Decoding of codes for applications to steganography

Carlos Munuera,
University of Valladolid (Spain)

Wilson Olaya León
Universidad Industrial de Santander (Colombia)

cmunuera@arq.uva.es

Abstract

Error-correcting codes are introduced and widely for correcting errors when information is transmited trhough noisy channels. A $[n, k]$ *linear error-correcting code* is a k-dimensional linear subspace $\mathcal{C} \subseteq \mathbb{F}^n$. Errors are corrected by using a *decoding map*. This is a mapping $dec : \mathcal{X} \to \mathcal{C}$, where $\mathcal{C} \subset \mathcal{X} \subseteq \mathbb{F}_2^n$ and $\mathrm{dec}(\mathbf{c}) = \mathbf{c}$ for all codeword $\mathbf{c} \in \mathcal{C}$.

Different criteria have been proposed for constructing decoding maps. The most common so far is *minimum distance*. Under this condition, $\mathrm{dec}(\mathbf{x})$ is taken as one of the nearest codewords to $\mathbf{x}$, with respect to the Hamming metric. It is well known that minimum distance decoding guarantees that we can recover the right information when the number of errors is not too big. The decoding dec is *complete* if $\mathcal{X} = \mathbb{F}_2^n$. Very few complete decoding methods are known, and except rare exceptions all of them are exponential in time and/or memory complexities. However completeness is not really a major problem in coding theory. Since the main goal is to recover the word sent by the sender, in most cases it is useless to obtain a different word as a result of our decoding, even being this word closest to the received vector. This is just the case when the nearest codeword is not unique. For this reason most efforts of coding-theorist have turned to find efficient bounded minimum distance decoding methods.

In recent times, new appplications of coding theory have been found. In this presentation we are interested in steganography. Roughly speaking, the purpose of a steganographic system is to hide as much secret information as possible in a innocuous-like cover object (like a digital image), making as few changes as posible in the cover, to reduce the chance of being detected by third parties. This is done by using error-correcting codes and decoding maps.

In this new scenario, the classical choice of coding theorists –to dispense with the condition of complete decoding– is no longer valid. Indeed, if cannot decode then we cannot embed information, and our stegosysmen does not work. Remark also that for error-correction purposes, errors of low weight are more probable, while for steganographic purposes, all vectors in $\mathbb{F}_2^n$ are equally probable as covers.

Therefore, it seems appropriate now to consider new decoding algorithms, by relaxing the condition of minimum distance. In this talk we present the first steps in this study.

Let $\mathcal{C}$ be a linear $[n, k]$ code with distance d and systematic parity-check matrix $\mathbf{H} = (\mathbf{H}'|\mathbf{I}_{n-k})$. Let $\mathbf{h}_1, \ldots, \mathbf{h}_n$ be the columns of $\mathbf{H}$ and let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the canonical basis of $\mathbb{F}_2^n$. The syndrome of $\mathbf{x} \in \mathbb{F}_2^n$ can be used to give an estimate of $d(\mathbf{x}, \mathcal{C})$. This leads to the following algorithm. Given a vector $\mathbf{x} \in \mathbb{F}_2^n$,

Input: *The vector* $\mathbf{x}$ *to be decoded, the matrix* $\mathbf{H}$ *in systematic form.*
0. [Initialization] $\mathbf{dec} \leftarrow \mathbf{x}$
1. [Iteration] Repeat until $\mathbf{s}(\mathbf{dec}) = \mathbf{0}$*:*
find a coordinate i *such that* $\mathrm{wt}(\mathbf{s}(\mathbf{dec}) + \mathbf{h}_i)$ *is minium among all columns of* $\mathbf{H}$
set $\mathbf{dec} \leftarrow \mathbf{dec} + \mathbf{e}_i$
recompute the syndrome $\mathbf{s}(\mathbf{dec})$
2. [Output] $\mathbf{dec}$

requires at most $\mathrm{wt}(\mathbf{s}(\mathbf{x})) \leq n - k$ iterations and provides a decoding $\mathrm{dec}(\mathbf{x}) = \mathbf{dec}$ of $\mathbf{x}$.

Some properties of this decoding algorithm are the following.
(a) $d(\mathbf{x}, \mathrm{dec}(\mathbf{x})) \leq \mathrm{wt}(s(\mathbf{x}))$.
(b) is $\mathrm{wt}(s(\mathbf{x})) \leq d/2$ then $\mathrm{dec}(\mathbf{x})$ is the nearest codeword to $\mathbf{x}$ in $\mathcal{C}$.
(c) $\rho(\mathrm{dec}) \leq n - k$.
(d) $\tilde{\rho}(\mathrm{dec}) \leq (n - k)/2$.

Keywords
Error-correcting code, decoding, steganography

References

[1] A. Barg, Complexity issues in coding theory, in Handbook of Coding Theory, vol. 1. Edited by V. Pless, W. Huffman, R. Brualdi. North-Holland, 1998.

[2] C. Munuera, Steganography from a coding theory point of view, in Algebraic geometry modeling in information theory. Edited by E. Martinez-Moro. World Scientific, 2012.