Geometria tensorów

Jarosław Buczyński

28 października 2014

1 Wstęp

W naukach ścisłych i inżynierii naukowcy analizują *skomplikowane* dane, aby wyizolować stosunkowo *proste* zjawiska, które mają kluczowe znaczenie w badanej sytuacji.

Jako przykład wyobraźmy sobie kilka osób rozmawiązjących przez telefon komórkowy w tym samym momencie. Stacja-odbiornik musi rozłożyć *skomplikowaną* falę elektromagnetyczną na *proste* pojedyncze sygnały, z których każdy przenosi jedną rozmowę.

Następny przykład pochodzi ze spektroskopii fluorescencyjnej. Jest to metoda służąca do analizowania stężenia związków chemicznych w próbkach roztworu. Każda próbka jest prześwietlana światłem o różnych długościach fali i badane są długości fal światła emitowanego. Zebrane dane mogą być *skomplikowane* i poszukiwany jest sposób rozłożenia danych na *proste* składniki pochodzące od pojedynczych związków chemicznych wchodzących w skład roztworu.

Problemy tego rodzaju są wszechobecne w nauce i stanowią motywację dla naszych badań nad *rozmaitościami siecznych* oraz nad powiązanymi pojęciami: rangą, rangą brzegową i rozkładem minimalnym.

Rozważmy skończenie wymiarową przestrzeń wektorową W nad ciałem liczb zespolonych $\mathbb{C}$ oraz podzbiór $\hat{X} \subset W$ rozpinający W, jako przestrzeń liniową. Niech $p \in W$. Definiujemy $\hat{X}$*-rangę* p jako najmniejszą liczbę całkowitą $r = r_{\hat{X}}(p)$, taką, że

$$p = \lambda_1 \hat{x}_1 + \lambda_2 \hat{x}_2 + \cdots + \lambda_r \hat{x}_r \text{ dla pewnych } \hat{x}_i \in \hat{X} \text{ i } \lambda_i \in \mathbb{C}.$$

Równoważnie r jest minimalną liczbą całkowitą taką, że $p \in \langle \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_r \rangle$, gdzie $\langle R \rangle$ oznacza przestrzeń liniową rozpiętą przez zbiór R.

Myślimy o V jak o zbiorze wszystkich możliwych stanów, zaś o $\hat{X}$ jak o zbiorze *prostych* stanów, oraz o p jak o *skomplikowanym* przypadku, który chcemy rozłożyć. Wobec tego ranga powinna być liczbą prostych składników, na które można rozłożyć nasz wybrany skomplkowany stan.

W powyższych przykładach, w optymalnych warunkach, ranga jest liczbą rozmów przez komórkę, lub liczbą substacji chemicznych w roztworze. Oczywiście kilka ważnych problemów może stanowić przeszkodę: jeśli składników/rozmów jest dużo, ranga może nie być pomocna. Nietrudno wyobrazić sobie, że gdy prowadzonych jest zbyt wiele rozmów jednocześnie na jednym odbiorniku, to nie jest możliwe oddzielenie pojedynczej rozmowy. Za moich czasów studenckich, tak się działo każdego roku na Sylwestra w Tatrach. Nie było możliwe zadzwonienie do krewnych z życzeniami. Podobno ten problem jeszcze występuje, ale sam nie spędzam już Sywestra w Tatrach, więc nie wiem. Podobnie, gdy dysponujemy niewielką liczbą pomiarów światła w porównaniu z liczbą składników roztworu, to nie będziemy potrafili ich zidentyfikować. Inny ważny problem to *zakłócenia*, (dane z odbiornika są nieco zniekształcone). Istnieją metody radzenia sobie z zakłóceniami, nie jest to jednak przedmiotem tego wykładu.

Rozważmy teraz kilka bardziej matematycznych przykładów.

1.1 Mnożenie macierzy

Mnożenie macierzy jest dwuliniowym odwzorowaniem $\mathbb{C}^{fg} \times \mathbb{C}^{gh} \to \mathbb{C}^{fh}$. Można więc myśleć o nim jako tensorze

$$M_{f,g,h} \in (\mathbb{C}^{fg})^* \otimes (\mathbb{C}^{gh})^* \otimes \mathbb{C}^{fh} = A \otimes B \otimes C = W.$$

Najprostszy, naiwny algorytm tego mnożenia potrzebuje fgh mnożeń liczb zespolonych i zapisuje się w postaci tensorowej jako:

$$M_{f,g,h} = \sum_{i,j,k} a_{ij} \otimes b_{jk} \otimes c_{ik}.$$

Jest to rozkład na sumę fgh *prostych tensorów* postaci $a \otimes b \otimes c$.

Niech $\hat{X} := \hat{S}eg(A \times B \times C) \subset W$ będzie zbiorem tensorów prostych. Wypisany powyżej rozkład daje oszacowanie na rangę $r_{\hat{X}}(M_{f,g,h}) \leq fgh$.

Strassen udowonił, że dwie macierze 2×2 można pomnożyć używając jedynie 7 mnożeń liczb zespolonych (zamiast $2 \cdot 2 \cdot 2 = 8$ mnożeń) [Stra69]:

$$\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \cdot \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix} = \begin{pmatrix} a_1b_1 + a_2b_3 & a_1b_2 + a_2b_4 \\ a_3b_1 + a_4b_3 & a_3b_2 + a_4b_4 \end{pmatrix}$$
$$= \begin{pmatrix} I + IV - V + VII & III + V \\ II + IV & I + III - II + VI \end{pmatrix}$$

gdzie:

$$
\begin{aligned}
I :=&(a_1 + a_4)(b_1 + b_4) & II :=&(a_3 + a_4)b_1 \\
III :=&a_1(b_2 - b_4) & IV :=&a_4(-b_1 + b_3) \\
V :=&(a_1 + a_2)b_4 & VI :=&(-a_1 + a_3)(b_1 + b_2) \\
VII :=&(a_2 - a_4)(b_3 + b_4), & &
\end{aligned}
$$

co w zapisie tensorowym przekłada się na:

$$
\begin{aligned}
M_{2,2,2} =&(a^1 + a^4) \otimes (b^1 + b^4) \otimes (c_1 + c_4) \\
&+(a^3 + a^4) \otimes b^1 \otimes (c_3 - c_4) \\
&+a^1 \otimes (b^2 - b^4) \otimes (c_2 + c_4) \\
&+a^4 \otimes (-b^1 + b^3) \otimes (c_1 + c_3) \\
&+(a^1 + a^2) \otimes b^4 \otimes (-c_1 + c_2) \\
&+(-a^1 + a^3) \otimes (b^1 + b^2) \otimes c_4 \\
&+(a^2 - a^4) \otimes (b^3 + b^4) \otimes c_1.
\end{aligned}
$$

Zatem $r_{\hat{X}}(M_{2,2,2}) \leq 7$ (a tak naprawdę X-ranga jest równa 7). Jeśli zastosujemy ten algorym wielokrotnie do macierzy blokowych, dostaniemy nowy sposób na pomnożenie dwóch $f \times f$ macierzy kwadratowych używając około $f^{\log_2 7} \simeq f^{2.81}$ mnożeń liczb zespolonych. Tę dyskusję można podsumować w następującym zdaniem: **Ranga mnożenia małych macierzy może dawać asymptotyczne ograniczenie górne na złożoność obliczeniową mnożenia dużych macierzy.**

Mnożenie macierzy jest tylko przykładem odwzorowania wieloliniowego, istnieją też inne bardzo ważne tego typu odwzorownia, które możemy badać w analogiczny sposób.

1.2 Ewaluacja wielomianów

Niech $W := S^d\mathbb{C}^n$ będzie przestrzenią wektorową wielomianów jednorodnych stopnia d w n zmiennych. Oznaczmy przez $\hat{X} := \hat{v}_d(\mathbb{C}^n) \subset W$ zbiór l^d po wszystkich $l \in \mathbb{C}^n$, czyli zbiór d-tych potęg form linowych. Zauważmy, że ewaluacja formy linowej w punkcie jest operacją dość tanią. Mając daną $p = l^d$ oraz n-tkę liczb zespolonych $a = (a_1, \ldots, a_n)$, możemy podstawić $l(a)$ i kolejno wyliczyć $l(a)^d = l(a)^{\lfloor d/2 \rfloor} \cdot l(a)^{\lceil d/2 \rceil}$. To oznacza, że stopień skomplikowania ewaluacji jest co najwyżej $\sim n \log_2 d$ w tym przypadku. Zatem złożoność ewaluacji wielomianu p wynosi co najwyżej $\sim r_X(p) \cdot n \log_2 d$.

W tym duchu, **ranga wielomianu mierzy obliczeniową złożoność ewaluacji wielomianów.**

Warto wspomnieć, że w przypadku wielomianów jednorodnych ranga jest również nazywana *rangą Waringa* w hołdzie matematykowi E. Waringowi, który w XVIII wieku zajmował się przedstawianiem liczb całkowitych jako sumy potęg, zobacz pracę przeglądową [VW02].

2 Wielomiany jednorodne w dwóch zmiennych

Ustalamy ciało $\mathbb{F}$. Zazwyczaj, będziemy sobie życzyli, aby $\mathbb{F}$ było algebraicznie domknięte, aby $\operatorname{char}\mathbb{F} = 0$, lub wręcz $\mathbb{F} = \mathbb{C}$. W trakcie ćwiczeń będziemy omawiali, jakie problemy i patologie możemy napotkać, gdy któreś z tych założeń o $\mathbb{F}$ nie jest spełnione, jak te problemy rozwiązać, lub dlaczego lepiej te patologie zostawić w spokoju.

W tym rozdziale wstępnie omówimy rozkłady wielomianów na sumy potęg. Na początek, niech $P(x, y)$ będzie wielomianem jednorodnym zależnym od dwóch zmiennych:

$$P(x, y) = a_d x^d + a_{d-1}x^{d-1}y + \cdots + a_1 xy^{d-1} + a_0 y^d$$

gdzie $a_0, a_1, \ldots, a_d$ są liczbami z ciała $\mathbb{F}$.

Problem. Jak znaleźć rangę wielomianu P oraz jego rozkład minimalny? To znaczy, niech r będzie najmniejszą liczbą całkowitą taką, że

$$P(x, y) = \ell_1^d + \ell_2^d + \cdots + \ell_r^d$$

dla pewnych liniowych wielomianów $\ell_i(x) = b_i x + c_i y$. Chcemy znaleźć r (rangę P) oraz ℓ_i (rozkład minimalny P).

Ćwiczenie 2.1. *Wyznacz rangę wielomianu* $P(x, y) = x^3 + 3xy^2$.

Różniczki $P(x, y)$ to następujące wielomiany:

$$\alpha \lrcorner P := \frac{\partial P(x, y)}{\partial x} = da_d x^{d-1} + (d-1)a_{d-1}x^{d-2}y + \cdots + 2a_2 xy^{d-2} + a_1 y^{d-1}$$
$$\beta \lrcorner P := \frac{\partial P(x, y)}{\partial y} = a_{d-1}x^{d-1} + 2a_{d-2}x^{d-2}y + \cdots + (d-1)a_1 xy^{d-2} + da_0 y^{d-1}$$

Ćwiczenie 2.2. *Policz* $\alpha \lrcorner (x^2 + xy + \frac{1}{2}y^2)$, $\beta \lrcorner (x^2 + xy + \frac{1}{2}y^2)$.

Wielokrotne pochodne będziemy oznaczać podobnie. Na przykład:

$$\alpha^2 \lrcorner P := d(d-1)a_d x^{d-2} + (d-1)(d-2)a_{d-1}x^{d-3}y + \cdots + a_2 y^{d-2}$$

i tak dalej. Ta konwencja jest wygodna dla operacji algebraicznych na różniczkach, przykładowo

$$
\begin{aligned}
(\alpha\beta - \alpha^2)\lrcorner P &= (\alpha\lrcorner(\beta\lrcorner P)) - \alpha\lrcorner(\alpha\lrcorner P)\\
(\alpha^2 + \alpha\beta - 2\beta^2)\lrcorner P &= \alpha\lrcorner(\alpha\lrcorner P) + (\alpha\lrcorner(\beta\lrcorner P)) - 2(\beta\lrcorner(\beta\lrcorner P))P(x)\\
(\alpha - \beta)(\alpha + 2\beta)\lrcorner P &= \alpha\lrcorner(\alpha\lrcorner P + 2\beta\lrcorner P) - \beta\lrcorner(\alpha\lrcorner P + 2\beta\lrcorner P).
\end{aligned}
$$

Ćwiczenie 2.3. *Policz*

$$(\alpha\beta - \alpha^2)\lrcorner(x^2y - xy^2).$$

Standardowe własności różniczek:

- $\Theta\lrcorner(P + R) = \Theta\lrcorner P + \Theta\lrcorner R$;
- $(\Theta + \Phi)\lrcorner P = \Theta\lrcorner P + \Phi\lrcorner P$;
- $(f\Theta)\lrcorner P = \Theta\lrcorner(fP) = f(\Theta\lrcorner P)$;
- $(\Theta + \Phi)\lrcorner(P + R) = \Theta\lrcorner P + \Theta\lrcorner R + \Phi\lrcorner P + \Phi\lrcorner R$;
- $\Theta\Phi\lrcorner P = \Theta\lrcorner(\Phi\lrcorner P) = \Phi\lrcorner(\Theta\lrcorner P)$;
- $(f\alpha + g\beta)\lrcorner(PR) = ((f\alpha + g\beta)\lrcorner P)R + P((f\alpha + g\beta)\lrcorner R)$.

Powyżej $P = P(x, y)$ i $R = R(x, y)$ są wielomianami w zmiennych x, y, natomiast Θ i Φ są wielomianami w zmiennych α i β. Ponadto $f, g \in \mathbb{F}$, a $(f\alpha + g\beta)$ jest dowolną liniową różniczką.

Ćwiczenie 2.4. *Pokaż, że* $(b\alpha - a\beta)\lrcorner\left[(ax + by)^d\right] = 0$ *dla dowolnych* $a, b \in \mathbb{F}$ *i dowolnego* $d > 0$.

Ćwiczenie 2.5. *Pokaż, że* $\beta(2\alpha - \beta)\lrcorner\left[(x + 2y)^d + x^d\right] = 0$ *dla dowolnego* $d > 0$.

Ćwiczenie 2.6. *Pokaż, że* $(\alpha + \beta)\beta(2\alpha - \beta)\lrcorner\left[(x + 2y)^d + x^d + (x - y)^d\right] = 0$ *dla dowolnego* $d > 0$.

Ćwiczenie 2.7. *Przypuśćmy, że* $(b\alpha - a\beta)\lrcorner P = 0$ *dla jakiegoś wielomianu* $P = P(x, y)$ *stopnia* d. *Pokaż, że* $P = c(ax + by)^d$ *dla pewnego* $c \in \mathbb{F}$.

Ćwiczenie 2.8. *Znajdź niezerowy wielomian kwadratowy* Θ *w zmiennych* α *i* β *postaci:*

$$\Theta = a\alpha^2 + b\alpha\beta + c\beta^2,$$

taki, że $\Theta\lrcorner(x^3y + xy^3) = 0$. *Jaką rangę ma* $P(x, y) = x^3y + xy^3$*?*

Ćwiczenie 2.9. *Załóżmy, że*

$$\Theta = (b_1\alpha - a_1\beta)(b_2\alpha - a_2\beta)\cdots(b_r\alpha - a_r\beta)$$

dla parami nieproporcjonalnych form liniowych $(b_i\alpha - a_i\beta)$. *Pokaż, że jeśli* $\Theta \lrcorner P = 0$ *dla wielomianu* $P \in \mathbb{F}[x,y]$, *to*

$$P(x,y) = c_1(a_1x + b_1y)^d + c_2(a_2x + b_2y)^d + \cdots + c_r(a_rx + b_ry)^d$$

dla pewnych $c_i \in \mathbb{F}$.

Ćwiczenie 2.10. *Ćwiczenie podsumowujące:*

(i) Opisz dobrą metodę*, którą można wyznaczyć rangę r dowolnego wielomianu jednorodnego dwóch zmiennych* $P(x,y)$ *oraz policzyć jego rozkład minimalny:*

$$P(x,y) = (a_1x + b_1y)^d + (a_2x + b_2y)^d + \cdots + (a_rx + b_ry)^d.$$

*(**Uwaga**: to zadanie można różnie interpretować i różnie rozwiązać; Uznamy, że metoda jest* wystarczająco dobra*, jeśli uda się ją przetestować na poniższych przykładach.)*

(ii) Uzasadnij poprawność swojej metody. Można korzystać z poprzednich ćwiczeń, pod warunkiem umieszczenia ich rozwiązań w swojej pracy.

(iii) Przetestuj swoją metodę na kilku przykładach (w szczególności znajdź rangi i rozkłady minimalne tych wielomianów):

(a) $3x^5y + 40x^3y^3 + 48xy^5$,

(b) $5x^4y + y^5$,

(c) $x^4 + 6x^2y^2 + 2y^4$,

(d) własny przykład.

(Można, jak ktoś chce, zaimplementować swoją metodę na komputerze. Jednakże te przykłady powinno się dać policzyć ręcznie. Jeśli są z tym problemy, to może to świadczyć o wątłej jakości metody.)

(iv) Co należałoby zdefiniować inaczej, żeby analogiczna metoda zadziałała,

(a) gdy $\mathbb{F}$ *nie jest algebraicznie domknięte;*

(b) gdy $\operatorname{char}\mathbb{F} > 0$?

Jest jasne, że rozkład prawie nigdy nie jest jednoznaczny, gdyż możemy zmieniać kolejność składników oraz przeskalowywać formy przez pierwiastki z jedności. Na przykład:

$$P(x,y) = 3x^6 + 6x^5y + 75x^4y^2 + 140x^3y^3 + 255x^2y^4 + 186xy^5 + 65y^6.$$

Wtedy:

$$\begin{aligned} P(x,y) &= (x+2y)^6 + x^6 + (x-y)^6 \\ &= x^6 + (x-y)^6 + (x+2y)^6 \\ &= (-x+y)^6 + (x+y)^6 + \left(\tfrac{-1+\sqrt{-3}}{2}x\right)^6 \end{aligned}$$

i tak dalej. Będziemy utożsamiać te rozkłady, mówiąc, że są one takie same z dokładnością do kolejności i przeskalowań. W rzeczy samej, powyższy wielomian P ma jednoznaczny rozkład z dokładnością do kolejności i przeskalowań.

Ćwiczenie 2.11. *Znajdź rangę $P(x,y) = x^3y$. Czy rozkład minimalny jest jednoznaczny z dokładnością do kolejności i przeskalowań?*

Ćwiczenie 2.12. *Udowodnij, że ranga wielomianu*

$$P(x,y) = x^4y + 8x^2y^3 + \tfrac{16}{5}y^5$$

jest równa 2. Czy rozkład minimalny jest jednoznaczny z dokładnością do kolejności i przeskalowań?

Ćwiczenie 2.13. *Czy istnieje wielomian (jednorodny, dwóch zmiennych) $P(x,y)$ rangi 2, który ma więcej niż jeden rozkład minimalny (z dokładnością do kolejności i przeskalowań)? Jeśli tak, to ile ma on takich rozkładów?*

3 Przestrzenie tensorów

Niech V będzie przestrzenią wektorową nad $\mathbb{F}$. Przez V^* oznaczamy *dualną przestrzeń wektorową*, czyli przestrzeń odwzorowań liniowych $V \to \mathbb{F}$.

Uwaga 3.1. Wykład dotyczy *geometrii*. To znaczy, że o przestrzeniach wektorowych myślimy jako o obiektach geometrycznych, często V będzie miało jakąś dodatkową strukturę. Pierwsze twierdzenie, którego uczymy się na kursie algebry liniowej nam mówi, że każde dwie przestrzenie wektorowe tego samego wymiaru są izomorficzne, więc niby V i V^* powinny być izomorficzne. Jednak te dwie przestrzenie mają zupełnie inne interpretacje geometryczne, o czym będziemy się przekonywać w trakcie tego wykładu.

Ćwiczenie 3.1. *Pokazać (naturalny) izomorfizm $V \simeq (V^*)^*$, tzn. taki, który nie zależy od żadnych wyborów, w szczególności od wyboru bazy.*

Literatura

[Stra69] Volker Strassen. Gaussian elimination is not optimal. *Numer. Math.*, 13:354–356, 1969.

[VW02] R. C. Vaughan and T. D. Wooley. Waring's problem: a survey. In *Number theory for the millennium, III (Urbana, IL, 2000)*, pages 301–340. A K Peters, Natick, MA, 2002.