5 Formy i przekształcenia kwadratowe

Przystępujemy do analizy złożonościowej problemu obliczania wartości m.in. przekształceń kwadratowych. Podstawowe pojęcia związane z modelem obliczeń i złożonością zostały zdefiniowane w rozdziale 2.3.

5.1 Formy w pierścieniu wielomianów

Formą liniową w pierścieniu $k[x_1, \ldots, x_n]$ nazywamy dowolny jednorodny wielomian stopnia 1, a więc wielomian definiowany wyrażeniem postaci

$$a_1x_1 + \ldots + a_nx_n.$$

Jest oczywiste, że złożoność multiplikatywna formy liniowej jest równa 0.

Formą kwadratową w pierścieniu $k[x_1, \ldots, x_n]$ nazywamy dowolny jednorodny wielomian stopnia 2, a więc wielomian definiowany wyrażeniem postaci

$$f(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} a_{i,j}x_ix_j.$$

Jest oczywiste, że jeżeli w powyższym wzorze jest m niezerowych współczynników $a_{i,j}$, to złożoność multiplikatywna $L(f)$ jest $\leq m$. Przykładem formy kwadratowej jest dowolne wyrażenie postaci

$$\sum_{i=1}^{k} a_if_i(x_1, \ldots, x_n)g_i(x_1, \ldots, x_n),$$

gdzie $f_1, \ldots, f_k$ i $g_1, \ldots, g_k$ są formami liniowymi.

Fakt 5.1 *Przypuśćmy, że istnieją formy liniowe $f_1, \ldots, f_k$ oraz $g_1, \ldots, g_k$ takie, że każda z form kwadratowych $h_1, \ldots, h_m$ spełnia równość*

$$h_j(x_1, \ldots, x_n) = \sum_{i=1}^{k} a_{i,j}f_i(x_1, \ldots, x_n)g_i(x_1, \ldots, x_n)$$

dla odpowiednio dobranych współczynników $a_{i,j}$. Wtedy złożoność multiplikatywna $L(h_1, \ldots, h_m)$ problemu obliczenia wartości form $h_1, \ldots, h_m$ jest $\leq k$. □

Lemat 5.2 *Załóżmy, że złożoność multiplikatywna $L(h_1, \ldots, h_m)$ problemu obliczenia wartości form kwadratowych $h_1, \ldots, h_m$ jest równa k. Wtedy istnieją formy liniowe $f_1, \ldots, f_k$ oraz $g_1, \ldots, g_k$ takie, że każda z form kwadratowych $h_1, \ldots, h_m$ spełnia równość*

$$h_j(x_1, \ldots, x_n) = \sum_{i=1}^{k} a_{i,j}f_i(x_1, \ldots, x_n)g_i(x_1, \ldots, x_n)$$

dla odpowiednio dobranych współczynników $a_{i,j}$.

Dowód. Niech P będzie programem liniowym obliczającym formy kwadratowe $h_1, \ldots, h_m$, wykonującym k mnożeń i korzystającym ze zmiennych wejściowych $x_1, \ldots, x_n$. Symbolem $w_x = w_x(x_1, \ldots, x_n)$ będziemy oznaczać wartość zmiennej x w programie P. Oczywiście, w_x jest pewnym wielomianem.

Dla dowolnego wielomianu $w = w(x_1, \ldots, x_n)$ definiujemy wielomiany $w^{(0)}$, $w^{(1)}$, $w^{(2)}$ i $w^{(3)}$. Wielomian $w^{(0)}$ jest wielomianem stałym równym wyrazowi wolnemu wielomianu w, $w^{(1)}$ – to suma wszystkich jednomianów liniowych występujących w w, a $w^{(2)}$ – to suma wszystkich jednomianów stopnia 2 występujących w w. Tak więc $w^{(0)}$ jest wielomianem stałym, $w^{(1)}$ jest formą liniową, a $w^{(2)}$ – formą kwadratową. Wielomian $w^{(3)}$ definiujemy tak, aby

$$w = w^{(0)} + w^{(1)} + w^{(2)} + w^{(3)}$$

(a więc $w^{(3)}$ jest sumą jednomianów stopnia ≥ 3 z wielomianu w).

Przypuśćmy, że $z_1, z_2, \ldots, z_t$ są kolejnymi zmiennymi (roboczymi) z programu P. Zdefiniujemy dwa ciągi form liniowych $f_1, f_2, \ldots, f_t$ i $g_1, g_2, \ldots, g_t$. Jeżeli i-ta instrukcja programu P jest postaci $z_i := x * y$ ze zmiennymi x i y (wejściowymi lub roboczymi), to $f_i(x_1, \ldots, x_n) = w_x^{(1)}(x_1, \ldots, x_n)$ oraz $g_i(x_1, \ldots, x_n) = w_y^{(1)}(x_1, \ldots, x_n)$. W przeciwnym razie (dla instrukcji innej postaci), formy f_i i g_i definiujemy jako stale równe 0.

Pokażemy przez indukcję ze względu na i, że

$$w_{z_i}^{(2)}(x_1, \ldots, x_n) = \sum_{l=1}^{i} a_{i,l} f_l(x_1, \ldots, x_n) g_l(x_1, \ldots, x_n)$$

dla odpowiednio dobranych współczynników $a_{i,l}$.

Zauważmy, że dla instrukcji $z_i := z_j + x_l$ lub $z_i := z_j + c$ zachodzi związek $w_{z_i}^{(2)} = w_{z_j}^{(2)}$, a dla instrukcji $z_i := z_j + z_l$ mamy $w_{z_i}^{(2)} = w_{z_j}^{(2)} + w_{z_l}^{(2)}$. Jeżeli natomiast w programie P znajduje się instrukcja $z_i := z_j * z_l$, to

$$w_{z_i}^{(2)} = w_{z_j}^{(0)} w_{z_l}^{(2)} + w_{z_j}^{(1)} w_{z_l}^{(1)} + w_{z_j}^{(2)} w_{z_l}^{(0)} = c_1 w_{z_l}^{(2)} + f_i g_i + c_2 w_{z_j}^{(2)}$$

dla pewnych stałych c_1 i c_2. W przypadku instrukcji $z_i := z_j * x_l$, odpowiednia równość przyjmuje postać

$$w_{z_i}^{(2)} = w_{z_j}^{(0)} x_l^{(2)} + w_{z_j}^{(1)} x_l^{(1)} + w_{z_j}^{(2)} x_l^{(0)} = f_i g_i.$$

Nietrudno sprawdzić, że w podanych przypadkach sposób obliczania $w_{z_i}^{(2)}$ zachowuje wymaganą postać. Podobnie jest we wszystkich pozostałych przypadkach.

Każda z form $h_1, \ldots, h_m$ jest wartością pewnej zmiennej. Jeżeli forma kwadratowa h_j jest wartością zmiennej z_i, to

$$h_j(x_1, \ldots, x_n) = w_{z_i}(x_1, \ldots, x_n) = w_{z_i}^{(2)}(x_1, \ldots, x_n) = \sum_{l=1}^{i} a_{i,l} f_l(x_1, \ldots, x_n) g_l(x_1, \ldots, x_n).$$

Wśród iloczynów $f_l(x_1, \ldots, x_n) g_l(x_1, \ldots, x_n)$ jest tylko k niezerowych. Usuwając składniki z iloczynami $f_l g_l$ równymi 0 i dodając brakujące składniki postaci $0 \cdot f_l g_l$ dla niezerowych iloczynów $f_l g_l$ otrzymujemy żądane przedstawienie formy h_j. $\square$

5.2 Przekształcenia kwadratowe przestrzeni liniowych

Niech U i W będą przestrzeniami liniowymi (skończonego wymiaru). Funkcję $f : U \to W$ nazywamy przekształceniem kwadratowym, jeżeli istnieją bazy $u_1, \ldots, u_n$ i $w_1, \ldots, w_m$ odpowiednio przestrzeni U i W oraz formy kwadratowe $f_1, \ldots, f_m$ takie, że

$$f(\sum_{i=1}^{n} \alpha_i u_i) = \sum_{j=1}^{m} f_j(\alpha_1, \ldots, \alpha_n) w_j \tag{27}$$

dla wszystkich $\alpha_1, \ldots, \alpha_n$.

Nietrudno zauważyć, że jeżeli funkcja f jest kwadratowa, to dla dowolnych baz $u_1, \ldots, u_n$ i $w_1, \ldots, w_m$ przestrzeni U i W istnieją formy kwadratowe $f_1, \ldots, f_m$, dla których równość (27) jest spełniona.

Obliczanie wartości f będzie polegać na obliczeniu wartości form kwadratowych $f_1, \ldots, f_m$. Tak więc złożoność multiplikatywna $L(f)$ zadania polegającego na obliczeniu wartości f jest rozumiana jako złożoność multiplikatywna $L(f_1, \ldots, f_m)$ zadania polegającego na obliczeniu wartości układu form $f_1, \ldots, f_m$. Zauważmy, że $L(f)$ nie zależy od wyboru baz przestrzeni U i W (gdyż jest to złożoność multiplikatywna).

Weźmy funkcjonały liniowe $g_1, \ldots, g_l : U \to k$ i $h_1, \ldots, h_l : U \to k$ oraz (dowolne) wektory $w_1, \ldots, w_l \in W$. Wtedy funkcja

$$f(u) = \sum_{i=1}^{l} g_i(u) h_i(u) w_i$$

jest przekształceniem kwadratowym o złożoności $L(f) \leq l$. Z lematu 5.2 wynika twierdzenie odwrotne:

Lemat 5.3 *Jeżeli przekształcenie kwadratowe $f : U \to W$ ma złożoność multiplikatywną $L(f) = l$, to można je przedstawić w postaci*

$$f(u) = \sum_{i=1}^{l} g_i(u) h_i(u) v_i$$

dla odpowiednich funkcjonałów liniowych $g_1, \ldots, g_l : U \to k$ i $h_1, \ldots, h_l : U \to k$ oraz wektorów $v_1, \ldots, v_l \in W$.

Dowód. Przypuśćmy, że

$$f(\sum_{i=1}^{n} \alpha_i u_i) = \sum_{j=1}^{m} f_j(\alpha_1, \ldots, \alpha_n) w_j$$

dla form kwadratowych f_j oraz baz $u_1, \ldots, u_n \in U$ i $w_1, \ldots, w_m \in W$. Korzystając z lematu 5.2 formy hf_j przedstawiamy w postaci

$$f_j(\alpha_1, \ldots, \alpha_n) = \sum_{i=1}^{l} a_{i,j} g'_i(\alpha_1, \ldots, \alpha_n) h'_i(\alpha_1, \ldots, \alpha_n)$$

dla pewnych form liniowych $g'_1, \ldots, g'_l$ i $h'_1, \ldots, h'_l$. Zdefiniujmy funkcjonały liniowe $g_1, \ldots, g_l : U \to k$ oraz $h_1, \ldots, h_l : U \to k$ przyjmując, że

$$g_j(\sum_{i=1}^{n} \alpha_i u_i) = g'_j(\alpha_1, \ldots, \alpha_n)$$

oraz

$$h_j(\sum_{i=1}^{n} \alpha_i u_i) = h'_j(\alpha_1, \ldots, \alpha_n).$$

W ten sposób otrzymujemy, że

$$f(u) = f(\sum_{t=1}^{n} \alpha_t u_t) = \sum_{j=1}^{m} f_j(\alpha_1, \ldots, \alpha_n) w_j = \sum_{j=1}^{m} \sum_{i=1}^{l} a_{i,j} g'_i(\alpha_1, \ldots, \alpha_n) h'_i(\alpha_1, \ldots, \alpha_n) w_j =$$
$$= \sum_{j=1}^{m} \sum_{i=1}^{l} a_{i,j} g_i(u) h_i(u) w_j = \sum_{i=1}^{l} g_i(u) h_i(u) (\sum_{j=1}^{m} a_{i,j} w_j) = \sum_{i=1}^{l} g_i(u) h_i(u) v_j$$

dla α_t takich, że $u = \sum_{t=1}^{n} \alpha_t u_t$ i $v_j = \sum_{j=1}^{m} a_{i,j} w_j$. $\square$

6 Rząd przekształceń dwuliniowych

6.1 Rząd przekształcenia

Funkcję $f : U \times W \to W$ nazywamy dwuliniową, jeżeli

$$f(\alpha u, v) = f(u, \alpha v) = \alpha f(u, v)$$

oraz

$$f(u_1 + u_2, v) = f(u_1, v) + f(u_2, v) \ \text{ i } \ f(u, v_1 + v_2) = f(u, v_1) + f(u, v_2)$$

dla wszystkich $u, u_1, u_2 \in U$, $v, v_1, v_2 \in V$ i $\alpha \in k$.

Przypuśćmy, że $u_1, \ldots, u_n$ i $v_1, \ldots, v_m$ są bazami odpowiednio przestrzeni U i V. Wtedy

$$f(\sum_{i=1}^{n} a_i u_i, \sum_{j=1}^{m} b_j v_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j f(u_i, v_j).$$

Wzór ten oznacza, że funkcje dwuliniowe określone na przestrzeniach U i V są odwzorowaniami kwadratowymi określonymi na przestrzeni $U \times V$.

Weźmy dwuliniowe odwzorowanie f i jego przedstawienie jako odwzorowania kwadratowego:

$$f(u, v) = \sum_{i=1}^{k} f_i(u, v) g_i(u, v) w_i$$

dla pewnych funkcjonałow liniowych f_i, g_i określonych w $U \times V$. Zauważmy, że

$$\begin{aligned}
f(u, v) &= \sum_{i=1}^{k} f_i(u, v) g_i(u, v) w_i = \sum_{i=1}^{k} (f_i(u, 0) + f_i(0, v))(g_i(u, 0) + g_i(0, v)) w_i = \\
&= \sum_{i=1}^{k} f_i(u, 0) g_i(u, 0) w_i + \sum_{i=1}^{k} f_i(u, 0) g_i(0, v) w_i + \sum_{i=1}^{k} f_i(0, v) g_i(u, 0) w_i + \sum_{i=1}^{k} f_i(0, v) g_i(0, v) w_i = \\
&= f(u, 0) + \sum_{i=1}^{k} f_i(u, 0) g_i(0, v) w_i + \sum_{i=1}^{k} f_i(0, v) g_i(u, 0) w_i + f(0, v) = \\
&= \sum_{i=1}^{k} f_i(u, 0) g_i(0, v) w_i + \sum_{i=1}^{k} f_i(0, v) g_i(u, 0) w_i.
\end{aligned}$$

Tak więc każde odwzorowanie dwuliniowe może zostać zdefiniowane wzorem postaci

$$f(u, v) = \sum_{i=1}^{r} f_i(u) g_i(v) w_i$$

dla pewnych funkcjonałów liniowych f_i określonych w przestrzeni U oraz funkcjonałów liniowych g_i określonych w przestrzeni liniowej V. Najmniejszą liczbą r taką, że f ma przedstawienie podanej postaci nazywamy rzędem f i oznaczamy symbolem $R(f)$.

Oczywiście, $L(f) \leq R(f)$. Z powyższych obliczeń wynika, że $R(f) \leq 2L(f)$.

Jeżeli interesuje nas funkcja przyporządkowująca liczbie n wartość $L(\langle n, n, n \rangle)$, to z przedstawionych rachunków wynika, że funkcja jest tego samego rzędu, co $R(\langle n, n, n \rangle)$.

Przypuśćmy, że $f : U \times V \to K$ jest formą dwuliniową. W tym przypadku łatwo można obliczyć $R(f)$. Weźmy funkcję $F : U \to V^*$ zdefiniowaną wzorem $F(u)(v) = f(u, v)$. F jest funkcją liniową, a $F(U)$ jest podprzestrzenią liniową V^*. Pokażemy, że $R(f)$ jest równy wymiarowi $F(U)$. Niech $r = R(f)$. Wtedy $f(u,v) = \sum_{i=1}^{r} f_i(u)g_i(v)$ Wobec tego $F(u)$ jest kombinacją liniową funkcji $g_1, \ldots, g_r$ o współczynnikach $f_1(u), \ldots, f_r(u)$. Tak więc $F(U)$ jest zawarty w przestrzeni generowanej przez r wektorów. Stąd wymiar $F(U)$ jest $\leq r$. Jeżeli przestrzeń $F(U)$ ma wymiar l, to ma bazę $g_1, \ldots, g_l$. Definiujemy $f_i(u)$ jako i-tą współrzędną wektora $F(u)$ w bazie $g_1, \ldots, g_l$. Oczywiście, $F(u) = \sum_{i=1}^{l} f_i(u)g_i$ oraz f_i są funkcjonałami liniowymi. Stąd otrzymujemy, że $R(f) \leq l$.

6.2 Obcięcie przekształcenia dwuliniowego

Lemat 6.1 *Przypuśćmy, że $f : U \times V \to W$ jest przekształceniem dwuliniowym oraz, że mamy dane przekształcenia liniowe $\alpha : U' \to U$, $\beta : V' \to V$ i $\gamma : W \to W'$. Jeżeli $f' : U' \times V' \to W'$ jest przekształceniem takim, że*

$$f'(u', v')) = \gamma(f(\alpha(u'), \beta(v'))) \tag{28}$$

dla dowolnych $u' \in U'$ i $v' \in V'$, to $R(f') \leq R(f)$.

Dowód. Jeżeli f ma przedstawienie postaci

$$f(u,v) = \sum_{i=1}^{r} f_i(u)g_i(v)w_i,$$

to mamy także

$$f'(u', v') = \gamma(f(\alpha(u'), \beta(v'))) = \sum_{i=1}^{r} f_i(\alpha(u'))g_i(\beta(v'))\gamma(w_i).$$

Stąd wynika teza lematu. □

Funkcję f' definiowaną wzorem (28) nazywamy obcięciem funkcji f. Jeżeli przyjmiemy, że U' i V' są podprzestrzeniami U i V, oraz weźmiemy identycznościowe funkcje α, β i γ, to funkcja f' będzie obcięciem funkcji f do zbioru $U' \times V'$ w zwykłym tego słowa znaczeniu.

Lemat 6.2 *Przypuśćmy, że $f : U \times V \to W$ i $f' : U' \times V' \to W'$ są przekształceniami dwuliniowymi oraz, że mamy dane przekształcenia liniowe $\alpha : U \to U'$, $\beta : V \to V'$ i $\gamma : W \to W'$ takie, że $\gamma(f(u,v)) = f'(\alpha(u), \beta(v))$ dla dowolnych $u \in U$ i $v \in V$. Wtedy, jeżeli α i β są typu na, to $R(f') \leq R(f)$. Jeżeli γ jest różnowartościowe, to $R(f) \leq R(f')$.*

Dowód. Jeżeli α i β są na, to istnieją przekształcenia liniowe α' i β' takie, że $\alpha\alpha' = id$ i $\beta\beta' = id$. Podstawiając w założonym wzorze otrzymujemy $\gamma(f(\alpha'(a), \beta'(b))) = f'(a,b)$. stąd obliczenie f' jest mniej skomplikowane, niż obliczenie f.

Podobnie dowodzimy drugą nierówność. □

Wniosek 6.3 *Jeżeli $m \leq m'$, $n \leq n'$ i $k \leq k'$, to $R(\langle m,n,k\rangle) \leq R(\langle m',n',k'\rangle)$.*

6.3 Rząd iloczynu tensorowego przekształceń dwuliniowych

Przypomnijmy, że danych $f \in Bil(U,V,W)$ i $g \in Bil(U',V',W')$, iloczynem tensorowym przekształceń f i g nazywamy przekształcenie dwuliniowe $f \otimes g \in Bil(U \otimes U', V \otimes V', W \otimes W')$ spełniające dla wszystkich $u \in U$, $u' \in U'$, $v \in V$ i $v' \in V'$ równości

$$(f \otimes g)(u \otimes u', v \otimes v') = f(u,v) \otimes g(u',v').$$

(patrz wzór (11)). Pokażemy

Lemat 6.4 *Jeżeli $f \in Bil(U, V, W)$ i $g \in Bil(U', V', W')$, to $R(f \otimes g) \leq R(f)R(g)$.*

Dowód. Przyjmijmy, że $r = R(f)$ i $s = R(g)$. Wtedy

$$f(u, v) = \sum_{i=1}^{r} f_i(u) g_i(v) w_i$$

oraz

$$g(u', v') = \sum_{j=1}^{s} f'_j(u') g'_j(v') w'_j$$

dla funkcjonałów $f_i, g_i : U \to k$ i $f'_i, g'_i : U' \to k$. Zauważmy, że

$$(f \otimes g)(u \otimes u', v \otimes v') = f(u, v) \otimes g(u', v') = (\sum_{i=1}^{r} f_i(u) g_i(v) w_i) \otimes (\sum_{j=1}^{s} f'_j(u') g'_j(v') w'_j) =$$
$$= \sum_{i,j=1}^{r,s} f_i(u) g_i(v) f'_j(u') g'_j(v') (w_i \otimes w'_j).$$

Funkcja przyporządkowująca parze $(u, u') \in U \times U'$ wartość $f_i(u) f'_j(u')$ jest funkcjonałem dwuliniowym. Z własności iloczynu tensorowego określonego na $U \times U'$ otrzymujemy, że istnieje funkcjonał liniowy $\varphi_{i,j} : U \otimes U' \to k$ taki, że $\varphi_{i,j}(u \otimes u') = f_i(u) f'_j(u')$. Podobnie znajdujemy funkcjonały liniowe $\psi_{i,j} : V \otimes V' \to k$ takie, że $\psi_{i,j}(v \otimes v') = g_i(v) g'_j(v')$. Tak więc

$$(f \otimes g)(u \otimes u', v \otimes v') = \sum_{i,j=1}^{r,s} \varphi_{i,j}(u \otimes u') \psi_{i,j}(v \otimes v')(w_i \otimes w'_j).$$

Zauważmy, że zarówno lewa jak i prawa strona powyższego wzoru jest dwuliniową funkcją zmiennych $u \otimes u'$ i $v \otimes v'$. Ponieważ wśród wektorów tej postaci są bazy przestrzeni $U \otimes U'$ i $V \otimes V'$ (patrz fakt 3.7), więc na podstawie faktu 3.3 stwierdzamy, że dla dowolnych $x \in U \otimes U'$ i $y \in V \otimes V'$ zachodzi równość

$$(f \otimes g)(x, y) = \sum_{i,j=1}^{r,s} \varphi_{i,j}(x) \psi_{i,j}(y)(w_i \otimes w'_j).$$

Ten ostatni wzór implikuje, że $R(f \otimes g) \leq rs$. □

Wniosek 6.5 *Dla dowolnych liczb naturalnych zachodzą nierówności*

$$R(\langle mm', nn', kk' \rangle) \leq R(\langle m, n, k \rangle) R(\langle m', n', k' \rangle) \text{ oraz } R(\langle 2^n, 2^n, 2^n \rangle) \leq (R(\langle 2, 2, 2 \rangle))^n.$$

Dowód. Jest to wniosek z lematu 6.4 i wzoru (21). □

Wniosek 6.6 *Funkcja $r(n) = R(\langle n, n, n \rangle)$ jest rzędu $\mathcal{O}(n^{\log_2 7})$.*

Dowód. Algorytm Strassena mnożenia macierzy o wymiarach 2×2 z rozdziału 1.5 świadczy o tym, że $R(\langle 2, 2, 2 \rangle) \leq 7$. Tak więc teza wynika z wniosków 6.3 i 6.5.□

6.4 Algorytm Strassena raz jeszcze

Algorytm Strassena z rozdziału 1.6 wykorzystywał mnożenie macierzy 2×2 z rozdziału 1.5 wymagające tylko mnożenia i dodawania elementów macierzy. Uogólnimy algorytm Strassena tak, aby wykorzystywał dowolny sposób mnożenia macierzy ustalonego rozmiaru, w szczególności sposób wymagający także mnożenia przez elementy ciała.

W przeciwieństwie do rozdziału 1.6, symbol Z będzie oznaczać k-algebrę. Jest to pierścień, w którym jest także określone mnożenie przez stałe z ciała k, i mnożenie to z dodawaniem mają własności wymagane od działań w przestrzeni liniowej. Najprostszym przykładem k-algebry jest ciało k. k-algebrą jest także zbiór macierzy $M_n(k)$ ze zwykłym dodawaniem i mnożeniem oraz mnożeniem przez stałą zdefiniowanym wzorem

$$c \cdot \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ & \ddots & \\ a_{n,1} & \dots & a_{n,n} \end{pmatrix} = \begin{pmatrix} c \cdot a_{1,1} & \dots & c \cdot a_{1,n} \\ & \ddots & \\ c \cdot a_{n,1} & \dots & c \cdot a_{n,n} \end{pmatrix}. \tag{29}$$

Dla dowolnej k-algebry Z zbiór $M_n(Z)$ wraz mnożeniem przez elementy ciała k zdefiniowanym wzorem (29) też jest k-algebrą. Tak więc kolejnym przykładem k-algebry jest $M_n(M_m(k))$. Co więcej, izomorfizm $i_{n,m} : M_{n \cdot m}(Z) \to M_n(M_m(Z))$ zdefiniowany w rozdziale 1.4 jest także izomorfizmem k-algebr, a więc zachowuje również mnożenie przez elementy ciała.

Przypuśćmy, że mamy dane programy D_Z, M_Z i C_Z obliczające sumę, iloczyn elementów Z oraz iloczyn elementu Z i stałej z k. Niech $D_{M_m(Z)}$ będzie naturalnym programem obliczającym sumę dwóch macierzy o wymiarach $m \times m$ wykonującym tylko m^2 dodawań (wywołującym m^2 razy program D_Z). Symbolem $C_{M_m(Z)}$ oznaczamy naturalny program obliczający iloczyn macierzy i stałej, wykonujący m^2 mnożeń elementów Z przez stałą z k. Jak zwykle, pierścień $M_1(Z)$ utożsamiamy z Z, $D_{M_1(Z)}$ i D_Z to ten sam program. Zdefiniujemy algorytm Strassena $\mathcal{S}[D_Z, M_Z, C_Z](A, B)$ (program będący funkcją) obliczający iloczyn $A \cdot B$, wykorzystujący do dodawania i mnożenia elementów Z odpowiednio algorytmy D_Z i M_Z, oraz program C_Z do mnożenia elementów Z przez stałe z k.

Definiowany algorytm będzie wykorzystywał algorytm $\mathcal{S}_m[D_Z, M_Z, C_Z](A, B)$ obliczający iloczyn macierzy A i B o wymiarach $m \times m$.

Algorytm $\mathcal{S}[D_Z, M_Z, C_Z](A, B)$ (dla macierzy A, B rozmiaru m^i):

1) $w :=$ wymiar A;
2) jeżeli $w = 1$, to mnożymy wyrazy A i B stosując procedurę M_Z, zwracamy macierz, której jedynym wyrazem jest otrzymany iloczyn i przerywamy wykonywanie algorytmu;
3) $\mathcal{S} := i^{-1}_{m,w/m}(\mathcal{S}_m[D_{M_{w/m}(Z)}, \mathcal{S}[D_Z, M_Z, C_Z], C_{M_{w/m}(Z)}](i_{m,w/m}(A), i_{m,w/m}(B)))$.

Poprawność algorytmu $\mathcal{S}$ wynika ze wspomnianego uogólnienia twierdzenia 1.1.

Niech $S'_{DC}(n)$ i $S'_M(n)$ oznacza odpowiednio liczbę dodawań i mnożeń przez stałe, oraz liczbę pozostałych mnożeń wykonanych przez algorytm $\mathcal{S}$ uruchomiony z danymi rozmiaru m^n.

Nietrudno zauważyć, że jeżeli algorytm $\mathcal{S}_m[D_Z, M_Z, C_Z](A, B)$ wykonuje $\le a$ mnożeń oraz b pozostałych instrukcji (dodawań i mnożeń przez stałe), to

$$S'_M(n+1) \le a \cdot S'_M(n) \text{ oraz } S'_M(0) = 1$$

oraz

$$S'_{DC}(n+1) \leq a \cdot S'_{DC}(n) + b \cdot (m^n)^2 = a \cdot S'_{DC}(n) + b \cdot m^{2n} \text{ oraz } S'_{DC}(0) = 0.$$

Stąd $S'_M(n) \leq a^n$. Łatwo też można dowieść przez indukcję, że jeżeli $a > m^2$, to

$$S'_{DC}(n) \leq c \cdot (a^n - m^{2n})$$

dla $c = \frac{b}{a-m^2}$. Mamy bowiem

$$S'_{DC}(n+1) \leq aS'_{DC}(n) + bm^{2n} \leq ca^{n+1} - cam^{2n} + bm^{2n} = ca^{n+1} + \frac{b-ca}{m^2}m^{2(n+1)} =$$
$$= c(a^{n+1} - m^{2(n+1)}) + \frac{b-ca+cm^2}{m^2}m^{2(n+1)} = c(a^{n+1} - m^{2(n+1)}).$$

Obliczenie iloczynu dowolnych macierzy można zrealizować korzystając z wzoru

$$\begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} A \cdot B & 0 \\ 0 & 0 \end{bmatrix}.$$

Niech $S_{DC}(n)$ i $S_M(n)$ oznacza odpowiednio liczby dodawań i mnożeń przez stałe, oraz pozostałych mnożeń wykonanych przez ogólny algorytm uruchomiony z macierzami rozmiaru $n \times n$. Wtedy $S_{DC}(n) = S'_{DC}(n')$ oraz $S_M(n) = S'_M(n')$ dla pewnej liczby n' spełniającej nierówność $m^{n'} < m \cdot n$. Wobec tego

$$S_M(n) = S'_M(n') \leq a^{n'} = m^{n' \log_m a} = (m^{n'})^{\log_m a} < (mn)^{\log_m a} = a \cdot n^{\log_m a}$$

oraz

$$S_{DC}(n) = S'_{DC}(n') \leq c \cdot (a^{n'} - m^{2n'}) < c \cdot a^{n'} = c \cdot (m^{n'})^{\log_m a} < c \cdot (mn)^{\log_m a} = ca \cdot n^{\log_m a}.$$

Wniosek 6.7 *Jeżeli mamy algorytm mnożący macierze $m \times m$ i wykonujący $a > m^2$ mnożeń (nie licząc mnożeń przez stałą), to*

$$L^{tot}(\langle n, n, n \rangle) = \mathcal{O}(n^{\log_m a})$$

(definicja L^{tot} znajduje się w rozdziale 2.3).□

Gdyby istniał algorytm mnożenia macierzy o wymiarach $m \times m$ wykonujący m^2 mnożeń, pomysł Strassena pozwalałby skonstruować algorytm, który dla macierzy $n \times n$ wykonuje $c \cdot \log_m n \cdot n^2$ operacji arytmetycznych. Jednak nie ma takiego algorytmu mnożenia macierzy o wymiarach $m \times m$. Można to pokazać stosując metody subtelniejsze niż w dowodzie twierdzenia 2.3. Stąd wynika także, że założenie $a > m^2$ we wniosku 6.7 jest niepotrzebne.

6.5 Rząd a złożoność, współczynnik ω

Jednym z celów badań nad mnożeniem macierzy jest ustalenie współczynnika

$$\omega = \inf\{\alpha : L^{tot}(\langle n, n, n \rangle) = \mathcal{O}(n^\alpha)\}.$$

Naturalny algorytm mnożenia macierzy świadczy o tym, że $\omega \leq 3$. Z twierdzenia 2.3 wynika, że $2 \leq \omega$. Algorytm Strassena z rozdziału 1.6 daje oszacowanie $\omega \leq \log_2 7 < 2,81$. Z wniosku 6.7 otrzymujemy następujące oszacowanie ω (prawdziwe także dla $a = m^2$):

Wniosek 6.8 *Jeżeli* $L(\langle m,m,m\rangle) \leq a$, *to* $\omega \leq \log_m a$. $\Box$

Zamiast współczynnika ω możemy badać współczynniki

$$\omega^R = \inf\{\alpha : R(\langle n,n,n\rangle) = \mathcal{O}(n^\alpha)\}$$

oraz

$$\omega^L = \inf\{\alpha : L(\langle n,n,n\rangle) = \mathcal{O}(n^\alpha)\}.$$

Z nierówności $L(f) \leq R(f) \leq 2L(f)$ z rozdziału 6.1 wynika, że $\omega^L = \omega^R$. Pokażemy także

Twierdzenie 6.9 *Współczynnik* ω^R *jest równy* ω.

Dowód. Z nierówności

$$\frac{1}{2}R(\langle n,n,n\rangle) \leq L(\langle n,n,n\rangle) \leq L^{tot}(\langle n,n,n\rangle)$$

wynika, że $\omega^R \leq \omega$.

Aby dowieść nierówność przeciwną weźmy dowolne $\varepsilon > 0$. Dla tego ε, dla pewnej stałej c i dla dostatecznie dużych n zachodzi nierówność

$$R(\langle n,n,n\rangle) \leq c \cdot n^{\omega^R+\varepsilon/2}.$$

Weźmy liczbę m, dla której prawdziwa jest powyższa nierówność i zachodzi

$$c \leq m^{\varepsilon/2}.$$

Dla takiej liczby m mamy

$$L(\langle m,m,m\rangle) \leq R(\langle m,m,m\rangle) \leq c \cdot m^{\omega^R+\varepsilon/2} \leq m^{\omega^R+\varepsilon}.$$

Oznacza to, że istnieje algorytm mnożenia macierzy o wymiarach $m \times m$ wykonujący $\leq m^{\omega^R+\varepsilon}$ mnożeń. Z wniosku 6.7 otrzymujemy, że

$$L^{tot}(\langle n,n,n\rangle) = \mathcal{O}(n^{\omega^R+\varepsilon}).$$

Tak więc $\omega \leq \omega^R + \varepsilon$, a ponieważ ε jest dowolną liczbą dodatnią, więc także $\omega \leq \omega^R$. $\Box$

6.6 Rzędy przekształceń $\langle m,n,k\rangle$ i $\langle k,m,n\rangle$

Lemat 6.10 *Rzędy* $R(\langle m,n,k\rangle)$ *i* $R(\langle k,m,n\rangle)$ *są równe.*

Dowód. Przypuśćmy, że $R(\langle m,n,k\rangle) = r$. Wtedy

$$\langle m,n,k\rangle(A,B) = \sum_{i=1}^{r} f_i(A)g_i(B)M_i$$

dla pewnych funkcjonałów $f_i : k^{m\times n} \to k$ oraz $g_i : k^{n\times k} \to k$, i dla pewnych macierzy $M_i \in k^{m\times k}$. Przekształcenie dwuliniowe, które macierzom A i B przyporządkowuje macierz $f_i(A)g_i(B)M_i$, jest równe $f_i \otimes g_i \otimes M_i$ ($\otimes$ oznacza tu standardowy iloczyn tensorowy zdefiniowany w przykładzie 3.12). Wobec tego,

$$\langle m,n,k\rangle(A,B) = \sum_{i=1}^{r}(f_i \otimes g_i \otimes M_i)(A,B)$$

i w konsekwencji

$$\langle m,n,k\rangle = \sum_{i=1}^{r}(f_i \otimes g_i \otimes M_i).$$

Weźmy izomorfizm Φ określony w rozdziale 4.4. Zauważmy, że

$$\langle k,m,n\rangle = \Phi(\langle m,n,k\rangle) = \Phi(\sum_{i=1}^{r} f_i \otimes g_i \otimes M_i) = \sum_{i=1}^{r}\Phi(f_i \otimes g_i \otimes M_i). \tag{30}$$

Jeżeli we wzorach (22) i (24) podstawimy $I_{m,n}(f_i)$ za A, $I_{n,k}(g_i)$ za B i M_i za C, to otrzymamy

$$T_1(I_{m,n}(f_i), I_{n,k}(g_i), M_i) = f_i \otimes g_i \otimes M_i$$

oraz

$$T_2(I_{m,n}(f_i), I_{n,k}(g_i), M_i) = I_{k,m}^{-1}(M_i^T) \otimes I_{m,n}^{-1}(I_{m,n}(f_i)) \otimes (I_{n,k}(g_i))^T.$$

Jeżeli przyjmiemy, że $f_i' = I_{k,m}^{-1}(M_i^T)$, $g_i' = I_{m,n}^{-1}(I_{m,n}(f_i)) = f_i$ oraz $M_i' = (I_{n,k}(g_i))^T$, to ostatniemu wzorowi będziemy mogli nadać postać

$$T_2(I_{m,n}(f_i), I_{n,k}(g_i), M_i) = f_i' \otimes g_i' \otimes M_i'.$$

Z wzoru (25) wynika, że

$$\Phi(f_i \otimes g_i \otimes M_i) = \Phi(T_1(I_{m,n}(f_i), I_{n,k}(g_i), M_i)) = T_2(I_{m,n}(f_i), I_{n,k}(g_i), M_i) = f_i' \otimes g_i' \otimes M_i'.$$

Stąd i z wzoru (30)

$$\langle k,m,n\rangle = \sum_{i=1}^{r}\Phi(f_i \otimes g_i \otimes M_i) = \sum_{i=1}^{r} f_i' \otimes g_i' \otimes M_i'.$$

Jeżeli teraz skorzystamy z definicji iloczynu $\otimes$, to otrzymamy

$$\langle k,m,n\rangle(A,B) = \sum_{i=1}^{r} f_i'(A) g_i'(B) M_i'.$$

Znalezione w ten sposób przedstawienie $\langle k,m,n\rangle$ świadczy o tym, że $R(\langle k,m,n\rangle) \leq r = R(\langle m,n,k\rangle)$.

Nierówność $R(\langle m,n,k\rangle) \leq R(\langle k,m,n\rangle)$ dowodzimy w ten sam sposób. $\Box$

6.7 Rzędy przekształceń $\langle m,n,k\rangle$ i $\langle k,n,m\rangle$

Lemat 6.11 *Rzędy $R(\langle m,n,k\rangle)$ i $R(\langle k,n,m\rangle)$ są równe.*

Dowód. Można ten lemat dowieść tą samą metodą, co lemat 6.10. Trzeba tylko odpowiednio zmodyfikować konstrukcję z rozdziału 4.4. Tym razem dowód można znacznie uprościć. Zauważmy, że

$$\langle k,n,m\rangle(A,B) = AB = ((AB)^T)^T = (B^TA^T)^T = (\langle m,n,k\rangle(B^T,A^T))^T.$$

Jeżeli więc $R(\langle n,m,k\rangle) = r$ i

$$\langle m,n,k\rangle(A,B) = \sum_{i=1}^{r} f_i(A) g_i(B) M_i,$$

to także

$$\langle k,n,m\rangle(A,B) = (\langle m,n,k\rangle(B^T,A^T))^T = (\sum_{i=1}^{r} f_i(A^T)g_i(B^T)M_i)^T = \sum_{i=1}^{r} f_i(A^T)g_i(B^T)M_i^T.$$

Z udowodnionego przedstawienia mnożenia $\langle k,n,m\rangle$ wynika, że $R(\langle k,n,m\rangle) \leq R(\langle m,n,k\rangle)$.

Przeciwną nierówność dowodzimy analogicznie. □

Twierdzenie 6.12 *Jeżeli σ jest dowolną permutacją zbioru $\{1,2,3\}$, to*

$$R(\langle n_1,n_2,n_3\rangle) = R(\langle n_{\sigma(1)},n_{\sigma(2)},n_{\sigma(3)}\rangle).$$

Dowód. Lemat 6.10 implikuje interesującą nas własność dla permutacji $\sigma_1 = (1,2,3)$, a lemat 6.11 – dla permutacji $\sigma_2 = (1,3)$. Teza twierdzenia zachodzi, ponieważ każda inna permutacja zbioru $\{1,2,3\}$ jest złożeniem permutacji σ_1 i σ_2. □

Wniosek 6.13 *Jeżeli $R(\langle m,n,k\rangle) \leq a$, to $\omega \leq 3\log_{mnk} a$.*

Dowód. Jeżeli $R(\langle m,n,k\rangle) \leq a$, to

$$\begin{aligned} R(\langle mnk,mnk,mnk\rangle) &= R(\langle m,n,k\rangle \otimes \langle k,m,n\rangle \otimes \langle n,k,m\rangle) \leq \\ &\leq R(\langle m,n,k\rangle)R(\langle k,m,n\rangle)R(\langle n,k,m\rangle) = R^3(\langle m,n,k\rangle) \leq a^3 \end{aligned}$$

(na podstawie wzoru 21 i lematu 6.4). Teza wynika z wniosku 6.8. □