Woher kommen die p-Normen?

Günther J. Wirsching

21. Juli 2011

1 Die Formel

Auf dem N-dimensionalen Vektorraum $\mathbb{R}^N$ ist die *p-Norm* der Vektoren wie folgt gegeben:

$$\forall \quad \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \in \mathbb{R}^N : \qquad \|\vec{x}\|_p := \left(\sum_{n=1}^{N} |x_n|^p \right)^{1/p}. \tag{1}$$

Sofern p eine reelle Zahl > 0 ist, ist diese Formel ohne weiteres auswertbar. Das Thema dieser Note ist, schlaglichtartig ein paar Hintergründe, mathematische Eigenschaften und Weiterentwicklungen dieser Formel zu beleuchten.

2 Minkowski

Hermann Minkowski wurde am 22. Juni 1864 in dem kleinen Ort Aleksotas auf dem Gebiet des heutigen Staates Litauen, damals zu Russland gehörend, geboren. Von 1887 bis 1909 lehrte er den Universitäten Bonn, Königsberg, Zürich und Göttingen, wo er am 12. Januar 1909 an den Folgen eines Blinddarmdurchbruchs verstarb.

Der entscheidene Term von Formel (1) taucht bei Minkowski auf, und zwar in seinem Buch *Geometrie der Zahlen* [1] auf Seite 115, in folgender Form:

$$\left(\frac{(\mathrm{abs}\, v_1)^p + \cdots + (\mathrm{abs}\, v_n)^p}{n} \right)^{\frac{1}{p}} = f(x_1, \ldots, x_n); \tag{2}$$

dabei hat er vorausgesetzt:

> *Es seien $v_1, \ldots, v_n$ n lineare Formen in $x_1, \ldots, x_n$* $\quad \ldots$

Was er unter einer *linearen Form* versteht, hat er ein paar Seiten vorher, zu Beginn seines vierten Kapitels, beschrieben [1, Seite 102]:

> ... *von linearen Formen in $x_1, \ldots, x_n$, also Ausdrücke von der Gestalt*
>
> $$\alpha_1 x_1 + \ldots + \alpha_n x_n,$$
>
> *und zwar durchweg mit lauter reellen Coefficienten;*

Also ist Formel (1) ein Spezialfall von Minkowskis Formel (2): Nimmt man als i-te Linearform ein geeignetes Vielfaches der i-ten Koordinate, $v_i := \sqrt[p]{n}\, x_i$, bezeichnet den Absolutbetrag gemäß heutiger Schreibweise mit zwei senkrechten Strichen, und ersetzt noch das kleine n durch ein großes N, so ist $f(x_1, \ldots, x_n)$ aus Formel (2) gleich $\|\vec{x}\|_p$ aus Formel (1).

Warum ist das vermutlich das erste Auftreten? Nun, der 1953 erschienene Druck seines Buchs *Geometrie der Zahlen* enthält eine Widmung:

HERRN
CHARLES HERMITE
ZUM SIEBZIGSTEN GEBURTSTAGE
IN GRÖSSTER VEREHRUNG
GEWIDMET
VOM
VERFASSER

Weil Hermite am 24. Dezember 1822 geboren wurde, sollte der Text um den 24. Dezember 1892 herum fertig geworden sein. Tatsächlich wurden Auszüge davon 1893 und 1896 publiziert; in den *Gesammelten Abhandlungen* [2] konnte ich die betreffende Stelle jedoch nicht finden. Es ist eher unwahrscheinlich, dass vorher schon jemand diese Formel untersucht haben, denn in welchem Zusammenhang sollte man das tun?

3 Die Minkowski'sche Ungleichung

Das erste, was Minkowski mit seiner in (2) definierten Funktion f macht, ist, die folgende Behauptung zu beweisen:

> *Es besteht nun ferner bei beliebigen reellen Werthen $a_1, \ldots, a_n;\ b_1, \ldots, b_n$ für ein $p \geq 1$ immer die Ungleichung*
>
> $$f(a_1 + b_1, \ldots, a_n + b_n) \leq f(a_1, \ldots, a_n) + f(b_1, \ldots, b_n),$$
>
> *und tritt darin im Falle eines $p > 1$ das Gleichheitszeichen nur dann ein, wenn entweder $b_1, \ldots, b_n$ sämmtlich Null sind oder aber Beziehungen*
>
> $$a_1 = (\tau - 1)b_1,\ \ldots\ a_n = (\tau - 1)b_n, \qquad \tau - 1 \geqq 0$$
>
> *gelten.* [1, Seite 115]

Also mit der in (1) eingeführten Schreibweise:

$$\forall\ p \in [1, \infty[,\ \vec{x}, \vec{y} \in \mathbb{R}^N : \begin{cases} \|\vec{x} + \vec{y}\|_p \leqslant \|\vec{x}\|_p + \|\vec{y}\|_p, & \text{und} \\ \|\vec{x} + \vec{y}\|_p = \|\vec{x}\|_p + \|\vec{y}\|_p & \Leftrightarrow \begin{cases} \vec{y} = 0 \quad \text{oder} \\ \exists \lambda \in [0, \infty[:\ \vec{x} = \lambda \vec{y}. \end{cases} \end{cases} \tag{3}$$

Diese Ungleichung ist heute unter dem Namen *Minkowski'sche Ungleichung* bekannt. Ihre Bedeutung ist, dass sie die Dreiecksungleichung für die Abbildung

$$\| \cdot \|_p : \mathbb{R}^N \to \mathbb{R} \tag{4}$$

ist, weswegen diese Abbildung zu einer *Norm* auf dem $\mathbb{R}^N$ wird. Diese Norm wird gelegentlich als *Minkowski-Norm* bezeichnet, obwohl dieser Ausdruck wegen der Minkowski'schen Untersuchungen zum *Minkowski-Raum* der speziellen Relativitätstheorie etwas irreführend aufgefasst werden könnte. In der Regel wird die Abbildung (4) als *p-Norm* bezeichnet – meist wird nur der Fall $p \geqslant 1$ betrachtet, wo es sich tatsächlich um eine *Norm* handelt. Wenn auch Werte $p < 1$ in Betracht kommen, ist eben eine *p-Norm* nicht unbedingt auch eine *Norm* (auf einem Vektorraum).

4 Mathematische Erweiterungen und Spezialfälle

Es ist irgendwie naheliegend, zu untersuchen, was passiert, wenn man für p noch andere Werte zulässt.

$p = \infty$ **.** Mit Hilfe einfacher Abschätzungen folgt aus (1) für alle $p \geqslant 1$ die Ungleichungskette

$$\max_{n \in \{1,\ldots,N\}} |x_n| \leqslant \|\vec{x}\|_p \leqslant \sqrt[p]{N} \cdot \max_{n \in \{1,\ldots,N\}} |x_n|; \tag{5}$$

daher ist folgende Bezeichnung sinnvoll:

$$\|\vec{x}\|_\infty := \lim_{p \to \infty} \|\vec{x}\|_p = \max_{n \in \{1,\ldots,N\}} |x_n|.$$

Weil die Minkowski'sche Ungleichung für alle $p \geqslant 1$ gültig ist, muss sie auch für die Grenzwert für $p \to \infty$ gelten, weswegen durch

$$\|\cdot\|_\infty : \mathbb{R}^N \to \mathbb{R} \tag{6}$$

auch eine *Norm* auf dem $\mathbb{R}^N$ gegeben ist. Diese Norm heißt meistens *Maximumsnorm* und manchmal auch *Tschebyschew-Norm*.

$p = 2$ **.** Das ist die bekannte *Euklidische Norm* auf dem $\mathbb{R}^N$.

$p = 1$ **.** Die Norm $\|\cdot\|_1$ heißt auch *Betragssummennorm*, oder *City-Block-Norm*, und spielt in vielen Anwendungen eine große Rolle.

$0 < p < 1$ **.** Für diese p ist die in (3) formulierte Aussage falsch, man erhält also keine *Norm*. Gegenbeispiele:

- Die Dreiecksungleichung gilt nicht, denn für $0 < p < 1$ ist:

$$\vec{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \|\vec{x} + \vec{y}\|_p = 2^{\frac{1}{p}} > 2 = 1 + 1 = \|\vec{x}\|_p + \|\vec{y}\|_p.$$

- Die umgekehrte Dreicksungleichung für nicht-kollineare Vektoren gilt auch nicht, denn für $0 < p < 1$ ist:

$$\vec{x} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \quad \Rightarrow \quad \|\vec{x} + \vec{y}\|_p = 4 < (5^p + 1)^{\frac{1}{p}} + 2^{\frac{1}{p}} = \|\vec{x}\|_p + \|\vec{y}\|_p.$$

Allerdings ist die durch

$$\widetilde{d}_p(\vec{x}, \vec{y}) := \sum_{n=1}^{N} |x_n - y_n|^p = \|\vec{x} - \vec{y}\|_p^p$$

definierte Abbildung auf $\mathbb{R}^N \times \mathbb{R}^N$ eine Metrik, weswegen es zu diesem Fall auch Fachartikel gibt.

$p = 0$ **.** Analog zum Fall $p = \infty$ kann man hier Grenzwerte betrachten: Wegen

$$\forall\, \xi \in [0, \infty[: \quad \lim_{p \downarrow 0} \xi^p = \begin{cases} \xi^0 = 1 & \text{für } \xi > 0, \\ 0 & \text{für } \xi = 0, \end{cases}$$

ist der Grenzwert

$$\lim_{p \downarrow 0} \|\vec{x}\|_p^p = \lim_{p \downarrow 0} \left(|x_1|^p + \ldots + |x_n|^p\right) \tag{7}$$

gleich der Anzahl der nicht-verschwindenden Komponenten von $\vec{x} = (x_1, \ldots, x_n)^T$. In manchen Arbeiten zu Themen in den Bereichen *Maschinelles Lernen* oder *Optimierung* wird dieser Grenzwert auch als *zero norm* bezeichnet – obwohl er keine *Norm* ist.

Aus (7) folgt

$$\lim_{p \downarrow 0} \|\vec{x}\|_p = \begin{cases} \infty & \text{falls } \vec{x} \text{ mindestens zwei Komponenten} \neq 0 \text{ besitzt,} \\ \|\vec{x}\|_1 = \|\vec{x}\|_\infty & \text{falls } \vec{x} \text{ höchstens eine Komponente} \neq 0 \text{ hat.} \end{cases}$$

$p < 0$ **.** Für $p < 0$ setzt man sinnvollerweise $0^p := \infty$ und $\infty^p := 0$. Dann lässt sich die Formel (1) auch für $p < 0$ auswerten. Man erhält insbesondere das folgende Resultat:

$$\text{Ist } p < 0 \text{ und ein } x_n = 0, \text{ dann ist } \|\vec{x}\|_p = 0. \tag{8}$$

- Die Dreiecksungleichung gilt nicht, denn für $p < 0$ ist:

$$\vec{x} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \quad \Rightarrow \quad \|\vec{x} + \vec{y}\|_p = 4 > (5^p + 1)^{\frac{1}{p}} + 2^{\frac{1}{p}} = \|\vec{x}\|_p + \|\vec{y}\|_p.$$

- Die umgekehrte Dreicksungleichung für nicht-kollineare Vektoren gilt auch nicht, denn für $p < 0$ ist:

$$\vec{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \|\vec{x} + \vec{y}\|_p = 2^{\frac{1}{p}} < 2 = 1 + 1 = \|\vec{x}\|_p + \|\vec{y}\|_p.$$

$p = -\infty$ **.** Analog zu (5) verifiziert man leicht die für $p \leqslant -1$ gültige Ungleichungskette

$$\sqrt[p]{N} \cdot \min_{n \in \{1,\ldots,N\}} |x_n| \leqslant \|\vec{x}\|_p \leqslant \min_{n \in \{1,\ldots,N\}} |x_n|, \tag{9}$$

woraus die Konvergenz von $\|\vec{x}\|_p$ für $p \to -\infty$ folgt. Also kann man setzen:

$$\|\vec{x}\|_{-\infty} := \lim_{p \to -\infty} \|\vec{x}\|_p = \min_{n \in \{1,\ldots,N\}} |x_n|.$$

Natürlich erhält man auf diese Weise keine *Norm* auf dem $\mathbb{R}^N$.

5 L^p-Räume

Richtig Karriere gemacht haben die p-Normen durch ihre (jedenfalls für F. Riesz) naheliegende Erweiterung auf integrierbare Funktionen: Seien zunächst eine Menge X, eine σ-Algebra $\mathcal{A}$ auf X und ein Maß $\mu : \mathcal{A} \to \mathbb{R}$ gegeben – damit ist definiert, was eine *meßbare* bzw. *integrierbare* Funktion ist. Hat eine meßbare Funktion $f : X \to \overline{\mathbb{R}}$ die Eigenschaft dann ist ihre p-Norm wohldefiniert:

$$\|f\|_p := \left(\int_X |f|^p \, d\mu \right)^{\frac{1}{p}}. \tag{10}$$

Was ist die Verbindung zu Minkowski's Formel (2) ? Nun, setzt man in Formel (10)

$$X := \{1, \ldots, n\} \qquad \text{und} \qquad \mu = \text{Gleichverteilungsmaß auf } X,$$

so geht die rechte Seite von (10) in die linke Seite von (2) mit $f(i)$ anstelle von v_i über. Das klingt so gesehen zwar etwas weit hergeholt, aber die Grundstruktur

„die p-te Wurzel aus dem Mittelwert der p-ten Potenzen der Beträge“

ist doch die gleiche.

Zurück zu Formel (10). Es stellt sich heraus, dass

(a) die Menge der meßbaren Funktionen $f : X \to \overline{\mathbb{R}}$ mit der Eigenschaft, dass $|f|^p$ integrierbar ist, ein linearer Raum (also ein *Vektorraum*) ist,

(b) im Falle $p \geqslant 1$ durch $\|\cdot\|_p$ eine *Halbnorm* auf diesem Raum gegeben ist, und

(c) die Gleichung $\|f\|_p = 0$ genau dann richtig ist, wenn f außerhalb einer μ-Nullmenge verschwindet.

Für die zweite Aussage muss man die Dreiecksungleichung $\|f+g\|_p \leqslant \|f\|_p + \|g\|_p$ beweisen; diese verallgemeinert die oben erwähnte Minkowski'sche Ungleichung – häufig wird diese Verallgemeinerung ebenfalls als *Minkowski'sche Ungleichung* bezeichnet.

Mathematisch interessant sind sicherlich Aussagen über die algebraischen Strukturen von Mengen integrierbarer Funktionen. Zu diesem Thema hat F. Riesz 1910 einen längeren Aufsatz [3] und ein dazu passende *Comptes Rendues*-Note [4] geschrieben, die folgendermaßen beginnt:

> *Appelons classe $[L^p]$ la totalité des fonctions $f(x)$, réelles ou non, définies sur l'intervalle (a, b), sommable et telles que $|f|^p$ est sommable.*

Das ist sehr wahrscheinlich die erste Verwendung der Bezeichnung L^p in der mathematischen Fachliteratur. Der Text des Aufsatzes [3] legt die Vermutung nahe, dass der Buchstabe L zu Ehren von *Henri Lebesgue* gewählt wurde. Die Wahl des Buchstaben p für den Exponenten bedarf bei Riesz keiner weiteren Erklärung; Minkowski hatte ja einige Jahre früher auch schon den Exponenten mit p bezeichnet.

Zurück zu fachlichen Inhalten. Es ist leicht zu zeigen, dass man wieder eine integrierbare Funktion erhält, wenn man eine integrierbare Funktion mit einem Skalar multipliziert, oder wenn man zwei integrierbare Funktionen addiert. Aber unter welchen zusätzlichen Bedingungen kann man zwei integrierbare Funktionen (punktweise) multplizieren, ohne dass die Integrierbarkeit verloren geht? In [3] hat Riesz darauf die folgende gefunden:

> *Danach ist das Produkt der Funktionen $f(x)$, $h(x)$ sicher integrierbar, wenn es eine Zahl $p > 1$ gibt, derart daß die Funktionen $|f(x)|^p$, $|h(x)|^{\frac{p}{1-p}}$ integrierbar ausfallen.*

Darüber hinaus fand er auch eine Art Umkehrung:

> *Ist das Produkt $f(x), h(x)$ für alle integrierbaren Funktionen $f(x)$, für welche die Potenz $|f(x)|^p$ $(p > 1)$ integrierbar ist, ebenfalls integrierbar, so ist es auch die Potenz $|h(x)|^{\frac{p}{p-1}}$.*

Literatur

[1] Hermann Minkowski. *Geometrie der Zahlen*. Chelsea Publ. Co., New York, 1953.

[2] Hermann Minkowski. *Gesammelte Abhandlungen*. Chelsea Publ. Co., New York, 1967.

[3] Friedrich Riesz. Untersuchungen über Systeme integrierbarer Funktionen. *Math. Annalen*, 69:449–497, 1910.

[4] Frédéric Riesz. Sur certains systèmes d'équations fonctionelles et l'approximation des fonctions continues. *Comptes Rendues*, 150:674–677, 1910.