

Una invitació a la matemàtica industrial

Discurs de presentació de Joaquim Bruna Floris
com a membre numerari de la Secció de Ciències
i Tecnologia, llegit el dia 14 d'octubre de 2019



Institut
d'Estudis
Catalans

SECCIÓ DE CIÈNCIES
I TECNOLOGIA

Una invitació
a la matemàtica industrial

Una invitació a la matemàtica industrial

Discurs de presentació de Joaquim Bruna Floris
com a membre numerari de la Secció de Ciències
i Tecnologia, llegit el dia 14 d'octubre de 2019

Barcelona, 2019



Institut
d'Estudis
Catalans

SECCIÓ DE CIÈNCIES
I TECNOLOGIA

Biblioteca de Catalunya. Dades CIP

Bruna, Joaquim, autor

Una Invitació a la matemàtica industrial. — Primera edició

Bibliografia

ISBN 9788499654874

I. Institut d'Estudis Catalans. Secció de Ciències i Tecnologia II. Títol

1. Matemàtica aplicada 2. Matemàtica — Aplicacions industrials 3. Codificació, Teoria de la
51-7

51:62/69

519.725/.728

© Joaquim Bruna Floris

© 2019, Institut d'Estudis Catalans, per a aquesta edició

Carrer del Carme, 47. 08001 Barcelona

Primera edició: octubre del 2019

Text revisat lingüísticament per la Unitat de Correcció del Servei Editorial de l'IEC

Disseny de la coberta: Azcunce | Ventura

Compost per Rosa Rodríguez

Imprès a Open Print, SL

ISBN: 978-84-9965-487-4

Dipòsit Legal: B 21753-2019

Són rigorosament prohibides, sense l'autorització escrita dels titulars del *copyright*, la reproducció total o parcial d'aquesta obra per qualsevol procediment i suport, incloent-hi la reprografia i el tractament informàtic, la distribució d'exemplars mitjançant lloguer o préstec comercial, la inclusió total o parcial en bases de dades i la consulta a través de xarxa telemàtica o d'Internet. Les infraccions d'aquests drets estan sotmeses a les sancions establertes per les lleis.

L'any 2012 jo era director del Centre de Recerca Matemàtica (CRM), un centre CERCA ubicat a la Facultat de Ciències de la Universitat Autònoma de Barcelona (UAB). Un dia vàrem rebre un correu electrònic d'ACCIÓ, l'agència del Departament d'Indústria que promou el contacte de les empreses amb l'acadèmia. Buscaven qui s'interessés per una sol·licitud de contacte que havia fet una empresa de Breda, Hohner Automáticos, SL. Aquesta empresa es dedica a la fabricació de codificadors (o *encoders*) òptics, i és la més important de l'Estat espanyol en el sector. Jo en aquell temps no tenia ni idea de què era un codificador òptic.

Des de fa un temps m'he interessat en el que ara en diuen *matemàtica industrial* (MI). Crec que una definició acceptada per tothom seria que la matemàtica industrial s'ocupa dels problemes amb contingut matemàtic que són d'interès industrial immediat, plantejats per necessitats de producció, i entenent el terme *indústria* i el terme *producció* en un sentit ampli. El matís que el diferencia del terme *transferència* és que la matemàtica industrial comporta, en general, fer recerca, crear models matemàtics nous i desenvolupar-los. Quan era director del Departament de Matemàtiques de la UAB, vaig fundar i dirigir el Servei de Consultoria Matemàtica, engrescant-hi el professorat que podia, i, de fet, vàrem tocar uns quants problemes força interessants. I un dels grups de recerca que durant la meva gestió es

va constituir al CRM fou precisament el de Matemàtica Industrial, que avui en dia té molta empenta. En definitiva, d'uns anys ençà he intentat acostar-me a aquest món de la transferència i la matemàtica industrial, tot mantenint al mateix temps l'activitat en la meua àrea d'especialització. Per això, i pensant també que al CRM, un centre fins llavors vist com a impermeable a les aplicacions, li aniria bé de cara a futures avaluacions, vaig pensar que valia la pena aprofitar l'ocasió i respondre a la crida.

Vaig comunicar la disposició del CRM a participar en una entrevista, a l'empresa mateixa, al cap de quinze dies. Durant aquests dies, vaig fracassar en l'intent d'engrescar cap investigador del CRM, així que m'hi vaig presentar jo mateix. Recordo que en arribar-hi vaig endur-me una sorpresa, ja que els d'ACCIÓ havien convocat també, simultàniament, una consultora privada que havia respost a la crida. L'enginyer en cap de l'empresa va explicar a la seva manera la problemàtica que tenien. No en podia precisar el marc científic, però jo vaig olorar de seguida de què anava la cosa. Com que era una mica violent expressar-se davant de possibles competidors, recordo que vaig dir que analitzaria el problema i al cap d'uns dies els diria si podia ser d'ajut. La consultora va enviar un dossier, que jo no vaig veure; jo simplement vaig trucar a l'enginyer i li vaig dir que el problema podia tractar-se matemàticament i que era interessant. Li vaig mencionar tan sols una paraula clau: *convolució*.

Al cap d'una setmana l'enginyer em va trucar per dir-me que ens havia triat a nosaltres i em va citar per a una primera reunió a l'empresa. Allà em vaig trobar que l'enginyer s'havia descarregat d'Internet un munt d'entrades relacionades amb la convolució, i ell, excitadíssim per la descoberta, ja començava a entendre que aquest era el marc del seu problema.

Així va començar la petita història que explico en aquest treball. És una història que pot considerar-se d'èxit, el punt final de la qual ha estat una innovació implementada reeixidament en la producció i una patent. Però l'he triada com a tema per a la Secció de Ciències i Tecnologia sobretot per fer evident, primer, que els científics que no traïem sovint el cap fora de la universitat, que som la majoria, tenim

un potencial enorme que desconeixem i, segon, que tenim oportunitats sense gaire esforç: per exemple, dins el Pla de Doctorats Industrials de la Generalitat de Catalunya.¹ Aquest és el perquè del títol.

Fer matemàtica industrial és quelcom molt diferent de fer investigació bàsica: les eines matemàtiques que cal utilitzar no són sofisticades (una excepció en seria la matemàtica financera) i l'objectiu no és provar teoremes, cosa que sovint és molt més difícil. Però és molt més costós en temps: deixant de banda les visites a l'empresa, l'esforç principal, també en temps, és entendre el llenguatge de l'empresa, entendre el problema i trobar el marc matemàtic, el model adequat. Quan això s'ha aconseguit, en certa forma tot ja fa baixada. En el cas que presentaré, vaig ser afortunat perquè l'àmbit matemàtic és l'anàlisi de Fourier, una de les meves especialitats; com veureu, l'altre àmbit rellevant és l'òptica.

D'altra banda, dedicar-se a la matemàtica industrial dona menys rendiment, acadèmicament parlant. Les regles de joc que tenim pel que fa a la progressió acadèmica fan que no sigui gaire atractiu dedicar-s'hi: és més fàcil i aporta més trams de recerca fer un més dels nostres articles especialitzats. Tractar d'aquesta problemàtica no és l'objectiu d'aquest treball, però.

La col·laboració amb l'empresa es va canalitzar a través d'un doctorat industrial. Part del contingut d'aquest treball es basa en els resultats de la tesi [3] que vaig dirigir en el marc d'aquest pla.

1. QUÈ ÉS UN CODIFICADOR?

Un codificador és un tipus de sensor que mesura el posicionament, la velocitat i l'acceleració d'un objecte determinat. Hohner Automáticos, SL es dedica a la fabricació d'aquest tipus de sensor. En concret, fabrica codificadors lineals i rotatius que operen amb sistemes òptics o magnètics; però també produeix altres sensors, com potenciòmetres o inclinòmetres. El codificador lineal fa mesures del desplaçament en una dimensió, com podria ser el moviment d'una eina al llarg d'una recta, mentre que el codificador rotatiu mesura el

¹ Agraeixo a l'amic Joan de Solà-Morales que m'hagi animat en aquesta tria.

desplaçament angular, com podria ser el gir d'un motor. Aquí tractarem amb codificadors rotatius, que poden ser *absoluts* o *incrementals* (vegeu la figura 1). La major part d'aquesta presentació tracta sobre codificadors incrementals; tan sols en l'apartat final parlarem d'una connexió interessant dels codificadors absoluts amb la teoria de codis.



FIGURA 1. Codificador.

FONT: Hohner Automáticos, SL, sèrie 10,
<https://www.encoderhohner.com/product/serie-10>.

Per explicar l'estructura bàsica dels codificadors incrementals, abans hem de parlar d'un concepte previ, que és el de la *resolució*, que en aquest context significa la capacitat del codificador, quantificada per un nombre N , de discriminar N posicions angulars en una volta, és a dir, de mesurar canvis de posició de fins a $\frac{2\pi}{N}$ radians. Un codificador ha de ser capaç de dir en quina posició angular es troba l'eix, i la precisió d'aquesta mesura està quantificada per N .

El codificador absolut codifica cada posició angular en la qual es troba l'eix. És un sistema molt útil per determinar en quina posició està el motor en tot moment; per exemple, després d'una apagada de llum. El sistema consisteix principalment en una fila de leds que enfoquen un disc giratori, una màscara fixa i una fila de fotodíodes receptors. Cada led il·lumina una pista circular diferent del disc, de

manera que la fila de leds il·lumina un radi en cada moment. Cada radi té un seguit d'obertures opaques o transparents, i en cada moment els leds llegeixen un codi diferent, el propi de cadascun dels N radis. La màscara té un conjunt d'obertures per pista que fan d'analitzadores, cadascuna de les quals està enfocada a un fotodíode diferent. Si s'enfoca una obertura opaca, el respectiu fotodíode no rep llum i s'obté un senyal digital 0. D'altra banda, si l'obertura és transparent, el fotodíode és il·luminat i s'obté un senyal 1. La combinació de bits per pista ha de ser tal que codifiqui i discrimini cadascuna de les N posicions angulars. Generalment s'utilitzen dos tipus de codificacions: la binària i el codi Gray. La configuració binària relaciona cada increment de posició amb el següent nombre binari; en canvi, el codi Gray és un algorisme dissenyat de tal manera que a cada increment de posició introdueix un canvi en només un dels bits. A la figura 2 es pot veure el codificador absolut esquematitzat. El problema que presenten els codificadors absoluts és que la resolució màxima que poden mesurar és bastant limitada, ja que depèn del nombre de pistes que es poden imprimir radialment. És a dir, si el disc té k pistes, el codificador podrà mesurar $N = 2^k$ posicions.

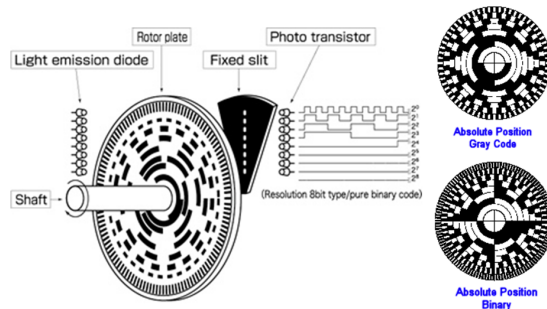


FIGURA 2. Esquema d'un codificador absolut i de les configuracions de codificació binària i de Gray.

FONT: Monolithic Power Systems,
<https://www.monolithicpower.com>.

En canvi, el codificador incremental mesura la posició angular *relativa respecte a un senyal de referència R*. Aquest tipus de codificador consta generalment d'un sol led, una lent col·limadora, un disc giratori

amb N obertures rectangulars a la perifèria, una màscara i diversos fotodíodes receptors. El led està enfocacat cap a una lent responsable de collimar un feix de llum. Aquest feix il·lumina una secció del disc giratori, el qual té disposat un seguit de petites obertures col·locades a la seva perifèria de forma periòdica. Per ser més precisos, les obertures formen un tren de N impulsos rectangulars equiseparats. A part, una segona pista interior s'encarrega de crear un sol impuls quan el disc dona una volta sencera, anomenat *senyal de referència* R i del qual parlarem després. Aquest senyal R fa, doncs, de compta-revolucions i interessa que en cada volta contingui un únic pic tan alt i localitzat com sigui possible. La màscara és fixa, i normalment conté sis finestres, anomenades A , $-A$, B , $-B$, R i $-R$. Les finestres també contenen una sèrie d'obertures rectangulars disposades de forma periòdica (excepte les de R i $-R$); ara bé, cada finestra introdueix un desfasament diferent. Quan la llum que ha passat per les obertures del disc arriba a la màscara, es troba amb les sis finestres i genera sis camps la intensitat dels quals és llegida pels fotodíodes. Així, es generen sis senyals que continuem anomenant A , $-A$, B , $-B$, R i $-R$, que seran senyals periòdics, perquè el disc va girant. Després explicarem amb detall com es pot modelitzar matemàticament aquest procediment. A la figura 3 s'exposa l'esquema gràfic d'un codificador incremental.

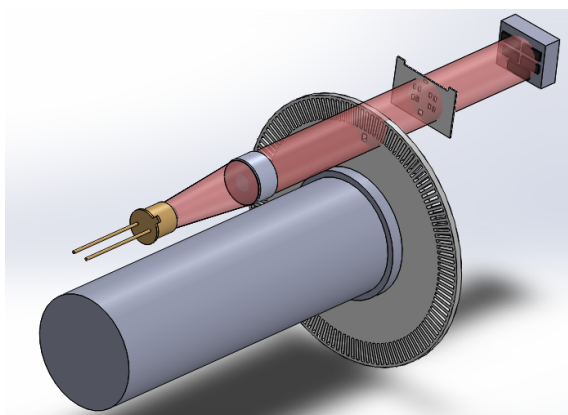


FIGURA 3. Esquema d'un codificador incremental.
 FONT: Encoder Products Company,
<https://encoder.com>.

El desfasament de 90 graus entre els senyals A i B permet saber en quin sentit està girant el codificador. Depenent de si primer s'encén A o B , es pot saber si el codificador gira en sentit horari o antihorari. D'altra banda, els senyals $-A$ i $-B$ serveixen per netejar possibles sorolls, introduint-los en un comparador diferencial amb els senyals A i B , respectivament. La màscara conté dues finestres més, anomenades R i $-R$, que registren el senyal de referència.

La gran virtut del codificador incremental és la capacitat de fer mesures amb una resolució molt alta. El nombre d'obertures pot variar des d'unes 50 fins a més de 10 000 al llarg dels 360 graus. Com més alta és la resolució, més obertures han d'estar impreses al llarg dels 360 graus, i, en conseqüència, han de ser més estretes. D'altra banda, existeix un procediment per augmentar la resolució del codificador, que és el de la *interpolació electrònica*. Mitjançant un xip, molt barat, es pot multiplicar el nombre d'impulsos per volta generats pel codificador. Per fer-ho, el xip interpolador necessita operar amb els senyals A i B (nets de soroll) com a entrades i que siguin el més sinusoidals possible. La qualitat de les sinusoides limitarà el grau d'interpolació al qual podrà arribar el codificador. Amb aquest procediment es pot aconseguir multiplicar la resolució per un factor 16 o 32.

Com en tots els camps industrials, en la fabricació de codificadors hi ha uns estàndards de procediment que segueixen tots els fabricants. En aquesta tradició, la majoria de codificadors tenen trens d'impulsos rectangulars (els més barats d'implementar) tant al disc com a la màscara. Tal com veurem amb detall després, això dona com a senyal de sortida, en una primera aproximació, un tren d'impulsos triangulars, lluny de ser sinusoidal. Aquest defecte es mira de compensar llavors dissenyant la interpolació electrònica de forma convenient. El fet, però, és que els xips que treballen amb entrades sinusoidals són els més eficients i barats.

Feta aquesta descripció bàsica de com funciona un codificador òptic incremental, procedim a fer-ne un model matemàtic, el més senzill possible. Primer de tot, cal descriure matemàticament la interacció entre el disc i la màscara, tant per generar els senyals A , B com el R . Les obertures practicades en el disc giratori que són il·luminades les

descrivim mitjançant una funció $f(x)$ de la posició angular x . Més precisament, f és una funció positiva i les obertures del disc corresponen al subgraf de f . Com que el gir té lloc a una velocitat angular constant, podem pensar també que f és una funció del temps. Aquesta funció f és el tren d'impulsos rectangulars, però de moment suposarem que pot ser qualsevol funció, ja que en principi podrien fabricar-se discs amb altres tipus d'obertures. Les obertures practicades a la màscara, a una distància z del disc, les descrivim anàlogament com el subgraf d'una altra funció positiva g , que suposarem que té un suport compacte, donat per la mida de la màscara fixa. En principi, doncs, f , g podrien ser arbitràries.

Ara descriurem de la forma més senzilla possible el senyal mesurat pels fotodíodes en cada instant. Els fotodíodes reben un senyal proporcional a la intensitat de la llum rebuda. Aquest tipus d'element consisteix en un semiconductor (normalment silici), construït amb una estructura d'unió PN o PIN, la qual és sensible a la llum en un cert rang de longituds d'ona λ . Idealment, cada fotó que incideix sobre el fotodíode excita un electró, i, per tant, es crea una intensitat de corrent elèctric. Aquesta intensitat és proporcional a la potència de la llum incident P ,

$$I = \frac{\eta_{qe}q}{h\nu}P,$$

on η_{qe} és l'eficiència quàntica del fotodíode, q la càrrega de l'electró, h la constant de Planck i ν la freqüència òptica.

Suposarem que la llum és una ona plana monocromàtica amb direcció (vector de Poynting) n perpendicular al disc i a la màscara. En aquest cas, la potència és constant en tots els punts situats a una mateixa distància z de la màscara en la direcció n . Si ignorem de moment l'efecte de la difracció, arribem a la conclusió que el senyal mesurat pels fotodíodes en cada instant t és proporcional a l'àrea comuna determinada per les obertures del disc i de la màscara en aquell instant. El gir del disc, el qual suposem uniforme, correspon a canviar $f(x)$ per $f(x - t)$. Per tant, el senyal generat és proporcional a

$$\text{CO}(f, g)(t) = \int \min[f(x - t), g(x)] dx = \int \min(f(x), g(x + t)) dx,$$

quelcom que podríem anomenar *convolució òptica* de f amb g . Com a operador en f i en g , la convolució òptica és invariant per translacions però no és lineal, fet que en fa difícil l'estudi matemàtic. Hem dit abans que en la indústria dels codificadors és un fet universal que al disc hi hagi un tren d'impulsos rectangulars (la funció f). L'explicació d'això és segurament de caire econòmic, però el fet és que des del punt de vista de la modelització matemàtica també té una explicació clara. Si f només pren els valors 0 i 1, com és el cas d'un tren d'impulsos rectangulars, i suposem que $0 \leq g \leq 1$ (que correspon a dir que les alçàries de les obertures de la màscara no depassen les del disc), aleshores $\min[f(x-t), g(x)] = f(x-t)g(x)$ i passem a tenir bàsicament la convolució

$$f * g(t) = \int_{-\infty}^{+\infty} f(x-t)g(x) dx = \int_{-\infty}^{+\infty} f(x)g(x+t) dx,$$

que és una operació lineal, invariant per translacions i, per tant, fàcilment manipulable utilitzant la transformació de Fourier.

Així, les obertures del disc que serviran per generar els senyals A , B , etc., es modelitzen amb el tren de N impulsos rectangulars equiespaiats

$$f(x) = \sum_{n=1}^N I(x - n\tau), \quad \tau = \frac{2\pi}{N},$$

on I és la funció característica de $[0, \frac{\tau}{2}]$. Aquest tren f d'impulsos rectangulars ja no el tocarem i continuarem, per tant, treballant amb els mateixos discs tradicionals.

En general, doncs, la interacció entre un senyal practicat al disc i un senyal practicat a la màscara es modelitza amb la convolució òptica. Si un d'ells tan sols consta d'impulsos (no necessàriament equiespaiats), aquesta convolució òptica passa a ser la convolució habitual. Això s'aplica tant a la generació dels senyals A , B com a la generació del senyal R : anomenem f , g els senyals de disc i de màscara, respectivament, per generar A , B , i k , h els senyals de disc i de màscara per generar R . Els senyals f , k constaran d'impulsos rectangulars.

2. LA GENERACIÓ DEL SENYAL R

Recordem que el senyal de referència R té per objectiu bàsic fer de *compta-revolucions*. A cada volta s'ha de generar, òpticament, un senyal que faci aquesta funció: per exemple, que tingui un pic i només un en cada volta, en una de les N posicions detectades pel codificador. Evidentment, el procediment més senzill seria practicar una obertura rectangular al disc i una altra de la mateixa mida a la màscara, de manera que k , h constarien d'un únic impuls. Quan la resolució del codificador comencés a ser alta, però, l'obertura hauria de ser tan estreta que la intensitat de llum que hi passés seria indetectable per al fotodíode. Per això, cal idear un nou mètode per crear un impuls tan alt i concentrat com sigui possible quan el disc dona una volta sencera.

Aquest era el problema plantejat inicialment per Hohner. Quan l'empresa rebia un encàrrec per fabricar codificadors d'una certa resolució, més alta que en encàrrecs precedents, no sabia com havia de procedir. Havien de demanar-ho a un consultor suís, que els facturava encàrrec per encàrrec, sense arribar a tenir la metodologia per generar-lo autònomament. Això és el que buscaven: obtenir el *know-how* per generar el senyal R .

Anomenem k la funció d'impulsos al disc que, juntament amb les obertures h de la màscara, ha de generar R ; serà de la forma

$$k(x) = \sum_i I(x - n_i\tau),$$

amb uns certs enters $n_i \leq M$. És, per tant, un tren d'impulsos rectangulars, però no necessàriament disposats de forma periòdica. A la figura 4 es mostra un exemple de com serien les obertures determinades per $k(x)$.

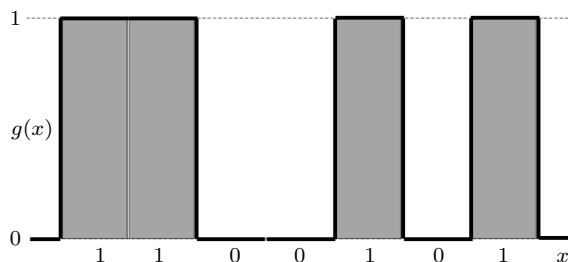


FIGURA 4. Funció $k(x)$ associada al senyal de referència R .

FONT: Elaboració pròpia.

La màscara portarà perforades unes obertures determinades pel subgraf d'una funció positiva h , que ha de ser una funció amb suport compacte, perquè ha de cabre dins la màscara fixa, que no gira. Si mirem els codificadors reals, veurem que tots els fabricants utilitzen $k = h$, però en principi k , h poden ser diferents, i de moment mantenim aquesta possibilitat.

En cada instant t les obertures del disc s'escriuen com a $k(x-t)$. En la situació que ens interessa podem evidentment ignorar la difracció, de manera que el senyal que genera el fotodíode encarregat de R és la convolució

$$k * h(t) = \int_{-\infty}^{+\infty} k(x-t)h(x) dx.$$

Matemàticament, el problema consisteix a triar k , h de forma que $k * h$ sigui el més alta i el més localitzada possible al voltant de $t = 0$. En el cas que triem $k = h$, es tracta de l'*autocorrelació* $k * k$. De moment, suposarem que estem en aquest cas; després, ja tornarem a la situació general. Tenim un senyal k per determinar; les mides físiques de la màscara imposen un límit $M = M(N)$ per al nombre màxim d'obertures de k .

És evident que el màxim de la funció d'autocorrelació serà quan ambdues funcions coincideixin, és a dir, quan no hi hagi desplaçament entre elles, que vol dir $t = 0$. El que voldríem és que $k * k$ fos el més petita possible per als altres valors angulars observats pel codificador, $t = i\frac{2\pi}{N}$, diferents de zero.

Un model equivalent i simplificat consisteix a substituir k per un codi binari $A = (a_n)_{n=1}^M$ de zeros i uns, posant $a_n = 1$ si $n = n_i$, i $a_n = 0$ si $n \neq n_i$; és a dir, $k(x) = \sum_{n=1}^M a_n I(x - n\tau)$. Aleshores l'objecte d'estudi és l'*autocorrelació discreta* d'aquest codi binari:

$$A * A(n) = S(n) = \sum_{i=n+1}^M a_{i-n}a_i, \quad n = 0, \dots, M-1.$$

Si redefinim a_n posant $a_n = 0$ si $n \notin [1, M]$, podem considerar

$$S(n) = \sum_i a_{i-n}a_i = \sum_i a_i a_{i+n}$$

com a funció parella en n . Observem que no es tracta de l'anomenada *autocorrelació circular*, en què considerariem A com a senyal M -periòdic. La pregunta és, ara, saber com cal triar A per tal que el valor màxim $S(0)$ (el nombre d'uns que conté, anomenat *màxim d'intensitat central*) estigui molt per sobre dels altres valors $S(n)$, $n \neq 0$. Per quantificar i formular el problema de forma precisa, introduïm

$$\sigma(A) = \max_{n \neq 0} S(n),$$

que és òbviament menor que $S(0)$, i introduïm el paràmetre de concentració

$$K(A) = \frac{S(0)}{\sigma(A)}.$$

Com més gran sigui aquest paràmetre, millor serà la relació entre $S(0)$ i $\sigma(A)$. Per tant, volem triar A de forma que $K(A)$ sigui el més gran possible entre tots els codis binaris A de longitud M (i que tinguin més d'un 1, amb $\sigma \geq 1$). Podem considerar

$$K(M) = \max(K(A)),$$

on A té longitud M i $\sigma(A) > 0$.

Tenim aleshores plantejats dos problemes: un de tipus matemàtic, teòric, que consisteix a avaluar com creix $K(M)$ asimptòticament, i un problema pràctic, que consisteix a trobar explícitament un codi A pel qual $K(A)$ sigui l'òptim o proper a l'òptim. Per a valors de M petits, hom pot examinar un per un els 2^M codis de longitud M , calcular $K(M)$ i constatar que no hi ha un únic codi òptim. Per a valors de M grans, això evidentment no és possible.

Com sovint passa en matemàtiques, la solució a aquests problemes la trobem en veure que són equivalents a d'altres de ben coneguts; això ens durà als *regles de Golomb* (*Golomb rulers*) i als *conjunts de Sidon* (*Sidon sets*). La reformulació comença per considerar el *conjunt de posicions* on hi ha un 1:

$$G = \{i = 1, 2, \dots, M : a_i = 1\}.$$

Llavors $S(n)$ és el cardinal (el nombre d'elements) del conjunt $G \cap (G - n)$, amb la qual cosa veiem que el problema té a veure

amb els conjunts de diferències (*difference sets*) ([7]), un camp molt estudiat en teoria additiva de nombres.

Com que ens interessa que $K(A)$ sigui el més gran possible, és natural fixar-se en els senyals pels quals el denominador de $K(A)$ és el més petit possible, $\sigma(A) = 1$, és a dir, $S(n) = 0, 1$ si $n \neq 0$. Això significa que per a cada n hi ha com a molt un i tal que $a_{i+n} = a_i = 1$, és a dir, $i+n, i \in G$. Dit d'una altra manera, totes les diferències $i-j$, $i, j \in G$, $i \neq j$ són diferents.

Aquesta és precisament la definició d'un regle de Golomb de longitud M , un subconjunt

$$G = \{i_1 = 1 < i_2 < \dots < i_K = M\},$$

de $\{1, 2, \dots, M\}$ pel qual totes les diferències entre els seus elements són nombres diferents. M s'anomena la *llargada* i K la *mida* de G . A Internet es poden trobar molts articles i taules referents als regles de Golomb ([4]). Aquests conjunts es consideren en radioastronomia (les distàncies entre radioestacions col·locades en una línia recta convé que segueixin un *ruler* per minimitzar certs paràmetres) i també en teoria de codis autocorrectors. El nom de *ruler* ('regle', en català) s'utilitza perquè si G ho és, posant en un llistó les marques corresponents als i_n , es poden mesurar llargades amb aquest regle d'una sola manera (en aquest sentit, els regles habituals són redundants, perquè una determinada longitud la podem mesurar utilitzant marques diferents). Un regle s'anomena *complet* si el conjunt de diferències és exactament $\{1, 2, \dots, M\}$: llavors es poden mesurar totes les longituds entre 1 i M d'una única manera. Una interpretació diferent i curiosa és pensar en una gran sala rectangular de conferències que hom voldria que fos convertible, a base de col·locar unes parets mòbils, en sales de capacitats més petites però diferents. Si les parets mòbils es col·loquen exactament en els punts indicats per un regle de Golomb, utilitzem el nombre mínim de panells i optimitzem els costos.

El problema que s'ha de resoldre és: donada una llargada M , quin és el regle de Golomb més gran que es pot obtenir, amb la K més gran? Podem considerar

$$F(M) = \max K,$$

on el màxim és entre tots els regles de Golomb de longitud M , i el problema és com abans: determinar com creix $F(M)$ i explicitar regles de Golomb òptims. És clar que podem plantejar el problema al revés: donada una mida K , quin és el regle de Golomb més curt que es pot construir amb aquesta mida?

Com que hem arribat als regles de Golomb particularitzant el cas $\sigma(A) = 1$, tenim òbviamment que

$$K(M) \geq F(M).$$

Una reformulació equivalent dels regles de Golomb és en termes de *conjunts de Sidon*. Un *conjunt de Sidon* és un conjunt G de nombres enters tals que totes les sumes $i + j$, $i \neq j$, $i, j \in G$ són diferents. Evidentment és un concepte equivalent al del regle de Golomb, perquè $i - j = i' - j'$ equival a $i + j' = j + i'$. Curiosament, però, les comunitats científiques que s'han ocupat del problema són diferents i bastant disjunctes. Sota aquesta formulació, aquests conjunts, en general infinits, van ser introduïts per Simon Sidon fa un segle, en relació amb les sèries de Fourier; les sèries de Fourier amb freqüències en un conjunt de Sidon tenen unes propietats especials: concretament, totes les normes hi són equivalents. Són objectes molt estudiats en anàlisi harmònica, matemàtica discreta i teoria de nombres. De fet, aquest era un dels temes favorits del famós (i excèntric) matemàtic hongarès Paul Erdős. L'estimació de $F(M)$ té moltes variants: per exemple, es pot relaxar la condició i establir que les sumes, en lloc de ser diferents, puguin repetir-se com a molt un nombre D prefixat de vegades. La determinació del creixement asimptòtic de $F(M)$ per a totes aquestes variants ha estat un problema obert en matemàtiques durant molts i molts anys. El matemàtic espanyol Javier Cilleruelo, traspasat fa tres anys, era un dels grans especialistes en el tema, i a ell i els seus col·laboradors es deu l'article que va tancar definitivament el problema ([2]).

És molt fàcil fer una primera estimació de $F(M)$, la qual resulta la bona. El nombre de diferències $i - j$, $i, j \in G$ no nul·les és exactament

$$\frac{K(K-1)}{2},$$

i com que totes han de ser diferents i són nombres entre 1 i M , tindrem

$$\frac{K(K-1)}{2} \leq M,$$

d'on s'obté

$$F(M) \leq \sqrt{2M} + 1.$$

Erdős i Turán ([5]) varen demostrar que

$$F(M) \leq \sqrt{M} + o(\sqrt{M}),$$

mentre que construccions explícites de Singer ([8]) varen provar que

$$F(M) \geq \sqrt{M} - o(\sqrt{M}),$$

de manera que

$$\lim_{M \rightarrow +\infty} \frac{F(M)}{\sqrt{M}} = 1.$$

És a dir, els regles de Golomb òptims tenen aproximadament \sqrt{M} uns. Els matemàtics, davant d'un problema com aquest, usualment volen anar més lluny, i es pregunten llavors de quin ordre és la diferència $F(M) - \sqrt{M}$, però la resposta no tindria rellevància per a la nostra situació. Ja s'ha comentat abans que

$$K(M) \geq F(M);$$

per tant, en principi, podria ser que $K(M)$ fos asimptòticament més gran que \sqrt{M} . En una comunicació privada, però, Noga Alon, un altre dels grans especialistes en el tema, ens ha fet arribar una prova del fet que $K(M)$ continua sent com \sqrt{M} .

La conclusió rellevant per a nosaltres és que el senyal A que generi el senyal R més potent possible dins de la mida M donada per les mides del fotodíode i la resolució és un tren d'impulsos rectangulars

$$k(x) = \sum_{n=1}^M a_n 1(x - n\tau),$$

amb $a_n = 1$ si $n \in G$, i $a_n = 0$ si $n \notin G$, on G és un regle de Golomb òptim de longitud M i mida aproximadament \sqrt{M} .

El problema matemàtic està, doncs, perfectament ben entès; tan sols hem necessitat un canvi de perspectiva. Una altra cosa és el problema pràctic. Analitzant els articles de Singer i Ruzsa, hom pot trobar un mètode, basat en l'aritmètica de cossos finits, per construir conjunts de Sidon quan M és una potència d'un primer. No hi ha mètodes generals, però. Hi ha en marxa projectes web cooperatius per trobar regles de Golomb, i s'ha arribat a trobar el regle de Golomb òptim per a $K = 27$ marques, el qual té una longitud de $M = 553$ (vegeu [4]). El que nosaltres vàrem fer per a l'empresa és un programa basat en algorismes genètics que troba per a cada valor de M un codi potser no òptim però proper a l'òptim. La informació qualitativa, teòrica, que els codis òptims tenen aproximadament \sqrt{M} uns és d'utilitat en el disseny d'aquest programa.

Abans s'ha comentat que no necessàriament cal que $k = h$; podrien ser diferents les obertures al disc i a la màscara per generar R . Trasl·ladant també el problema a codis discrets, això ens porta al problema de trobar dos codis binaris A, B de mida M tal que la seva correlació

$$A * B(n) = S(n) = \sum_{i=n+1}^M a_{i-n} b_i, \quad n = 0, \dots, M - 1,$$

sigui el més alta i el més concentrada possible al voltant del zero. Això representa un nou grau de llibertat i podria fer més fàcil el problema. Actualment no en conec la resposta.

3. UN PROJECTE DE DOCTORAT INDUSTRIAL

El programa per trobar regles de Golomb gairebé òptims va funcionar i funciona perfectament; aquest èxit inicial va animar l'empresa a plantejar altres problemes relacionats amb la fabricació de codificadors òptics incrementals. Em van semblar suficientment interessants, i que requerien prou dedicació per plantejar a l'empresa la possibilitat d'un doctorat industrial, dins el pla que justament llavors iniciava el Govern de la Generalitat. Així vàrem fer-ho: el projecte va ser aprovat en la primera convocatòria dels doctorats industrials, però, ai las!, no vàrem tenir cap candidat, malgrat la difusió que se'n va fer entre els graduats

en física i matemàtiques. Ens hi vàrem tornar a presentar el segon any i, aquest cop sí, trobarem un bon candidat. Part del que explicarem en els apartats següents es troba més desenvolupat en la tesi [3].

Els problemes plantejats per l'empresa en aquesta segona fase eren dos.

Recordem que el tren de N impulsos rectangulars equiespaiats del disc giratori s'escriu

$$f(x) = \sum_{n=1}^N 1(x - n\tau), \quad \tau = \frac{2\pi}{N},$$

on I és la funció característica de $[0, \frac{\tau}{2}]$.

Les obertures de la màscara són el subgraf d'una funció positiva g amb suport compacte. Recordem també que si s'ignora la difracció i si $g \leq 1$, el senyal captat pel fotodíode és proporcional a

$$f_p * g(t) = \int_{-\infty}^{+\infty} f_p(x - t)g(x) dx = \int_{-\infty}^{+\infty} f_p(x)g(x + t) dx,$$

on f_p és la τ -perioditzada de f , $\tau = \frac{2\pi}{N}$, perquè el disc va girant. Per conveniència del càlcul, farem una translació en t i suposarem que f_p és la funció τ -periòdica que en l'interval $[-\frac{\tau}{2}, +\frac{\tau}{2}]$ coincideix amb la funció característica de $[-\frac{\tau}{4}, +\frac{\tau}{4}]$; com que f_p és parella, l'expressió anterior és la convolució de f_p amb g .

Si prenguéssim com a g també un tren d'impulsos rectangulars, és a dir, que a la màscara es practiquessin les mateixes obertures que al disc, aleshores el senyal obtingut seria un tren d'impulsos triangulars, com a la figura 5, perquè si dos rectangles iguals es desplacen l'un sobre l'altre a una velocitat constant, l'àrea comuna que determinen varia linealment.



FIGURA 5. Tren de polsos triangulars.

FONT: Elaboració pròpia.

Tal com s'ha explicat en el primer apartat, per poder multiplicar la resolució electrònicament d'una forma barata, l'ideal fora que el senyal obtingut fos una sinusoide pura en lloc d'un tren d'impulsos triangulars. Per tant, la primera pregunta òbvia feta per l'empresa fou: *mantenint els trens d'impulsos rectangulars als discs, com han de ser les obertures a la màscara per tal que el senyal de sortida sigui exactament una sinusoide?*

Passem ara a formular la segona pregunta. Els codificadors treballen sovint sota unes condicions extremes que pertorben molts dels paràmetres de fabricació. Per exemple, la vibració de l'eix rotor pot fer que la distància entre el disc i la màscara s'alteri. L'experiència acumulada per l'empresa els ha ensenyat que la distància (*gap*) z entre el disc i la màscara és crítica, en el sentit que la qualitat del senyal òptic de sortida es deteriora molt quan canvia aquesta distància. L'empresa es troba que bastants clients retornen codificadors per tal que siguin reajustats. La segona pregunta fou: *és possible dissenyar una màscara que, a més de produir un senyal sinusoidal, sigui robusta envers els canvis en la distància entre el disc i la màscara?*

És intuïtivament evident que el tractament d'aquest segon problema implica entendre bé els fenòmens de difracció de la llum que tenen lloc al codificador.

4. LA MÀSCARA IDEAL

L'objectiu d'aquest apartat és buscar una funció g apropiada, positiva amb suport compacte, que convolucionada amb f_p doni un senyal sinusoidal/cosinusoidal pur, positiu, amb el mateix període bàsic τ . La buscarem també parella, i plantegem l'equació

$$(f_p * g)(t) = \int_{-\infty}^{+\infty} f_p(x-t)g(x) dx = A \left(1 + \cos \frac{2\pi}{\tau} t \right).$$

Per resoldre aquest problema analíticament, utilitzarem la transformada de Fourier, ja que és l'eina adequada per manipular les operacions lineals invariants per translacions, com és el cas de les convolucions.

La convolució de f_p amb g és evidentment τ -periòdica,

$$\begin{aligned}(f_p * g)(t + \tau) &= \int_{-\infty}^{+\infty} f_p(x - t - \tau)g(x) dx \\ &= \int_{-\infty}^{+\infty} f_p(x - t)g(x) dx = (f_p * g)(t),\end{aligned}$$

i, per tant, té un desenvolupament en sèrie de Fourier

$$(f_p * g)(t) = \sum_n c_n(f_p * g)e^{2\pi i \frac{n}{\tau} t}.$$

Els coeficients $c_n(f_p * g)$ són

$$\begin{aligned}c_n(f_p * g) &= \frac{1}{\tau} \int_0^\tau (f_p * g)(t)e^{-2\pi i \frac{n}{\tau} t} dt \\ &= \frac{1}{\tau} \int_0^\tau \left\{ \int_{-\infty}^{+\infty} f_p(x - t)g(x) dx \right\} e^{-2\pi i \frac{n}{\tau} t} dt,\end{aligned}$$

i aplicant Fubini,

$$\begin{aligned}c_n(f_p * g) &= \frac{1}{\tau} \int_{-\infty}^{+\infty} \left\{ \int_0^\tau f_p(x - t)e^{-2\pi i \frac{n}{\tau}(t-x)} dt \right\} g(x)e^{-2\pi i \frac{n}{\tau} x} dx \\ &= c_n(f_p) \int_{-\infty}^{+\infty} g(x)e^{-2\pi i \frac{n}{\tau} x} dx = c_n(f_p)\widehat{g}\left(\frac{n}{\tau}\right).\end{aligned}$$

En definitiva, els coeficients de la sèrie de Fourier de la convolució són el producte dels dos coeficients de Fourier, el de f_p com a funció periòdica i el de g com a funció amb suport compacte:

$$c_n(f_p * g) = c_n(f_p)\widehat{g}\left(\frac{n}{\tau}\right).$$

Calculem $c_n(f_p)$ utilitzant un període bàsic, un pols rectangular

$$\text{per a } n \neq 0 \quad c_n(f_p) = \frac{1}{\tau} \int_{-\frac{\tau}{4}}^{+\frac{\tau}{4}} e^{-2\pi i \frac{n}{\tau} t} dt \quad (1)$$

$$= \frac{1}{\pi n} \sin\left(\frac{\pi n}{2}\right) = \begin{cases} 0, & \text{si } n = 2k, \\ \frac{(-1)^k}{\pi n}, & \text{si } n = 2k + 1, \end{cases}$$

$$\text{per a } n = 0 \quad c_0 = \frac{1}{\tau} \tau = 1. \quad (2)$$

L'objectiu és obtenir una solució sinusoidal positiva

$$(f_p * g)(t) = 1 + \cos\left(2\pi \frac{t}{\tau}\right) = 1 + \frac{1}{2} (e^{2\pi i \frac{t}{\tau}} + e^{-2\pi i \frac{t}{\tau}})$$

amb una tria de g convenient, positiva. Comparant ambdós desenvolupaments, veiem que g ha de complir

$$c_n(f_p) \widehat{g}\left(\frac{n}{\tau}\right) = \begin{cases} 1, & \text{si } n = 0, \\ \frac{1}{2}, & \text{si } n = \pm 1, \\ 0, & \text{si } n \neq 0, \pm 1. \end{cases}$$

Com que els coeficients $c_n(f_p)$ són zero per als n parells diferents de zero, això implica que la transformada de Fourier de g ha de complir

$$\widehat{g}\left(\frac{n}{\tau}\right) = \begin{cases} \widehat{g}(0) = 2, \\ \widehat{g}\left(\pm \frac{1}{\tau}\right) = \frac{\pi}{2}, \\ \widehat{g}\left(\frac{n}{\tau}\right) = 0 \text{ per a } n \text{ senar, } n \neq \pm 1, \\ \text{cap condició per a } n \text{ parell, } n \neq 0. \end{cases} \quad (3)$$

Considerem ara la funció més senzilla que s'anul·la en els punts $\frac{n}{\tau}$ per a n senar, que és $\cos \frac{\pi}{2} \tau \xi$. Aquesta funció té una factorització com a producte infinit

$$\cos \frac{\pi}{2} \tau \xi = \prod_{k=1}^{\infty} \left(1 - \frac{\tau^2 \xi^2}{(2k+1)^2}\right) = \prod_{n \text{ senar}} \left(1 - \frac{\tau \xi}{n}\right).$$

Com que no es vol que \widehat{g} s'anul·li en els punts $\pm\frac{1}{\tau}$, factoritzem aquests zeros i ho provem amb

$$\widehat{g}(\xi) = 2 \left(\cos \frac{\pi}{2} \tau \xi \right) \frac{1}{(1 - \tau^2 \xi^2)} \quad (4)$$

(aquesta expressió està motivada pel fet que la transformada de Fourier d'una funció amb suport compacte és una funció entera amb aquesta expressió general). Amb això, $\widehat{g}(0) = 2$, mentre que en els punts $\pm\frac{1}{\tau}$ es té

$$\widehat{g}\left(\pm\frac{1}{\tau}\right) = 2 \lim_{\xi \rightarrow \pm\frac{1}{\tau}} \left\{ \left(\cos \frac{\pi}{2} \tau \xi \right) \frac{1}{(1 - \tau^2 \xi^2)} \right\}.$$

Calculant el límit amb la regla de L'Hôpital, resulta

$$\widehat{g}\left(\pm\frac{1}{\tau}\right) = \frac{\pi}{2},$$

com volíem. Així, la funció (4) compleix totes les condicions requerides per als coeficients, i en certa forma és la tria més senzilla. Ens falta, però, quelcom: g ha de ser positiva. El següent pas és calcular $g(x)$ fent la transformada de Fourier inversa de (4), és a dir,

$$g(x) = 2 \int_{-\infty}^{+\infty} e^{2\pi i x \xi} \cos\left(\frac{\pi}{2} \tau \xi\right) \frac{d\xi}{(1 - \tau^2 \xi^2)}.$$

Definim $k(x)$ com

$$k(x) = \int_{-\infty}^{+\infty} e^{2\pi i x \xi} \frac{d\xi}{1 - \frac{\xi^2}{v^2}} = \frac{v}{2} \left[\int_{-\infty}^{+\infty} e^{2\pi i x \xi} \frac{d\xi}{v - \xi} + \int_{-\infty}^{+\infty} e^{2\pi i x \xi} \frac{d\xi}{v + \xi} \right],$$

amb $v = \frac{\Delta}{\tau}$, de manera que

$$g(x) = k\left(x + \frac{\tau}{4}\right) + k\left(x - \frac{\tau}{4}\right).$$

Per calcular k , partim de

$$\int \frac{e^{2\pi i x \xi}}{\xi} d\xi = H(x) = \begin{cases} \pi i, & \text{si } x > 0, \\ -\pi i, & \text{si } x < 0. \end{cases}$$

La funció H de la dreta té derivada en el sentit de les distribucions igual a $2\pi i \delta_0$ (on δ_0 indica la delta de Dirac en el zero); per tant, té transformada \widehat{H} que compleix $\xi \widehat{H} = 1$; la funció $\frac{1}{\xi}$ no és, però, localment integrable i la integral s'ha d'interpretar en el sentit del valor principal. Aleshores,

$$\int_{-\infty}^{+\infty} \frac{e^{2\pi i x \xi}}{v - \xi} d\xi = \begin{cases} -\pi i e^{2\pi i x v}, & \text{si } x > 0, \\ \pi i e^{2\pi i x v}, & \text{si } x < 0, \end{cases}$$

$$\int_{-\infty}^{+\infty} \frac{e^{2\pi i x \xi}}{v + \xi} d\xi = \begin{cases} \pi i e^{-2\pi i x v}, & \text{si } x > 0, \\ -\pi i e^{-2\pi i x v}, & \text{si } x < 0. \end{cases}$$

Així, s'obté

$$k(x) = \begin{cases} \frac{v}{2}(-\pi i e^{2\pi i x v} + \pi i e^{-2\pi i x v}) = v\pi \sin(2\pi x v), & \text{si } x > 0, \\ \frac{v}{2}(\pi i e^{2\pi i x v} - \pi i e^{-2\pi i x v}) = -v\pi \sin(2\pi x v), & \text{si } x < 0. \end{cases}$$

Substituint-ho a $g(x) = k\left(x + \frac{\tau}{4}\right) + k\left(x - \frac{\tau}{4}\right)$, com que

$$\sin\left(\frac{2\pi}{\tau}\left(x \pm \frac{\tau}{4}\right)\right) = \sin\left(\frac{2\pi}{\tau}x \pm \frac{\pi}{2}\right) = \pm \cos \frac{2\pi}{\tau}x,$$

resulta que $g(x) = 0$ si $x + \frac{\tau}{4} < 0$ o $x - \frac{\tau}{4} > 0$, i a l'interval $\left[-\frac{\tau}{4}, +\frac{\tau}{4}\right]$ val

$$g(x) = \frac{2\pi}{\tau} \cos\left(\frac{2\pi}{\tau}x\right). \quad (5)$$

En definitiva, s'obté com a g , llevat de les constants, la funció que consisteix a prendre la part positiva de $\cos\left(2\pi\frac{x}{\tau}\right)$. Òbviament, es poden afegir a g tantes τ -traslladades seves com càpiguen a la finestra de la màscara, tal com es mostra a la figura 6. Anomenem aquesta funció *màscara ideal* o *de primera generació*.

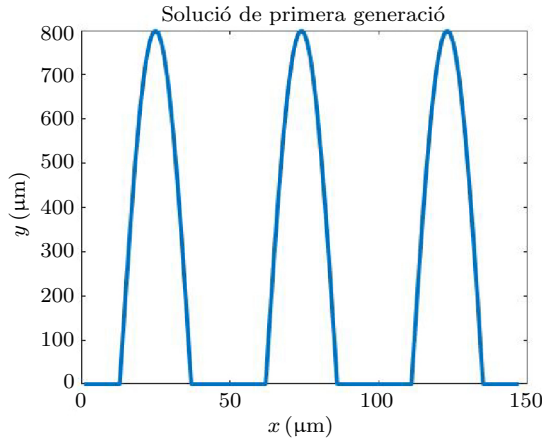


FIGURA 6. Màscara ideal.

FONT: Elaboració pròpia.

Per reafirmar que aquesta màscara és bona, calculem ara directament la convolució de g i f_p . Com que f_p és la τ -perioditzada de la funció f que en l'interval $[-\frac{\tau}{2}, +\frac{\tau}{2}]$ coincideix amb la funció característica de $[-\frac{\tau}{4}, +\frac{\tau}{4}]$, el resultat és la τ -perioditzada de

$$\int_{-\frac{\tau}{4}}^{+\frac{\tau}{4}} \left(\cos \frac{2\pi}{\tau} x \right) f(x-t) dx = \int_{\max(-\frac{\tau}{4}, t-\frac{\tau}{4})}^{\min(\frac{\tau}{4}, t+\frac{\tau}{4})} \cos \frac{2\pi}{\tau} x dx.$$

Aquesta funció és zero si $|t| \geq \frac{\tau}{2}$, mentre que si $0 \leq t \leq \frac{\tau}{2}$ val

$$\int_{t-\frac{\tau}{4}}^{\frac{\tau}{4}} \cos \frac{2\pi}{\tau} x dx = \frac{\tau}{2\pi} \left(1 + \cos \frac{2\pi}{\tau} t \right),$$

i anàlogament si $-\frac{\tau}{2} \leq t \leq 0$. En definitiva, s'obté un múltiple d' $1 + \cos \frac{2\pi}{\tau} t$, i es verifica així que la solució és perfectament vàlida.

Amb això hem trobat una resposta a la primera pregunta. Però als matemàtics ens agrada aprofundir, encara que a vegades no sigui útil, i sorgeix la següent pregunta: *hi ha altres funcions g que serveixin?*

Encara que resulti una mica sorprenent, la resposta és que no; només hi ha aquesta solució, essencialment. Aquest «essencialment» ve de l'observació següent, que podem formular de dues maneres diferents. A part de ser positiva, la propietat desitjada per a g és equivalent a (3), que són condicions per als coeficients de Fourier $\hat{g}(\frac{n}{\tau})$.

Aquests coeficients no determinen g ; són, de fet, els coeficients de Fourier de la τ -perioditzada de g

$$\sum_n g(x - n\tau).$$

Evidentment, podem tenir dues funcions diferents amb la mateixa perioditzada. Per exemple, si repartim la massa de g en el seu suport en dues parts i cada part la traslладem un múltiple enter de τ , obtindrem una altra funció amb la mateixa perioditzada. Veiem, doncs, que les condicions (3) determinen g si, a més, imposem que g tingui un suport de longitud τ com a màxim.

Això significa que el que té sentit és preguntar-se per la unicitat de g entre totes les funcions positives amb suport en un període, diguem en $[-\frac{\tau}{2}, +\frac{\tau}{2}]$. Hem vist que $g(x) = \frac{2\pi}{\tau} \cos(\frac{2\pi}{\tau}x)$, $|x| \leq \frac{\tau}{4}$ compleix l'equació de convolució

$$(f_p * g)(t) = \int_{-\frac{\tau}{2}, \frac{\tau}{2}} f(x - t)g(x) dx = 1 + \cos \frac{2\pi}{\tau}t, \quad t \in \mathbb{R}.$$

Es pot demostrar que la funció g de (5) és l'única funció positiva amb suport en $[-\frac{\tau}{2}, +\frac{\tau}{2}]$ que compleix aquesta equació: és l'única màscara ideal. A tall d'anècdota, aquest «teorema d'unicitat» ens va servir per dir a la sollicitud que cap altra patent podia competir amb la nostra.

5. LA MÀSCARA ROBUSTA

Quan l'empresa va provar experimentalment la màscara ideal, els resultats foren satisfactoris, però no del tot. La qualitat del senyal sinusoidal captat pels fotodíodes depenia molt de la distància z entre el disc i la màscara. Per a certes distàncies els resultats eren espectacularment bons, però si es modificava lleugerament, el senyal es deteriorava, és a dir, la màscara ideal no era robusta. El següent problema era, per tant, explicar aquest fet i trobar una màscara més robusta. La criticitat de z és intuïtivament explicable: com més reduïdes són les obertures, més s'ha de reduir la separació entre el disc i la màscara perquè el fotodíode registri les variacions del senyal i, per tant, pugui generar senyals que es converteixin en polsos. Quan l'ample de les

obertures és de l'ordre de desenes de micres, z ha de ser del mateix ordre, i, en conseqüència, el procés de fabricació és difícil i el producte final té un baix nivell de robustesa. Per intuïció, hom pot veure que hi ha dos factors principals que contribueixen a aquest problema: la qualitat de col·limació del sistema òptic, i els efectes de difracció i interferència que pateix la llum després de travessar les obertures del disc. El primer problema és fàcilment tractable, canviant si convé a llum làser en la font emissora. El problema important és la difracció, que fins ara hem ignorat. Per fer el nostre petit model basat en la convolució òptica, hem suposat que la llum es propaga com una ona plana, abans i després de travessar les obertures del disc. Evidentment la realitat no és així, perquè en passar per les obertures la llum es difracta i el camp no és el mateix a un costat i l'altre. Per a la generació del senyal R , és raonable ignorar la difracció, perquè al capdavall es busca una intensitat màxima, un pic, i un pic distorsionat continuarà sent similar a un pic. Però per a la generació del senyal sinusoidal —una sortida molt concreta i delicada—, és clar que cal tenir-la en compte. Ja s'ha dit que els xips interpoladors basen la seva eficàcia en el fet que els senyals entrants siguin el més sinusoidals possible.

El tractament matemàtic clàssic de la difracció, a partir de les equacions de Maxwell del camp electromagnètic, tal com pot trobar-se, per exemple, a la referència [6], és delicat, i molt més si es vol fer rigorosament. El que és clar és que cal revisar el nostre model tenint en compte la difracció. Considerarem que tenim una ona monocromàtica plana i molt ben col·limada enfocant de forma totalment normal les obertures del disc, amb longitud d'ona λ . D'aquesta manera, el camp electromagnètic d'origen (E_0, B_0) té els sis components $U_0(x, y)$ proporcionals a la funció

$$U_0(x, y) = I\left(\frac{x}{c}\right) I(y) f_p(x),$$

on I és la funció característica de l'interval $[-1, +1]$, i f_p és la funció de període τ que correspon a les obertures del disc, que a $[-\frac{\tau}{2}, +\frac{\tau}{2}]$ val 1 si $|x| \leq \frac{\tau}{4}$ i 0 a fora. D'altra banda, $2c$ acota quina és la zona il·luminada, és a dir, quantes obertures contribueixen als efectes difractius i d'interferència. Això és important per entendre que no estem tractant amb una zona infinitament il·luminada, sinó que la contribució

és d'una petita part, marcada pel diàmetre del feix incident. El producte $I\left(\frac{x}{c}\right)I(y)$ modelitza la regió il·luminada per la llum incident. En l'anàlisi següent podem considerar una funció general $f(x)$ en lloc de $f_p(x)$.

L'operació $U_0 \rightarrow U_z$, on U_z designa qualsevol dels components del camp en un pla paral·lel a una distància z del patró de difracció, és invariant per translacions, i consisteix, per tant, en una convolució: $U_z = U_0 * K_z$. A partir de l'equació de Helmholtz, hom pot veure que el nucli de convolució és

$$K_z(x, y) = \frac{1}{2\pi} \frac{e^{ikR}}{R} \left(\frac{1}{R} - ik \right) \frac{z}{R},$$

amb $R = \sqrt{x^2 + y^2 + z^2}$ i $k = \frac{2\pi}{\lambda}$ com a nombre d'ona. És habitual en òptica fer unes aproximacions que simplifiquen aquestes expressions. En primer lloc, per a valors de z (i, per tant, de R) molt més grans que λ , com és el cas que ens ocupa, s'ignora el terme $\frac{1}{R}$ del parèntesi i es considera l'aproximació

$$K_z(x, y) = -\frac{ik}{2\pi} \frac{ze^{ikR}}{R^2} = \frac{z}{i\lambda} \frac{e^{ikR}}{R^2}. \quad (6)$$

Amb això tenim l'expressió del camp difractat quan $z \gg \lambda$,

$$U_z(x, y) = (U_0 * K_z)(x, y) = \frac{z}{i\lambda} \int_{\mathbb{R}^2} U_0(x - x', y - y') \frac{e^{ikR'}}{R'^2} dx' dy',$$

on ara $R'^2 = x'^2 + y'^2 + z^2$. En polars,

$$\frac{z}{i\lambda} \int_0^{+\infty} \frac{e^{ik\sqrt{r^2+z^2}}}{r^2+z^2} \left(\int_0^{2\pi} U_0(x - r \cos \theta, y - r \sin \theta) d\theta \right) r dr,$$

i ara fent el canvi $r = tz$,

$$\frac{z}{i\lambda} \int_0^{+\infty} \frac{e^{ikz\sqrt{1+t^2}}}{1+t^2} \left(\int_0^{2\pi} U_0(x - tz \cos \theta, y - tz \sin \theta) d\theta \right) t dt. \quad (7)$$

Aquesta integral és de tipus *oscil·latori*,

$$I(\mu) = \int_a^b e^{i\mu\Phi(t)} \Psi(t) dt.$$

En el nostre cas, $\Phi(t) = \sqrt{1+t^2}$ i $\mu = kz$ és gran. Un principi bàsic ([9]) d'aquestes integrals és el *principi de localització*, segons el qual el comportament asimptòtic quan μ és gran està determinat pels punts on $\Phi'(t) = 0$. Concretament, si $\Phi'(t) \neq 0$ en $[a, b]$, aleshores

$$I(\mu) = O(\mu^{-N}),$$

per a tot N . En canvi, si $\Phi'(t)$ s'anul·la en un punt t_0 , el comportament ve donat per l'ordre k de la primera derivada $\Phi^{(k)}(t_0) \neq 0$. Concretament [9, p. 334],

$$I(\mu) \approx \mu^{-\frac{1}{k}}.$$

El que això significa és que en una integral oscil·latòria, la contribució essencial a la integral és la d'un interval al voltant dels punts t_0 on la fase Φ té derivada zero, i la resta és d'ordre molt més petit. A més a més, podem substituir $\Phi(t)$ pel seu desenvolupament en t_0 fins a l'ordre k .

En el nostre cas, $\Phi(t) = \sqrt{1+t^2}$ té el punt estacionari $t = 0$, i, per tant, la contribució essencial en (7) és la d'un interval $[0, \varepsilon]$. Utilitzant el desenvolupament

$$\sqrt{1+t^2} = 1 + \frac{1}{2}t^2 + \dots,$$

el nucli de (7), $e^{ikz\sqrt{1+t^2}}$, se simplifica per $e^{ikz(1+\frac{1}{2}t^2)}$. Desfent els canvis en (6), el resultat final és l'anomenada *aproximació de Fresnel*, utilitzada comunament en òptica: U_z és la convolució de U_0 amb el *nucli de Fresnel*

$$K_z(x, y) = \frac{e^{ikz}}{i\lambda z} e^{i\frac{k}{2z}[x^2+y^2]},$$

és a dir,

$$U_z(x, y) = \frac{e^{ikz}}{i\lambda z} \iint_{-\infty}^{+\infty} U_0(x', y') \exp\left\{i\frac{k}{2z}[(x-x')^2 + (y-y')^2]\right\} dx' dy', \quad (8)$$

vàlida quan $|x-x'| + |y-y'| \leq \varepsilon z$ i $z \gg \lambda$. Aquesta és una bona aproximació en la regió $z \gg \lambda$, si el quadrat de centre (x, y) i radi εz està inclòs en el suport de U_0 .

De fet, és possible afinar molt més analitzant amb detall el nucli de Fresnel. Suposem per simplificar que $U_0(x', y')$ és separable (que és

el cas que ens interessarà després) en un producte $A(x')B(y')$, cosa que ens porta a una integral en una variable (llevat de les constants)

$$\int_{-\infty}^{+\infty} A(x')e^{i\frac{\pi}{\lambda z}(x-x')^2} dx' = \int_{-\infty}^{+\infty} A(x-x')e^{i\frac{\pi}{\lambda z}(x')^2} dx'.$$

La idea intuïtiva és que $e^{i\frac{\pi}{\lambda z}(x')^2}$ no oscil·la mentre es compleixi $|x'|^2 \leq O\left(\frac{\pi}{\lambda z}\right)$ i, en canvi, oscil·la molt ràpidament en la resta de casos, amb la qual cosa hi ha molta cancel·lació, de manera que la contribució essencial és la de $|x'| \leq O\left(\sqrt{\frac{\pi}{\lambda z}}\right)$.

Tornant a (8), veiem que la part essencial en la integral és la del quadrat

$$|x - x'|, |y - y'| \leq O(\sqrt{\lambda z}),$$

que és molt més petit que el quadrat considerat fins ara

$$|x - x'|, |y - y'| \leq \varepsilon z,$$

perquè $z \gg \lambda$. El significat d'aquesta aproximació és que si el camp $U_0(x, y)$ té suport en una regió G del pla, aleshores l'aproximació de Fresnel és acurada en els punts (x, y, z) tals que el quadrat de centre $(x, y, z = 0)$ i aresta $\sqrt{\lambda z}$ està tot inclòs dins G . Aquesta *regió d'influència* té un perfil de tipus parabòlic ([6, p. 70-72]). Podem considerar com a zona fosca el conjunt de punts (x, y, z) per als quals aquest quadrat és disjunt amb G . La figura 7 il·lustra aquesta situació.

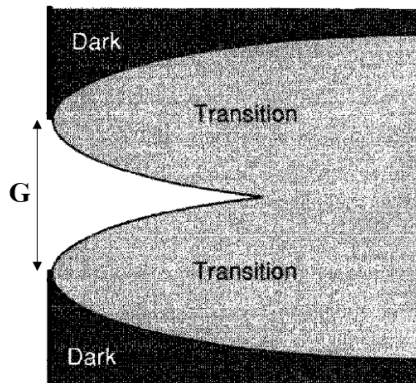


FIGURA 7. Propagació transversal de la llum per difracció.

FONT: J. W. GOODMAN, *Introduction to Fourier optics*. Englewood, CO: Roberts and Company Publishers, 2005.

Una altra forma d'expressar l'aproximació de Fresnel és

$$U_z(x, y) = \frac{e^{ikz}}{i\lambda z} e^{i\frac{k}{2z}(x^2+y^2)} \iint_{-\infty}^{+\infty} \{U_0(x', y') e^{i\frac{k}{2z}(x'^2+y'^2)}\} e^{-i\frac{2\pi}{\lambda z}(xx'+yy')} dx' dy',$$

on a la integral hi reconeixem la transformada de Fourier, avaluada a (x, y) , de la funció

$$U_0(x', y') e^{i\frac{k}{2z}(x'^2+y'^2)}.$$

Apliquem ara l'aproximació de Fresnel a la nostra situació, per a la qual $U_0(x, y) = I\left(\frac{x}{c}\right)I(y)f(x)$. Designem per (E_z, B_z) el camp a distància z del disc i per

$$A_z(x) = e^{i\frac{\pi}{\lambda z}x^2},$$

el nucli de Fresnel unidimensional; cadascun dels components és proporcional a

$$\begin{aligned} U_z(x, y) &= \iint U_0(x', y') K_z(x - x', y - y') dx' dy' \\ &= \frac{e^{ikz}}{i\lambda z} \left(\int_{-1}^1 A_z(y - y') dy' \right) \left(\int_{-c}^c f(x) A_z(x - x') dx' \right). \end{aligned}$$

En aquest punt fem la hipòtesi que z és tal que l'interval de centre x i radi $\sqrt{\lambda z}$ està inclòs a $[-c, c]$. Segons hem vist, això ens permet simplificar la integral en x prescindint del factor $I\left(\frac{x}{c}\right)$ i obtenim

$$U_z(x, y) = \frac{e^{ikz}}{i\lambda z} \phi_z(y) \int_{-\infty}^{+\infty} f(x') A_z(x - x') dx' = \frac{e^{ikz}}{i\lambda z} \phi_z(y) f_z(x) \quad (9)$$

amb

$$\phi_z(y) = \int_{-c}^c A_z(y - y') dy' = \int_{-c}^c e^{i\frac{\pi}{\lambda z}(y-y')^2} dy'$$

i $f_z = f * A_z$. Ara bé, la pertorbació vertical $\phi_z(y)$ és negligible, perquè és essencialment constant. En efecte, la màscara té unes dimensions

verticals més petites que la zona il·luminada, és a dir, $|y| \ll c$. Aleshores, com que la part essencial és, com hem mencionat abans, la de l'interval $[y - O(\sqrt{\lambda z}), y + O(\sqrt{\lambda z})]$, que és tot dins $[-c, c]$, hom veu que $|\phi_z(y)|$ és essencialment constant sobre la màscara.

Suposem que la xarxa unidimensional és periòdica, és a dir, que f és una funció τ -periòdica. Explicarem a continuació com en aquest cas es produeix un efecte ben conegut en òptica, l'*efecte Talbot*. Tindrem un desenvolupament

$$f(x) = \sum_n c_n(f) e^{2\pi i \frac{n}{\tau} x}.$$

Per tant, repetint els càlculs de l'apartat anterior, la convolució $f_z = f * A_z$ també és τ -periòdica amb desenvolupament

$$f_z(x) = \sum_n c_n(f) \widehat{A}_z\left(\frac{n}{\tau}\right) e^{2\pi i \frac{n}{\tau} x}.$$

És ben conegut que la transformada de Fourier d'una gaussiana és també gaussiana. Per prolongació analítica, això també és cert per a paràmetres complexos. De fet,

$$\widehat{A}_z(\xi) = \sqrt{\lambda z} \exp[-i\pi \lambda z \xi^2].$$

Per tant,

$$f_z(x) = \sum_n c_n(f) e^{-i\pi \lambda z \frac{n^2}{\tau^2}} e^{2\pi i \frac{n}{\tau} x},$$

i hom pot acabar reescrivint (9) com a

$$U_z(x, y) = \frac{e^{ikz}}{i\sqrt{\lambda z}} \phi_z(y) \sum_n c_n(f) e^{-i\pi \lambda z \frac{n^2}{\tau^2}} e^{2\pi i \frac{n}{\tau} x}. \quad (10)$$

Si es pren z de forma que $\frac{\lambda z}{\tau^2}$ sigui parell,

$$\frac{\lambda z}{\tau^2} = 2m,$$

és a dir, que z sigui un múltiple enter de

$$z_T = \frac{2\tau^2}{\lambda},$$

anomenat *distància de Talbot* ([6, p. 89-91]), aleshores $e^{-i\pi\lambda z \frac{n^2}{\tau^2}} = 1$ per a tot n ; per tant, $f_z = f$ i

$$U_z = \frac{e^{ikz}}{i\sqrt{\lambda z}} \phi_z(y) f(x).$$

És a dir, per a múltiples enters de la distància de Talbot apareix una distribució de camp essencialment idèntica al camp d'origen (perquè ϕ_z és essencialment constant), amb només una variació de fase.

Si, en canvi,

$$\frac{\lambda z}{\tau^2} = (2m + 1), \quad z = \left(m + \frac{1}{2}\right) z_T,$$

aleshores $e^{-i\pi\lambda z \frac{n^2}{\tau^2}} = e^{-i\pi n^2} = (-1)^{n^2} = (-1)^n$, amb la qual cosa

$$U_z(x, y) = \frac{e^{ikz}}{i\sqrt{\lambda z}} \phi_z(y) \sum_n c_n(f) (-1)^n e^{2\pi i \frac{n}{\tau} x} = \frac{e^{ikz}}{i\sqrt{\lambda z}} \phi_z(y) f\left(x + \frac{\tau}{2}\right).$$

Això significa que hi ha un desfasament de 180 graus. En el cas $f = f_p$, les zones opaques s'intercanvien amb les il·luminades. Finalment, si

$$\frac{\lambda z}{\tau^2} = (2m + 1), \quad z = (2m + 1) \frac{z_T}{4},$$

aleshores $e^{-i\pi\lambda z \frac{n^2}{\tau^2}} = (-1)^{mn^2} (-i)^{n^2}$. Aquesta expressió val 1 si n és parell i $(-1)^{m+1}i$ si n és senar. Això significa que

$$f_z(x) = \sum_{n \text{ parell}} c_n(f) e^{2\pi i \frac{n}{\tau} x} + (-1)^{m+1} i \sum_{n \text{ senar}} c_n(f) e^{2\pi i \frac{n}{\tau} x}.$$

El primer sumatori és una funció $\frac{\tau}{2}$ -periòdica, mentre que el segon sumatori $S(x)$ compleix $S\left(x + \frac{\tau}{2}\right) = -S(x)$. En definitiva, $|f_z|^2$, i, per tant, la potència, és una funció $\frac{\tau}{2}$ -periòdica: té doble freqüència.

A la taula 1 es mostren les diferents distàncies de Talbot calculades per a resolucions de 500, 1 024, 2 500 i 4 096 polsos per volta (en disc de 19 mm de radi), en funció de la font d'il·luminació utilitzada.

TAULA 1. Distàncies de Talbot

Font d'il·luminació	Resolució (polsos per volta)			
	500	1 024	2 500	4 096
Làser vermell ($\lambda = 635$ nm)	179,55 mm	42,81 mm	7,14 mm	2,66 mm
Làser blau ($\lambda = 405$ nm)	281,52 mm	67,12 mm	11,26 mm	4,2 mm
Led ($\lambda = 470$ nm)	242,58 mm	57,84 mm	9,70 mm	3,62 mm

Continuem analitzant el camp difractat per una xarxa periòdica unidimensional $I\left(\frac{y}{c}\right)f(x)$, donat per (9). Com que tots els components del camp (E_z, B_z) són proporcionals a U_z , la intensitat o potència P_z del camp en el punt (x, y, z) és proporcional a $|U_z(x, y)|^2$,

$$P_z(x, y) \propto |\phi_z(y)|^2 |f_z(x)|^2.$$

Considerem ara la màscara, modelitzada per la funció g amb suport compacte. Aquí pensarem per conveniència que està simetritzada, duplicada, i que les obertures de la màscara ocupen, a una distància z , la regió del pla donada per

$$G = \{(x, y), 0 \leq |y| \leq g(x)\}.$$

Considerant també que els fotodíodes són solidaris amb la màscara, la intensitat que capten és

$$P_z = \int_G |\phi_z(y)|^2 |f_z(x)|^2 dx dy = \int_{-\infty}^{+\infty} |f_z(x)|^2 \left(\int_{|y| \leq g(x)} |\phi_z(y)|^2 dy \right) dx.$$

Això ens porta a introduir la funció \tilde{g} , també amb suport compacte, definida per

$$\tilde{g}(x) = \int_{|y| \leq g(x)} |\phi_z(y)|^2 dy, \quad x \in \mathbb{R},$$

de forma que

$$P_z = \int_{\mathbb{R}} |f_z(x)|^2 \tilde{g}(x) dx.$$

El següent pas és considerar com variarà la potència lumínica $P_z(t)$ quan giri el disc. En l'instant t cal canviar $f(x)$ per $f(x - t)$; llavors $f_z(x)$ canvia a $f_z(x - t)$ i obtenim

$$P_z(t) = \int_{\mathbb{R}} |f_z(x - t)|^2 \tilde{g}(x) dx = (|f_z|^2 * \tilde{g})(t).$$

En definitiva, si el disc està donat per f i la màscara per g , el senyal captat pel fotodíode ve donat per l'expressió anterior. Hi destaquen dues característiques: la no-linealitat deguda al terme quadràtic i la substitució de g per \tilde{g} . Ara bé, ja s'ha comentat abans que ϕ_z és essencialment constant en la màscara; per tant, $\tilde{g} \approx g$. Amb aquesta simplificació, en definitiva, llevat de les constants, hom té

$$P_z(t) = (|f_z|^2 * g)(t).$$

Quan z és un múltiple de la distància de Talbot, aleshores $f_z = f$, $|f_z|^2 = f^2$. En el nostre cas, a més, $f = f_p$ i $f_p^2 = f_p$. Això significa que *l'anàlisi feta en l'apartat anterior és vàlida, i la màscara de primera generació donarà una sortida sinusoidal pura*. No la donarà, però, per a altres valors de z . Això explica perfectament el comportament de la màscara ideal observat per l'empresa: la màscara ideal funciona perfectament a distàncies z múltiples de la distància de Talbot.

D'altra banda, es pot veure que la font de llum està molt relacionada amb la distància de Talbot. Quan estem tractant amb longitud d'ona λ curta (com la blava), z_T augmenta. D'aquesta manera, el sistema admet toleràncies més altes a l'hora de col·locar la màscara al voltant d'una determinada z_T . No obstant això, també pot ser positiu utilitzar longitud d'ona λ llarga (com la infraroja), ja que, en una llargada reduïda, es poden abastar diferents múltiples de z_T . Dependent de la periodicitat entre obertures, relacionada amb la resolució del codificador, pot interessar aprofitar una propietat o una altra.

Quan s'analitza la funció difractada en qualsevol distància diferent d'un múltiple de la de Talbot, $z \neq z_T$, no tenim $f_z = f$. El nostre objectiu és trobar g tal que P_z sigui una sinusoide, en qualsevol distància z . Com que aquesta és una equació de convolució, hem

de procedir igual que abans, però hem d'utilitzar $|f_z|^2$ quan $f = f_p$. Partim de (5), on els c_n són els de (1) i (2). Tindrem

$$|f_z(x)|^2 = f_z(x)\overline{f_z(x)} = \sum_{n,m} c_n \overline{c_m} e^{i\pi\lambda z \frac{1}{\tau^2}(m^2 - n^2)} e^{2\pi i \frac{x}{\tau}(n-m)}.$$

Agrupant els termes amb $n - m = l$, s'obté

$$|f_z(x)|^2 = \sum_l \left(e^{-i\pi\lambda z \frac{l^2}{\tau^2}} \sum_m \overline{c_m} c_{m+l} e^{-i\pi\lambda z \frac{2ml}{\tau^2}} \right) e^{2\pi i \frac{x}{\tau} l}.$$

Definim la funció $d_l(z)$ per

$$d_l(z) = \sum_m \overline{c_m} c_{m+l} e^{-i\pi\lambda z \frac{2ml}{\tau^2}}, \quad (11)$$

i s'obté, doncs,

$$|f_z(x)|^2 = \sum_l d_l(z) e^{-i\pi\lambda z \frac{l^2}{\tau^2}} e^{2\pi i \frac{x}{\tau} l}.$$

Recordant (1) i (2),

$$c_m = \begin{cases} \frac{1}{2} & \text{per a } m = 0, \\ 0 & \text{per a } m \text{ parell } \neq 0, \\ \frac{(-1)^k}{\pi m} & \text{si } m = 2k + 1, \end{cases}$$

es pot calcular $d_l(z)$ segons el valor de l :

1. Cas $l = 0$. En aquest cas es té

$$d_0(z) = \sum_m |c_m|^2.$$

Per Parseval es pot desenvolupar

$$\frac{\tau}{2} = \int_0^\tau |f_0(x)|^2 dx = \sum_m |c_m|^2 \int_0^\tau |e^{2\pi i \frac{x}{\tau} m}|^2 dx = \tau \sum_m |c_m|^2,$$

per concloure que $d_0 = c_0 = \frac{1}{2}$.

2. Cas $l = 2p \neq 0$. A l'equació (11) es tenen en compte, llavors, els casos $m = 2k + 1$ i $m = 0$:

$$\begin{aligned} d_{2p}(z) &= \bar{c}_0 c_{2p} + \sum_k \frac{(-1)^k}{\pi(2k+1)} \frac{(-1)^{k+p}}{\pi(2(p+k)+1)} e^{-i\pi\lambda z \frac{2(2k+1)2p}{\tau^2}} \\ &= \frac{(-1)^p}{\pi^2} e^{-i\pi\lambda z \frac{4p}{\tau^2}} \sum_{k=-\infty}^{+\infty} \frac{e^{-i\pi\lambda z \frac{8pk}{\tau^2}}}{(2k+1)(2k+2p+1)}. \end{aligned}$$

3. Cas l senar. Aquí es considera només $m = 0$ i $m = -l$, ja que en els altres casos $c_{m+l} = 0$, i $m + l$ és un nombre parell no nul, de manera que

$$d_l(z) = \frac{1}{2}c_l + \frac{1}{2}\bar{c}_{-l}e^{i\pi\lambda z \frac{2l^2}{\tau^2}} = \frac{1}{2}c_l(1 + e^{i\pi\lambda z \frac{2l^2}{\tau^2}}),$$

ja que $\bar{c}_{-l} = c_l$.

En definitiva, els coeficients de Fourier \tilde{c}_l de $|f_z|^2$ són

$$\begin{aligned} \tilde{c}_0 &= c_0 = \frac{1}{2}, \\ \tilde{c}_l &= c_l \cos\left(\pi\lambda z \frac{l^2}{\tau^2}\right), \\ \tilde{c}_l &= \frac{(-1)^p}{\pi^2} e^{-i\pi\lambda z \frac{l^2+2l}{\tau^2}} \sum_{-\infty}^{+\infty} \frac{e^{-i\pi\lambda z \frac{8pk}{\tau^2}}}{(2k+1)(2k+2p+1)}. \end{aligned}$$

El desenvolupament de Fourier de $P_z = |f_z|^2 * g$ és llavors

$$P_z(t) = \sum_n \tilde{c}_n \hat{g}\left(\frac{n}{\tau}\right) e^{2\pi i \frac{nt}{\tau}}.$$

El nostre problema és trobar g tal que P_z sigui una sinusoide pura de període τ . Com que ara cap dels \tilde{c}_n són zero en general, $\hat{g}\left(\frac{n}{\tau}\right)$ ha de ser zero per a tot $n \neq 0, 1, -1$; per tant, de la forma (fent una translació si és necessari)

$$g(x) = 1 + A \cos \frac{x}{\tau},$$

llevat de les constants multiplicatives. Com que es vol g positiva i amb suport compacte, triem senzillament la funció g amb suport a $[0, \tau]$ que val $1 + A \cos 2\pi \frac{z}{\tau}$ amb $0 \leq A \leq 1$. Evidentment,

$$\widehat{g}(0) = \tau, \quad \widehat{g}(\pm 1) = \frac{\tau A}{2},$$

amb la qual cosa, tenint en compte que

$$\tilde{c}_1 = c_1 \cos\left(\pi \lambda z \frac{1}{\tau^2}\right) = \frac{1}{\pi} \cos\left(\pi \lambda z \frac{1}{\tau^2}\right),$$

resulta

$$P_z(t) = \frac{\tau}{2} + \frac{A\tau}{\pi} \cos\left(\pi \lambda z \frac{1}{\tau^2}\right) \cos\left(2\pi \frac{t}{\tau}\right).$$

En definitiva, la propietat que explotem en la tria de la màscara robusta és senzillament que els sinus i els cosinus són vectors propis per a tots els operadors de convolució. En aquest cas veiem que no és possible obtenir una sinusoide pura que passi per zero. Consultada l'empresa, el fet que no passi per zero no representa cap inconvenient seriós, i, de fet, té algun avantatge tècnic, perquè representa que en tot moment el fotodiode rep fotons i això facilita que la seva resposta sigui més ràpida quan hi ha un increment d'intensitat.

Com es pot apreciar, el resultat esdevé sinusoidal per a qualsevol valor de z , sense que apareguin harmònics secundaris. Però, alhora, hom també pot veure que hi ha una dependència en z que afecta l'amplitud del senyal observat. L'amplitud de la sinusoide,

$$A(z) = \left| \cos\left(\pi \lambda z \frac{1}{\tau^2}\right) \right|,$$

és màxima en múltiples enters de la meitat de la distància de Talbot, i zero en aquests valors desplaçats $\frac{z\tau}{4}$.

La primera màscara té transformada

$$2 \cos\left(\frac{\pi}{2} \tau \xi\right) \frac{1}{1 - \tau^2 \xi^2};$$

per tant, té coeficients $2(-1)^n \frac{1}{1-n^2}$ en els parells, que per (3) són irrellevants; si a la primera màscara se li pertorben aquests harmònics, continuarà funcionant: en aquest sentit, és més robusta que la segona, però en un altre aspecte.

6. CODIFICADORS ABSOLUTS, CODIS ÒPTIMS I UN DICIONARI UNIVERSAL

En aquest darrer apartat tornem a considerar els codificadors absoluts per explicar una altra relació interessant amb la teoria de codis. Recordem que la finalitat d'un codificador absolut és codificar amb un codi binari cadascuna de les N posicions angulars. Si es vol mantenir la mida del disc dins d'un límit raonable, el problema és que per a resolucions N molt altes no hi ha prou espai físic al disc per gravar un codi per a cada radi. Si hi ha un límit k en el nombre de corones concèntriques, hi ha un límit 2^k en la resolució.

Aquest problema es tracta codificant la posició dels radis d'una altra forma. Pensem en un únic codi binari A de longitud N que ocupi el total de la perifèria del disc. Podem pensar que A és una successió N -periòdica de zeros i uns. Volem que aquest codi compleixi una propietat ben especial relativa a un número $n < N$: que tots els N subcodis de A consistents en n dígitos consecutius de A siguin diferents entre si. En particular, $N \leq 2^n$. Aquests codis existeixen i han estat extensament estudiats en combinatòria i teoria de grafs. Els més interessants són els que, a més, són *complets*, en el sentit que hi apareixen *tots* els codis binaris de longitud n . Aquests codis, de longitud $N = 2^n$, es coneixen com a *codis de Bruijn*: són codis binaris de longitud 2^n tals que tots els codis binaris de longitud n apareixen una única vegada com a subcodi consecutiu de A . Aquesta propietat permet codificar N radis: cada radi es codifica pel subcodi de A de longitud n centrat a l'extrem del radi.

I el mateix és cert per a codis no binaris que utilitzin un alfabet de m símbols: si $N \leq m^n$, hi ha codis de longitud N tals que els N subcodis de longitud n són tots diferents. L'article [1], i les seves referències, conté molta informació sobre aquest tema i un algorisme per generar aquests codis.

Com a aplicació curiosa en un altre context, podem pensar en les lletres de l'abecedari, amb $m = 28$ caràcters (incloent-hi la ζ i la l/l); hi afegim un caràcter addicional per separar paraules. Ara pensem en totes les frases que incloguin n caràcters: n'hi ha 29^n . Evidentment,

moltes d'aquestes frases no tindran significat. El que diu el resultat anterior, que està implementat en els codificadors absoluts, és que hi ha un llibre en català, amb 29^n caràcters, dins el qual es pot llegir *qualsevol frase possible*, tingui sentit o no, i és el més curt possible amb aquesta propietat.

Acabo amb un problema per als companys de la Secció Filològica: si ens ocupem només de frases amb sentit, i convenim que $n = 200$ és la longitud màxima d'una frase en català que es pot pronunciar sense respirar, quina és la longitud mínima del diccionari universal català dins el qual es pot llegir qualsevol frase?

REFERÈNCIES

- [1] BALLE, B.; VENTURA, E.; FUERTES, J. M. «An algorithm to design prescribed length codes for single-tracked shaft encoders». A: *Proceedings of the 2009 IEEE International Conference on Mechatronics (ICM 2009)* (Màlaga, abril 2009).
- [2] CILLERUELO, J.; RUZSA, I.; VINUESA, C. «Generalized Sidon sets». *Adv. Math.*, 225 (5) (2010), p. 2786-2807.
- [3] COSTA, N. *Innovació en el disseny i construcció de sensors rotatius òptics*. Tesi doctoral. UAB. Departament de Matemàtiques, 2018.
- [4] DIMITROMANOLAKIS, A. *Analysis of the Golomb ruler and the Sidon set problems and determination of large near-optimal Golomb rulers* [en línia]. <<http://www.cs.toronto.edu/~apostol/golomb/main.pdf>>.
- [5] ERDŐS, P.; TURÁN, P. «On a problem of Sidon in additive number theory, and on some related problems». *J. London Math. Soc.*, 16 (1941), p. 212-215.
- [6] GOODMAN, J. W. *Introduction to Fourier optics*. 3a ed. Englewood, CO: Roberts and Company Publishers, 2005.

- [7] MOORE, E. H.; POLLATSEK, H. S. *Difference sets: Connecting algebra, combinatorics, and geometry*. Providence, RI: American Mathematical Society, 2013. (Student Mathematical Library; 67)
- [8] SINGER, J. «A theorem in finite projective geometry and some applications to number theory». *Trans. Amer. Math. Soc.*, 43 (3) (1938), p. 377-385.
- [9] STEIN, E. M. *Harmonic analysis: Real-variable methods, orthogonality, and oscillatory integrals*. Princeton, NJ: Princeton University Press, 1993. (Princeton Mathematical Series; 43. Monographs in Harmonic Analysis; III)

