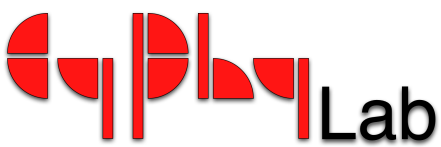
Secure state-estimation for dynamical systems under active adversaries

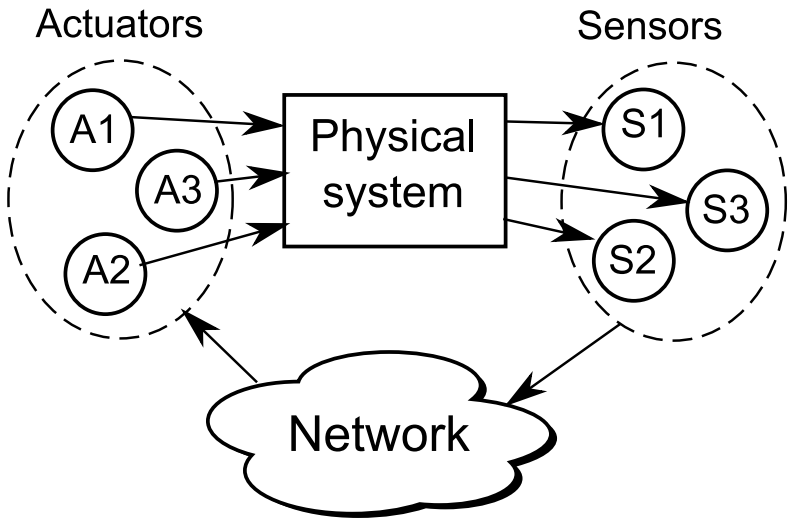
Hamza Fawzi

Joint work with Paulo Tabuada and Suhas Diggavi



Why security for control systems?

- Control systems are *physical* processes (chemical plants, power grid, mechanical system, etc.)



- Control systems becoming larger (large sensor networks) and increasingly open to the *cyber*-world (e.g., internet) $\Rightarrow$ **increased vulnerability to attacks**
- Examples of real attacks: Sewage control system (Queensland, Australia, 2000), Natural gas pipelines (Russia, 2000), Stuxnet (2010), ...
- **Need efficient ways to detect attacks on control systems...**

For more info on security for control systems see *[Cardenas, Amin, Sastry, 2008]*

Security for control systems

- (Some of the) existing works on adversarial, malicious attacks:
  - Optimal control in the presence of intelligent jammer (cf. Gupta, Langbort and Basar, 2010)
    - game-theoretic approach; attacker's objective is to maximize cost function
  - Secure state-estimation for power network against malicious attacks (cf. Pasqualetti, Dorfler, Bullo (2011))
    - attack-detection filter is proposed, but computationally expensive (combinatorial, test all possible attack sets)
- **This talk: efficient algorithm to estimate the state of a linear dynamical system when sensors are attacked**

The setup

- Physical process modeled as a linear dynamical system

$$x^{(t+1)} = Ax^{(t)}$$

The setup

- Physical process modeled as a linear dynamical system

$$x^{(t+1)} = Ax^{(t)}$$

- A total of p sensors monitor state of plant: $(y^{(t)} \in \mathbf{R}^p)$

$$y^{(t)} = Cx^{(t)}$$

The setup

- Physical process modeled as a linear dynamical system

$$x^{(t+1)} = Ax^{(t)}$$

- A total of p sensors monitor state of plant: $(y^{(t)} \in \mathbf{R}^p)$

$$y^{(t)} = Cx^{(t)} + \underbrace{e^{(t)}}_{\substack{\textit{attack} \\ \textit{vector}}}$$

- Some sensors are attacked
  - $e_i^{(t)} \neq 0 \longrightarrow$ sensor i is attacked at time t

The setup

- Physical process modeled as a linear dynamical system

$$x^{(t+1)} = Ax^{(t)}$$

- A total of p sensors monitor state of plant: $(y^{(t)} \in \mathbf{R}^p)$

$$y^{(t)} = Cx^{(t)} + \underbrace{e^{(t)}}_{\substack{\textit{attack} \\ \textit{vector}}}$$

- Some sensors are attacked
  - $e_i^{(t)} \neq 0 \longrightarrow$ sensor i is attacked at time t
  - If sensor i is attacked, $e_i^{(t)}$ can be **arbitrary** (no boundedness assumption, no stochastic model, etc.)

The setup

- Physical process modeled as a linear dynamical system

$$x^{(t+1)} = Ax^{(t)}$$

- A total of p sensors monitor state of plant: $(y^{(t)} \in \mathbf{R}^p)$

$$y^{(t)} = Cx^{(t)} + \underbrace{e^{(t)}}_{\substack{\textit{attack} \\ \textit{vector}}}$$

- Some sensors are attacked
  - $e_i^{(t)} \neq 0 \longrightarrow$ sensor i is attacked at time t
  - If sensor i is attacked, $e_i^{(t)}$ can be **arbitrary** (no boundedness assumption, no stochastic model, etc.)
- Set of attacked sensors (**unknown**) is denoted by $K \subset \{1, \ldots, p\}$:

$$\text{support}(e^{(t)}) = K \quad \forall t = 0, 1, \ldots$$

The setup

- Physical process modeled as a linear dynamical system

$$x^{(t+1)} = Ax^{(t)}$$

- A total of p sensors monitor state of plant: $(y^{(t)} \in \mathbf{R}^p)$

$$y^{(t)} = Cx^{(t)} + \underbrace{e^{(t)}}_{\substack{\textit{attack} \\ \textit{vector}}}$$

- Some sensors are attacked
  - $e_i^{(t)} \neq 0 \longrightarrow$ sensor i is attacked at time t
  - If sensor i is attacked, $e_i^{(t)}$ can be **arbitrary** (no boundedness assumption, no stochastic model, etc.)
- Set of attacked sensors (**unknown**) is denoted by $K \subset \{1, \ldots, p\}$:

$$\text{support}(e^{(t)}) = K \quad \forall t = 0, 1, \ldots$$

- Number of attacked sensors will be denoted by q: $|K| = q$

The setup

- Physical process modeled as a linear dynamical system

$$x^{(t+1)} = Ax^{(t)}$$

- A total of p sensors monitor state of plant: $(y^{(t)} \in \mathbf{R}^p)$

$$y^{(t)} = Cx^{(t)} + \underbrace{e^{(t)}}_{\substack{\textit{attack} \\ \textit{vector}}}$$

- Some sensors are attacked
  - $e_i^{(t)} \neq 0 \longrightarrow$ sensor i is attacked at time t
  - If sensor i is attacked, $e_i^{(t)}$ can be **arbitrary** (no boundedness assumption, no stochastic model, etc.)
- Set of attacked sensors (**unknown**) is denoted by $K \subset \{1, \ldots, p\}$:

$$\text{support}(e^{(t)}) = K \quad \forall t = 0, 1, \ldots$$

- Number of attacked sensors will be denoted by q: $|K| = q$
- **Objective: Given observations $y^{(0)}, \ldots, y^{(T-1)}$: recover state $x^{(0)}$ of physical plant from observations (attack set K is unknown)**

Error correction

$$x^{(t+1)} = Ax^{(t)}$$
$$y^{(t)} = Cx^{(t)} + e^{(t)}$$

- A **decoder** D_T takes observations $y^{(0)}, \ldots, y^{(T-1)}$ and produces an estimate of the initial state $x^{(0)}$

Error correction

$$x^{(t+1)} = Ax^{(t)}$$
$$y^{(t)} = Cx^{(t)} + e^{(t)}$$

- A **decoder** D_T takes observations $y^{(0)}, \ldots, y^{(T-1)}$ and produces an estimate of the initial state $x^{(0)}$

- We say that a decoder $D_T : (\mathbf{R}^p)^T \to \mathbf{R}^n$ **corrects q errors** if it is resilient against any attack of q sensors, i.e., if for any initial condition $x^{(0)} \in \mathbf{R}^n$, and for any attack vectors $e^{(0)}, \ldots, e^{(T-1)}$ corresponding to q attacked sensors, we have

$$D_T(y^{(0)}, \ldots, y^{(T-1)}) = x^{(0)}.$$

Error correction

$$x^{(t+1)} = Ax^{(t)}$$
$$y^{(t)} = Cx^{(t)} + e^{(t)}$$

- A **decoder** D_T takes observations $y^{(0)}, \ldots, y^{(T-1)}$ and produces an estimate of the initial state $x^{(0)}$
- We say that a decoder $D_T : (\mathbf{R}^p)^T \to \mathbf{R}^n$ **corrects q errors** if it is resilient against any attack of q sensors, i.e., if for any initial condition $x^{(0)} \in \mathbf{R}^n$, and for any attack vectors $e^{(0)}, \ldots, e^{(T-1)}$ corresponding to q attacked sensors, we have

$$D_T(y^{(0)}, \ldots, y^{(T-1)}) = x^{(0)}.$$

- We say that q **errors are correctable** after T steps (for the system (A, C)) if there exists a decoder that can correct q errors

Error correction

$$x^{(t+1)} = Ax^{(t)}$$
$$y^{(t)} = Cx^{(t)} + e^{(t)}$$

- A **decoder** D_T takes observations $y^{(0)}, \ldots, y^{(T-1)}$ and produces an estimate of the initial state $x^{(0)}$
- We say that a decoder $D_T : (\mathbf{R}^p)^T \to \mathbf{R}^n$ **corrects q errors** if it is resilient against any attack of q sensors, i.e., if for any initial condition $x^{(0)} \in \mathbf{R}^n$, and for any attack vectors $e^{(0)}, \ldots, e^{(T-1)}$ corresponding to q attacked sensors, we have

$$D_T(y^{(0)}, \ldots, y^{(T-1)}) = x^{(0)}.$$

- We say that q **errors are correctable** after T steps (for the system (A, C)) if there exists a decoder that can correct q errors
- Note:

  can correct $q = 0$ errors $\equiv$ can recover $x^{(0)}$ from $(Cx^{(0)}, \ldots, CA^{T-1}x^{(0)}) \equiv (A, C)$ observable

Necessary and sufficient condition for correction of q errors

Proposition

Let $T > 0$ be fixed. Then q errors are correctable after T steps iff

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q \tag{1}$$

Necessary and sufficient condition for correction of q errors

Proposition

Let $T > 0$ be fixed. Then q errors are correctable after T steps iff

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q \tag{1}$$

- Interpretation of condition (1): $C, CA, \ldots, CA^{T-1}$ have to *spread* the components of the state x.

Necessary and sufficient condition for correction of q errors

Proposition

Let $T > 0$ be fixed. Then q errors are correctable after T steps iff

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q \tag{1}$$

- Interpretation of condition (1): $C, CA, \ldots, CA^{T-1}$ have to *spread* the components of the state x.
  - Example of a good pair (A, C):

$$A = \begin{bmatrix} 010 \\ 001 \\ 100 \end{bmatrix} \text{ (circular permutation)}, \quad C = \text{identity}$$

Necessary and sufficient condition for correction of q errors

Proposition

Let $T > 0$ be fixed. Then q errors are correctable after T steps iff

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q \tag{1}$$

- Interpretation of condition (1): $C, CA, \ldots, CA^{T-1}$ have to *spread* the components of the state x.
  - Example of a good pair (A, C):

$$A = \begin{bmatrix} 010 \\ 001 \\ 100 \end{bmatrix} \text{ (circular permutation)}, \quad C = \text{identity}$$

$$\text{For } x = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow Ax = \begin{bmatrix} 0 \\ x_1 \\ 0 \end{bmatrix}, \quad A^2x = \begin{bmatrix} 0 \\ 0 \\ x_1 \end{bmatrix}$$

Necessary and sufficient condition for correction of q errors

Proposition

Let $T > 0$ be fixed. Then q errors are correctable after T steps iff

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q \tag{1}$$

- Interpretation of condition (1): $C, CA, \ldots, CA^{T-1}$ have to *spread* the components of the state x.
  - Example of a good pair (A, C):

$$A = \begin{bmatrix} 010 \\ 001 \\ 100 \end{bmatrix} \text{ (circular permutation)}, \quad C = \text{identity}$$

$$\text{For } x = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow Ax = \begin{bmatrix} 0 \\ x_1 \\ 0 \end{bmatrix}, \quad A^2x = \begin{bmatrix} 0 \\ 0 \\ x_1 \end{bmatrix}$$

$$|\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \mathrm{supp}(CA^2x)| = |\mathrm{supp}(\begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix}) \cup \mathrm{supp}(\begin{bmatrix} 0 \\ x_1 \\ 0 \end{bmatrix}) \cup \mathrm{supp}(\begin{bmatrix} 0 \\ 0 \\ x_1 \end{bmatrix})| = 3 \text{ (maximal)}$$

Necessary and sufficient condition for correction of q errors

Proposition

Let $T > 0$ be fixed. Then q errors are correctable after T steps iff

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q \tag{1}$$

- Interpretation of condition (1): $C, CA, \ldots, CA^{T-1}$ have to *spread* the components of the state x.
  - Example of a good pair (A, C):

$$A = \begin{bmatrix} 010 \\ 001 \\ 100 \end{bmatrix} \text{ (circular permutation)}, \quad C = \text{identity}$$

$$\text{For } x = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow Ax = \begin{bmatrix} 0 \\ x_1 \\ 0 \end{bmatrix}, \quad A^2x = \begin{bmatrix} 0 \\ 0 \\ x_1 \end{bmatrix}$$

$$|\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \mathrm{supp}(CA^2x)| = |\mathrm{supp}(\begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix}) \cup \mathrm{supp}(\begin{bmatrix} 0 \\ x_1 \\ 0 \end{bmatrix}) \cup \mathrm{supp}(\begin{bmatrix} 0 \\ 0 \\ x_1 \end{bmatrix})| = 3 \text{ (maximal)}$$

  - Example of a very bad pair (A, C): $A =$ identity, $C =$ identity (easy to see that even $q = 1$ does not satisfy condition above: take x to be supported on one component)

Why "spread" the components?

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q$$

- If, for some $x \neq 0$

$$Cx = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}, \quad CAx = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix}, \quad CA^2x = \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix}$$

Why "spread" the components?

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q$$

- If, for some $x \neq 0$

$$Cx = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}, \quad CAx = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix}, \quad CA^2x = \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix}$$

then if only sensor 1 is attacked and attacker chooses

$$e^{(0)} = \begin{bmatrix} -a \\ 0 \\ 0 \end{bmatrix}, \quad e^{(1)} = \begin{bmatrix} -b \\ 0 \\ 0 \end{bmatrix}, \quad e^{(2)} = \begin{bmatrix} -c \\ 0 \\ 0 \end{bmatrix}$$

Why "spread" the components?

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q$$

- If, for some $x \neq 0$

$$Cx = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}, \quad CAx = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix}, \quad CA^2x = \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix}$$

then if only sensor 1 is attacked and attacker chooses

$$e^{(0)} = \begin{bmatrix} -a \\ 0 \\ 0 \end{bmatrix}, \quad e^{(1)} = \begin{bmatrix} -b \\ 0 \\ 0 \end{bmatrix}, \quad e^{(2)} = \begin{bmatrix} -c \\ 0 \\ 0 \end{bmatrix}$$

then we observe

$$y^{(0)} = 0 \qquad , y^{(1)} = 0, \qquad y^{(2)} = 0$$

Why "spread" the components?

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q$$

- If, for some $x \neq 0$

$$Cx = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}, \quad CAx = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix}, \quad CA^2x = \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix}$$

then if only sensor 1 is attacked and attacker chooses

$$e^{(0)} = \begin{bmatrix} -a \\ 0 \\ 0 \end{bmatrix}, \quad e^{(1)} = \begin{bmatrix} -b \\ 0 \\ 0 \end{bmatrix}, \quad e^{(2)} = \begin{bmatrix} -c \\ 0 \\ 0 \end{bmatrix}$$

then we observe

$$y^{(0)} = 0 \qquad , y^{(1)} = 0, \qquad y^{(2)} = 0$$

- $\Rightarrow$ **We cannot know, by simply looking at the observations:**
  - **the true initial state was 0 and no sensor was attacked; or**
  - **the true initial state was x and sensor 1 was attacked**

Why "spread" the components?

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q$$

- If, for some $x \neq 0$

$$Cx = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}, \quad CAx = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix}, \quad CA^2x = \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix}$$

then if only sensor 1 is attacked and attacker chooses

$$e^{(0)} = \begin{bmatrix} -a \\ 0 \\ 0 \end{bmatrix}, \quad e^{(1)} = \begin{bmatrix} -b \\ 0 \\ 0 \end{bmatrix}, \quad e^{(2)} = \begin{bmatrix} -c \\ 0 \\ 0 \end{bmatrix}$$

then we observe

$$y^{(0)} = 0 \qquad , y^{(1)} = 0, \qquad y^{(2)} = 0$$

- $\Rightarrow$ **We cannot know, by simply looking at the observations:**
  - **the true initial state was 0 and no sensor was attacked; or**
  - **the true initial state was x and sensor 1 was attacked**
- $\Rightarrow$ $q = 1$ **error is NOT correctable in this case**

Comments

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q$$

- Number of correctable errors does not increase beyond $T = n$ steps (Cayley-Hamilton theorem)
- No more than $p/2$ errors can be corrected (q is necessarily $< p/2$)

Comments

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q$$

- Number of correctable errors does not increase beyond $T = n$ steps (Cayley-Hamilton theorem)
- No more than $p/2$ errors can be corrected (q is necessarily $< p/2$)

Proposition

For *almost all* systems (A, C), the number of correctable errors is maximal (equal to $\lceil p/2 - 1 \rceil$).

Comments

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q$$

- Number of correctable errors does not increase beyond $T = n$ steps (Cayley-Hamilton theorem)
- No more than $p/2$ errors can be corrected (q is necessarily $< p/2$)

Proposition

For *almost all* systems (A, C), the number of correctable errors is maximal (equal to $\lceil p/2 - 1 \rceil$).

- Given a specific system (A, C), what is the number of correctable errors q?
  - Unfortunately, this is still unsolved and is likely to be hard:

Comments

$$\forall x \neq 0,\ |\mathrm{supp}(Cx) \cup \mathrm{supp}(CAx) \cup \cdots \cup \mathrm{supp}(CA^{T-1}x)| > 2q$$

- Number of correctable errors does not increase beyond $T = n$ steps (Cayley-Hamilton theorem)
- No more than $p/2$ errors can be corrected (q is necessarily $< p/2$)

Proposition

For *almost all* systems (A, C), the number of correctable errors is maximal (equal to $\lceil p/2 - 1 \rceil$).

- Given a specific system (A, C), what is the number of correctable errors q?
  - Unfortunately, this is still unsolved and is likely to be hard: cf. computation of *sparsest nonzero element of a subspace* L (when $L = \ker A$, this is the *spark* of A)

An explicit decoder

- We received observations $y^{(0)}, \ldots, y^{(T-1)}$.
  Objective: Find $x^{(0)}$ (state) and K (set of attacked sensors) that generated these observations.

An explicit decoder

- We received observations $y^{(0)}, \ldots, y^{(T-1)}$.
  Objective: Find $x^{(0)}$ (state) and K (set of attacked sensors) that generated these observations.
- Consider looking for the smallest possible attack set $\hat{K}$ that is *consistent* with observations $y^{(0)}, \ldots, y^{(T-1)}$.

An explicit decoder

- We received observations $y^{(0)}, \ldots, y^{(T-1)}$.
  Objective: Find $x^{(0)}$ (state) and K (set of attacked sensors) that generated these observations.
- Consider looking for the smallest possible attack set $\hat{K}$ that is *consistent* with observations $y^{(0)}, \ldots, y^{(T-1)}$.
  - $\hat{K} \subseteq \{1, \ldots, p\}$ is *consistent* with observations $y^{(0)}, \ldots, y^{(T-1)}$ if we can write
    $$y^{(t)} = CA^t \hat{x}^{(0)} + \hat{e}^{(t)}, \quad t = 0, \ldots, T-1$$
    for some $\hat{x}^{(0)}$ and attack vectors $\hat{e}^{(0)}, \ldots, \hat{e}^{(T-1)}$ supported on $\hat{K}$.

An explicit decoder

- We received observations $y^{(0)}, \ldots, y^{(T-1)}$.
  Objective: Find $x^{(0)}$ (state) and K (set of attacked sensors) that generated these observations.
- Consider looking for the smallest possible attack set $\hat{K}$ that is *consistent* with observations $y^{(0)}, \ldots, y^{(T-1)}$.
  - $\hat{K} \subseteq \{1, \ldots, p\}$ is *consistent* with observations $y^{(0)}, \ldots, y^{(T-1)}$ if we can write
    $$y^{(t)} = CA^t \hat{x}^{(0)} + \hat{e}^{(t)}, \quad t = 0, \ldots, T-1$$
    for some $\hat{x}^{(0)}$ and attack vectors $\hat{e}^{(0)}, \ldots, \hat{e}^{(T-1)}$ supported on $\hat{K}$.
  - Decoder:

$$\begin{array}{ll} \underset{\hat{x}^{(0)}, \hat{K}}{\text{minimize}} & |\hat{K}| \\ \text{subject to} & y^{(t)} = CA^t \hat{x}^{(0)} + \hat{e}^{(t)} \quad , t = 0, \ldots, T-1 \\ & \text{supp}(e^{(t)}) \subseteq \hat{K} \end{array} \tag{2}$$

An explicit decoder

- We received observations $y^{(0)}, \ldots, y^{(T-1)}$.
  Objective: Find $x^{(0)}$ (state) and K (set of attacked sensors) that generated these observations.
- Consider looking for the smallest possible attack set $\hat{K}$ that is *consistent* with observations $y^{(0)}, \ldots, y^{(T-1)}$.
  - $\hat{K} \subseteq \{1, \ldots, p\}$ is *consistent* with observations $y^{(0)}, \ldots, y^{(T-1)}$ if we can write
    $$y^{(t)} = CA^t \hat{x}^{(0)} + \hat{e}^{(t)}, \quad t = 0, \ldots, T-1$$
    for some $\hat{x}^{(0)}$ and attack vectors $\hat{e}^{(0)}, \ldots, \hat{e}^{(T-1)}$ supported on $\hat{K}$.
  - Decoder:

$$\begin{array}{ll} \underset{\hat{x}^{(0)}, \hat{K}}{\text{minimize}} & |\hat{K}| \\ \text{subject to} & y^{(t)} = CA^t \hat{x}^{(0)} + \hat{e}^{(t)} \quad , t = 0, \ldots, T-1 \\ & \text{supp}(e^{(t)}) \subseteq \hat{K} \end{array} \tag{2}$$

Proposition

If q errors can be corrected by *some* decoder, then the *above decoder* can correct q errors.

An explicit decoder

- We received observations $y^{(0)}, \ldots, y^{(T-1)}$.
  Objective: Find $x^{(0)}$ (state) and K (set of attacked sensors) that generated these observations.
- Consider looking for the smallest possible attack set $\hat{K}$ that is *consistent* with observations $y^{(0)}, \ldots, y^{(T-1)}$.
  - $\hat{K} \subseteq \{1, \ldots, p\}$ is *consistent* with observations $y^{(0)}, \ldots, y^{(T-1)}$ if we can write
    $$y^{(t)} = CA^t \hat{x}^{(0)} + \hat{e}^{(t)}, \quad t = 0, \ldots, T-1$$
    for some $\hat{x}^{(0)}$ and attack vectors $\hat{e}^{(0)}, \ldots, \hat{e}^{(T-1)}$ supported on $\hat{K}$.
  - Decoder:

$$\begin{array}{ll} \underset{\hat{x}^{(0)}, \hat{K}}{\text{minimize}} & |\hat{K}| \\ \text{subject to} & y^{(t)} = CA^t \hat{x}^{(0)} + \hat{e}^{(t)} \quad , t = 0, \ldots, T-1 \\ & \text{supp}(e^{(t)}) \subseteq \hat{K} \end{array} \tag{2}$$

Proposition

If q errors can be corrected by *some* decoder, then the *above decoder* can correct q errors.

- Interpretation: The above decoder is, in some sense, *unbeatable*...

An explicit decoder

- We received observations $y^{(0)}, \ldots, y^{(T-1)}$.
  Objective: Find $x^{(0)}$ (state) and K (set of attacked sensors) that generated these observations.
- Consider looking for the smallest possible attack set $\hat{K}$ that is *consistent* with observations $y^{(0)}, \ldots, y^{(T-1)}$.
  - $\hat{K} \subseteq \{1, \ldots, p\}$ is *consistent* with observations $y^{(0)}, \ldots, y^{(T-1)}$ if we can write
    $$y^{(t)} = CA^t \hat{x}^{(0)} + \hat{e}^{(t)}, \quad t = 0, \ldots, T-1$$
    for some $\hat{x}^{(0)}$ and attack vectors $\hat{e}^{(0)}, \ldots, \hat{e}^{(T-1)}$ supported on $\hat{K}$.
  - Decoder:

$$\begin{array}{ll} \underset{\hat{x}^{(0)}, \hat{K}}{\text{minimize}} & |\hat{K}| \\ \text{subject to} & y^{(t)} = CA^t \hat{x}^{(0)} + \hat{e}^{(t)} \quad , t = 0, \ldots, T-1 \\ & \text{supp}(e^{(t)}) \subseteq \hat{K} \end{array} \tag{2}$$

Proposition

If q errors can be corrected by *some* decoder, then the *above decoder* can correct q errors.

- Interpretation: The above decoder is, in some sense, *unbeatable*...
- One little problem: It is NP-hard... :-(

An explicit decoder

- Idea: Relax the previous decoder to make it computationally tractable
  - $\rightarrow$ use ℓ_1-relaxation techniques from compressed sensing and error correction over the reals

An explicit decoder

- Idea: Relax the previous decoder to make it computationally tractable
  - $\rightarrow$ use ℓ_1-relaxation techniques from compressed sensing and error correction over the reals
- Some notations first:
  - Collect observations from $t = 0$ to $t = T - 1$ in a $p \times T$ matrix:

$$\underbrace{\left[\begin{array}{c|c|c} y^{(0)} & \dots & y^{(T-1)} \end{array}\right]}_{Y^{(T)} \in \mathbf{R}^{p \times T}} = \underbrace{\left[\begin{array}{c|c|c} Cx & \dots & CA^{T-1}x \end{array}\right]}_{\Phi^{(T)}x} + \underbrace{\left[\begin{array}{c|c|c} e^{(0)} & \dots & e^{(T-1)} \end{array}\right]}_{E^{(T)} \in \mathbf{R}^{p \times T}}$$

An explicit decoder

- Idea: Relax the previous decoder to make it computationally tractable
  - $\rightarrow$ use ℓ_1-relaxation techniques from compressed sensing and error correction over the reals
- Some notations first:
  - Collect observations from $t = 0$ to $t = T - 1$ in a $p \times T$ matrix:

$$\underbrace{\left[\begin{array}{c|c|c} y^{(0)} & \dots & y^{(T-1)} \end{array}\right]}_{Y^{(T)} \in \mathbf{R}^{p \times T}} = \underbrace{\left[\begin{array}{c|c|c} Cx & \dots & CA^{T-1}x \end{array}\right]}_{\Phi^{(T)} x} + \underbrace{\left[\begin{array}{c|c|c} e^{(0)} & \dots & e^{(T-1)} \end{array}\right]}_{E^{(T)} \in \mathbf{R}^{p \times T}}$$

- Define ℓ_0 norm of $E^{(T)}$ as the number of nonzero rows of $E^{(T)}$ ($=$ number of attacked sensors):

$$\|E^{(T)}\|_{\ell_0} = |\text{rowsupport}(E^{(T)})|$$

An explicit decoder

- Idea: Relax the previous decoder to make it computationally tractable
  - $\rightarrow$ use ℓ_1-relaxation techniques from compressed sensing and error correction over the reals
- Some notations first:
  - Collect observations from $t = 0$ to $t = T - 1$ in a $p \times T$ matrix:

$$\underbrace{\left[\begin{array}{c|c|c} y^{(0)} & \dots & y^{(T-1)} \end{array}\right]}_{Y^{(T)} \in \mathbf{R}^{p \times T}} = \underbrace{\left[\begin{array}{c|c|c} Cx & \dots & CA^{T-1}x \end{array}\right]}_{\Phi^{(T)} x} + \underbrace{\left[\begin{array}{c|c|c} e^{(0)} & \dots & e^{(T-1)} \end{array}\right]}_{E^{(T)} \in \mathbf{R}^{p \times T}}$$

- Define ℓ_0 norm of $E^{(T)}$ as the number of nonzero rows of $E^{(T)}$ (= number of attacked sensors):

$$\|E^{(T)}\|_{\ell_0} = |\text{rowsupport}(E^{(T)})|$$

- Let's rewrite the previous "*unbeatable*" decoder using these notations:
  - smallest number of attacked sensors that explain the received observations:

$$\underset{x}{\text{minimize}} \; \| \underbrace{Y^{(T)} - \Phi^{(T)} x}_{E^{(T)}} \|_{\ell_0}$$

The "ℓ_1" relaxation

- ℓ_0 decoder: NP-hard

$$\underset{x}{\text{minimize}} \; \| Y^{(T)} - \Phi^{(T)} x \|_{\ell_0}$$

- Relaxation idea: Instead of "ℓ_0 norm" (intractable), use ℓ_1 norm (convex program, tractable)

The "ℓ_1" relaxation

- ℓ_0 decoder: NP-hard

$$\underset{x}{\text{minimize}} \; \| Y^{(T)} - \Phi^{(T)} x \|_{\ell_0}$$

- Relaxation idea: Instead of "ℓ_0 norm" (intractable), use ℓ_1 norm (convex program, tractable)
  - i.e., replace number of nonzero rows of $E^{(T)}$, by $\underbrace{\text{sum}}_{\ell_1}$ of the $\underbrace{\textbf{magnitudes}}_{\ell_r}$ of the rows of $E^{(T)}$

The "ℓ_1" relaxation

- ℓ_0 decoder: NP-hard

$$\underset{x}{\text{minimize}} \; \| Y^{(T)} - \Phi^{(T)} x \|_{\ell_0}$$

- Relaxation idea: Instead of "ℓ_0 norm" (intractable), use ℓ_1 norm (convex program, tractable)
  - i.e., replace number of nonzero rows of $E^{(T)}$, by $\underbrace{\text{sum}}_{\ell_1}$ of the $\underbrace{\textbf{magnitudes}}_{\ell_r}$ of the rows of $E^{(T)}$

$$\ell_1/\ell_r \text{ decoder: } \underset{x}{\text{minimize}} \; \| \underbrace{Y^{(T)} - \Phi^{(T)} x}_{\in \mathbf{R}^{p \times T}} \|_{\ell_1/\ell_r} = \sum_{i=1}^{p} \| \underbrace{\left(Y^{(T)} - \Phi^{(T)} x \right)_i}_{\in \mathbf{R}^T} \|_{\ell_r}$$

The "ℓ_1" relaxation

- ℓ_0 decoder: NP-hard

$$\underset{x}{\text{minimize}} \; \| Y^{(T)} - \Phi^{(T)} x \|_{\ell_0}$$

- Relaxation idea: Instead of "ℓ_0 norm" (intractable), use ℓ_1 norm (convex program, tractable)
  - i.e., replace number of nonzero rows of $E^{(T)}$, by $\underbrace{\text{sum}}_{\ell_1}$ of the $\underbrace{\textbf{magnitudes}}_{\ell_r}$ of the rows of $E^{(T)}$

$$\ell_1/\ell_r \text{ decoder: } \underset{x}{\text{minimize}} \; \| \underbrace{Y^{(T)} - \Phi^{(T)} x}_{\in \mathbf{R}^{p \times T}} \|_{\ell_1/\ell_r} = \sum_{i=1}^{p} \| \underbrace{\left(Y^{(T)} - \Phi^{(T)} x\right)_i}_{\in \mathbf{R}^T} \|_{\ell_r}$$

- **Magnitude** of a row of $E^{(T)}$ measured by its ℓ_r norm (in $\mathbf{R}^T$), for any $r \geq 1$.

The "ℓ_1" relaxation

- ℓ_0 decoder: NP-hard

$$\underset{x}{\text{minimize}} \; \| Y^{(T)} - \Phi^{(T)} x \|_{\ell_0}$$

- Relaxation idea: Instead of "ℓ_0 norm" (intractable), use ℓ_1 norm (convex program, tractable)
  - i.e., replace number of nonzero rows of $E^{(T)}$, by $\underbrace{\text{sum}}_{\ell_1}$ of the $\underbrace{\textbf{magnitudes}}_{\ell_r}$ of the rows of $E^{(T)}$

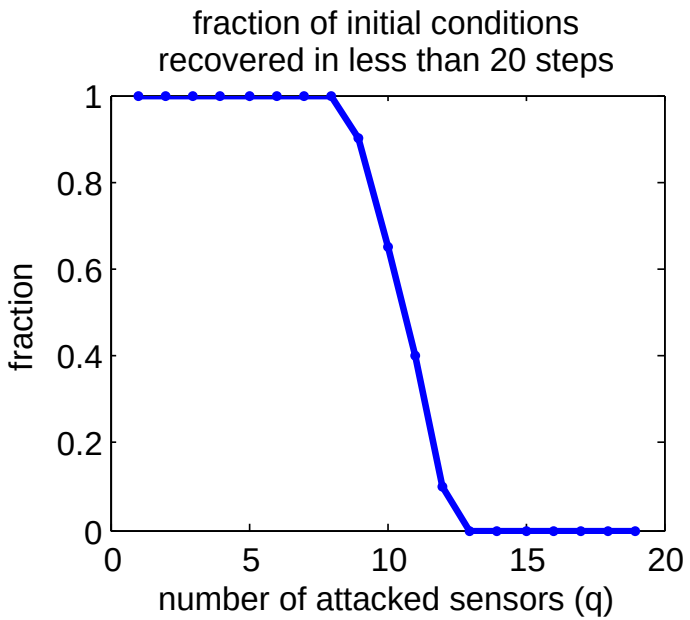
$$\ell_1/\ell_r \text{ decoder: } \underset{x}{\text{minimize}} \; \| \underbrace{Y^{(T)} - \Phi^{(T)} x}_{\in \mathbf{R}^{p \times T}} \|_{\ell_1/\ell_r} = \sum_{i=1}^{p} \| \underbrace{(Y^{(T)} - \Phi^{(T)} x)_i}_{\in \mathbf{R}^T} \|_{\ell_r}$$

- **Magnitude** of a row of $E^{(T)}$ measured by its ℓ_r norm (in $\mathbf{R}^T$), for any $r \geq 1$.

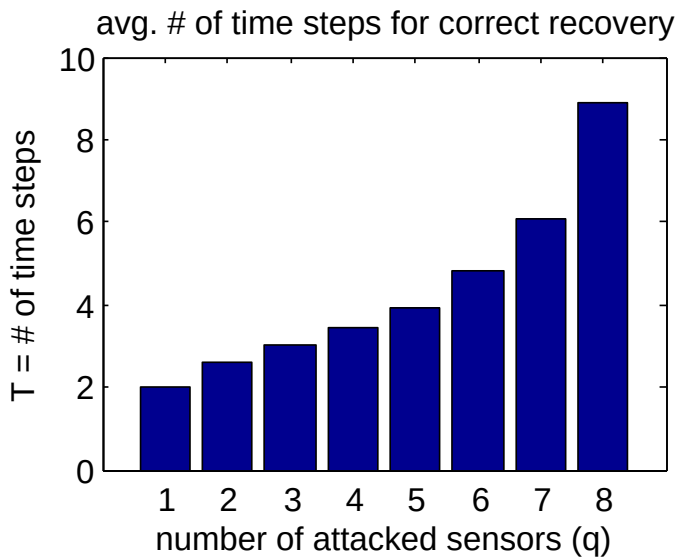
- $\ell_0 \to \ell_1$ relaxation idea used in compressed sensing (recovery of sparse signals), and error correction over the reals (cf. Candes, Tao, Donoho, etc.)

Numerical example 1

- Randomly generated system (A, C) with $n = 30$ and $p = 20$ (Gaussian entries)
- Used ℓ_1/ℓ_2 decoder



(a)



(b)

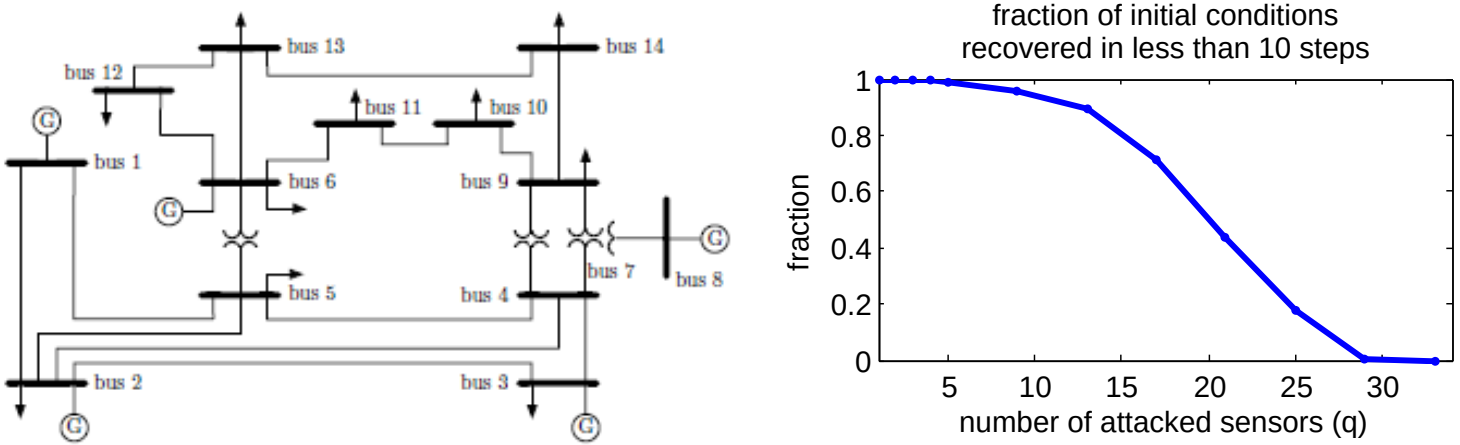
Figure: (a) Fraction of initial conditions (out of 20) that were correctly recovered in less than $T = 20$ time steps, for different values of q. (b) Average number of time steps it took to correctly recover the initial state, as a function of the number of corrupted components.

Numerical example 2

Electric power network: IEEE 14-bus power network (5 generators, 14 buses)

- $n = 2 \times 5 = 10$ states for the rotor angles δ_i and the frequencies $d\delta_i/dt$ of each generator i
- $p = 35$ sensors to measure: real power injections at every bus (14 sensors), real power flows along every branch (20 sensors), rotor angle at generator 1 (1 sensor) [1]

Used ℓ_1/ℓ_∞ decoder



(a) **(b)**

Figure: (a) IEEE 14-bus power network (b) Fraction of initial conditions that were correctly recovered in less than $T = 10$ steps. For each value of q, 200 simulations were carried out with different initial conditions and different sets of attacked sensors.

[1] cf. [Pasqualetti, Dorfler, Bullo 2010]. Thanks to Fabio Pasqualetti from UCSB for the data!

The ℓ_1/ℓ_r decoder

- How suboptimal is the ℓ_1/ℓ_r decoder compared to the ℓ_0 decoder?

The ℓ_1/ℓ_r decoder

- How suboptimal is the ℓ_1/ℓ_r decoder compared to the ℓ_0 decoder?

Proposition

Let $T > 0$ be fixed. Then the ℓ_1/ℓ_r decoder can correct q errors after T steps iff

$$\|(\Phi x)_K\|_{\ell_1/\ell_r} < \|(\Phi x)_{K^c}\|_{\ell_1/\ell_r} \quad \forall x \neq 0 \; \forall K \text{ s.t. } |K| = q$$

(recall that $\Phi x = \left[\begin{array}{c|c|c} Cx & \cdots & CA^{T-1}x \end{array}\right]$)

The ℓ_1/ℓ_r decoder

- How suboptimal is the ℓ_1/ℓ_r decoder compared to the ℓ_0 decoder?

Proposition

Let $T > 0$ be fixed. Then the ℓ_1/ℓ_r decoder can correct q errors after T steps iff

$$\|(\Phi x)_K\|_{\ell_1/\ell_r} < \|(\Phi x)_{K^c}\|_{\ell_1/\ell_r} \quad \forall x \neq 0 \; \forall K \text{ s.t. } |K| = q$$

(recall that $\Phi x = \left[\begin{array}{c|c|c} Cx & \cdots & CA^{T-1}x \end{array}\right]$)

- Interpretation: The (row) components of Φx must be well *spread*.

The ℓ_1/ℓ_r decoder

- How suboptimal is the ℓ_1/ℓ_r decoder compared to the ℓ_0 decoder?

Proposition

Let $T > 0$ be fixed. Then the ℓ_1/ℓ_r decoder can correct q errors after T steps iff

$$\|(\Phi x)_K\|_{\ell_1/\ell_r} < \|(\Phi x)_{K^c}\|_{\ell_1/\ell_r} \quad \forall x \neq 0 \; \forall K \text{ s.t. } |K| = q$$

$$\text{(recall that } \Phi x = \left[\begin{array}{c|c|c} Cx & \cdots & CA^{T-1}x \end{array}\right]\text{)}$$

- Interpretation: The (row) components of Φx must be well *spread*.
  - Condition for ℓ_1/ℓ_r decoder is stronger than condition $|\text{supp}(Cx) \cup \cdots \cup \text{supp}(CA^{T-1}x)| > 2q$

The ℓ_1/ℓ_r decoder

- How suboptimal is the ℓ_1/ℓ_r decoder compared to the ℓ_0 decoder?

Proposition

Let $T > 0$ be fixed. Then the ℓ_1/ℓ_r decoder can correct q errors after T steps iff

$$\|(\Phi x)_K\|_{\ell_1/\ell_r} < \|(\Phi x)_{K^c}\|_{\ell_1/\ell_r} \quad \forall x \neq 0 \; \forall K \text{ s.t. } |K| = q$$

$$\text{(recall that } \Phi x = \left[\begin{array}{c|c|c} Cx & \cdots & CA^{T-1}x \end{array}\right]\text{)}$$

- Interpretation: The (row) components of Φx must be well *spread*.
  - Condition for ℓ_1/ℓ_r decoder is stronger than condition $|\text{supp}(Cx) \cup \cdots \cup \text{supp}(CA^{T-1}x)| > 2q$
- Question: Given (A, C) how to check above condition? no known efficient way...

Summary and Conclusion

Summary:

- Study of linear dynamical systems with attacked sensors
- Efficient algorithm for estimating the state of the system despite the attacked sensors
- Algorithm performs very well in practice

Summary and Conclusion

Summary:

- Study of linear dynamical systems with attacked sensors
- Efficient algorithm for estimating the state of the system despite the attacked sensors
- Algorithm performs very well in practice

Open questions:

- Find efficient way to compute the maximum number of errors that can be corrected for a given system (A, C) (i.e., number of errors that the ℓ_0 decoder can handle).
- Same question for the ℓ_1/ℓ_r decoder...

Extensions:

- Generalize to control systems with inputs
- Study robustness (noise in unattacked sensors, disturbance in state-evolution equation, etc.)

Summary and Conclusion

Summary:

- Study of linear dynamical systems with attacked sensors
- Efficient algorithm for estimating the state of the system despite the attacked sensors
- Algorithm performs very well in practice

Open questions:

- Find efficient way to compute the maximum number of errors that can be corrected for a given system (A, C) (i.e., number of errors that the ℓ_0 decoder can handle).
- Same question for the ℓ_1/ℓ_r decoder...

Extensions:

- Generalize to control systems with inputs
- Study robustness (noise in unattacked sensors, disturbance in state-evolution equation, etc.)

Thank you!

Extra: Error correction over the reals

- Information message: $x \in \mathbf{R}^n$

Decoding by linear programming, Candes and Tao, IEEE Transactions on Information Theory, 2005

Extra: Error correction over the reals

- Information message: $x \in \mathbf{R}^n$
- Add redundancy and transmit $Cx \in \mathbf{R}^N$ where $N > n$ ($C \in \mathbf{R}^{N \times n}$)

Decoding by linear programming, Candes and Tao, IEEE Transactions on Information Theory, 2005

Extra: Error correction over the reals

- Information message: $x \in \mathbf{R}^n$
- Add redundancy and transmit $Cx \in \mathbf{R}^N$ where $N > n$ ($C \in \mathbf{R}^{N \times n}$)
- Receiver receives $y = Cx + e$ where e is q-sparse (q components were corrupted)

Decoding by linear programming, Candes and Tao, IEEE Transactions on Information Theory, 2005

Extra: Error correction over the reals

- Information message: $x \in \mathbf{R}^n$
- Add redundancy and transmit $Cx \in \mathbf{R}^N$ where $N > n$ ($C \in \mathbf{R}^{N \times n}$)
- Receiver receives $y = Cx + e$ where e is q-sparse (q components were corrupted)
- Optimal ℓ_0 decoder:

$$\underset{\hat{x}}{\text{minimize}} \quad \|y - C\hat{x}\|_{\ell_0}$$

Decoding by linear programming, Candes and Tao, IEEE Transactions on Information Theory, 2005

Extra: Error correction over the reals

- Information message: $x \in \mathbf{R}^n$
- Add redundancy and transmit $Cx \in \mathbf{R}^N$ where $N > n$ ($C \in \mathbf{R}^{N \times n}$)
- Receiver receives $y = Cx + e$ where e is q-sparse (q components were corrupted)
- Optimal ℓ_0 decoder:

$$\underset{\hat{x}}{\text{minimize}} \quad \|y - C\hat{x}\|_{\ell_0}$$

- ℓ_1-relaxation:

$$\underset{\hat{x}}{\text{minimize}} \quad \|y - C\hat{x}\|_{\ell_1}$$

Decoding by linear programming, Candes and Tao, IEEE Transactions on Information Theory, 2005