TP11 : Codes correcteurs d'erreurs algébriques

Résumé du TP

Bertrand Meyer

18 mai 2020

Codes linéaires

Définitition

Un $[n, k, d]_q$-code correcteur d'erreurs linéaire $\mathcal{C}$ est un sous-espace vectoriel de $\mathbb{F}_q^n$ de dimension k tel que pour tout couple de mots de codes $\mathbf{c}$ et $\mathbf{c}'$,

$$d_H(\mathbf{c}, \mathbf{c}') \geq d$$

où d_H est la distance de Hamming.

Matrice génératrice $\mathbf{G}$: $\mathcal{C} = \{\mathbf{m}\mathbf{G}; \mathbf{m} \in \mathbb{F}_q^k\}$.
Matrice de contrôle $\mathbf{H}$: $\mathcal{C} = \{\mathbf{c} \in \mathbb{F}_q^n; \mathbf{H}\mathbf{c} = 0\}$.

Le décodage

Décoder un message reçu $\mathbf{r}$ revient à trouver $\mathbf{c} \in \mathcal{C}$ tel que la probabilité que $\mathbf{c}$ se soit transformé en $\mathbf{r}$ est maximale. Les maths montrent qu'il faut calculer $\min\{d_H(\mathbf{r}, \mathbf{c}); \mathbf{c} \in \mathcal{C}\}$ ($\to$ décodage par maximum de vraisemblance).

En pratique on dispose de techniques qui recherchent l'ensemble

$$\mathsf{C} = \{\mathbf{c} \in \mathcal{C}; d_H(\mathbf{r}, \mathbf{c}) \leq t\}.$$

à t fixé.

- Si $2t < d$, on est sûr que $|\mathsf{C}| = 0$ ou 1. On parle de décodage unique.
- Sinon, C est de taille quelconque. On parle de décodage en liste.

Le décodage des codes de Reed Solomon

Les codes de Reed-Solomon

Soit α un élément primitif de $\mathbb{F}_q$, $n = |\mathbb{F}_q^\times| = q - 1$, le code de Reed-Solomon s'obtient par évaluation de polynômes

$$\mathcal{RS}_{k,q} = \left\{ \left(f(\alpha^0), f(\alpha^1), \cdots, f(\alpha^{q-2})\right), f \in \mathbb{F}_q[X]_{<k} \right\}.$$

Code $[n, k, n-k+1]_q$, car un polynôme f n'annule au plus $\deg f$ fois.

Matrice génératrice : $\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\ 1 & \alpha^2 & \alpha^4 & \cdots & \alpha^{2n-2} \\ \vdots & & & & \\ 1 & \alpha^{k-1} & \alpha^{2k-2} & \cdots & \alpha^{kn-k} \end{pmatrix}$

Matrice de contrôle : $\begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\ 1 & \alpha^2 & \alpha^4 & \cdots & \alpha^{2n-2} \\ \vdots & & & & \\ 1 & \alpha^{n-k} & \alpha^{2n-2k} & \cdots & \alpha^{(n-k)(n-1)} \end{pmatrix}$

Le décodage de Berlekamp-Massey

Soient

$$\begin{aligned}
\mathbf{e} = \mathbf{r} - \mathbf{c} &= (e_0, \dots, e_{n-1}), \text{ le vecteur d'erreur,} \\
e(x) &= e_0 + e_1 x + \cdots + e_{n-1} x^{n-1}, \\
\mathbf{s} = \mathbf{H}\mathbf{r} = \mathbf{H}\mathbf{e} &= (s_0, \dots, s_{n-k}), \text{ le syndrôme,} \\
M &= \{i \in [\![0, n-1]\!]; e_i \neq 0\} \text{ la position des erreurs.}
\end{aligned}$$

Alors

$$\begin{aligned}
s_j &= e(\alpha^j) \\
\sum_{\ell=0}^{\infty} e(\alpha^\ell) x^\ell &= \frac{[\cdots]}{\underbrace{\prod_{i \in M} (1 - \alpha^i x)}_{\sigma(x)}}
\end{aligned}$$

L'algorithme de Berlekamp Massey calcule σ à partir des $2|M|$ premiers termes de $\left(e(\alpha^\ell)\right)_{\ell \in \mathbb{N}}$.

Le décodage de Sudan

1. Trouver un polynôme bivarié $Q(x, y) \in \mathbb{F}_q[x, y]$ tel que $Q(\alpha^i, r_i) = 0$ pour tout i
2. Chercher ses facteurs $y - f(x)$.
3. Renvoyer $f(x)$ quand f est de degré $< k$ et $(f(\alpha^i))_{i<n}$ coïncide avec $\mathbf{r}$ au moins $n - t$ fois.

Les codes algébriques

Les limites des codes de Reed-Solomon

Les codes de Reed-Solomon sont

- très bons car MDS (maximum distance separable), i.e. $n - 1 = k + d$. (Un polynôme a un nombre restreint de zéros)
- mais de longueur limitée, car $n \leq q - 1$ ne croit pas à alphabet constant.

Les codes algébriques :

- même ingrédients : évaluer une partie bornée des fonctions rationnelle sur les points d'une courbe ou variété ($\rightarrow$ on contrôle le nombre de zéro).
- possibilité de faire croitre le nombre de points (i.e. la taille du code).

Rappels de géométrie algébrique

Un diviseur est une notation

$$D = \sum_{Q \in \mathcal{X}} n_Q(Q)$$

pour dénoter un ensemble de points avec des *multiplicités*.

Le degré d'un diviseur est l'entier

$$\deg D = \sum_{Q \in \mathcal{X}} n_Q \in \mathbb{Z}.$$

Le diviseur principal d'un fonction rationnelle $g \in \mathbb{F}_q(\mathcal{X})$ dénote la multiplicité des zéros/pôles de g.

$$\mathrm{div}(g) = \sum_{Q \in \mathcal{X}} \mathrm{ord}_Q(f)(Q)$$

Il est toujours de degré 0.

L'espace des fonctions rationnelles associées à D est

$$\mathcal{L}(D) = \{g \in \mathbb{F}_q ; \operatorname{div}(g) + D \geq 0\} \cup \{0\}.$$

(Théorème de Riemann-Roch) C'est un espace vectoriel de dimension

$$\dim \mathcal{L}(D) \geq \deg D + 1 - g.$$

avec égalité si $\deg D \geq 2g - 2$.

Codes de Goppa

On fixe

- $\mathcal{X}$ une courbe projective lisse
- des points $\mathcal{S} = \{P_1, P_2, \ldots, P_n\}$
- un diviseur D à support disjoint de $\mathcal{S}$

On évalue les éléments de $f \in \mathcal{L}(D)$:

$$(f(P_1), f(P_2), \ldots, f(P_n)).$$

On obtient un code :

- de longueur n,
- de dimension $\deg D + 1 - g$ (si $2g - 2 < \deg D < n$)
- et de distance minimale $\geq n - \deg D$.

Bilan sur les codes construits

Avec des codes algébriques, on parvient à construire des codes vérifiant

$$k + d \geq n + 1 - g$$

soit asymptotiquement

$$R + \delta \geq 1 - \frac{g-1}{n}.$$

où R est le rendement, δ la distance minimale relative, g le genre et n le nombre de points sur la courbe.

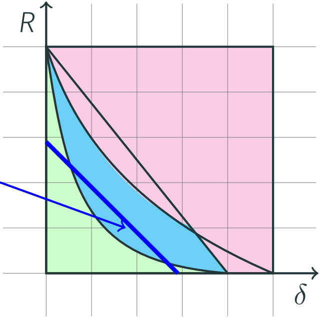
Les codes algébriques battent la borne de Gilbert-Varshamov

On pose

$$N_q(g) = \max\left\{|\mathcal{X}(\mathbb{F}_q)|\ ;\ \mathcal{X} \text{ est une courbe sur } \mathbb{F}_q \text{ de genre } g\right\},$$

$$A(q) = \limsup_{g \to +\infty} \frac{N_q(g)}{g}.$$

Droite $R + \delta = 1 - \dfrac{1}{A(q)}$



On obtient les meilleures familles de codes connues à ce jours pour de grands alphabets.