GLOBAL OPTIMALITY FOR NON-LINEAR CONSTRAINED RESTORATION PROBLEMS VIA INVEXITY

Samuel Pinilla, & Jeyan Thiyagalingam
*Scientific Computing Department, Science and Technology Facilities Council
samuel.pinilla@stfc.ac.uk

ABSTRACT

Signal restoration is an important constrained optimization problem with significant applications in various domains. Although non-convex constrained optimization problems have been shown to perform better than convex counterparts in terms of reconstruction quality, convex constrained optimization problems have been preferred for their global optima guarantees. Despite the success of non-convex methods in many applications, it is not an overstatement to say that there is little or no hope for non-convex problems to ensure global optima. In this paper, for the first time, we propose a family of *invex functions* for handling constrained inverse problems using the non-convex setting along with guarantees for their global optima - *invex function* is a mapping where any critical point is a global minimizer. We also develop relevant theories to extend the global optima guarantee to a family of quasi-invex functions - the largest set of optimizable mappings. Our theoretical results show that the proposed family of invex and quasi-invex functions can aid in extending existing convex optimization algorithms, such as the alternating direction of multipliers and accelerated proximal gradient methods. Moreover, our numerical tests show that the proposed family of invex/quasi-invex functions overcome the performance of convex mappings in terms of reconstruction quality across several relevant metrics and imaging tasks.

1 INTRODUCTION

Signal restoration is an optimization problem that seeks to estimate an unknown original signal $\boldsymbol{x} \in \mathbb{R}^n$ from a set of noisy observations $\boldsymbol{y} \in \mathbb{R}^m$. This optimization problem falls under the purview of inverse problems, which are common across all disciplines, from physics (Mountrakis et al., 2011; Vouras et al., 2022), to medical imaging (Ginat & Gupta, 2014; Pinilla et al., 2022b) to signal processing (Pinilla et al., 2023), and to machine learning (Harris et al., 2020; Wei et al., 2022). In general, an inverse problem can be formulated as follows:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \;\; g(\boldsymbol{x}) \;\; \text{subject to } f(\boldsymbol{x}) \leq \epsilon, \tag{1}$$

where $\epsilon > 0$, $f(\boldsymbol{x})$ is a mapping constructed as a reconstruction error (fidelity term), and $g(\boldsymbol{x})$ is a regularizer to model the priors of the signal such as sparsity, low-rankness, and smoothness (Beck, 2017, Chapter 10). Here, $f(\boldsymbol{x})$ is often considered to be differentiable (i.e., smooth), while $g(\boldsymbol{x})$ is not. For some imaging tasks, the fidelity term has the form of $f(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y})$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. When $m < n$ for $\boldsymbol{A}$, the inverse problem reduces to *compressive sensing*. This technique is extensively exploited in many areas such as microscopy, optical imaging, and spectroscopy (Arce et al., 2013; Jerez et al., 2020; Guerrero et al., 2020).

Solving (1), a guarantee for global optimality is of paramount importance to ensure that the best-possible solution is found subject to the relevant constraints (Parikh & Boyd, 2014; Pinilla et al., 2022a). Convex constructs of $f(\boldsymbol{x}), g(\boldsymbol{x})$, as demonstrated in the literature (Beck & Teboulle, 2009; Beck, 2017), are helpful for seeking guaranteed global optima, when either the measurements can be considered as ideal and noiseless (i.e., $\epsilon = 0$) (Parikh & Boyd, 2014) or when the regularizer $g(\boldsymbol{x})$ exactly models the priors of the data. However, real-world problems do not fit this assumption, where noise is rather prevalent in all measurements (i.e., $\epsilon > 0$), and convex regularizers can easily inaccurately model the sparsity, low-rankness, smoothness, and abnormal patterns in the data (Candes et al., 2008; Laghrib et al., 2023). Although approaches with assumptions tied to noise models (such as Gaussianity) perform better than plain convex approaches, they too fail as the real-world noise models often deviate from these assumptions (Sun & Pong, 2022; Tankovich et al., 2021).

*Rutherford Appleton Laboratory, Harwell, UK.

Table 1: State-of-the-art proposition for $f(\boldsymbol{x})$, and $g(\boldsymbol{x})$ in optimization problem (1).

Ref	**Type of function**		**Global optima**
	$f(\boldsymbol{x})$	$g(\boldsymbol{x})$	
(Sun & Pong, 2022)	Geman-McClure	SCAD	No
(Carrillo et al., 2010)	Huber loss	Log-sum penalty	No
(Carrillo et al., 2016)	Cauchy loss function	ℓ_1-norm	No
(Pinilla et al., 2022a)	ℓ_2-norm	invex	**Yes**
Proposed	invex/quasi-invex	invex/quasi-invex	**Yes**

An alternative approach is to consider non-convex mappings for $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$. From (Kaul & Kaur, 1985), it can be seen that there exists a large collection of mappings that can be well-optimized compared to pure convex settings. We summarize important non-convex mappings from the literature in Table 1. These are a set of well-accepted mappings that perform better than convex ones in terms of restoration quality (Pinilla et al., 2022a; Sun & Pong, 2022). Various regularizers have been considered in the literature to overcome the limitations of pure convex approaches, such as the ℓ_1-norm, total-variation (Chambolle, 2004; Selesnick et al., 2020) and nuclear norm (Candès & Recht, 2006), which are the continuous and convex surrogates of the ℓ_0-pseudo norm and rank, respectively (Fu, 2014). More specifically, several interpolations between the ℓ_0-pseudonorm and the ℓ_1-norm have been explored as regularizers including the ℓ_p-quasinorms (where $0 < p < 1$) (Marjanovic & Solo, 2012), Capped-ℓ_1 penalty (Zhang et al., 2007), Log-Sum Penalty (Candes et al., 2008), Convex-concave penalty (Kovalev & Gasnikov, 2022), Minimax Concave Penalty (Zhang, 2010), Geman Penalty (Geman & Yang, 1995), and the Smoothly Clipped Absolute Deviation (SCAD) (Fan & Li, 2001). Furthermore, ℓ_2-norm incorporated as part of the constraint in (1) is often replaced by other loss functions to reflect different noise models or robustness requirements in the noisy data $\boldsymbol{y}$, such as those outlined in (Carrillo et al., 2016; Zoubir et al., 2012). Concrete alternative reconstruction error functions include the Cauchy loss (Carrillo et al., 2016), the Huber function (Hedman et al., 2021), and the Tukey biweight loss (Mosteller & Tukey, 1977). However, while offering outstanding restoration quality, none of these can offer a guarantee for global optima, which would have been ideal for providing the soundness of these approaches.

The lack of approaches for guaranteeing the global minima in non-convex settings leaves any optimization-based algorithmic solution incomplete and non-unique. It hence cannot be categorically accepted as the best possible solution despite their improved performance over convex functions. Motivated by this problem of seeking guarantees for global optimality when solving non-convex problems, we investigate broader classes of functions that exploit the concept of *invexity* and *quasi-invexity* (Hanson, 1981; Reiland, 1990), which have been initially defined in early 1980s (Hanson, 1981). The specialty of invex functions is that *any critical point in an invex function is a global minimizer itself* (Pinilla et al., 2022a). Quasi-invex mappings generalize invex functions and hence are the largest set of optimizable functions as outlined in (Kaul & Kaur, 1985). This places the quasi-invex functions in a special and unique position for handling inverse problems. Although mathematical implications of quasi-invex (and hence invex) functions are well developed (Syed et al., 2013; Barik & Honorio, 2021; Pinilla et al., 2022a), their practical utility is less understood (Zălinescu, 2014). To the best of our knowledge, there is no existing body of work in the invex literature on applying and developing quasi-invex (and invex)-based methods for seeking global optimality in constrained optimization problems, such as signal restoration.

In this paper, we focus on demonstrating the practical utility of assuming $f(\boldsymbol{x}), g(\boldsymbol{x})$ to be invex/quasi-invex to handle inverse problems as in (1), with guarantees for global optima. Hence, in contrast with results in (Pinilla et al., 2022a), we make the following key contributions in:

- We develop a set of mathematical tools to construct a family of $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ in (1) with proved invexity/quasi-invexity for model-based inverse imaging problems. Additionally, we prove that our proposed family of invex/quasi-invex functions is closed under summation - a result that has not been available before in the invex literature of optimization (Zălinescu, 2014).
- We prove global optima under the invex/quasi-invex settings. These results extend the capability of *alternating direction method of multipliers* and *accelerated proximal gradient* methods to handle non-convex cases otherwise limited to convex settings. We also report their convergence rates, which show both methods using invex/quasi-invex mappings are as efficient as solving the well-established convex optimization problems;
- We conduct an effective evaluation of the proposed approach in handling a number of signal restoration problems against state-of-the-art algorithms and baselines.

2 PRELIMINARIES

The i-th entry of a vector $\boldsymbol{w}$, is $\boldsymbol{w}[i]$. For vectors, $\|\boldsymbol{w}\|_p$ is the ℓ_p-norm. An open ball is defined as $B(\boldsymbol{x}; r) = \{\boldsymbol{y} \in \mathbb{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\|_2 < r\}$. The operation conv$(\mathcal{A})$ represents the convex hull of the set $\mathcal{A}$. We use $\sigma_i(\boldsymbol{W})$ to denote the i-th singular value of a matrix $\boldsymbol{W}$ in descending order.

We present several concepts needed for the development of this paper starting with the definition of a locally Lipschitz continuous function.

Definition 1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz continuous at a point $\boldsymbol{x} \in \mathbb{R}^n$ if there exist scalars $K > 0$ and $\epsilon > 0$ such that for all $\boldsymbol{y}, \boldsymbol{z} \in B(\boldsymbol{x}, \epsilon)$ we have $|f(\boldsymbol{y}) - f(\boldsymbol{z})| \leq K\|\boldsymbol{y} - \boldsymbol{z}\|_2$.

Since the ordinary directional derivative being the most important tool in optimization does not necessarily exist for locally Lipschitz continuous functions, it is required to introduce the concept of subdifferential (Bagirov et al., 2014) which is calculated in practice as follows.

Theorem 1. (Bagirov et al., 2014, Theorem 3.9) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function at $\boldsymbol{x} \in \mathbb{R}^n$, and define $\Omega_f = \{\boldsymbol{x} \in \mathbb{R}^n | f \text{ is not differentiable at the point } \boldsymbol{x}\}$. Then the subdifferential of f is given by

$$\partial f(\boldsymbol{x}) = \text{ conv } \left(\{\boldsymbol{\zeta} \in \mathbb{R}^n | \text{ exists } (\boldsymbol{x}_i) \in \mathbb{R}^n \setminus \Omega_f \text{ such that } \boldsymbol{x}_i \to \boldsymbol{x} \text{ and } \nabla f(\boldsymbol{x}_i) \to \boldsymbol{\zeta}\}\right). \quad (2)$$

The notion of subdifferential is given for locally Lipschitz continuous functions because it is always nonempty (Bagirov et al., 2014, Theorem 3.3). Based on this, we present the concept of invexity.

Definition 2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz; then f is invex if there exists a function $\eta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ such that $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $\forall \boldsymbol{\zeta} \in \partial f(\boldsymbol{y})$ we have $f(\boldsymbol{x}) - f(\boldsymbol{y}) \geq \boldsymbol{\zeta}^T \eta(\boldsymbol{x}, \boldsymbol{y})$.

It is well known that a convex function simply satisfies this definition for $\eta(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x} - \boldsymbol{y}$. The following theorem (Mishra & Giorgi, 2008, Theorem 4.33) makes connection between invexity and its well-known optimum property which motivates the design of invex optimization problems.

Theorem 2. (Mishra & Giorgi, 2008, Theorem 4.33) Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz. Then the following statements are equivalent: 1) f is invex. 2) Every point $\boldsymbol{y} \in \mathbb{R}^n$ that satisfies $\mathbf{0} \in \partial f(\boldsymbol{y})$ is a global minimizer of f.

We finalize this section by introducing a generalization of invexity; indeed (Reiland, 1990) introduced also the following classes of generalized convex functions.

Definition 3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz; then f is quasi-invex if there exists $\eta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ such that $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\forall \boldsymbol{\zeta} \in \partial f(\mathbf{y})$ we have $f(\mathbf{x}) - f(\mathbf{y}) \leq 0 \implies \boldsymbol{\zeta}^T \eta(\mathbf{x}, \mathbf{y}) \leq 0$.

3 INVEX AND QUASI-INVEX FUNCTION SETS

We address the necessity of handling the issues around real-world data by constructing invex/quasi-invex functions for data fidelity loss $f(\boldsymbol{x})$, and regularizer $g(\boldsymbol{x})$ functions. To this end, we first define the family of functions this paper investigates and their properties.

Definition 4. (**Admissible Function**) Let $h : \mathbb{R}^n \to \mathbb{R}$ such that $h(\boldsymbol{x}) = \sum_{i=1}^n s(|\boldsymbol{x}[i]|)$, where $s : [0, \infty) \to [0, \infty)$ and $s'(w) > 0$ for $w \in (0, \infty)$. If s with $s(0) = 0$ such that $s(w)/w^2$ is non-increasing on $(0, \infty)$, then $h(\boldsymbol{x})$ is said to be an *admissible function*.

Definition 4 places a requirement on $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ in (1) to operate element-wise on positive values. This condition is mild since it holds for the vast majority of inverse problem formulations, including loss functions in deep learning, whether they are convex or non-convex (Carrillo et al., 2016; Wen et al., 2018; Sun & Pong, 2022). Using element-wise mapping offers notable advantages, especially in developing high-performance algorithms. One key advantage lies in the ability to parallelize, effectively reducing runtimes, making them suitable for addressing large-scale problems (Parikh & Boyd, 2014; Wang & Chen, 2022; Zoubir et al., 2012). Other conditions in Definition 4 are crucial as they establish a vital connection to invexity, which we discuss next in Theorem 3 (see Appendix A for the proof).

Theorem 3. Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be two admissible functions as in Definition 4, such that $f(\boldsymbol{x}) = \sum_{i=1}^n s_f(|\boldsymbol{x}[i]|)$, and $g(\boldsymbol{x}) = \sum_{i=1}^n s_g(|\boldsymbol{x}[i]|)$. Then the following holds: i) $f(\boldsymbol{x})$, and $g(\boldsymbol{x})$ are invex; ii) $h_s(\boldsymbol{x}) = \alpha f(\boldsymbol{x}) + \beta g(\boldsymbol{x})$ is an admissible function (therefore invex) for every $\alpha, \beta \geq 0$; iii) $h_c(\boldsymbol{x}) = \sum_{i=1}^n (s_f \circ s_g)(|\boldsymbol{x}[i]|)$ is an admissible function. In addition to these, if we assume that $s'_f(w), s'_g(w) \geq 0$ for $w \in (0, \infty)$, then $f(\boldsymbol{x}), g(\boldsymbol{x})$, and $h_s(\boldsymbol{x}), h_c(\boldsymbol{x})$ are all quasi-invex.

The essential breakthrough of Theorem 3 compared with results in the invex literature such as (Pinilla et al., 2022a) lies in the fact that Definition 4 introduces *the first family of invex/quasi-invex functions closed under summation*. This implies that the sum of invex/quasi-invex functions is also invex/quasi-invex. The significance of this result manifests on two fronts. Firstly, Theorem 3 paves the way for establishing theoretical guarantees for global optima of (1) beyond the realm of convexity. Secondly, it bestows practical benefits to practitioners, as both Definition 4 and the third statement of Theorem 3 offer a systematic methodology for constructing invex/quasi-invex functions. These results also give away another side consequence: for any admissible function $f(\boldsymbol{x}) = \sum_{i=1}^{n} s_f(|\boldsymbol{x}[i]|)$ such that $s'_f(w) \geq 0$, then f is not invex. The proof stems from the fact that $s'_f(w)$ can be zero in non-minimizer values, leading to its quasi-invexity. In the following sections, we leverage the results in Theorem 3 to present illustrative examples of invex/quasi-invex functions that apply in diverse scientific domains.

3.1 Invex Regularizers

In addition to the list of original invex regularizer functions presented in (Pinilla et al., 2022a, Lemma 1), consider the following Theorem 4 (See Appendix B for proof).

Theorem 4. For $p \in (0, 1]$, $\epsilon > 0$, the following mapping

$$\textit{(LL}_p\textit{-regularizer)} \quad g(\boldsymbol{x}) = \sum_{i=1}^{n} \log(1 + (|\boldsymbol{x}[i]| + \epsilon)^p), \tag{3}$$

is an admissible function. Additionally, let $\boldsymbol{D} \in \mathbb{R}^{(n-1)\times n}$ be the first-order difference matrix (further details in Appendix B). If g is an admissible function, then $g_{TV}(\boldsymbol{x}) = g(\boldsymbol{D}\boldsymbol{x})$ is invex.

The key to prove Theorem 4 is centred around the fact that any admissible function is invex, and it is a point-wise non-increasing when divided by a quadratic mapping. It is worth mentioning that the invex function presented in equation (3) is one of the pivotal functions in the domain of compressive sensing (Keshavarzian et al., 2019). Although the analysis of equation (3) is valid with or without ϵ, we prefer to retain ϵ to satisfy Definition 1. Although the literature identified this function as non-convex (Wang & Chen, 2022; Liu et al., 2018), there is no notion of invexity or proof of invexity.

The second part of Theorem 4 presents invex generalizations of the classical convex total-variation (TV) regularizer (Chambolle, 2004; Selesnick et al., 2020) with the ℓ_1-norm replaced by invex functions - missing result in the invex literature such as (Pinilla et al., 2022a). TV-based regularization is effective in various applications such as computer tomography (Sun et al., 2019), hyperspectral imaging (Iordache et al., 2012; Wang et al., 2023), and computer vision (Estrela et al., 2016), where the goal is purely signal restoration. Classical TV-based algorithms suffer from an issue where the ℓ_1-norm tends to underestimate the amplitudes of signal discontinuities owing to the limitation of the ℓ_1-norm regularization itself (Selesnick et al., 2020). Several studies in the literature (Liu et al., 2018; Selesnick et al., 2020) show that ℓ_p-quasinorm instead of ℓ_1-norm can significantly address this issue, albeit the literature not demonstrating the benefits ascribed to the invexity. Again, similar to equation (3), the literature identified this function as non-convex (Wang & Chen, 2022; Liu et al., 2018) without any notion of these functions being invex or proof of invexity.

Driven by the success of equation (3) in compressive sensing (Keshavarzian et al., 2019), we propose its TV-like regularizer version. Further insights of Theorem 4, including the extension of TV-like regularizers to two- and three-dimensional signals, are presented in Section B.1.

3.2 Invex Fidelity Losses

Now, to complement the list of invex regularizers stated in Section 3.1, we present five invex data fidelity functions $f(\boldsymbol{x})$ in the following theorem.

Theorem 5. All the following functions for $c, \delta > 0$, and $\alpha \in \mathbb{R}$ are admissible

$$\textit{(Cauchy)} \quad f(\boldsymbol{x}) = \sum_{i=1}^{m} \log\left(1 + \frac{\boldsymbol{x}^2[i]}{\delta^2}\right) \tag{4}$$

$$\textit{(Geman-McClure)} \quad f(\boldsymbol{x}) = \sum_{i=1}^{m} \frac{2\boldsymbol{x}^2[i]}{\boldsymbol{x}^2[i] + 4\delta^2} \tag{5}$$

$$\textit{(Welsh)} \quad f(\boldsymbol{x}) = \sum_{i=1}^{n} 1 - \exp\left(\frac{-\boldsymbol{x}^2[i]}{2\delta^2}\right) \tag{6}$$

$$\textit{(Adaptive robust)} \quad f(\boldsymbol{x}) = \sum_{i=1}^{n} \frac{|\alpha-2|}{\alpha}\left(\left(\frac{(\boldsymbol{x}[i]/c)^2}{|\alpha-2|}+1\right)^{\alpha/2}-1\right) \tag{7}$$

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} \log\left(1+\boldsymbol{x}^2[i]\right) - \frac{\boldsymbol{x}^2[i]}{2\boldsymbol{x}^2[i]+2} \tag{8}$$

Fidelity losses given in equations (4)-(7) have existed in the literature without ever being identified as invex functions or the benefits being ascribed to their invexity. Equation (6) was identified as an invex function, and its invexity was proved in (Syed et al., 2013; 2014). Equations (4)-(7) are found in the literature replacing the traditional ℓ_2-norm. This was to account for different noise models and improve the solution's robustness towards noisy measurements $\boldsymbol{y}$ in (1). Equations (4)-(7) have been used to improve the performance of the original Variational Autoencoder (Kingma & Welling, 2014) by incorporating equations (4)-(6) as robust loss functions in (Barron, 2019). Finally, equation (8) is a novel invex function we propose to provide more flexibility around losses (harder to achieve in a convex setting). The proof of Theorem 5 is deferred to Appendix C.

3.3 Quasi-invex Regularizers

We list three quasi-invex function constructs for $g(\boldsymbol{x})$ in Theorem 6 below.

Theorem 6. For constants $\lambda, p \in (0,1]$, and $\epsilon > 0$, the following mappings are quasi-invex

$$g_\theta(\boldsymbol{x}) = \lambda \sum_{i=1}^{n} r_\theta(\boldsymbol{x}[i]);\ \theta \geq 1,\ r_\theta(\boldsymbol{x}[i]) = \begin{cases} \lambda|\boldsymbol{x}[i]| - \boldsymbol{x}^2[i]/(2\theta) & \text{if } |\boldsymbol{x}[i]| < \theta\lambda \\ \theta\lambda^2/2 & \text{if } |\boldsymbol{x}[i]| \geq \theta\lambda \end{cases} \tag{9}$$

$$g_\theta(\boldsymbol{x}) = \sum_{i=1}^{n} \min\left(\log(1+(|\boldsymbol{x}[i]|+\epsilon)^p), \theta\right);\ \theta > 0. \tag{10}$$

Further, let $\boldsymbol{D} \in \mathbb{R}^{(n-1)\times n}$ be the first-order difference matrix. If g is admissible such that $g(\boldsymbol{x}) = \sum_{i=1}^{n} s_g(|\boldsymbol{x}[i]|)$ with $s_g'(w) \geq 0$ for $w \in (0,\infty)$, then $g_{TV}(\boldsymbol{x}) = g(\boldsymbol{D}\boldsymbol{x})$ is quasi-invex.

The critical aspect of proving Theorem 6 is centred around the quasi-invex construct in Theorem 3 and that it is a point-wise non-increasing when divided by a quadratic mapping. Equation (9) referred to as the Minimax Concave Penalty (MCP) (Zhang, 2010) in the non-convex literature, with a range of applications in various domains (Kovalev & Gasnikov, 2022; Zhang, 2010). Although this mapping function is standard in imaging applications (Ding et al., 2018; Song et al., 2022), the knowledge of its quasi-invexity has never been identified nor proven. For the first time, we provide formal proof of MCP being quasi-invex in Appendix D. Additional functions, such as the Clipped Absolute Deviation (SCAD) (Fan & Li, 2001), are proven to be quasi-invex in Section D. Furthermore, following an analysis analogous to Theorem 4, we can extend to two- and three-dimensional TV-like quasi-invex regularizers. In addition to these three quasi-invex constructs, we propose another way to produce quasi-invex loss functions. We present this in Lemma 1 below.

3.4 Quasi-invex Fidelity Losses

We finalize the exposition on quasi-invex functions by presenting the following Lemma 1 for constructing a quasi-invex function for $f(\boldsymbol{x})$. This result states that the maximum of two invex functions (not necessarily with respect to the same η) is quasi-invex. This lemma has applications in robust optimization or robust imaging problems (Vargas et al., 2018b), where controlling the accuracy of the final solution is crucial. The route to robustness is achieved by minimizing the maximum between two invex functions, which will bound the highest error to a desirable value (Boyd et al., 2004, Section 1.2.2). The proof of this lemma is deferred to Appendix E.

Lemma 1. Let $f, h : \mathbb{R}^n \to \mathbb{R}$ be two invex functions as in Definition 2 (not necessarily with respect to the same η). Then, $q(\boldsymbol{x}) = \max(f(\boldsymbol{x}), h(\boldsymbol{x}))$ is quasi-invex.

4 Applications

In this section, we demonstrate the use of invex/quasi-invex functions on several applications and guarantees for global optima, which has only been possible for convex functions. To this end, we commence by (1) ensures global optima when $f(\boldsymbol{x}), g(\boldsymbol{x})$ are admissible functions - missing result

in the invex literature such as (Pinilla et al., 2022a). The importance of this result, which we prove in Appendix F, lies in the fact that it is the first time that global optima is achieved for a family of invex constrained optimization problems (for the noiseless case $\epsilon = 0$). The key to the proof is that both f, g functions are admissible.

To complement this result, we provide an extensive list of applications that would directly benefit from the global optima guarantees of (1), and the family of invex/quasi-invex constructs in Theorem 3 summarized in Table 2, and for the reasons of brevity, we selectively discuss three of the important applications. These include two imaging tasks, namely, compressive image restoration and total variation filtering, and two algorithmic applications, namely, extended Alternating Direction Method of Multipliers (ADMM) in Section 4.3, and extended accelerated proximal gradient method (APGM) which we present due to space limitations in Appendix J. Additionally, the extended guarantees of ADMM allows to state the global optimality of solving program (1) when $f(\boldsymbol{x}), g(\boldsymbol{x})$ are the family of quasi-invex functions introduced in Theorem 3.

Table 2: List of Invex and Quasi-invex functions studied in this work.

Reference	Invex	Application
(Wang et al., 2022)	equation (3)	Compressive sensing
This work	Invex $g_{TV}(\boldsymbol{x})$	Tomography Total variation filtering
(Tankovich et al., 2021)	equation (4)	Robust Learning
(Park et al., 2021)	equation (5)	Neural Radiance Fields
(Psaros et al., 2022)	equations (6),(7)	Adaptive Filtering
This work	equation (8)	Supervised Learning

Reference	Quasi-invex	Application
(Song et al., 2022)	equation (9)	Model selection, ADMM
This work	equation (10)	Dictionary Learning
This work	Quasi-invex $g_{TV}(\boldsymbol{x})$	Inpainting

4.1 Compressive Image Restoration

The reconstruction of a signal/image from a compressed set of measurements is essential in a range of scientific domains that rely on the analysis and interpretation of image content (De los Reyes et al., 2017). In this context, program (1) becomes an inverse problem that aims at recovering an image $\boldsymbol{f} \in \mathbb{R}^n$ assuming that $\boldsymbol{f}$ has a k-sparse representation $\boldsymbol{\theta} \in \mathbb{R}^n$ ($k \ll n$ non-zero elements) in a basis $\boldsymbol{\Psi} \in \mathbb{R}^{n\times n}$, that is $\boldsymbol{f} = \boldsymbol{\Psi}\boldsymbol{x}$, in order to ensure uniqueness under some conditions. Examples of this sparse basis $\boldsymbol{\Psi}$ in imaging are the Wavelet (also Haar Wavelet) transform, cosine, and Fourier representations (Foucart & Rauhut, 2013). Hence, one can work with the abstract noisy model $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{\Psi}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}+\boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is the noise, $\boldsymbol{A}$ encapsulates the product between acquisition matrix $\boldsymbol{H}$, and $\boldsymbol{\Psi}$, with ℓ_2-normalized columns (Arce et al., 2013; Candès & Wakin, 2008). Under this setup, compressive sensing enables the recovery of $\boldsymbol{x}$ using much fewer samples than predicted by the Nyquist criterion (Candès & Wakin, 2008). When the regularizer $g(\boldsymbol{x})$ takes the convex form of ℓ_1-norm, $f(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y})$ is the ℓ_2-norm, and when the sampling matrix $\boldsymbol{A}$ satisfies the *restricted isometry property* (RIP) for any k-sparse vector $\boldsymbol{x} \in \mathbb{R}^n$, i.e., $(1-\delta_{2k})\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leq (1+\delta_{2k})\|\boldsymbol{x}\|_2^2$ for $\delta_{2k} < \frac{1}{3}$ (Foucart & Rauhut, 2013, Theorem 6.9), it has been proved that $\boldsymbol{x}$ can be exactly recovered by solving (1) with $\epsilon = 0$.

It is classically known that (1) has unique solution under convexity assumptions on $g(\boldsymbol{x})$, and $f(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})$ for $\epsilon = 0$. We are interested in applying invex and quasi-invex regularizers and fidelity functions. It has been proven that, when $g(\boldsymbol{x})$ takes the particular invex form in (Pinilla et al., 2022a, Theorem 4), and $f(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})$ the ℓ_2-norm, $\boldsymbol{x}$ can be exactly recovered by program (1) for $\epsilon = 0$ (Pinilla et al., 2022a). Below in Theorem 7, we further generalize this result to all the invex/quasi-invex constructs as in Theorem 3 - missing result in the invex literature such as (Pinilla et al., 2022a).

Theorem 7. Assume $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\eta} = \boldsymbol{y}$, where $\boldsymbol{x} \in \mathbb{R}^n$ is k-sparse, the matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ ($m < n$) with ℓ_2-normalized columns that satisfies the RIP condition with $\delta_{2k} < \frac{4}{\sqrt{41}}$, the noise vector $\boldsymbol{\eta}$ satisfies $f(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) < \epsilon$ for $\epsilon > 0$ and $\boldsymbol{y} \in \mathbb{R}^m$ is a noisy measurement vector. If $g(\boldsymbol{x})$, and $f(\boldsymbol{x})$ are admissible functions where $g(\boldsymbol{x})$ satisfies the triangle inequality (further details in Appendix G), then the solution $\boldsymbol{x}^\star$ of (1) holds $\|\boldsymbol{x} - \boldsymbol{x}^\star\|_1 \leq C\beta_{k,g}(\boldsymbol{x}) + D\sqrt{\epsilon}\upsilon_{f,\eta}(\boldsymbol{x})$, where the constants $C, D > 0$ depend only on δ_{2k}, and $\beta_{k,g}(\boldsymbol{x})$, $\upsilon_{f,\eta}(\boldsymbol{x})$. We show that a similar result holds for quasi-invex constructs as in Theorem 3. In addition, if $\epsilon = 0$ then $\boldsymbol{x}$ can be exactly recovered by solving (1) using both invex/quasi-invex constructs.

The key to prove Theorem 7 is centred around the properties of invex/quasi-invex constructs in Theorem 3. We present the proof of Theorem 7 in Appendix G. It also includes insights on using Lemma 1 to have more quasi-invex functions, for which Theorem 7 is also valid. We highlight that Theorem 7 is in the same form of convex recovery result (Foucart & Rauhut, 2013, Theorem 6.12), suggesting the effectiveness of our proposed family of invex/quasi-invex functions.

4.2 TOTAL VARIATION FILTERING

Total Variation (TV) regularization is a deterministic technique that safeguards discontinuities in images (Ehrhardt & Betcke, 2016; Selesnick et al., 2020). It is based on the principle that images with sharp edges have high variation while smooth transitions will show low variations (Lin et al., 2019; Liu et al., 2018). These variations are modeled using the first-order difference matrix in Theorem 4 (more details in Appendix B) and have been instrumental on a wide range of image restoration problems, including denoising (Hu & Selesnick, 2020), deblurring (Selesnick et al., 2020), inpainting (Schönlieb et al., 2009), infrared super-resolution (Liu et al., 2018), and magnetic resonance imaging (Ehrhardt & Betcke, 2016), among others (Chambolle, 2004).

Mathematically, classical TV regularization of an image $\boldsymbol{u} \in \mathbb{R}^n$ is the solution of

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad g_{TV}(\boldsymbol{x}) + \frac{1}{2\lambda}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2, \tag{11}$$

for $\lambda \in (0, 1]$ (typical choice in practice), where $g_{TV}(\boldsymbol{x})$ is the sum of the absolute value differences between adjacent pixel values, i.e., ℓ_1-norm of differences between adjacent pixel values.By solving for $\boldsymbol{u}$, TV regularization encourages the output image to have smooth regions while maintaining sharp edges. The main advantage of (11) is that it has no extraneous local minima, and the minimizer is unique (Selesnick et al., 2020). However, owing to the limitations of ℓ_1-norm (Selesnick et al., 2020; Liu et al., 2018), optimization problem in (11) can have undesirable effects, for example, in denoising, by underestimating signal discontinuities' amplitudes.

We improve program (11) by incorporating the invex/quasi-invex regularizers $g_{TV}(\boldsymbol{x})$ presented in Theorems 4 and 6 as follows (extensions for two-dimensional in Section B.1).

Theorem 8. Consider the optimization problem in (11) for invex and quasi-invex TV regularizers as in Theorems 4, and 6, respectively. Then: 1) The function $h(\boldsymbol{x}) = g_{TV}(\boldsymbol{x}) + \frac{1}{2\lambda}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2$ is convex for a proper range of λ. 2) The resolvent operator $(\boldsymbol{I} + \lambda\partial g_{TV})^{-1}$ is a singleton.

It is classically known that the sum of two invex functions (or quasi-invex and invex) is not necessarily invex (Mishra & Giorgi, 2008). Then, the above result is helpful to both the invexity and imaging communities, complementing Theorem 3. We present the proof of Theorem 8 in Appendix H, which uses the fact that any admissible function is point-wise non-increasing when divided by a quadratic mapping. This analysis helps extend the original ADMM algorithm in Section 4.3. A set of mappings called *prox-regular* or *weakly convex*, can be found in the literature (Wang et al., 2019b; Rockafellar & Wets, 2009), for which Theorem 8 holds - a function is prox-regular if program (11) is convex for some λ. However, it is crucial to discern a critical disparity between prox-regular functions and admissible mappings. Prox-regularity implies quasi-invexity (proven for the first time in Appendix H). This highlights the intrinsic advantage of admissible functions in attaining global optima, showcasing their prowess in contrast to prox-regular functions.

4.3 EXTENDED ALTERNATING DIRECTION METHOD OF MULTIPLIERS (ADMM)

The ADMM algorithm has a special place in a range of signal restoration and imaging problems under convex settings with linear equality constraints (Boyd et al., 2011). It is particularly favored for handling large-scale problems (Glowinski, 2014) with many constraints given its amenability for parallelization (Boyd et al., 2011). Some notable applications of ADMM include blind ptychography (Li, 2023), phase retrieval (Liang et al., 2017), computer tomography (Wei et al., 2022), unrolling (Xie et al., 2019), and magnetic resonance imaging (Sun et al., 2016).

The ADMM method solves problems such as in (1), where the constraint has the form $f(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y})$ (e.g. Section 4.1), and program (11) which is rewritten using $\boldsymbol{S} \in \mathbb{R}^{m\times n}$, $\boldsymbol{P} \in \mathbb{R}^{m\times p}$ as

$$\underset{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{z}\in\mathbb{R}^p}{\text{minimize}} \quad h_1(\boldsymbol{x}) + h_2(\boldsymbol{z}) \;\; \text{subject to } \boldsymbol{S}\boldsymbol{x} + \boldsymbol{P}\boldsymbol{z} = \boldsymbol{y}. \tag{12}$$

We provide details on how to cast (1) into program (12)in Appendix I. To solve (12), we form the scaled augmented Lagrangian as $\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v}) = h_1(\boldsymbol{x}) + h_2(\boldsymbol{z}) + \frac{\rho}{2}\|\boldsymbol{S}\boldsymbol{x} + \boldsymbol{P}\boldsymbol{z} - \boldsymbol{y} + \boldsymbol{v}\|_2^2$, where $\boldsymbol{v} \in \mathbb{R}^m$ is the dual variable, and $\rho > 0$. The optimization of $\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v})$ is summarized as

$$\begin{aligned}
\boldsymbol{x}^{(t+1)} &:= \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min} \left(h_1(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{S}\boldsymbol{x} + \boldsymbol{P}\boldsymbol{z}^{(t)} - \boldsymbol{y} + \boldsymbol{v}^{(t)}\|_2^2 \right) \\
\boldsymbol{z}^{(t+1)} &:= \underset{\boldsymbol{z}\in\mathbb{R}^p}{\arg\min} \left(h_2(\boldsymbol{z}) + \frac{\rho}{2}\|\boldsymbol{S}\boldsymbol{x}^{(t+1)} + \boldsymbol{P}\boldsymbol{z} - \boldsymbol{y} + \boldsymbol{v}^{(t)}\|_2^2 \right) \\
\boldsymbol{v}^{(t+1)} &:= \boldsymbol{v}^{(t)} + \boldsymbol{S}\boldsymbol{x}^{(t+1)} + \boldsymbol{P}\boldsymbol{z}^{(t+1)} - \boldsymbol{y}.
\end{aligned} \tag{13}$$

The convergence of the above algorithm is well-known and established for convex $h_1(\boldsymbol{x})$, and $h_2(\boldsymbol{z})$ (Boyd et al., 2011). However, the convergence properties of ADMM are unknown in the quasi-invex (hence invex) space. Here, we extend the ADMM algorithm so that its benefits are available to the signal restoration problems under settings based on Theorem 3, as follows.

Theorem 9. Let $h_1(\boldsymbol{x}), h_2(\boldsymbol{z})$ be admissible functions or quasi-invex constructs in Theorem 3, with $\rho\sigma_n(\boldsymbol{S}) \geq 1$, and $\rho\sigma_p(\boldsymbol{P}) \geq 1$. Assume there exists $(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{v}^*)$ for which $\mathcal{L}_\rho(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{v}) \leq \mathcal{L}_\rho(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{v}^*) \leq \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v}^*)$, for all $\boldsymbol{x}, \boldsymbol{z}$, and $\boldsymbol{v}$. Then, equation (13) satisfies: i) Residual $\|\boldsymbol{r}^{(t)}\|_2 = \|\boldsymbol{S}\boldsymbol{x}^{(t)} + \boldsymbol{P}\boldsymbol{z}^{(t)} - \boldsymbol{y}\|_2 \to 0$ and $\boldsymbol{v}^{(t)} \to \boldsymbol{v}^*$ as $t \to \infty$ at a rate of $\mathcal{O}(1/t)$, where $\boldsymbol{v}^*$ is the dual optimal point; ii) $h_1(\boldsymbol{x}^{(t)}) + h_2(\boldsymbol{z}^{(t)})$ approaches the optimal value.

Proof of Theorem 9 relies on Theorem 8 which ensures uniqueness in the estimations of $\boldsymbol{x}^{(t+1)}, \boldsymbol{z}^{(t+1)}$ in equation (13). This uniqueness is not obtained by prox-regular functions which highlights the significance of our proposed family of invex/quasi-invex functions. The proof is presented in Appendix I along with applications that satisfy the mild constraints on $\boldsymbol{S}, \boldsymbol{P}$. Appendix I also provides the proximal solution for the studied invex/quasi-invex functions along with their computational time. These solutions are used to perform the numerical experiments in Section 5.

5 Experimental Evaluation

We performed three experiments to evaluate the utility of the proposed invex/quasi-invex functions against the state-of-the-art, and we provide only the key results in Table 3. Additional set of results can be found in Appendix K. Across the experiments below we use the following metrics: absolute error (AbsError), peak signal-to-noise ratio (PSNR), root mean square error (RMSE), and structural similarity index measure (SSIM). Lower (higher) values for AbsError and RMSE (PSNR and SSIM) imply better signal restoration quality. In addition to AbsError and RMSE, we have also used square root of the AbsError (SqrtError), and log of the RMSE as additional metrics.

Experiment 1: Here, we focus on reconstructing three-dimensional volumes for Computed Tomography (CT). CT is a powerful imaging technique with applications to various domains, including medical and material sciences (Ginat & Gupta, 2014; Jørgensen et al., 2021). Mathematically, CT is an inverse program modeled as in (1) where the constraint is given by $f(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) \leq \epsilon$ with $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ describing the image formation model of the noisy measurements $\boldsymbol{y} \in \mathbb{R}^m$ (Xiang et al., 2021). Many model-based recovery algorithms have been proposed to solve (1) using $g(\boldsymbol{x})$ as the ℓ_1-norm such as the Iterative Shrinkage/Thresholding Algorithm, ADMM, and Primal-Dual Hybrid Gradient algorithm (Kamilov et al., 2023). However, the performance of the model-based strategies has been overcome by hybrid methods (i.e., unrolled networks) such as the Fast Iterative Shrinkage Thresholding Network (FISTA-Net) (Xiang et al., 2021). Here, we consider FISTA-Net as the state-of-the-art in CT due to its performance and because FISTA-Net solves (1) where $g(\boldsymbol{x})$ is the convex ℓ_1-norm and the training loss function $f(\boldsymbol{x})$ is the ℓ_2-norm. Then, the invex/quasi-invex regularizers and losses to be tested accordingly in this experiment are equations (3), (9), (10), and equations (4)-(8), respectively. To train FISTA-Net (with invex/quasi-invex functions above), we follow the experimental setting as in (Xiang et al., 2021). The training image dataset contains 2,378 full-dose CT images from eight of ten material sciences objects[1]. The reference images were reconstructed by `iradon` operator in MATLAB using all 720 views (Han & Ye, 2018; Jin et al., 2017). The validation set consists of 409 images and 330 slices for the test set.

Experiment 2: Here we focus on image spectral reconstruction (ST), which has profound applications in medical imaging, remote sensing, and material sciences (Wang et al., 2023). ST focuses on estimating the spectral reflectance of scenes from a set of RGB images captured at different wavelengths (Cai et al., 2022). Mathematically, spectral imaging is an inverse program modeled as in (1) where the constraint is given by $f(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) \leq \epsilon$ with $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ describing the image formation model of the noisy measurements $\boldsymbol{y} \in \mathbb{R}^m$ (Arce et al., 2013). Conventional model-based methods adopt hand-crafted priors such as sparsity, total variation, and non-local similarity to regularize the reconstruction procedure Arguello et al. (2023). However, these methods result in poor generalization ability due to convex regularizers like ℓ_1-norm, and unsatisfactory reconstruction quality (Cai et al., 2022). Deep learning methods such as the Multi-stage Spectral-wise Transformer (MST++) (Cai et al., 2022) has shown to exhibit the best performance because it does not employ regularizer. Originally the fidelity term $f(\boldsymbol{x})$ of MST++ is the ℓ_2-norm. Then, the invex losses to be tested here are equations (4)-(8). To train MST++, we followed the setting in (Cai et al., 2022)

[1]Acquired with the ISIS Neutron and Muon Source at the Rutherford Appleton Laboratory

Table 3: Performance Results: Best: green, and the worst: red.

Experiment 1 (Combination of functions used to train FISTA-Net)							
Regularizer	Metrics	equation (4)	equation (5)	equation (6)	equation (7)	equation (8)	ℓ_2-norm
equation (3)	AbsError	0.1676	0.1331	0.1543	0.1246	0.1429	0.1834
	PSNR	39.33	40.40	39.68	40.78	40.04	38.99
	RMSE	0.0101	0.0075	0.0086	0.0080	0.0093	0.0110
	SSIM	0.9590	0.9602	0.9608	0.9620	0.9614	0.9596
equation (9)	AbsError	0.3307	0.2909	0.3112	0.2798	0.3079	0.3464
	PSNR	37.75	38.32	38.00	38.65	38.47	37.50
	RMSE	0.0221	0.0176	0.0156	0.0132	0.0116	0.0286
	SSIM	0.9523	0.9551	0.9542	0.9570	0.9561	0.9533
equation (10)	AbsError	0.1859	0.1716	0.1493	0.1401	0.1601	0.2025
	PSNR	39.00	39.96	39.33	40.71	39.55	38.68
	RMSE	0.0113	0.0081	0.0101	0.0087	0.0096	0.0125
	SSIM	0.9577	0.9590	0.9597	0.9610	0.9603	0.9583
(Xiang et al., 2021) ℓ_1-norm	AbsError	0.4106	0.4004	0.3814	0.3726	0.3906	0.4213
	PSNR	37.45	37.94	37.69	38.44	38.18	37.22
	RMSE	0.0290	0.0130	0.0222	0.0180	0.0151	0.0420
	SSIM	0.9510	0.9540	0.9530	0.9560	0.9550	0.9520
Experiment 2 (Loss functions used to train MST++)							
Metrics		equation (4)	equation (5)	equation (6)	equation (7)	equation (8)	ℓ_2-norm
AbsError		0.1975	0.2056	0.2241	0.1830	0.1900	0.2145
SqrtError		1.3752	1.3982	1.4465	1.4219	1.3530	1.3315
RMSE		2.6847	2.9390	3.2465	2.1315	2.4710	2.2887
LogRMSE		0.2683	0.2851	0.2943	0.2534	0.2606	0.2765
Experiment 3 (Total variation filtering using ADMM)							
Metrics		TV-ℓ_p	Invex TV-LL_p	Quasi-invex TV-LL_p	ℓ_1-norm		
SSIM		0.6137	0.6320	0.6227	0.6050		
MS-SSIM		0.9192	0.9235	0.9149	0.9106		
ADMM-residual		2.3×10^{-3}	1.8×10^{-3}	2.0×10^{-3}	2.8×10^{-3}		

and Appendix K. The image dataset is from (Arad et al., 2022), which contains 1,000 RGB-spectral pairs. This dataset is split into the train, valid, and test subsets in the ratio of 18:1:1.

Experiment 3: This experiment focuses on solving (11). As we mentioned in Section 4.2 the classical total variation filtering, with global optima guarantees, is defined using the ℓ_1-norm, then the invex/quasi-invex regularizers to be tested are the TV-regularizer version of ℓ_p-quasinorm (TV-ℓ_p), and equation (3) (Invex TV-LL_p) equation (10) (Quasi-invex TV-LL_p). We defer the exact details around the ADMM implementation to the supplementary material (Section K). We fixed the number of iterations of the ADMM method to 1,000 for all tested regularizers, and we included the smallest residual value $\|\boldsymbol{r}^{(t)}\|_2$ across iterations to characterize the accuracy of the restored image. The dataset employed here has 40 images[1]. These images contain the neutron attenuation properties of the object, which helps analyze the material structure. Here, in addition to SSIM, we also use Multi-scale SSIM (MS-SSIM) and residual error of the ADMM (ADMM-residual) as metrics.

6 Discussion, Limitations and Conclusions

In this paper, by defining a family of admissible functions with relevant properties, we identified invex/quasi-invex functions to support real-world signal processing problems - signal restoration. We then provided proof for these functions' invex/quasi-invex behaviors and global optimality beyond the realm of convexity with their convergence rate(s). Specifically, the breakthrough of Theorem 3 compared with current invex literature lies in that the sum of two proposed invex/quasi-invex functions is also invex/quasi-invex. And the results in Theorem 9 and Appendix J show that both ADMM and APGM methods with invex/quasi-invex losses, *are as efficient as solving the well-established convex optimization problems*.

We provided detailed evaluations in Table 3 and Appendix K of these functions in imaging tasks such as computer tomography, spectral imaging, and total-variation regularization. These results show significant benefits of using the proposed family of invex/quasi-invex functions from theoretical and empirical aspects. Specifically, they overcome the performance of convex mappings in terms of reconstruction quality across several relevant metrics and the studied imaging tasks. The superiority of the tested invex/quasi-invex functions is because they provide a more accurate model of the sparsity and abnormal patterns in the data than convex mappings (Sun & Pong, 2022).

While our theoretical analysis to handle signal restoration problems is the first of its kind and performs well on a number of tasks, its applications beyond imaging tasks are yet to be explored. More specifically, the application of these algorithms around deep learning research yet to be explored, which can pave a way to improve several downstream applications, such as low-rank matrix recovery and alike. As a pure algorithm, a number of things can be improved, such as decoupling the η from $\upsilon_{f,\eta}(x)$, improving the efficiency of the algorithm beyond what ADMM and APGM can offer.

REFERENCES

Cupy library. https://cupy.dev/. Accessed: 2023-09-11.

Cupy kernel. https://docs.cupy.dev/en/stable/user_guide/kernel.html. Accessed: 2023-09-11.

Kodak image dataset. http://r0k.us/graphics/kodak/. Accessed: 2023-08-08.

Pytorch and cupy. https://docs.cupy.dev/en/stable/user_guide/interoperability.html#using-custom-kernels-in-pytorch. Accessed: 2023-09-11.

Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Luc Van Gool, Shuai Liu, Yongqiang Li, Chaoyu Feng, Lei Lei, Jiaojiao Li, Songcheng Du, Chaoxiong Wu, Yihong Leng, Rui Song, Mingwei Zhang, Chongxing Song, Shuyi Zhao, Zhiqiang Lang, Wei Wei, Lei Zhang, Renwei Dian, Tianci Shan, Anjing Guo, Chengguo Feng, Jinyang Liu, Mirko Agarla, Simone Bianco, Marco Buzzelli, Luigi Celona, Raimondo Schettini, Jiang He, Yi Xiao, Jiajun Xiao, Qiangqiang Yuan, Jie Li, Liangpei Zhang, Taesung Kwon, Dohoon Ryu, Hyokyoung Bae, Hao-Hsiang Yang, Hua-En Chang, Zhi-Kai Huang, Wei-Ting Chen, Sy-Yen Kuo, Junyu Chen, Haiwei Li, Song Liu, Sabarinathan, K Uma, B Sathya Bama, and S. Mohamed Mansoor Roomi. NTIRE 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 863–881, June 2022.

Gonzalo R Arce, David J Brady, Lawrence Carin, Henry Arguello, and David S Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1): 105–115, 2013.

Henry Arguello, Jorge Bacca, Hasindu Kariyawasam, Edwin Vargas, Miguel Marquez, Ramith Hettiarachchi, Hans Garcia, Kithmini Herath, Udith Haputhanthri, Balpreet Singh Ahluwalia, et al. Deep optical coding design in computational imaging: a data-driven framework. *IEEE Signal Processing Magazine*, 40(2):75–88, 2023.

Adil Bagirov, Napsu Karmitsa, and Marko M Mäkelä. *Introduction to Nonsmooth Optimization: theory, practice and software*. Springer, 2014.

Adarsh Barik and Jean Honorio. Fair sparse regression with clustering: An invex relaxation for a combinatorial problem. *Advances in Neural Information Processing Systems*, 34, 2021.

Jonathan T Barron. A general and adaptive robust loss function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4331–4339, 2019.

Mokhtar S Bazaraa and Chitharanjan M Shetty. *Foundations of optimization*, volume 122. Springer Science & Business Media, 2012.

Amir Beck. *First-order methods in optimization*. SIAM, 2017.

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

Radu Ioan Boţ, Ernö Robert Csetnek, and Szilárd Csaba László. An inertial forward–backward algorithm for the minimization of the sum of two nonconvex functions. *EURO Journal on Computational Optimization*, 4(1):3–25, 2016.

Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

Richard L Burden and J Douglas Faires. 2.1 the bisection algorithm. *Numerical analysis*, 3, 1985.

T Tony Cai and Lie Wang. Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Transactions on Information theory*, 57(7):4680–4688, 2011.

Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 745–755, 2022.

Ariolfo Camacho, Edwin Vargas, and Henry Arguello. Hyperspectral and multispectral image fusion addressing spectral variability by an augmented linear mixing model. *International Journal of Remote Sensing*, 43(5):1577–1608, 2022.

EJ Candès and Benjamin Recht. The trace norm and nuclear norm of a matrix. *Communications on Pure and Applied Mathematics*, 59(7):849–883, 2006.

Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 14(5):877–905, 2008.

Rafael E Carrillo, Kenneth E Barner, and Tuncer C Aysal. Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):392–408, 2010.

Rafael E Carrillo, Ana B Ramirez, Gonzalo R Arce, Kenneth E Barner, and Brian M Sadler. Robust compressive sensing of sparse signals: A review. *EURASIP Journal on Advances in Signal Processing*, 2016:1–17, 2016.

Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20:89–97, 2004.

Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

Juan Carlos De los Reyes, C-B Schönlieb, and Tuomo Valkonen. Bilevel parameter learning for higher-order total variation regularisation models. *Journal of Mathematical Imaging and Vision*, 57(1):1–25, 2017.

Wei Deng, Ming-Jun Lai, Zhimin Peng, and Wotao Yin. Parallel multi-block admm with $\mathcal{O}(1/k)$ convergence. *Journal of Scientific Computing*, 71:712–736, 2017.

Jie Ding, Vahid Tarokh, and Yuhong Yang. Model selection techniques: An overview. *IEEE Signal Processing Magazine*, 35(6):16–34, 2018.

Matthias J Ehrhardt and Marta M Betcke. Multicontrast MRI reconstruction with structure-guided total variation. *SIAM Journal on Imaging Sciences*, 9(3):1084–1106, 2016.

Vania Vieira Estrela, Hermes Aguiar Magalhães, and Osamu Saotome. Total variation applications in computer vision. In *Handbook of Research on Emerging Perspectives in Intelligent Pattern Recognition, Analysis, and Image Processing*, pp. 41–64. IGI Global, 2016.

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. 2013. ISBN 9780817649487 0817649484. doi: 10.1007/978-0-8176-4948-7.

Pierre Frankel, Guillaume Garrigos, and Juan Peypouquet. Splitting methods with variable metric for kurdyka–łojasiewicz functions and general convergence rates. *Journal of Optimization Theory and Applications*, 165(3):874–900, 2015.

Yun Fu. *Low-rank and sparse modeling for visual analysis*. Springer, 2014.

Donald Geman and Chengda Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE transactions on Image Processing*, 4(7):932–946, 1995.

Daniel Thomas Ginat and Rajiv Gupta. Advances in computed tomography imaging technology. *Annual review of biomedical engineering*, 16:431–453, 2014.

Roland Glowinski. On alternating direction methods of multipliers: a historical perspective. *Modeling, simulation and optimization for science and technology*, pp. 59–82, 2014.

Rémi Gribonval and Morten Nielsen. Highly sparse representations from dictionaries are unique and independent of the sparseness measure. *Applied and Computational Harmonic Analysis*, 22 (3):335–355, 2007.

Andrés Guerrero, Samuel Pinilla, and Henry Arguello. Phase recovery guarantees from designed coded diffraction patterns in optical imaging. *IEEE Transactions on Image Processing*, 29:5687–5697, 2020.

Yoseob Han and Jong Chul Ye. Framing U-Net via deep convolutional framelets: Application to sparse-view CT. *IEEE transactions on medical imaging*, 37(6):1418–1429, 2018.

Morgan A Hanson. On sufficiency of the Kuhn-Tucker conditions. *Journal of Mathematical analysis and applications*, 80(2):545–550, 1981.

Charles R Harris, K Jarrod Millman, Stéfan J Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, et al. Array programming with numpy. *Nature*, 585(7825):357–362, 2020.

Peter Hedman, Pratul P Srinivasan, Ben Mildenhall, Jonathan T Barron, and Paul Debevec. Baking neural radiance fields for real-time view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5875–5884, 2021.

Yinan Hu and Ivan W Selesnick. Nonconvex Haar-TV denoising. *Digital Signal Processing*, 107: 102855, 2020.

Marian-Daniel Iordache, José M Bioucas-Dias, and Antonio Plaza. Total variation spatial regularization for sparse hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4484–4502, 2012.

Andrés Jerez, Samuel Pinilla, and Henry Arguello. Fast target detection via template matching in compressive phase retrieval. *IEEE Transactions on Computational Imaging*, 6:934–944, 2020.

Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9): 4509–4522, 2017.

Jeremy Johnston, Yinchuan Li, Marco Lops, and Xiaodong Wang. ADMM-Net for communication interference removal in stepped-frequency radar. *IEEE Transactions on Signal Processing*, 69: 2818–2832, 2021.

Jakob S Jørgensen, Evelina Ametova, Genoveva Burca, Gemma Fardell, Evangelos Papoutsellis, Edoardo Pasca, Kris Thielemans, Martin Turner, Ryan Warr, William RB Lionheart, et al. Core imaging library-part I: A versatile python framework for tomographic imaging. *Philosophical Transactions of the Royal Society A*, 379(2204):20200192, 2021.

Ulugbek S Kamilov, Charles A Bouman, Gregery T Buzzard, and Brendt Wohlberg. Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 40(1):85–97, 2023.

RN Kaul and Surjeet Kaur. Optimality criteria in nonlinear programming involving nonconvex functions. *Journal of Mathematical Analysis and Applications*, 105(1):104–112, 1985.

Razieh Keshavarzian, Ali Aghagolzadeh, and Tohid Yousefi Rezaii. ll_p norm regularization based group sparse representation for image compressed sensing recovery. *Signal Processing: Image Communication*, 78:477–493, 2019.

Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

Dmitry Kovalev and Alexander Gasnikov. The first optimal algorithm for smooth and strongly-convex-strongly-concave minimax optimization. *arXiv preprint arXiv:2205.05653*, 2022.

Amine Laghrib, Lekbir Afraites, Aissam Hadri, and Mourad Nachaoui. A non-convex PDE-constrained denoising model for impulse and gaussian noise mixture reduction. *Inverse Problems and Imaging*, 17(1):23–67, 2023.

Huan Li and Zhouchen Lin. Accelerated proximal gradient methods for nonconvex programming. *Advances in neural information processing systems*, 28, 2015.

Ji Li. Solving blind ptychography effectively via linearized alternating direction method of multipliers. *Journal of Scientific Computing*, 94(1):19, 2023.

Junli Liang, Petre Stoica, Yang Jing, and Jian Li. Phase retrieval via the alternating direction method of multipliers. *IEEE Signal Processing Letters*, 25(1):5–9, 2017.

Fan Lin, Yingpin Chen, Lingzhi Wang, Yuqun Chen, Wei Zhu, and Fei Yu. An efficient image reconstruction framework using total variation regularization with ℓ_p-quasinorm and group gradient sparsity. *Information*, 10(3):115, 2019.

Xingguo Liu, Yingpin Chen, Zhenming Peng, Juan Wu, and Zhuoran Wang. Infrared image super-resolution reconstruction based on quaternion fractional order total variation with l_p quasinorm. *Applied Sciences*, 8(10):1864, 2018.

Trang C Mai, Hien Quoc Ngo, and Le-Nam Tran. Energy-efficient power allocation in cell-free massive MIMO with zero-forcing: First order methods. *Physical Communication*, 51:101540, 2022.

Goran Marjanovic and Victor Solo. On ℓ_q optimization and matrix completion. *IEEE Transactions on signal processing*, 60(11):5714–5724, 2012.

Anke Meyer-Bäse, Anke Meyer-Baese, and Volker J Schmid. *Pattern Recognition and Signal Analysis in Medical Imaging*. Academic Press, 2004.

Shashi K Mishra and Giorgio Giorgi. *Invexity and optimization*, volume 88. Springer Science & Business Media, 2008.

Frederick Mosteller and John W Tukey. Data analysis and regression. A second course in statistics. *Addison-Wesley series in behavioral science: quantitative methods*, 1977.

Giorgos Mountrakis, Jungho Im, and Caesar Ogole. Support vector machines in remote sensing: A review. *ISPRS journal of photogrammetry and remote sensing*, 66(3):247–259, 2011.

Peter Ochs, Yunjin Chen, Thomas Brox, and Thomas Pock. iPiano: Inertial proximal algorithm for nonconvex optimization. *SIAM Journal on Imaging Sciences*, 7(2):1388–1419, 2014.

Yuyuan Ouyang, Yunmei Chen, Guanghui Lan, and Eduardo Pasiliao Jr. An accelerated linearized alternating direction method of multipliers. *SIAM Journal on Imaging Sciences*, 8(1):644–681, 2015.

Diethard Ernst Pallaschke and Stefan Rolewicz. *Foundations of mathematical optimization: convex analysis without linearity*, volume 388. Springer Science & Business Media, 2013.

Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3): 127–239, 2014.

Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5865–5874, 2021.

Samuel Pinilla, Tingting Mu, Neil Bourne, and Jeyan Thiyagalingam. Improved imaging by invex regularizers with global optima guarantees. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 10780–10794. Curran Associates, Inc., 2022a. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/45f7927942098d14e473fc5d000031e2-Paper-Conference.pdf.

Samuel Pinilla, Seyyed Reza Miri Rostami, Igor Shevkunov, Vladimir Katkovnik, and Karen Egiazarian. Hybrid diffractive optics design via hardware-in-the-loop methodology for achromatic extended-depth-of-field imaging. *Optics Express*, 30(18):32633–32649, 2022b.

Samuel Pinilla, Kumar Vijay Mishra, Igor Shevkunov, Mojtaba Soltanalian, Vladimir Katkovnik, and Karen Egiazarian. Unfolding-aided bootstrapped phase retrieval in optical imaging: Explainable AI reveals new imaging frontiers. *IEEE Signal Processing Magazine*, 40(2):46–60, 2023.

Apostolos F Psaros, Kenji Kawaguchi, and George Em Karniadakis. Meta-learning PINN loss functions. *Journal of Computational Physics*, 458:111121, 2022.

Juan Marcos Ramirez, José Ignacio Martínez-Torre, and Henry Arguello. LADMM-Net: An unrolled deep network for spectral image fusion from compressive data. *Signal Processing*, 189: 108239, 2021.

Thomas W Reiland. Nonsmooth invexity. *Bulletin of the Australian Mathematical Society*, 42(3): 437–446, 1990.

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

Carola-Bibiane Schönlieb, Andrea Bertozzi, Martin Burger, and Lin He. Image inpainting using a fourth-order total variation flow. In *SAMPTA'09*, pp. Special–session, 2009.

Ivan Selesnick, Alessandro Lanza, Serena Morigi, and Fiorella Sgallari. Non-convex total variation regularization for convex denoising of signals. *Journal of Mathematical Imaging and Vision*, 62 (6-7):825–841, 2020.

Yunquan Song, Zitong Li, and Minglu Fang. Robust variable selection based on penalized composite quantile regression for high-dimensional single-index models. *Mathematics*, 10(12):2000, 2022.

Jian Sun, Huibin Li, Zongben Xu, et al. Deep ADMM-Net for compressive sensing MRI. *Advances in neural information processing systems*, 29, 2016.

Shuqin Sun and Ting Kei Pong. Doubly iteratively reweighted algorithm for constrained compressed sensing models. *arXiv preprint arXiv:2206.08205*, 2022.

Yuli Sun, Hao Chen, Jinxu Tao, and Lin Lei. Computed tomography image reconstruction from few views via log-norm total variation minimization. *Digital Signal Processing*, 88:172–181, 2019.

Mujahid Syed, Panos Pardalos, and Jose Principe. Invexity of the minimum error entropy criterion. *IEEE Signal Processing Letters*, 20(12):1159–1162, 2013.

Mujahid N Syed, Panos M Pardalos, and Jose C Principe. On the optimization properties of the correntropic loss function in data analysis. *Optimization Letters*, 8(3):823–839, 2014.

Vladimir Tankovich, Christian Hane, Yinda Zhang, Adarsh Kowdle, Sean Fanello, and Sofien Bouaziz. HITNET: Hierarchical iterative tile refinement network for real-time stereo matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14362–14372, 2021.

Edwin Vargas, Oscar Espitia, Henry Arguello, and Jean-Yves Tourneret. Spectral image fusion from compressive measurements. *IEEE Transactions on Image Processing*, 28(5):2271–2282, 2018a.

Edwin Vargas, Samuel Pinilla, Jorge Bacca, and Henry Arguello. Robust formulation for solving underdetermined random linear system of equations via ADMM. In *2018 IEEE Statistical Signal Processing Workshop (SSP)*, pp. 283–287. IEEE, 2018b.

Edwin Vargas, Henry Arguello, and Jean-Yves Tourneret. Spectral image fusion from compressive measurements using spectral unmixing and a sparse representation of abundance maps. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7):5043–5053, 2019.

Peter Vouras, Kumar Vijay Mishra, Alexandra Artusio-Glimpse, Samuel Pinilla, Angeliki Xenaki, David W Griffith, and Karen Egiazarian. An overview of advances in signal processing techniques for classical and quantum wideband synthetic apertures. *arXiv preprint arXiv:2205.05602*, 2022.

Jiaxi Wang, Li Zeng, Chengxiang Wang, and Yumeng Guo. Admm-based deep reconstruction for limited-angle CT. *Physics in Medicine & Biology*, 64(11):115011, 2019a.

Mengyuan Wang, Wei He, and Hongyan Zhang. A spatial-spectral transformer network with total variation loss for hyperspectral image denoising. *IEEE Geoscience and Remote Sensing Letters*, 2023.

Rongfang Wang, Yali Qin, Zhenbiao Wang, and Huan Zheng. Group-based sparse representation for compressed sensing image reconstruction with joint regularization. *Electronics*, 11(2):182, 2022.

Weina Wang and Yunmei Chen. An accelerated smoothing gradient method for nonconvex nonsmooth minimization in image processing. *Journal of Scientific Computing*, 90(1):1–28, 2022.

Yu Wang, Wotao Yin, and Jinshan Zeng. Global convergence of ADMM in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, 78:29–63, 2019b.

Kaixuan Wei, Angelica Aviles-Rivero, Jingwei Liang, Ying Fu, Hua Huang, and Carola-Bibiane Schönlieb. TFPNP: Tuning-free plug-and-play proximal algorithms with applications to inverse imaging problems. *Journal of Machine Learning Research*, 23(16):1–48, 2022.

Fei Wen, Lei Chu, Peilin Liu, and Robert C Qiu. A survey on nonconvex regularization-based sparse and low-rank recovery in signal processing, statistics, and machine learning. *IEEE Access*, 6:69883–69906, 2018.

Joseph Woodworth and Rick Chartrand. Compressed sensing recovery via nonconvex shrinkage penalties. *Inverse Problems*, 32(7):075004, 2016.

Rui Wu and Di-Rong Chen. The improved bounds of restricted isometry constant for recovery via ℓ_p-minimization. *IEEE transactions on information theory*, 59(9):6142–6147, 2013.

Jinxi Xiang, Yonggui Dong, and Yunjie Yang. FISTA-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE Transactions on Medical Imaging*, 40(5): 1329–1339, 2021.

Xingyu Xie, Jianlong Wu, Guangcan Liu, Zhisheng Zhong, and Zhouchen Lin. Differentiable linearized ADMM. In *International Conference on Machine Learning*, pp. 6902–6911. PMLR, 2019.

Constantin Zălinescu. A critical view on invexity. *Journal of Optimization Theory and Applications*, 162(3):695–704, 2014.

Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics*, 38(2):894–942, 2010.

Zhihua Zhang, James T Kwok, and Dit-Yan Yeung. Surrogate maximization/minimization algorithms and extensions. *Machine Learning*, 69(1):1–33, 2007.

Abdelhak M Zoubir, Visa Koivunen, Yacine Chakhchoukh, and Michael Muma. Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts. *IEEE Signal Processing Magazine*, 29(4):61–80, 2012.