On Bounds for Batch Codes

Jens Zumbrägel

Institute of Algebra

TU Dresden

with Vitaly Skachek, University of Tartu

Algebraic **Com**binatorics and **A**pplications

ALCOMA 15 · Kloster Banz · 15 – 20 March 2015

Outline

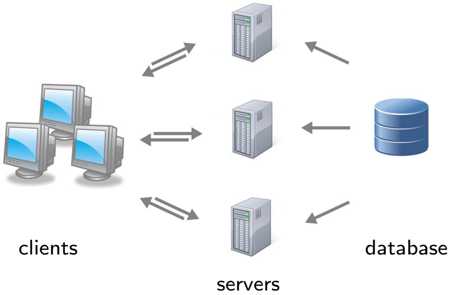
Introduction

Batch Codes, Definition and Examples

Relation to Error-Correcting Codes

Background

- Scenario: one or more clients want to receive many elements from a large database $\rightsquigarrow$ issue of **load balancing**.
- Batch codes, introduced in 2004 by Ishai et al [1], provide this
  - by dividing the database into several servers,
  - so that the client(s) need only to make few queries to each server in order to *reconstruct* all desired elements.



- These codes are of use in **distributed storage** systems [2].

Recent work

- Several works on so-called combinatorial batch codes, e.g., Stinson, Wei, Paterson [3], or Silberstein, Gál [4].
- Lipmaa and Skachek [5] recently studied linear batch codes.
  - They show that a generator matrix of a binary linear batch code is also a generator matrix of classical binary linear error-correcting code with lower-bounded minimum distance.
  - This immediately yields that coding theoretic upper bounds on the code size can be applied to binary linear batch codes.

We provide a precise *mathematical definition* of batch codes and *generalise* this result to general, nonlinear nonbinary codes.

Outline

Introduction

Batch Codes, Definition and Examples

Relation to Error-Correcting Codes

Standard batch codes

Definition

An (n, N, m, M, t) batch encoder over the alphabet F w. r. t. a partition $[N] = \bigcup_{j\in[M]} P_j$ is a map

$$\varphi : F^n \to F^N$$

such that for any $I \subseteq [n]$, $\#I = m$ there exists $T \subseteq [N]$ with

1. $\#(T \cap P_j) \leq t$ for all $j \in [M]$,
2. $\varphi(\mathbf{x})|_T$ "determines" $\mathbf{x}|_I$, i.e., there is a map $\psi : F^T \to F^I$ with $\psi(\varphi(\mathbf{x})|_T) = \mathbf{x}|_I$ for all $\mathbf{x} \in F^n$.

Interpretation:

- n size of data base, I "batch" of m queries,
- N total storage, M number of "buckets"/servers,
- t maximal load.

Batch code example

Example

An $(n, N, m, M, t) = (6, 9, 2, 3, 1)$ batch encoder is

$$\begin{aligned} \varphi : (x_1, \ldots, x_6) &\mapsto (y_1, \ldots, y_9) \\ &= (x_1, x_2, x_3,\ x_4, x_5, x_6,\ x_1+x_4, x_2+x_5, x_3+x_6)\,. \end{aligned}$$

Say, if $I = \{1, 2\}$ then take $T = \{1, 5, 8\}$. Then

1. $\#(T \cap \{1, 2, 3\}) = 1$, $\#(T \cap \{4, 5, 6\}) = 1$ and $\#(T \cap \{7, 8, 9\}) = 1$,
2. we retrieve $y_1 = x_1$ and $y_8 - y_5 = x_2 + x_5 - x_5 = x_2$.

Multi-user setup

Definition

An (n, N, m, M, t) multiset batch encoder over F w. r. t. a partition $[N] = \bigcup_{j\in[M]} P_j$, is a map

$$\varphi : F^n \to F^N$$

such that for any $\mathbf{i} : [m] \to [n]$ there exist $T_1, \ldots, T_m \subseteq [N]$ with

1. $\sum_{\ell\in[m]} \#(T_\ell \cap P_j) \leq t$ for all $j \in [M]$,
2. $\varphi(\mathbf{x})|_{T_\ell}$ "determines" x_{i_ℓ}, i.e., there is a map $\psi_\ell : F^{T_\ell} \to F$ with $\psi_\ell(\varphi(\mathbf{x})|_{T_\ell}) = x_{i_\ell}$ for all $\mathbf{x} \in F^n$, for all $\ell \in [m]$.

Remark

- Any multiset batch encoder is also a (standard) batch encoder with same parameters.
- Any batch encoder φ is injective; call $\varphi(F^n)$ the *batch code*.

Primitive batch codes

We state a single definition for multiset batch codes in the important special case where $t = 1$ and $M = N$.

Definition

An (N, n, m) primitive batch encoder over F is a map

$$\varphi : F^n \to F^N$$

such that for any $\mathbf{i} : [m] \to [n]$ there are

1. *disjoint* sets $T_1, \ldots, T_m \subseteq [N]$, such that
2. $\varphi(\mathbf{x})|_{T_\ell}$ "determines" x_{i_ℓ}, i.e., there is a map $\psi_\ell : F^{T_\ell} \to F$ with $\psi_\ell(\varphi(\mathbf{x})|_{T_\ell}) = x_{i_\ell}$ for all $\mathbf{x} \in F^n$, for all $\ell \in [m]$.

Linear batch codes

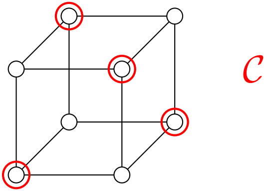
Let F now be a finite field. Then any *linear* batch encoder $\varphi : F^n \to F^N$ is specified by an $n \times N$ *generator matrix* G.

Example

The map $\varphi : \mathbb{F}_2^2 \to \mathbb{F}_2^3$, $(x_1, x_2) \mapsto (x_1, x_2, x_1 + x_2)$, i.e.,

$$\varphi(\mathbf{x}) = \mathbf{x} \cdot G\,, \quad \text{where } G = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

defines a $(3, 2, 2)$ primitive batch code ("subcube code") $\mathcal{C}$.



A criterion for generator matrices

Let $\varphi : F^n \to F^N$ be a linear map, let $I \subseteq [n]$ and $T \subseteq [N]$. If there is a map $\psi : F^T \to F^I$ with $\psi(\varphi(\mathbf{x})|_T) = \mathbf{x}|_I$ for all $\mathbf{x} \in F^n$, then ψ can be chosen to be linear.

Proposition

Let $\varphi : F^n \to F^N$, $\varphi(\mathbf{x}) = \mathbf{x} \cdot G$ be a linear encoder. Then φ is an (N, n, m) batch encoder if and only if for all $\mathbf{i} : [m] \to [n]$ there are disjoint sets $T_1, \ldots, T_m \subseteq [N]$ such that

$$\forall \ell \in [m] : \ \mathbf{e}_{i_\ell} \in \mathrm{colspan}(G|_{T_\ell}) .$$

Example

A generator matrix for a binary $(N, n, m) = (3, 2, 2)$ batch code is

$$G = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} .$$

Outline

Introduction

Batch Codes, Definition and Examples

Relation to Error-Correcting Codes

Nonlinear nonbinary batch codes

Theorem

Let $\varphi : F^n \to F^N$ be an (N, n, m) primitive batch encoder over some alphabet F. Then $\mathcal{C} = \varphi(F^n) \subseteq F^N$ is an error-correcting code of minimum distance at least m.

Proof.

Let $\mathbf{x}, \mathbf{x}' \in F^n$ with $d_H(\varphi(\mathbf{x}), \varphi(\mathbf{x}')) < m$. We show that $\mathbf{x} = \mathbf{x}'$.

- Fix $j \in [n]$.
- Let the batch $\mathbf{i} : [m] \to [n]$ be $\mathbf{i}(\ell) = i_\ell = j$ for all ℓ.
- There are disjoint sets $T_1, \ldots, T_m \subseteq [N]$ and maps $\psi_\ell : F^{T_\ell} \to F$ with $\psi_\ell(\varphi(\mathbf{x})|_{T_\ell}) = x_{i_\ell}$ for all $\ell \in [m]$.
- There must exist $\ell \in [m]$ with $\varphi(\mathbf{x})|_{T_\ell} = \varphi(\mathbf{x}')|_{T_\ell}$, hence $x_j = x_{i_\ell} = \psi_\ell(\varphi(\mathbf{x})|_{T_\ell}) = \psi_\ell(\varphi(\mathbf{x}')|_{T_\ell}) = x'_{i_\ell} = x'_j$.

Hence $\mathbf{x} = \mathbf{x}'$ as desired. $\square$

Remarks and future work

Example

A $(7,3,4)$ primitive batch code is defined by

$$\varphi(a,b,c) = (a\,, b\,, c\,, a{+}b\,, a{+}c\,, b{+}c\,, a{+}b{+}c)\,.$$

In this case the coding theory lower bound is tight.

Open problems:

- Find other lower bounds by combinatorial counting arguments.
- Shorten the gap between lower bounds and constructions of (primitive) batch codes.

References

- Y. Ishai, E. Kushilevitz, R. Ostrovsky, A. Sahai, "Batch Codes and their Applications," Proc. 36th ACM Symposium on Theory of Computing (STOC), ACM, 2004.
- A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, K. Ramchandran, "Network Coding for Distributed Storage Systems," IEEE Transactions on Information Theory, vol. 56, no. 9 (2010), pp. 4539–4551.
- D. R. Stinson, R. Wei, and M. B. Paterson, "Combinatorial batch codes," Advances in Mathematics of Communications, vol. 3 (2009), pp. 13–27.
- N. Silberstein, A. Gál, "Optimal Combinatorial Batch Codes based on Block Designs," Designs, Codes and Cryptography (2014), pp. 1–16.
- H. Lipmaa, V. Skachek, "Linear Batch Codes," Proc. 4th International Castle Meeting on Coding Theory and Applications, Palmela, Portugal, Sep 2014.
- A. S. Rawat, D. S. Papailiopoulos, A. G. Dimakis, S. Vishwanath, "Locality and Availability in Distributed Storage," Preprint, arXiv:1402.2011 (2014).
- A. G. Dimakis, A. Gal, A. S. Rawat, Z. Song, "Batch Codes through Dense Graphs without Short Cycles," Preprint, arxiv:1410.2920 (2014).