UNIVERSITÀ DEGLI STUDI ROMA TRE

FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI

Tesi di laurea magistrale di

Andrea Cova

Numeri lisci: teoria numerica computazionale ed applicazioni in ambito crittografico

Sintesi

Relatore

Professor Francesco Pappalardi

Il candidato Il relatore

Classificazione AMS: 11N25, 11Y16, 68Q25, 11Y05, 94A60, 11N36

Parole chiave: distribuzione di interi con specifiche proprietà moltiplicative, analisi di algoritmi e complessità computazionale, crittografia, fattorizzazione di interi, applicazione di metodi di crivello

ANNO ACCADEMICO 2007-2008

MAGGIO 2009

Sintesi

Nell'analisi di numerosi algoritmi di Teoria dei Numeri gioca frequentemente un ruolo essenziale una peculiare classe di interi, contraddistinti dal possedere unicamente fattori primi di piccola entità: tali interi vengono consuetamente denominati *numeri lisci*. Allo scopo di confrontare la complessità computazionale di molteplici algoritmi numerici e crittografici e con la finalità di ottimizzarne le prestazioni e dunque l'efficienza, si è rivelato estremamente significativo ed importante il pervenire a stime sufficientemente accurate per la percentuale di numeri lisci che compaiono in svariate sequenze. Difatti sorprendentemente numerosi sono stati i recenti sviluppi (e i possibili scenari aperti a ulteriori scoperte) della ricerca crittografica fondata sulle proprietà di questa specifica classe di numeri: metodologie algoritmiche per la fattorizzazione di interi, test di primalità, calcolo di logaritmi discreti, protezione di sistemi di cifratura dei messaggi da possibili attacchi malevoli sono solamente alcune delle più rilevanti problematiche, di fondamentale importanza in ambito crittografico, per le quali i numeri lisci interpretano un ruolo di cardinale importanza.

Innumerevoli sono stati gli studiosi che hanno dedicato attenzione ed energia allo studio delle proprietà dei numeri lisci: per citare solo alcuni dei nomi maggiormente rappresentativi si ricordino i contributi allo sviluppo della materia apportati da Adolf Hildebrand e Gerald Tenenbaum, Karl Dickman, Nicolaas Govert de Bruijn, Alexander Adol'fovich Buchstab, Robert Alexander Rankin, Andrew Granville, Pieter Moree, Carl Pomerance, Igor Shparlinski, Simon Hunter, Jonathan Sorenson e Daniel Julius Bernstein.

In questa tesi passeremo in rassegna gli studi maggiormente interessanti che sono stati sinora condotti su tale tipologia di numeri interi: dapprima analizzeremo le più rilevanti stime applicabili nell'ambito di questioni di Teoria dei Numeri computazionale, prendendo in considerazione sia risultati compiutamente e dettagliatamente dimostrati sia congetture non ancora provate ma comunque plausibilmente veritiere; in seguito descriveremo alcune applicazioni dei numeri lisci a svariati problemi relativi a differenti settori della Teoria dei Numeri e concluderemo infine la nostra

analisi dedicando una particolare attenzione all'ambito crittografico.

Il **primo capitolo** è interamente focalizzato sull'introduzione delle definizioni di base concernenti i numeri lisci e sull'analisi delle più rilevanti stime che da un lato consentono di analizzarne le proprietà ed il peculiare andamento e dall'altro ne permettono l'applicazione pratica a molteplici ambiti di carattere numerico-crittografico. Si può anzitutto introdurre la basilare nozione di numero liscio, fulcro della nostra trattazione, nella seguente maniera.

Definizione 1.1. *(i) Un intero positivo si dice y-liscio se non possiede alcun fattore primo maggiore di y ovvero, più formalmente, possiamo affermare che*

$$x \text{ è un numero } y\text{-liscio} \Leftrightarrow p \leq y, \forall p \mid x, \text{ con } p \text{ primo.}$$

(ii) La funzione $\psi(x, y)$ viene definita come il numero di interi positivi y-lisci che risultino essere $\leq x$, ossia

$$\psi(x, y) := \sum_{n \leq x, P(n) \leq y} 1,$$

dove si è denotato con $P(n)$ il più grande fattore primo di n (avendo posto per definizione $P(1) = 1$).

Il passo naturale immediatamente successivo è quello di prendere in considerazione il fondamentale risultato teorico, dovuto a Karl Dickman (e dimostrato in [1]), che fornisce una tra le più importanti stime asintotiche alle quali è possibile ricorrere per approssimare efficacemente la funzione $\psi(x, y)$ appena introdotta.

Teorema 1.1. (Dickman) *Sia $u > 1$ un fissato numero reale e supponiamo che $x = y^u$. Allora esiste un opportuno numero reale $\rho(u) > 0$ tale che:*

$$\psi(x, y) \sim x\rho(u) \qquad \text{allorchè } x \to \infty.$$

La funzione $\rho(u)$ che compare nell'enunciato precedente viene denominata **funzione di Dickman-de Bruijn** e può essere definita elegantemente attraverso la seguente equazione differenziale.

Definizione 1.2. *La* funzione di Dickman $\rho(u)$ *è individuata, per* $u \geq 0$, *dall' unica soluzione continua del seguente sistema:*

$$\begin{cases} \rho(u) = 1 & per\ 0 \leq u \leq 1 \\ u\rho'(u) + \rho(u-1) = 0 & per\ u > 1. \end{cases}$$

Inoltre si pone $\rho(u) = 0$ *per* $u < 0$, *cosicchè tale funzione* ρ *possa venire ad essere definita sull'intero asse reale.*

Si è poi ritenuto opportuno prendere immediatamente in considerazione ed analizzare in dettaglio talune proprietà elementari della funzione di Dickman $\rho(u)$, sintetizzate nella seguente proposizione.

Proposizione 1.1. *La funzione di Dickman* $\rho(u)$ *soddisfa le seguenti proprietà:*

$$\begin{array}{ll} \textit{(i)} & u\rho(u) = \displaystyle\int_{u-1}^{u} \rho(v)dv \qquad (u \in \mathbb{R}), \\ \textit{(ii)} & \rho(u) > 0 \qquad (u \geq 0), \\ \textit{(iii)} & \rho'(u) < 0 \qquad (u > 1), \\ \textit{(iv)} & 0 < \rho(u) \leq \dfrac{1}{\Gamma(u+1)} \qquad (u \geq 0), \end{array}$$

dove si denoti, come di consueto, con Γ *la funzione gamma di Eulero.*

Successivamente si è analizzata la letteratura degli ultimi tre decenni desumendone talune stime di questa peculiare funzione particolarmente interessanti da un punto di vista teorico ed al contempo anche piuttosto utili nelle applicazioni pratiche, per poi passare a prendere in considerazione alcuni risultati interamente espliciti concernenti i numeri lisci e la funzione $\psi(x, y)$. In primo luogo, a partire da argomentazioni di calcolo combinatorio, si è introdotta e dimostrata in dettaglio la seguente stima.

Proposizione 1.2. *Risulta verificata la disuguaglianza*

$$\binom{[\frac{\log x}{\log 2}] + \pi(y)}{\pi(y)} \geq \psi(x, y) \geq \binom{[u] + \pi(y)}{\pi(y)} \geq \left(\frac{\pi(y)}{[u]}\right)^{[u]},$$

dove, come di consueto, si sia posto $u = \log x / \log y$.

Si è quindi proposto un ulteriore metodo di carattere eminentemente geometrico che consente di ricavare limiti inferiore e superiore per la funzione $\psi(x,y)$, determinando il numero dei punti di reticolo a coordinate intere contenuti all'interno di opportuni tetraedri multidimensionali. Si è pervenuti in questa maniera ad una stima che può essere formalizzata nella maniera seguente.

Proposizione 1.3. *Risultano verificate le disuguaglianze*

$$\frac{(\log x)^{\pi(y)}}{(\pi(y))!\prod_{p\le y}\log p}\le\psi(x,y)\le\frac{(\log X)^{\pi(y)}}{(\pi(y))!\prod_{p\le y}\log p}\ ,$$

dove si sia posto $X=x\prod_{p\le y}p$ *.*

Si è poi avuto modo di constatare come l'applicazione di opportuni metodi numerici di setacciamento possa consentire di ricavare delle stime per la funzione $\psi(x,y)$ su cui si è focalizzata la nostra analisi. In tale frangente si è ad esempio avuto modo di osservare che risulta verificata la seguente stima asintotica, che fornisce ulteriori informazioni relative all'andamento della funzione di Dickman (seppure per il momento solamente all'interno di un intervallo dell'asse reale di ampiezza notevolmente limitata).

Proposizione 1.5. *Se* $x=y^u$*, allora sussiste la seguente relazione*

$$\psi(x,x^{1/u})\sim x(1-\log u)\qquad \textit{per } 1\le u\le 2.$$

Ritornando poi ad un approccio di carattere combinatorio, si è pervenuti ad una prima stima della funzione $\psi(x,y)$ di carattere generale, per valori maggiori del parametro $u:=\frac{\log x}{\log y}$; difatti applicando al problema in esame il principio di inclusione-esclusione si ottiene agevolmente che

$$\psi(x,y)=x\prod_{y<p\le x}(1-1/p)-\sum_{p|d\Rightarrow y<p\le x}\mu(d)\Big\{\frac{x}{d}\Big\},$$

avendo denotato con μ la funzione aritmetica di Moebius. È anche possibile raffinare ulteriormente quest'ultima stima sino ad ottenere

$$\psi(x,y)=x\sum_{\substack{d\le x,\\ p|d\Rightarrow y<p\le x}}\frac{\mu(d)}{d}+O\Big(\frac{x}{\log y}\Big).$$

Il passo successivo, compiuto con l'intento di ricavare stime dall'applicabilità più generale per quanto concerne la funzione $\psi(x,y)$, è consistito nell'introdurre due equazioni funzionali di basilare rilevanza. Su tali identità è stata poi fondata la dimostrazione del risultato teorico principale nell'ambito dell'analisi delle proprietà della funzione $\psi(x,y)$, centro essenziale della nostra trattazione, ovvero il Teorema di Dickman sopra enunciato.
Storicamente la prima equazione funzionale ad essere derivata per $\psi(x,y)$ è stata l'**Identità di Buchstab-de Bruijn**, scoperta da Buchstab nel 1949 ([2]), la quale ha rappresentato il punto di partenza della stragrande maggioranza della ricerca in questo specifico settore della Teoria dei Numeri nei quattro decenni successivi e può essere enunciata nella seguente maniera.
Proposizione 1.6. (*identità di Buchstab-de Bruijn*):

$$\psi(x,y)=\psi(x,z)-\sum_{y<p\leq z}\psi\Big(\frac{x}{p},p\Big)\qquad (con\ 1\leq y<z\leq x);$$

nel caso speciale in cui si ponga $y=1$ otterremo in particolare la seguente relazione

$$\psi(x,z)=1+\sum_{p\leq z}\psi\Big(\frac{x}{p},p\Big)\qquad (con\ z>1).$$

Successivamente, in un articolo particolarmente rappresentativo per la moderna psixyologia (questo il termine originalmente ed ironicamente coniato per la materia oggetto della nostra trattazione da parte dell'olandese Pieter Moree nel 1993 in [3]) pubblicato nel 1986 ([4]), Hildebrand ha ricavato un'altra fondamentale equazione funzionale, la cui deduzione era stata ispirata da similari equazioni comparse nel lavoro di Delange sulle funzioni aritmetiche moltiplicative: tale equazione va consuetamente sotto il nome di ***Identità di (Chebyshev)-Hildebrand***.
Proposizione 1.7. (*identità di (Chebyshev)-Hildebrand*):

$$\psi(x,y)\log x=\int_1^x\psi(t,y)\frac{dt}{t}+\sum_{p^m\leq x,p\leq y}\psi\Big(\frac{x}{p^m},y\Big)\log p\ .$$

Si è quindi tornati a prendere in considerazione il Teorema di Dickman, che storicamente e computazionalmente ha assunto un ruolo centrale nel problema di stimare $\psi(x,y)$ e, basandosi su tali identità, se ne sono proposte due dimostrazioni alternative.

I paragrafi successivi sono dedicati ad ulteriori approfondimenti concernenti l' andamento di $\psi(x,y)$. Dapprima si è mostrato che è possibile valutare la funzione di Dickman $\rho(u)$ mediante un semplice procedimento iterativo, il quale conduce alla seguente formula che consente di approssimarla con notevole efficacia:

$$\begin{aligned}\rho(u) &= 1-\int_{t_1=1}^{u}\frac{dt_1}{t_1}+\frac{1}{2!}\int_{\substack{t_1=1,\\ t_1+t_2\le u}}^{u}\int_{t_2=1}^{u}\frac{dt_1dt_2}{t_1t_2}-\cdots\\ &+ \frac{(-1)^k}{k!}\int_{\substack{t_1=1,\\ t_1+t_2+\cdots+t_k\le u}}^{u}\int_{t_2=1}^{u}\cdots\int_{t_k=1}^{u}\frac{dt_1dt_2\cdots dt_k}{t_1t_2\cdots t_k}+\cdots\end{aligned}$$

Si è successivamente osservato che una maniera indubbiamente più sofisticata per valutare numericamente la funzione di Dickman si fonda invece sulla trasformazione di Laplace di $\rho(u)$, definita da

$$\widehat{\rho}(s)=\int_0^{\infty}\rho(u)e^{-us}du.$$

Procedendo secondo questa strategia operativa si può dimostrare che

$$\begin{aligned}\rho(u) &= \frac{1}{2i\pi}\int_{\Re(s)=\alpha}\widehat{\rho}(s)e^{us}ds\\ &= c\int_{\Re(s)=\alpha}\exp\Big(us-\int_0^s\frac{1-e^{-t}}{t}dt\Big)ds,\end{aligned}$$

dove si è posto $c=\widehat{\rho}(0)/(2i\pi)(=e^{\gamma}/(2i\pi))$ (con $\gamma=\lim_{n\to\infty}\big(\sum_{k=1}^{n}\frac{1}{k}-\log(n)\big)\approx 0,57721566$, costante di Eulero-Mascheroni) ed α rappresenta un opportuno numero reale più grande della parte reale di tutte le singolarità della funzione $\widehat{\rho}(s)$. Anche in questo caso si tratta sfortunatamente di una formula molto difficile da utilizzare da un punto di vista computazionale, sebbene sia possibile dedurne la stima asintotica

$$\rho(u)=\Big(1+O\Big(\frac{1}{u}\Big)\Big)\sqrt{\frac{\xi'(u)}{2\pi}}\,\exp\Big(\gamma-u\xi+\int_0^{\xi}\frac{e^t-1}{t}dt\Big),$$

dove $\xi=\xi(u)$ rappresenta l'unica soluzione positiva dell'equazione trascendente individuata da $e^{\xi}=1+\xi u$ (con $u>0, u\neq 1$).

Infine, a conclusione di questo primo capitolo in cui si è offerta una panoramica sullo stato attuale delle conoscenze relative ai numeri lisci e alle modalità di approssimazione della funzione $\psi(x, y)$ sono state presentate altre due basilari tecniche analitiche che vengono frequentemente adottate con tale finalità: il metodo di Rankin ed il metodo del punto di sella. Nel 1938 difatti l'eminente studioso scozzese di Teoria dei Numeri R.A. Rankin, focalizzando la propria attenzione sugli intervalli che intercorrono tra primi consecutivi, introdusse (in [5]) una semplice ma particolarmente interessante metodologia per stimare la funzione oggetto della nostra trattazione, la quale si è rivelata estremamente efficace ed al contempo suscettibile di applicazioni in numerose situazioni. Tale tecnica analitica, consuetamente denominata *metodo di Rankin* dal nome del suo illustre ideatore, permette di individuare una semplice, ma ciò nonostante utile, limitazione superiore per la funzione $\psi(x, y)$ e si basa sull'immediata constatazione che, per ogni $\sigma > 0$, $x \geq 1$ ed $y \geq 2$, si ha (osservando banalmente che, se $n \leq x$ e $\sigma > 0$, allora risulta essere $(x/n)^\sigma \geq 1$) che:

$$\psi(x, y) \leq \sum_{\substack{1 \leq n \leq x, \\ P(n) \leq y}} \left(\frac{x}{n}\right)^\sigma = x^\sigma \sum_{\substack{1 \leq n \leq x, \\ P(n) \leq y}} \frac{1}{n^\sigma} = x^\sigma \prod_{p \leq y} \left(1 - \frac{1}{p^\sigma}\right)^{-1} = x^\sigma \zeta(\sigma, y),$$

dove $\zeta(s, y)$ rappresenta il prodotto di Eulero parziale sino ad y per la funzione zeta di Riemann, individuata da $\zeta(s) := \Sigma_{n=1}^{\infty} \frac{1}{n^s}$. Minimizzando il membro di destra della relazione precedente allo scopo di ottenere una stima il più possibile accurata, si è avuto modo di osservare che esiste un'unica soluzione $\sigma = \alpha(x, y)$ che presenta il seguente andamento

$$\alpha(x, y) = \frac{\log(1 + y/\log x)}{\log y}\left(1 + O\left(\frac{\log\log(1 + y)}{\log y}\right)\right) \approx 1 - \frac{u \log u}{\log y}.$$

Si potrebbe obiettare che questa formula per esprimere la funzione $\alpha(x, y)$ sia piuttosto tecnica ed artificiosa da sostituire nell'espressione precedente ed in particolare si riveli asintoticamente più grande rispetto alla soluzione esatta di un esiguo fattore pari ad $\alpha(x, y)\sqrt{2\pi\varphi_2(\alpha, y)}$, dove si sia posto $\varphi_k(s, y) := \frac{\partial^k}{\partial s^k} \log \zeta(s, y)$ per $k \in \mathbb{N}$; comunque tale formula presenta il rilevante vantaggio di costituire un metodo relativamente semplice per ottenere un efficace limite superiore per la funzione $\psi(x, y)$ in tutti gli intervalli per le variabili x ed y, risultato che chiaramente si rivela estremamente utile per le applicazioni pratiche. Tale limite superiore $x^\alpha\zeta(\alpha, y)$ è solitamente conosciuto come *limite superiore di Rankin* per la funzione $\psi(x, y)$.

Per quanto concerne invece il metodo del punto di sella, Hildebrand e Tenenbaum, sviluppando nella loro pubblicazione [6] del 1986 un approccio già introdotto dall'olandese de Bruijn, hanno usato questo metodo numerico allo scopo di ottenere una approssimazione asintotica per la funzione $\psi(x, y)$ in tutti quegli intervalli all'interno dei quali non se ne sarebbe potuto analizzare l'andamento sfruttando altre strategie operative. Partendo della formula integrale di Perron, si può dedurre che

$$\psi(x, y) = \sum_{\substack{n\geq 1,\\ p|n\Rightarrow p\leq y}} \int_{\Re s=\alpha} \frac{(x/n)^s}{s}\, ds + O(1) = \int_{\Re s=\alpha} \zeta(s, y)\frac{x^s}{s}\, ds + O(1).$$

Il metodo del punto di sella sfrutta sostanzialmente l'assunzione che il parametro α possa essere scelto in modo essenzialmente arbitrario: la linea di integrazione $\Re s = \alpha$ viene selezionata in maniera tale che il contributo principale all'integrale provenga da un intorno di ampiezza estremamente limitata localizzato attorno al punto $s = \alpha$ (da cui appunto la denominazione di metodo del "punto di sella"). Si suppone a questo punto di scegliere $\alpha = \alpha(x, y)$ (il punto di ottimizzazione precedentemente individuato nell'ambito della trattazione finalizzata a ricavare il limite superiore di Rankin) e si definiscono le funzioni $\varphi(s, y) = \log\zeta(s, y)$, $\varphi_k = \varphi_k(s, y) = \frac{d^k}{ds^k}\,\varphi(s, y)$ (per $k \geq 1$), con $\varphi_0 = \varphi$. In questa maniera è immediatamente possibile asserire che risulta essere $\log x + \varphi_1(\alpha, y) = 0$ e si potrebbe anche mostrare facilmente che il più consistente contributo a questo integrale deriva da un segmento di ampiezza notevolmente limitata e centrato in corrispondenza di $\alpha(x, y)$, il corrispondente *punto di sella*, cosicchè si viene ad avere che

$$\psi(x, y) = \frac{1}{2\pi i}\int_{\alpha(x,y)-i/\log y}^{\alpha(x,y)+i/\log y} \zeta(s, y)x^s\frac{ds}{s} + \text{(piccolo errore)}$$

ed, a meno di manipolazioni analitiche, si è pertanto arrivati a concludere che, per ogni $x \geq y \geq 2$ è verificata la stima

$$\psi(x, y) = \frac{x^\alpha\zeta(\alpha, y)}{\alpha\sqrt{2\pi\varphi_2(\alpha, y)}}\left(1 + O\left(\frac{1}{u} + \frac{\log y}{y}\right)\right)$$

(risultato inferiore all'estremo superiore ricavato mediante il metodo di Rankin di un fattore approssimativamente pari ad $\alpha\sqrt{\varphi_2(\alpha, y)} \ll \log x$).

Dopo questa analisi dei basilari fondamenti della psixyologia e delle più interessanti proprietà teorico-matematiche dei numeri lisci, nel **secondo capitolo** si è iniziato a passare in rassegna talune applicazioni algoritmiche. Innanzitutto è stata presentata un'efficace metodologia numerica, ideata dagli studiosi Hunter e Sorenson (si veda [7] del 1997), che consente di approssimare la funzione $\psi(x, y)$ in maniera accurata e computazionalmente accessibile, nella fattispecie pervenendo ad un'approssimazione sino ad un ordine di precisione $1 + O\left(\frac{1}{u} + \frac{\log y}{y}\right)$ in $O\left(y\left\{\frac{\log\log x}{\log y} + \frac{1}{\log\log y}\right\}\right)$ operazioni floating point. Tale algoritmo è fondato essenzialmente sul basilare risultato teorico ricavato da Hildebrand e Tenenbaum (sempre in [6]) ricorrendo al metodo del punto di sella, stima già presa in considerazione in precedenza e che può essere formalizzata in maniera rigorosa enunciando il seguente:

Teorema 2.2. (Teorema di Hildebrand e Tenenbaum)
Sia $\bar{u} := \bar{u}(x, y) = \min\{\log x, y\}/\log y = \min\{u, y/\log y\}$ e si ponga $HT(x, y, s) := \frac{x^s\zeta(s,y)}{s\sqrt{2\pi\varphi_2(s,y)}}$. Sia $\alpha = \alpha(x, y)$ l'unica soluzione dell'equazione $\varphi_1(\alpha, y) + \log x = 0$. Allora si avrà che

$$\psi(x, y) = HT(x, y, \alpha(x, y))(1 + O(1/\bar{u}))$$

uniformemente per $2 \leq y \leq x$.

L'algoritmo proposto da Hunter e Sorenson (e da loro denotato con HT dalle iniziali di coloro che hanno fornito la base teorica per la sua realizzazione, appunto Hildebrand e Tenenbaum) consente pertanto di approssimare la funzione $\psi(x, y)$ semplicemente calcolando la grandezza $HT(x, y, \alpha')$, dove α' sia un'opportuna approssimazione di α. Una sua pseudocodifica può essere allora articolata nelle seguenti operazioni algoritmiche elementari.

<u>**Algoritmo HT.**</u> I passi principali dell'algoritmo di approssimazione fondato sul teorema di Hildebrand e Tenenbaum sono i seguenti.

1. individuare tutti i numeri primi $p \leq y$;

2. calcolare un'approssimazione α' per la soluzione $s = \alpha$ dell'equazione $f(s) = 0$, dove si sia posto $f(s) := \varphi_1(s, y) + \log x$; affinchè tale approssimazione risulti essere

sufficientemente accurata, si richiede che

$$|\alpha' - \alpha| < \min\{0.0001, 1/(\bar{u}\log x)\};$$

3. restituire in output $HT(x, y, \alpha')$.

Comunemente la soluzione α può essere efficacemente approssimata nell'ambito del secondo passo del procedimento algoritmico appena proposto ricorrendo al metodo numerico di bisezione; allo scopo di ottimizzare il tempo di esecuzione dell'algoritmo HT, il metodo di Newton potrebbe però essere sostituito alla meno performante strategia di bisezione; difatti, a patto di operare seguendo opportuni accorgimenti, è possibile dimostrare effettivamente che il metodo di Newton converge quadraticamente alla soluzione $s = \alpha$ dell'equazione $f(s) = 0$ e questo consentirà in taluni casi particolari (essenzialmente allorchè si abbia che $y \geq \log x$) di migliorare leggermente il tempo di esecuzione che contraddistingue l'algoritmo HT sino ad $O(y/(\log\log y))$ operazioni elementari.

Successivamente si è passati ad analizzare due algoritmi per determinare tutti i numeri lisci presenti all'interno di un determinato intervallo, obiettivo sicuramente di notevole rilevanza per le applicazioni pratiche di carattere numerico, in special modo concernenti l'ambito crittografico. Il primo metodo presentato è quello più naturale ed immediato e consiste essenzialmente in un'operazione di setacciamento progressivo analogo a quello usualmente realizzato nel crivello di Eratostene, applicato con l'intento di identificare con precisione tutti i numeri y-lisci presenti in un dato intervallo $(x, x+z)$, con $z \leq x$. La semplicità di tale procedimento algoritmico comporta però come controparte il fatto che è caratterizzato da un tempo di esecuzione $\ll z\log\log y + uy$, dunque piuttosto proibitivo (d'altra parte appare inevitabile che un qualunque algoritmo di siffatta tipologia debba presentare un tempo di esecuzione necessariamente $\geq \pi(y)$, dal momento che evidentemente i numeri primi minori o uguali ad y rappresentano parte integrante (ed essenziale) dei dati di input del problema considerato). Sicuramente è invece al contempo più costruttiva e computazionalmente meno onerosa la strategia operativa sostanzialmente diversa proposta da Dan Boneh, nell'ambito di una sua pubblicazione [8] del 2002, per l' individuazione dei numeri lisci presenti in un determinato intervallo, elegante strategia che conduce ad un algoritmo contraddistinto da un tempo di esecuzione di approssimativamente $(y \log x)^{O(1)}$ operazioni elementari.

Il passo successivo è stato poi quello di proporre opportune metodologie computazionali per determinare limiti superiori ed inferiori alla funzione $\psi(x, y)$. In proposito si è rammentato come Bernstein (nel suo articolo [9] del 2002) abbia avuto modo di sottolineare che talune semplici argomentazioni geometriche fondate sulla nozione di punti di reticolo possano consentire di ottenere piuttosto agevolmente delle significative limitazioni superiori ed inferiori. In questo paragrafo si è pertanto dimostrato in dettaglio che, dato un intero N di considerevoli dimensioni, selezionando per ciascun numero primo $p_j \leq y$ il più piccolo intero m_j per cui si abbia che $m_j \geq N \log p_j$, allora il numero delle soluzioni della disuguaglianza individuata da

$$a_1 m_1 + \cdots + a_k m_k \leq d$$

soddisfacenti la condizione $d = [N \log x]$ fornisce un interessante limite inferiore per la funzione $\psi(x, y)$; in maniera sostanzialmente analoga si può altresì immediatamente osservare che il numero delle soluzioni della medesima relazione soddisfacenti la condizione $d = [N \log X]$ (dove si sia posto $X = x^{1+1/(N \log 2)}$), restituisce anche il limite superiore desiderato per la funzione $\psi(x, y)$. Si è inoltre avuto modo di evidenziare che, al crescere del parametro N inizialmente assunto in maniera arbitraria, le limitazioni ricavate per la funzione $\psi(x, y)$ secondo questo metodo divengono sempre più accurate.

La parte conclusiva di questo secondo capitolo è stata infine dedicata all'analisi delle proprietà dei cosiddetti *primi traslati lisci*, ossia di quei peculiari numeri primi p, per i quali $p-1$ risulta essere un numero liscio oppure possiede un fattore liscio di ragguardevoli dimensioni. A tale proposito si sono presentate limitazioni superiori estremamente interessanti al numero di primi $p \leq x$ per i quali $p-1$ risulti essere y-liscio e, come corollario, si sono dedotte stime concernenti la cardinalità dell'insieme dei primi dispari $p \leq x$ per i quali l'ordine moltiplicativo $l(p)$ di 2 modulo p risulti essere y-liscio. Si è poi anche avuta l'occasione di passare in rassegna talune interessanti applicazioni crittografiche delle considerazioni effettuate relativamente a questa specifica tipologia di interi. Dalla sostanziale esiguità del numero dei primi traslati lisci scaturisce difatti ad esempio il fatto che molteplici rilevanti algoritmi numerici, tra i quali è possibile annoverare il ben conosciuto metodo $p-1$ di Pollard e numerose tipologie di algoritmi per la fattorizzazione di polinomi, non possano essere quasi mai contraddistinti da un tempo di esecuzione di carattere polinomiale. Alla luce delle stime precedentemente ricavate per gli ordini moltiplicativi di 2 modulo p, si è poi valutata l'opportunità di utilizzare proprio $g = 2$ in qualità di generatore per costruzioni crittografiche fondate su di un procedimento di esponenziazione (tra le quali rammentiamo lo schema basato sullo scambio di chiavi Diffie-Hellman, il crittosistema El Gamal oppure l'Algoritmo di Firma Digitale DSA) e si è osservato che in questo modo si riduce nettamente il costo relativo all'esponenziazione e dunque la complessità computazionale complessiva degli algoritmi in questione.

Infine nel **terzo ed ultimo capitolo** è stata focalizzata l'attenzione su questioni più eminentemente concernenti la Teoria dei Numeri Computazionale e la crittografia. Difatti i numeri lisci rappresentano un prezioso strumento in Teoria dei Numeri dal momento che, non solo sono caratterizzati da una semplice struttura moltiplicativa, ma risultano anche essere piuttosto numerosi; l'accoppiamento di queste due proprietà dei numeri lisci rappresenta la motivazione principale che ne giustifica il ruolo chiave in pressochè ogni singolo moderno algoritmo per la fattorizzazione di interi.

In primo luogo si è constatato che un passo significativo in numerosi algoritmi di Teoria dei Numeri Computazionale è quello di determinare, il più velocemente possibile, un opportuno sottoinsieme non vuoto di una sequenza $m_1, m_2, \ldots$ di interi minori o uguali di una prefissata soglia x, il cui prodotto rappresenti un quadrato perfetto ed in proposito si è avuto modo di constatare come i numeri lisci svolgano un ruolo essenziale nella soluzione sia teorica che algoritmica di questo problema. Si supponga di costruire una sequenza numerica scegliendo interi indipendentemente e con distribuzione uniforme nell'intervallo $[1, x]$: è legittimo domandarsi quanti numeri andranno individuati in questa maniera affinchè quasi sicuramente esista un opportuno sottoinsieme non vuoto degli interi prescelti, il cui prodotto risulti essere un quadrato perfetto. Si è a tal proposito osservato che la risposta a tale interrogativo dipende dall'andamento della funzione $L(x) = \exp\left(\sqrt{\log x \log\log x}\right)$, come stabilito dal seguente risultato teorico.

Proposizione 3.1. *Sia ε un numero positivo arbitrariamente piccolo. Se si selezionano $L(x)^{\sqrt{2}+\varepsilon}$ interi appartenenti all'intervallo $[1, x]$ (indipendentemente e con distribuzione uniforme), allora, per $x \to \infty$, la probabilità che esista un opportuno sottoinsieme non vuoto di questa collezione di interi, il cui prodotto sia un quadrato perfetto, tende ad 1, mentre se invece si selezionano $L(x)^{\sqrt{2}-\varepsilon}$ interi, allora tale probabilità tende a 0.*

Alla luce della Proposizione 3.1 sì è pertanto evidenziato che, qualora venga costruita una sequenza numerica costituita da $N = L(x)^{\sqrt{2}+\varepsilon}$ interi $m_1, \ldots, m_N$, allora è possibile asserire che, con una probabilità tendente ad uno, ne esisterà un opportuno sottoinsieme non vuoto costituito esclusivamente da interi $L(x)^{1/\sqrt{2}}$-lisci il cui prodotto rappresenti un quadrato perfetto. A partire da quest'ultima osservazione si è inoltre pervenuti alla realizzazione di un algoritmo di algebra lineare estremamente semplice ed intuitivo che consente effettivamente l'individuazione di tale peculiare sottoinsieme. A ciascun intero n, che sia y-liscio, è possibile difatti associare un determinato *vettore esponente* $\overrightarrow{v}(n)$ di dimensione esattamente pari al numero dei primi minori o uguali ad y: infatti, se $p \leq y$ risulta essere primo, allora la coordinata in $\overrightarrow{v}(n)$ corrispondente a p viene individuata dall'esponente di p nella fattorizzazione in primi distinti dell'intero n. Si può conseguentemente asserire che, qualora si abbiano a disposizione più di $\pi(y)$ di tali vettori esponente, allora questi risultano necessariamente linearmente dipendenti ed, in particolare, sono dipendenti sullo spazio vettoriale $\mathbb{F}_2^{\pi(y)}$. La relazione di dipendenza in questione può essere pertanto rappresentata da una sommatoria, effettuata su un opportuno sottoinsieme non vuoto di tali vettori, e che fornisca come risultato il vettore nullo; i vettori esponente appartenenti al sottoinsieme individuato in questa maniera corrisponderanno allora esattamente ai numeri interi il cui prodotto costituisca un quadrato perfetto, che ci si era sin dall'inizio riproposti di determinare esplicitamente.

Alla luce di queste considerazioni si è poi iniziato ad analizzare in dettaglio la stretta connessione che intercorre tra le proprietà caratteristiche dei numeri lisci e i problemi numerici di fattorizzazione, cominciando ad analizzare alcuni tra i metodi per la fattorizzazione di interi che storicamente hanno assunto maggiore rilevanza sia da un punto di vista teorico che nelle applicazioni pratiche: il metodo di Fermat, il metodo dei quadrati casuali di Dixon ([10]) ed il metodo delle frazioni continue di Brillhart e Morrison ([11]).

Per quanto concerne gli ultimi due procedimenti algoritmici appena citati si è rivelata in particolare un'operazione di importanza cruciale quella di stabilire se taluni opportuni interi ausiliari generati nell'ambito dell'elaborazione numerica godano o meno della proprietà di essere lisci: si è posta allora la necessità di domandarsi, allorchè si prenda in considerazione un generico intero $n \leq x$, quale sia la complessità computazionale associata a quest'operazione consistente nello stabilire se tale numero n risulti essere y-liscio. Una prima immediata ipotesi operativa potrebbe essere l'applicazione del semplice algoritmo "trial division", tuttavia tale procedimento risulta indiscutibilmente troppo oneroso da un punto di vista computazionale perchè possa rivelarsi una scelta effettivamente valida. Per ovviare a tale problema sono stati dunque presentati ed analizzati in dettaglio due metodi numerici, la *strategia di interruzione anticipata* (si veda Pomerance, [12]) ed il *metodo di valutazione polinomiale (o dei "fattoriali veloci") di Pollard-Strassen* (si veda Strassen, [13]), la cui applicazione combinata in luogo del rudimentale algoritmo "trial division" consente di abbattere drasticamente il tempo di esecuzione della fase di individuazione di ogni singolo numero liscio, comportando perciò una riduzione assolutamente significativa della complessità computazionale degli algoritmi di fattorizzazione considerati.

Si è d'altra parte avuto modo di constatare che, in talune circostanze, per rispondere efficacemente all'esigenza di stabilire se un determinato intero risulta essere liscio, non è necessario ricorrere a metodologie tanto sofisticate e potenti come quelle appena prese in considerazione; difatti, qualora ad esempio si prenda in esame una sequenza pseudorandomica i cui elementi rappresentino i valori consecutivi di un opportuno polinomio a coefficienti interi, è possibile adottare un semplice test dalle prestazioni alquanto soddisfacenti per l'individuazione esplicita dei numeri lisci. Tale strategia numerica è fondata essenzialmente sul *crivello di Eratostene*, universalmente conosciuto come un semplice metodo per individuare tutti i numeri primi sino ad una prefissata soglia. Apportando una lieve modifica a questo procedimento algoritmico iterativo, oltre a permettere di discernere i numeri primi da quelli composti presenti in un determinato intervallo dell'asse reale, è possibile anche costruire delle fattorizzazioni complete degli interi composti individuati in questa maniera. Si è infine osservato che, mediante un'opportuna semplificazione, questo procedimento può essere anche efficacemente trasformato in una tecnica algoritmica in grado di riconoscere tutti i numeri lisci presenti all'interno dell'intervallo considerato. Inoltre si è constatato che il metodo ottenuto operando in questo modo prevede in definitiva l'esecuzione di operazioni elementari rappresentate da istruzioni particolarmente semplici, conducendo ad una complessità computazionale dell'ordine $O(\log \log y)$ per ciascun candidato di cui ci si proponga di stabilire se si tratta di un intero liscio, miglioramento ragguardevolissimo ed immediatamente ravvisabile se confrontato con le $\pi(y)$ operazioni per ciascun candidato previste dall'algoritmo "trial division".

Il paragrafo conclusivo del capitolo descrive infine in dettaglio due tra i metodi per la fattorizzazione di interi maggiormente efficienti ed interessanti, il *Metodo del Crivello Quadratico* ([12],[14]) ed il *Metodo del Crivello del Campo Numerico* ([15],[16],[17]), i quali, pur essendo diretti discendenti di altri algoritmi precedenti (tra i quali ad esempio lo stesso metodo delle frazioni continue di Brillhart e Morrison), hanno consentito un netto miglioramento delle capacità di fattorizzazione e nel contempo anche una considerevolissima diminuzione della complessità computazionale. Per entrambi gli algoritmi si è dunque ritenuto opportuno proporre una descrizione estremamente dettagliata della metodologia operativa che li contraddistingue, un calcolo esaustivo e preciso del costo computazionale ad essi associato ed un'analisi approfondita sia dei principali inconvenienti che si manifestano nella loro implementazione sia delle soluzioni che comunemente si adottano per risolverli.

Bibliografia

[1] K. Dickman, *On the frequency of numbers containing prime factors of a certain relative magnitude*, Ark. Mat. Astr. Fys. **22** (1930), 1-14.

[2] A. A. Buchstab, *On those numbers in an arithmetic progression all prime factors of which are small in order of magnitude*,
Doklady Akad. Nauk SSSR (N.S.) **67** (1949), 5-8.

[3] P. Moree, *Psixyology and Diophantine equations*,
Ph.D. Thesis, Rijksuniversiteit Leiden, 1993.

[4] A. Hildebrand, *On the number of positive integers* $\leq x$ *and free of prime factors* $> y$, Journal of Number Theory **22**:3 (1986), 289-307.

[5] R.A. Rankin, *The difference between consecutive prime numbers*, J. London Math. Soc. **13** (1938), pagg.242-247.

[6] A. Hildebrand e G. Tenenbaum, *On integers free of large prime factors*, Trans. Amer. Math. Soc. **296** (1986), pagg.265-290.

[7] S. Hunter e J. Sorenson, *Approximating the number of integers free of large prime factors*, Mathematics of computation, Volume 66, **220** (1997), pagg.1729-1741.

[8] D. Boneh, *Finding smooth integers in short intervals using CRT decoding*, J.Comput.System Sci. **64**:4 (2002), pagg.768-784.

[9] D.J. Bernstein, *Arbitrarily tight bounds on the distribution of smooth integers*, Number Theory for the millennium, I (Urbana, IL, 2000), pubblicato da M.A.Bennett, A K Peters, Natick, MA(2002), pagg.49-66.

[10] J. Dixon, *Asymptotically fast factorization of integers*, Math. Comp. **36**, (1981) pagg.255-260.

[11] M. Morrison e J. Brillhart, *A method of factoring and the factorization of F_7*, Math. Comp., **29** (1975) pagg. 183-205. Collection of articles dedicated to Derrick Henry Lehmer on the occasion of his seventieth birthday.

[12] C. Pomerance, *Analysis and comparison of some integer factoring algorithms*, Computational Methods in Number Theory (H.W. Lenstra,Jr. ed R. Tijdeman, eds.), Math. Centre Tracts **154/155**, Mathematisch Centrum, Amsterdam (1982) pagg. 89-139.

[13] V. Strassen, *Einige Resultate über Berechnungskomplexität*, Jahresber. Deutsch. Math. Verein **78** (1976/77) pagg. 1-8.

[14] C. Pomerance, *The quadratic sieve factoring algorithm*, Advances in Cryptology - Proceedings of EUROCRYPT 84 (T. Beth ed altri,eds), Lecture Notes in Computer Science **209**, Springer-Verlag, Berlin and New York (1985) pagg. 169-182.

[15] J.P. Buhler, H.W. Lenstra e C. Pomerance, *Factoring integers with the number field sieve*, in [16] pagg.50-94.

[16] A.K. Lenstra ed H.W. Lenstra, *The development of the number field sieve*, Lecture Notes in Mathematics, Springer-Verlag, Berlin and New York **1554** (1993)

[17] C. Pomerance, *The number field sieve*, Mathematics of Computation 1943-1993: A Half-Century of Computational Mathematics (W. Gautschi,ed.), Proc. Sympos. Appl. Math. **48**, Amer. Math. Soc. Providence (1994) pagg.465-480.