L'algoritmo di fattorizzazione di Berlekamp

Primi ingredienti:

1. Siano f, a, b polinomi di $\mathbb{Z}_p[x]$ tali che a e b sono primi tra loro. Allora vale:
$$\gcd(f, ab) = \gcd(f, a) \cdot \gcd(f, b)$$
2. Il polinomio $x^p - x \in \mathbb{Z}_p[x]$ si spezza nel prodotto
$$(x - 0) \cdot (x - 1) \cdots (x - (p - 1))$$
3. Se $g \in \mathbb{Z}_p[x]$, allora vale:
$$g^p - g = (g - 0) \cdot (g - 1) \cdots (g - (p - 1))$$

Per vedere il primo punto, basta scrivere ciascuno dei tre polinomi come prodotto di potenze di polinomi irriducibili; il secondo punto segue dal piccolo teorema di Fermat. Il terzo punto segue dal secondo e dal fatto che l'applicazione $\phi : \mathbb{Z}_p[x] \longrightarrow \mathbb{Z}_p[x]$ definita da $\phi(a_0 + a_1x + \cdots + a_nx^n) = a_0 + a_1g + \cdots + a_ng^n$ è un omomorfismo di anelli.

Teorema 1 *Sia f un polinomio di $\mathbb{Z}_p[x]$ e sia $g \in \mathbb{Z}_p[x]$ un polinomio tale che:*

1. $1 \leq \deg(g) < \deg(f)$;
2. *f divide $g^p - g$;*

allora vale:

$$f = \gcd(f, g - 0) \cdot \gcd(f, g - 1) \cdots \gcd(f, g - (p - 1))$$

e la fattorizzazione di f scritta sopra è effettiva, cioè ci sono fattori propri di f.

Dim.: Se f divide $g^p - g$ allora abbiamo:

$$\begin{aligned} f &= \gcd(f, g^p - g) \\ &= \gcd(f, (g - 0) \cdot (g - 1) \cdots (g - (p - 1)) \\ &= \gcd(f, g - 0) \cdot \gcd(f, g - 1) \cdots \gcd(f, g - (p - 1)) \end{aligned}$$

L'ultimo passaggio segue dal fatto che $g - i$ e $g - j$ sono, se $i \neq j$, primi tra loro (infatti se un polinomio divide entrambi, deve dividere la loro differenza che è la costante $i - j$). Poichè il grado di g è inferiore al grado di f, ognuno dei polinomi $\gcd(f, g - i)$ ha grado inferiore al grado di f; inoltre, poichè g non è una costante, non può succedere che $\gcd(f, g - i)$ sia, per qualche $i \in \mathbb{Z}_p$, il polinomio f stesso, quindi la fattorizzazione scritta sopra è effettiva. □

Per riuscire a fattorizzare un polinomio f di $\mathbb{Z}_p[x]$ si deve quindi trovare un polinomio g con le proprietà indicate nel precedente teorema. Il problema diventa allora quello di trovare i polinomi g che soddisfano alle seguenti due condizioni:

1. $\deg(g) < d$ (dove $d = \deg(f)$);
2. $f \mid g^p - g$.

(per ora non richiediamo la condizione $deg(g) \geq 1$, come richiesta nel teorema precedente).

Sia allora $g = b_0 + b_1 x + \cdots + b_{d-1} x^{d-1}$ il generico polinomio di grado al più $d-1$ con coefficienti $b_0, b_1, \ldots, b_{d-1} \in \mathbb{Z}_p$. Vale:

$$\begin{aligned} g^p &= \left(b_0 + b_1 x + \cdots + b_{d-1} x^{d-1}\right)^p \\ &= b_0^p + b_1^p x^p + \cdots + b_{d-1}^p x^{(d-1)p} \qquad (1) \\ &= b_0 + b_1 x^p + \cdots + b_{d-1} x^{(d-1)p} \qquad (2) \end{aligned}$$

dove (1) deriva dal fatto che nell'anello $\mathbb{Z}_p[x]$, che è di caratteristica p, vale $(\alpha + \beta)^p = \alpha^p + \beta^p$, mentre l'uguaglianza (2) segue dal piccolo teorema di Fermat, cioè dal fatto che ogni elemento di $\mathbb{Z}_p$ soddisfa all'equazione $a^p = a$. Se dividiamo $g^p - g$ per f otteniamo:

$$g^p - g = q \cdot f + r,$$

dove $\deg(r) < \deg(f)$, allora per avere che f divida $g^p - g$ bisogna avere che r sia 0. Per calcolare in modo più esplicito il resto r dividiamo intanto i monomi x^{jp} che compaiono in g^p; sia dunque:

$$x^{jp} = q_j \cdot f + r_j, \quad \text{per } j = 0, \ldots, d-1$$

dove $\deg(r_j) < d$. Sostituendo questi valori di x^{jp} in g^p otteniamo:

$$\begin{aligned} g^p - g &= b_0 x^0 + b_1 x^p + \cdots + b_{d-1} x^{(d-1)p} - b_0 - b_1 x - \cdots - b_{d-1} x^{d-1} \\ &= b_0(q_0 \cdot f + r_0) + b_1(q_1 \cdot f + r_1) + \cdots + b_{d-1}(q_{d-1} \cdot f + r_{d-1}) \\ &\quad - b_0 - b_1 x - \cdots - b_{d-1} x^{d-1} \\ &= (b_0 q_0 + \cdots + b_{d-1} q_{d-1}) \cdot f + b_0 r_0 + \cdots + b_{d-1} r_{d-1} \\ &\quad - b_0 - b_1 x - \cdots - b_{d-1} x^{d-1} \end{aligned}$$

Pertanto il resto r della divisione di $g^p - g$ per f vale:

$$r = b_0 r_0 + \cdots + b_{d-1} r_{d-1} - b_0 - b_1 x - \cdots - b_{d-1} x^{d-1} \qquad (3)$$

I polinomi r_j sono polinomi di grado al più $d-1$, quindi sarà:

$$r_j = r_{0\,j} + r_{1\,j} x + \cdots + r_{d-1\,j} x^{d-1} \quad \text{per } j = 0, \ldots, d-1$$

Sostituendo questi valori per r_j in (3) si ottiene:

$$\begin{aligned} r &= b_0(r_{0\,0} + r_{1\,0} x + \cdots + r_{d-1\,0} x^{d-1}) \\ &\quad + b_1(r_{0\,1} + r_{1\,1} x + \cdots + r_{d-1\,1} x^{d-1}) \\ &\quad \cdots \\ &\quad + b_{d-1}(r_{0\,d-1} + r_{1\,d-1} x + \cdots + r_{d-1\,d-1} x^{d-1}) \\ &\quad - b_0 - b_1 x - \cdots - b_{d-1} x^{d-1} \\ &= (b_0 r_{0\,0} + \cdots + b_{d-1} r_{0\,d-1} - b_0) + \\ &\quad (b_0 r_{1\,0} + \cdots + b_{d-1} r_{1\,d-1} - b_1) x + \\ &\quad \cdots \\ &\quad (b_0 r_{d-1\,0} + \cdots + b_{d-1} r_{d-1\,d-1} - b_{d-1}) x^{d-1} \end{aligned}$$

Affinché il polinomio r sia zero, per il principio di identità dei polinomi, devono essere zero tutti i suoi coefficienti, quindi abbiamo che: f divide $g^p - g$ se e solo se $b_0, \ldots, b_{d-1}$ soddisfano il sistema lineare:

$$\begin{cases} (b_0-1)r_{0\,0} & + & b_1 r_{0\,1} & + & \cdots & + & b_{d-1} r_{0\,d-1} & = & 0 \\ b_0 r_{1\,0} & + & (b_1-1)r_{1\,1} & + & \cdots & + & b_{d-1} r_{1\,d-1} & = & 0 \\ & \cdots & & & \cdots & & & \cdots & \\ b_0 r_{d-1\,0} & + & b_1 r_{d-1\,1} & + & \cdots & + & (b_{d-1}-1) r_{d-1\,d-1} & = & 0 \end{cases}$$

Detta Q la matrice seguente:

$$Q = \begin{pmatrix} r_{0\,0} & r_{0\,1} & \cdots & r_{0\,d-1} \\ r_{1\,0} & r_{1\,1} & \cdots & r_{1\,d-1} \\ & \cdots & \cdots & \\ r_{d-1\,0} & r_{d-1\,1} & \cdots & r_{d-1\,d-1} \end{pmatrix}$$

otteniamo allora che f divide $g^p - g$ se e solo se il vettore ${}^t(b_0, b_1, \ldots, b_{d-1})$ appartiene al nucleo di $Q - I$, dove I è la matrice identica di ordine d. Pertanto abbiamo trovato che vi è una corrispondenza biunivoca tra l'insieme

$$G = \{g \in \mathbb{Z}_p[x] \ : \ \deg(g) < d \ \text{ e } \ f \mid g^p - g\} \tag{4}$$

e $\ker(Q - I)$. Nella corrispondenza biunivoca l'elemento $g = b_0 + b_1 x + \cdots + b_{d-1}x^{d-1}$ viene mandato nel vettore ${}^t(b_0, b_1, \ldots, b_{d-1})$. Si verifica poi facilmente che l'insieme G è uno $\mathbb{Z}_p$–spazio vettoriale, come anche $\ker(Q-I)$ e si vede subito che la corrispondenza biunivoca in realtà è un isomorfismo di spazi vettoriali. Le considerazioni fin qui fatte si possono riassumere nel:

Teorema 2 *Sia f un polinomio di $\mathbb{Z}_p[x]$ di grado d. Allora vi è un isomorfismo di spazi vettoriali tra i polinomi g dello spazio vettoriale G definito in (4) e $\ker(Q - I)$ dove Q è la matrice le cui colonne sono i coefficienti dei resti della divisione di x^{jp} per f (dove j vale $0, 1, \ldots, d-1$)* □

Dai teoremi 1 e 2 si ha quindi un metodo per fattorizzare i polinomi di $\mathbb{Z}_p[x]$.

Osservazione L'insieme G, definito in (4), contiene sempre almeno tutte le costanti, cioè $\mathbb{Z}_p$ (come segue subito dal piccolo teorema di Fermat). Può succedere però che G contenga *solo* le costanti. In questo caso tutti i polinomi di G hanno grado 0 e quindi non si è riusciti a costruire alcun polinomio di grado almeno 1, come richiesto dal teorema 1. Quando G contiene solamente le costanti, la sua dimensione, come $\mathbb{Z}_p$–spazio vettoriale, è 1 (e quindi è 1 anche la dimensione di $\ker(Q - I)$). Questo è il punto di partenza per un successivo teorema di Berlekamp.

Teorema 3 *Sia f un polinomio di $\mathbb{Z}_p[x]$ di grado d e sia Q la matrice sopra costruita. allora vale:*

1. *La dimensione di $\ker(Q - I)$ come $\mathbb{Z}_p$–spazio vettoriale coincide con il numero dei fattori irriducibili distinti di f;*
2. *f è irriducibile se e solo se $\dim(\ker(Q - I)) = 1$ e f e f' sono primi tra loro.*

DIM.: Supponiamo che il numero di fattori irriducibili distinti di f sia k, quindi il polinomio f può essere scomposto in:

$$f = q_1^{\alpha_1} \cdot q_2^{\alpha_2} \cdots q_k^{\alpha_k}$$

(con $q_1, q_2, \ldots, q_k$ polinomi irriducibili distinti di $\mathbb{Z}_p[x]$). Sia G definito come in (4) e prendiamo $g \in G$. Quindi $f \mid g^p - g$, perciò $q_i \mid g^p - g = (g-0) \cdot (g-1) \cdots (g-(p-1))$ per ogni $i = 1, \ldots, k$, da cui segue, essendo q_i irriducibile, che per ogni i esiste un $s_i \in \mathbb{Z}_p$ tale che q_i divide $g - s_i$ (inoltre s_i è univocamente determinato da q_i, in quanto i polinomi $g-i$ e $g-j$, se $i \neq j$, sono primi tra loro). (Da quest'ultima osservazione segue anche che in realtà $q_i^{\alpha_i} \mid (g - s_i)$). Pertanto abbiamo costruito un'applicazione $\phi : G \longrightarrow \mathbb{Z}_p^k$ tale che $\phi(g) = (s_1, \ldots, s_k)$. Verifichiamo che ϕ è suriettiva: sia $(s_1, \ldots, s_k) \in \mathbb{Z}_p^k$, allora, per il teorema cinese del resto, esiste un unico polinomio $g \in \mathbb{Z}_p[x]$ tale che $g \equiv s_i \bmod(q_i^{\alpha_i})$ e $\deg(g) < \alpha_1 + \cdots + \alpha_k = \deg(f)$. Inoltre, da $q_i^{\alpha_i} \mid g - s_i$ otteniamo che $q_i^{\alpha_i} \mid g^p - g$ e quindi, essendo i polinomi $q_i^{\alpha_i}$ a due a due coprimi, anche il loro prodotto, che è f, deve dividere $g^p - g$. Questo prova che $g \in G$. Si verifica poi subito che $\phi(g) = (s_1, \ldots, s_k)$ e quindi ϕ è suriettiva; proviamo ora che è iniettiva: siano $g_1, g_2 \in G$ tali che $\phi(g_1) = \phi(g_2) = (s_1, \ldots, s_k)$. Allora, per ogni i, $q_i \mid g_1 - s_i$ e $q_i \mid g_2 - s_i$, quindi $q_i^{\alpha_i} \mid g_1 - s_i$ e $q_i^{\alpha_i} \mid g_2 - s_i$ (quest'ultima affermazione segue dal fatto che se $q_i \mid g_1 - s_i$, allora q_i è relativamente primo con $g_1 - s_j$ e analogamente per g_2). Da questo risultato si ottiene che $q_i^{\alpha_i} \mid g_1 - g_2$ (per ogni i) e quindi $f \mid g_1 - g_2$, ma questo succede solo se $g_1 = g_2$ perché il grado di $g_1 - g_2$ è minore del grado di f. Quindi ϕ è una biiezione e G ha lo stesso numero di elementi di $\mathbb{Z}_p^k$, cioè G ha p^k elementi. Si è precedentemente dimostrato che G è uno spazio vettoriale sul campo $\mathbb{Z}_p$, quindi per avere p^k elementi deve necessariamente avere una base di k elementi. Cioè la sua dimensione, come $\mathbb{Z}_p$–spazio vettoriale, è k. Essendo $\ker(Q - I)$ isomorfo a G, anche $\dim(\ker(Q - I)) = k$. Questo prova il primo punto. Il secondo punto segue facilmente dal primo punto e dal fatto che un polinomio (in $\mathbb{Z}_p[x]$) non ha fattori multipli se e solo se f e f' sono primi tra loro. □