Analyzing Least Squares & Kalman Filtered Compressed Sensing

Namrata Vaswani
Dept. of Electrical & Computer Engineering
Iowa State University
http://www.ece.iastate.edu/~namrata/
Goal

- **Goal:** Reconstruct
  - a time sequences of (spatially) sparse signals
  - with slowly changing sparsity patterns (support sets)
  - from a limited number of incoherent measurements
  - in real-time (recursively and causally)

- **Examples:** real-time dynamic MR imaging, video compression, single-pixel video, sensor nets for real-time estimation of time-varying fields

- **Key assumption:** sparsity pattern changes slowly over time
Slowly Changing Sparsity

- **Approx. Sparsity.** Size of 99%-energy support set: less than 7% for the larynx sequence and less than 9% for the cardiac sequence.

- **Slow Change in Sparsity Pattern.** Maximum size of change in support: less than 2% of minimum sparsity size in both cases
Outline

• Problem definition and background
• LS CS-residual (LS-CS) & KF CS-residual (KF-CS)
• Bounding LS CS error
• Conditions for stability of LS and KF CS
• Simulations
Problem Definition

Recursively reconstruct a sparse vector, \( x_t \), from the current observation, \( y_t := Ax_t + w_t \), & all past observations, \( y_{1:t-1} \)

- \( \dim(y_t) = n < \dim(x_t) = m \)
- \( x_t \) is \( S_t \)-sparse with support set, \( N_t \)
- the support, \( N_t \), changes slowly over time
- \( A \) is \( S_* \)-“approximately orthonormal” (\( \delta_{S_*} < 1/2 \)) and \( S_* > S_t \)
  - i.e. \( \|Ax\|_2 \) between 0.7 and 1.2 for \( S_* \) or less sparse vectors \( x \)
RIP and ROP constants [Candes,Tao]

- Restricted Isometry constant, $\delta_S$: smallest real number satisfying

$$(1 - \delta_S)\|c\|_2^2 \leq \|A_T c\|_2^2 \leq (1 + \delta_S)\|c\|_2^2$$

for all subsets $T$ with $|T| \leq S$ and for all $c$

  - Easy to see: $\|(A_T' A_T)^{-1}\|_2 \leq 1/(1 - \delta_{|T|})$

- Restricted Orthogonality constant, $\theta_{S,S'}$: smallest real number satisfying

$$|c_1' A_{T_1} A_{T_2} c_2| \leq \theta_{S,S'}\|c_1\|_2\|c_2\|_2$$

for all disjoint sets $T_1, T_2$ with $|T_1| \leq S$, $|T_2| \leq S'$ and for all $c_1, c_2$

  - Easy to see: $\|A_{T_1}' A_{T_2}\|_2 \leq \theta_{|T_1|,|T_2|}$
Compressed Sensing [Candes, Romberg, Tao] [Donoho]

- CS (noiseless) [Candes, Romberg, Tao '05] [Donoho'05]

- CS (noisy - BPDN) [Chen, Donoho] [Tropp'06]

- CS (noisy - Dantzig Selector) [Candes, Tao '06]

\[
\min_{\beta} \|\beta\|_1 \text{ s.t. } \|A'(y_t - A\beta)\|_\infty < \lambda
\]

We use \( \hat{x}_t = \text{CS}(y_t) \) to denote the solution of above.

- If noise is bounded between \( \pm \lambda/\|A\|_1 \) in each dimension, and \( \delta_{3S_t} < 1/2 \),

\[
\|x_t - \hat{x}_t\|_2^2 \leq C_1(S_t) S_t \lambda^2
\]

\[
C_1(S) := 16/(1 - \delta_{2S} - \theta_{S,2S})^2
\]

(simplification of Theorem 1.1 of [Dantzig Selector])
The Question

- Most existing work: Batch-CS on entire time sequence
  [Gamper et al ’08 (dynamic MRI)], [Wakin et al (video)]
  - Offline and very slow, but uses few measurements

- Alternative: CS at each time separately (simple CS)
  - Causal and fast, but needs many more measurements

- The Question: How can we
  - improve simple CS by using past observations, and
  - how can we do it recursively, i.e. by only using the previous signal estimate and the current observation?
Finding a Recursive Solution

- Given $y_t := Ax_t + w_t$, $x_t$ is sparse with support $N_t$, $N_t$ changes slowly over time, $A$ satisfies $\delta_{S_t} < 1/2$, $S_t := |N_t|

- **If $N_t$ known:** easy to compute a restricted-LS estimate

  \[ \hat{x}_t = \text{restrictedLS}(y_t, N_t) := (\hat{x}_t)_{N_t} = A_{N_t}^\dagger y_t, (\hat{x}_t)_{N_c} = 0 \]

- **If $N_t$ unknown:** an option is to estimate it by thresholding CS output

  \[ \hat{N}_t = \text{threshold}(\text{CS}(y_t)) \text{ threshold}(x) := \{i : x_i^2 > \alpha\} \]

  and then do the same thing

  - **But: not using past observations:** large error
CS-residual idea [Vaswani, ICIP’08, ICASSP’09]

- Let $T := \hat{N}_{t-1}$ (estimated support at $t - 1$) and $\Delta := N_t \setminus \hat{N}_{t-1}$
- Assume that the undetected set, $\Delta$, is small, i.e.
  - the support changes slowly, and
  - the support at $t - 1$ is well estimated
- Use $T := \hat{N}_{t-1}$ to compute restricted LS estimate, & observation residual
  
  $\hat{x}_{t,\text{init}} = \text{restrictedLS}(y_t, T)$
  $y_{t,\text{res}} = y_t - A\hat{x}_{t,\text{init}}$

- CS-residual: $\hat{x}_t = \hat{x}_{t,\text{init}} + \text{CS}(y_{t,\text{res}})$

  - $y_{t,\text{res}}$ is a noisy measurement of an approx. $|\Delta|$ sparse vector
Why CS-residual works?

- Notice that $y_{t,\text{res}} = A\beta_t + w_t$ and $\beta_t := x_t - \hat{x}_{t,\text{init}}$ satisfies

$$
(\beta_t)_\Delta = (x_t)_\Delta \\
(\beta_t)_T = -A_T^\dagger (A_\Delta (x_t)_\Delta + w_t) \\
(\beta_t)_{(T \cup \Delta)^c} = 0
$$

- If $|\Delta|$ small enough s.t. $||A_T' A_\Delta||_2 < \theta_{|T|, |\Delta|}$ small:
  - $\beta_t$ small along $T$, i.e. it is only $|\Delta|$-approx-sparse

- CS error strongly depends on approx. sparsity size
  - **CS-residual**: much smaller error than CS on $y_t$ (simple CS)
LS CS-residual (LS-CS) algorithm

Initial LS
Compute LS estimate, & residual using T

\[ \tilde{y}_t, \tilde{\chi}_t, \text{init} \]

CS-residual
Do CS on \( \tilde{y}_t, \text{res} \) to estimate \( \beta_t := x_t - \tilde{\chi}_t, \text{init} \)

\[ \tilde{\chi}_t, CS_{res} = \tilde{\chi}_t, \text{init} + \beta_t \]

Estimate Support

Final LS
estimate using \( \hat{N}_t \)

Delay
\[ t \leftarrow t + 1 \]

Estimate Support. Either do

- \( \hat{N}_t = \text{threshold}(\tilde{\chi}_t, CS_{res}) \) (add and delete indices at the same time)
- Or \( \hat{N}_t = T \cup \text{threshold}(\tilde{\chi}_t, CS_{res}) \) (only add new indices) (easier to analyze)
Bounding LS CS-residual error [Vaswani, ICASSP’09]

- **Assume that**
  1. noise bounded b/w $\pm \lambda / ||A||_1$ and has variance $\sigma^2$ in each dimension
  2. number of false additions, $|T \setminus N_t| \leq S_{fa}$ and $\delta_{S_t+S_{fa}} < 1/2$
  3. number of new plus undetected additions, $|\Delta| < S_t/3$

  - recall: $T := \hat{N}_{t-1}$, $\Delta := N_t \setminus T$

- **Expected CS-residual error given past := $y_{1:t-1}$ is bounded by**

  $$C_2(|\Delta|) |\Delta| \lambda^2 + C_3(|\Delta|) \frac{|T|}{|\Delta|} [4\theta^2_{|T|,|\Delta|} E[||x_t\Delta||^2|\text{past}]] + 2|T|\sigma^2]$$

  - $C_2(S), C_3(S)$: linear functions of $1/(1 - \delta_{2S} - \theta_{S,2S})^2$

- **Final LS usually further improves performance**
Comparison with CS

- If noise bounded and $\delta_{3S_t} < 1/2$, CS error bounded by

$$C_1(S_t) \Delta \lambda^2 + C_1(S_t) (S_t - \Delta) \lambda^2$$

$C_1(S)$: linear function of $1/(1 - \delta_{2S} - \theta_{S,2S})^2$

- holds under much stronger assumption
- the constant much larger and second term not much smaller

- If noise bounded and $\delta_{3S} < 1/2$ ($S < S_t$), CS error bounded by

$$\min_{S: \delta_{3S} < 1/2} C_2(S) S \lambda^2 + C_3(S) \frac{S_t - S}{S} E[\| (x_t)_{\text{rest}} \|^2 | \text{past}]$$

rest: $(S_t - S)$ nonzero elements of $x_t$ which did not get estimated

- constants same, but second term much larger
Comparison with CS: special case

- Let $S = |\Delta|$ is the optimal $S$ for CS bound. Also, let $S_{f,a} = 0$

- **CS error bound:**

\[ C_2 |\Delta| \lambda^2 + C_3' E[\| (x_t)_{\text{rest}} \|^2 |\text{past}] \]

- **CS-residual error bound:**

\[ C_2 |\Delta| \lambda^2 + C_3' [4\theta^2 E[\| (x_t)_\Delta \|^2 |\text{past}] + 2|T|\sigma^2] \]

- $T := \tilde{N}_{t-1}$, $\Delta := N_t \setminus T$, $\theta := \theta_{|T|,|\Delta|}$

- **CS-residual bound much smaller when**
  - $|\Delta|$ is small $\rightarrow$ much smaller than $|\text{rest}| = S_t - |\Delta|$
  - $|\Delta|$ small $\rightarrow \theta^2 = \theta_{|T|,|\Delta|}^2$ is small
  - SNR large (when SNR too small, nothing works) $\rightarrow \sigma^2$ small
Disclaimer

• We are only comparing upper bounds

• The upper bound on CS error being larger does not necessarily mean that CS-residual is better
Kalman filtered CS-residual (KF-CS)

[Vaswani, ICIP’08]

- So far only used $\hat{N}_{t-1}$ to improve accuracy of CS at $t$: did not use $\hat{x}_{t-1}$
- If a prior dynamic model for nonzero coefficients of $x_t$ is available: do this by replacing initial LS by a KF for $(x_t)_T$
- A possible prior model: random-walk on $(x_t)_N$ starting with $x_0 = 0$
  \[ x_t = x_{t-1} + \nu_t, \quad \nu_t \sim \mathcal{N}(0, Q_t), \quad Q_t = \sigma_s^2 I_{N_t} \]
  $I_T$: diagonal matrix, 1’s at diagonal locations from set $T$, zero elsewhere

- **KF CS-residual:**
  - dimension-varying KF with current states’ set being $T := \hat{N}_{t-1}$
  - compute $\hat{N}_t$ by thresholding output of CS on KF residual
KF CS-residual (KF-CS) algorithm

- **Initial KF.** Let \( T = \hat{N}_{t-1} \). Run a Kalman prediction and update step using \( \hat{Q}_t = \sigma^2_{sys} I_T \) and compute the KF residual, \( y_{t,\text{res}} \), i.e. compute

\[
\begin{align*}
\hat{P}_{t|t-1} &= P_{t-1} + \hat{Q}_t, \quad \text{where} \quad \hat{Q}_t := \sigma^2_{sys} I_T \\
K_t &= P_{t|t-1} A' (A P_{t|t-1} A' + \sigma^2 I)^{-1}, \quad P_t = (I - K_t A) P_{t|t-1} \\
\hat{x}_{t,\text{init}} &= (I - K_t A) \hat{x}_{t-1} + K_t y_t \\
y_{t,\text{res}} &= y_t - A \hat{x}_{t,\text{init}}
\end{align*}
\]

- **CS-residual.** Compute \( \hat{x}_{t,CSres} = \hat{x}_{t,\text{init}} + \text{CS}(y_{t,\text{res}}) \)

- **Estimate Support.** \( \hat{N}_t = T \cup \text{Threshold}(\hat{x}_{t,CSres}) \)

- **Final LS.** If \( \hat{N}_t \) is equal to \( \hat{N}_{t-1} \), set \( \hat{x}_t = \hat{x}_{t,\text{init}} \),

  else compute restricted LS estimate using \( \hat{N}_t \) and update \( P_t \), i.e.

\[
\begin{align*}
\hat{x}_t &= \text{restrictedLS}(y_t, \hat{N}_t) \\
(P_t)_{\hat{N}_t,\hat{N}_t} &= \left( A_{\hat{N}_t}' A_{\hat{N}_t} \right)^{-1} \sigma^2
\end{align*}
\]
Convergence to Genie KF/LS [Vaswani, ICASSP’09]

- **Assume that**
  1. the noise is bounded
  2. \((x_t)^N_t\) follows the Gaussian random walk model
  3. \(A\) is incoherent enough to ensure that \(\delta_{3S_t} < 1/2\) for all \(t\)
  4. addition threshold, \(\alpha\), set large enough to prevent false additions
  5. all additions occur before a finite time

- **KF (LS) CS-residual estimate converges to the genie-aided KF (LS) estimate in probability**
Corollary: Stability [Vaswani, ICASSP’09]

- If replace “all additions before a finite time” by the following
  - number of additions at a given time less than $S_{a,max} << S_t$, and
  - delay between two addition times is large enough

  above: one way to quantify “slowly changing sparsity pattern”

- w.h.p. KF (LS) CS-residual gets to within a small error of the genie KF (LS), within a finite delay after a new addition time
  - (and remains that way until next addition)
Main Idea of Proof

- Bounded noise and $\delta_{3S_t} < 1/2 \rightarrow$ CS-residual error bounded by a constant
- Addition threshold = error bound $\rightarrow$ ensures no false additions
- Gaussian random-walk model on $(x_t)_{N_t} \rightarrow$ expected value of the square of any nonzero coefficient keeps increasing linearly with $t$
  - w.h.p. it will exceed addition threshold plus error bound within a finite delay after being added, i.e. will get detected
- All additions before finite time $\rightarrow$ w.h.p. all detected before a finite time
  - LS CS-residual converges to genie LS in probability
- When all nonzero elements detected KF CS-residual runs a time-invariant KF with correct model parameters
  - will converge to genie KF in mean square and hence in probability
Simulations: random-Gaussian meas.

- Signal length: $m=256$
- Sparsity size: $S_1 = 8$, 2 new additions every 5 time units from $t=10$ to 50
- Observations: $n=72$, $\sigma^2 = 16/9n$
- To answer a question raised during the talk
  - The above is just one simulation example used to make it similar to the assumptions used by our theorem
  - There could be new additions or deletions at every frame (this happens for real image sequences, e.g. MRI – see last page) and our algorithm still works
  - Even though the above simulation may appear to be solvable using MUSIC, the general scenario is not. Also MUSIC is NON-CAUSAL, we want a CASUAL algorithm.
  - Also, to use MUSIC one would need to know how long there has been no addition (this may be possible to detect by thresholding norm of residual)
Related Work

- Our Kalman filtered CS work first appeared in ICIP’08

- Works not using the current observation to compute the initial estimate that is then used to compute the observation residual
  - k-t FOCUSS \cite{JungYeISBI08}
  - Locally Competitive Algorithms for sparse coding \cite{RozellICIP07}

- Very recent work
  - Recursive Lasso \cite{AngelosanteGiannakisICASSP09}
  - Dynamic l1 minimization \cite{AsifRombergCISS09}
  - KF-CS for dynamic MRI \cite{QiuLuVaswaniICASSP09}
  - Modified-CS \cite{VaswaniLuISIT09}
Summary

- LS CS-residual and KF CS-residual

- Bounded LS CS-residual error under mild assumptions
  - bound much tighter than CS if sparsity pattern changes slowly enough

- KF (LS) CS-residual gets to within a small error of the genie-aided KF (LS) within a finite delay after a new addition
  - proved this under stronger assumptions
Ongoing/Future Work

- **Modified-CS** [Vaswani, Lu, ISIT’09]. $\hat{x}_t$ is the solution of

$$\min_{\beta} ||\beta^c||_1 \quad \text{s.t.} \quad y_t = A\beta$$

  - an approach for provably exact reconstruction from noiseless measurements using partly known support, $T := \hat{N}_{t-1}$
  - **exact reconstruction if** $\delta_{|T|+2|\Delta|} < 1/5$ (**much weaker than CS**)

- Combine Modified-CS with CS-residual for noisy/compressible cases

- **Prove stability under weaker assumptions**

- **KF CS-residual for dynamic MRI** [Qiu, Lu, Vaswani, ICASSP’09]
- Combine Modified-CS with CS-residual for noisy/compressible cases

\[
\min_{\beta} \gamma \|\beta_{Tc}\|_1 + (1/2)\|y_{t,res} - A\beta\|_2^2
\]

\[
\min_{\beta} \gamma \|\beta_{Tc}\|_1 + \|\beta_T - (\hat{x}_{t-1})_T\|_{(P_{t|t-1})_T,T}^2 + (1/2\sigma^2)\|y_t - A\beta\|_2^2
\]

- A good way to delete zero coefficients
Simulated MRI [Qiu, Lu, Vaswani, ICASSP’09]

- Observations: $n = m/2$, $m = 128$ (one column at a time)
- Support size $\sim 0.26m$, change in support $\sim 0.03m$
- Variable density undersampling in ky, full resolution in kx
- Select $\gamma$ using the error bound of [Tropp’06]

Plot of normalized MSE against time

Cardiac sequence: reconstructed last frame