

Fehler korrigierende Codes

Felix Ulmer

Institut de Recherche Mathématiques de Rennes

IRMAR, UMR 6625 du CNRS

Université de Rennes 1

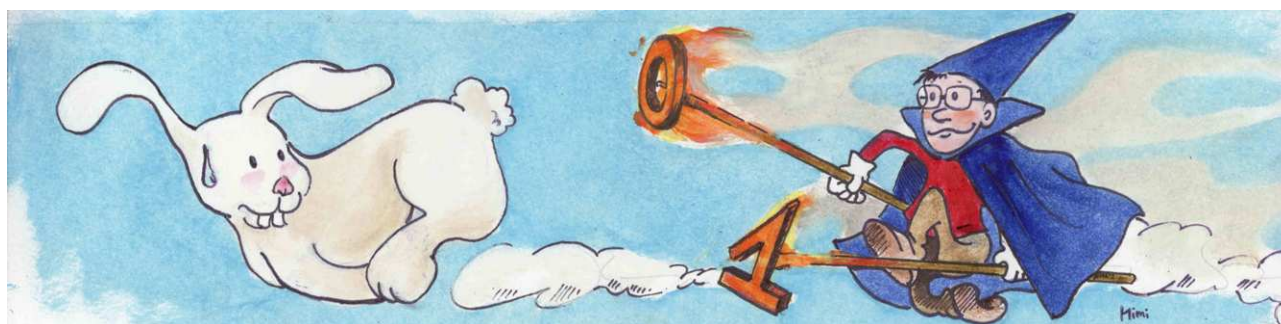
Campus de Beaulieu

F-35042 Rennes (Frankreich)

<http://perso.univ-rennes1.fr/felix.ulmer>



Fehler korrigierende Codes



1 Fehler bzw. Fälschungen erkennen

Codes gibt es im elektronischen Zeitalter fast überall: CD's, Festplatten, Digitale Bilder,...

Die Rückseite eines Euroscheins enthält ein Codewort der Länge 12 : Links einen Buchstaben, gefolgt von 11 Ziffern (z.B. U24263273615). Der Buchstabe gibt Aufschluss über das Herkunftsland (X für Deutschland, U für Frankreich,...). Wenn man den Buchstaben durch seine Reihenfolge im Alphabet ersetzt (also $X = 24, U = 21$), muss die so gewonnene Zahl z (z.B. 2124263273615) stets Rest 8 bei der Division durch 9 ergeben.

Die Zahlen $z + 1$ und $z - 8$ sind genau dann durch 9 teilbar, wenn die Summe ihrer Ziffern durch 9 teilbar ist. Dieser Prozess kann *rekursiv* angewendet werden.

2 Fehler korrigieren: Der Zaubertrick

Der Zauberer *Hamming* bittet einen Zuschauer, sich eine ganze Zahl Z mit $0 \leq Z \leq 15$ auszudenken. Er stellt dem Zuschauer nun 7 Fragen. Bei der Beantwortung der Fragen darf der Zuschauer höchstens einmal lügen, muss aber nicht.

- Frage 1 = ist $Z \geq 8$?
- Frage 2 = ist Z in der Menge $\{4, 5, 6, 7, 12, 13, 14, 15\}$?
- Frage 3 = ist Z in der Menge $\{2, 3, 6, 7, 10, 11, 14, 15\}$?
- Frage 4 = ist Z eine ungerade Zahl ?
- Frage 5 = ist Z in der Menge $\{1, 2, 4, 7, 9, 10, 12, 15\}$?
- Frage 6 = ist Z in der Menge $\{1, 2, 5, 6, 8, 11, 12, 15\}$?
- Frage 7 = ist Z in der Menge $\{1, 3, 4, 6, 8, 10, 13, 15\}$?

Die Schwierigkeit liegt natürlich darin, dass der Zuschauer bei einer beliebigen Frage gelogen haben kann, oder auch nicht. Man kann beweisen, dass es unter diesen Umständen unmöglich ist, diesen Trick mit weniger als 7 Fragen durchzuführen.

Nun soll $f_i = 0$ sein, wenn die Antwort auf die i -te Frage Nein ist, und $f_i = 1$, wenn die Antwort auf die i -te Frage Ja ist. Hamming berechnet nun folgende Zahlen modulo 2 (das heißt, er teilt die gewonnenen Zahlen durch 2 und betrachtet den Rest):

$$\begin{aligned} b_1 &= f_4 + f_5 + f_6 + f_7 \pmod{2} \\ b_2 &= f_2 + f_3 + f_6 + f_7 \pmod{2} \\ b_3 &= f_1 + f_3 + f_5 + f_7 \pmod{2} \end{aligned}$$

Er bekommt so die binäre Zahl $T = \overline{b_1 b_2 b_3}^2$, also die in Basis zwei geschriebene ganze Zahl

$$T = b_1 \cdot 2^2 + b_2 \cdot 2^1 + b_3 \cdot 2^0 \leq 7.$$

Ist T gleich Null, so wurde nie gelogen, sonst wurde genau bei der Frage Nummer T gelogen. Falls gelogen wurde, so korrigiert man nun die Antwort auf die Frage T . Die ausgedachte Zahl ist die (ggfs. korrigierte) binäre Zahl $Z = \overline{f_1 f_2 f_3 f_4}^2$, also die Zahl $Z = f_1 \cdot 2^3 + f_2 \cdot 2^2 + f_3 \cdot 2 + f_4$.

Beispiel 1 Wenn sich der Zuschauer die Zahl 6 ausdenkt und bei der dritten Frage lügt, so lauten seine 7 Antworten 0100011. Daher ergibt sich $b_3 = 0, b_2 = 1$ und $b_1 = 1$. Hamming berechnet $\overline{011}^2 = 3$ und weiß nun, dass bei der dritten Frage gelogen wurde. Er korrigiert die Antworten zu 0110011. Er berechnet nun $\overline{0110}^2 = 6$.

3 Fehler korrigierende Codes

Hammings Trick liefert ein Mittel, um eine in Basis 2 geschriebene ganze Zahl $\overline{f_1 f_2 f_3 f_4}^2$ (also eine ganze Zahl zwischen 0 und $15 = 2^4 - 1$) so zu übertragen, dass der Empfänger, bei höchstens einem Übertragungsfehler, erkennen kann, ob es einen Fehler gab, und, falls ja, ihn zu korrigieren. Man sendet dafür statt $\overline{f_1 f_2 f_3 f_4}^2$ die Zahl $w = \overline{f_1 f_2 f_3 f_4 f_5 f_6 f_7}^2$, also eine Zahl $0 \leq w \leq 255 = 2^8 - 1$. Man beachte, dass die vier Antworten auf die ersten Fragen des Zauberers, eigentlich genau die binäre Darstellung der ausgedachten Zahl sind (wenn da nicht gelogen wurde). Die Anzahl der möglichen Übertragungen ist also 256, davon sind jedoch nur 16 Übertragungen korrekt, das heißt es gibt nur 16 Codewörter in unserem Code der Länge 7.

4 Der Wiederholungscode

Man kann ein Wort auch dadurch codieren, dass man es mehrmals wiederholt. Schickt man statt 1011 das Codewort 10111011, so schließt man beim Empfang auf einen Fehler, wenn beide Hälften nicht identisch sind. Schickt man statt 1011 nun 101110111011, und wird das Wort 100110111011 empfangen, so korrigiert man diesen einen ersichtlichen Fehler. Wenn bei der Übertragung höchstens einmal *gelogen* wird, so kann man diesen einen Fehler immer korrigieren. Hätten wir im obigen Zaubertrick einen *Wiederholungscode* verwendet, so hätten wir 12 statt 7 Fragen stellen müssen.

5 Das Prinzip des Korrigierens

Der *Hamming Abstand* $d(w_1, w_2)$ zwischen zwei Codewörtern w_1 und w_2 misst, in wie vielen Stellen sich zwei Codewörter eines Codes unterscheiden. Wenn der kleinste Abstand zwischen zwei beliebigen Codewörtern $d = 2e + 1$ ist, so lässt sich bei höchstens $d - 1$ Fehler ein fehlerhaftes Wort als solches erkennen und bei höchstens $e = (d - 1)/2$ Fehler sogar korrigieren. Hierbei wird als korrektes Wort stets, das *am nächsten gelegene* Wort angegeben. Beim Zaubertrick gibt es 16 Codewörter. Die Zahl 6 ergibt dort richtig codiert 0110011, und die Zahl 14 entspricht 1110000. Diese Zahlen haben den Abstand 3, weil sie sich in 3 Stellen unterscheiden. Empfängt man die Zahl 0100011, so stellt man fest, dass diese Zahl kein Codewort ist und ersetzt die Zahl durch das nächstgelegene Codewort 0110011. Der Zaubertrick ist deshalb so interessant, weil er, ohne alle Codewörter heran zu ziehen, direkt korrigieren kann. Zwei verschiedene Codewörter des dreimaligen Wiederholungscode haben natürlich auch Abstand 3, was beweist, dass ein solcher Code ebenfalls einen Fehler korrigieren kann. Der Hamming Code ist in dem Sinn optimal, dass man mindestens drei zusätzliche Symbole braucht, also die Länge 7, um eine vierstellige binäre Zahl so zu übertragen, dass man gegebenenfalls einen einzelnen Fehler korrigieren kann.

Wenn mehr als ein Fehler passiert, gibt es wahrscheinlich einen Übertragungsfehler, der unentdeckt bleibt.

6 Mathematik statt Zauberei



Die Welt der Informatik besteht aus Nullen und Einsen. In dieser Welt wollen wir immer noch addieren und multiplizieren können, und dazu begeben wir uns nun in eine Welt \mathbb{F}_2 , in der $1 + 1 = 0$ sein soll, in der also -1 gleich 1 ist. Konkret bedeutet dies für uns, dass wir immer modulo zwei rechnen: $1 + 1 = 2 = 0 \pmod{2}$. Polynome sind die Grundbausteine der *Computer-Algebra*. Deshalb wollen wir nun ein Codewort, wie 0110011, mit einem Polynom in der Variable X identifizieren :

$$0 \cdot X^6 + 1 \cdot X^5 + 1 \cdot X^4 + 0 \cdot X^3 + 0 \cdot X^2 + 1 \cdot X^1 + 1 \cdot X^0$$

Wie üblich schreiben wir einfach nur $X^5 + X^4 + X + 1$, wobei die Koeffizienten des Polynoms hier in \mathbb{F}_2 sind und nach den obigen Regeln addiert und multipliziert werden.

Die **Codierung** einer binären Zahl $6 = \overline{0110}^2$, also des Polynoms $X^2 + X$, kann nun (statt vieler Fragen) folgendermaßen erfolgen: Man multipliziert das zu übertragende Polynom dazu immer mit dem selben *Erzeuger-Polynom* $g = X^3 + X + 1$. Die Codierung von 6, also $X^2 + X$, ist dann

$$(X^2 + X)(X^3 + X + 1) = X^5 + X^4 + X^3 + X$$

(Man beachte, dass $X^2 + X^2 = (1+1)X^2 = 0 \cdot X^2 = 0$ ist). Die zu übermittelnde Zahl ist also 0111010.

Um das empfangene Wort $w = X^5 + X^4 + X^3 + X$ zu **entschlüsseln**, muss eine Division mit Rest des Polynoms $w = X^5 + X^4 + X^3 + X$ durch $g = X^3 + X + 1$ durchgeführt werden. Dies erfolgt im Prinzip wie bei ganzen Zahlen. Man betrachtet die höchsten Koeffizienten von w und g und überlegt nun, *wie oft geht g in w*, und es geht natürlich X^2 Mal. Daher zieht man X^2 Mal g , also $X^5 + X^3 + X^2$ von w ab und bekommt $w - X^2 \cdot g = X^4 + X^2 + X$ (man beachte, dass -1 gleich 1 ist). Schließlich *geht X mal g in w - X^2 · g* und man bekommt $(w - X^2 \cdot g) - X \cdot g = 0$. Also ist $w = (X^2 + X) \cdot g$, und wir erhalten wieder $X^2 + X$, also $6 = \overline{0110}^2$. Das Codieren entspricht einer Multiplikation und das Erkennen der Nachricht einer Division.

Hätten wir $\tilde{w} = X^5 + X^4 + X$ statt w empfangen, so hätten wir dank $\tilde{w} = (X^2 + X + 1) \cdot g + (X + 1)$ erkannt, dass der Rest nicht Null ist und somit ein Fehler aufgetreten ist. Diesen gilt es dann zu korrigieren.

Eine Eigenschaft des *Polynom-Codes* ist, dass die Summe zweier Codewörter (d.h. Polynome) wieder ein Codewort ist: der Code ist ein *linearer Code*. Dies folgt aus der Tatsache, dass die Summe zweier durch g teilbarer Polynome, wieder durch g teilbar ist. Diese lineare Eigenschaft hat zum Beispiel auch der *Hamming Code*. Die additive Eigenschaft erlaubt nun, schneller den minimalen Abstand zwischen zwei beliebigen Codewörtern zu ermitteln. Dieser entspricht der minimalen Anzahl von Einsen in einem von Null verschiedenen Codewort. Man spricht vom *Gewicht* des Codewortes, und so ist zum Beispiel das Gewicht von 0111010 einfach 4. Es genügt also, den Abstand zum Nullwort 0000000 zu messen.

Um dies zu beweisen beachte man, dass zwei Codewörter w_1 und w_2 genau dann den Abstand d haben, wenn es d Einsen im Codewort $w_1 - w_2$ gibt (hier werden die Polynome Koeffizientenweise substrahiert).

Beispiel 2 Da die Codewörter des Polynom-Codes in aufsteigender Reihenfolge 0000000, 0001011, 0010110, 0011101, 0101100, 0100111, 0111010, 0110001, 1011000, 1010011, 1001110, 1000101, 1110100, 1111111, 1100010, 1101001 sind, ergibt sich, dass der kleinste Abstand zwischen zwei Codewörtern 3 ist. Der *Polynom-Code* kann also ebenfalls einen Fehler korrigieren.

Es kann also mehr, als nur einen optimalen Code geben.

Unser Polynom-Code ist zudem noch ein *zyklischer Code*: genau dann ist $a_6a_5a_4a_3a_2a_1a_0$ ein Codewort, wenn $a_0a_6a_5a_4a_3a_2a_1$ ein Codewort ist.

7 Zauberlehrlinge der Mathematik

Zur Fehlerkorrektur des *Polynom-Codes* werden wir nun Mathematik benutzen, deren Begründung in der Regel im Rahmen eines Mathematikstudiums stattfindet. Ist man bereit einige Tatsachen vorab zu akzeptieren, so ist der Zugang jedoch recht einfach.

Als erstes wollen wir akzeptieren, dass in einer Welt, in der $1+1=0$ ist, es immer noch Nullstellen von Polynomen gibt. Falls α eine solche Nullstelle des *Erzeuger-Polynoms* $g = X^3 + X + 1$ ist, so ist $\alpha^3 + \alpha + 1 = 0$ oder, wegen $1 = -1$, auch $\alpha^3 = \alpha + 1$. Die Potenzen α und α^2 können nicht vereinfacht werden. Für höhere Potenzen ergeben sich jedoch folgende Gleichungen:

$$\begin{array}{ll} \alpha^3 &= \alpha + 1 & \alpha^6 &= \alpha^2 + 1 \\ \alpha^4 &= \alpha^2 + \alpha & \alpha^7 &= 1 \\ \alpha^5 &= \alpha^2 + \alpha + 1 & & \end{array}$$

Man kann nachrechnen, dass, in der $1+1=0$ Welt, auch α^2 und α^4 Wurzeln von $X^3 + X + 1$ sind. In der Tat ist in dieser Welt

$$X^7 - 1 = (X^3 + X^2 + 1)(X + 1)(X^3 + X + 1).$$

In der *Normalen Welt*, in der $1+1=2$ ist, rechnet man nach, dass dies nicht der Fall ist. In der Welt in der $1+1=0$ ist, passieren also merkwürdige Dinge. Wir wollen nun auch akzeptieren, dass in der neuen Welt $\alpha^i \neq 0$ ist.

Nun möchten wir unsere seltsame Welt benutzen, um einen Übertragungsfehler zu korrigieren. Wichtig hierbei ist, dass alle Codewörter w des Polynom-Codes Vielfache des Erzeuger-Polynoms g sind, und somit die Form $w = h \cdot g$ haben. Da wir annehmen, dass höchstens ein Fehler bei der Übertragung aufgetreten ist, ist das empfangene \tilde{w} entweder w oder $w + X^i$. Im letzteren Fall ist das richtige Wort eigentlich $\tilde{w} + X^i$ (immer wegen $1+1=0$). Wir schreiben $\tilde{w}(X) = h(X) \cdot g(X)$ oder $\tilde{w}(X) = h(X) \cdot g(X) + X^i$ um die Variable X hervorzuheben. Nun setzen wir die Nullstelle α von g in ein \tilde{w} ein. Falls $\tilde{w}(X) = h(X) \cdot g(X)$ ist, so ist $\tilde{w}(\alpha) = h(\alpha) \cdot g(\alpha) = h(\alpha) \cdot 0 = 0$, und falls $\tilde{w}(X) = h(X) \cdot g(X) + X^i$, so ist $\tilde{w}(\alpha) = \alpha^i \neq 0$. Man kann durch Einsetzen also sofort erkennen, ob die Übertragung korrekt oder fehlerhaft ist.

Beispiel 3 (Korrektur von $\tilde{w} = X^5 + X^4 + X$)

Da $1 \cdot \alpha + 1 \cdot \alpha = 1 \cdot \alpha + (-1) \cdot \alpha = 0$ und $\alpha^2 + \alpha^2 = 0$, gilt

$$\begin{aligned} \tilde{w}(\alpha) &= \alpha^5 + \alpha^4 + \alpha \\ &= (\alpha^2 + \alpha + 1) + (\alpha^2 + \alpha) + \alpha \\ &= \alpha + 1 \end{aligned}$$

Es gibt also einen Fehler der Form X^i mit $\alpha^i = \alpha + 1$. Unsere obigen Berechnungen zeigen, dass $i = 3$ sein muss, und der Fehler also X^3 ist. Das korrekte Wort ist somit $w = \tilde{w} + X^3 = (X^5 + X^4 + X) + X^3 = X^5 + X^4 + X^3 + X$. Nun ergibt sich aus der Division, dass $w = (X^2 + X) \cdot g$ ist. Daher war das gesendete Wort $X^2 + X$, was der Zahl $6 = \overline{0110}^2$ entspricht.

8 Die Tricks des Mathematikers

Im Gegensatz zu Zauberern, verraten Mathematiker ihre Tricks nur zu gerne, hier ausnahmsweise ohne Beweise. Wichtig beim Polynom-Code ist, dass, in der Welt, in der $1+1=0$ ist, der Erzeuger-Polynom g ein Teiler von $X^7 - 1$ ist. Den Abstand **3** zwischen zwei Codewörtern erreicht man dann dadurch, dass die $2 = \mathbf{3} - 1$ Potenzen α und α^2 Nullstellen von g sind.

Einen Polynom-Code der Länge 15, der bei der Übertragung einer 7 stelligen binären Zahl maximal $2 = (\mathbf{5} - 1)/2$ Fehler korrigieren kann, erreicht man nun mit dem Erzeuger-Polynom $\tilde{g} = X^8 + X^7 + X^6 + X^4 + 1$. Wenn $1+1=0$ ist, so ist \tilde{g} ein Teiler von $X^{15} - 1$, und wenn $\tilde{\alpha}$ eine Nullstelle des Teilers $X^4 + X + 1$ von \tilde{g} ist, sind auch die $4 = \mathbf{5} - 1$ Potenzen α , α^2 , α^3 , α^4 Nullstellen von \tilde{g} . Codieren und Decodieren erfolgen wieder durch Multiplikation und Division. Das Korrigieren kann man in Beispiel 20.5 in [1] nachlesen.

Literaturverzeichnis

- [1] Lidl, R., Pilz, G., 1998. Applied abstract algebra, Springer Verlag.
- [2] *Codes correcteurs d'erreurs*, Agrégation externe de mathématiques, 2005, Épreuve de modélisation.

<http://agreg.dnsalias.org/Textes/527.pdf>