A Matemática e a Transmissão de Informação Uma introdução aos códigos *

Juscelino Bezerra † ‡
Patricia Helena Araújo da Silva Nogueira §

Resumo

Nesse artigo apresentamos um pequena introdução à teoria dos códigos corretores de erros e a sua surpreendente relação com uma das mais abstratas áreas da matemática através dos códigos algébricos-geométricos de Goppa.

Abstract

In this article we present a short introduction to the error correcting codes and theirs amazing relation with one of the most abstract mathematical theory by the algebraic-geometric Goppa codes.

1 Introdução

Em nosso mundo físico, desde muito tempo, o homem, com a sua capacidade de percepção do que julga ser um problema e o desejo de significar uma ação capaz de resolvê-lo, usa a matemática como um sistema formal de pensamento para reconhecer, classificar e explorar padrões. Formada por signos linguísticos que passam idéias e significados, a matemática é uma linguagem universal, e no método matemático, o qual a partir de indícios, deduzem-se regras, temos um tipo de beleza que se aplica às idéias e não somente aos objetos.

Esse sentido etéreo e atemporal das idéias matemáticas já pode ser encontrado nos antigos gregos como Platão, não se fazendo durante muito tempo uma distinção clara entre a matemática útil/aplicável e a abstração pura. Tal dicotomia começa a se fortalecer no final do século dezenove culminando no início do século vinte com o crescente grau de abstração e formalismo matemático.

Daí em diante as teorias matemáticas foram se tornando cada vez mais complexas e aparentemente distantes do "homem comum" a ponto do famoso matemático G.H.Hardy em seu livro "Apologia do Matemático" escrever: "se um jogo de xadrez é, num sentido rude, "inútil", então isto é igualmente verdade para a maior parte da mais refinada matemática (...). Eu nunca fiz nada "útil". Nenhuma descoberta que fiz já produziu, direta ou indiretamente, para o bem ou para o mal a menor diferença na melhoria do mundo." Hardy estudava Teoria de Números e se orgulhava, em meados dos anos 40, de que não havia nenhuma utilidade bélica para a mesma. Hoje porém existem grandes aplicações da Teoria

* *Palavras chave:* Códigos corretores de erros, curvas algébricas, corpos finitos, códigos de Goppa
† Departamento de Análise Matemática, IME/UERJ, bezerra@ime.uerj.br
‡ O autor foi parcialmente financiado pela FAPERJ
§ Departamento de Matemática, Estatística e Computação, FAT/UERJ, patricia@fat.uerj.br

de Números na segurança da transmissão de informação (criptografia) o que tem obviamente relações estreitas com o mundo "bélico" e "capitalista"!

Na verdade, a maior parte dos matemáticos hoje talvez compactue com as idéias expostas pelo físico matemático Roger Penrose no livro "A mente nova do rei"([9]): "De um lado a matemática deve ser estudada e compreendida por si mesma, e não devemos querer uma aplicabilidade totalmente exata aos objetos da experiência física, do outro, o funcionamente do mundo externo concreto só pode, em última análise, ser entendido em termos da matemática precisa."

Seguindo tal raciocínio, não é estranho aos matemáticos que certos padrões e estruturas matemáticas possam ser identificados no mundo concreto, ainda que tais padrões façam parte da Álgebra, uma das áreas mais abstratas da matemática. Através desse artigo queremos exemplificar tais idéias através de uma das mais inusitadas aplicações do pensamento puro matemático propiciada pelo nosso moderno mundo digital.

2 Códigos Corretores de Erros

Quando escutamos uma música através de um CD estamos nesse processo utilizando um sistema de reprodução sofisticado que traduz uma música que está digitalizada para uma música em formato de som correta. Os reprodutores digitais foram desenvolvidos no final dos anos oitenta, com a promessa de melhorar a qualidade de áudio, da alta fidelidade, reduzindo os ruídos e chiados das fitas cassete e dos discos de vinil. CDs e discos de vinil são semelhantes no que se refere aos seus objetivos, ambos são dispositivos de armazenamento de aúdio. Enquanto para os discos de vinil, a agulha do toca-discos reproduz mecanicamente os sinais analógicos que os geraram, no CD as informações são gravadas por traços. A medida do comprimento de cada traço corresponde a cada informação. A leitura é feita por um finíssimo feixe de laser que focaliza a linha tracejada no disco e é dirigido a um conjunto de detectores. Dessa forma, esses detectores podem medir o comprimento dos traços, tornando possível a leitura da informação, além de manter o feixe na trilha correta. Os CDs podem reproduzir qualquer sinal digitalizado, ou seja, transformado em dígitos binários, além dos sinais de áudio ([13]).

Este processo de reprodução, seja de um CD ou de um disco de vinil, consiste da leitura de uma informação codificada, por meio digital no caso do CD, e decodificada por meio de aúdio. Codificação significa a modificação das características de uma informação para torná-la mais apropriada para uma aplicação específica, como por exemplo transmissão ou armazenamento de dados. Garantir que uma grande quantidade de informação transmitida pelos mais variáveis meios chegue intacta ao receptor e possa ser lida, escutada com nitidez no caso de uma mensagem de aúdio, é um problema abordado pela área da teoria da informação chamada de Teoria dos códigos, mais especificamente pelos Códigos Corretores de Erros. Pode-se afirmar que hoje praticamente todo sistema de envio de informações possui algum tipo de código corretor de erros. Como exemplos típicos, a telefonia digital, a transmissão de dados via satélite, a comunicação interna em computadores, armazenamento ótico de dados e armazenamento de dados em fitas ou disquetes.

A teoria da informação aborda os aspectos quantitativos de armazenamento e transmissão das mensagens e tem como um de seus objetivos principais garantir a autenticidade dos dados enviados através de algum tipo de canal. Na manipulação das mensagens, dois obstáculos são encontrados:

(a) baixa capacidade de armazenamento ou de transmissão das mensagens enviadas;

(b) surgimento aleatório de erros nas mensagens enviadas, isto é, ruídos na transmissão.

Para resolver estes problemas a teoria dos códigos estuda técnicas que tentam detectar e corrigir tais erros, duas tarefas independentes, pois dependendo do objetivo, certos modelos podem ser mais eficientes. A teoria dos códigos corretores de erros surgiu nos laboratórios de empresas de telefonia em 1948 com um trabalho de C. E. Shannon ([10]), e posteriormente se transformou em uma teoria matemática completa com interações com várias áreas como, por exemplo, Geometria Algébrica, e trata de ferramentas que visam recuperar informações que no processo de emissão tenham sofrido algum tipo de ruído.

Ruídos podem sugir das mais variadas formas. Pro exemplo, quando usamos as palavras "cama" e "cana" estamos nos referindo a verbetes distintos, mas que de certa forma estão bastante "próximos". Basta-se trocar a terceira letra para que se mude completamente a palavra. Este tipo de anomalia pode ocorrer no processo de envio de uma mensagem e pode ser resolvido quando introduzimos mais símbolos (letras) nas palavras e/ou determinamos para nosso dicionário (ou melhor, o conjunto de palavras utilizadas na mensagem) uma distância entre as palavras que possa identificar quando determinada palavra faz parte ou não da mensagem enviada, estimando-se qual a palavra transmitida a partir da palavra mais próxima da que foi recebida.

Neste contexto, o tamanho do alfabeto utilizado na formação das palavras e o tamanho das palavras são fundamentais na hora da decodificação. Quanto menos letras trocadas, mais autenticidade garantimos para a mensagem transmitida.

Para estabelecermos algum tipo de algoritmo de detecção e correção de erros temos que começar a traduzir nosso problema para a linguagem matemática. Representamos as letras do alfabeto a ser utilizado como elementos de um certo conjunto K e as palavras em seqüências numéricas de mesmo tamanho n, completando, por exemplo, as palavras de tamanho menor com zeros, e onde alguns dígitos desta seqüência são utilizados para verificação. Obtemos assim um grande conjunto $\mathcal{P}$ de todas as palavras possíveis de serem construídas com nosso alfabeto (de um certo tamanho pré-determinado e mesmo que não tenham sentido semântico). Observe que $\mathcal{P}$ nada mais é que $K^n = K \times K \times \cdots \times K$. O código então é definido como um subconjunto $\mathcal{C} \subset \mathcal{P}$ das palavras que estão presentes na nossa mensagem.

A noção de "proximidade"entre palavras é traduzida em termos matemáticos pela *distância de Hamming.* Se $u = (u_1, u_2, \cdots, u_n)$ e $v = (v_1, v_2, \cdots, v_n)$ são duas palavras distintas em $\mathcal{P}$ definimos a distância entre elas como sendo:

$$d(u, v) = \sharp\{i \,/\, u_i \neq v_i, i = 1, \cdots, n\}$$

onde $\sharp$ significa cardinalidade. Não é difícil ver que d assim definida é realmente uma métrica em $\mathcal{P}$, valendo por exemplo que a distância entre duas palavras é sempre positiva e é nula se, e somente se, elas

são iguais (além de serem válidas a simetria e a desigualdade triangular). Vemos que o nosso código $\mathcal{C}$ já alcança assim o status de um espaço métrico (finito!).

Chamando de $\mathbf{r}$ a palavra recebida (possivelmente com algum erro em algum caractere, produzido por algum ruído no canal de transmissão) podemos estabelecer a nossa primeira tentativa de algoritmo:

(i) Calcule todas as distâncias $d(\mathbf{r}, u)$ fazendo u percorrer o código $\mathcal{C}$.

(iii) Troque $\mathbf{r}$ por um elemento $\mathbf{c}$ de $\mathcal{C}$ possuindo a menor distância de $d(\mathbf{r}, \mathbf{c})$.

Observe que se a palavra recebida pertencer ao código, ela permanece inalterada (pois nesse caso a menor distância será zero). Além disso, vemos que o sucesso do último passo depende do fato de haver palavras do código suficientemente próximas da enviada e que haja uma que seja realmente sua melhor aproximação. Ainda assim, não podemos garantir, a priori, que detectamos e corrigimos todos os erros cometidos. A fim de tirarmos algumas conclusões fazemos uso da estrutura de espaço métrico já identificada.

Para isso precisaremos da noção de *distância mínima d,* que nada mais é que a menor distância detectada entre dois elementos distintos de $\mathcal{C}$. Portanto, caso ocorram até $d-1$ erros na palavra enviada, estes serão detectados (mais que isso, pode-se achar uma palavra do código igual a palavra recebida e considerá-la correta).

Definindo κ como sendo a parte inteira do número real $\frac{d-1}{2}$, podemos mostrar (usando a desigualdade triangular e a minimalidade de d) que discos de raio κ centrados em palavras distintas do código são sempre disjuntos. Supondo então que no máximo κ erros foram cometidos, podemos melhorar o nosso algoritmo:

(i) Calcule todas as distâncias $d(\mathbf{r}, u)$ fazendo u percorrer o código $\mathcal{C}$.

(ii) Encontre $\mathbf{c} \in \mathcal{C}$ satisfazendo $d(\mathbf{r}, \mathbf{c}) \leq \kappa$.

(iii) Troque $\mathbf{r}$ por $\mathbf{c}$.

Pelo que já vimos, caso exista, $\mathbf{c}$ encontrado no segundo passo é único. O problema é que $\mathbf{r}$ pode não pertencer a nenhum disco de raio κ centrado em elementos de $\mathcal{C}$ e nesse caso não saberíamos com certeza por qual elemento do código devemos trocar a palavra recebida. Analisando o algoritmo, outros problemas aparecem: ele simplesmente compara a palavra recebida com cada palavra do código $\mathcal{C}$, implicando em um custo computacional imenso caso o número de elementos de $\mathcal{C}$ e n (o tamanho das palavras) sejam muito grande. Porém a utilidade de um código corretor de erros está intimamente ligada a quantidade de informação transmitida! Entra aqui, portanto, a idéia de um *bom código,* o que além de tornar possível a detecção e correção de muitos erros, seja "grande o suficiente" para podermos mandar muita informação sem perdermos o controle.

Novamente, trabalhando somente com a estrutura de espaço métrico adquirida, conseguimos estabelecer uma relação matemática (a *Cota de Singleton*) entre os chamados *parâmetros do código*, a saber:

$$M \leq q^{n-d+1}$$

onde q é o número de símbolos do alfabeto K, M é o número de elementos de $\mathcal{C}$, n é o comprimento das palavras de $\mathcal{C}$ e d a sua distância mínima (a qual nos informa a real capacidade de correção do código).

Quando medimos a eficiência de um código tais parâmetros têm claramente importância fundamental. De fato, definindo a taxa de informação de $\mathcal{C}$ como sendo $\frac{log_q M}{n}$ e a de correção de erros de $\mathcal{C}$ como sendo $\frac{d}{n}$, vemos que, quanto maiores forem essas taxas, melhor será o código. Dessa forma, tendo em vista a relação que os parâmetros do código satisfazem, estamos olhando para o comportamento assintótico destes parâmetros em função do tamanho das palavras, isto é, queremos olhar como os outros parâmetros se comportam quanto n cresce indefinidamente. Neste estudo assintótico várias cotas surgem na tentativa de se determinar bons códigos, dentre estas cotas a mais importante é a de Gilbert-Varshamov (Vide [7], [8]).

Para podermos avaliar e controlar todas essa variáveis, melhorando inclusive nosso algoritmo de correção de erros é que se torna imperiosa a detecção de alguma outra estrutura matemática em $\mathcal{C}$ além da que já expusemos. Nesse sentido, a Álgebra aparece de forma natural (ao menos a um matemático!), pois estamos lidando com conjuntos finitos de seqüências de símbolos.

Por exemplo, ao alfabeto K pode-se, em vários casos (mais especificamente quando o número de elementos q é uma potência de um número primo), ser atribuído uma estrutura algébrica denominada *corpo.* Grosso modo, tal conjunto de símbolos é munido de certas operações que o tornam muito semelhante ao conjunto dos números reais (a não ser pela sua finitude). Dessa maneira, o conjunto de todas as palavras possıveis K^n torna-se um *espaço vetorial* tendo como conjunto de escalares o alfabeto K (agora um corpo finito). Para que $\mathcal{C}$ herde tal estrutura, pede-se então que este seja um subespaço vetorial de K^n, passando a ser chamado então de um Código Linear. Aqui, surge um novo parâmetro denotado por k dado pela dimensão de $\mathcal{C}$ como espaço vetorial e relacionado com o parâmetro M (número de elementos de $\mathcal{C}$) por $k = \log_q M$, onde q é a cardinalidade do corpo K. Passa-se a ter nesse caso um controle maior sobre os parâmetros de $\mathcal{C}$, melhorando cotas, otimizando cálculos (como o da distância mínima d) e tornando muito mais eficiente o algoritmo de correção de erros. Por exemplo, a fim de verificar se a palavra recebida $\mathbf{r}$ pertence ao código basta verificar se esta pertence ao núcleo da chamada *matriz de paridade* H de $\mathcal{C}$. No caso em que Hu for não nulo, tal elemento (denominado *síndrome*) ajuda de forma algorítmica a calcular o vetor $\mathbf{c}$ o qual deverá substituir $\mathbf{r}$. Um Código Linear fica assim completamente determinado pela sua matriz de paridade, pois passa a ser visto como o núcleo de uma certa transformação linear.

Como exemplos de códigos lineares temos os *códigos de Hamming*, uma importante família de códigos de fácil codificação e decodificação. Um código binário de Hamming de comprimento $n = 2^r - 1$ é determinado por uma matriz de paridade cujo as colunas consistem de todos os vetores não-nulos de comprimento r. Para estes códigos temos os seguintes parâmetros: $n = 2^r - 1$, $k = 2^r - 1 - r$ e $d = 3$ (vide [7] Capítulo 5). Outra família importante de códigos lineares é conhecida como *códigos BCH*, ou códigos de Bose-Chaudhuri-Hocquenghem. Estes códigos generalizam, em algum sentido, os códigos de Hamming e são determinados a partir de uma matriz de Vandermonde com entradas dadas em termos de uma n-ésima raiz primitiva da unidade no corpo finito sobre o qual consideramos o alfabeto.

Na ânsia de encontrarmos melhores códigos o processo continua e estruturas cada vez mais complexas e abstratas acabam "surgindo", chegando, por exemplo a interagir com subáreas da Álgebra antes "confinadas" ao mundo das idéias platônico. Um desses casos é o aparecimento de códigos (lineares) associados às chamadas *curvas algébricas* definidas sobre corpos finitos, estruturas essas que fazem parte da Geometria Algébrica.

3 Códigos Algébricos-Geométricos de Goppa

Os códigos de Goppa surgiram na década de 70 como um modelo especializado dos códigos BCH. Estes códigos revitalizaram o estudo de curvas algébricas projetivas, irredutíveis e não-singulares definidas sobre um corpo finito[1]. Para um estudo detalhado sobre tal construção veja [11].

A construção de Goppa buscava determinar uma seqüência de códigos atingindo a cota de Gilbert-Varshamov. Durante muito tempo os pesquisadores da teoria dos códigos corretores de erros foram incapazes de exibir construções de códigos com parâmetros relativos (taxa de informação e taxa de correção de erros) que atingissem ou superassem esta cota, o que os levou a pensar que o gráfico da função que define esta cota limitasse a região do plano real conhecida como domínio dos códigos. Para surpresa de tais pesquisadores, Tsfasman, Vladut e Zink conseguiram uma seqüência de códigos de Goppa que superava a cota de Gilbert-Varshamov em alguns casos (Vide [12]). Uma tal seqüência de códigos de Goppa é construída a partir de uma seqüência de curvas algébricas definidas sobre um corpo finito, chamada torre de curvas.

Existe um parâmetro assintótico a partir do qual se determina se uma torre de corpos dá origem, ou não, a uma boa seqüência de códigos corretores de erros. Este parâmetro, definido a partir do gênero e do número de lugares racionais de cada curva que forma a torre, é chamado de limite da torre. Torres com limite positivo, conhecidas como torres assintoticamente boas, podem medir a precisão dos parâmetros relativos de códigos lineares, na medida em que melhoram a cota de Gilbert-Varshamov. O estudo de torres assintoticamente boas tem se desenvolvido muito nos últimos anos e exemplos explícitos destas torres, isto é, torres onde as curvas são determinadas por equações explícitas, têm sido explorados em vista de serem essenciais em termos de aplicações. Para um apanhado geral sobre a teoria de torres de curvas definidas explicitamente por equações veja [6]. Em [3, 4, 5], Garcia e Stichtenoth apresentam exemplos de torres assintoticamente boas definidas sobre corpos com q^2 elementos. Um exemplo interessante definido sobre o mesmo corpo também foi dado por Bezerra e Garcia em [1]. Em [2], um exemplo importante de uma torre explícita assintoticamente boa definido sobre um corpo com q^3 elementos, foi dado por Bezerra, Garcia e Stichtenoth. Vale observar que alguns códigos construidos a partir destes exemplos, melhoram a cota de Gilbert-Varshamov.

Por fim, vale observar que existem outros tipos de códigos, obtidos através de outras construções, interagindo com outras áreas da Matemática[2] e fortalecendo assim a interdisciplinaridade e tornando

[1] Uma curva algébrica plana é o lugar geométrico dos zeros de um polinômio de duas variáveis e apesar de termos curvas algébricas vivendo em ambientes mais sofisticados do que o plano, toda curva algébrica tem um modelo plano.

[2] Levando-se em conta aspectos probabilísticos por exemplo. Os códigos citados nesse artigo partem da premissa que o

ainda mais tênue a distinção entre a matemática pura e a aplicada.

canal é simétrico, significando que todos os caracteres têm a mesma probabilidade de serem trocados por qualquer outro do alfabeto.

Referências

[1] BEZERRA, J. GARCIA, A. A tower with non-Galois steps which attains the Drinfeld-Vladut bound, J. Number Theory, Vol. 106, 2004, p. 142-154.

[2] BEZERRA, J. GARCIA, A. STICHTENOTH, H. An explicit tower of function fields over cubic finite fields and Zinks lower bound, J. Reine Angew. Math., Vol. 589, n. 8, 2005, p. 159-199.

[3] GARCIA, A. STICHTENOTH, H. A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vladut bound, Invent. Math., Vol. 121, 1995, p. 211-222.

[4] GARCIA, A. STICHTENOTH, H. Asymptotically good towers of function fields over finite fields, C.R. Acad. Sci. Paris, t. 322, Sr. I, 1996, p. 1067-1070.

[5] GARCIA, A. STICHTENOTH, H. On the asymptotic behaviour of some towers of function fields over finite fields, J. Number Theory, Vol. 61, 1996, 248-273.

[6] GARCIA, A. STICHTENOTH, H. Topics in Geometry, Coding Theory and Cryptography, Springer.

[7] HEFEZ, A., VILLELA, M.L.T, Códigos Corretores de Erros, Coleção Computação e Matemática, SBM.

[8] MACWILLIAMS, F.J., SLOANE, N.J.A. The Theory of Error-Correcting Codes, North-Holland.

[9] PENROSE, R., A mente nova do rei, Ed.Campus.

[10] SHANNON, C.E., A mathematical theory of communication, Bell Syst. Tech. J., 27, pp.379-423, 623-656 (1948).

[11] STICHTENOTH, H. Algebraic Function Fields and Codes, Springer-Verlag.

[12] TSTASMAN, M.A. VLADUT, S.G. ZINK,T. Modular curves, Shimura curves and Goppa codes better than the Varshamov-Gilbert bound, Math. Nachr., Vol. 109, 1982, p. 21-28.

[13] http://pt.wikipedia.org/wiki/Leitor_de_CD