Université Paris 13 Année 2021–2022

M2 Mathématiques des données

Codes correcteurs — Sujet de projet n°1
Partage de secret

11 janvier 2022

Description du sujet. Dans un schéma de partage de secret, on souhaite répartir une valeur secrète parmi plusieurs participants, de sorte que l'on ne puisse révéler le secret que si un certain nombre (appelé seuil) de participants le souhaite. En termes d'application, on peut penser au partage d'une clef d'accès (à des données communes, à du matériel commun), qui ne peut être utilisée que si une majorité des collaborateurs donne son accord.

Objectifs. Il est proposé de traiter les points suivants :

1. Décrire le protocole partage de secret de Shamir [2] avec un formalisme « polynomial ». En démontrer les propriétés.
2. Abstraire et généraliser ce protocole avec un formalisme de codes correcteurs. Donner et démontrer formellement les propriétés du protocole (seuil de reconstruction, sécurité, nombre de participants, complexité de calcul) en fonction des paramètres du code. On pourra s'aider de [1].
3. Dans un contexte que l'on précisera (polynomial, codes correcteurs), implanter entre autres les fonctions de :
   - génération des parts,
   - reconstruction du secret par un sous-ensemble de participants de cardinal supérieur au seuil de reconstruction.

   Proposer un script d'exemple, facilement exécutable, pour illustrer votre code.

Références

[1] James L. Massey. Minimal Codewords and Secret Sharing. In *Proceedings of the 6th Joint Swedish-Russian International Workshop on Information Theory*, pages 276–279, 1993.

[2] Adi Shamir. How to Share a Secret. *Commun. ACM*, 22(11) :612–613, 1979.

Université Paris 13 Année 2021–2022

M2 Mathématiques des données

Codes correcteurs — Sujet de projet n°2
Calcul matriciel distribué

11 janvier 2022

Description du sujet. Dans beaucoup d'entreprises ou de laboratoires, il est de plus en plus courant de déporter des calculs sur des grappes de machines distantes. Il est même courant qu'il faille distribuer les calculs (c'est-à-dire les séparer en sous-tâches, données à différentes machines), car les données du calcul à effectuer sont trop massives pour être stockées sur une seule machine. La distribution des calculs sur plusieurs machines pose alors la question de la récupération de la solution lorsqu'une (ou plusieurs) machine est défaillante ou trop lente.

Dans ce projet, on s'intéresse au problème calcul matriciel distribué. L'idée est alors d'encoder les matrices afin d'ajouter de la redondance au calcul, de distribuer les calculs sur plusieurs serveurs, puis d'être capable de retrouver de résultat final malgré la présence de machines défaillantes.

Objectifs. Il est proposé de traiter les points suivants :

1. Décrire un ou plusieurs protocoles de calcul d'algèbre linéaire distribué (par exemple : produit matriciel, produit matrice-vecteur, etc.). On se réfèrera aux documents suivants : [2, 3, 1].
2. Implanter (entre autres) les fonctions de :
   - encodage et distribution des données sur lesquelles mener les calculs
   - calcul « partiel » à mener par une machine
   - reconstruction de la solution dans le cas où une (ou plusieurs) machine ne rend pas de résultat à temps

Références

[1] Sanghamitra Dutta, Mohammad Fahim, Farzin Haddadpour, Haewon Jeong, Viveck R. Cadambe, and Pulkit Grover. On the optimal recovery threshold of coded matrix multiplication. *IEEE Trans. Inf. Theory*, 66(1) :278–301, 2020.

[2] Kuang-Hua Huang and Jacob A. Abraham. Algorithm-based fault tolerance for matrix operations. *IEEE Trans. Computers*, 33(6) :518–528, 1984.

[3] Qian Yu, Mohammad Ali Maddah-Ali, and Salman Avestimehr. Polynomial codes : an optimal design for high-dimensional coded matrix multiplication. In *Advances in Neural Information Processing Systems 30 : Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 4403–4413, 2017.

Université Paris 13 Année 2021–2022

M2 Mathématiques des données

Codes correcteurs — Sujet de projet n°3
Premiers protocoles de PIR

11 janvier 2022

Description du sujet. Le retrait confidentiel d'information (*private information retrieval*, PIR) permet d'extraire une entrée d'une base de données stockée à distance, sans révéler au serveur l'identité de l'entrée désirée.

Dans un modèle où l'on souhaite une sécurité inconditionnelle, la base de donnée est répliquée sur plusieurs serveurs. L'utilisateur qui souhaite retrouver une entrée de la base de données transmet alors à chaque serveur une requête qui semble aléatoire. Grâce aux réponses des serveurs (en accord avec les requêtes), l'utilisateur peut alors reconstruire l'entrée désirée.

Objectifs. Il est proposé de traiter les points suivants :

1. Définir formellement un protocole de PIR dans un modèle de sécurité inconditionnelle.
2. Décrire le protocole de Chor, Goldreich, Kushilevitz et Sudan [1], d'abord avec un exemple, puis dans un cadre général.
3. En se basant sur [1], implanter :
   - une fonction qui engendre les requêtes aux serveurs,
   - une fonction qui calcule la réponse d'un serveur,
   - une fonction qui reconstruit l'information à partir des réponses des serveurs.

Références

[1] Benny Chor, Eyal Kushilevitz, Oded Goldreich, and Madhu Sudan. Private Information Retrieval. *J. ACM*, 45(6) :965–981, 1998.

Université Paris 13 Année 2021–2022

M2 Mathématiques des données

Codes correcteurs — Sujet de projet n°4
Codes de Goppa binaires

11 janvier 2022

Description du sujet. Les codes de Goppa binaires forment une famille de correcteurs sur l'alphabet $\mathbb{F}_2$, qui a trouvé des applications en télécommunication et en cryptographie. Il sont notamment utilisés dans le schéma de chiffrement de McEliece.

Ces codes disposent d'algorithmes de décodage efficaces jusqu'à leur rayon de décodage unique (par exemple, algorithme de Patterson), et possèdent une structure algébrique qui les lient aux codes de Reed-Solomon. Pour plus de détails, voir notamment l'article originel de Patterson [2], ou les trois premières sections de [1].

Objectifs. Il est proposé de traiter les points suivants :

1. Définir formellement les codes de Goppa binaires. Donner leurs paramètres et en faire la preuve.
2. Présenter un algorithme de décodage (par exemple celui de Patterson).
3. Implanter :
   - une fonction qui produit la matrice génératrice d'un code de Goppa binaire,
   - un algorithme de décodage pour les codes de Goppa binaire, que l'on pourra mettre en application.

Références

[1] Daniel J. Bernstein. List decoding for binary goppa codes. In *Coding and Cryptology - Third International Workshop, IWCC 2011, Qingdao, China, May 30-June 3, 2011. Proceedings*, volume 6639 of *Lecture Notes in Computer Science*, pages 62–80. Springer, 2011.

[2] Nicholas J. Patterson. The algebraic decoding of goppa codes. *IEEE Trans. Inf. Theory*, 21(2) :203–207, 1975.

Université Paris 13 Année 2021–2022

M2 Mathématiques des données

Codes correcteurs — Sujet de projet n°5
Codes elliptiques dans le système de McEliece

11 janvier 2022

Description du sujet. Une variation du cryptosystème de McEliece propose d'utiliser une famille de codes elliptiques pour engendrer les clefs de chiffrement et de déchiffrement.

Néanmoins, comme pour le cas de la famille des codes de Reed–Solomon avec l'algorithme de Sidelnikov–Shestakov, une attaque sur la clé privée est possible. Autrement dit, à partir d'une matrice génératrice d'un code ellitique, un algorithme permet de retrouver les paramètres (courbe elliptique, points d'évaluation, diviseur) qui engendre le code. Cet algorithme est dû à Minder, voir Chapitre 3 de [1].

Objectifs. Il est proposé de traiter les points suivants :

1. Présenter le système de McEliece dans le cadre de codes elliptiques.
2. Décrire l'attaque de Minder [1]. On donnera la preuve de certaines propriétés que l'on trouve primordiales.
3. Proposer une implantation de certaines sous-fonctions de l'attaque sur le système.

Références

[1] Lorenz Minder. *Cryptography based on Error Correcting Codes*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2007.

Université Paris 13 Année 2021–2022

M2 Mathématiques des données

Codes correcteurs — Sujet de projet n°6
Disponibilité pour les codes localement recouvrables

11 janvier 2022

Description du sujet. Les codes localement recouvrables permettent de retrouver tout symbole effacé de tout mot du code en ne contactant qu'un petit nombre d'autres symboles du mot. Ce type de codes a des applications naturelles en stockage distribué.

Dans le cas où ces autres symboles ne sont pas accessibles, il est souhaitable d'avoir une *seconde* manière de retrouver ce symbole en contactant peu d'autres symboles. On parle alors de code à disponibilité égale à 2. Le but de ce projet est présenter la notion de disponibilité en théorie des codes, et son application dans les systèmes de stockage distribués.

Pour cela, on pourra se référer aux articles suivants : [3, 1, 2].

Objectifs. Il est proposé de traiter les points suivants :

1. Décrire formellement le concept de disponibilité pour les codes localement recouvrables.
2. Donner des bornes sur les paramètres des codes à disponibilité avec $t = 2$ et/ou avec $t \geq 2$ quelconque. On proposera au moins une preuve.
3. Donner une construction de bon code ou de code optimal relativement à ces bornes.
4. Implanter une matrice génératrice de ce(s) code(s).

Références

[1] Ankit Singh Rawat, Dimitris S. Papailiopoulos, Alexandros G. Dimakis, and Sriram Vishwanath. Locality and Availability in Distributed Storage. *IEEE Trans. Information Theory*, 62(8) :4481–4493, 2016.

[2] Itzhak Tamo, Alexander Barg, and Alexey A. Frolov. Bounds on the Parameters of Locally Recoverable Codes. *IEEE Trans. Information Theory*, 62(6) :3070–3083, 2016.

[3] Anyu Wang and Zhifang Zhang. Repair Locality With Multiple Erasure Tolerance. *IEEE Trans. Information Theory*, 60(11) :6979–6987, 2014.

Université Paris 13 Année 2021–2022

M2 Mathématiques des données

Codes correcteurs — Sujet de projet n°7
Algorithme de Guruswami–Sudan

11 janvier 2022

Description du sujet. Un algorithme de décodage en liste a pour but de retrouver la liste de tous les mots de code à distance $\leq w$ d'un mot donné. Si w est plus grand que le rayon de décodage unique et si la liste est de petite taille, on peut donc espérer retrouver le mot émis parmi la liste des mots retournés par l'algorithme.

Sudan a proposé l'un des premiers algorithmes de décodage en liste (voir cours), pour les codes de Reed–Solomon, avec un paramètre $w/n \lesssim 1 - \sqrt{2k/n}$. Dans ce projet, on propose d'étudier l'algorithme de Guruswami–Sudan, qui permet d'atteindre $w/n \lesssim 1 - \sqrt{k/n}$. Les algorithmes de Sudan et de Guruswami–Sudan sont notamment présentés dans le Chapitre 15 de [1].

Objectifs. Il est proposé de traiter les points suivants :

1. Rappeler l'algorithme de Sudan pour le décodage en liste de codes de Reed–Solomon, et présenter sa généralisation due à Guruswami et Sudan.
2. Donner une preuve de validité de l'algorithme de Guruswami–Sudan.
3. Implanter (au moins une partie) de l'algorithme de décodage de Guruswami–Sudan.

Références

[1] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. `https://cse.buffalo.edu/faculty/atri/courses/coding-theory/book/web-coding-book.pdf`, 2019.

Université Paris 13 Année 2021–2022

M2 Mathématiques des données

Codes correcteurs — Sujet de projet n°8
Group testing

11 janvier 2022

Description du sujet. Dans le but de dépister une population, une méthode simple consiste à tester tous les individus séparément. Ceci implique un nombre de tests égal au nombre d'individus à tester, qui peut être de trop grande taille par rapport au nombre de tests disponibles.

Une autre méthode consiste à prélever tous les individus, puis à effectuer des tests par groupes : si le test d'un groupe est positif, cela signifie qu'au moins un individu du groupe l'est également, et on procède alors à des tests individuels. Si le test est négatif, tous les individus sont considérés comme négatifs.

Connaissant un estimation de la fraction de la population contaminée, on peut alors essayer de déterminer la taille et le nombre de groupes à mettre en place, afin de minimiser le nombre de tests à effectuer et le temps de retrouver les personnes positives.

Les codes correcteurs permettent essentiellement de formaliser et d'optimiser la procédure décrite ci-dessus. On réfèrera notamment par exemple à [1, 4, 2], au chapitre 19 de [3] pour des détails.

Objectifs. Il est proposé de traiter les points suivants :

1. Décrire une méthode de *group testing* qui utilise le formalisme des codes correcteurs.
2. Démontrer les propriétés de la méthodes (nombre de tests à effectuer, probabilité d'échec, etc.) en fonction des paramètres du code.
3. Implanter un exemple d'utilisation.

Références

[1] Alexander Barg and Arya Mazumdar. Group testing schemes from codes and designs. *IEEE Trans. Inf. Theory*, 63(11) :7131–7141, 2017.

[2] Arkadii G. D'yachkov, Anthony J. Macula, and Vyacheslav V. Rykov. *New Applications and Results of Superimposed Code Theory Arising from the Potentialities of Molecular Biology*, pages 265–282. Springer US, 2000.

[3] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. `https://cse.buffalo.edu/faculty/atri/courses/coding-theory/book/web-coding-book.pdf`, 2019.

[4] William H. Kautz and Richard C. Singleton. Nonrandom binary superimposed codes. *IEEE Trans. Inf. Theory*, 10(4) :363–377, 1964.

Université Paris 13 Année 2021–2022

M2 Mathématiques des données

Codes correcteurs — Sujet de projet n°9
Codes LDPC

11 janvier 2022

Description du sujet. Les codes LDPC (*Low Density Parity Check codes*) forment une famille de codes binaires particulièrement adaptés pour la transmission de messages dans des canaux bruités. L'idée de la construction est la suivante : on construit une matrice de parité du code dont chaque ligne et tirée aléatoirement et contient un petit nombre de 1.

Ensuite, pour décoder ces codes, on utilise un algorithme dit de *bit-flipping*, du à Gallager [1]. Il consiste à modifier itérativement des bits du mots erroné afin de satisfaire un maximum d'équations de parité. Le but est alors de faire converge l'algorithme vers l'état où toutes équations de parité sont satisfaites. Voir également [2].

Objectifs. Il est proposé de traiter les points suivants :

1. Définir les codes LDPC.
2. Présenter l'algorithme de *bit-flipping* de Gallager [1].
3. Implanter :
   - une fonction qui construit un code LDPC de dimension et « poids par ligne » donnés,
   - l'algorithme de décodage par *bit-flipping*.

   On pourra également proposer une série de tests, voire une analyse expérimentale de l'algorithme de décodage.

Références

[1] Robert G. Gallager. Low-density parity-check codes. *IRE Trans. Inf. Theory*, 8(1) :21–28, 1962.

[2] Tom Richardson and Rüdiger L. Urbanke. The renaissance of gallager's low-density parity-check codes. *IEEE Commun. Mag.*, 41(8) :126–131, 2003.

Université Paris 13 Année 2021–2022

M2 Mathématiques des données

Codes correcteurs — Sujet de projet n°10
Codes en métrique rang

11 janvier 2022

Description du sujet. En théorie des codes, on utilise usuellement la distance de Hamming — c'est-à-dire, le nombre de symboles distincts — pour quantifier l'erreur commise sur un mot de code. Dans certains contextes d'applications (cryptographie, transmission dans des réseaux), il s'avère intéressant de définir une autre distance entre mots, appelée métrique « rang ».

L'idée est la suivante : on assimile tout mot de code à une matrice, et la distance entre deux mots est égale au rang de la différence de ces matrices. Les paramètres des codes en métrique rang satisfont alors de nouvelles bornes, et l'on peut construire de nouvelles familles de codes optimales.

Objectifs. Il est proposé de traiter les points suivants :

1. Définir formellement les codes en métrique rang.
2. Donner et démontrer l'équivalent des bornes fondamentales sur les codes en métrique rang (par exemple borne de Singleton).
3. Donner, démontrer et implanter la famille de codes de Gabidulin [2]. Si le temps le permet, on pourra également proposer un algorithme de décodage.

On pourra également s'aider des documents suivants : un texte d'agrégation [1], les premiers chapitres du manuscrit d'HDR [3].

Références

[1] Anonyme. Texte d'agrégation externe. `https://agreg.org/Textes/public2017-C3.pdf`, 2017.

[2] Ernst Gabidulin. Theory of codes with maximum rank distance. *Problems of Information Transmission*, 21(1) :1–12, 1985.

[3] Pierre Loidreau. *Métrique rang et cryptographie*. Habilitation à diriger des recherches, Université Paris 6, 2007.