# QUADRATIC ENVELOPE REGULARIZATION FOR STRUCTURED LOW RANK APPROXIMATION

*Jamie Caprani, Marcus Carlsson*

Centre for Mathematical Sciences, Lund University, Box 118, 22100 Lund, Sweden

**ABSTRACT**

Compressed sensing techniques, such as nuclear norm minimization, can be used for structured low rank approximation, but it is well known that these methods lead to suboptimal results. In this article we consider how to improve this approach by use of so called "quadratic envelopes". The new feature is the extension to weighted matrix spaces relying on tensors, and we show how this can be used for improved accuracy of complex frequency estimation methods.

***Index Terms*—** Complex frequency estimation, compressed sensing, convex envelopes.

## 1. INTRODUCTION

For concreteness we begin with the familiar Low Rank Hankel approximation problem/Complex Frequency Estimation Problem. Let $\mathbb{M}_{m,n}$ be the set of $m \times n$ matrices and let $\mathcal{H}$ be the linear subspace of Hankel matrices. Given a sequence $y$ we denote by $H_y$ the corresponding Hankel matrix. We consider the classical problem of finding a matrix $X = H_y$ of fixed (a priori known) rank $K$ on $\mathcal{H}$ that minimizes the distance $\|y-d\|$ to some fixed measurement $d$. The routine way of approaching this problem via compressed sensing techniques is to solve

$$\underset{X=H_y}{\arg\min}\ \lambda\|X\|_{nuc} + \frac{1}{2}\|y-d\|^2, \tag{1}$$

where $\lambda$ is a parameter which has to be tuned until the right rank is found [1, 2]. Beyond the drawback with having to tune $\lambda$, it is well known within the community that this approach leads to a bias [3].

On the other extreme, it is known that ADMM works fairly well on non-convex problems, and hence one may try to minimize e.g.

$$\underset{X=H_y}{\arg\min}\ \iota_{\mathcal{M}_K}(X) + \frac{1}{2}\|y-d\|^2 \tag{2}$$

using ADMM, where $\iota_{\mathcal{M}_K}$ is the indicator functional of matrices with rank $\leq K$ (i.e. $\iota_{\mathcal{M}_K}(X)$ equals 0 if

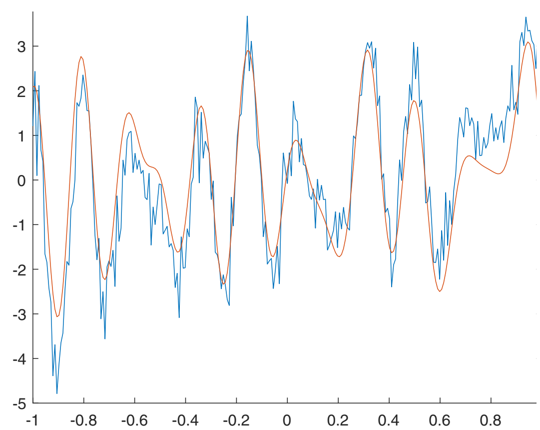


**Fig. 1**. Illustration to the example in Section 6.

$\mathsf{rank}(X) \leq K$ and $\infty$ else). Successful performance in comparison with (1), (as well as a number of competing methods), was reported in [4]. Moreover, upon trading the term $\|y-d\|^2$ for $\|A(y)-d\|^2$, where $A$ is an interpolation operator, it was shown that this method can handle missing data and unequally spaced sampling. However, there are no guarantees that this method will converge, and for difficult problems we have indeed observed that the corresponding algorithm may diverge.

To find a balance between these extremes, we have in a series of articles developed a machinery to convexify non-convex problems, which led to quadratic envelopes and the so called $\mathcal{Q}_\gamma$-transform [5, 6]. We introduce this in a general framework, since everything in this paper applies just as well if $\mathcal{H}$ is any linear subspace of $\mathbb{M}_{m,n}$ and $y \mapsto H_y$ is a parametrization of this subspace. Moreover the functional $\iota_{\mathcal{M}_K}$ can be swapped for many other penalties, such as $\mathsf{rank}(X)$, or even any general functional $f$ on some Hilbert space $\mathcal{V}$. We have chosen to focus on the Hankel matrix/frequency estimation problem mainly for concreteness.

978-1-5386-4658-8/18/$31.00 ©2019 IEEE

## 2. THE QUADRATIC ENVELOPE

Given any functional $f$ on a Hilbert space, consider the set $S$ of all quadratic functionals that satisfy $\alpha - \frac{\gamma}{2}\|\cdot - y\|^2 \leq f$, where $\gamma > 0$ is a parameter. The $\mathcal{Q}_\gamma$-transform is defined as

$$\mathcal{Q}_\gamma(f)(x) = \sup_{\alpha\in\mathbb{R}, y\in\mathcal{V}} \{\alpha - \frac{\gamma}{2}\|x-y\|^2 \in S.\}$$

It is designed such that $\mathcal{Q}_\gamma(f)(x) + \frac{\gamma}{2}\|x\|^2$ is the l.s.c. convex envelope of $f(x) + \frac{\gamma}{2}\|x\|^2$. The transform has a number of desirable features, we refer to [5, 6] for details and figures. Most notably $\mathcal{Q}_\gamma(f)$ is continuous, satisfies $0 \leq \mathcal{Q}_\gamma(f) \leq f$ and it often happens that $\mathcal{Q}_\gamma(f)(x) = f(x)$, which makes $\mathcal{Q}_\gamma(f)$ suitable as a regularizer. Also, for unconstrained minimization problems of the type

$$\arg\min\ f(x) + \frac{1}{2}\|A(x) - d\|^2$$

it is shown in [6] that regularizing with $\mathcal{Q}_\gamma$, i.e. swapping $f$ for $\mathcal{Q}_\gamma(f)$ has the very nice property that global minima coincide, as long as $\gamma > \|A\|^2$. See [7] for applications sparse (vector)-estimation.

## 3. SLRA AND $\mathcal{Q}_\gamma$

Let $c$ be a convex functional on $\mathcal{V}$ incorporating prior information known about the problem in question, and suppose we wish to minimize

$$\arg\min_x\ f(x) + \frac{1}{2}\|x - d\|^2 + c(x). \qquad (3)$$

For concrete examples of this form we refer e.g. to the overview article [8], or Section 4 of [3] which contains applications to structure from motion and system identification. In particular, we can take $c$ as the indicator function of some closed convex subset $\mathcal{H}$ of $\mathcal{V}$, i.e. $c = \iota_\mathcal{H}$, to retrieve the problem (2).

We suppose now that (3) does not have a closed form solution, and consider replacing it by

$$\arg\min_x\ \mathcal{Q}_\gamma(f)(x) + \frac{1}{2}\|x - d\|^2 + c(x) \qquad (4)$$

to obtain a strongly convex problem (for $\gamma < 1$, just convex if $\gamma = 1$). The key result about this regularization, which is a modification of Theorem 5.1 in [6] (or Theorem 3.1 of [5]), reads as follows;

**Proposition 3.1.** *Let $f$ be a $[0,\infty]$-valued functional on a separable Hilbert space $\mathcal{V}$, and let $c \geq 0$ be a l.s.c. convex function such that* $\mathsf{dom}f \cap \mathsf{dom}c \neq \emptyset$. *Given $\gamma < 1$, the functional in* (4) *is strongly convex and supercoercive. The solution is thus a unique point $\hat{x}$, which solves* (3) *whenever $\mathcal{Q}_\gamma(f)(\hat{x}) = f(\hat{x})$.*

Most noteworthy, by the design of the $\mathcal{Q}_\gamma$-transform it frequently happens that $\mathcal{Q}_\gamma(f)(x) = f(x)$, and this is easy to check upon convergence, so one of the benefits of the proposed method is that it often solves the non-convex problem (3). The rationale behind replacing (3) by (4) is pragmatical; since the latter is convex the solution may be found using convex optimization routines. This may seem ad hoc but we remind the reader that replacing e.g. $\mathsf{rank}(X)$ by the nuclear norm $\|X\|_*$ has had a substantial impact, and that for these concrete cases the modification $\mathcal{Q}_\gamma(f)$ is much closer to the original functional $f$. The proposed framework is also more flexible, since we can work e.g. with $\mathcal{Q}_\gamma(\mathsf{rank}(X))$ or $\mathcal{Q}_\gamma(\iota_{\mathcal{M}_K})$ depending on whether we have a priori information about the model order. We refer to [9] for more information on $\mathcal{Q}_\gamma(\iota_{\mathcal{M}_K})$ and to [5] for many other possible penalties $f$. For a variety of "Hankel-type" subspaces used in multidimensional frequency estimation, see [10].

## 4. $\mathcal{Q}_\gamma$-TRANSFORMS IN WEIGHTED SPACES

It is sometimes desirable to work in weighted spaces. In this section we show how this can be done for a particular class of weights. In order to find a framework that works for both vectors and matrices we consider the more general situation of tensor products. Given $\mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{N}^d$ we let $\mathbb{M}_\mathbf{n}$ denote the set of tensors $X = (x_\mathbf{j})_{\mathbf{1}\leq\mathbf{j}\leq\mathbf{n}}$ where $\mathbf{1} \leq \mathbf{j} \leq \mathbf{n}$ means that $1 \leq j_i \leq n_i$ for all $1 \leq i \leq d$.

Given $W \in \mathbb{M}_\mathbf{n}$ with positive entries, we let $\mathbb{M}_\mathbf{n}^W$ be the Hilbert space obtained by introducing the norm

$$\|X\|_W^2 = \sum_\mathbf{j} w_\mathbf{j}|x_\mathbf{j}|^2.$$

In the case $W = \mathbf{1}$, i.e. when $W$ is equal to one componentwise, we will simply write $\mathbb{M}_\mathbf{n}$ as earlier, and the corresponding norm is written $\|X\|$ as opposed to $\|X\|_{\mathbb{M}_\mathbf{n}^\mathbf{1}}$. Suppose now that we are interested in computing $\mathcal{Q}_\gamma(f)$ in the weighted space $\mathbb{M}_{m,n}^W$, where $f$ is such that $\mathcal{Q}_\gamma(f)$ has an explicit expression in the standard space $\mathbb{M}_{m,n}$. To keep notation separate, we refer to the former as $\mathcal{Q}_{W,\gamma}(f)$. In general, this will only be possible if $W$ is a direct tensor, i.e. of the form

$$w_\mathbf{j} = w_{1,j_1} \ldots w_{d,j_d} \qquad (5)$$

where $w_1, \ldots, w_d$ sequences.

Let $\sqrt{W}$ be the pointwise square root of $W$, define $\mathscr{I} : \mathbb{M}_\mathbf{n}^W \to \mathbb{M}_\mathbf{n}$ by $(\mathscr{I}(X))_\mathbf{j} = \sqrt{W}_\mathbf{j} X_\mathbf{j}$ and note that this linear operator is isometric and bijective.

**Proposition 4.1.** *Let $f$ be a $[0,\infty]$-valued functional on $\mathbb{M}_\mathbf{n}^W$ where $W$ is an outer product as in* (5). *Then*

$$\mathcal{Q}_{W,\gamma}(f)(X) = \mathcal{Q}_\gamma(f \circ \mathscr{I}^{-1})(\mathscr{I}(X)).$$

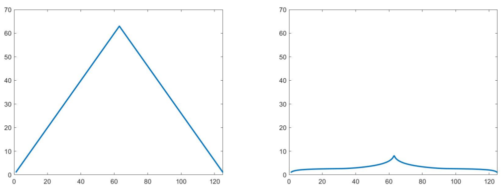


**Fig. 2**. Left; standard triangular weight. Right; corresponding weight for (8), with $n = 63$.

*Proof.* Let $\mathcal{S}_{W,\gamma}$ be the negative of the Moreau envelope, i.e. $\sup_{X\in\mathbb{M}_{\mathbf{n}}^W} -f(X) - \frac{\gamma}{2}\|X - Y\|^2_{\mathbb{M}_{\mathbf{n}}^W}$. Then $Q_{W,\gamma}(f) = \mathcal{S}_{W,\gamma}(\mathcal{S}_{W,\gamma}(f))$ (see Example 1.44 in [11]) and

$$\mathcal{S}_{W,\gamma}(f)(Y) = \sup_{X\in\mathbb{M}_{\mathbf{n}}^W} -f(X) - \frac{\gamma}{2}\|X - Y\|^2_{\mathbb{M}_{\mathbf{n}}^W}$$
$$= \sup_{X\in\mathbb{M}_{\mathbf{n}}^W} -f(\mathscr{I}^{-1}(\mathscr{I}X)) - \frac{\gamma}{2}\|\mathscr{I}(X - Y)\|^2$$
$$= \mathcal{S}_{\gamma}(f\circ\mathscr{I}^{-1})(\mathscr{I}(Y)).$$

The sought identity follows by applying this identity twice. □

## 5. THE LOW RANK HANKEL PROBLEM

At the end of Section 1 we posed the problem (2) and noted that [3] had investigated solving this by computing its convex envelope. By the above theory we know that the convex envelope can be expressed as $\mathcal{Q}_1(\iota_{\mathcal{M}_K})(X) + \frac{1}{2}\|X - H_d\|^2_{\mathbb{M}_{n,n}}$. However, convergence proofs for e.g. ADMM often require some sort of strict convexity [12, 13], which we achieve if we lower $\gamma = 1$ slightly. We have also observed that this greatly improves convergence rates, so henceforth we let $\gamma < 1$ be a number near 1.

In order to minimize the corresponding functional, one may use ADMM on the splitting

$$\operatorname*{arg\,min}_{X=Y,\ Y\in\mathcal{H}} \mathcal{Q}_{\gamma}(\iota_{\mathcal{M}_K})(X) + \frac{1}{2}\|X - H_d\|^2_{\mathbb{M}_{n,n}} + \iota_{\mathcal{H}}(Y), \tag{6}$$

which is guaranteed to converge e.g. by [13]. However, this has the drawback that the term $\|X - H_d\|^2_{\mathbb{M}_{n,n}}$ gives rise to the undesired triangular weight in Figure 2, left. (The issue with the triangular weight is pertinent to many Hankel based frequency estimation schemes, see e.g. [14, 15, 16, 17].) Moreover, since for ADMM we have to compute an $\arg\min_X$ for the corresponding Lagrangian, it is impossible to put any operator in front of $X$ if we want a closed form solution to the $X$-update step, which rules out considering missing data or unequally spaced data. It is therefore tempting to swap $X$ for $Y = H_y$ in the middle term, giving the problem

$$\operatorname*{arg\,min}_{X=H_y} \mathcal{Q}_{\gamma}(\iota_{\mathcal{M}_K})(X) + \frac{1}{2}\|H_y - H_d\|^2_{\mathbb{M}_{n,n}} \tag{7}$$

and the corresponding steps in ADMM are efficiently computed, even if we e.g. introduce weights in the middle term to get rid of the triangular weight (at the cost of losing convexity).

However, it turns out that the ADMM scheme for (7) can diverge, despite the problem being convex over the subspace defined by $X = H_y$ (but not in the full space). This is rather surprising in the light of several recent contributions showing that ADMM converges in many nonconvex settings [18]. However, these issues seem to cease if one takes $\rho$ sufficiently large, and therefore we will explore this option in the numerical section. The issue of lacking convexity of (7) led to the introduction of the more complicated functional

$$\mathcal{Q}_{\gamma}(\iota_{\mathcal{M}_K})(X) + \frac{p}{2}\|X - H_y\|^2_{\mathbb{M}_{n,n}} + \frac{q}{2}\|H_y - H_d\|^2_{\mathbb{M}_{n,n}},$$

see [17] where it was shown that this is convex if and only if $1/p + 1/q \leq 2/\gamma$. The new ingredient for this article is to consider minimization of the functional $\iota_{\mathcal{M}_K}(X) + \frac{1}{2}\|X - H_d\|^2_W$ over the set of Hankel matrices, where $W_{i,j} = u_i u_j$ with $u_i = \frac{1}{\sqrt{k-|i-k|}}$. By the above theory the l.s.c. convex envelope is given by

$$\mathcal{Q}_1(\iota_{\mathcal{M}_K})(\mathscr{I}(X)) + \frac{1}{2}\|X - H_d\|^2_W$$

Inserting $X = H_y$ in the quadratic term gives

$$\|H_y - H_d\|^2_W = \sum_{j=1}^{2n-1} \omega_j |y_j - d_j|^2, \tag{8}$$

where $\omega_j$ is depicted in Figure 2, right. This weight is clearly much closer to a uniform flat weight than the one arising from (7) (see also [16]).

## 6. NUMERICAL RESULTS

We have run extensive tests comparing minimization of the functionals discussed earlier, as well as

$$\mathcal{Q}_{0.98}(\iota_{\mathcal{M}_K})(\mathscr{I}(X)) + \frac{1}{2}\|A(X - H_d)\|^2_W \tag{9}$$

where $A$ either is identity or an optional operator designed so that $\|A(H_y - H_d)\|^2_W = \|y - d\|^2$, and $W$ either

| 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 1**. 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 1. 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 number of times "nuclear" needs to be repeated to find a suitable $\lambda$.

| 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 2**. Same as before but $K = 5$.

equals $\mathbf{1}$ or the weight considered in the end of the previous section.

We share a few representative examples and some general observations. A tempting way to obtain a more flat weight in (7) is to consider rectangular matrices $m \neq n$, but we saw no substantial improvement in estimation accuracy, so this option is not demonstrated. We tried all combinations of $A$ and $W$ in (9), and concluded that $W = \mathbf{1}$, $A = Id$ (labeled $\mathcal{Q}$-standard) gives fastest convergence and good accuracy, whereas letting both be "weighted" (labeled $\mathcal{Q}$-weights) gives best accuracy consistently. We therefore display only these results along with "hard cutting" (HC), i.e. running the same algorithm without $\mathcal{Q}_{0.98}$ applied to $\iota_{\mathcal{M}_K}$, in which case the corresponding proximal operator "cuts" the singular values after index $K$ (c.f. eq. (2)). The results of one concrete test are displayed in Table 1, the corresponding noisy function and its approximation via $\mathcal{Q}$-weighted is shown in Fig. 1. The stopping criteria used was $\|X^{k+1} - X^k\| < 10^{-12}$ for two successive iterates, or $k = 2000$.

In Table 1 we also display the result of using ESPRIT and nuclear norm $\lambda\|X\|_{nuc} = \lambda\|\sigma(X)\|_1$. A major drawback of the latter is that one needs to run the algorithm multiple times to find an appropriate value of $\lambda$ which gives the desired rank, so this algorithm is in reality far slower. We only display this method using the misfit $\frac{1}{2}\|H_y - H_d\|^2$ (i.e. using the triangle weight from Figure 2) as opposed to the (flat-weight) $\frac{1}{2}\|y-d\|^2$ appearing in (1). The reason is that this gives best results (in terms of "flat-weight" misfit!) and far superior convergence rates. The latter is symptomatic for all penalties (although less pronounced), unclear to us why.

Fastest is of course ESPRIT, which is not iterative, and based on Table 1 one may ask whether iterative methods are worthwhile at all. However, ESPRIT is inadequate to treat e.g. missing data, whereas the above algorithms are readily modified to this setting, see [2, 4]. Moreover, we shall soon see that ESPRIT performs worse if the model-order is poorly estimated.

It is interesting to note that $\mathcal{Q}$-standard and $HC$-standard finds exactly the same point (in accordance with Proposition 3.1) whereas this is not the case for the weighted versions, demonstrating the ability of $\mathcal{Q}_\gamma$-transform based techniques to avoid local minima. Moreover, for more difficult scenarios, the $HC$-algorithms sometimes do not converge at all, which also favors the use of the $\mathcal{Q}_\gamma$-transform.

Next we ran the same example but seeking rank 5, see Table 2. Now all iterative methods outperform ESPRIT with good margin, and best is again $\mathcal{Q}$-weights. Interesting to note is that $\mathcal{Q}-$standard has a better misfit but found the "wrong" rank. This can be corrected by increasing $\gamma$ (at the cost of entering the non-convex regime) which plays a similar role as $\lambda$ for the nuclear norm. In fact, due to the simple structure of $\iota_{\mathcal{M}_K}$ we have $\mathcal{Q}_\gamma(\iota_{\mathcal{M}_K}) = \gamma\mathcal{Q}_1(\iota_{\mathcal{M}_K})$. We reran the experiment with $\gamma = 1.2$, and then $\mathcal{Q}$-standard diverges and finds the same point as $HC$-standard, again underlining that $\mathcal{Q}_\gamma(f)$ often equals $f$. By "diverges" we mean that $\|X^{k+1} - X^k\|$ seems to converge to a non-zero number.

| Name | $\|y-d\|$ | Conv. Rate | Rank |
|---|---|---|---|
| $\mathcal{Q}$-weights | 11.65 | 588 | 8 |
| $HC$-weights | 11.66 | 567 | 8 |
| $\mathcal{Q}$-standard | 12.28 | 88 | 8 |
| $HC$-standard | 12.28 | 90 | 8 |
| ESPRIT | 12.96 | 1 | 8 |
| nuclear | 14.06 | iter$*$236 | 8 |

**Table 3**. Same as Table 1 but averaged over 69 trials.

Finally we ran the same setup as in the first example, but repeated 100 times. On 19 occasions the $\mathcal{Q}$-standard method got the wrong rank and on 12 occasions the weighted versions (both) diverged, so we discarded these and averaged the rest, see Table 3. As is plain to see, the pattern from Table 1 is typical.

## 7. REFERENCES


[1] Maryam Fazel, *Matrix rank minimization with applications*, Ph.D. thesis, PhD thesis, Stanford University, 2002.

[2] Maryam Fazel, Ting Kei Pong, Defeng Sun, and Paul Tseng, "Hankel matrix rank minimization with applications to system identification and realization," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 946–977, 2013.

[3] Viktor Larsson and Carl Olsson, "Convex low rank approximation," *International Journal of Computer Vision*, pp. 1–21, 2016.

[4] Fredrik Andersson, Marcus Carlsson, Jean-Yves Tourneret, and Herwig Wendt, "A new frequency estimation method for equally and unequally spaced data," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5761–5774, 2014.

[5] Marcus Carlsson, "On convexification/optimization of functionals including an l2-misfit term," *arXiv preprint arXiv:1609.09378*, 2016.

[6] Marcus Carlsson, "On convex envelopes and regularization of non-convex functionals without moving global minima," *Journal of Optimization Theory and Applications, to appear*.

[7] Marcus Carlsson, Daniele Gerosa, and Carl Olsson, "An unbiased approach to compressed sensing," *arXiv preprint arXiv:1806.05283*, 2018.

[8] Paul Tseng, "Approximation accuracy, gradient methods, and error bound for structured convex optimization," *Mathematical Programming*, vol. 125, no. 2, pp. 263–295, 2010.

[9] Fredrik Andersson, Marcus Carlsson, and Carl Olsson, "Convex envelopes for fixed rank approximation," *Optimization Letters*, vol. 11, no. 8, pp. 1783–1795, 2017.

[10] Fredrik Andersson and Marcus Carlsson, "Esprit for multidimensional general grids," *SIAM Journal on Matrix Analysis and Applications*, vol. 39, no. 3, pp. 1470–1488, 2018.

[11] R Tyrrell Rockafellar and Roger J-B Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.

[12] Tom Goldstein, Brendan O'Donoghue, Simon Setzer, and Richard Baraniuk, "Fast alternating direction optimization methods," *SIAM Journal on Imaging Sciences*, vol. 7, no. 3, pp. 1588–1623, 2014.

[13] Pontus Giselsson and Stephen Boyd, "Linear convergence and metric selection for douglas-rachford splitting and admm," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 532–544, 2017.

[14] James A Cadzow, "Signal enhancement-a composite property mapping algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 1, pp. 49–62, 1988.

[15] Jonathan Gillard and Anatoly Zhigljavsky, "Optimization challenges in the structured low rank approximation problem," *Journal of Global Optimization*, vol. 57, no. 3, pp. 733–751, 2013.

[16] JW Gillard and Anatoly A Zhigljavsky, "Weighted norms in subspace-based methods for time series analysis," *Numerical Linear Algebra with Applications*, vol. 23, no. 5, pp. 947–967, 2016.

[17] Fredrik Andersson and Marcus Carlsson, "Fixed-point algorithms for frequency estimation and structured low rank approximation," *Applied and Computational Harmonic Analysis*, 2017.

[18] Yu Wang, Wotao Yin, and Jinshan Zeng, "Global convergence of admm in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, pp. 1–35, 2015.