**Hausaufgaben zur Vorlesung** | **SoSe 2016**
**Berechenbarkeit und Komplexität** | **TU Berlin**
**Niedermeier/Chen/Froese/Sorge** | **08.07.2016**

# 2. Hausaufgabenblatt

Abgabe **im Sekretariat TEL 509B** jeweils
**20.07., 10:00-13:00**, oder **21.07., 10:00-13:00**, oder **22.07., 09:00-10:00**
**Alle Antworten sind zu begründen!**
**Antworten ohne Begründung erhalten 0 Punkte.**
**Bitte beschränken Sie Ihre Abgabe auf maximal (!) 5 Seiten.**

**Aufgabe 1. Ein NP-vollständiges Matrixproblem** **6+4(P)**

Betrachten Sie das folgende Matrixproblem:

**MP**
**Eingabe:** Eine Matrix $M \in \mathbb{N}^{n \times m}$ in der alle Zeilen paarweise verschieden sind und eine natürliche Zahl $k \in \mathbb{N}$.
**Frage:** Ist es möglich, $k' \leq k$ Spalten von $M$ so auszuwählen, dass in der Matrix $M' \in \mathbb{N}^{n \times k'}$, die genau diese $k'$ Spalten enthält, alle Zeilen paarweise verschieden sind?

Als Beispiel betrachten Sie folgende $(4 \times 5)$-Matrix $M$:

| | | | | |
|---|---|---|---|---|
| 0 | 2 | 0 | 5 | 1 |
| 0 | 4 | 1 | 5 | 2 |
| 1 | 3 | 0 | 6 | 1 |
| 1 | 3 | 1 | 6 | 2 |

Die Instanz $(M, k)$ mit $k \geq 2$ ist eine Ja-Instanz für MP, denn die letzten zwei Spalten bilden eine $(4 \times 2)$-Matrix, in der alle Zeilen paarweise verschieden sind.

1. Zeigen Sie, dass MP NP-vollständig ist. Sie dürfen hierbei davon ausgehen, dass das Hitting Set-Problem NP-vollständig ist, welches wie folgt definiert ist.

   **Hitting Set**
   **Eingabe:** Eine Grundmenge $X = \{x_1, \ldots, x_m\}$, eine Familie $\mathcal{F} = \{S_i \subseteq X \mid 1 \leq i \leq n\}$ von Teilmengen von $X$ und eine natürliche Zahl $k \in \mathbb{N}$.
   **Frage:** Existiert eine Teilmenge $X' \subseteq X$ mit $|X'| \leq k$, sodass für jedes $S_i \in \mathcal{F}$ gilt, dass $X' \cap S_i \neq \emptyset$?

2. Mit BMP bezeichnen wir im Folgenden die spezielle Variante des MP-Problems, in der die Eingabematrix $M$ folgende Eigenschaften erfüllt:

   (i) $M$ ist eine binäre Matrix, d.h. $M \in \{0,1\}^{n \times m}$ und

   (ii) jedes Zeilenpaar in $M$ unterscheidet sich in höchstens zwei Spalten.

   Zeigen Sie, dass BMP in P liegt, indem Sie zunächst zeigen, dass BMP in polynomieller Zeit auf den Spezialfall, in dem die Eingabematrix eine Zeile enthält, die nur aus 0'en besteht, reduziert werden kann. Zeigen Sie daraufhin, wie dieser Spezialfall in Polynomzeit gelöst werden kann. Überlegen Sie dazu, welche Form eine solche Matrix hat, wenn sie mehr als drei Zeilen enthält, von denen jede mehr als eine 1 enthält.

——————Lösungsskizze——————

1. Um zu zeigen, dass MP NP-vollständig ist, müssen wir zeigen, dass MP in NP enthalten ist, und, dass MP NP-schwer ist.

   Wir zeigen zunächst, dass MP$\in$ NP gilt. Hierfür verwenden wir die „Zertifikatdefinition“ von NP. Sei also $(M, k)$ eine Ja-Instanz von MP. Dann existieren $k' \leq k$ Spalten in $M$, sodass die $(n \times k')$-Submatrix $M'$ paarweise verschiedene Zeilen enthält. Wir wählen also die $n \times k'$-Matrix $M'$ der Größe $O(nk')$ als Zertifikat. Zum Verifizieren des Zertifiakts wird zunächst überprüft, ob $k' \leq k$ gilt und, ob $M'$ tatächlich eine Submatrix der Eingabematrix $M$ ist. Dies kann in $O(nkm)$ Zeit durchgeführt werden. Nun kann in $O(n^2k')$ Zeit überprüft werden, ob jedes der $O(n^2)$ Zeilenpaare in $M'$ verschieden ist, indem wir für jedes Paar überprüfen, ob in einer der $k'$ Spalten unterschiedliche Werte stehen.

   Es bleibt zu zeigen, dass MP NP-schwer ist. Hierfür zeigen wir Hitting Set $\leq_m^{\mathrm{p}}$ MP. Sei $I := (X = \{x_1, \ldots, x_m\}, \mathcal{F} = \{S_1, \ldots, S_n\}, k)$ eine Hitting Set-Instanz. Falls $X = \emptyset$ oder $\emptyset \in \mathcal{F}$, dann ist $I$ eine Nein-Instanz von Hitting Set und wir bilden $I$ auf eine triviale Nein-Instanz $f(I) := (M, 1)$ von MP ab, wobei

   $$M := \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

   Dieser Fall kann in $O(nm)$ Zeit behandelt werden.

   Andernfalls reduzieren wir $I$ auf eine MP-Instanz $I' := f(I) = (M, k)$, wobei die $((n + 1) \times m)$-Matrix $M$ wie folgt definiert ist:

   Für alle $i \in \{1, \ldots, n\}$ und $j \in \{1, \ldots, m\}$ setze

   $$M_{i,j} := \begin{cases} i, & \text{falls } x_j \in S_i \\ 0, & \text{sonst} \end{cases}.$$

   Außerdem setze $M_{n+1,j} := 0$ für alle $j \in \{1, \ldots, m\}$.

   Die so definierte Reduktionsfunktion $f$ kann in $O(nm)$ Zeit berechnet werden.

   Wir zeigen nun die Korrektheit, also dass $I$ genau dann eine Ja-Instanz von Hitting Set ist, wenn $I'$ eine Ja-Instanz von MP ist.

   - „$\Rightarrow$“: Sei $I$ eine Ja-Instanz von Hitting Set. Dann existiert eine Teilmenge $X' \subseteq X$ mit $|X'| \leq k$, sodass $X' \cap S_i \neq \emptyset$ für alle $S_i \in \mathcal{F}$. Wir wählen nun als Lösung für die MP-Instanz $I'$ genau die $k' := |X'| \leq k$ Spalten in $M$ aus, deren Indizes den Elementen aus $X'$ entsprechen, und behaupten, dass in der resultierenden Submatrix $M'$ alle Zeilen paarweise verschieden sind. Dazu stellen wir fest:

     Die $(n + 1)$-te Zeile in $M'$ enthält nur 0'en, da $M$ in dieser Zeile bereits nur 0'en enthielt. Jede Zeile $i = 1, \ldots, n$ in $M'$ enthält jedoch in mindestens einer Spalte den Eintrag $i$. Dies gilt, da für jedes $S_i \in \mathcal{F}$ ein $x_j$ in $X'$ existiert, sodass $x_j \in S_i$. Aus der Definition von $M$ folgt also $M_{i,j} = i$.

     Daraus folgt aber schon, dass alle $n + 1$ Zeilen in $M'$ paarweise verschieden sind, womit gezeigt ist, dass $I'$ eine Ja-Instanz ist.

   - „$\Leftarrow$“: Sei $I'$ eine Ja-Instanz von MP. Dann können wir $k' \leq k$ Spalten in $M$ so auswählen, dass die zugehörige Submatrix $M'$ paarweise verschiedene Zeilen enthält. Als Lösung für die Hitting Set-Instanz $I$ wählen wir die Teilmenge $X' \subseteq X$, die genau die $k'$ Elemente enthält, deren Indizes den ausgewählten Spalten entsprechen. Um zu zeigen, dass dies eine Lösung darstellt, betrachten wir die ausgewählten Spalten in $M'$. Da die $(n + 1)$-te Zeile nur 0'en enthalten kann, muss für jede andere Zeile $i$, $1 \leq i \leq n$, mindestens eine Spalte $j$ in $M$ ausgewählt worden sein, in der $M_{i,j} \neq 0$, also $M_{i,j} = i$ gilt. Nach der Definition von $M$ folgt also, dass für jedes $S_i \in \mathcal{F}$ ein $x_j$ in $X'$ existiert, sodass $x_j \in S_i$.

     Somit ist $I$ auch eine Ja-Instanz.

2. Wir zeigen, wie BMP in polynomieller Zeit gelöst werden kann.

   Dazu reduzieren wir eine gegebene BMP-Instanz ($M \in \{0,1\}^{n\times m}, k$) zuerst auf eine äquivalente Instanz $(N, k)$, wobei $N$ eine binäre Matrix mit einer Nullzeile ist. Dazu wählen wir eine beliebige (o.B.d.A. die erste) Zeile von $M$ aus und invertieren in $M$ jede Spalte $j$, in der in dieser Zeile eine 1 steht. Das heißt, wir definieren $N$ wie folgt:

$$N_{i,j} := \begin{cases} 1 - M_{i,j}, & \text{falls } M_{1,j} = 1 \\ M_{i,j}, & \text{sonst} \end{cases}$$

   für alle $i = 1, \ldots, n$, $j = 1, \ldots, m$. Wir erhalten also in $O(nm)$ Zeit eine Matrix $N$, in der die erste Zeile nur 0'en enthält. Da sich jedes Zeilenpaar in $N$ genau in den selben Spalten unterscheidet wie in $M$, folgt, dass $(N, k)$ genau dann eine Ja-Instanz ist, wenn $(M, k)$ eine Ja-Instanz ist. Außerdem gilt, dass $N$ analog zu $M$ die Eigenschaft (ii) erfüllt, also dass jedes Zeilenpaar sich nur in höchstens zwei Spalten unterscheidet.

   Aus (ii) folgt wiederum, dass jede Zeile $i = 2, \ldots, n$ in $N$ mindestens eine und maximal zwei 1'en enthält (da die erste Zeile nur 0'en enthält). Wir definieren für jede Zeile $i$ die Menge $Z_i := \{j \in \{1, \ldots, m\} \mid N_{i,j} = 1\}$ der Inidizes der Spalten, in denen in Zeile $i$ eine 1 steht. Da alle Zeilen paarweise verschieden sind, gilt $Z_i \neq Z_\ell$ für alle $i \neq \ell \in \{1, \ldots, n\}$. Aus (ii) lässt sich formal also Folgendes ableiten:

   (a) $|Z_1| = 0$ und $|Z_i| \in \{1, 2\}$ für alle $i = 2, \ldots, n$ und

   (b) $(|Z_i| \geq 1) \wedge (|Z_\ell| = 2) \;\Rightarrow\; |Z_i \cap Z_\ell| = 1$ für alle $\{i, \ell\} \subseteq \{1, \ldots, n\}$.

   Die Implikation (b) gilt hierbei, da $|Z_i \cap Z_\ell| \in \{0, 2\}$ bedeuten würde, dass entweder Zeile $i$ und Zeile $\ell$ sich in mindestens 3 Spalten, nämlich alle Spalten in $Z_i \cup Z_\ell$ (für den Fall $Z_i \cap Z_\ell = \emptyset$) unterscheiden, oder dass Zeile $i$ und $\ell$ identisch sind (falls $Z_i = Z_\ell$). Mit Hilfe dieser strukturellen Eigenschaften können wir nun eine Fallunterscheidung auf Basis der Anzahl der vorhandenen Zeilen mit zwei 1'en (also Zeilen $i$ mit $|Z_i| = 2$) in $N$ durchführen. Für jeden Fall werden wir zeigen, dass dieser in polynomieller Zeit gelöst werden kann. Im Folgenden bezeichne $I_x := \{i \in \{1, \ldots, n\} : |Z_i| = x\}$ die Menge der Zeilen mit $x \in \{0, 1, 2\}$ 1'en in $N$. Somit gilt, dass $|I_0| + |I_1| + |I_2| = n$ und $|I_0| = 1$.

   Fall 1: $|I_2| = 0$. Die Matrix $N$ enthält die Nullzeile und alle anderen Zeilen enthalten jeweils genau eine 1. Falls $M$ nur eine Zeile enthält, dann ist jede Spalte eine Lösung für $M$. Ansonsten muss für jede Zeile $i \in I_1$ die eine Spalte aus $Z_i$ ausgewählt werden, um Zeile $i$ von der Nullzeile zu unterscheiden. Die Instanz ist demnach genau dann eine Ja-Instanz, wenn $k \geq |I_1| = n - 1$ gilt.

   Fall 2: $1 \leq |I_2| \leq 3$. Aus (b) folgt, dass $|I_1| \leq 2 \cdot |I_2| \leq 6$, denn für jede Zeile $i$ mit $|Z_i| = 1$ und jede Zeile $\ell$ mit $|Z_\ell| = 2$ gilt, dass $Z_i \subset Z_\ell$. Demnach gibt es nur $|\bigcup_{i=1}^{n} Z_i| \leq 6$ Spalten, die jeweils mindestens eine 1 enthalten (Spalten, die nur 0'en enthalten, können wir für die Lösung ignorieren, da sie kein Paar von Zeilen voneinander unterscheiden) und höchstens $n = |I_2| + |I_1| + |I_0| \leq 3 + 6 + 1 = 10$ Zeilen in $N$. Somit können wir das Problem also in konstanter Zeit $O(1)$ lösen.

   Fall 3: $|I_2| > 3$. Wir zeigen, dass $(N, k)$ eine Ja-Instanz ist gdw. $k \geq n - 1$. Sei $i \in I_2$ eine beliebige Zeile mit zwei 1'en. Dann gilt für jedes andere Zeilenpaar $a, b \in I_2 \setminus \{i\}$, $a \neq b$, mit zwei 1'en, dass $(Z_i \cap Z_a) = (Z_i \cap Z_b)$. Wir beweisen dies per Widerspruch. Sei $Z_i = \{j, j'\} \subseteq \{1, \ldots, m\}$ mit $j \neq j'$ und $(Z_i \cap Z_a) \neq (Z_i \cap Z_b)$. Dann folgt aus (b), dass o.B.d.A.

$$\{j\} = (Z_i \cap Z_a) \neq (Z_i \cap Z_b) = \{j'\}.$$

   Desweiteren impliziert (b), dass $Z_a \cap Z_b = \{j''\}$ mit $j'' \notin \{j, j'\}$. Daraus ergeben sich also folgende Mengen

$$Z_i = \{j, j'\}, \quad Z_a = \{j, j''\}, \quad Z_b = \{j', j''\}.$$

Da aber $|I_2| \geq 4$, existiert also eine vierte Zeile $c \in I_2$, sodass sich $Z_c$ mit jeder der drei obigen Mengen in genau einem Element schneidet (wegen (b)). Dies ist aber nicht möglich da $|Z_c| = 2$. Aufgrund des Widerspruchs folgt also $(Z_i \cap Z_a) = (Z_i \cap Z_b)$ für alle $a, b \in I_2 \setminus \{i\}$ mit $a \neq b$.

Daraus folgt wiederum $\bigcap_{i \in I_2} Z_i \neq \emptyset$. Es existiert also genau eine Spalte $j$, sodass $j \in Z_i$ für alle $i \in I_2$. In der $j$-ten Spalte steht für jede Zeile mit zwei 1'en eine der beiden 1'en. Somit gilt $Z_a \cap Z_b = \{j\}$ für alle $a \neq b \in I_2$. Aus (b) folgt nun auch, dass $|I_1| \leq 1$, da eine Zeile mit genau einer 1 diese in genau einer möglichen Spalte, nämlich $j$, haben kann. Falls also $|I_1| = 1$, so müssen alle Spalten in $\bigcup_{i \in I_2} Z_i$ ausgewählt werden. Es muss also $k \geq |I_2| + 1 = n - 1$ gelten. Andernfalls, reicht es für $|I_1| = 0$ z.B. die $|I_2| = n - 1$ Spalten in $\bigcup_{i \in I_2} Z_i \setminus \{j\}$ auszuwählen. Zudem sieht man leicht, dass es für $k < |I_2|$ keine Lösung gibt, da sonst entweder eine Zeile aus $I_2$ nicht von der Nullzeile unterschieden wird, oder zwei Zeilen aus $I_2$ nicht unterschieden werden. Die Instanz ist also genau dann eine Ja-Instanz, wenn $k \geq n - 1$ gilt.

Mit der obigen Fallunterscheidung erhalten wir also folgenden Algorithmus, um die Instanz $(N, k)$ zu lösen: Zähle die Anzahl der Zeilen mit zwei 1'en in $N$ in Zeit $O(nm)$. Entscheide je nach Fall, ob eine Lösung existiert. In allen drei Fällen, kann eine Antwort in konstanter Zeit gegeben werden.

---