\textbf{\large $\ell_1$-Norm Minimization with Regula Falsi Type Root Finding Methods}

Metin Vural, Aleksandr Y. Aravkin, and Sławomir Stańczak

Abstract—Sparse level-set formulations allow practitioners to find the minimum $1$-norm solution subject to likelihood constraints. Prior art requires this constraint to be convex. In this letter, we develop an efficient approach for nonconvex likelihoods, using Regula Falsi root-finding techniques to solve the level-set formulation. Regula Falsi methods are simple, derivative-free, and efficient, and the approach provably extends level-set methods to the broader class of nonconvex inverse problems. Practical performance is illustrated using $\ell_1$-regularized Student’s $t$ inversion, which is a nonconvex approach used to develop outlier-robust formulations.

Index Terms—$\ell_1$-norm minimization, nonconvex models, Regula-Falsi, root-finding

I. INTRODUCTION

S\textsc{parse} recovery using $\ell_1$-norm minimization plays a major role in many signal processing applications. Denoting $y \in \mathbb{R}^M$ as a measurement vector, $D \in \mathbb{R}^{M \times N}$ as an overcomplete matrix with $M < N$, and $\rho$ as the penalty that measures the data misfit, the ‘noise-aware’ level-set problem is to minimize $\ell_1$-norm subject to a misfit or likelihood constraint:

$$ \text{(P}_\sigma) \quad \text{minimize} \quad \|x\|_1 \quad \text{s.t.} \quad \rho(y - Dx) \leq \sigma,$$

where $\sigma$ indicates the noise level. $P_\sigma$ is used in many applications, including compressed sensing [1]–[2], overcomplete signal representation [3], [4], coding theory [5], and image processing [6].

An efficient way to solve $(P_\sigma)$ is to develop an explicit relationship with a simpler problem that can be directly solved with primal-only methods, such as the prox-gradient algorithm [7]:

$$ \text{(P}_\tau) \quad \text{minimize} \quad \rho(y - Dx) \quad \text{s.t.} \quad \|x\|_1 \leq \tau.$$

Exploiting the relationship between $P_\sigma$ and $P_\tau$ allows one to specify the noise tolerance $\sigma$, and then find the solution by inexactly optimizing a sequence of simpler $(P_\tau)$ problems.

It has been known for a long time that $(P_\tau)$ and $(P_\sigma)$ can provide equivalent solutions [8], and the idea of solving $(P_\tau)$ to obtain the solution of $(P_\sigma)$ was first proposed by [9], [10]. Their idea follows the optimality trade-off between the minimum $\ell_1$-norm and the least squares data misfit, which generates a differentiable convex Pareto frontier. This optimality tracing is formulated as a non-linear equation root finding problem, i.e. getting the exact $\tau$ for a given noise tolerance $\sigma$, and is solved by an inexact Newton Method. The resulting level-set approach has been generalized to other instances of convex programming by [2], [11].

In the most general case, the relationship between $(P_\tau)$ and $(P_\sigma)$ does not require convexity [11, Theorem 2.1]. However, practical implementations of the root-finding approach require convexity of the Pareto frontier to guarantee success of the root finding procedure, limiting the approach to the convex case. Current implementations favor Newton’s method, which requires derivatives. To address this issue, an extension using an inexact secant method has also been developed [7].

In this paper, we introduce Regula Falsi type derivative-free non-linear equation root finding schemes to solve $(P_\sigma)$. They are bracketing type methods that offer convergence guarantee for convex and nonconvex models with the proper choices of root searching interval: two initial points with the opposite signs assures convergence [12]. Regula Falsi type methods do not require convexity to trace the root, allowing nonconvex loss functions in the $(P_\sigma)$ formulations. Finally, these methods are also derivative free. All of these advantages allow Regula Falsi type methods to be applied to cases where Newton, secant, and their variants are not guaranteed to converge.

Moving outside of the convex class opens the way for using many useful nonconvex models in $(P_\sigma)$ formulations. For example, in [13] and [14] consider mixture models whose negative log-likelihood are nonconvex, with applications to high-dimensional inhomogeneous data where number of covariates could be larger than sample size. A second application area uses nonconvex Student’s $t$ likelihoods to develop outlier-robust approaches [15]–[17]. In this paper, we show how Regula Falsi type root finding methods can be used with the nonconvex Student’s $t$ loss, as well the convex least-squares and Huber losses.

This paper is organized as follows. In Section II a Pareto frontier that reveals the relation between $(P_\tau)$ and $(P_\sigma)$ is defined and Regula Falsi type methods are introduced. Section III presents the proposed $(P_\sigma)$ solver while the simulation results are discussed in Section IV.

II. PARETO FRONTIER AND REGULA FALSI-TYPE ROOT FINDING METHODS

Under simple ‘active constraint’ conditions, problems $(P_\tau)$ and $(P_\sigma)$ are equivalent for some pair $(\tau, \sigma)$ [11]. Pareto frontier approaches use root finding and inexact solutions to a sequence of $(P_\tau)$ to solve $(P_\sigma)$.

A. Pareto Optimality

Definition 1: i) Pareto optimal is the minimal achievable feasible point of a feasible set. ii) The set that comprised Pareto optimal points is called the Pareto frontier.

In this work, we also seek to solve $(P_\sigma)$ by working with $(P_\tau)$. Specifically, we are interested in the optimal objective...
value of the (Pₜ) for a given y and τ which can be expressed with following
\[ \nu(\tau) := \inf_x \{ \rho(Dx - y) : \|x\|_1 \leq \tau \}, \]  
(1)
and the corresponding Pareto frontier can be defined as
\[ \psi(\tau) := \nu(\tau) - \sigma. \]  
(2)

**Theorem 1:** i) If ρ is a convex function (e.g. ℓ₂-norm, Huber function), then so is ψ. ii) If ρ is a nonconvex function, convexity of ψ does not follow. When ρ is quasi-convex function, then so is ψ.

*Proof 1:* Let us consider any two solutions x₁ and x₂ of (Pₜ) for any τ₁ and τ₂ respectively. Since ℓ₁-norm is convex, for any β ∈ [0, 1] following holds
\[ \|βx₁ + (1 - β)x₂\|_1 \leq β\|x₁\|_1 + (1 - β)\|x₂\|_1 \]
(3)
An immediate outcome of eq. (3) is that βx₁ + (1 - β)x₂ is a feasible point of (Pₜ) with τ = βτ₁ + (1 - β)τ₂. Thus we can write the following inequality
\[ \nu(βτ₁ + (1 - β)τ₂) \leq ρ(D(βx₁ + (1 - β)x₂) - y) \]
\[ = ρ(β(Dx₁ - y) + (1 - β)(Dx₂ - y)). \]
(4)
i) If ρ is convex, then
\[ ρ(β(Dx₁ - y) + (1 - β)(Dx₂ - y)) \leq ρ(Dx₁ - y) + (1 - β)ρ(Dx₂ - y), \]
that shows ψ is convex as well as ψ.

ii) If ρ is quasi-convex, then
\[ ρ(β(Dx₁ - y) + (1 - β)(Dx₂ - y)) \leq \max\{ρ(Dx₁ - y), ρ(Dx₂ - y)\} = \max\{ν(τ₁), ν(τ₂)\}, \]
(6)
that shows ψ is quasi-convex as well as ψ.

Pareto optimal points are unique for (Pₜ) with convex and quasi-convex losses ρ that can be inferred from [18] Theorem 1.1, Theorem 1.2, [19]. Also, the feasible set of (Pₜ) enlarges as τ increases, thus ψ(τ) is nonincreasing [20].

In Fig. 1 an abstract pareto frontier (vertical line) is depicted for convex and quasi-convex losses ρ where the red line represents the τ level.

Obtaining the solution of (Pₜ) by solving (Pₜ) proceeds as follows. We start with a τ parameter to solve (Pₜ), and using the solution of (Pₜ), find a new τ value. We proceed iteratively until ψ(τₐ) → 0. τₐ occurs at the intersection of the red line and black curve in Fig 1 where we immediately see that the solution of (Pₜ) is also a solution of the (Pₜ), a fact proven formally by [11]. Finding τₐ can be formulated as a nonlinear root finding problem.

**B. Regula Falsi Type Methods**

Our aim is to find \( \tau \) such that \( \psi(\tau) = 0 \).

If ρ is nonconvex, neither Newton’s method nor secant variants are guaranteed to solve [7]. In particular, the tangent lines may cross in the feasible region, and secant lines may not bracket the feasible area. In contrast, regardless of shape of the ρ, bracketing type root finding methods are guaranteed to solve [7]. Here, we develop Regula Falsi type methods for [7].

We denote the solution of a nonlinear equation of \( f \) by \( x^* \), i.e \( f(x^*) = 0 \). With this notation, Regula Falsi type methods starting with the points \( a \) and \( b \) proceed as follows.

1) Calculate the secant line between \( a \) and \( b \),
\[ s_{ab} = \frac{f(b) - f(a)}{b - a}, \]
(8)
and find the point where \( s_{ab} \) intersects the x-axis, which is \( c = b - \frac{f(b)}{s_{ab}} \).

2) Calculate \( f(c) \). If \( f(c) = 0 \) then \( x^* = c \), otherwise continue.

3) Adjust the new interval: if \( f(c)f(b) < 0 \), \( x^* \) should be in between \( b \) and \( c \). Set
\[ a = b, \quad b = c, \quad \text{and} \quad f(a) = f(b), \]
(9)
if \( f(c)f(b) > 0 \), \( x^* \) should be in between \( a \) and \( c \). Set
\[ b = c, \quad \text{and} \quad f(a) = \mu f(a), \]
(10)
where \( \mu \) is the scaling factor.

4) Check the ending condition: if \( |b - a| \leq \epsilon \), stop the iteration. Take
\[ x^* = \begin{cases} b, & \text{if } |f(b)| \leq |f(a)| \\ a, & \text{if } |f(b)| > |f(a)| \end{cases}, \]
(11)
if \( |b - a| > \epsilon \), continue the iteration, go back to 1) with the values \( a, b \) and \( f(a), f(b) \) from 3).

Regula Falsi type methods differ from each other in the choice of the scaling factor \( \mu \). Several commonly considered \( \mu \) in the literature is summarized in Table I. Additional options for \( \mu \) are studied in [21, 22].

**III. SOLVING (Pₜ)**

**A. (Pₜ) Solver**

In order to solve (Pₜ), we repeatedly solve (Pₜ). (Pₜ) can be solved using the simple projected gradient method
\[ x^{(k)} = \text{proj}_{(y)\mathbb{R}_+} \left( x^{(k-1)} + \gamma D^T \nabla \rho(y - Dx^{(k-1)}) \right), \]
(12)
1) Projection onto the $l_1$-ball: Projection of a vector $\mathbf{a} = [a_1, a_2, ..., a_N]$ onto the $l_1$-ball can be written as following

$$\text{proj}(\mathbf{a}, \tau) = \begin{cases} 
\mathbf{a}, & \text{if } \|\mathbf{a}\|_1 \leq \tau \\
\text{sgn}(a_i)\max\{|a_i| - \kappa, 0\}, & \text{else}
\end{cases}$$

(13)

where $\kappa$ is the Lagrangian multiplier of $\text{proj}_{1}(\mathbf{a}, \tau)$ [23]. The tricky part of the projection is to find the $\kappa$ that satisfies Karush-Kuhn-Tucker optimality condition $\sum_i a_i = \tau^N$ in an efficient way.

To find $\kappa$, we utilized the simple, sorting based approach introduced in [24]. Additional variations of this method are described in [25].

To find $\tau$:
- Sort $|a|$ as: $c_1 \geq c_2 \geq ... \geq c_N$.
- Find $K = \max_{1 \leq k \leq N} \{k \mid \sum_{j=1}^{k} c_j - \tau \leq c_k\}$.
- Calculate $\kappa = \left(\sum_{k=1}^{K} c_k - \tau\right) / K$.

To solve $(P_\tau)$, we used a projected gradient method with the spectral line search strategy discussed by [9], with the projection steps given in [III-A1].

B. Solving $(P_\sigma)$

1) Bracketing: In order to choose the root searching interval, we consider a well-known decomposition method called method of frames (mof) [26]. The mof decomposition of a signal $\mathbf{y}$ can be obtained with the inverse linear mapping such that $\mathbf{x}_{MF} = \mathbf{D}^T \mathbf{D}^{-1} \mathbf{y} = \arg \min \{\|\mathbf{x}\|_2 \mid \mathbf{D} \mathbf{x} = \mathbf{y}\}$. Many common loss functions $\rho$ are nonnegative and vanish at the origin, including gauges and nonconvex losses considered in this study. For these losses, under the assumption that $\mathbf{D}$ is full row-rank, $\rho(\mathbf{y} - \mathbf{D}\mathbf{x}_{MF}) = 0$ and $\psi(\tau_{MF}) = -\sigma$ with $\tau_{MF} = \|\mathbf{x}_{MF}\|_1$. For the left endpoint, we consider $\mathbf{x} = 0$, and $\tau = 0$, with loss equal to $\rho(\mathbf{y})$. Bracketing the root searching interval between the points $\mathbf{x}_{MF}$ and $0$ ensures finding a solution for (7) since they provide two initial points with opposite signs for $\psi$, as long as $\rho(\mathbf{y}) > \sigma$.

2) Solving $(P_\sigma)$: We combine the proposed Regula Falsi methods and a $(P_\tau)$ solver to solve $(P_\sigma)$ as follows:
- Choose initial $\tau$ values.
- Choose two initial values with opposite signs to ensure convergence of Regula Falsi type Methods. The default choice is given by $\tau = 0$ and $\tau = \tau_{MF}$.
- Apply the steps of Regula Falsi type methods.
- Every iteration of the Regula Falsi-type methods requires solving $(P_\tau)$, except at $\tau = 0$ and $\tau = \tau_{MF}$.
- Terminate once the stopping criteria are met.

IV. SIMULATIONS

In order to examine the performances of Regula Falsi type methods given in Table I, we created a test environment and benchmark by using Sparco framework [27]. Real-valued problems are chosen from Sparco for the simulations among this collection of test problems that includes many examples from the literature. Details about the problems and related publications can be found in [27].

| Problems   | id | $M$ | $N$ | $\rho(y)$ | $p_h(y)$ | $p_s(y)$ |
|------------|----|-----|-----|------------|----------|----------|
| cos-spike  | 3  | 1024| 2048| 102.2423   | 2378.8   | 25.09    |
| gauss-en   | 11 | 256 | 1024| 99.9055    | 1272.5   | 8.957    |
| jitter     | 902| 200 | 1000| 0.4476     | 4.6881   | 0.0901   |

TABLE II: $N, M, \rho(y)$ values for the problem setups.

Performances of the Regula Falsi type methods are investigated for three different loss functions $\rho$ which are Least squares $\rho_{ls}(\mathbf{x}) = \|\mathbf{x}\|_2$, Huber $\rho_h(\mathbf{x}) = \sum_i \left\{ \frac{x_i^2}{\delta^2}, \text{ if } |x_i| \leq \delta \right\}$, and Student’s $t$ $\rho_s(\mathbf{x}) = \sum_i \nu \log \left(1 + \frac{x_i^2}{\nu}\right)$ where $\delta$ and $\nu$ are the tuning parameters for $\rho_h$ and $\rho_s$, respectively; we take $\delta = 5 \times 10^{-3}$ and $\nu = 10^{-2}$ in our simulations. $N, M$ and $\rho(y)$ values for the chosen problems are given in Table II.

We solve $(P_\tau)$ for a range of $\sigma$ values, chosen relative to $\rho(y)$, in particular $5 \times 10^{-3} \rho(y)$, $5 \times 10^{-2} \rho(y)$ and $5 \times 10^{-3} \rho(y)$. We compute residuals $\rho_r(y) = \rho(y - \mathbf{D}\mathbf{x})$, norms $\|\mathbf{x}_r\|_1$, the number of nonzero (nnz) for each solution $\mathbf{x}_r$, and $\text{iter}$ the total solves of $(P_\tau)$ to reach these solutions; results are displayed in Table III.

Newton’s method requires fewer $(P_\tau)$ solves (see Table III). However, solving (7) is more expensive for Newton’s method than for Regula Falsi-type methods since the derivative calculation of the nonlinear equation is required along with the function evaluation, while Regula Falsi-type methods need only the function evaluation. From a robustness standpoint, Newton, secant, and their variants are not guaranteed to converge for nonconvex loss functions, in contrast to Regula Falsi-type methods.

The Pareto frontier $\nu(\tau)$ is shown for the gauss-en problem with losses $\rho_{ls}, \rho_h$ and $\rho_s$ in Figure 2. We expect similar patterns in many problems. As expected, the Pareto frontier is nonconvex, and Newton iterations will not stay in the feasible area for $p_s$ in this example, since the tangent lines can leave the feasible region.

Figure 3 shows a typical compressed sensing example. A 20-sparse vector is recovered using a normally distributed Parseval frame $\mathbf{D} \in \mathbb{R}^{175 \times 600}$. A measurement is generated according to $y = \mathbf{D}\mathbf{x} + \mathbf{w} + \zeta$, where the noise $\mathbf{w}$ is zero mean normal error with the variance of 0.005 and $\zeta$ has five randomly placed outliers with a zero mean normal distribution variance of 4. $(P_\tau)$ solved with the $\sigma = \rho(\zeta)$ for a fair comparison of $\rho_{ls}, \rho_h$ and $\rho_s$. Huber loss is less sensitive to outliers in measurement data than least-squares, but Student’s $t$ outperforms Huber loss since it grows sublinearly as outliers increase, a property also noted by [11].

V. CONCLUSION

In this note, we developed a new approach using bracketing Regula Falsi-type methods, that enable level-set methods to be applied to sparse optimization problems with nonconvex likelihood constraints, significantly expanding their usability. These methods achieve comparable performance to Newton’s method for root finding on convex problems, and are guaranteed to converge in the nonconvex case, where Newton and secant variants may fail.
Fig. 2: From left to right, \( \nu(\tau) \) for the gauss-en problem with \( \rho_i, \rho_h \) and \( \rho_s \).

Fig. 3: Left, top to bottom: true signal, reconstructions with least squares, Huber and Student’s \( t \) losses. Right, top to bottom: true errors, least squares, Huber and Student’s \( t \) residuals.

| Problems | Methods | least squares | Huber (\( \delta = 5 \times 10^{-3} \)) | Student’s \( t \) (\( \nu = 10^{-3} \)) |
|---------|---------|---------------|-----------------|-----------------|
|          | \( \sigma / \rho \) | \( p_i(\tau_i) \) | \( \| x_{opt} - x \|_1 \) | \( n_{tr} \) | \( \text{iter} \) | \( p_i(\tau_i) \) | \( \| x_{opt} - x \|_1 \) | \( n_{tr} \) | \( \text{iter} \) | \( p_i(\tau_i) \) | \( \| x_{opt} - x \|_1 \) | \( n_{tr} \) | \( \text{iter} \) |
| cos-spike | 0.5 | Regula Falsi | 51.12 | 66.24 | 2 | 10 | 1189.4 | 68.328 | 2 | 8 | 12.545 | 94.289 | 10 | 14 | 1.2106 | 174.09 | 79 | 72 |
|          | 0.5 | Illinois | 51.12 | 66.24 | 2 | 7 | 1189.4 | 68.328 | 2 | 7 | 12.545 | 94.289 | 10 | 7 | 1.2106 | 174.09 | 79 | 72 |
|          | 0.5 | Pegasus | 51.12 | 66.24 | 2 | 7 | 1189.4 | 68.328 | 3 | 6 | 12.545 | 94.289 | 10 | 6 | 1.2106 | 174.09 | 79 | 72 |
|          | 0.5 | And.-Björck | 51.12 | 66.24 | 2 | 9 | 1189.4 | 68.328 | 3 | 9 | 12.545 | 94.289 | 10 | 10 | 1.2106 | 174.09 | 79 | 72 |
|          | 0.5 | Newton | 51.12 | 66.24 | 2 | 3 | 1189.4 | 68.328 | 2 | 3 | 12.545 | 94.289 | 10 | 10 | 1.2106 | 174.09 | 79 | 72 |
|          | 0.5 | cos-spike | 51.12 | 188.4 | 75 | 12 | 1189.4 | 134.74 | 2 | 8 | 12.545 | 94.289 | 10 | 10 | 1.2106 | 174.09 | 79 | 72 |
|          | 0.05 | Regula Falsi | 51.12 | 188.4 | 75 | 13 | 1189.4 | 134.74 | 2 | 15 | 1.2214 | 172.65 | 78 | 16 | 1.2214 | 172.65 | 78 | 16 |
|          | 0.05 | Illinois | 51.12 | 188.4 | 75 | 12 | 1189.4 | 134.74 | 2 | 14 | 1.259 | 170.95 | 76 | 34 | 1.259 | 170.95 | 76 | 34 |
|          | 0.05 | Pegasus | 51.12 | 188.4 | 75 | 12 | 1189.4 | 134.74 | 2 | 14 | 1.259 | 170.95 | 76 | 34 | 1.259 | 170.95 | 76 | 34 |
|          | 0.05 | And.-Björck | 51.12 | 188.4 | 75 | 27 | 1189.4 | 134.74 | 2 | 32 | 1.251 | 170.85 | 75 | 50 | 1.251 | 170.85 | 75 | 50 |
|          | 0.05 | Newton | 51.12 | 188.4 | 75 | 5 | 1189.4 | 134.74 | 2 | 5 | 1189.4 | 134.74 | 2 | 5 | 1189.4 | 134.74 | 2 | 5 |
|          | 0.05 | gauss-en | 4.995 | 26.45 | 35 | 26 | 63.63 | 26.406 | 50 | 27 | 0.448 | 27.82 | 39 | 15 | 0.448 | 27.82 | 39 | 15 |
|          | 0.05 | Regula Falsi | 4.995 | 26.45 | 35 | 11 | 63.63 | 26.406 | 50 | 14 | 0.448 | 27.82 | 39 | 15 | 0.448 | 27.82 | 39 | 15 |
|          | 0.05 | Illinois | 4.995 | 26.45 | 35 | 19 | 63.63 | 26.406 | 50 | 11 | 0.448 | 27.82 | 39 | 15 | 0.448 | 27.82 | 39 | 15 |
|          | 0.05 | Pegasus | 4.995 | 26.45 | 35 | 11 | 63.63 | 26.406 | 50 | 11 | 0.448 | 27.82 | 39 | 15 | 0.448 | 27.82 | 39 | 15 |
|          | 0.05 | And.-Björck | 4.995 | 26.45 | 35 | 15 | 63.63 | 26.406 | 50 | 27 | 0.448 | 27.82 | 39 | 28 | 0.448 | 27.82 | 39 | 28 |
|          | 0.05 | Newton | 4.995 | 26.45 | 35 | 4 | 63.63 | 26.406 | 50 | 4 | 63.63 | 26.406 | 50 | 4 | 63.63 | 26.406 | 50 | 4 |
|         | jitter | 0.002 | 1.614 | 3 | 17 | 0.0234 | 1.5644 | 3 | 193 | 0.0005 | 1.2686 | 3 | 419 | 0.0005 | 1.2686 | 3 | 419 |
|         | jitter | 0.005 | Regula Falsi | 0.002 | 1.614 | 3 | 19 | 0.0234 | 1.5644 | 3 | 16 | 0.0005 | 1.2686 | 3 | 16 | 0.0005 | 1.2686 | 3 | 16 |
|         | jitter | 0.005 | Illinois | 0.002 | 1.614 | 3 | 20 | 0.0234 | 1.5644 | 3 | 14 | 0.0005 | 1.2686 | 3 | 16 | 0.0005 | 1.2686 | 3 | 16 |
|         | jitter | 0.005 | Pegasus | 0.002 | 1.614 | 3 | 35 | 0.0234 | 1.5644 | 3 | 27 | 0.0005 | 1.2686 | 3 | 107 | 0.0005 | 1.2686 | 3 | 107 |
|         | jitter | 0.005 | And.-Björck | 0.002 | 1.614 | 3 | 2 | 0.0235 | 1.5643 | 3 | 5 | 0.0005 | 1.2686 | 3 | 16 | 0.0005 | 1.2686 | 3 | 16 |
|         | jitter | 0.005 | Newton | 0.002 | 1.614 | 3 | 2 | 0.0235 | 1.5643 | 3 | 5 | 0.0005 | 1.2686 | 3 | 16 | 0.0005 | 1.2686 | 3 | 16 |
REFERENCES

[1] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” IEEE Transactions on Information Theory, vol. 52, no. 2, pp. 489–509, 2006.

[2] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theor., vol. 52, no. 4, p. 1289–1306, Apr. 2006. [Online]. Available: https://doi.org/10.1109/TIT.2006.871582

[3] D. L. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition,” IEEE Transactions on Information Theory, vol. 47, no. 7, pp. 2845–2862, 2001.

[4] J. Fuchs, “On sparse representations in arbitrary redundant bases,” IEEE Transactions on Information Theory, vol. 50, no. 6, pp. 1341–1344, 2004.

[5] E. J. Candes and T. Tao, “Decoding by linear programming,” IEEE Transactions on Information Theory, vol. 51, no. 12, pp. 4203–4215, 2005.

[6] M. Lustig, J. H. Lee, D. L. Donoho, and J. M. Pauly, “Faster imaging with randomly perturbed, undersampled spirals and —l—1 reconstruction,” in Proceedings of the 13th Annual Meeting of ISMRM, 2005, p. 685.

[7] A. Y. Aravkin, J. V. Burke, D. Drusvyatskiy, M. P. Friedlander, and S. Roy, “Level-set methods for convex optimization,” Mathematical Programming, vol. 174, no. 1, pp. 359–390, 2019.

[8] S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing, ser. Applied and Numerical Harmonic Analysis. Springer New York, 2013.

[9] E. van den Berg and M. Friedlander, “Probing the pareto frontier for basis pursuit solutions,” SIAM Journal on Scientific Computing, vol. 31, no. 2, pp. 890–912, 2009. [Online]. Available: https://doi.org/10.1137/08074488

[10] ———, “Sparse optimization with least-squares constraints,” SIAM Journal on Optimization, vol. 21, no. 4, pp. 1201–1229, 2011. [Online]. Available: https://doi.org/10.1137/100785028

[11] A. Y. Aravkin, J. V. Burke, and M. P. Friedlander, “Variational properties of value functions,” SIAM Journal on Optimization, vol. 23, no. 3, pp. 1689–1717, 2013.

[12] G. Dahlquist and Å. Björck, Numerical Methods, ser. Dover Books on Mathematics. Dover Publications, 2003. [Online]. Available: https://books.google.de/books?id=armfeHPlhwAC

[13] N. Stüdl, P. Buehlmann, and S. Van De Geer, “l1-penalization for mixture regression models,” Test, vol. 19, no. 2, pp. 209–256, 2010.

[14] P. Buehlmann and S. Geer, “Non-convex loss functions and l1-regularization,” Statistics for High-Dimensional Data, pp. 293–338, 2011.

[15] “Robust and trend-following kalman smoothers using student’s t*,” IFAC Proceedings Volumes, vol. 45, no. 16, pp. 1215 – 1220, 2012, 16th IFAC Symposium on System Identification.

[16] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, “Sparse/robust estimation and kalman smoothing with nonsmooth log-concave densities: Modeling, computation, and theory,” The Journal of Machine Learning Research, vol. 14, no. 1, pp. 2689–2728, 2013.

[17] A. Aravkin, M. P. Friedlander, F. J. Herrmann, and T. Van Leeuwen, “Robust inversion, dimensionality reduction, and randomized sampling,” Mathematical Programming, vol. 134, no. 1, pp. 101–125, 2012.

[18] P. M. Pardalos, A. Žilinskas, and J. Žilinskas, Non-convex multi-objective optimization. Springer, 2017.

[19] K. Miettinen, “Some methods for nonlinear multi-objective optimization,” in International conference on evolutionary multi-criterion optimization. Springer, 2001, pp. 1–20.

[20] E. Van den Berg, “Convex optimization for generalized sparse recovery,” Ph.D. dissertation, University of British Columbia, 2009.

[21] S. Galdino, “A family of regula falsi root-finding methods.”

[22] J. A. Ford, “Improved algorithms of illinois-type for the numerical solution of nonlinear equations.”

[23] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, “Efficient projections onto the l1-ball for learning in high dimensions,” in ICML ’08: Proceedings of the 25th international conference on Machine learning, New York, NY, USA, 2008, pp. 272–279. [Online]. Available: http://doi.acm.org/10.1145/1390156.1390191

[24] M. Held, P. Wolfe, and H. P. Crowder, “Validation of subgradient optimization,” Math. Program., vol. 6, no. 1, pp. 62–88, Dec. 1974. [Online]. Available: https://doi.org/10.1007/BF01580223

[25] L. Condat, “Fast Projection onto the Simplex and the l1 Ball,” Mathematical Programming, Series A, vol. 158, no. 1, pp. 575–585, Jul. 2016. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01056171

[26] I. Daubechies, “Time-frequency localization operators: a geometric phase space approach,” IEEE Transactions on Information Theory, vol. 34, no. 4, pp. 605–612, July 1988.

[27] E. Van den Berg, M. P. Friedlander, G. Hennenfent, F. J. Herrmann, R. Saab, and O. Yilmaz, “Algorithm 890: Sparco: A testing framework for sparse reconstruction,” ACM Trans. Math. Softw., vol. 35, no. 4, Feb. 2009. [Online]. Available: https://doi.org/10.1145/1462173.1462178