

Wstęp do metod numerycznych

10. Miejsca zerowe wielomianów

P. F. Góra

https://zfs.fais.uj.edu.pl/pawel_gora

19 grudnia 2023

Podstawowe Twierdzenie Algebry

Rozwiązywanie równań wielomianowych

$$P_n(z) = a_n z^n + a_{n-1} z^{n-1} + \dots + a_1 z + a_0 = 0 \quad (1)$$

jest jednym z nielicznych przypadków, w których chcemy poznać **wszystkie** pierwiastki równania nieliniowego. Dzieje się tak z dwu powodów: Po pierwsze, równania wielomianowe mają duże znaczenie praktyczne. Po drugie, mamy potężne narzędzie teoretyczne:

Twierdzenie 1. Podstawowe Twierdzenie Algebry: *Wielomian stopnia n ma na płaszczyźnie zespolonej dokładnie n pierwiastków, przy czym pierwiastki wielokrotne liczy się z ich krotnościami.*

Podstawowe twierdzenie algebry nie jest konstruktywne — nie daje sposobu poszukiwania pierwiastków — ale zapewnia przynajmniej to, że wiemy, czego szukać. Jest to częsty przykład sytuacji, w której silny, choć niekonstruktywny wynik teoretyczny, znacznie ułatwia stosowanie metod numerycznych.

Ponieważ Podstawowe Twierdzenie Algebry nie wyróżnia w żaden sposób pierwiastków rzeczywistych, w tym wykładzie będziemy, w zasadzie, traktować wszystkie pierwiastki (i wielomiany) jak zespolone.

Obliczanie wartości wielomianu

Obliczanie wartości wielomianu za pomocą wielokrotnego wywoływania funkcji bibliotecznej obliczającej potęgę jest *bardzo kosztowne*, zwłaszcza w przypadku argumentu zespolonego! Znacznie bardziej efektywne jest wyliczanie kolejnych potęg za pomocą mnożenia i obliczanie wielomianu począwszy od wyrazu wolnego:

$$0: P = a_0, y = 1, k = 0$$

$$1: k = k + 1$$

$$2: y = z \cdot y$$

$$3: P = P + a_k \cdot y$$

4: Jeżeli $k < n$ GOTO 1, jeżeli $k = n$, KONIEC.

Algorytm Hornera

Powyższy algorytm można jeszcze uprościć. W tym celu zauważmy, że

$$\begin{aligned} P_n(z) &= a_n z^n + a_{n-1} z^{n-1} + \dots + a_2 z^2 + a_1 z + a_0 = \\ &= (a_n z^{n-1} + a_{n-1} z^{n-2} + \dots + a_2 z + a_1) z + a_0 = \\ &= ((a_n z^{n-2} + a_{n-1} z^{n-3} + \dots + a_2) z + a_1) z + a_0 = \dots = \\ &= (\dots ((a_n z + a_{n-1}) z + a_{n-2}) z + \dots + a_1) z + a_0. \end{aligned} \quad (2)$$

Algorytm Hornera ma o połowę mniej mnożeń, niż poprzedni algorytm:

0: $P = a_n, k = n$

1: $k = k - 1$

2: $P = P \cdot z + a_k$

3: Jeżeli $k > 0$ GOTO 1, jeżeli $k = 0$ KONIEC.

Wpływ zaburzeń współczynników

W obliczeniach praktycznych współczynniki wielomianów, których pierwiastków poszukujemy, rzadko znamy w sposób dokładny. Najczęściej są one wynikiem jakichś poprzednich obliczeń, są zatem obarczone pewnymi błędami. Jaki jest wpływ błędów współczynników na wartości znalezionych numerycznie miejsc zerowych?

Niech dokładny wielomian ma postać jak w (1) i niech z_0 będzie jego dokładnym pierwiastkiem. Przypuśćmy dalej, że *dokładne* wartości współczynników a_k nie są znane — zamiast tego znamy wartości przybliżone $\tilde{a}_k = a_k + \delta_k$, przy czym $\forall k: |\delta_k| \ll 1$. Poszukujemy pierwiastków wielomianu zaburzonego:

$$\tilde{P}_n(z) = \tilde{a}_n z^n + \tilde{a}_{n-1} z^{n-1} + \dots + \tilde{a}_1 z + \tilde{a}_0 = 0 \quad (3)$$

Spodziewamy się, że miejsce zerowe wielomianu (3) jest zaburzonym miejscem zerowym wielomianu (1): $\tilde{P}_n(\tilde{z}_0) = 0 \rightarrow \tilde{z}_0 = z_0 + \varepsilon, |\varepsilon| \ll 1$.

Mamy

$$\begin{aligned}
 0 = P_n(z_0) &= a_n z_0^n + a_{n-1} z_0^{n-1} + \dots + a_1 z_0 + a_0 \\
 (\tilde{a}_n - \delta_n)(\tilde{z}_0 - \varepsilon)^n &+ (\tilde{a}_{n-1} - \delta_{n-1})(\tilde{z}_0 - \varepsilon)^{n-1} + \dots \\
 &+ (\tilde{a}_1 - \delta_1)(\tilde{z}_0 - \varepsilon) + (\tilde{a}_0 - \delta_0). \tag{4}
 \end{aligned}$$

Zauważmy, że

$$(\tilde{z}_0 - \varepsilon)^k = \sum_{l=0}^k \binom{k}{l} \tilde{z}_0^{k-l} (-1)^l \varepsilon^l \simeq \tilde{z}_0^k - k \tilde{z}_0^{k-1} \varepsilon, \tag{5}$$

gdyż wyższe potęgi ε możemy zaniedbać. Zaniedbujemy też iloczyn $\delta_k \varepsilon$.

Zatem

$$\begin{aligned}
 0 &= P_n(z_0) \simeq \\
 &\quad \tilde{a}_n \tilde{z}_0^n - n \tilde{a}_n \tilde{z}_0^{n-1} \varepsilon + \delta_n \tilde{z}_0^n + \\
 &\quad \tilde{a}_{n-1} \tilde{z}_0^{n-1} - (n-1) \tilde{a}_{n-1} \tilde{z}_0^{n-2} \varepsilon + \delta_{n-1} \tilde{z}_0^{n-1} + \dots \\
 &= \underbrace{\sum_{k=0}^n \tilde{a}_k \tilde{z}_0^k}_{\tilde{P}_n(\tilde{z}_0)=0} - \underbrace{\left(\sum_{k=1}^n k \tilde{a}_k \tilde{z}_0^{k-1} \right)}_{\tilde{P}'_n(\tilde{z}_0)} \varepsilon + \sum_{k=0}^n \delta_k \tilde{z}_0^k. \tag{6}
 \end{aligned}$$

Ostatecznie otrzymujemy następujące **oszacowanie wpływu zaburzeń współczynników na zaburzenie miejsca zerowego wielomianu**:

$$|\varepsilon| \simeq \frac{\left| \sum_{k=0}^n \delta_k \tilde{z}_0^k \right|}{\left| \tilde{P}'_n(\tilde{z}_0) \right|}. \tag{7}$$

Przykład Wilkinsona

Wilkinson podał następujący przykład: Rozpatrujemy wielomian

$$W(z) = (z + 1)(z + 2) \cdots (z + 20). \quad (8)$$

Jego miejscami zerowymi są liczby całkowite ujemne $-1, -2, \dots, -20$. Załóżmy, że w wielomianie (8) zaburzamy tylko jeden współczynnik: $\delta_{19} = 2^{-23} \simeq 10^{-7}$, $\delta_{k \neq 19} = 0$. Jak zmieni się położenie miejsca zerowego $z_0 = -20$? $W'(-20) = -19!$ Oszacowanie (7) daje

$$|\varepsilon| \simeq \frac{10^{-7} \cdot 20^{19}}{19!} \simeq 4.4. \quad (9)$$

Zaburzenie miejsca zerowego jest *siedem rzędów wielkości* większe od zaburzenia pojedynczego współczynnika! (W rzeczywistości miejsca zerowe tak zaburzonego wielomianu stają się nawet zespolone.) Zagadnienie znajdywania miejsc zerowych wielomianów może być **źle uwarunkowane!**

Zaburzenia wielokrotnych miejsc zerowych

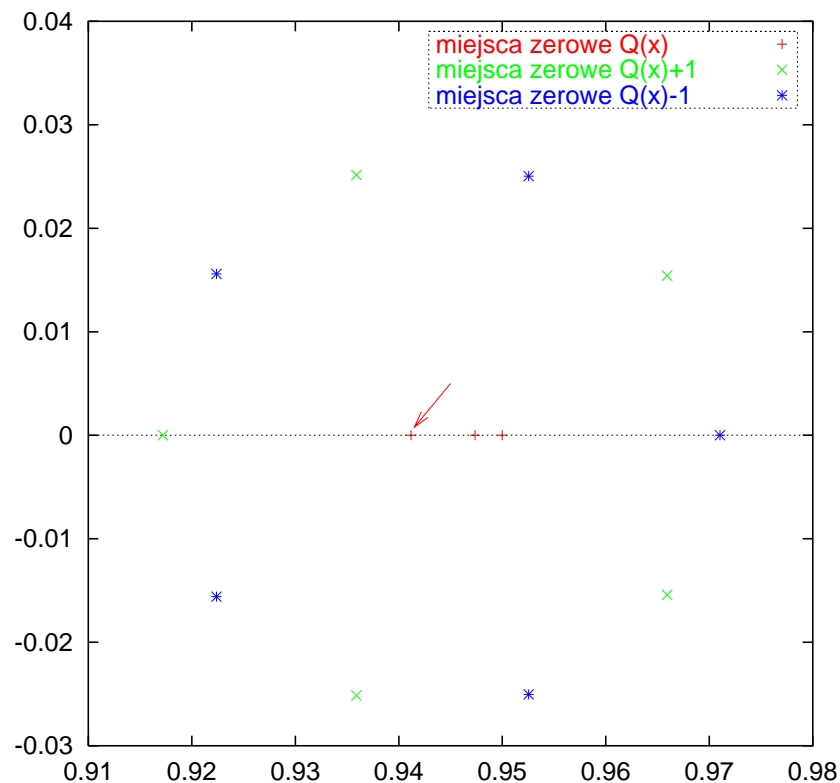
Oszacowanie (7) załamuje się dla miejsc zerowych o krotności większej od jeden (w takich miejscach zerowych znika także pochodna wielomianu). Oznacza to, że z wielokrotnym miejscem zerowym może stać się coś strasznego po niewielkim zaburzeniu współczynników wielomianu — istotnie, w takiej sytuacji **wielokrotne miejsce zerowe** na ogół **rozszczenia się** na miejsca jednokrotne, leżące w pewnym niewielkim dysku na płaszczyźnie zespolonej. Rozpatrzmy następujący **przykład**: Rozważmy wielomian

$$Q(x) = 39205740x^6 - 147747493x^5 + 173235338x^4 + 2869080x^3 - 158495872x^2 + 118949888x - 28016640 \quad (10a)$$

$$= 17^3 \cdot 19 \cdot 20 \cdot 21 \left(x + \frac{20}{21}\right) \left(x - \frac{16}{17}\right)^3 \left(x - \frac{18}{19}\right) \left(x - \frac{19}{20}\right). \quad (10b)$$

Znalezienie miejsc zerowych wielomianu w postaci (10b) jest trywialne. Numeryczne znalezienie miejsc zerowych *tego samego* wielomianu danego w postaci (10a) może być bardzo trudne. Co więcej, jeżeli wielomian (10) zaburzymy zwiększając lub zmniejszając wyraz wolny o 1 — jest to *mała* zmiana, gdyż $1/28016640 < 10^{-8}$ — miejsca zerowe, i to nie tylko potrójne miejsca zerowe, przeniosą się z osi rzeczywistej na płaszczyznę zespoloną.

W praktyce numerycznej jeśli otrzymamy grupę leżących blisko “numerycznych miejsc zerowych”, czasami trudno jest rozstrzygnąć, czy są one naprawdę różne, czy też na skutek skończonej dokładności, z jaką znamy współczynniki, reprezentują one “rozszczipione” miejsce wielokrotne.



Dokładne miejsca zerowe wielomianu (10) oraz znalezione numerycznie miejsca zerowe wielomianów $Q(x) \pm 1$. Strzałka wskazuje trzykrotne miejsca zerowe wielomianu $Q(x)$. Rysunek pokazuje tylko miejsca zerowe o dodatnich częściach rzeczywistych — miejsca zerowe leżące w $-20/21$ nie zmienia się zauważalnie przy takich zaburzeniach wielomianu.

Co robić?!

Potrzebujemy dwu rzeczy:

1. Metody numerycznego poszukiwania miejsc zerowych dedykowanej do poszukiwania miejsc zerowych wielomianów. Metodą taką jest *metoda Laguerre'a*.
2. Właściwej strategii postępowania. Polega ona na obniżaniu stopnia (deflacji) wielomianu po każdorazowym znalezieniu jego miejsca zerowego poprzez dzieleniu wielomianu przez $(z - z_0)$, gdzie z_0 jest znalezionym miejscem zerowym i wygładzaniu znalezionych miejsc zerowych przy pomocy pierwotnego, niewydzielonego wielomianu.

Metoda Laguerre'a

Niech $P_n(z)$ będzie wielomianem stopnia n . *Metoda Laguerre'a* zadana jest następującą iteracją:

$$z_{i+1} = z_i - \frac{n P_n(z_i)}{P_n'(z_i) \pm \sqrt{(n-1) \left((n-1) [P_n'(z_i)]^2 - n P_n(z_i) P_n''(z_i) \right)}}, \quad (11)$$

gdzie znak w mianowniku wybieramy tak, aby moduł mianownika był większy. Jeżeli wszystkie miejsca zerowe P_n są pojedyncze i rzeczywiste, można pokazać, że metoda (11) jest zbieżna sześciennie dla dowolnego (rzeczywistego) przybliżenia początkowego, to znaczy, że jeśli $|z_i - \bar{z}| < \varepsilon \ll 1$, to $|z_{i+1} - \bar{z}| \sim \varepsilon^3$, gdzie \bar{z} jest poszukiwanym miejscem zerowym. Bardzo techniczny dowód tego faktu można znaleźć u Ralstona.

W wypadku ogólnym

- metoda jest zbieżna sześcienne do wszystkich pojedynczych miejsc zerowych (rzeczywistych i zespolonych),
- metoda jest zbieżna liniowo do wielokrotnych miejsc zerowych,
- przypadki braku zbieżności są bardzo rzadkie; w nielicznych przypadkach, w których metodzie grozi stagnacja, można ją przerwać wykonując jeden-dwa kroki (zespoloną) metodą Newtona, a potem powrócić do metody Laguerre'a.

Metoda Laguerre'a jest metodą “z wyboru” przy poszukiwaniu (rzeczywistych i zespolonych, pojedynczych i wielokrotnych) miejsc zerowych wielomianów.

Można jej także używać do poszukiwania miejsc zerowych funkcji analitycznych, lokalnie rozwijalnych w szereg potęgowy do wyrazów rzędu n .

- Jeżeli porównamy metodę Laguerre'a (11) z wyprowadzoną w poprzednim wykładzie metodą opartą o rozwinięcie w szereg Taylora do drugiego rzędu

$$z_{i+1} = z_i - \frac{2f(z_i)}{f'(z_i) \pm \sqrt{[f'(z_i)]^2 - 2f(z_i)f''(z_i)}}. \quad (12)$$

widzimy, że metoda Laguerre'a jest do niej podobna, ale *lepsz*a, gdyż uwzględnia stopień wielomianu.

- Metoda Laguerre'a wymaga obliczania drugiej pochodnej, ale w wypadku wielomianów jest to bardzo proste.
- Metoda Laguerre'a, nawet dla rzeczywistych punktów początkowych, może prowadzić do zespolonych miejsc zerowych. Jednak z uwagi na specyfikę wielomianów, nie ma sensu upieranie się przy operowaniu na liczbach rzeczywistych.

Obniżanie stopnia wielomianu i wygładzanie

Gdy numerycznie poszukujemy miejsca zerowego wielomianu, może się zdarzyć, że metoda, nawet zstartowana z innego punktu, zbiegnie się do już wcześniej znalezionego miejsca zerowego. Aby tego uniknąć, **obniżamy stopień wielomianu**, to znaczy znajdujemy faktoryzację $P_n(z) = (z - z_1)P_{n-1}(z)$, gdzie z_1 jest wcześniej znalezionym miejscem zerowym wielomianu $P_n(z)$. Następnie szukamy miejsca zerowego wielomianu $P_{n-1}(z)$. Wiemy jednak, że nawet drobne zaburzenia współczynników wielomianu, powstałe na przykład na skutek obliczania współczynników ze skończoną precyzją, mogą znacznie zaburzyć miejsce zerowe, dlatego znalezione miejsce zerowe **poprawiamy** (wygładzamy) za pomocą pełnego, niewydzielonego wielomianu.

Powiedzmy, że znalezionym numerycznie miejscem zerowym wielomianu P_{n-1} jest \tilde{z}_2 . Tę wartość traktujemy jako warunek początkowy dla metody Laguerre'a zastosowanej do wielomianu P_n — spodziewamy się, że \tilde{z}_2 leży blisko prawdziwego miejsca zerowego, więc zbieżność będzie szybka. W ten sposób znajdujemy wygładzone miejsce zerowe z_2 . Następnie znajdujemy faktoryzację $P_{n-1}(z) = (z - z_2)P_{n-2}(z)$ (to znaczy $P_n(z) = (z - z_1)(z - z_2)P_{n-2}(z)$) itd.

Deflacja wielomianu

Przypuśćmy, że z_0 jest pierwiatkiem wielomianu (1). Musi zachodzić

$$\begin{aligned} & (z - z_0) \cdot P_{n-1}(z) \\ = & (z - z_0) (b_{n-1}z^{n-1} + b_{n-2}z^{n-2} + \dots + b_1z + b_0) \\ = & a_nz^n + a_{n-1}z^{n-1} + \dots + a_1z + a_0 = P_n(z). \end{aligned} \quad (13)$$

Trzeba znaleźć współczynniki b_k . Rozwijając (13) znajdujemy

$$\begin{aligned} b_{n-1} &= a_n \\ -z_0b_{n-1} + b_{n-2} &= a_{n-1} \\ -z_0b_{n-2} + b_{n-3} &= a_{n-2} \\ &\vdots \\ -z_0b_2 + b_1 &= a_2 \\ -z_0b_1 + b_0 &= a_1 \\ -z_0b_0 &= a_0 \end{aligned} \quad (14)$$

Należy się chwilę zastanowić: (14) stanowi układ $n+1$ *nieliniowych* równań na $n+1$ nieznanymi parametrów: b_0, b_1, \dots, b_{n-1} oraz z_0 . Równania te jednak zawierają w sobie ukryty warunek, że z_0 jest pierwiastkiem wyjściowego wielomianu. Istotnie, mnożąc to z równań (14), którego prawą stroną jest a_k , przez z^k , postępując tak z każdym z tych równań i dodając wymnożone równania stronami, otrzymujemy poszukiwany warunek. Możemy zatem potraktować z_0 jako znany parametr i wyeliminować jedno z równań (14). Eliminując ostatnie otrzymujemy następujący układ równań *liniowych*:

$$\begin{bmatrix} 1 & & & & & & & \\ -z_0 & 1 & & & & & & \\ & -z_0 & 1 & & & & & \\ & & \dots & \dots & & & & \\ & & & -z_0 & 1 & & & \\ & & & & -z_0 & 1 & & \end{bmatrix} \begin{bmatrix} b_{n-1} \\ b_{n-2} \\ b_{n-3} \\ \vdots \\ b_1 \\ b_0 \end{bmatrix} = \begin{bmatrix} a_n \\ a_{n-1} \\ a_{n-2} \\ \vdots \\ a_2 \\ a_1 \end{bmatrix} \quad (15)$$

Układ ten można bardzo łatwo rozwiązać metodą *forward substitution*.

Strategia postępowania

Przypuśćmy, że znaleźliśmy już miejsca zerowe z_1, z_2, \dots, z_k wielomianu $P_n(z)$ i faktoryzację $P_n(z) = (z - z_1)(z - z_2) \cdots (z - z_k)P_{n-k}(z)$. Teraz

1. Biorąc dowolny zespolony warunek początkowy, znajdujemy za pomocą metody Laguerre'a miejsce zerowe \tilde{z}_{k+1} wielomianu P_{n-k} .
2. W celu wygładzenia tego miejsca zerowego, liczby \tilde{z}_{k+1} używamy jako warunku początkowego dla metody Laguerre'a zastosowanej do *pełnego* wielomianu $P_n(z)$. Spodziewamy się, że w ciągu kilku (nie-wielu) iteracji znajdziemy jego miejsce zerowe z_{k+1} .
3. Faktoryzujemy wielomian $P_{n-k}(z)$, to znaczy obliczamy $P_{n-k}(z) = (z - z_{k+1})P_{n-k-1}(z)$.
4. Postępujemy tak dopóki nie dojdziemy do wielomianu stopnia 2, którego pierwiastki obliczamy według znanego wzoru.

Uwagi końcowe

1. Jeżeli wyjściowy wielomian $P_n(z)$ ma współczynniki rzeczywiste, wiemy, że jego miejsca zerowe są albo rzeczywiste, albo tworzą zespolone pary sprzężone. Jeżeli zatem za pomocą powyższej strategii znajdziemy jego wygładzone, zespolone miejsce zerowe $z_k = x_k + iy_k$, $x_k, y_k \in \mathbb{R}$, **wiemy**, że miejscem zerowym jest także $z_{k+1} = x_k - iy_k$, a więc możemy obziżyć stopień wielomianu od razu o dwa.

2. Jeżeli wyjściowy wielomian ma współczynniki całkowite (i niezbyt wielkie), przed przystąpieniem do obliczeń numerycznych, warto za pomocą znanego twierdzenia poszukać jego pierwiastków wymiernych.

Przykład

Szukamy miejsc zerowych wielomianu z dokładnością do 10^{-8}

$$P_8(z) = 6561x^8 - 2187x^7 - 243x^5 + 27x^3 + 3x - 1. \quad (16a)$$

Stosując przedstawioną wyżej strategię otrzymuję następujące miejsca zerowe:

	przed wygładzeniem	wygładzone
1	$0.33333310 + 0.00000040i$	
2	$0.33333310 - 0.00000040i$	$0.33333309 - 0.00000043i$
3	$0.00000000 + 0.33333333i$	$-0.00000000 + 0.33333333i$
4	$0.33333381 + 0.00000002i$	$0.33333363 + 0.00000051i$
5	$-0.33333479 - 0.00000055i$	$-0.33333333 - 0.00000000i$
6	$-0.16666756 - 0.28867623i$	$-0.16666667 - 0.28867513i$
7	$0.00000042 - 0.33333352i$	$0.00000000 - 0.33333333i$
8	$-0.16666648 + 0.28867465i$	$-0.16666667 + 0.28867513i$

Pierwszego znalezionej miejsca zerowego *nie wygładzam*, bo zostało znalezione przy pomocy niewydzielonego wielomianu.

Wielomian (16a) jest równy

$$P_8(x) = (3x - 1)^3(3x + 1)(9x^2 + 3x + 1)(9x^2 + 1) \quad (16b)$$

a więc pierwiastkami są $1/3$ (trzykrotny), $-1/3$, $\pm 1/3i$, $-1/6 \pm \sqrt{3}/6i$.

Choć w powyższym przykładzie wygładzanie miejsc zerowych wielomianów niższego stopnia przy pomocy pełnego (niewydzielonego) wielomianu (16a) nie przyniosło znaczących zmian, krok ten może być bardzo ważny.

Zupełnie inne podejście 😊

Ze względu na to, że numeryczne znajdowanie miejsc zerowych wielomianów może być źle uwarunkowane ze względu na małe zaburzenia współczynników, znajdowanie wartości własnych macierzy — poza szczególnymi przypadkami, wykraczającymi poza zakres tego wykładu — za pomocą numerycznego znajdowania pierwiastków wielomianu charakterystycznego no ogół jest numerycznie **złym pomysłem**. Czasami robi się wręcz na odwrót: aby znaleźć pierwiastki wielomianu, szukamy wartości własnych macierzy stowarzyszonej (ang. *companion matrix*)

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & -a_3 & \cdots & -a_{n-2} & -a_{n-1} \end{bmatrix} \quad (17)$$

Równaniem charakterystycznym macierzy stowarzyszonej (17) jest (z dokładnością do znaku)

$$x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \dots + a_1x + a_0 = 0. \quad (18)$$

Lewa strona tego równania różni się od ogólnego wielomianu (1) tylko tym, że współczynnik przy najwyższej potędze wynosi 1. Każdy wielomian stopnia n można sprowadzić do tej postaci *dzieląc* przez współczynnik przy najwyższej potędze. Nie zmienia to pierwiastków wielomianu.

Istnieje ciekawy związek pomiędzy równaniem charakterystycznym (17) a algorytmem Hornera.

Przy numerycznym znajdowaniu wartości własnych macierzy (17), ze względów praktycznych łatwiej jest diagonalizować jej macierz transponowaną (ich widma są identyczne), ale *niekiedy* zbieżność algorytmu QR i tak jest bardzo wolna. Znajomość wektorów własnych nie jest potrzebna.

W wypadku ogólnym omówiona poprzednio strategia — metoda Laguerre'a, połączona z deflacją wielomianu i wygładzaniem miejsc zerowych — jest, moim zdaniem, numerycznie bardziej wydajna.