Minimum Sum of Distances Estimator: Robustness and Stability

Yoav Sharon, John Wright, Yi Ma

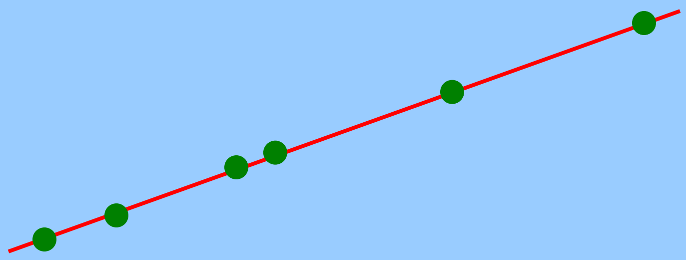
Coordinated Science Laboratory and
Dept. of Electrical & Computer Eng.,
Univ. of Illinois at Urbana-Champaign

Problem Formulation

- Estimate the (vector) state $x_0 \in \mathbb{R}^n$,
  Given m (scalar) measurements, $y_1, \ldots, y_m$,

  $$y_i \approx a_i^T x_0$$

  Obviously, $m > n$
- Example:
  - state – line
  - measurements – points

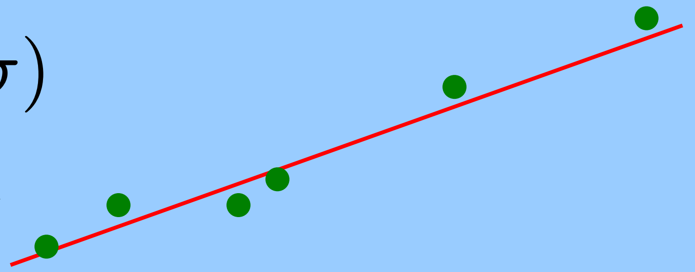


Problem Formulation

- **Different types of errors:**
  - (small) white ***noise***:

    $$y_i = a_i^T x_0 + z_i \quad z_i \sim \mathcal{N}(0, \sigma)$$

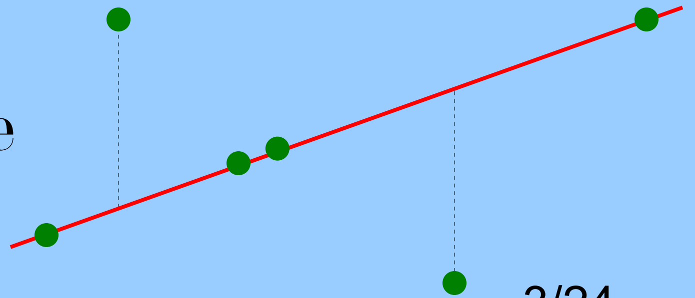
    ➔ Use least squares (LS) – optimal



  - ***corruption*** of some (small number of) measurements:

    $$y_i = a_i^T x_0 + e_i$$

    where e_i can be arbitrarily large

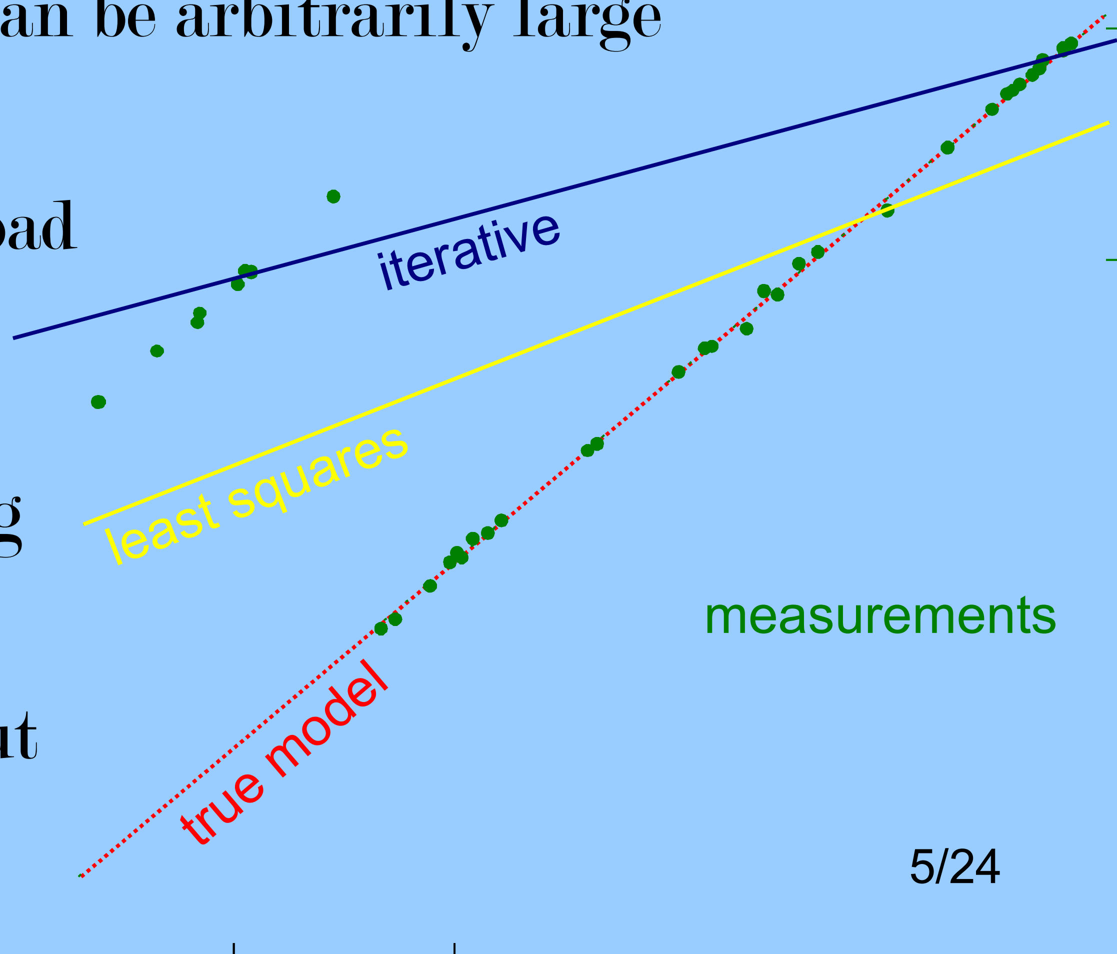


Problem Formulation

- **In matrix-vector notation (with both errors):**
  $y = Ax_0 + z + e, \quad x_0 \in \mathbb{R}^n,\ y, z, e \in \mathbb{R}^m$
- $z \sim \mathcal{N}(0, \sigma)$
- e **is arbitrary but sparse:**
  $p \ll m$ **where** $p = \|e\|_0$ - **#nonzero entries of** e

Other Solutions

- Least Squares – closed form solution, but
  - estimation error can be arbitrarily large
- Iterative
  - add/remove good/bad measurements
  - local minimum
- Random Sampling (RANSAC)
  - very successful, but
  - exponential



Desired Properties

$$y = Ax_0 + z + e$$

- Polynomial complexity
- Guaranteed ***robustness***:

existence of a ***breakdown point***-T^* such that

if $\quad \|e_0\|_0 < T \leq T^*$

then $\quad \|\hat{x} - x_0\| \leq \alpha(A, T)\|z\|$

- ***Stability***: $\forall z : T^*(A, z) = T^*(A, 0) \doteq T^*(A)$

The ℓ^1 Approach

$$\hat{x} = \arg\min_{x} \|y - Ax\|_1$$

First introduced by Candes&Tao and by Donoho

Property	Proved By
polynomial	linear programming
robustness, no noise	Candes&Tao, Donoho
robustness, with noise	this work weaker results exist in compressed sensing
derivation of $\alpha(A,T)$ $\|\hat{x} - x_0\| \leq \alpha(\cdot,\cdot)\|z\|$	this work
stability	this work

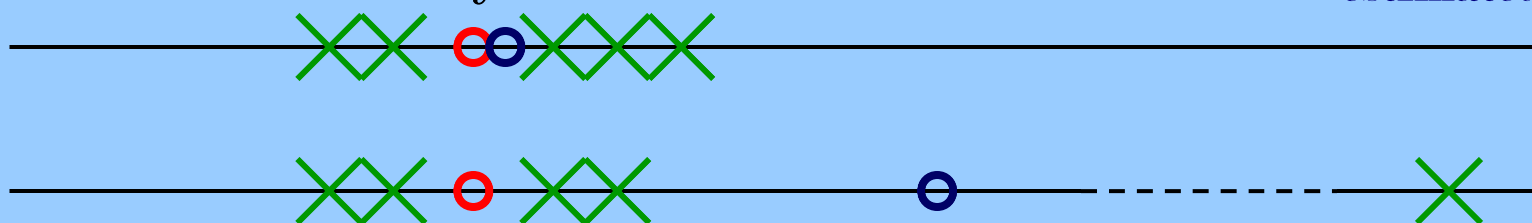
More Results

- Efficient algorithm to compute T^* (this work)
  - Faster than any existing alternative (exponentially) (for certain classes of systems)
  - Polynomial in # of measurements if $\dim(x)$ is fixed
  - Exponential in general
- Existing result by Candes&Tao, Donoho:
  - For randomly generated A, in the limit as $m, n \to \infty$ $T^*(A) \to \rho m$ with high probability $(\rho \in (0,1))$
  - Limited use for deterministic A

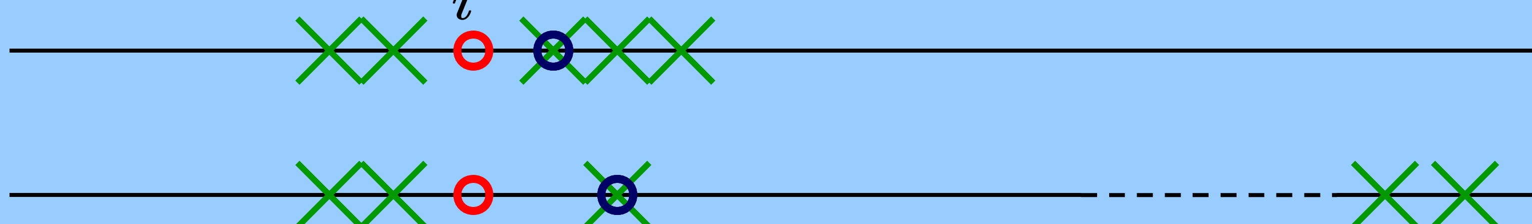
Average vs. Median (1D)

- Average, $\min_x \sum_i |y_i - x|^2 \ (T^* = 0\%)$

true
measurements
estimated



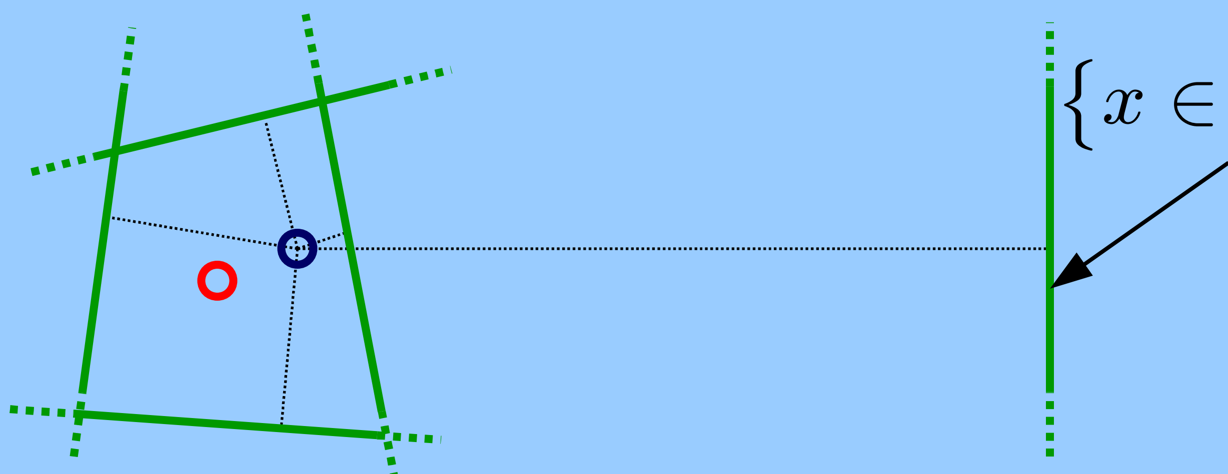
- Median, $\min_x \sum_i |y_i - x| \ \ (T^* = 50\%)$



Extension to Multidimensional

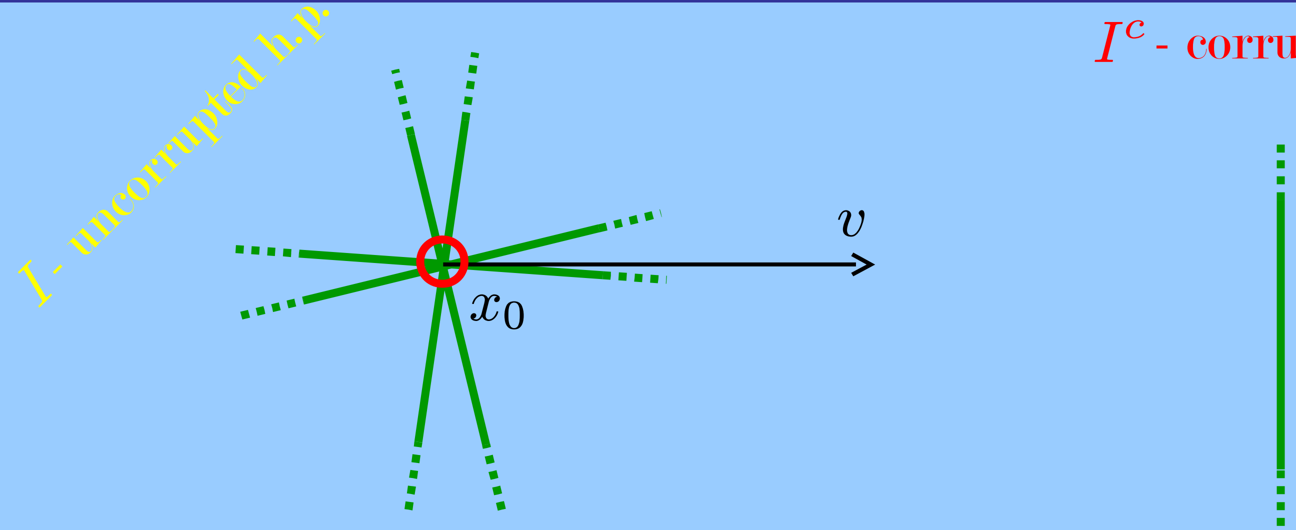
- Measurements as points ($T^* = 50\%$)
  - (different problem than of this paper)

- Measurements as hyperplanes (h.p.) ($T^* = ?$)



$$\{x \in \mathbb{R}^n \,|\, y_i = a_i^T x\}$$

Noiseless Case

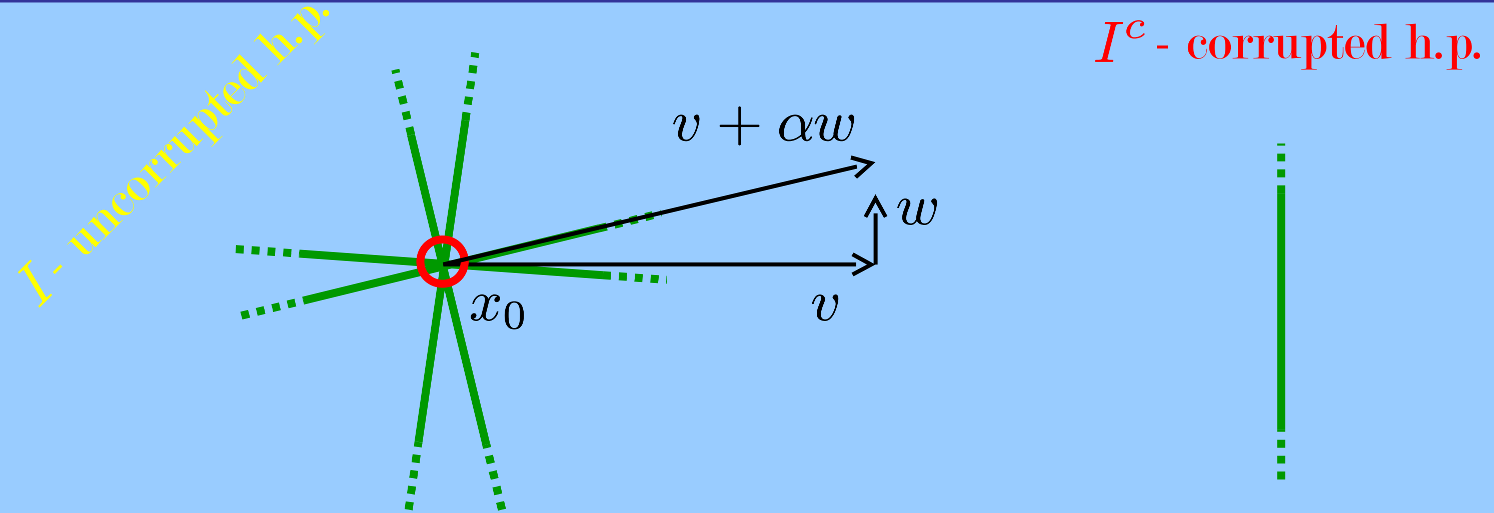


- $x_0 \neq \arg\min_x \|y - Ax\|_1$ **iff** $\exists v$
  $$\sum_{i \in I} |a_i^T v| \leq \sum_{i \in I^c} |a_i^T v| \qquad (1)$$
- $T^* = \min T$ **such that** $\exists v \in \mathbb{R}^m \exists\, |I^c| = T$ **for which (1) holds**

How to Calculate T^*?

- $T^* = \min T$ such that $\exists v \in \mathbb{R}^m \exists \, |I^c| = T$:
  $\sum_{i \in I} \left| a_i^T v \right| \leq \sum_{i \in I^c} \left| a_i^T v \right| \quad$ **(1)**
- *Proposition* : (proof next slide)
  If $\exists v$ s.t. (1) holds, then $\exists v'$ s.t. (1) holds,
  and $\exists \, |J| = n - 1$ s.t. $\forall i \in J : \; a_i^T v' = 0$
- J defines v' (up to multiplication by a scalar)
  finite number of J's
- Need to check (1) for $\forall \, |J| = n-1, \; \forall \, |I^c| < T$
- With sorting don't need to check $\forall \, |I^c| < T$

How to Calculate T^*?



- Assume $\exists v$ such that $\sum_{i \in I} |a_i^T v| \leq \sum_{i \in I^c} |a_i^T v|$ and $\forall i \in J : \ a_i^T v = 0$
- If $|J| < n - 1$, can find $\sum_{i \in I^c} a_i^T w = 0$ and $\forall i \in J : a_i^T w = 0$
- Choose α such that $\operatorname{sgn}(\alpha) \sum_{i \in I} a_i^T w < 0$ and $a_i^T (v + \alpha w) = 0, \ i \in I \setminus J$

Complexity of Calculating T*

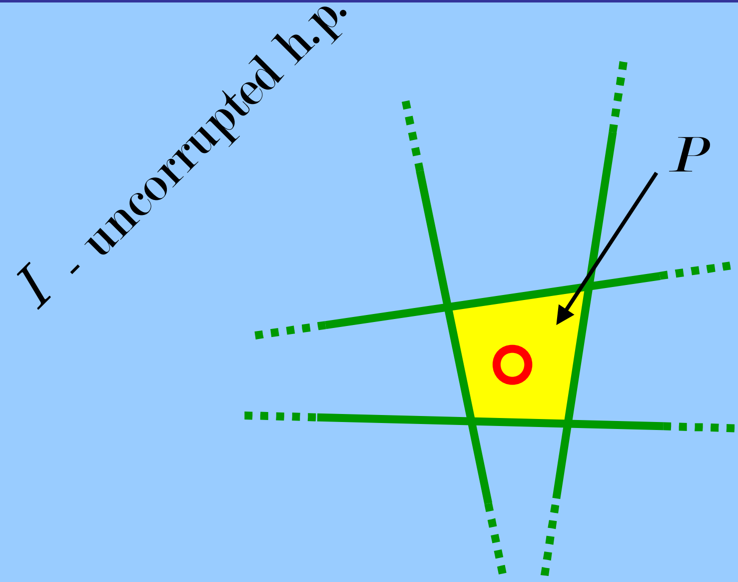
- **Complexity of proposed method:**

$$\binom{m}{n-1}\left(O\left(n^3\right) + O\left(mn\right) + O\left(m \log m\right)\right)$$

iterate over all $|J| = n-1$ | find $v' \in \mathrm{null}\left(A_J\right)$ | compute $\forall i : \left|a_i^T v'\right|$ | sort $\left|a_i^T v'\right|$

- **Exponential in general, polynomial for fixed *n***
- **Alternative method (Candes & Tao):**
  - Whether $\hat{x} = x_0$ only depends on signs of e
  - Complexity: $\sum_{T=1}^{T^*} 2^T \binom{m}{T-1} t_{lp}\left(m, n\right)$
  - Requires solving linear programming for each e

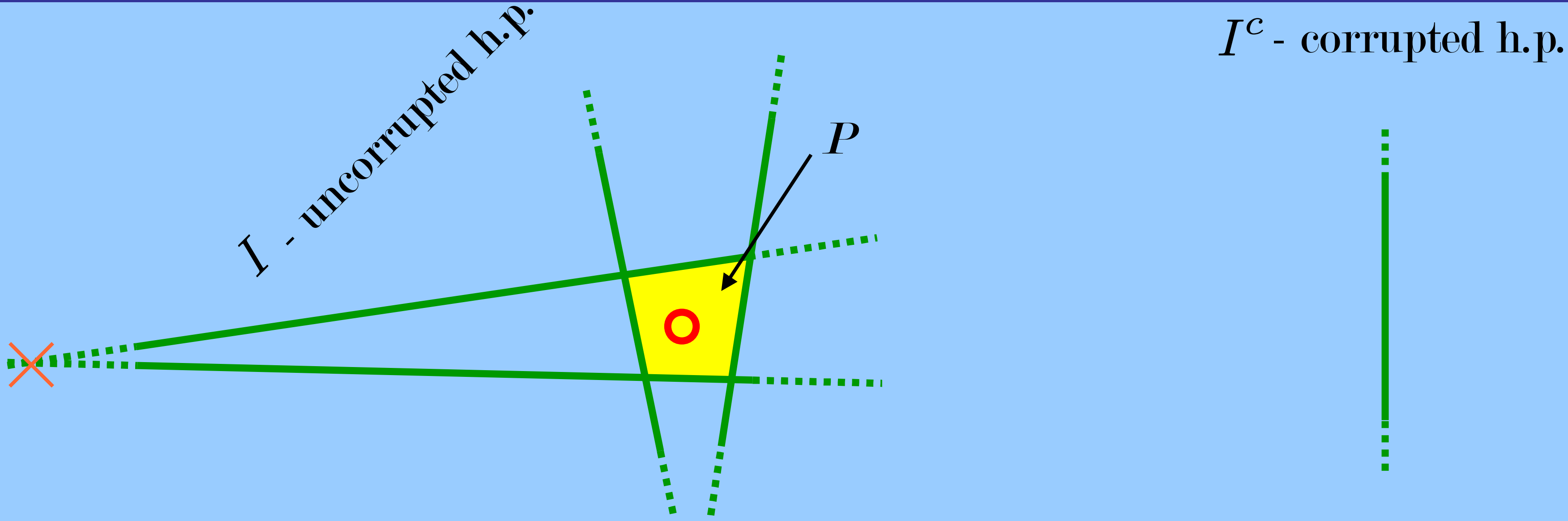
Noisy Case



- Noiseless: all "good" hyperplanes (h.p.) pull to x_0
- Noisy: all "good" h.p. pull to P
  $P-$ convex hull of n-wise intersections of good h.p.

➔ If $|I^c| < T^*(A, z = 0)$ then $\|\hat{x} - x_0\| \leq \text{size}(P)$

Noisy Case



- Which n-wise intersections are in P?
- If we include all n-wise intersections, then with almost parallel hyperplanes (h.p.) $\text{size}\,(P) \to \infty$ even if $\|z\|$ is small

Noisy Case

- A set of h.p., J, is called *possibly extreme set* if
  $\sum_{i \in J \cup I^c} \left| a_i^T \nu_J \right| \leq \sum_{i \in I \setminus J} \left| a_i^T \nu_J \right|$
  ν_J – singular vector associated with $\sigma_{\min}(A_J)$

$$A_J \doteq \begin{bmatrix} a_{J_1}^T \\ \vdots \\ a_{J_n}^T \end{bmatrix}$$

- *Proposition*: (proof in paper)
  - $\hat{x}$ = intersection J' iff:
  - J' is extreme, or
  - There exists z' and an extreme set J'' such that $\hat{x}'$ = intersection J'' and $\|\hat{x} - x_0\| \leq \|\hat{x}' - x_0\|$

Noisy Case

- Assume $\hat{x}$ = intersection J
- Then, $A_J \tilde{x} \doteq A_J (\hat{x} - x_0) = y_J - A_J x_0 = z_J$
- Therefore,

$$\|z_J\|_2 \geq \sigma_{\min}(A_J) \|\tilde{x}\|_2 \Rightarrow \|\tilde{x}\|_2 \leq \frac{\|z_J\|_2}{\sigma_{\min}(A_J)}$$

- **Main** ***Theorem*****:** $\|\hat{x} - x_0\|_2 \leq \alpha(A,T) \|z\|$

  where $\alpha(A,T) = \min_{J' \in Q_T} 1/\sigma_{\min}(A_{J'})$

  and Q_T – set of all extreme sets given $|I^c| < T$

Remarks

- Main *Theorem*: $\|\hat{x} - x_0\|_2 \leq \alpha(A, T)\,\|z\|$
  where $\alpha(A, T) = \min_{J' \in Q_T} 1/\sigma_{\min}(A_{J'})$
  and Q_T – set of all extreme sets given $|I^c| < T$

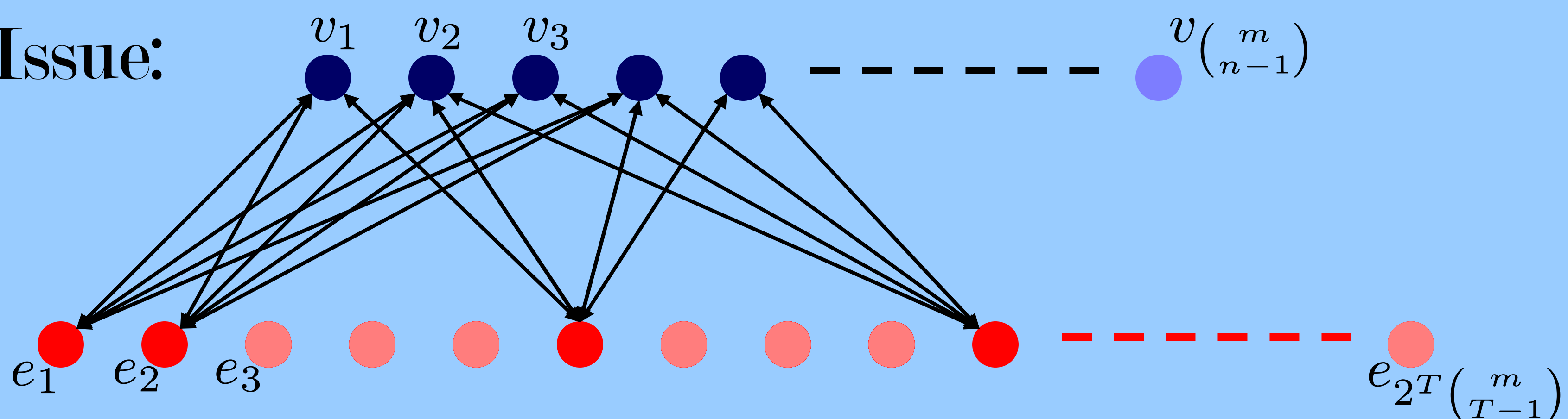
- If $J' \in Q_T$ and $T \leq T^*(A, z = 0)$
  then $\sigma_{\min}(A_{J'}) > 0$

➔ *Corollary*: Stability (and robustness)

- Smaller $T \longrightarrow$ less extreme sets $\longrightarrow$ better bound

Remarks

- Worst case analysis
  - Recovery of x_0 may succeed even if $|I^c| \geq T^*$
- Desire: distribution of $\tilde{x} = \hat{x} - x_0$ given probabilistic models of e, z
- Issue:



One “bad” v can cause many e's to fail

Remarks

- RANSAC - <u>exponential</u> complexity
- ℓ^1 - <u>polynomial</u> complexity

➔ Thus: main advantage of ℓ^1 is in large problems

  – Recovery of x_0 is computationally feasible

  – Random matrices: high T^* is expected (Comp. Sens.)

  – Deterministic matrices: calculating T^* is expensive

  – Calculating $\alpha\left(A, T\right)$ is even more expensive

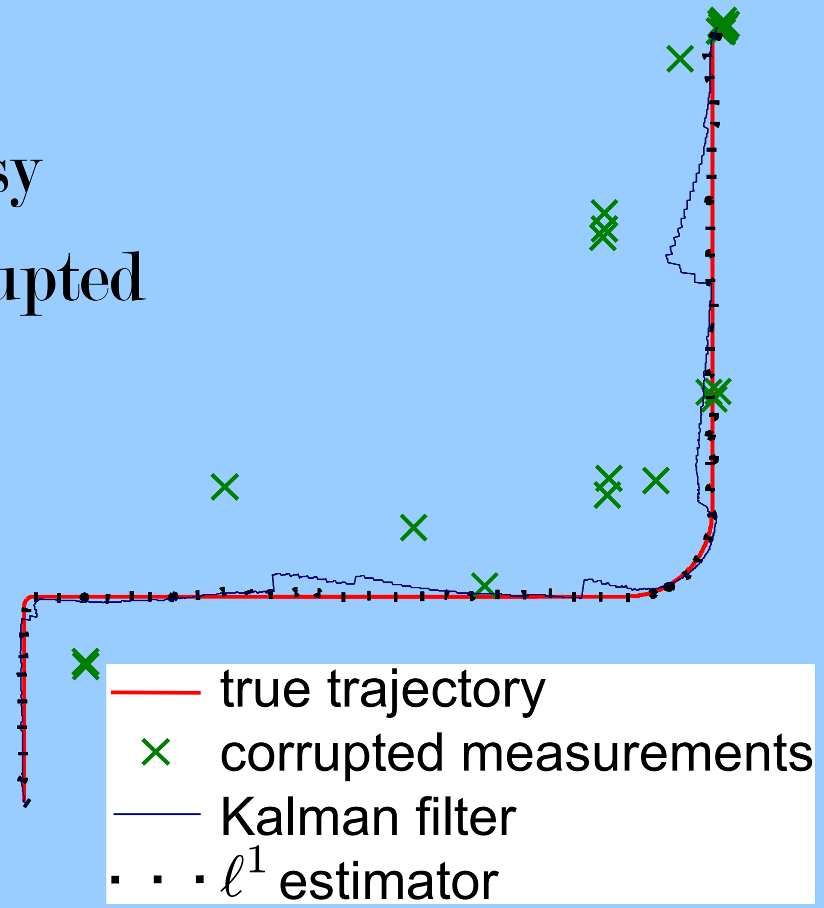
- Optimal weighting for maximizing T^*
  – Possible due to this work, too expensive in practice

Open Questions

- Calculating T^* for special structures:
  - $A = \left[C; CF; CF^2; \ldots; CF^n\right], \quad T^*(C, F, n) = ?$
- Tradeoff between $\|\cdot\|_1$ and $\|\cdot\|_2$
  - $\|\cdot\|_1$ good for error correction
  - $\|\cdot\|_2$ better for noise attenuation
  - $\|\cdot\|_1 + \lambda\|\cdot\|_2$?
- Extension to mixture of points, hyperplanes and other linear subspaces

Applications

- State – position (2D), velocity, orientation
- Measurements:
  - Inertial (acceleration): noisy
  - GPS (position): noisy, corrupted
- Small scale problem
  - Other methods may perform better
- Large scale problems?



Thank You