QUADRATIC ENVELOPE REGULARIZATION FOR STRUCTURED LOW RANK APPROXIMATION.

Marcus Carlsson

May 10, 2019

Examples of uses:

Low rank approximation.

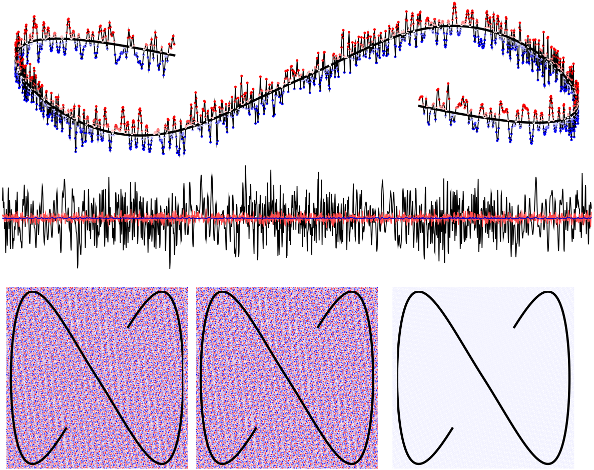


Figure: Frequency estimation experiment.

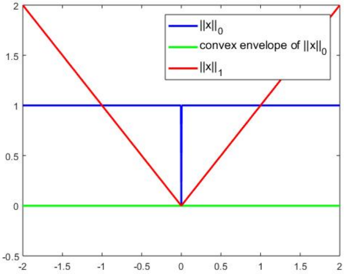
Compressed sensing

Problem 1; minimize $\|x\|_0$ given $Ax = d$. (Matrix A and measurement d known.)

Problem 2; minimize $\text{rank}(X)$ given $\mathcal{A}X = d$. (Operator $\mathcal{A}$ and measurement d known.)

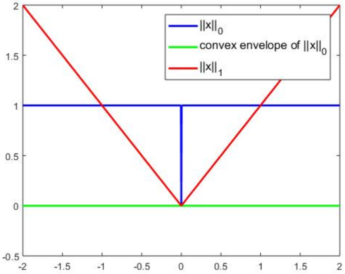
$x \mapsto \|x\|_0$ and $X \mapsto \text{rank}(X)$ are highly discontinuous and non-convex. Convex envelopes are identically 0.

Popular approach; replace $\|x\|_0$ by $\|x\|_1$, replace $\text{rank}(X)$ with $\|X\|_{nuc}$.



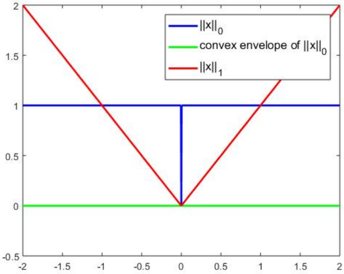
Compressed sensing

Problem 1; minimize $\|x\|_0$ given $Ax = d$. (Matrix A and measurement d known.)
Problem 2; minimize $\text{rank}(X)$ given $\mathcal{A}X = d$. (Operator $\mathcal{A}$ and measurement d known.)
$x \mapsto \|x\|_0$ and $X \mapsto \text{rank}(X)$ are highly discontinuous and non-convex. Convex envelopes are identically 0.
Popular approach; replace $\|x\|_0$ by $\|x\|_1$, replace $\text{rank}(X)$ with $\|X\|_{nuc}$.



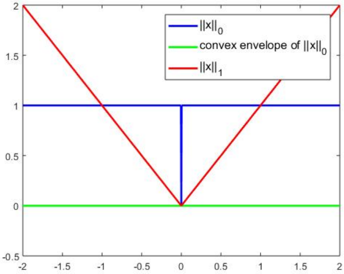
Compressed sensing

Problem 1; minimize $\|x\|_0$ given $Ax = d$. (Matrix A and measurement d known.)
Problem 2; minimize $\text{rank}(X)$ given $\mathcal{A}X = d$. (Operator $\mathcal{A}$ and measurement d known.)
$x \mapsto \|x\|_0$ and $X \mapsto \text{rank}(X)$ are highly discontinuous and non-convex. Convex envelopes are identically 0.
Popular approach; replace $\|x\|_0$ by $\|x\|_1$, replace $\text{rank}(X)$ with $\|X\|_{nuc}$.



Compressed sensing

Problem 1; minimize $\|x\|_0$ given $Ax = d$. (Matrix A and measurement d known.)
Problem 2; minimize $\text{rank}(X)$ given $\mathcal{A}X = d$. (Operator $\mathcal{A}$ and measurement d known.)
$x \mapsto \|x\|_0$ and $X \mapsto \text{rank}(X)$ are highly discontinuous and non-convex. Convex envelopes are identically 0.
Popular approach; replace $\|x\|_0$ by $\|x\|_1$, replace $\text{rank}(X)$ with $\|X\|_{nuc}$.



Oracle solution

We consider vector case first. Let x_0 be sparse, let d be given by $Ax_0 + noise$. Let T be the support of x_0 and A_T the matrix obtained by deleting all rows in A outside T. Then x_S is obtained by solving

$$d = A_T x.$$

$x_S \neq x_0$ but is the best possible!

Finding the "oracle solution" x_S

Two methods based on the quadratic envelope applied to the vector sparse recovery problem.

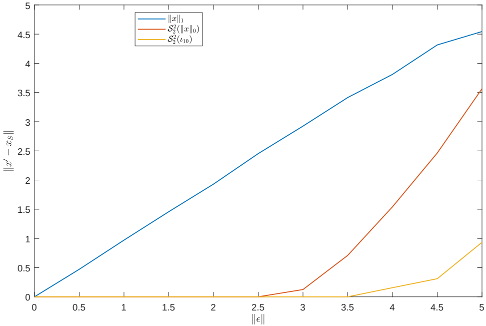


Figure: Distance to oracle solution x_S as a function of noise. Here $\|x_0\| \approx 10$ and A is 100x200 with normalized random columns.

For every sparse vector method there is a corresponding low rank method for matrices!

Finding the "oracle solution" x_S

Two methods based on the quadratic envelope applied to the vector sparse recovery problem.

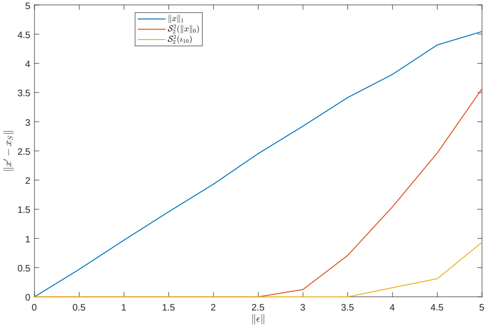


Figure: Distance to oracle solution x_S as a function of noise. Here $\|x_0\| \approx 10$ and A is $100x200$ with normalized random columns.

For every sparse vector method there is a corresponding low rank method for matrices!

Dual formulation

Solve

$$\arg\min_x \quad \lambda \|x\|_0 + \frac{1}{2}\|Ax - d\|_2^2$$

(impossible) or convex problem

$$\arg\min_x \quad \lambda \|x\|_1 + \frac{1}{2}\|Ax - d\|_2^2$$

where λ is a parameter.

Dual formulation

Solve

$$\arg\min_{x} \quad \lambda \|x\|_0 + \frac{1}{2}\|Ax - d\|_2^2$$

(impossible) or convex problem

$$\arg\min_{x} \quad \lambda \|x\|_1 + \frac{1}{2}\|Ax - d\|_2^2$$

where λ is a parameter.

Similar but different problems

Consider

$$\underset{X}{\arg\min} \quad \operatorname{rank}(X) + \frac{1}{2}\|X - D\|_F^2 \tag{1}$$

where $X \in \mathbb{M}_{m,n}$ (the space of $m \times n$-matrices with the Frobenius norm). $\operatorname{rank}(X)$ can be replaced by e.g.

$$\iota_K(X) = \begin{cases} 0 & \operatorname{rank}(X) \leq K, \\ \infty & \text{else.} \end{cases} \tag{2}$$

Maybe we only know scalar products $\langle X, Y_j \rangle \approx d_j, j = 1 \dots m$. This leads to

$$\underset{X}{\arg\min} \quad \operatorname{rank}(X) + \frac{1}{2}\|\mathcal{A}(X) - d\|_F^2 \tag{3}$$

Similar but different problems

Consider

$$\arg\min_X \quad \text{rank}(X) + \frac{1}{2}\|X - D\|_F^2 \tag{1}$$

where $X \in \mathbb{M}_{m,n}$ (the space of $m \times n$-matrices with the Frobenius norm). rank(X) can be replaced by e.g.

$$\iota_K(X) = \begin{cases} 0 & \text{rank}(X) \leqslant K, \\ \infty & \text{else.} \end{cases} \tag{2}$$

Maybe we only know scalar products $\langle X, Y_j \rangle \approx d_j, j = 1 \ldots m$. This leads to

$$\arg\min_X \quad \text{rank}(X) + \frac{1}{2}\|\mathcal{A}(X) - d\|_F^2 \tag{3}$$

Similar but different problems

Consider

$$\arg\min_X \quad \text{rank}(X) + \frac{1}{2}\|X - D\|_F^2 \tag{1}$$

where $X \in \mathbb{M}_{m,n}$ (the space of $m \times n$-matrices with the Frobenius norm). $\text{rank}(X)$ can be replaced by e.g.

$$\iota_K(X) = \begin{cases} 0 & \text{rank}(X) \leqslant K, \\ \infty & \text{else.} \end{cases} \tag{2}$$

Maybe we only know scalar products $\langle X, Y_j \rangle \approx d_j, j = 1 \dots m$. This leads to

$$\arg\min_X \quad \text{rank}(X) + \frac{1}{2}\|\mathcal{A}(X) - d\|_F^2 \tag{3}$$

General question

Given a problem

$$\arg\min_x \quad f(x) + \frac{1}{2}\|\mathcal{A}x - d\|^2,$$

where x is in some linear space, how to replace f to get a convex (or at least continuous) functional with same or similar global minima?

General setting:

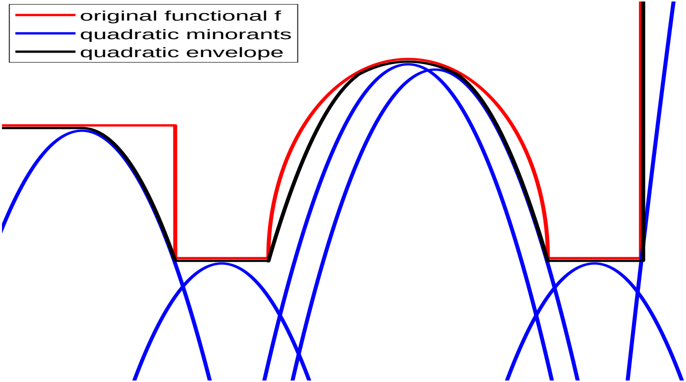
$$f(x) + \frac{1}{2}\|x - d\|_{\mathcal{V}}^2 \tag{4}$$

$\mathcal{V}$- Hilbert space, f -bounded below.
"New" transform, quadratic envelope $\mathcal{Q}(f)$. The (l.s.c) convex envelope of the functional in (4) is

$$\mathcal{Q}(f)(x) + \frac{1}{2}\|x - d\|_{\mathcal{V}}^2. \tag{5}$$

Note: form independent of data d.

Illustration:



Example 1: $\mathcal{Q}$ of cardinality ($CE\ell 0$)

$$\mathcal{Q}(\|\cdot\|_0)(x) = 1 - \left(\max\{1 - \frac{|x|}{\sqrt{2}}, 0\}\right)^2$$

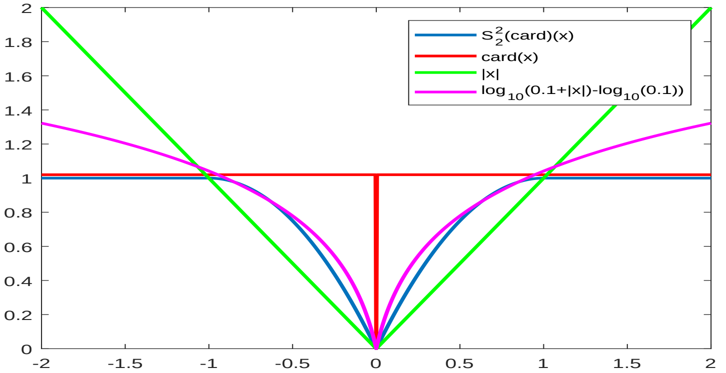
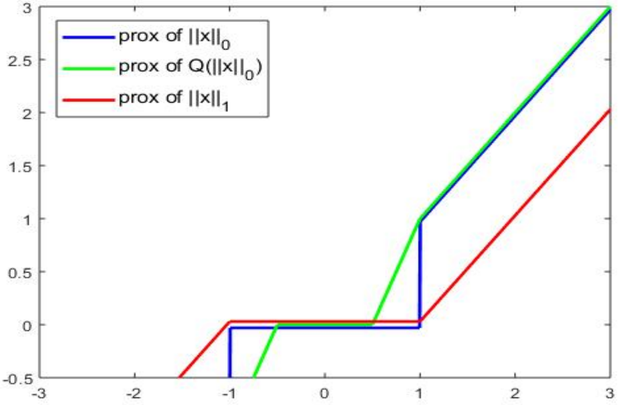
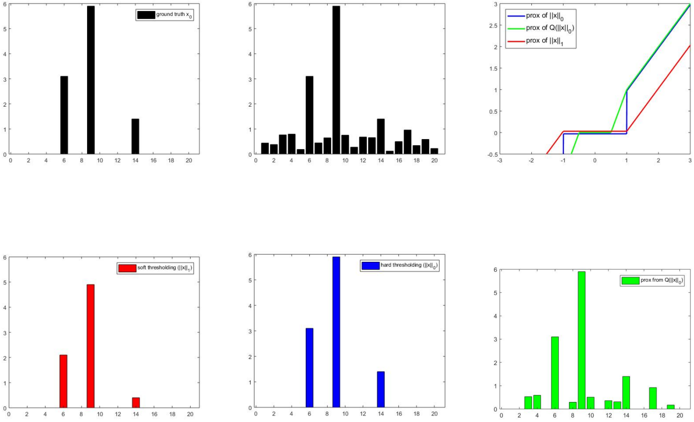


Figure: Illustration of $\|x\|_0$ (red) along with its $\mathcal{Q}$-transform (blue).

The proximal operator



Toy example


ground truth x_0
prox of $||x||_0$
prox of $Q(||x||_0)$
prox of $||x||_1$
soft thresholding ($||x||_1$)
hard thresholding ($||x||_0$)
prox from $Q(||x||_0)$

Examples 2, known rank K:

$$f_K(X) = \begin{cases} 0 & \text{rank}(X) \leqslant K, \\ \infty & \text{else.} \end{cases}$$

$$\mathcal{Q}(f_K)(x) = \frac{1}{2k_*}\left(\sum_{j>K-k_*} |\sigma_j|\right)^2 - \frac{1}{2}\sum_{j>K-k_*} |\sigma_j|^2$$

where k_* is a particular number between 1 and K and σ the singular values.

Use of $\mathcal{Q}(f)$:

Two types, different challenges:

$$\arg\min_{x} \quad f(x) + \frac{1}{2}\|Ax - d\|_2^2 \tag{6}$$

(compressed sensing)

$$\arg\min_{x \in \mathcal{M}} \quad f(x) + \frac{1}{2}\|x - d\|_2^2 \tag{7}$$

(structured low rank approximation)

(6): convex envelope not computable.

(7): convex envelope

$$\mathcal{Q}(f)(x) + \frac{1}{2}\|x - d\|_F^2 \tag{8}$$

Use of $\mathcal{Q}(f)$:

Two types, different challenges:

$$\operatorname*{arg\,min}_{x} \quad f(x) + \frac{1}{2}\|Ax - d\|_2^2 \tag{6}$$

(compressed sensing)

$$\operatorname*{arg\,min}_{x \in \mathcal{M}} \quad f(x) + \frac{1}{2}\|x - d\|_2^2 \tag{7}$$

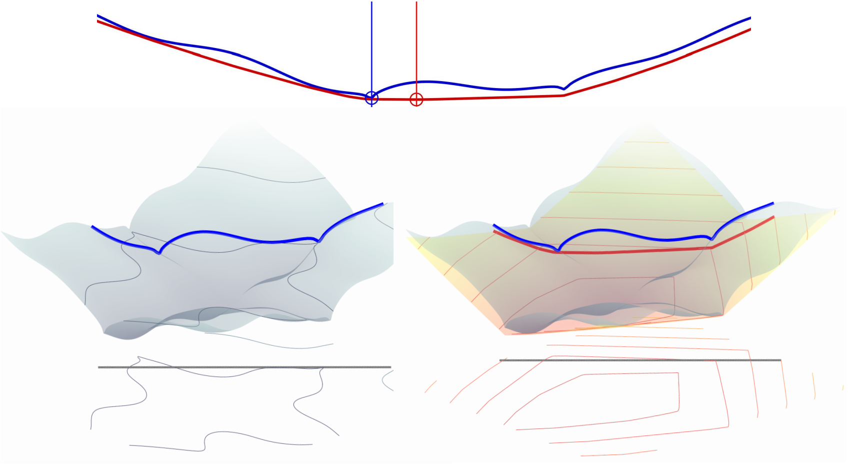
(structured low rank approximation)
(6): convex envelope not computable.
(7): convex envelope

$$\mathcal{Q}(f)(x) + \frac{1}{2}\|x - d\|_F^2 \tag{8}$$

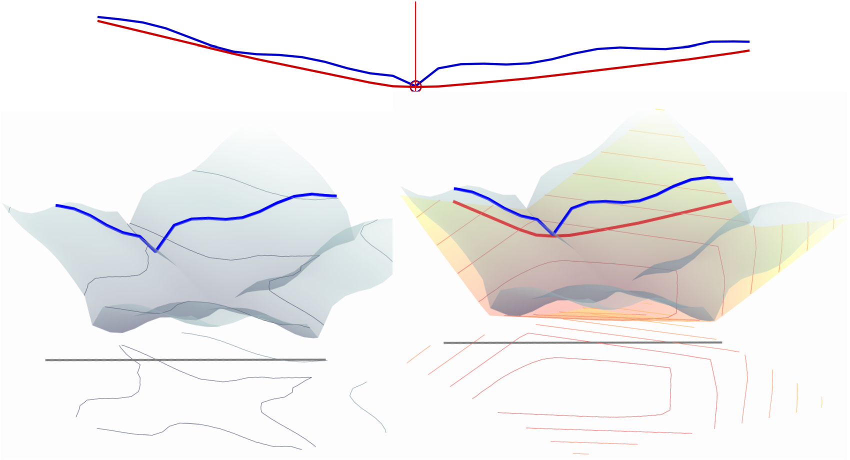
$\mathcal{M}$: subspace. Consider

$$\underset{x \in \mathcal{M}}{\arg\min} \quad \mathcal{Q}(f)(x) + \frac{1}{2}\|x - d\|^2$$

No A, minimization over subspace



No A, minimization over subspace



Matrix A in the l^2-misfit

$$\arg\min_x \quad f(x) + \frac{1}{2}\|Ax - d\|_{\mathcal{W}}^2 \tag{9}$$

Note that $\mathcal{Q}(f)(x) + \frac{1}{2}\|x - d\|_{\mathcal{V}}^2$ is the convex envelope of $f(x) + \frac{1}{2}\|x - d\|_{\mathcal{V}}^2$.
Can we use

$$\mathcal{Q}(f)(x) + \frac{1}{2}\|Ax - d\|_{\mathcal{W}}^2?$$

By rescaling, we may assume that $A^*A \leqslant I$ or $A^*A \geqslant I$.

Matrix A in the l^2-misfit

$$\arg\min_x \quad f(x) + \frac{1}{2}\|Ax - d\|_{\mathcal{W}}^2 \tag{9}$$

Note that $\mathcal{Q}(f)(x) + \frac{1}{2}\|x - d\|_{\mathcal{V}}^2$ is the convex envelope of $f(x) + \frac{1}{2}\|x - d\|_{\mathcal{V}}^2$.

Can we use

$$\mathcal{Q}(f)(x) + \frac{1}{2}\|Ax - d\|_{\mathcal{W}}^2?$$

By rescaling, we may assume that $A^*A \leqslant I$ or $A^*A \geqslant I$.

Matrix A in the l^2-misfit

$$\operatorname*{arg\,min}_x \quad f(x) + \frac{1}{2}\|Ax - d\|_{\mathcal{W}}^2 \tag{9}$$

Note that $\mathcal{Q}(f)(x) + \frac{1}{2}\|x - d\|_{\mathcal{V}}^2$ is the convex envelope of $f(x) + \frac{1}{2}\|x - d\|_{\mathcal{V}}^2$.
Can we use

$$\mathcal{Q}(f)(x) + \frac{1}{2}\|Ax - d\|_{\mathcal{W}}^2?$$

By rescaling, we may assume that $A^*A \leqslant I$ or $A^*A \geqslant I$.

Case $A^*A \geqslant I$

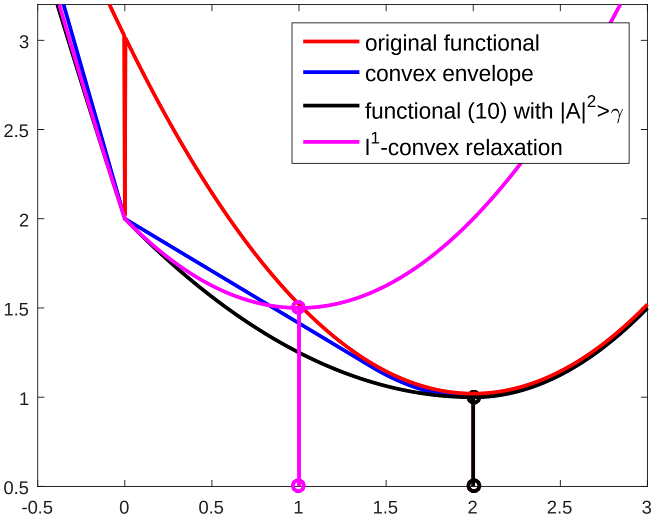


Figure: Illustration. Left graph $A = 2$ and $d = 1$.

Case $A^*A \leqslant I$

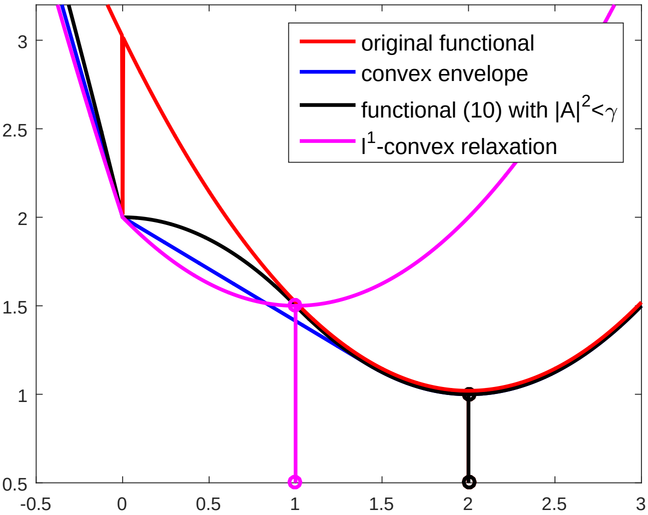


Figure: Illustration. $A = 1/2$, $d = 1$.

Observations

$$\mathcal{Q}(f)(x) + \frac{1}{2}\|Ax - d\|_{\mathcal{W}}^2$$

is not convex. Minima does not move.

And often are fewer in number.

Observations

$$\mathcal{Q}(f)(x) + \frac{1}{2}\|Ax - d\|_{\mathcal{W}}^2$$

is not convex. Minima does not move.
And often are fewer in number.

Key theorem

Theorem

Suppose that $\|A\| < 1$ and that f is a weakly l.s.c $[0,\infty]$-valued functional. Set

$$g(x) = f(x) + \frac{1}{2}\|Ax - d\|_{\mathcal{W}}^2$$

and

$$h(x) = \mathcal{Q}(f)(x) + \frac{1}{2}\|Ax - d\|_{\mathcal{W}}^2.$$

If x is a local minimizer of h, then it is also a local minimizer of g, and $h(x) = g(x)$. In particular the global minimizers coincide.

Key theorem

Theorem

Suppose that $\|A\| < 1$ and that f is a weakly l.s.c $[0, \infty]$-valued functional. Set

$$g(x) = f(x) + \frac{1}{2}\|Ax - d\|_{\mathcal{W}}^2$$

and

$$h(x) = \mathcal{Q}(f)(x) + \frac{1}{2}\|Ax - d\|_{\mathcal{W}}^2.$$

If x is a local minimizer of h, then it is also a local minimizer of g, and $h(x) = g(x)$. In particular the global minimizers coincide.

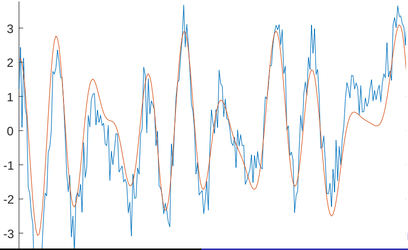
ICASSP paper

:

- We provide formulas for working with $\mathcal{Q}$ and tensor weights, which increases the flexibility and uses of $\mathcal{Q}$.
- We study low rank Hankel approximation by minimizing variants of

$$\mathcal{Q}(\mu\text{rank})(X) + \|X - D\|_2^2$$

over the space of Hankel matrices.
- Application; frequency estimation when number of frequencies K, a.k.a. model order, is known.



Numerical test, correct model order

Name	$\|y - d\|$	Conv. Rate	Rank
$\mathcal{Q}$-weights	11.40	315	8
HC-weights	11.67	429	8
$\mathcal{Q}$-standard	11.96	86	8
HC-standard	11.96	86	8
ESPRIT	12.00	1	8
nuclear	13.18	iter$*$92	8

Table: Performance with $SNR = 5$ and $K = 8$. Tests use a function d with 8 exponential functions plus noise (noise level $\|\varepsilon\| = 12.5$), see Figure 6. Second column displays distance to input signal, interesting to note is that most methods beat "ground truth" (i.e. 12.5). The third column displays the amount of iterations for each method, where *iter* is the the number of times "nuclear" needs to be repeated to find a suitable λ.

Numerical test, wrong model order

Name	$\|y-d\|$	Conv. Rate	Rank
$\mathcal{Q}$-weights	23.23	288	5
HC-weights	26.69	352	5
$\mathcal{Q}$-standard	21.56	150	6
HC-standard	23.72	diverges	5
ESPRIT	40.43	1	5
nuclear	26.71	iter$*$236	5

Table: Same as before but $K = 5$.

The code is available at:
https://github.com/J-Son89/QuadraticEnvelopeTransform

References compressed sensing techniques:

- Quadratic envelope: Carlsson, Marcus. "On Convex Envelopes and Regularization of Non-Convex Functionals without moving Global Minima." To appear in JOTA, arXiv:1811.03439 (2018).
- Finding the oracle solution:Carlsson, Marcus, Daniele Gerosa, and Carl Olsson. "An unbiased approach to compressed sensing." arXiv:1806.05283 (2018).
- Crossover method: Olsson, Carl, Marcus Carlsson, and Daniele Gerosa. "Bias Reduction in Compressed Sensing." arXiv preprint arXiv:1812.11329 (2018).
- Oracle solution for matrix problems: Work in progress
- Optimization with linear constraints: Caprani, Jamie, and Marcus Carlsson. "Quadratic Envelope Regularization for Structured Low Rank Approximation." ICASSP, IEEE, 2019.

marcus.carlsson@math.lu.se

References frequency estimation:

- Working with irregular grids: Andersson, F., Carlsson, M., Tourneret, J. Y., Wendt, H. "A new frequency estimation method for equally and unequally spaced data". IEEE Transactions on Signal Processing, 62(21), 5761-5774.
- ESPRIT in several variables, strange domains: ""Andersson, Fredrik, and Marcus Carlsson. "ESPRIT for multidimensional general grids." SIAM Journal on Matrix Analysis and Applications 39.3 (2018): 1470-1488.
- Quadratic envelopes and SLRA: Andersson, Fredrik, and Marcus Carlsson. "Fixed-point algorithms for frequency estimation and structured low rank approximation." Applied and Computational Harmonic Analysis 46.1 (2019): 40-65.

marcus.carlsson@math.lu.se