Datenkompression (Rote), SS 2003, 4. Übungsblatt
Abgabe: bis Freitag, 16. 5. 2003, in der Vorlesung

18. Optimale Kodes als ganzahliges lineares Optimierungsproblem (20 Punkte)

    (a) (5 Punkte) B_l und I_l sei die Anzahl der Blätter und der inneren Knoten auf Tiefe l in einem binären Kode-Baum. Zeigen Sie, dass diese Zahlen durch folgende Gleichungen charakterisiert sind.

    $$B_0 + I_0 = 1, \text{ und } B_l + I_l = 2I_{l-1} \text{ für } l = 1, 2, \ldots$$

    (b) (5 Punkte) R_l ist die Anzahl der Blätter auf Tiefe l und tiefer:

    $$R_0 = n, \quad R_l = R_{l-1} - B_{l-1} \text{ für } l = 1, 2, \ldots$$

    Das Gewicht (die mittlere gewichtete Pfadlänge) eines optimalen Baumes mit den Parametern n, B_l und I_l für die Gewichte $p_1 \leq p_2 \leq \cdots \leq p_n$ ist durch

    $$\sum_{l=0}^{\infty} f(R_l)$$

    gegeben, wobei $f(i) = p_1 + p_2 + \cdots + p_i$ ist.

    (c) (5 Punkte) Das Gewicht kann man auch durch folgende Zielfunktion ausdrücken:

    $$\text{minimiere } \sum_{l=0}^{\infty} y_l$$

    unter den Bedingungen

    $$y_l \geq p_1 + p_2 + \cdots + p_i + p_{i+1}(R_l - i), \text{ für } i = 0, 1, \ldots, n-1 \text{ und } l \geq 0$$

    Hinweis: Zeichnen Sie ein Diagramm für die Beziehung zwischen y_l und R_l.

    (d) (5 Punkte) Manchmal möchte man einen *längenbeschränkten* optimalen Kode, einen Kode, bei dem alle Kodewörte durch eine vorgegebene Länge L beschränkt sind. Durch welche zusätzlichen Bedingungen lässt sich das obige Modell für diesen Fall anpassen?

19. Optimale Kodes mit zusätzlicher Nebenzielfunktion (0 Punkte)
    Manchmal gibt es mehrere optimale Kodes, die die Zielfunktion $\sum_{i=1}^{n} p_i l_i$ minimieren. Dann sucht man unter den optimalen Kodes einen, der möglichst keine allzu langen Kodewörter enthält. (Siehe auch Aufgabe 14.) Eine Möglichkeit ist, unter den optimalen Kodes jenen auszuwählen, der einfach die *ungewichtete* mittlere Länge, d.h, die Zielfunktion $\sum_{i=1}^{n} l_i$ minimiert. Dieses Problem kann man durch die Zielfunktion

    $$\text{minimiere } \sum_{i=1}^{n} p_i l_i + \varepsilon \cdot \left(\sum_{i=1}^{n} l_i \right)$$

    ausdrücken, für ein genügend kleines $\varepsilon > 0$.

    (a) Wie klein muss man ε wählen, damit dieser Ansatz funktioniert?

    (b) Wie kann man dieses Problem durch einen geeignete Variante des Huffman-Algorithmus lösen?

    (c) (Zusatzfrage, 10 Punkte) Wird durch die Optimallösung automatisch auch das *längste* Kodewort minimiert? (Geben Sie einen Beweis oder ein Gegenbeispiel.)