\mu_0 = 0$ schließlich muss ja trotzdem ein anderes μ_i von 0 verschieden sein, und man kann wieder ein passendes t wählen. ■

Zum Schluss ein paar anschauliche Grafiken zu konvexen Funktionen und Mengen:



Konvexe Hüllen endlicher Punktmengen der Ebene: *konvexe Polygone*; des Raumes: *konvexe Polytope*. (Klassiker: B. Grünbaum, *Convex Polytopes, 2nd ed.*, Springer 2003)

5 Einige Literaturhinweise

Es gibt viele lesenswerte Bücher zur Approximationstheorie. Einige wenige, die relativ viel über Bernstein-Polynome enthalten, seien genannt:

[1] Isidor P. Natanson: *Konstruktive Funktionentheorie*, Akademie-Verlag 1955

[2] Philip J. Davis: *Interpolation and Approximation*, Dover 1975

[3] George G. Lorentz: *Bernstein Polynomials, 2nd Ed.*, Chelsea 1986

[4] George G. Lorentz: *Approximation of Functions, 2nd Ed.*, Chelsea 1986

[5] Ronald A. DeVore, George G. Lorentz: *Constructive Approximation*, Springer 1993 (Grundlehren der mathematischen Wissenschaften 303)

In [1] sind alle Beweise sehr ausführlich ausgearbeitet und für Leserinnen und Leser mit Grundkenntnissen in Analysis verständlich; aber auch [2], ursprünglich 1963 erschienen und immer noch aktuell, ist für Mathematik-Studierende mittlerer Semester *sehr* zu empfehlen. Die Bücher [3], [4], [5], alle drei inhaltlich schwergewichtige Standardwerke (obwohl die ersten beiden äußerlich nur schmal sind), sind insgesamt recht anspruchsvoll.

Zum Thema Konvexität, in der gesamten Mathematik von überragender Bedeutung, seien nur die folgenden Standardreferenzen genannt. Die ersten beiden sind gut lesbare Einführungen mit jeweils umfangreichen Literaturverzeichnissen. Das dritte setzt insgesamt mathematisch sehr fortgeschrittene Leserinnen und Leser voraus, behandelt neuere Anwendungen „höherer“ Konvexitätsbegriffe in der Theorie der partiellen Differentialgleichungen und der komplexen Analysis; aber die ersten gut 100 (von insgesamt ca. 400) Seiten stellen auf brillante Weise *alle* wesentlichen klassischen Resultate dar und sind leichter zu verdauen als der Rest. Hörmander (1931–2012), Träger der Fields-Medaille (1962), des Wolf-Preises (1988) und des Steele-Preises (2006), war einer der bedeutendsten Mathematiker des 20. Jahrhunderts.

[6] Frederick A. Valentine: *Konvexe Mengen*, Bibliogr. Institut 1968

[7] A. Wayne Roberts, Dale E. Varberg: *Convex Functions*, Academic Press 1973

[8] Lars Hörmander: *Notions of Convexity*, Birkhäuser 1994