Tractable downfall of basis pursuit in structured sparse optimization

Maya V. Marmary Christian Grussler

Technion - Israel Institute of Technology

IAAC Graduate Students in Systems and Control Conference, July 2025

Table of Contents

Table of Contents

Sparse minimization

$$V \in \mathbb{R}^{m \times n},\ m < n,\ y \in \mathbb{R}^n$$

ℓ_0 Minimization

$$\begin{aligned} \underset{u \in \mathbb{R}^n}{\text{minimize}} \quad & \|u\|_{\ell_0} := \sum_{i=1}^{n} |\mathrm{sign}(u_i)| \\ \text{subject to} \quad & Vu = y. \end{aligned}$$

- Compressed Sensing & ML
- System Id. & Model Reduction
- Sensor & Actuator Selection
- Fuel Optimal Control

Sparse minimization

$$V \in \mathbb{R}^{m \times n},\ m < n,\ y \in \mathbb{R}^n$$

ℓ_0 Minimization

$$\begin{aligned} \underset{u \in \mathbb{R}^n}{\text{minimize}} \quad & \|u\|_{\ell_0} := \sum_{i=1}^{n} |\text{sign}(u_i)| \\ \text{subject to} \quad & Vu = y. \end{aligned}$$

$$\stackrel{?}{=}$$

ℓ_1 Minimization

$$\begin{aligned} \underset{u \in \mathbb{R}^n}{\text{minimize}} \quad & \|u\|_{\ell_1} := \sum_{i=1}^{n} |u_i| \\ \text{subject to} \quad & Vu = y. \end{aligned}$$

- Compressed Sensing & ML
- System Id. & Model Reduction
- Sensor & Actuator Selection
- Fuel Optimal Control

Mutual Coherence and Restricted Isometry Property

- **Mutual Coherence:** For a matrix $V \in \mathbb{R}^{m \times n}$ with columns $v_1, \ldots, v_n$, the *mutual coherence* is defined as

$$\mu(V) = \max_{1 \le i \ne j \le n} \frac{|\langle v_i, v_j \rangle|}{\|v_i\|_2 \, \|v_j\|_2} = \cos \theta_{ij}.$$

Lower $\mu(V)$ indicates that the columns of V are less correlated. A solution is guaranteed to be sparse if $\|u^*\|_0 < 1/2(1 + 1/\mu(V))$

Mutual Coherence and Restricted Isometry Property

- **Mutual Coherence:** For a matrix $V \in \mathbb{R}^{m \times n}$ with columns $v_1, \ldots, v_n$, the *mutual coherence* is defined as

$$\mu(V) = \max_{1 \le i \ne j \le n} \frac{|\langle v_i, v_j \rangle|}{\|v_i\|_2 \, \|v_j\|_2} = \cos\theta_{ij}.$$

  Lower $\mu(V)$ indicates that the columns of V are less correlated. A solution is guaranteed to be sparse if $\|u^*\|_0 < 1/2(1 + 1/\mu(V))$

- **Restricted Isometry Property (RIP):** A matrix V satisfies the RIP of order s with constant $\delta_s \in (0,1)$ if, for every s-sparse vector x,

$$(1-\delta_s)\|x\|_2^2 \le \|Vx\|_2^2 \le (1+\delta_s)\|x\|_2^2.$$

  This property ensures that V approximately preserves the ℓ_2-norm of sparse signals.

  $\Longrightarrow$ **Problem:** NP-hard to verify and often not practical!

Fuel Optimal Control

L_0 Minimization

$$\begin{aligned} &\text{minimize} && \|u\|_{L_0} := \int_0^T |\text{sign}(u(t)|dt \\ &\text{subject to} && \dot{x} = Ax + bu \\ &&& x(0) = \xi,\ x(T) = 0 \end{aligned}$$

$=$

L_1 Minimization

$$\begin{aligned} &\text{minimize} && \|u\|_{L_1} := \int_0^T |u(t)|dt \\ &\text{subject to} && \dot{x} = Ax + bu \\ &&& x(0) = \xi,\ x(T) = 0 \end{aligned}$$

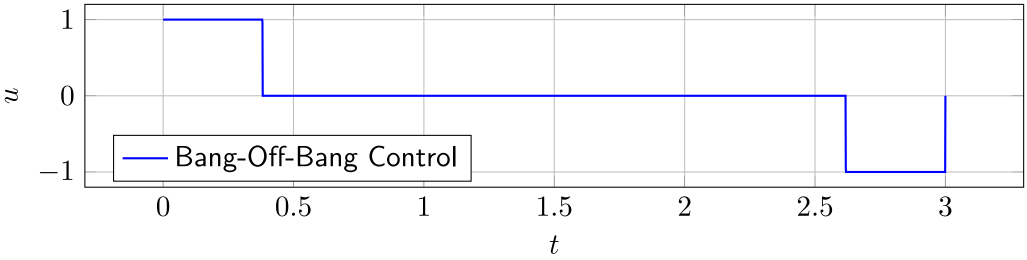


Table of Contents

Main result - Failure Guarantees

Main result - Failure Guarantees

1. Alignment property: $\|V^\mathsf{T}\beta^*\|_\infty = {u^*}^\mathsf{T}\underbrace{V^\mathsf{T}\beta^*}_{\tilde{\beta}^*} = \|u^*\|_1$

Main result - Failure Guarantees

1. $$\text{Alignment property: } \|V^\mathsf{T}\beta^*\|_\infty = {u^*}^\mathsf{T} \underbrace{V^\mathsf{T}\beta^*}_{\tilde{\beta}^*} = \|u^*\|_1$$

2. $$\tilde{\beta}^* \in \mathsf{Im}(V^\mathsf{T}) = \mathsf{Im}\left(\begin{bmatrix} I_m \\ \left[V_{(1:m,:)}^{-1} V_{(:,m+1:n)}\right]^\mathsf{T} \end{bmatrix}\right).$$

Main result - Failure Guarantees

Theorem

Let $u^* \in \mathbb{R}^n$ be a solution to ℓ_0 problem, where $V \in \mathbb{R}^{m\times n}$ is such that $V_{(:,1:m)}$ is invertible. Then, u^* is not a solution of ℓ_1 problem if there exists an $i^* \in (m+1:n)$ with $u_i^* \neq 0$ and $\|V_{(:,1:m)}^{-1} V_{(:,i^*)}\|_{\ell_1} < 1$.

1. $$\text{Alignment property: } \|V^\mathsf{T}\beta^*\|_\infty = {u^*}^\mathsf{T} \underbrace{V^\mathsf{T}\beta^*}_{\tilde{\beta}^*} = \|u^*\|_1$$

2. $$\tilde{\beta}^* \in \mathrm{Im}(V^\mathsf{T}) = \mathrm{Im}\left(\begin{bmatrix} I_m \\ \left[V_{(1:m,:)}^{-1} V_{(:,m+1:n)}\right]^\mathsf{T} \end{bmatrix}\right).$$

Tractable removal of "unhelpful" entries

Goal: Systematic removal of "unhelpful" entries.

Observation:

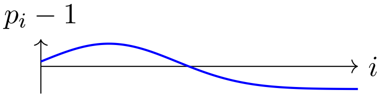
- If $b = \mathbf{1}_m \in \mathbb{R}^m$, $A = \text{diag}(\lambda_1, \ldots, \lambda_m)$, $\lambda_i \geq 0$ and $V = \begin{bmatrix} b & Ab & \ldots & A^N b \end{bmatrix}$, then

  $$p_i = \|V_{(1:m,:)}^{-1} V_{(:,m+1:i)}\|_1 = \mathbf{1}_m^\mathsf{T} W V_{(1:m,:)}^{-1} V_{(:,m+1:i)}$$

  is part of an impulse response.

  $\Longrightarrow$ (Schur-)Stable A and large enough N, $p_i < 1$ for $i \geq I^*$.

- Tractability by unimodality of p

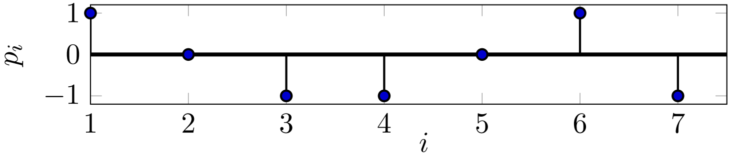


  $\Longrightarrow$ Simple binary search to find smallest I^* with $p_{I^*} < 1$.

Variation & Unimodality

- **Variation of a Vector:** For $p \in \mathbb{R}^n$

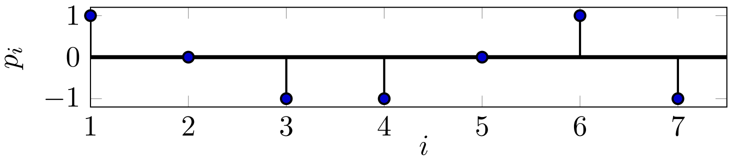
$$\mathsf{S}(p) := \#\text{ of sign changes (after deleting zeros)}$$



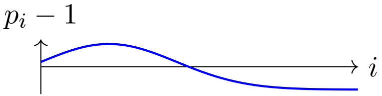
Variation & Unimodality

- **Variation of a Vector:** For $p \in \mathbb{R}^n$

$$\mathsf{S}(p) := \# \text{ of sign changes (after deleting zeros)}$$



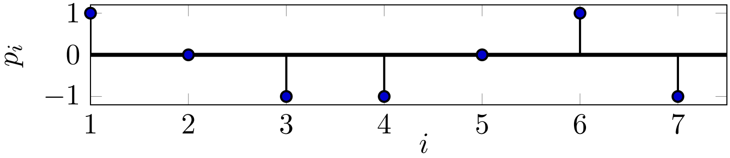
- **Unimodality:** Let $(\Delta p)_i := p_{i+1} - p_i$, then p is *unimodal* if $\mathsf{S}(\Delta p) \leq 1$ and a possible sign change in Δp occurs from positive to negative.



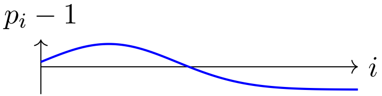
Variation & Unimodality

- **Variation of a Vector:** For $p \in \mathbb{R}^n$

$$\mathsf{S}(p) := \#\text{ of sign changes (after deleting zeros)}$$



- **Unimodality:** Let $(\Delta p)_i := p_{i+1} - p_i$, then p is *unimodal* if $\mathsf{S}(\Delta p) \le 1$ and a possible sign change in Δp occurs from positive to negative.



- k**-Variation Bounding (VB$_k$):** A matrix $X \in \mathbb{R}^{k \times n}$ with $k > n$ is *k-variation bounding* if for every $u \in \mathbb{R}^n \setminus \{0\}$,

$$\mathsf{S}(Xu) \le k - 1.$$

Sign Consistency & Variation Bounding

For $X \in \mathbb{R}^{m \times n}$, $m > n$,

- m-**Variation Bounding (VB$_m$):** X is m-*variation bounding* if for every $u \in \mathbb{R}^n \setminus \{0\}$,

$$\mathsf{S}(Xu) \leq m - 1.$$

Sign Consistency & Variation Bounding

For $X \in \mathbb{R}^{m \times n}$, $m > n$,

- m-**Variation Bounding (VB$_m$):** X is m-*variation bounding* if for every $u \in \mathbb{R}^n \setminus \{0\}$,

$$\mathsf{S}(Xu) \leq m - 1.$$

- X is m-sign consistent if all its minors of order m share the same sign.

Sign Consistency & Variation Bounding

For $X \in \mathbb{R}^{m \times n}$, $m > n$,

- m-**Variation Bounding (VB$_m$):** X is *m-variation bounding* if for every $u \in \mathbb{R}^n \setminus \{0\}$,

$$\mathsf{S}(Xu) \leq m - 1.$$

- X is m-sign consistent if all its minors of order m share the same sign.

VB$_m$ versus SC$_m$ (Lemma): For $X \in \mathbb{R}^{m \times n}$ with $m > n$ and $\mathrm{rk}(X) = n$, X is m-variation bounding (VB$_m$) if and only if it is m-sign consistent (SC$_m$).

Main result - Unimodality

Theorem

Let $V \in \mathbb{R}^{m\times n}$, $m < n$, *be such that* $V \in SC_m$, $\Delta((\mathbf{1}_m^\mathsf{T} V)^\mathsf{T}) \in VB_m$ *and* $\det(V_{(:,1:m)}) \neq 0$. *Then,* $p \in \mathbb{R}^n$ *defined by*

$$p_k := \|V_{(:,1:m)}^{-1} V_{(:,\{k\})}\|_{\ell_1},\ k \in (1:n)$$

is unimodal .

Proof outline

- By definition $p_{(1:m)} = \mathbf{1}_m$, does this implies $\mathsf{S}(\Delta p_{(m:n)}) \leq 1$.

Proof outline

- By definition $p_{(1:m)} = \mathbf{1}_m$, does this implies $\mathsf{S}(\Delta p_{(m:n)}) \leq 1$.
- Define

$$W := V^{\mathsf{T}} \quad \text{and} \quad Q := W\,W_{(1:m,:)}^{-1} K_m.$$

Proof outline

- By definition $p_{(1:m)} = \mathbf{1}_m$, does this implies $\mathsf{S}(\Delta p_{(m:n)}) \leq 1$.
- Define
$$W := V^{\mathsf{T}} \quad \text{and} \quad Q := W\, W_{(1:m,:)}^{-1} K_m.$$
- Because $W \in \mathsf{SC}_m$, then from TP properties $\Rightarrow Q_{(m+1:n,:)}$ is totally positive.

Proof outline

- By definition $p_{(1:m)} = \mathbf{1}_m$, does this implies $\mathsf{S}(\Delta p_{(m:n)}) \leq 1$.
- Define
$$W := V^{\mathsf{T}} \quad \text{and} \quad Q := W\, W^{-1}_{(1:m,:)} K_m.$$
- Because $W \in \mathsf{SC}_m$, then from TP properties $\Rightarrow Q_{(m+1:n,:)}$ is totally positive.
- $\Delta \mathbf{1}_m^{\mathsf{T}} W \in \mathsf{VB}_m \Rightarrow$
$$\mathsf{S}(\Delta Q\, \mathbf{1}_m) = \mathsf{S}(\Delta W\, W^{-1}_{(1:m,:)} K_m\, \mathbf{1}_m) \leq m - 1.$$

Proof outline

- By definition $p_{(1:m)} = \mathbf{1}_m$, does this implies $\mathsf{S}(\Delta p_{(m:n)}) \leq 1$.
- Define
$$W := V^{\mathsf{T}} \quad \text{and} \quad Q := W\, W_{(1:m,:)}^{-1} K_m.$$
- Because $W \in \mathsf{SC}_m$, then from TP properties $\Rightarrow Q_{(m+1:n,:)}$ is totally positive.
- $\Delta \mathbf{1}_m^{\mathsf{T}} W \in \mathsf{VB}_m \Rightarrow$
$$\mathsf{S}(\Delta Q\, \mathbf{1}_m) = \mathsf{S}(\Delta W\, W_{(1:m,:)}^{-1} K_m\, \mathbf{1}_m) \leq m-1.$$
- The first m rows of $\Delta Q\, \mathbf{1}_m$ contribute $m-2$ sign changes
$$(\Delta Q)_{(1:m,:)}\, \mathbf{1}_m = \Delta K_m\, \mathbf{1}_m = \begin{bmatrix} \vdots \\ 2 \\ -2 \\ 2 \end{bmatrix}.$$

Proof outline

- By definition $p_{(1:m)} = \mathbf{1}_m$, does this implies $\mathsf{S}(\Delta p_{(m:n)}) \le 1$.
- Define
$$W := V^{\mathsf{T}} \quad \text{and} \quad Q := W\, W^{-1}_{(1:m,:)} K_m.$$
- Because $W \in \mathsf{SC}_m$, then from TP properties $\Rightarrow Q_{(m+1:n,:)}$ is totally positive.
- $\Delta \mathbf{1}_m^{\mathsf{T}} W \in \mathsf{VB}_m \Rightarrow$
$$\mathsf{S}(\Delta Q\, \mathbf{1}_m) = \mathsf{S}(\Delta W\, W^{-1}_{(1:m,:)} K_m\, \mathbf{1}_m) \le m - 1.$$
- The first m rows of $\Delta Q\, \mathbf{1}_m$ contribute $m - 2$ sign changes
$$(\Delta Q)_{(1:m,:)}\, \mathbf{1}_m = \Delta K_m\, \mathbf{1}_m = \begin{bmatrix} \vdots \\ 2 \\ -2 \\ 2 \end{bmatrix}.$$
- $p_{(m+1:n)} = Q_{(m+1:n,:)}, \Rightarrow \mathsf{S}(\Delta p(m:n)) \le 1.$

Proof outline

- By definition $p_{(1:m)} = \mathbf{1}_m$, does this implies $\mathsf{S}(\Delta p_{(m:n)}) \leq 1$.
- Define
$$W := V^{\mathsf{T}} \quad \text{and} \quad Q := W\, W_{(1:m,:)}^{-1} K_m.$$
- Because $W \in \mathsf{SC}_m$, then from TP properties $\Rightarrow Q_{(m+1:n,:)}$ is totally positive.
- $\Delta \mathbf{1}_m^{\mathsf{T}} W \in \mathsf{VB}_m \Rightarrow$
$$\mathsf{S}(\Delta Q\, \mathbf{1}_m) = \mathsf{S}(\Delta W\, W_{(1:m,:)}^{-1} K_m\, \mathbf{1}_m) \leq m - 1.$$
- The first m rows of $\Delta Q\, \mathbf{1}_m$ contribute $m - 2$ sign changes
$$(\Delta Q)_{(1:m,:)}\, \mathbf{1}_m = \Delta K_m\, \mathbf{1}_m = \begin{bmatrix} \vdots \\ 2 \\ -2 \\ 2 \end{bmatrix}.$$
- $p_{(m+1:n)} = Q_{(m+1:n,:)}, \Rightarrow \mathsf{S}(\Delta p(m:n)) \leq 1$.
- Finally, the single sign change in $\Delta p(m:n)$ must be due to a negative entry, which implies that p is unimodal.

Corollary

Corollary

Let $V \in \mathbb{R}^{m \times m}$ defined as in the previous theorem. Further, let $T \in \mathbb{R}^{m \times m}$ be invertible, $\bar{V} = TV$ and

$$p(k) := \|\mathbf{1}_m^\top \bar{V}_{(:,1:m)}^{-1} \bar{V}_{(:,m+k)}\|_{\ell_1}$$

for all $k \geq 1$. Then, the sequence $\{p(k)\}_{k \geq 1}$ is unimodal.

- In other words: mpulse response is realization independent
- Applies to Hankel, Page, Toeplitz matrices, ...

Log-Concavity

Theorem

Let $A \in \mathbb{R}^{m\times m}$ and $b \in \mathbb{R}^m$ be such that A is diagonalizable and (A, b) controllable. Further, let $V = \mathcal{C}(A, b) \in SC_m$ and

$$g_{(i)}(k) := W_k^T e_i, \; i \in (1:m)$$
$$p(k) := W_k^T \mathbf{1}_m$$

where

$$W_k := K_m^T V_{(:,1:m)}^{-1} V_{(:,m+k)}, \; k \geq 1.$$

Then, the sequences $\{g_{(i)}(k)\}_{k\geq 1}$ for each $i \in (1:m)$ and $\{p(k)\}_{k\geq 1}$ are log-concave.

Log-Concavity

Theorem

Let $A \in \mathbb{R}^{m\times m}$ and $b \in \mathbb{R}^m$ be such that A is diagonalizable and (A, b) controllable. Further, let $V = \mathcal{C}(A, b) \in SC_m$ and

$$g_{(i)}(k) := W_k^T e_i,\ i \in (1:m)$$
$$p(k) := W_k^T \mathbf{1}_m$$

where

$$W_k := K_m^T V_{(:,1:m)}^{-1} V_{(:,m+k)},\ k \geq 1.$$

Then, the sequences $\{g_{(i)}(k)\}_{k\geq 1}$ for each $i \in (1:m)$ and $\{p(k)\}_{k\geq 1}$ are log-concave.

Hidden Gem: The transformation induced by W_k yields the companion matrix of A.

Characteristic Polynomial

Theorem

Let $A \in \mathbb{R}^{m\times m}$ *have characteristic polynomial*

$$\det(sI - A) \;=\; s^m \;+\; \alpha_{m-1}s^{m-1} \;+\; \cdots + \alpha_1 s + \alpha_0,$$

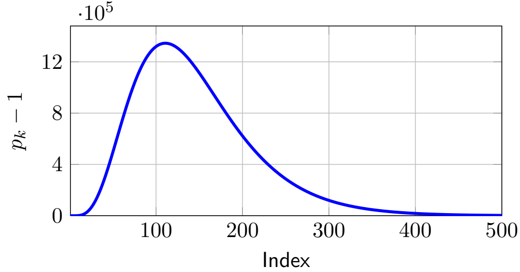
with $\sum_{i=0}^{m-1} |\alpha_i| < 1$, *and let* $b \in \mathbb{R}^m$ *be such that the pair* (A, b) *is controllable. Then, for the controllability matrix* $V = \mathcal{C}(A, b)$ *it holds*

$$\left\| V^{-1}_{(:,1:m)}\, V_{(:,\,m+k)} \right\|_{\ell_1} \;<\; 1 \qquad \text{for all } k \geq 0.$$

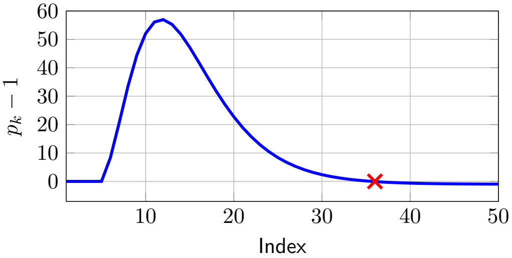
Table of Contents

Sparse minimization (G)

$$V = \begin{bmatrix} b & Ab & \dots & A^{N-1}b \end{bmatrix} \begin{bmatrix} u(N-1) & \dots & u(0) \end{bmatrix}^{\mathsf{T}}$$



$$A = \text{diag}\left(\begin{bmatrix} 0.98 & 0.97 & 0.96 & 0.95 & 0.94 \end{bmatrix}\right)$$



$$A = \text{diag}\left(\begin{bmatrix} 0.8 & 0.7 & 0.6 & 0.5 & 0.4 \end{bmatrix}\right)$$

Table of Contents

Concluding remarks

- Deterministic framework that provides failure guarantees for basis pursuit for special matrix structures.
- Complements (conservative) success guarantees
- New avenue for study of sparse/low-rank optimization problems
- Future work: sparsity in terms of singular values and rank minimization.