INEXACT HALF-QUADRATIC OPTIMIZATION FOR LINEAR INVERSE PROBLEMS

Marc C Robini, Feng Yang, Yuemin Zhu

INEXACT HALF-QUADRATIC OPTIMIZATION FOR LINEAR INVERSE PROBLEMS

MARC C. ROBINI, FENG YANG, AND YUEMIN ZHU

Abstract. We study the convergence of a generic half-quadratic algorithm for minimizing a wide class of objective functions that occur in inverse imaging problems; this algorithm amounts to solving a sequence of positive definite systems (the inner systems) and has the advantages of simplicity and versatility. Half-quadratic optimization has been meticulously studied, both theoretically and experimentally, but two difficulties remain: first, the practical solutions of the inner systems are generally approximate, which may hamper convergence, and, second, convergence to a stationary point of the objective is not guaranteed if the set of such points contains a continuum. We present new results that do not suffer from these limitations and hence extend our work in [SIAM J. Imaging Sci. **8** (2015), no. 3, 1752–1797]. We consider the inexact process in which the inner systems are solved to a fixed arbitrary accuracy defined in terms of the energy norm of the error. We show that this process converges to a stationary point of the objective under minimal conditions ubiquitous in regularized reconstruction and restoration. Our main results are based on the assumption that the objective has the Kurdyka-Łojasiewicz property, for which we provide constructing rules using the concept of tameness from the theory of o-minimal structures. We also propose an implementation using a truncated conjugate gradient method that controls the accuracy at negligible additional cost. Experiments on three different inverse problems show that the resulting algorithm performs well in various nonconvex scenarios and converges to solutions accurate to full machine precision.

1. Introduction

1.1. **Motivations and background.** We consider the problem of recovering an original signal $\boldsymbol{x}^\bullet \in \mathbb{R}^n$ (e.g., a temporal sequence, an image, or a volume) given some data

$$\boldsymbol{d} := \chi(\boldsymbol{D}\boldsymbol{x}^\bullet + \boldsymbol{\nu}), \tag{1.1}$$

where $\boldsymbol{D} \in \mathbb{R}^{m\times n}$ models the deterministic part of the observation process, $\boldsymbol{\nu} \in \mathbb{R}^m$ is a noise vector whose components are realizations of independent zero-mean random variables, and $\chi : \mathbb{R}^m \longrightarrow \mathbb{R}^m$ is a component-wise sample function representing contamination by impulse noise. A common approach to this general inverse

Date: September 4, 2017.

2010 *Mathematics Subject Classification.* 03C64, 26D10, 49N45, 54C60, 65F22, 65K05, 65K10, 68U10, 68W40, 90C26.

Key words and phrases. half-quadratic optimization, nonconvex optimization, tame optimization, Kurdyka-Łojasiewicz inequality, o-minimal structure, inverse problems, regularization, image reconstruction, image restoration.

problem is to look for minimizers of an objective function $\Theta : \mathbb{R}^n \longrightarrow \mathbb{R}$ of the form

$$\Theta(\boldsymbol{x}) := \sum_{k\in[1\,..\,m]} \theta_k\big(|[\boldsymbol{D}\boldsymbol{x}-\boldsymbol{d}]_k|\big) + \sum_{k\in[m+1\,..\,K]} \theta_k\big(\|\boldsymbol{R}_k\boldsymbol{x}\|\big), \tag{1.2}$$

where $\theta_1,\dots,\theta_K$ are increasing functions on $\mathbb{R}_+$, $[\,\cdot\,]_k$ is the kth coordinate projection, and $\|\cdot\|$ denotes the ℓ_2-norm. The first sum—the *data-fidelity* term—favors solutions consistent with the observations, and the second sum—the *regularization* term—imposes prior constraints modeled by the matrices $\boldsymbol{R}_k$ (see, e.g., [1, 2] and the references therein). The functions θ_k are called *potential functions*, or simply *potentials*, in reference to the Bayesian interpretation of regularization [3].

The function (1.2) can be written in the general form

$$\Theta(\boldsymbol{x}) := \sum_{k\in[1\,..\,K]} \theta_k\big(\|\boldsymbol{A}_k\boldsymbol{x}-\boldsymbol{a}_k\|\big), \tag{1.3}$$

where, for each k, $\boldsymbol{A}_k$ is a matrix in $\mathbb{R}^{n_k\times n}$ and $\boldsymbol{a}_k$ is a vector in $\mathbb{R}^{n_k}$. (We obtain (1.2) by setting $(\boldsymbol{A}_k,\boldsymbol{a}_k) := (\boldsymbol{D}(k,:),[\boldsymbol{d}]_k)$ if $k\in[1\,..\,m]$, where $\boldsymbol{D}(k,:)$ denotes the kth row of $\boldsymbol{D}$, and $(\boldsymbol{A}_k,\boldsymbol{a}_k) := (\boldsymbol{R}_k,\boldsymbol{0})$ otherwise.) We focus on the problem of minimizing objectives of the type (1.3) under the assumption that every potential satisfies the following conditions:

(C1) θ_k is increasing and nonconstant on $\mathbb{R}_+$,

(C2) θ_k is C^1 on $(0,+\infty)$ and continuous at zero,

(C3) the function $\theta_k^\dagger : t\in(0,+\infty) \longmapsto t^{-1}\theta_k'(t)$ is decreasing and bounded.

We call a potential function *admissible* if it satisfies (C1)–(C3). Common admissible potentials are listed in Appendix A; they can be nonconvex and even eventually constant. Note that the boundedness of $\theta_k^\dagger$ implies that the derivative of θ_k vanishes at zero. However, a function θ satisfying all the admissibility conditions except the boundedness of $\theta^\dagger$ can be approximated arbitrarily closely by admissible potentials (e.g., $\theta\big((\delta^2+t^2)^{1/2}\big)$ with $\delta>0$ "small enough").

A popular tool for the task of minimizing Θ is the *fully* half-quadratic (FHQ) optimization algorithm, which consists of generating a sequence $(\boldsymbol{x}^{(p)})_{p\in\mathbb{N}}$ using the recurrence relation

$$\boldsymbol{x}^{(p+1)} := \operatorname*{arg\,min}_{\boldsymbol{y}\in\mathbb{R}^n} \sum_{k\in[1\,..\,K]} \varepsilon_k(\boldsymbol{x}^{(p)})\,\|\boldsymbol{A}_k\boldsymbol{y}-\boldsymbol{a}_k\|^2, \tag{1.4}$$

where the weighting coefficients $\varepsilon_k(\boldsymbol{x}^{(p)})$ are defined by

$$\varepsilon_k(\boldsymbol{x}) := \theta_k^\dagger\big(\|\boldsymbol{A}_k\boldsymbol{x}-\boldsymbol{a}_k\|\big) \tag{1.5}$$

with the convention that $\theta_k^\dagger(0) = \lim_{t\to 0^+} t^{-1}\theta_k'(t)$. We employ the term "fully" to indicate that the inner optimization problem (1.4) is quadratic, as distinguished from the mere use of half-quadratic modeling to develop an alternating minimization algorithm [4]. As discussed in [2], the FHQ algorithm has different interpretations: fixed-point iteration, quasi-Newton optimization, alternating minimization, and majorization-minimization. It also has the advantages of simplicity, versatility, and ease of implementation. The FHQ algorithm is well-defined if the sum in (1.4) is positive definite, which is equivalent to the condition

(C4) $\displaystyle\bigcap_{k\in\mathcal{K}_+} \operatorname{null}(\boldsymbol{A}_k) = \{\boldsymbol{0}\}, \qquad \mathcal{K}_+ := \big\{k\in[1\,..\,K] : \theta_k \text{ is strictly increasing}\big\}.$

In this case, the inner optimization problem can be solved efficiently using either a direct method such as Cholesky decomposition or an iterative method such as conjugate gradient (CG), the choice of one over the other being guided by the overall sparsity of the matrices $\boldsymbol{A}_k$.

The global convergence of FHQ algorithms has been studied in [1, 5–9] under various conditions on Θ in addition to (C1)–(C4). We have generalized these results in a more recent paper [2], in which we impose only Conditions (C1)–(C4) and include the problematic cases where Θ is nonconvex and the set of stationary points $\mathcal{S}_\Theta := \{\boldsymbol{x} \in \mathbb{R}^n : \nabla\Theta(\boldsymbol{x}) = \boldsymbol{0}\}$ is nondiscrete. (Note that the results in [2] are also stronger than those obtained by applying Theorem 2 in [10] to FHQ optimization; indeed, the results derived from [10] further require the potentials to have locally Lipschitz continuous derivatives, which is not guaranteed by (C1)–(C3).) However, two limitations remain. First, it is assumed implicitly that the inner optimization problem is solved exactly at each iteration, which is not the case in practice, especially when using an iterative solver. Second, if $\mathcal{S}_\Theta$ is not level-discrete (that is, if there exists $h \in \mathbb{R}$ such that $\mathcal{S}_\Theta \cap \{\boldsymbol{x} \in \mathbb{R}^n : \Theta(\boldsymbol{x}) = h\}$ is not discrete), then the convergence is guaranteed only in terms of point-to-set distance; in other words, the FHQ sequence $(\boldsymbol{x}^{(p)})_p$ may diverge while "hugging" the boundary of $\mathcal{S}_\Theta$ in the sense that $\lim_p \inf_{\boldsymbol{y}\in\mathcal{S}_\Theta} \|\boldsymbol{y} - \boldsymbol{x}^{(p)}\| = 0$. The present work addresses both issues. In a nutshell, we derive new results for the global convergence to a single stationary point when the inner optimization problem is solved approximately and $\mathcal{S}_\Theta$ is not level-discrete, and we propose an efficient and numerically stable implementation based on CG.

1.2. **Contributions and organization.** The *inexact* FHQ algorithm, described in Section 3, is obtained by replacing the exact iteration map $\boldsymbol{x}^{(p)} \longmapsto \boldsymbol{x}^{(p+1)}$ defined in (1.4) by a set-valued map whose definition is based on the majorization-minimization interpretation of FHQ optimization. We first show in Section 4 that this set-valued map satisfies the conditions of the monotone convergence theorem of Meyer [11, Theorem 3.1] and hence shares the convergence properties derived in [2]. Our approach to establishing global convergence when $\mathcal{S}_\Theta$ is not level-discrete is presented in Section 5; it consists of excluding pathological cases by assuming that Θ satisfies the ubiquitous Kurdyka-Łojasiewicz (KL) inequality (see, e.g., [12, 13]). In this case, a sequence generated by the inexact FHQ algorithm converges to a stationary point and forms a finite-length trajectory. Section 6 is devoted to the construction of objectives satisfying the KL inequality. Our starting point is that a real C^1 function on $\mathbb{R}^n$ has the KL property if it is *tame*, meaning that its restrictions to balls are definable in an o-minimal structure (we refer to [14–16] and Appendix B for an introduction to o-minimal geometry). Of special interest is the subclass of objectives with piecewise analytic potentials, for which we can specify the convergence rate of the inexact FHQ algorithm. To our knowledge, the potentials used in the literature are piecewise analytic; yet we also provide practical rules for constructing general tame objectives.

The implementation of the inexact FHQ algorithm is discussed in Section 7. We solve the inner optimization problem using truncated CG, and we propose a CG stopping criterion that controls the accuracy of the iterates. The resulting CG-based FHQ algorithm is numerically stable in that it converges to a solution

accurate to full machine precision. Its versatility is illustrated in Section 8 with numerical experiments on three inverse imaging problems—reconstruction from limited projection data, deblurring in the presence of mixed Gaussian-impulse noise, and inpainting from moderately sparse data—under various scenarios involving quadratic or nonquadratic data-fidelity, nonconvex potentials, and different regularization operators (gradient, tight frame, and patch-based sparsifying transform). Section 9 concludes the paper.

2. Notation

We denote matrices by bold upper-case roman letters (e.g., $\boldsymbol{M}$), vectors by bold lower-case roman letters (e.g., $\boldsymbol{x}$), sets by caligraphic upper-case letters (e.g., $\mathcal{A}$), and set-valued maps by bold greek letters (e.g., $\boldsymbol{\Phi}$). The kth coordinate (or component) of a vector $\boldsymbol{v}$ is denoted by v_k or $[\boldsymbol{v}]_k$ depending on the context, and $\|\cdot\|$ is the ℓ_2-norm. We let $\|\cdot\|_{\boldsymbol{M}}$ denote the ℓ_2-norm weighted by the matrix $\boldsymbol{M}$, that is, $\|\boldsymbol{z}\|_{\boldsymbol{M}} := (\boldsymbol{z}^T\boldsymbol{M}\boldsymbol{z})^{1/2}$ (we will omit the parentheses for the square of such a norm, so $\|\boldsymbol{z}\|_{\boldsymbol{M}}^2 := (\|\boldsymbol{z}\|_{\boldsymbol{M}})^2 = \boldsymbol{z}^T\boldsymbol{M}\boldsymbol{z}$). The distance from a point $\boldsymbol{x} \in \mathbb{R}^n$ to a nonempty set $\mathcal{A} \subseteq \mathbb{R}^n$ is denoted by $\operatorname{dist}(\boldsymbol{x}, \mathcal{A})$, that is,

$$\operatorname{dist}(\boldsymbol{x}, \mathcal{A}) := \inf_{\boldsymbol{y} \in \mathcal{A}} \|\boldsymbol{y} - \boldsymbol{x}\|. \tag{2.1}$$

The interior, the closure, and the boundary of $\mathcal{A}$ are denoted by $\mathcal{A}^\circ$, $\overline{\mathcal{A}}$, and $\partial\mathcal{A}$, respectively, and $\mathscr{B}(\boldsymbol{x}, r)$ denotes the open ball with center $\boldsymbol{x}$ and radius r.

Let f be a real function on a $\mathcal{A} \subseteq \mathbb{R}^n$. The restriction of f to some set $\mathcal{A}' \subset \mathcal{A}$ is denoted by $f|_{\mathcal{A}'}$. Given a binary relation $\mathcal{R}$ on $\mathbb{R}$ and a real number h, we use the shortcut notation

$$\{f \mathcal{R} h\} := \big\{\boldsymbol{x} \in \mathcal{A} : f(\boldsymbol{x}) \mathcal{R} h\big\}. \tag{2.2}$$

For example, $\{f = h\}$ and $\{f \leqslant h\}$ are the level and sublevel sets of f at height h, respectively. When convenient, we denote the preimage of a set $\mathcal{B} \subseteq \mathbb{R}$ under f by $\{f \in \mathcal{B}\}$ rather than $f^{-1}(\mathcal{B})$. The set of stationary points of a function f in $C^1(\mathbb{R}^n)$ (the class of real C^1 functions on $\mathbb{R}^n$) is denoted by $\mathcal{S}_f$, that is,

$$\mathcal{S}_f := \{\|\nabla f\| = 0\}. \tag{2.3}$$

Finally, we let $\mathscr{C}_{\text{obj}}(\mathbb{R}^n)$ denote the class of objective functions $\Theta : \mathbb{R}^n \longrightarrow \mathbb{R}$ of the form (1.3) such that Conditions (C1)–(C4) hold, and we define $\mathscr{C}_{\text{obj}} := \bigcup_{n \in [1..\infty)} \mathscr{C}_{\text{obj}}(\mathbb{R}^n)$. Other notation will be introduced as needed.

3. The inexact fully half-quadratic algorithm

We assume throughout this section that $\Theta \in \mathscr{C}_{\text{obj}}(\mathbb{R}^n)$. Let $\boldsymbol{A}$ and $\boldsymbol{a}$ be the vertical concatenations of the matrices $\boldsymbol{A}_k \in \mathbb{R}^{n_k \times n}$ and of the vectors $\boldsymbol{a}_k \in \mathbb{R}^{n_k}$, that is,

$$\boldsymbol{A} := \begin{pmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_K \end{pmatrix} \in \mathbb{R}^{N \times n} \quad \text{and} \quad \boldsymbol{a} := \begin{pmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_K \end{pmatrix} \in \mathbb{R}^N, \tag{3.1}$$

where $N := \sum_{k \in [1..K]} n_k$, and let

$$\boldsymbol{E}(\boldsymbol{x}) := \operatorname{diag}\big(\varepsilon_1(\boldsymbol{x})\boldsymbol{I}_{n_1}, \ldots, \varepsilon_K(\boldsymbol{x})\boldsymbol{I}_{n_K}\big), \tag{3.2}$$

where the weighting coefficients $\varepsilon_k(\boldsymbol{x})$ are defined in (1.5) and $\boldsymbol{I}_n$ denotes the identity matrix of order n (so $\boldsymbol{E}(\boldsymbol{x})$ is an $N \times N$ nonnegative diagonal matrix). With this notation, the gradient of Θ is given by

$$\nabla\Theta(\boldsymbol{x}) \;=\; \boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{a}). \tag{3.3}$$

Let $\overline{\Theta} : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ be defined by

$$\overline{\Theta}(\boldsymbol{x},\boldsymbol{y}) \;:=\; \Theta(\boldsymbol{x}) + (\boldsymbol{y}-\boldsymbol{x})^T\nabla\Theta(\boldsymbol{x}) + \tfrac{1}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}}. \tag{3.4}$$

By [2, Proposition 2.4], the matrix $\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}$ is positive definite for all $\boldsymbol{x} \in \mathbb{R}^n$, and hence so is the quadratic function $\overline{\Theta}(\boldsymbol{x},\cdot) : \boldsymbol{y} \in \mathbb{R}^n \longmapsto \overline{\Theta}(\boldsymbol{x},\boldsymbol{y})$. It is easy to see that the exact FHQ iterations (1.4) are equivalent to

$$\boldsymbol{x}^{(p+1)} \;:=\; \boldsymbol{\Phi}_0(\boldsymbol{x}^{(p)}), \tag{3.5a}$$

$$\boldsymbol{\Phi}_0(\boldsymbol{x}) \;:=\; \arg\min_{\boldsymbol{y}\in\mathbb{R}^n} \overline{\Theta}(\boldsymbol{x},\boldsymbol{y}). \tag{3.5b}$$

Furthermore, we readily have $\overline{\Theta}(\boldsymbol{x},\boldsymbol{x}) = \Theta(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$, and it is shown in [2, Proposition 2.6] that

$$\overline{\Theta}(\boldsymbol{x},\boldsymbol{y}) \;\geqslant\; \Theta(\boldsymbol{y}) \quad \text{for all } \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n. \tag{3.6}$$

Therefore the exact FHQ algorithm is a majorization-minimization algorithm [17, 18] with surrogate function $\overline{\Theta}$.

From now on, we focus on a relaxed version of (3.5) in which $\boldsymbol{x}^{(p+1)}$ is any point $\boldsymbol{y}$ such that the difference $\overline{\Theta}(\boldsymbol{x}^{(p)},\boldsymbol{y}) - \min\overline{\Theta}(\boldsymbol{x}^{(p)},\cdot)$ is smaller than some fraction of $\Theta(\boldsymbol{x}^{(p)}) - \min\overline{\Theta}(\boldsymbol{x}^{(p)},\cdot)$. More precisely, the exact FHQ map $\boldsymbol{\Phi}_0$ is replaced by the set-valued map $\boldsymbol{\Phi}_\mu : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined by

$$\boldsymbol{\Phi}_\mu(\boldsymbol{x}) \;:=\; \big\{\boldsymbol{y} \in \mathbb{R}^n : \overline{\Theta}(\boldsymbol{x},\boldsymbol{y}) - \min\overline{\Theta}(\boldsymbol{x},\cdot) \leqslant h_\mu(\boldsymbol{x})\big\}, \tag{3.7a}$$

$$h_\mu(\boldsymbol{x}) \;:=\; \mu\big(\Theta(\boldsymbol{x}) - \min\overline{\Theta}(\boldsymbol{x},\cdot)\big), \tag{3.7b}$$

where the constant $\mu \in (0,1)$ sets the accuracy of the approximation to $\boldsymbol{\Phi}_0(\boldsymbol{x})$.

Definition 3.1 (Inexact FHQ process)**.** We call a sequence $(\boldsymbol{x}^{(p)})_{p\in\mathbb{N}}$ an *inexact FHQ sequence* if there exists a constant $\mu \in (0,1)$ such that

$$\boldsymbol{x}^{(p+1)} \;\in\; \boldsymbol{\Phi}_\mu(\boldsymbol{x}^{(p)}) \quad \text{for all } p \in \mathbb{N}. \tag{3.8}$$

The iterative process (3.8) starting from a given point $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$ is called an *inexact FHQ algorithm*.

The inexact FHQ optimization process is illustrated in Figure 1. As we will see in Section 7, inexact FHQ sequences occur in particular when the inner optimization problem (3.5) is solved approximately by CG with a fixed minimum number of iterations.

4. Monotonic nonlinear programming interpretation

The results presented in this section extend those in [2] to inexact FHQ sequences. They rely on a general convergence theorem of Meyer characterizing the behavior of sequences generated by set-valued maps under some compactness, continuity, and monotonicity assumptions [11, Theorem 3.1]. This approach is similar to that in [2], except we deal with a set-valued rather than a point-valued iteration map; so we do not reproduce the parts of the proofs unaffected by this change.

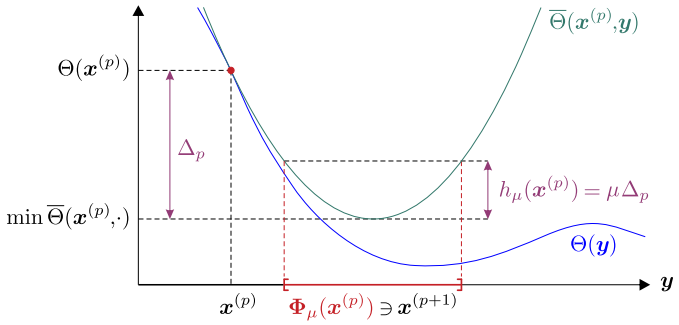


FIGURE 1. Inexact FHQ optimization process: given the current iterate $\boldsymbol{x}^{(p)}$, the next iterate $\boldsymbol{x}^{(p+1)}$ is any point in the sublevel set of the surrogate $\overline{\Theta}(\boldsymbol{x}^{(p)},\cdot)$ at an height exceeding its minimum value by a fraction μ of the difference $\Theta(\boldsymbol{x}^{(p)}) - \min \overline{\Theta}(\boldsymbol{x}^{(p)},\cdot)$.

We start with some terminology on set-valued maps in Definition 4.1 and recall Meyer's results in Theorem 4.2. Lemma 4.5 then states the properties of the inexact FHQ map that are needed to apply Meyer's theorem, from which we derive our convergence results. In brief, the inexact FHQ algorithm converges to a stationary point if $\mathcal{S}_\Theta$ is discrete in each level set of Θ (Theorem 4.9(i)) and otherwise converges to $\mathcal{S}_\Theta$ in three weaker senses: in terms of point-to-set distance (Theorem 4.9(ii)), gradient (Theorem 4.9(iii)), and objective (Theorem 4.7(iv)). Theorem 4.10 characterizes the basins of attraction of isolated local minimizers regardless of the stucture of $\mathcal{S}_\Theta$.

Definition 4.1 (Terminology of set-valued maps)**.** Let $\mathcal{H}$ be a closed set in $\mathbb{R}^n$ and let $\boldsymbol{\Psi}$ be a set-valued map from $\mathcal{H}$ to itself.

(i) A sequence $(\boldsymbol{x}^{(p)})_{p\in\mathbb{N}}$ in $\mathbb{R}^n$ is said to be *generated by* $\boldsymbol{\Psi}$ if $\boldsymbol{x}^{(0)} \in \mathcal{H}$ and $\boldsymbol{x}^{(p+1)} \in \boldsymbol{\Psi}(\boldsymbol{x}^{(p)})$ for all $p \in \mathbb{N}$.

(ii) A point $\boldsymbol{x} \in \mathcal{H}$ is called a *fixed point* of $\boldsymbol{\Psi}$ if $\boldsymbol{\Psi}(\boldsymbol{x}) = \{\boldsymbol{x}\}$. The set of fixed points of $\boldsymbol{\Psi}$ is denoted by $\mathcal{F}_{\boldsymbol{\Psi}}$.

(iii) $\boldsymbol{\Psi}$ is called *upper semicontinuous* if for all convergent sequences $(\boldsymbol{x}^{(p)})_p$ and $(\boldsymbol{y}^{(p)})_p$ in $\mathcal{H}$ such that $\boldsymbol{y}^{(p)} \in \boldsymbol{\Psi}(\boldsymbol{x}^{(p)})$ for all p, we have $\lim_p \boldsymbol{y}^{(p)} \in \boldsymbol{\Psi}(\lim_p \boldsymbol{x}^{(p)})$.

(iv) $\boldsymbol{\Psi}$ is said to be *strictly monotonic* with respect to a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ if for every $\boldsymbol{x} \in \mathcal{H} \setminus \mathcal{F}_{\boldsymbol{\Psi}}$, we have $f(\boldsymbol{y}) < f(\boldsymbol{x})$ for all $\boldsymbol{y} \in \boldsymbol{\Psi}(\boldsymbol{x})$.

Theorem 4.2 (Meyer [11])**.** *Let $\mathcal{H}$ be a closed set in $\mathbb{R}^n$ and let $\boldsymbol{\Psi}$ be a set-valued map from $\mathcal{H}$ to itself. Let $\mathcal{X} := (\boldsymbol{x}^{(p)})_p$ be a sequence generated by $\boldsymbol{\Psi}$ and denote by $\mathcal{C}_{\mathcal{X}}$ its set of cluster points. If $\boldsymbol{\Psi}$ is upper semicontinuous and strictly monotonic with respect to a function f, and if $\mathcal{X}$ is bounded, then*

(i) $\emptyset \neq \mathcal{C}_{\mathcal{X}} \subseteq \mathcal{F}_{\boldsymbol{\Psi}}$,

(ii) $\lim_p \|\boldsymbol{x}^{(p+1)} - \boldsymbol{x}^{(p)}\| = 0$,

(iii) $\lim_p f(\boldsymbol{x}^{(p)}) = f(\boldsymbol{x}^*)$ *for some* $\boldsymbol{x}^* \in \mathcal{C}_{\mathcal{X}}$.

Definition 4.3 (Flat set)**.** Let f be a real function on $\mathbb{R}^n$. A nonempty set $\mathcal{A} \subseteq \mathbb{R}^n$ is said to be *flat* with respect to f (or simply flat if f is clear from the context) if f is constant on $\mathcal{A}$.

Definition 4.4 (Generalized cup)**.** Let f be a real continuous function on $\mathbb{R}^n$. We call a set $\mathcal{G} \subset \mathbb{R}^n$ a *generalized cup* of f if there exists a bounded set $\mathcal{B}$ such that

$$\emptyset \neq \mathcal{G} = \mathcal{B} \cap \big\{ f < \min_{\partial \mathcal{B}} f \big\}. \tag{4.1}$$

In other words, $\mathcal{G}$ is a bounded open set with flat boundary.

Lemma 4.5 (Properties of the inexact FHQ map)**.** *Let $\Theta \in \mathscr{C}_{\mathrm{obj}}(\mathbb{R}^n)$ and let $\mathcal{G}$ be a generalized cup of Θ. For every $\mu \in (0,1)$, the set-valued map $\boldsymbol{\Phi}_\mu$ defined in* (3.7) *has the following properties:*

(i) $\boldsymbol{\Phi}_\mu(\overline{\mathcal{G}}) := \bigcup_{\boldsymbol{x} \in \overline{\mathcal{G}}} \boldsymbol{\Phi}_\mu(\boldsymbol{x}) \subseteq \overline{\mathcal{G}}$ *(that is, $\boldsymbol{\Phi}_\mu$ maps $\overline{\mathcal{G}}$ into itself),*

(ii) $\boldsymbol{\Phi}_\mu(\overline{\mathcal{G}})$ *is bounded,*

(iii) $\mathcal{F}_{\boldsymbol{\Phi}_\mu} = \mathcal{S}_\Theta$ *(that is, the fixed points of $\boldsymbol{\Phi}_\mu$ are the stationary points of Θ),*

(iv) $\boldsymbol{\Phi}_\mu$ *is upper semicontinuous,*

(v) $\boldsymbol{\Phi}_\mu$ *is strictly monotonic with respect to Θ.*

Proof. (i) Let $\boldsymbol{x} \in \mathcal{G}$. We have

$$\boldsymbol{\Phi}_\mu(\boldsymbol{x}) \subset \big\{ \overline{\Theta}(\boldsymbol{x}, \cdot) \leqslant \Theta(\boldsymbol{x}) \big\} \subseteq \mathcal{G},$$

where the second inclusion follows from [19, Proposition 6.3] (indeed, by the proof of Lemma 4.12 in [2], a generalized cup is the set of points that are well-contained in a generalized basin). So $\boldsymbol{\Phi}_\mu(\mathcal{G}) \subseteq \mathcal{G}$.

Let $\boldsymbol{x} \in \partial\mathcal{G}$ and set

$$H_\mu(\boldsymbol{x}) := h_\mu(\boldsymbol{x}) + \min \overline{\Theta}(\boldsymbol{x}, \cdot).$$

Since $\overline{\Theta}(\boldsymbol{x}, \cdot)$ is a positive definite quadratic function, $\boldsymbol{\Phi}_\mu(\boldsymbol{x})$ is a closed ellipsoid in $\mathbb{R}^n$, and therefore

$$(\boldsymbol{\Phi}_\mu(\boldsymbol{x}))^\circ = \big\{ \overline{\Theta}(\boldsymbol{x}, \cdot) < H_\mu(\boldsymbol{x}) \big\} \quad \text{and} \quad \overline{(\boldsymbol{\Phi}_\mu(\boldsymbol{x}))^\circ} = \boldsymbol{\Phi}_\mu(\boldsymbol{x}).$$

Let $\boldsymbol{y} \in (\boldsymbol{\Phi}_\mu(\boldsymbol{x}))^\circ$ and let $(\boldsymbol{x}^{(p)})_p$ be a sequence in $\mathcal{G}$ converging to $\boldsymbol{x}$. Then, for sufficiently large p, $\boldsymbol{y} \in (\boldsymbol{\Phi}_\mu(\boldsymbol{x}^{(p)}))^\circ \subset \mathcal{G}$. Hence $(\boldsymbol{\Phi}_\mu(\boldsymbol{x}))^\circ \subseteq \mathcal{G}$, and taking the closure on both sides of this inclusion yields $\boldsymbol{\Phi}_\mu(\boldsymbol{x}) \subseteq \overline{\mathcal{G}}$. So $\boldsymbol{\Phi}_\mu(\partial\mathcal{G}) \subseteq \overline{\mathcal{G}}$.

(ii) Let $\boldsymbol{x} \in \overline{\mathcal{G}}$. Since the function $\overline{\Theta}(\boldsymbol{x}, \cdot)$ is coercive, its sublevel sets are bounded, and so $\boldsymbol{\Phi}_\mu(\boldsymbol{x}) \subseteq \overline{\mathscr{B}(\boldsymbol{0}, r_{\boldsymbol{x}})}$ for some $r_{\boldsymbol{x}} > 0$. Since $\overline{\mathcal{G}}$ is compact, it follows that $\boldsymbol{\Phi}_\mu(\overline{\mathcal{G}}) \subseteq \overline{\mathscr{B}(\boldsymbol{0}, R)}$, with $R := \sup\{ r_{\boldsymbol{x}} : \boldsymbol{x} \in \overline{\mathcal{G}} \} < +\infty$.

(iii) Let $\boldsymbol{x} \in \mathbb{R}^n$. We have

$$\boldsymbol{x} \in \mathcal{F}_{\boldsymbol{\Phi}_\mu} \iff \big\{ \overline{\Theta}(\boldsymbol{x}, \cdot) \leqslant H_\mu(\boldsymbol{x}) \big\} = \{\boldsymbol{x}\} \iff H_\mu(\boldsymbol{x}) = \min \overline{\Theta}(\boldsymbol{x}, \cdot).$$

So we must show that $h_\mu(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{x} \in \mathcal{S}_\Theta$. The gradient of the function $\overline{\Theta}(\boldsymbol{x}, \cdot)$ is given by

$$\nabla\overline{\Theta}(\boldsymbol{x}, \cdot)(\boldsymbol{y}) = \boldsymbol{A}^T \boldsymbol{E}(\boldsymbol{x}) \boldsymbol{A}(\boldsymbol{y} - \boldsymbol{x}) + \nabla\Theta(\boldsymbol{x}).$$

Therefore the global minimizer of $\overline{\Theta}(\boldsymbol{x}, \cdot)$ is

$$\boldsymbol{\Phi}_0(\boldsymbol{x}) = \boldsymbol{x} - (\boldsymbol{A}^T \boldsymbol{E}(\boldsymbol{x}) \boldsymbol{A})^{-1} \nabla\Theta(\boldsymbol{x}), \tag{4.2}$$

and hence

$$\min \overline{\Theta}(\boldsymbol{x}, \cdot) = \overline{\Theta}(\boldsymbol{x}, \boldsymbol{\Phi}_0(\boldsymbol{x})) = \Theta(\boldsymbol{x}) - \tfrac{1}{2} \|\nabla\Theta(\boldsymbol{x})\|^2_{(\boldsymbol{A}^T \boldsymbol{E}(\boldsymbol{x}) \boldsymbol{A})^{-1}}. \tag{4.3}$$

Substituting into the definition of h_μ, we obtain

$$h_\mu(\boldsymbol{x}) = \tfrac{1}{2} \mu \|\nabla\Theta(\boldsymbol{x})\|^2_{(\boldsymbol{A}^T \boldsymbol{E}(\boldsymbol{x}) \boldsymbol{A})^{-1}}, \tag{4.4}$$

which is zero if and only if $\nabla\Theta(\boldsymbol{x}) = \mathbf{0}$.

(iv) Let $(\boldsymbol{x}^{(p)})_p$ and $(\boldsymbol{y}^{(p)})_p$ be convergent sequences in $\mathbb{R}^n$ such that $\boldsymbol{y}^{(p)} \in \boldsymbol{\Phi}_\mu(\boldsymbol{x}^{(p)})$ for all p, and let $\boldsymbol{x}$ and $\boldsymbol{y}$ be their respective limits. Since $\overline{\Theta}$ is continuous on the product space $\mathbb{R}^n \times \mathbb{R}^n$, we have

$$\overline{\Theta}(\boldsymbol{x}, \boldsymbol{y}) \;=\; \lim_p \overline{\Theta}(\boldsymbol{x}^{(p)}, \boldsymbol{y}^{(p)}) \;\leqslant\; \lim_p \min \overline{\Theta}(\boldsymbol{x}^{(p)}, \cdot) + \lim_p h_\mu(\boldsymbol{x}^{(p)}).$$

From (4.3) and (4.4), the functions $\boldsymbol{z} \in \mathbb{R}^n \longmapsto \min \overline{\Theta}(\boldsymbol{z}, \cdot)$ and h_μ are continuous (because the matrix inverse function is continuous at every point that represents an invertible matrix). So $\overline{\Theta}(\boldsymbol{x}, \boldsymbol{y}) \leqslant \min \overline{\Theta}(\boldsymbol{x}, \cdot) + h_\mu(\boldsymbol{x})$, that is, $\boldsymbol{y} \in \boldsymbol{\Phi}_\mu(\boldsymbol{x})$.

(v) For every $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \boldsymbol{\Phi}_\mu(\boldsymbol{x})$,

$$\begin{aligned} \Theta(\boldsymbol{y}) \;\leqslant\; \overline{\Theta}(\boldsymbol{x}, \boldsymbol{y}) \;&\leqslant\; \min \overline{\Theta}(\boldsymbol{x}, \cdot) + h_\mu(\boldsymbol{x}) \\ &=\; \Theta(\boldsymbol{x}) - \tfrac{1}{2}(1-\mu)\|\nabla\Theta(\boldsymbol{x})\|^2_{(\boldsymbol{A}^T \boldsymbol{E}(\boldsymbol{x}) \boldsymbol{A})^{-1}}. \end{aligned}$$

If $\boldsymbol{x} \notin \mathcal{F}_{\boldsymbol{\Phi}_\mu}$, then $\nabla\Theta(\boldsymbol{x}) \neq \mathbf{0}$ by (iii), and so $\Theta(\boldsymbol{y}) < \Theta(\boldsymbol{x})$. □

Definition 4.6 (Continuum)**.** A nonempty compact and connected set in $\mathbb{R}^n$ is called a *continuum*. A continuum is called *nondegenerate* if it contains more than one point.

Theorem 4.7 (Monotone convergence)**.** *Let $\Theta \in \mathscr{C}_{\mathrm{obj}}(\mathbb{R}^n)$ and let $\mathcal{X} := (\boldsymbol{x}^{(p)})_p$ be an inexact FHQ sequence starting from a point $\boldsymbol{x}^{(0)}$ in a generalized cup of Θ.*

(i) *$\mathcal{C}_\mathcal{X} \subseteq \mathcal{S}_\Theta$ (that is, the cluster points of $\mathcal{X}$ are stationary points of Θ).*

(ii) *$\mathcal{C}_\mathcal{X}$ is either a singleton or a flat nondegenerate continuum.*

(iii) $\lim_p \|\boldsymbol{x}^{(p+1)} - \boldsymbol{x}^{(p)}\| = 0$.

(iv) *The objective sequence $(\Theta(\boldsymbol{x}^{(p)}))_p$ decreases to the value of Θ on $\mathcal{C}_\mathcal{X}$ and is strictly decreasing as long as $\boldsymbol{x}^{(p)} \notin \mathcal{S}_\Theta$.*

Proof. Let $\mu \in (0,1)$ be such that $\mathcal{X}$ is generated by $\boldsymbol{\Phi}_\mu$, and let $\mathcal{G}$ be a generalized cup of Θ containing $\boldsymbol{x}^{(0)}$. By Lemma 4.5, Theorem 4.2 applies to the restrictions of $\boldsymbol{\Phi}_\mu$ and Θ to $\overline{\mathcal{G}}$. Therefore, $\mathcal{C}_\mathcal{X}$ is nonempty and included in $\mathcal{S}_\Theta \cap \overline{\mathcal{G}}$, the difference $\boldsymbol{x}^{(p+1)} - \boldsymbol{x}^{(p)}$ goes to zero, and $(\Theta(\boldsymbol{x}^{(p)}))_p$ converges to $\Theta(\boldsymbol{x}^*)$ for some $\boldsymbol{x}^* \in \mathcal{C}_\mathcal{X}$. The proof that $\mathcal{C}_\mathcal{X}$ is a flat continuum is the same as in [2, Theorem 3.7], and the fact that $\Theta(\boldsymbol{x}^{(p+1)}) < \Theta(\boldsymbol{x}^{(p)})$ for all $\boldsymbol{x}^{(p)} \notin \mathcal{S}_\Theta$ follows from the strict monotonocity of $\boldsymbol{\Phi}_\mu$. □

Definition 4.8 (Level-discrete stationary points)**.** Let $f \in C^1(\mathbb{R}^n)$. We call the set of stationary points $\mathcal{S}_f$ *level-discrete* if for every $h \in \mathbb{R}$, the intersection of $\mathcal{S}_f$ with the level set $\{f = h\}$ is either discrete or empty.

Theorem 4.9 (Global convergence to stationary points)**.** *Let Θ and $\mathcal{X}$ be as in Theorem 4.7.*

(i) *If $\mathcal{S}_\Theta$ is level-discrete, then $\mathcal{X}$ converges to a point in $\mathcal{S}_\Theta$.*

(ii) $\lim_p \operatorname{dist}(\boldsymbol{x}^{(p)}, \partial\mathcal{S}_\Theta) = 0$ *(that is, $\mathcal{X}$ converges to the boundary of $\mathcal{S}_\Theta$).*

(iii) $\lim_p \|\nabla\Theta(\boldsymbol{x}^{(p)})\| = 0$.

Proof. The proofs of the three items of Theorem 4.9 rely on Theorem 4.7. They are similar to those of Theorems 3.9 and 3.12 in [2], but with the sublevel set $\{\Theta \leqslant \Theta(\boldsymbol{x}^{(0)})\}$ replaced by $\overline{\mathcal{G}}$. □

Remark 1. In Theorems 4.7 and 4.9, the condition that the starting point $\boldsymbol{x}^{(0)}$ be in a generalized cup can be replaced by any condition ensuring the boundedness of $\mathcal{X}$. Indeed, if $\mathcal{X}$ is bounded, we can apply Theorem 4.2 to $\boldsymbol{\Phi}_\mu$ and Θ without restricting their domain, and so we just have to replace the generalized cup $\mathcal{G}$ in the proofs by any bounded open set containing $\mathcal{X}$. It follows that $\mathcal{X}$ either goes to infinity or converges to $\partial\mathcal{S}_\Theta$.

Remark 2. In [2], the condition on $\boldsymbol{x}^{(0)}$ is that the sublevel set $\{\Theta \leqslant \Theta(\boldsymbol{x}^{(0)})\}$ be bounded. In this case, there is a bounded set $\mathcal{B}$ such that $\{\Theta \leqslant \Theta(\boldsymbol{x}^{(0)})\} \subset \mathcal{B}^\circ$, so $\partial\mathcal{B}$ is a compact subset of $\{\Theta > \Theta(\boldsymbol{x}^{(0)})\}$ and thus $\Theta(\boldsymbol{x}^{(0)}) < \min_{\partial\mathcal{B}} \Theta$. Therefore the boundedness of $\{\Theta \leqslant \Theta(\boldsymbol{x}^{(0)})\}$ implies that $\boldsymbol{x}^{(0)}$ is in a generalized cup. The converse implication, however, is not true, as illustrated by the objective function in Example 1 after Theorem 4.10.

Remark 3. If Θ is coercive, then all its sublevel sets are bounded and so the conclusions of Theorems 4.7 and 4.9 hold independently of the starting point. This makes coercivity an attractive property that should be imposed whenever possible. Let $\mathcal{K}_\infty$ be the set of indices $k \in [1\,..\,K]$ such that $\lim_{t\to+\infty} \theta_k(t) = +\infty$. As shown in [2, Proposition 2.5], if the potentials $\theta_1, \ldots, \theta_K$ are admissible, then Θ is coercive if and only if

$$\mathcal{N}_\infty := \bigcap_{k\in\mathcal{K}_\infty} \operatorname{null}(\boldsymbol{A}_k) = \{\boldsymbol{0}\} \tag{4.5}$$

(in particular, Condition (C4) holds if Θ is coercive). Therefore, we can always enforce the coercivity of the objective by adding a term of the form

$$\Theta_0(\boldsymbol{x}) := \sum_{k\in[1\,..\,n_0]} \theta_0(|[\boldsymbol{A}_0\boldsymbol{x}]_k|), \tag{4.6}$$

where θ_0 is admissible and such that $\lim_{t\to+\infty} \theta_0(t) = +\infty$, and where $\boldsymbol{A}_0 \in \mathbb{R}^{n_0\times n}$ satisfies $\operatorname{null}(\boldsymbol{A}_0) \cap \mathcal{N}_\infty = \{\boldsymbol{0}\}$. This technique is used for example in [20], [1], and [2], where Θ_0 is respectively a Tikhonov penalty, a multiresolution contour-line smoothing penalty, and a wavelet-sparsity penalty.

Theorem 4.9 does not guarantee that all isolated local minimizers are attractors; that is, a local minimizer isolated from the other stationary points may not be a limit point no matter how closely it is approached. The following theorem shows that, regardless of the configuration of the stationary points, every isolated local minimizer is an attractor and every generalized cup containing a single stationary point is a basin of attraction. Furthermore, such generalized cups are "cups" in the sense defined in [2, Definition 4.4], so the water-flooding interpretation of the basins of attraction developed therein holds for the inexact FHQ algorithm.

Theorem 4.10 (Local convergence to isolated minimizers)**.** *Let $\Theta \in \mathscr{C}_{\text{obj}}(\mathbb{R}^n)$, let $\mathcal{G}$ be a generalized cup of Θ such that $\mathcal{G} \cap \mathcal{S}_\Theta = \{\boldsymbol{x}^*\}$, and let $\mathcal{X} := (\boldsymbol{x}^{(p)})_p$ be an inexact FHQ sequence. The point $\boldsymbol{x}^*$ is a local minimizer of Θ, and if $\boldsymbol{x}^{(p)} \in \mathcal{G}$ for some p, then $\mathcal{X}$ converges to $\boldsymbol{x}^*$.*

Proof. It follows from the definition of a generalized cup that $\mathcal{G}$ is open and $\emptyset \neq \arg\min_{\overline{\mathcal{G}}} \Theta \subset \mathcal{G}$. Therefore the global minimizers of Θ on $\overline{\mathcal{G}}$ are stationary points in $\mathcal{G}$. But $\boldsymbol{x}^*$ is the only stationary point in $\mathcal{G}$, so $\boldsymbol{x}^*$ is a local minimizer.

Let p be such that $\boldsymbol{x}^{(p)} \in \mathcal{G}$. By Lemma 4.5(i), the shifted sequence $\mathcal{X}_{+p} := (\boldsymbol{x}^{(q+p)})_{q\in\mathbb{N}}$ is in the compact set $\overline{\mathcal{G}}$, and thus $\mathcal{X}$ converges to $\boldsymbol{x}^*$ if $\mathcal{C}_\mathcal{X} = \{\boldsymbol{x}^*\}$.

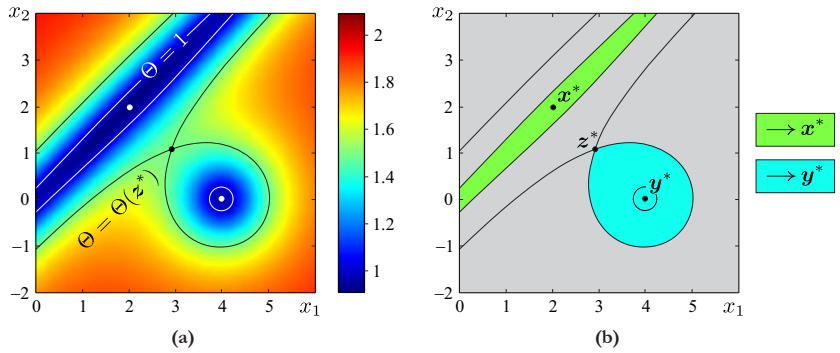


FIGURE 2. Two-dimensional noncoercive objective: (a) function Θ defined in (4.7) and level sets at heights 1 and $\Theta(\boldsymbol{z}^*)$; (b) basins of attraction of the minimizers $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ (there is no theoretical convergence guarantee when starting outside these basins).

Let $\mathcal{B}$ be a bounded set such that $\mathcal{G} = \mathcal{B} \cap \{\Theta < h\}$, where $h := \min_{\partial\mathcal{B}} \Theta$. By Theorem 4.7 applied to $\mathcal{X}_{+p}$, we have

$$\emptyset \neq \mathcal{C}_{\mathcal{X}} \subseteq \mathcal{S}_\Theta \cap \overline{\mathcal{G}} \subseteq \{\boldsymbol{x}^*\} \cup \partial\mathcal{G},$$

and $(\Theta(\boldsymbol{x}^{(q+p)}))_q$ is a decreasing sequence starting from $\Theta(\boldsymbol{x}^{(p)}) < h$. Hence

$$\mathcal{C}_{\mathcal{X}} \subseteq (\{\boldsymbol{x}^*\} \cup \partial\mathcal{G}) \cap \{\Theta < h\},$$

and since

$$\partial\mathcal{G} \cap \{\Theta < h\} \subseteq (\partial\mathcal{B} \cup \partial\{\Theta < h\}) \cap \{\Theta < h\} = \partial\mathcal{B} \cap \{\Theta < h\} = \emptyset,$$

it follows that $\mathcal{C}_{\mathcal{X}} = \{\boldsymbol{x}^*\}$. □

The following two examples illustrate the scope of Theorems 4.9(i) and 4.10.

Example 1. Let $\Theta \in \mathscr{C}_{\mathrm{obj}}(\mathbb{R}^2)$ be defined by

$$\Theta(\boldsymbol{x}) := \vartheta_{\mathrm{GM}}\big(\|(x_1, x_2) - (4, 0)\|\big) + \vartheta_{\mathrm{GM}}\big(|x_2 - x_1|\big), \tag{4.7}$$

where ϑ_{GM} is the Geman and McClure function defined in (A.10). This function has three isolated stationary points: two minimizers $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ near $(2, 2)$ and $(4, 0)$, and one saddle point $\boldsymbol{z}^*$ near $(3, 1)$. Figure 2(a) shows the level sets $\{\Theta = 1\}$ and $\{\Theta = \Theta(\boldsymbol{z}^*)\}$ superimposed on Θ; both are unbounded, as $\Theta(x, x) < 1$ for all $x \in \mathbb{R}$ and $\lim_{x\to\infty} \Theta(x, x) = 1$. Let $\mathcal{G}_h(\boldsymbol{x})$ denote the connected component of a point $\boldsymbol{x}$ in the sublevel set $\{\Theta < h\}$. For every $h \in (\Theta(\boldsymbol{x}^*), 1)$, the set $\mathcal{G}_h(\boldsymbol{x}^*)$ is a generalized cup, and $\mathcal{G}_h(\boldsymbol{x}^*) \cap \mathcal{S}_\Theta = \{\boldsymbol{x}^*\}$. It follows from Theorem 4.10 that these cups are basins of attraction of $\boldsymbol{x}^*$, and hence so is $\mathcal{G}_1(\boldsymbol{x}^*)$. Likewise, $\mathcal{G}_{\Theta(\boldsymbol{z}^*)}(\boldsymbol{y}^*)$ is a basin of attraction of $\boldsymbol{y}^*$. (We also observe that a point in a cup is not necessarily in a bounded sublevel set, as $\{\Theta \leqslant \Theta(\boldsymbol{x})\}$ is unbounded if $\boldsymbol{x} \in \mathcal{G}_{\Theta(\boldsymbol{z}^*)}(\boldsymbol{y}^*) \cap \{\Theta \geqslant 1\}$.) The basins $\mathcal{G}_1(\boldsymbol{x}^*)$ and $\mathcal{G}_{\Theta(\boldsymbol{z}^*)}(\boldsymbol{y}^*)$ are shown in Figure 2(b). Since their union contains all the generalized cups of Θ, Theorem 4.9 does not tell us more about the convergence of the inexact FHQ algorithm; so there is neither global nor local convergence guarantee when the starting point is outside $\mathcal{G}_1(\boldsymbol{x}^*) \cup \mathcal{G}_{\Theta(\boldsymbol{z}^*)}(\boldsymbol{y}^*)$.

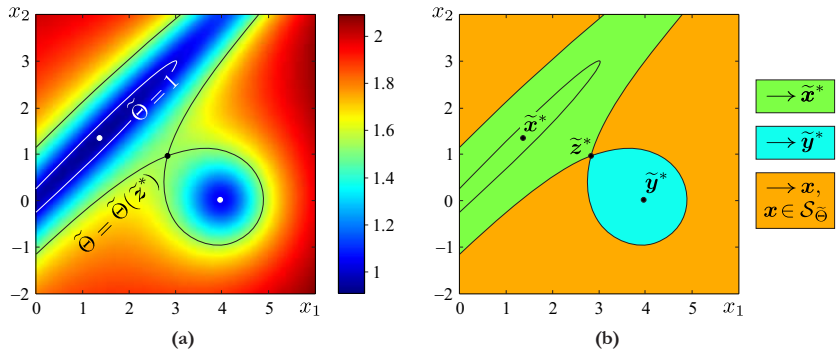


FIGURE 3. Two-dimensional coercive objective: (a) function $\widetilde{\Theta}$ defined in (4.8) and level sets at heights 1 and $\widetilde{\Theta}(\widetilde{\boldsymbol{z}}^*)$; (b) basins of attraction of the minimizers $\widetilde{\boldsymbol{x}}^*$ and $\widetilde{\boldsymbol{y}}^*$ (when started outside these basins, the inexact FHQ algorithm converges to either $\widetilde{\boldsymbol{x}}^*$ or $\widetilde{\boldsymbol{y}}^*$, or to the saddle point $\widetilde{\boldsymbol{z}}^*$).

Example 2. Consider the coercive function

$$\widetilde{\Theta}(\boldsymbol{x}) := \Theta(\boldsymbol{x}) + \gamma_0 \|\boldsymbol{x}\|^2, \tag{4.8}$$

where Θ is as in Example 1 and $\gamma_0 := 5 \times 10^{-3}$ (a value small enough so that the Tikhonov penalty $\gamma_0\|\boldsymbol{x}\|^2$ slightly pulls the stationary points of Θ towards $\mathbf{0}$ without changing their nature). The function $\widetilde{\Theta}$ is shown in Figure 3(a) and the convergence guarantees provided by Theorems 4.9 and 4.10 are depicted in Figure 3(b). By Theorem 4.9(i), the inexact FHQ algorithm converges to a stationary point independently of the starting point. By Theorem 4.10, the basins of attraction of $\widetilde{\boldsymbol{x}}^*$ and $\widetilde{\boldsymbol{y}}^*$ are their respective connected components in $\{\widetilde{\Theta} < \widetilde{\Theta}(\widetilde{\boldsymbol{z}}^*)\}$.

5. Exploiting the Kurdyka-Łojasiewicz property

We know from Theorem 4.9 that the inexact FHQ algorithm converges globally to a stationary point if no level set of the objective contains a nondiscrete set of stationary points. Here, we drop this assumption and show that an inexact FHQ sequence forms a finite-length trajectory to a stationary point if the objective satisfies the KL inequality. This result is stated in Theorem 5.5, preceded by the definition of the KL inequality for C^1 functions and two lemmas.

For any $\alpha \in (0, +\infty)$, we let $\mathscr{C}^\star(\alpha)$ denote the class of real C^1 functions on $(0, \alpha)$ that are positive, strictly increasing, and concave.

Definition 5.1 (KL inequality, KL function)**.** A function $f \in C^1(\mathbb{R}^n)$ is said to satisfy the *Kurdyka-Łojasiewicz (KL) inequality* at a stationary point $\boldsymbol{x} \in \mathcal{S}_f$ if there exist positive constants α and r and a function $\varphi \in \mathscr{C}^\star(\alpha)$ such that

$$\varphi'\big(f(\boldsymbol{y}) - f(\boldsymbol{x})\big)\|\nabla f(\boldsymbol{y})\| \geqslant 1 \quad \text{for all } \boldsymbol{y} \in \mathscr{B}(\boldsymbol{x}, r) \cap \{f - f(\boldsymbol{x}) \in (0, \alpha)\}. \tag{5.1}$$

We call f a *KL function* if it satisfies the KL inequality at each of its stationary points. We denote by $\mathscr{C}_{\mathrm{KL}}(\mathbb{R}^n)$ the class of KL functions in $C^1(\mathbb{R}^n)$, and we define $\mathscr{C}_{\mathrm{KL}} := \bigcup_{n \in [1..\infty)} \mathscr{C}_{\mathrm{KL}}(\mathbb{R}^n)$.

Remark 4. If $f(\boldsymbol{x}) = 0$, the inequality in (5.1) can be written as $\|\nabla(\varphi \circ f)(\boldsymbol{y})\| \geqslant 1$. Therefore, if $\boldsymbol{x}$ is an isolated local minimizer, the informal meaning of the KL inequality is that the values of f can be reparameterized so that the resulting composite function is steep around $\boldsymbol{x}$. Note also that the KL inequality is satisfied at all maximizers and all nonstationary points: if $\boldsymbol{x}$ is a local maximizer, then the set $\mathscr{B}(\boldsymbol{x}, r) \cap \{ f > f(\boldsymbol{x}) \}$ is empty for sufficiently small r, and so (5.1) is trivial; and if $\nabla f(\boldsymbol{x}) \neq \boldsymbol{0}$, then $\|\nabla f(\boldsymbol{y})\| \geqslant \frac{1}{2}\|\nabla f(\boldsymbol{x})\|$ in a neighborhood of $\boldsymbol{x}$, and so (5.1) holds with $\varphi(t) = 2t\|\nabla f(\boldsymbol{x})\|^{-1}$.

In the definition of a KL function, the constants α and r and the function φ depend on the stationary point considered. In fact, as stated in the following lemma, KL functions satisfy the KL inequality uniformly on flat compact sets of stationary points.

Lemma 5.2 (Uniform KL inequality on a flat compact set)**.** *Let $f \in \mathscr{C}_{\mathrm{KL}}(\mathbb{R}^n)$, let $\mathcal{C}$ be a flat compact subset of $\mathcal{S}_f$, and denote by h the value of f on $\mathcal{C}$. There exist positive constants α and r and a function $\varphi \in \mathscr{C}^\star(\alpha)$ such that*

$$\varphi'\big(f(\boldsymbol{y}) - h\big)\|\nabla f(\boldsymbol{y})\| \geqslant 1 \quad \textit{for all } \boldsymbol{y} \in \mathscr{B}(\mathcal{C}, r) \cap \{ f - h \in (0, \alpha) \}, \tag{5.2}$$

where $\mathscr{B}(\mathcal{C}, r) := \{ \boldsymbol{z} \in \mathbb{R}^n : \operatorname{dist}(\boldsymbol{z}, \mathcal{C}) < r \}$.

Proof. By Definition 5.1, for every $\boldsymbol{x} \in \mathcal{C}$, there exist positive constants $\alpha_{\boldsymbol{x}}$ and $r_{\boldsymbol{x}}$ and a function $\varphi_{\boldsymbol{x}} \in \mathscr{C}^\star(\alpha_{\boldsymbol{x}})$ such that

$$\varphi'_{\boldsymbol{x}}\big(f(\boldsymbol{y}) - h\big)\|\nabla f(\boldsymbol{y})\| \geqslant 1 \quad \text{for all } \boldsymbol{y} \in \mathscr{B}(\boldsymbol{x}, r_{\boldsymbol{x}}) \cap \{ f - h \in (0, \alpha_{\boldsymbol{x}}) \}. \tag{5.3}$$

Since $\mathcal{C}$ is compact, there is a finite set $\mathcal{S} \subseteq \mathcal{C}$ such that $\{ \mathscr{B}(\boldsymbol{x}, r_{\boldsymbol{x}}) : \boldsymbol{x} \in \mathcal{S} \}$ is a cover of $\mathcal{C}$. Let

$$\mathscr{B}_{\mathcal{S}} := \bigcup_{\boldsymbol{x} \in \mathcal{S}} \mathscr{B}(\boldsymbol{x}, r_{\boldsymbol{x}})$$

and define the function $\rho : \mathcal{C} \longrightarrow \mathbb{R}_+$ and the radius $r \in \mathbb{R}_+$ by

$$\rho(\boldsymbol{x}) := \sup\big\{ \varrho \in \mathbb{R}_+ : \mathscr{B}(\boldsymbol{x}, \varrho) \subseteq \mathscr{B}_{\mathcal{S}} \big\} \quad \text{and} \quad r := \inf_{\boldsymbol{x} \in \mathcal{C}} \rho(\boldsymbol{x}).$$

Suppose that $r = 0$. Then there is a sequence $(\boldsymbol{x}^{(p)})_p$ in $\mathcal{C}$ such that $\lim_p \rho(\boldsymbol{x}^{(p)}) = 0$. From the compactness of $\mathcal{C}$, $(\boldsymbol{x}^{(p)})_p$ has a subsequence $(\boldsymbol{y}^{(q)})_q$ converging to a point $\boldsymbol{y} \in \mathcal{C} \subset \mathscr{B}_{\mathcal{S}}$. Let $\varrho > 0$ be such that $\mathscr{B}(\boldsymbol{y}, \varrho) \subseteq \mathscr{B}_{\mathcal{S}}$. For sufficiently large q, we have $\mathscr{B}(\boldsymbol{y}^{(q)}, \frac{1}{2}\varrho) \subset \mathscr{B}(\boldsymbol{y}, \varrho)$ and hence $\rho(\boldsymbol{y}^{(q)}) \geqslant \frac{1}{2}\varrho$. This contradicts the fact that $\lim_p \rho(\boldsymbol{x}^{(p)}) = 0$, so $r > 0$.

Let $\boldsymbol{y} \in \mathscr{B}(\mathcal{C}, r)$. There is a point $\boldsymbol{z} \in \mathcal{C}$ such that $\|\boldsymbol{y} - \boldsymbol{z}\| < r \leqslant \rho(\boldsymbol{z})$, and thus $\boldsymbol{y} \in \mathscr{B}_{\mathcal{S}}$ by the definition of ρ. So $\boldsymbol{y} \in \mathscr{B}(\boldsymbol{x}, r_{\boldsymbol{x}})$ for some $\boldsymbol{x} \in \mathcal{S}$. Define the function

$$\psi : t \in (0, \alpha) \longmapsto \max_{\boldsymbol{x} \in \mathcal{S}} \varphi'_{\boldsymbol{x}}(t), \qquad \alpha := \min_{\boldsymbol{x} \in \mathcal{S}} \alpha_{\boldsymbol{x}}.$$

It follows from (5.3) that

$$\psi\big(f(\boldsymbol{y}) - h\big)\|\nabla f(\boldsymbol{y})\| \geqslant 1 \quad \text{if } \boldsymbol{y} \in \{ f - h \in (0, \alpha) \}.$$

The function ψ is positive, decreasing, and continuous; moreover, for every $t \in (0, \alpha)$ and $\epsilon \in (0, t)$,

$$\int_\epsilon^t \psi \leqslant \int_\epsilon^t \sum_{\boldsymbol{x} \in \mathcal{S}} \varphi'_{\boldsymbol{x}}(t) < \sum_{\boldsymbol{x} \in \mathcal{S}} \varphi_{\boldsymbol{x}}(t) < +\infty.$$

Therefore the function

$$\varphi : t \in (0, \alpha) \longmapsto \lim_{\epsilon \to 0^+} \int_{\epsilon}^{t} \psi$$

is well-defined and belongs to $\mathscr{C}^\star(\alpha)$, which completes the proof. □

The next lemma shows that the distance between two successive iterates of a bounded inexact FHQ sequence is bounded below by the norm of the gradient and above by the square root of the objective difference (up to constant factors).

Lemma 5.3 (Further properties of the inexact FHQ map)**.** *Let $\Theta \in \mathscr{C}_{\mathrm{obj}}(\mathbb{R}^n)$ and $\mu \in (0,1)$. For every compact set $\mathcal{C} \subset \mathbb{R}^n$, there exist positive constants c_1 and c_2 such that for all $(\boldsymbol{x}, \boldsymbol{y}) \in \bigcup_{\boldsymbol{z} \in \mathcal{C}} \{\boldsymbol{z}\} \times \boldsymbol{\Phi}_\mu(\boldsymbol{z})$,*

$$c_1 \|\nabla\Theta(\boldsymbol{x})\| \leqslant \|\boldsymbol{y} - \boldsymbol{x}\|, \tag{5.4}$$

$$c_2 \|\boldsymbol{y} - \boldsymbol{x}\|^2 \leqslant \Theta(\boldsymbol{x}) - \Theta(\boldsymbol{y}). \tag{5.5}$$

Proof. Let $\mathcal{C}$ be a compact set in $\mathbb{R}^n$ and let $\boldsymbol{x} \in \mathcal{C}$. We assume that $\boldsymbol{x} \notin \mathcal{S}_\Theta$, for otherwise $\boldsymbol{x}$ is a fixed point of $\boldsymbol{\Phi}_\mu$ (by Lemma 4.5(iii)) and the inequalities (5.4) and (5.5) are trivial. Substituting from (4.3) and (4.4) into the definition of $\boldsymbol{\Phi}_\mu$, we obtain

$$\begin{aligned} \boldsymbol{y} \in \boldsymbol{\Phi}_\mu(\boldsymbol{x}) \iff \|\boldsymbol{y} - \boldsymbol{x}\|^2_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}} + 2(\boldsymbol{y} - \boldsymbol{x})^T \nabla\Theta(\boldsymbol{x}) \\ + (1-\mu)\|\nabla\Theta(\boldsymbol{x})\|^2_{(\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A})^{-1}} \leqslant 0. \end{aligned} \tag{5.6}$$

For every $\boldsymbol{z} \in \mathbb{R}^n$, we have

$$\|\boldsymbol{z}\|^2_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}} \geqslant \lambda_1(\boldsymbol{x})\|\boldsymbol{z}\|^2 \quad \text{and} \quad \|\boldsymbol{z}\|^2_{(\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A})^{-1}} \geqslant (\lambda_n(\boldsymbol{x}))^{-1}\|\boldsymbol{z}\|^2, \tag{5.7}$$

where $\lambda_1(\boldsymbol{x})$ and $\lambda_n(\boldsymbol{x})$ denote the smallest and largest eigenvalues of the $n \times n$ positive definite matrix $\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}$, respectively. Since $\mathcal{C}$ is compact,

$$0 < \inf_{\boldsymbol{z} \in \mathcal{C}} \lambda_1(\boldsymbol{z}) =: \lambda_1(\mathcal{C}) \leqslant \sup_{\boldsymbol{z} \in \mathcal{C}} \lambda_n(\boldsymbol{z}) =: \lambda_n(\mathcal{C}) < \infty. \tag{5.8}$$

Let $\boldsymbol{y} \in \boldsymbol{\Phi}_\mu(\boldsymbol{x})$. From (5.6)–(5.8) and the Cauchy-Schwarz inequality, we deduce that

$$\lambda_1(\mathcal{C})\big(\|\boldsymbol{y} - \boldsymbol{x}\|\,\|\nabla\Theta(\boldsymbol{x})\|^{-1}\big)^2 - 2\|\boldsymbol{y} - \boldsymbol{x}\|\,\|\nabla\Theta(\boldsymbol{x})\|^{-1} + (1-\mu)(\lambda_n(\mathcal{C}))^{-1} \leqslant 0.$$

The quadratic function of $\|\boldsymbol{y}-\boldsymbol{x}\|\,\|\nabla\Theta(\boldsymbol{x})\|^{-1}$ on the left-hand side has two positive roots, which shows that there are positive constants c_1 and $\tilde{c}_1$ independent of $\boldsymbol{x}$ and $\boldsymbol{y}$ such that

$$c_1\|\nabla\Theta(\boldsymbol{x})\| \leqslant \|\boldsymbol{y} - \boldsymbol{x}\| \leqslant \tilde{c}_1\|\nabla\Theta(\boldsymbol{x})\|.$$

Furthermore, from the definition of $\boldsymbol{\Phi}_\mu$ and using (4.3), we have

$$\begin{aligned} \Theta(\boldsymbol{x}) - \Theta(\boldsymbol{y}) &\geqslant \Theta(\boldsymbol{x}) - \overline{\Theta}(\boldsymbol{x}, \boldsymbol{y}) \\ &\geqslant (1-\mu)\big(\Theta(\boldsymbol{x}) - \min \overline{\Theta}(\boldsymbol{x}, \cdot)\big) \\ &= \tfrac{1}{2}(1-\mu)\|\nabla\Theta(\boldsymbol{x})\|^2_{(\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A})^{-1}} \\ &\geqslant \tfrac{1}{2}(1-\mu)(\lambda_n(\mathcal{C})\tilde{c}_1^2)^{-1}\|\boldsymbol{y} - \boldsymbol{x}\|^2. \end{aligned}$$

This completes the proof. □

Definition 5.4 (Finite-length sequence)**.** We say that a sequence $(\boldsymbol{x}^{(p)})_{p \in \mathbb{N}}$ in $\mathbb{R}^n$ has *finite length* if the series $\sum_{p=0}^{\infty} \|\boldsymbol{x}^{(p+1)} - \boldsymbol{x}^{(p)}\|$ converges.

Theorem 5.5 (Finite-length convergence)**.** *Let $\Theta \in \mathscr{C}_{\mathrm{obj}}(\mathbb{R}^n)$ and let $\mathcal{X} := (\boldsymbol{x}^{(p)})_p$ be an inexact FHQ sequence starting in a generalized cup of Θ. If Θ is a KL function, then $\mathcal{X}$ has finite length and converges to a point in $\mathcal{S}_\Theta$.*

Proof. Suppose $\Theta \in \mathscr{C}_{\mathrm{KL}}(\mathbb{R}^n)$ and let $\mu \in (0,1)$ be such that $\mathcal{X}$ is generated by $\boldsymbol{\Phi}_\mu$. Since the fixed points of $\boldsymbol{\Phi}_\mu$ are the stationary points of Θ (see Lemma 4.5(iii)), the theorem is trivial if $\boldsymbol{x}^{(p)} \in \mathcal{S}_\Theta$ for some p. We assume from now on that $\boldsymbol{x}^{(p)} \notin \mathcal{S}_\Theta$ for all p. By Theorems 4.7 and 4.9,

(a) the set $\mathcal{C}_\mathcal{X}$ of cluster points of $\mathcal{X}$ is a flat compact subset of $\mathcal{S}_\Theta$,
(b) $(\Theta(\boldsymbol{x}^{(p)}))_p$ is strictly decreasing to the value h of Θ on $\mathcal{C}_\mathcal{X}$,
(c) $\lim_p \operatorname{dist}(\boldsymbol{x}^{(p)}, \mathcal{C}_\mathcal{X}) = 0$.

It follows from (a) that Lemma 5.2 applies with $f = \Theta$ and $\mathcal{C} = \mathcal{C}_\mathcal{X}$. Let α, r, and φ be as in the conclusion of Lemma 5.2, and set $t_p := \Theta(\boldsymbol{x}^{(p)}) - h$ for all p. From (b) and (c), we have $t_p \in (0, \alpha)$ and $\boldsymbol{x}^{(p)} \in \mathscr{B}(\mathcal{C}_\mathcal{X}, r)$ for all p sufficiently large; so there exists $q \in \mathbb{N}$ such that the inequality in (5.2) holds for all $\boldsymbol{y} \in \{\boldsymbol{x}^{(p)} : p \geqslant q\}$. Let $p \geqslant q$. Using the concavity of φ, we deduce that

$$\varphi(t_p) - \varphi(t_{p+1}) \;\geqslant\; \varphi'(t_p)(t_p - t_{p+1}) \;\geqslant\; \|\nabla\Theta(\boldsymbol{x}^{(p)})\|^{-1}(t_p - t_{p+1}).$$

Since $\mathcal{X}$ is bounded by items (i) and (ii) of Lemma 4.5, it follows from Lemma 5.3 that there exists a positive constant c independent of p such that

$$\varphi(t_p) - \varphi(t_{p+1}) \;\geqslant\; c\|\boldsymbol{x}^{(p+1)} - \boldsymbol{x}^{(p)}\|.$$

Therefore

$$\begin{aligned} c \sum_{p \in [q..\infty)} \|\boldsymbol{x}^{(p+1)} - \boldsymbol{x}^{(p)}\| \;&\leqslant\; \lim_m \sum_{p \in [q..m]} \big(\varphi(t_p) - \varphi(t_{p+1})\big) \\ &=\; \varphi(t_q) - \lim_m \varphi(t_{m+1}) \\ &\leqslant\; \varphi(t_q), \end{aligned}$$

where the last inequality follows from (b) and the fact that φ is positive and strictly increasing. So $\mathcal{X}$ has finite length and hence converges. Furthermore, since $\mathcal{C}_\mathcal{X} \subseteq \mathcal{S}_\Theta$, the limit point of $\mathcal{X}$ is a stationary point. □

Remark 5. A bounded inexact FHQ sequence $\mathcal{X}$ is a descent sequence generated by Algorithm 1 in [21] if the following conditions hold:

(i) there is a constant $\mathfrak{K} > 0$ and a compact set $\mathcal{C}$ containing $\mathcal{X}$ such that $\nabla\Theta$ is $\mathfrak{K}$-Lipschitz continuous on $\mathcal{C}$;
(ii) there is a constant $c > \mathfrak{K}$ such that for all $\boldsymbol{x} \in \mathcal{C}$ and $\boldsymbol{y} \in \boldsymbol{\Phi}_\mu(\boldsymbol{x}) \cap \mathcal{C}$,

$$c\|\boldsymbol{y} - \boldsymbol{x}\|^2 + 2(\boldsymbol{y} - \boldsymbol{x})^T \nabla\Theta(\boldsymbol{x}) \;\leqslant\; 0. \tag{5.9}$$

It can be shown that (ii) is satisfied if

$$\mathfrak{K} \;<\; 2\lambda_1(\mathcal{C})\Big(1 + \big[1 - (1-\mu)\lambda_1(\mathcal{C})(\lambda_n(\mathcal{C}))^{-1}\big]^{1/2}\Big)^{-1} \;=:\; \mathfrak{K}_\mu, \tag{5.10}$$

where $\lambda_1(\mathcal{C})$ and $\lambda_n(\mathcal{C})$ are defined as in (5.8). The quantity $\mathfrak{K}_\mu$ decreases with increasing μ and is bounded by $2\lambda_1(\mathcal{C})$. It follows that if (i) holds with $\mathfrak{K}$ sufficiently small, then Theorem 5.5 is a special case of the general convergence result in [21, Theorem 3.2].

Two important classes of KL functions are those of real analytic functions and real semialgebraic functions [22], for which the KL inequality holds with $\varphi(t) \propto t^{\upsilon}$ for some $\upsilon \in (0,1]$. Recall that a set in $\mathbb{R}^n$ is called semialgebraic if it is the union of finitely many sets of the form

$$\bigcap_{i \in [1..l]} \big\{ \boldsymbol{x} \in \mathbb{R}^n \,:\, P_i(\boldsymbol{x}) = 0,\, Q_i(\boldsymbol{x}) > 0 \big\}, \tag{5.11}$$

where $P_1, Q_1, \ldots, P_l, Q_l$ are real polynomial functions on $\mathbb{R}^n$. A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called semialgebraic if its graph $\{(\boldsymbol{x}, h) \in \mathbb{R}^n \times \mathbb{R} \,:\, f(\boldsymbol{x}) = h\}$ is a semialgebraic set in $\mathbb{R}^{n+1}$. Elementary examples of semialgebraic functions are polynomial and rational functions, the absolute value function, and the square root function. Semialgebraic sets and functions have numerous stability properties [23]; in particular, finite sums and products of semialgebraic functions are semialgebraic, the composition of semialgebraic functions is semialgebraic, and images and preimages of semialgebraic sets under semialgebraic functions are semialgebraic. As for the objective Θ, the functions $\boldsymbol{x} \in \mathbb{R}^n \longmapsto \|\boldsymbol{A}_k \boldsymbol{x} - \boldsymbol{a}_k\|$ are semialgebraic, and so Theorem 5.5 applies if the potentials are semialgebraic. Examples of semialgebraic potentials among those given in Appendix A are the function of minimal surface (A.2), the Lange function (A.4), the Huber function (A.8), the Geman and McClure function (A.10), and the Tukey's biweight function (A.13).

The scope of Theorem 5.5 is not limited to analytic and semialgebraic objectives: as we will see in the next section, the class $\mathscr{C}_{\mathrm{KL}}$ contains all the objective functions that are definable in some o-minimal structure on the real field. Before that, we give simple examples showing that Theorems 4.9 and 5.5 complement each other. On one hand, Theorem 5.5 extends Theorem 4.9 because it guarantees a stronger mode of convergence and also because $\mathscr{C}_{\mathrm{obj}} \cap \mathscr{C}_{\mathrm{KL}}$ contains functions with non-level-discrete sets of stationary points (Example 3). On the other hand, Theorem 5.5 does not generalize Theorem 4.9 because $\mathscr{C}_{\mathrm{obj}} \setminus \mathscr{C}_{\mathrm{KL}}$ contains functions with level-discrete as well as non-level-discrete sets of stationary points (Examples 4 and 5).

Example 3 (KL objective with a flat continuum of stationary points). Let $a, b \in \mathbb{R}$ be such that $a < b$, and let $\Theta_{[a,b]} \in \mathscr{C}_{\mathrm{obj}}(\mathbb{R})$ be defined by

$$\Theta_{[a,b]}(x) \,:=\, \vartheta_{\mathrm{Hu}}\big(|x - (b+1)|\big) + \vartheta_{\mathrm{Hu}}\big(|x - (a-1)|\big), \tag{5.12}$$

where ϑ_{Hu} is the Huber function defined in (A.8). The function $\Theta_{[a,b]}$ is semialgebraic, strictly decreasing on $(-\infty, a)$, constant on $[a,b]$, and strictly increasing on $(b, +\infty)$; so it is a KL function and its set of stationary points is the flat continuum $[a,b]$.

Example 4 (Non-KL objective with a level-discrete set of stationary points). Let θ_1 and θ_2 be the admissible potentials defined as follows:

$$\theta_1(t) \,:=\, \begin{cases} t^2(1 - \frac{1}{3}t) & \text{if } t \in [0,1], \\ t - \frac{1}{3} & \text{if } t > 1, \end{cases}$$

$$\theta_2(t) \,:=\, \begin{cases} \theta_1(t) & \text{if } t \in [0,1], \\ \frac{2}{3} + \displaystyle\int_1^t u\,\eta(u-1)\,\mathrm{d}u & \text{if } t > 1, \end{cases} \tag{5.13}$$

where $\eta : (0, +\infty) \longrightarrow \mathbb{R}$ is given by

$$\eta(t) \,:=\, \big(1 + t\sin(t^{-1})\big)\big(1 + t + \sin(t^{-1})\big)^{-1}$$

(θ_2 is admissible because η is positive and decreasing and $\lim_{t\to 0^+}\eta(t)=1$). Let $\Theta\in\mathscr{C}_{\mathrm{obj}}(\mathbb{R})$ be defined by

$$\Theta(x) \;:=\; \theta_1(|x-2|)+\theta_2(|x|).$$

This function is decreasing on $(-\infty,1]$ and increasing on $[1,+\infty)$, and its set of stationary points is

$$\mathcal{S}_\Theta \;=\; \{1\}\cup\{x^{(k)}:k\in\mathbb{N}\},\qquad x^{(k)} \;:=\; 1+2((4k+3)\pi)^{-1}.$$

Therefore $\mathcal{S}_\Theta$ is level-discrete and Θ does not satisfy the KL inequality at $x=1$ (since otherwise its derivative at $x^{(k)}$ would be nonzero for large k).

Example 5 (Non-KL objective with a flat continuum of stationary points). Let Θ be as in Example 4, and let $\widetilde{\Theta}\in\mathscr{C}_{\mathrm{obj}}(\mathbb{R})$ be defined as Θ but with

$$\theta_1(t) \;:=\; \begin{cases} t^2(1-\frac{1}{3}t) & \text{if } t\in[0,2],\\ \frac{4}{3} & \text{if } t>2.\end{cases}$$

The function $\widetilde{\Theta}$ is constant on $[0,1]$ and equal to Θ on $[1,3]$; therefore $\mathcal{S}_{\widetilde{\Theta}}$ contains the continuum $[0,1]$ and $\widetilde{\Theta}$ does not satisfy the KL inequality at $x=1$.

6. Tame half-quadratic optimization

The KL property is difficult to verify when the objective is neither analytic nor semialgebraic. The theory of o-minimal structures, based on an axiomatization of semialgebraic geometry [14–16], provides useful tools to overcome this problem. (The fundamental definitions and results concerning o-minimal structures are recalled in Appendix B.) We build here on the notion of *tameness* adapted from [24], which refers to functions whose restrictions to balls are o-minimal. First, in Section 6.1, we show that tame potentials yield KL objectives with "nice" sets of stationary points. Next, in Section 6.2, we focus on the special case of piecewise analytic potentials, for which we can specify the convergence rate of the inexact FHQ algorithm. Finally, in Section 6.3, we propose a rule-based model for constructing tame potentials and hence KL objectives.

6.1. Ensuring the KL property.

We begin by showing that tameness implies the KL property (Proposition 6.2) and is guaranteed by tame potentials (Theorem 6.3). Next, we give some structural properties of the sets of stationary points of tame functions (Proposition 6.5), followed by an overview of the global convergence properties of the inexact FHQ algorithm.

Definition 6.1 (Tame function)**.** Let $\mathcal{M}$ be an o-minimal structure on $(\mathbb{R},+,\cdot)$. A function $f:\mathcal{A}\subseteq\mathbb{R}^n\longrightarrow\mathbb{R}$ is said to be *tame* in $\mathcal{M}$ (or simply tame) if for all $r\in(0,+\infty)$, the restriction $f|_{\mathcal{A}\cap\mathscr{B}(\mathbf{0},r)}$ is $\mathcal{M}$-definable.

Proposition 6.2 (Tame functions are KL functions)**.** *Let $f\in C^1(\mathbb{R}^n)$. If f is tame, then f is a KL function.*

Proof. Let $f\in C^1(\mathbb{R}^n)$ be tame in an o-minimal structure $\mathcal{M}$, and let $\boldsymbol{x}$ be a local maximizer or a saddle point of f (recall that the KL inequality is trivial at all maximizers). By the definition of a tame function, $f|_{\mathscr{B}(\mathbf{0},2(\|\boldsymbol{x}\|+1))}$ is $\mathcal{M}$-definable, and thus so is the restriction of f to the set

$$\mathcal{O} \;:=\; \mathscr{B}(\boldsymbol{x},\|\boldsymbol{x}\|+1)\cap\{f>f(\boldsymbol{x})\}.$$

Applying Theorem B.6 to the function $\boldsymbol{y} \in \mathcal{O} \longmapsto f(\boldsymbol{y}) - f(\boldsymbol{x})$, we obtain that there exist a positive constant α and a strictly increasing, $\mathcal{M}$-definable C^1 function $\varphi : (0, +\infty) \longrightarrow (0, +\infty)$ such that

$$\varphi'\big(f(\boldsymbol{y}) - f(\boldsymbol{x})\big)\|\nabla f(\boldsymbol{y})\| \geqslant 1 \quad \text{for all } \boldsymbol{y} \in \mathcal{O} \cap \{f < f(\boldsymbol{x}) + \alpha\}. \tag{6.1}$$

Furthermore, since the derivative of a differentiable definable function of a real variable is definable, we have from Theorem B.5 that φ' is either strictly monotone or constant on an interval $(0, \alpha')$, $\alpha' > 0$. So it remains to show that $\lim_{t \to 0^+} \varphi'(t) = +\infty$, because then φ' is strictly decreasing on $(0, \alpha')$ and hence φ is concave on this interval. Since $\boldsymbol{x}$ is not a local maximizer, there exists a sequence $(\boldsymbol{y}^{(p)})_{p \in [1..\infty)}$ such that

$$\boldsymbol{y}^{(p)} \in \mathscr{B}(\boldsymbol{x}, p^{-1}) \cap \{f > f(\boldsymbol{x})\} \quad \text{for all } p \geqslant 1.$$

Let $(t_p)_{p \in [1..\infty)}$ be the sequence defined by $t_p := f(\boldsymbol{y}^{(p)}) - f(\boldsymbol{x})$ for all p. Then $\lim_p t_p = 0^+$, and it follows from (6.1) that

$$\|\nabla f(\boldsymbol{y}^{(p)})\| \geqslant \big(\varphi'(t_p)\big)^{-1} > 0 \quad \text{for all } p \text{ sufficiently large.}$$

But $\lim_p \nabla f(\boldsymbol{y}^{(p)}) = \nabla f(\boldsymbol{x}) = \boldsymbol{0}$, and so $\lim_p \varphi'(t_p) = +\infty$. Hence, from the monotonicity of φ' on $(0, \alpha')$, we have $\lim_{t \to 0^+} \varphi'(t) = +\infty$. □

Theorem 6.3 (Tame potentials yield KL objectives)**.** *Let $\Theta \in C^1(\mathbb{R}^n)$ be of the form* (1.3) *and suppose the potentials $\theta_1, \ldots, \theta_K$ are tame in an o-minimal structure $\mathcal{M}$. Then Θ is tame in $\mathcal{M}$ and hence is a KL function.*

Proof. Suppose $\theta_1, \ldots, \theta_K$ are tame in an o-minimal structure $\mathcal{M}$. Let $r > 0$ and define, for every $k \in [1\,..\,K]$,

$$\rho_k := 1 + \max_{\boldsymbol{x} \in \overline{\mathscr{B}(\boldsymbol{0}, r)}} \|\boldsymbol{A}_k \boldsymbol{x} - \boldsymbol{a}_k\| < +\infty.$$

Each function $\boldsymbol{x} \in \mathscr{B}(\boldsymbol{0}, r) \longmapsto \theta_k\big(\|\boldsymbol{A}_k \boldsymbol{x} - \boldsymbol{a}_k\|\big)$ is $\mathcal{M}$-definable as the composition of the restriction $\theta_k|_{[0, \rho_k)}$, which is $\mathcal{M}$-definable by the tameness assumption, with a semialgebraic function. So $\Theta|_{\mathscr{B}(\boldsymbol{0}, r)}$ is a sum of $\mathcal{M}$-definable functions and hence is $\mathcal{M}$-definable. This shows that Θ is tame in $\mathcal{M}$. The KL property then follows from Proposition 6.2. □

From Theorems 5.5 and 6.3, tame admissible potentials and Condition (C4) guarantee that an inexact FHQ sequence started in a generalized cup has finite length and converges to a stationary point. Tame objectives locally enjoy the properties of o-minimal functions and therefore behave "nicely"; in particular, by Theorem B.5, their restrictions to lines are piecewise smooth and monotone. Tameness also impacts the set of stationary points, as Proposition 6.5 below shows.

Lemma 6.4 (Finiteness of the stationary values)**.** *Let f be a tame C^1 function. For every $r \in (0, +\infty)$, the set $\{f(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{S}_f \cap \mathscr{B}(\boldsymbol{0}, r)\}$ is finite.*

Proof. Let $f \in C^1(\mathbb{R}^n)$ be tame in an o-minimal structure $\mathcal{M}$, let $r > 0$, and define $\mathcal{B} := f(\mathcal{S}_f \cap \mathscr{B}(\boldsymbol{0}, r))$. The set $\mathcal{B}$ is $\mathcal{M}$-definable and hence a finite union of intervals and points, so it suffices to show that it cannot contain a nondegenerate interval. Let $\boldsymbol{x} \in \mathcal{S}_f \cap \mathscr{B}(\boldsymbol{0}, r)$ and $\mathcal{O} := \mathscr{B}(\boldsymbol{0}, r) \cap \{f > f(\boldsymbol{x})\}$. Applying Theorem B.6 to the function $\boldsymbol{y} \in \mathcal{O} \longmapsto f(\boldsymbol{y}) - f(\boldsymbol{x})$, we obtain that there exists a positive constant α such that $\|\nabla f(\boldsymbol{y})\| > 0$ for all $\boldsymbol{y} \in \mathcal{O} \cap \{f < f(\boldsymbol{x}) + \alpha\}$. It follows that $f(\boldsymbol{y}) \geqslant f(\boldsymbol{x}) + \alpha$ for all $\boldsymbol{y} \in \mathcal{S}_f \cap \mathcal{O}$, and thus $\mathcal{B} \cap (f(\boldsymbol{x}), f(\boldsymbol{x}) + \alpha) = \emptyset$. This

shows that for every $h \in \mathcal{B}$, there is an $\alpha > 0$ such that $\mathcal{B} \cap (h, h+\alpha) = \emptyset$, which completes the proof. □

Proposition 6.5 (Tame functions have "nice" sets of stationary points)**.** *Let f be a tame C^1 function.*

(i) *The intersection of $\mathcal{S}_f$ with any ball has finitely many connected components, each of which is path-connected.*

(ii) *Either $\mathcal{S}_f$ is discrete or it contains a nondegenerate continuum.*

(iii) *The connected components of $\mathcal{S}_f$ are flat*[1]*.*

Proof. Let $f \in C^1(\mathbb{R}^n)$ be tame in some o-minimal structure $\mathcal{M}$.

(i) $\mathcal{S}_f$ is the preimage of the singleton $\{0\}$ under ∇f and hence is $\mathcal{M}$-definable. Therefore the intersections $\mathcal{S}_f \cap \mathscr{B}(\mathbf{0}, r)$ and $\mathcal{S}_f \cap \overline{\mathscr{B}(\mathbf{0}, r)}$ are $\mathcal{M}$-definable for all $r > 0$, and the result follows from Theorem B.4.

(ii) Suppose $\mathcal{S}_f$ is nondiscrete, so there exists a convergent sequence $(\boldsymbol{x}^{(p)})_p$ in $\mathcal{S}_f$. Let $r := \|\lim_p \boldsymbol{x}^{(p)}\| + 1$. The set $\mathcal{S}_f \cap \overline{\mathscr{B}(\mathbf{0}, r)}$ has finitely many connected components, each of which is compact, and it contains all the points $\boldsymbol{x}^{(p)}$ for sufficiently large p. Therefore at least one of the connected components of $\mathcal{S}_f \cap \overline{\mathscr{B}(\mathbf{0}, r)}$, and hence of $\mathcal{S}_f$, must be a nondegenerate continuum.

(iii) We first show that for every $\boldsymbol{x} \in \mathcal{S}_f$, there is an $r > 0$ such that $\mathcal{S}_f \cap \mathscr{B}(\boldsymbol{x}, r)$ is flat. Let $\boldsymbol{x} \in \mathcal{S}_f$ and suppose for contradiction that for every $r > 0$, there are two points $\boldsymbol{u}_r$ and $\boldsymbol{v}_r$ in $\mathcal{S}_f \cap \mathscr{B}(\boldsymbol{x}, r)$ such that $f(\boldsymbol{u}_r) \neq f(\boldsymbol{v}_r)$. Define the sequence $\mathcal{X} := (\boldsymbol{x}^{(p)})_{p \in [2..\infty)}$ by $(\boldsymbol{x}^{(2q)}, \boldsymbol{x}^{(2q+1)}) = (\boldsymbol{u}_{1/q}, \boldsymbol{v}_{1/q})$ for all $q \geqslant 1$. Since $\mathcal{X}$ converges to $\boldsymbol{x}$ and $(f(\boldsymbol{x}^{(p)}))_p$ is not eventually constant, $\mathcal{X}$ has a subsequence $(\boldsymbol{y}^{(q)})_q$ such that $|f(\boldsymbol{y}^{(q)}) - f(\boldsymbol{x})|$ is strictly decreasing, which contradicts Lemma 6.4.

Let $\mathcal{C}$ be a connected component of $\mathcal{S}_f$. For every $\boldsymbol{x} \in \mathcal{C}$, there is an $r > 0$ such that f is constant on the intersection of $\mathcal{C}$ with $\mathscr{B}(\boldsymbol{x}, r) =: \mathscr{B}_{\boldsymbol{x}}$. Consider the equivalence relation $\sim$ on $\mathcal{C}$ defined as follows: $\boldsymbol{x} \sim \boldsymbol{y}$ if and only if there exists a finite sequence $(\boldsymbol{x}^{(i)})_{i \in [1..k]}$ in $\mathcal{C}$ such that $\boldsymbol{x} \in \mathscr{B}_{\boldsymbol{x}^{(1)}}$, $\boldsymbol{y} \in \mathscr{B}_{\boldsymbol{x}^{(k)}}$, and $\mathscr{B}_{\boldsymbol{x}^{(i)}} \cap \mathscr{B}_{\boldsymbol{x}^{(i+1)}} \cap \mathcal{C} \neq \emptyset$ for all $i \in [1..k-1]$ (this latter condition is void if $k = 1$). Each equivalence class of $\sim$ is flat, so we must show that the quotient set $\mathcal{C}/\sim$ contains only one equivalence class. Let $\mathcal{E} \in \mathcal{C}/\sim$ and let $(\boldsymbol{y}^{(p)})_p$ be a sequence in $\mathcal{E}$ converging to some point $\boldsymbol{y} \in \mathbb{R}^n$. Since $\mathcal{C}$ is closed, $\boldsymbol{y} \in \mathcal{C}$ and so f is constant on the intersection of $\mathcal{C}$ with an open ball $\mathscr{B}_{\boldsymbol{y}}$ centered at $\boldsymbol{y}$. For sufficiently large p, $\boldsymbol{y}^{(p)} \in \mathscr{B}_{\boldsymbol{y}}$ and thus $\boldsymbol{y} \sim \boldsymbol{y}^{(p)}$. This shows that the equivalence classes of $\mathcal{C}$ by $\sim$ are closed. It follows that $\mathcal{C}/\sim$ contains only one equivalence class, for otherwise $\mathcal{C}/\sim$ would be a nontrivial partition of $\mathcal{C}$ into closed subsets, in contradiction to the connectedness of $\mathcal{C}$. □

Let $\mathscr{C}_{\text{tame}}(\mathbb{R}^n)$ denote the class of tame C^1 functions on $\mathbb{R}^n$, let $\mathscr{C}_{\text{LD}}(\mathbb{R}^n)$ denote the class of real C^1 functions on $\mathbb{R}^n$ whose set of stationary points is level-discrete, and define $\mathscr{C}_{\text{tame}} := \bigcup_{n \in [1..\infty)} \mathscr{C}_{\text{tame}}(\mathbb{R}^n)$ and $\mathscr{C}_{\text{LD}} := \bigcup_{n \in [1..\infty)} \mathscr{C}_{\text{LD}}(\mathbb{R}^n)$. Figure 4 summarizes the global convergence properties of the inexact FHQ algorithm depending on the conditions imposed on the objective Θ. We know from Examples 3–5 that the sets $\mathscr{C}_{\text{obj}} \cap \mathscr{C}_{\text{tame}} \setminus \mathscr{C}_{\text{LD}}$ (yellow), $\mathscr{C}_{\text{obj}} \cap \mathscr{C}_{\text{LD}} \setminus \mathscr{C}_{\text{KL}}$ (green), and $\mathscr{C}_{\text{obj}} \setminus \mathscr{C}_{\text{KL}} \setminus \mathscr{C}_{\text{LD}}$ (blue) are nonempty. Example 6 below shows that this is also

[1]The connected components of the set of stationary points of a C^1 function are not necessarily flat—see the famous example of Whitney [25].

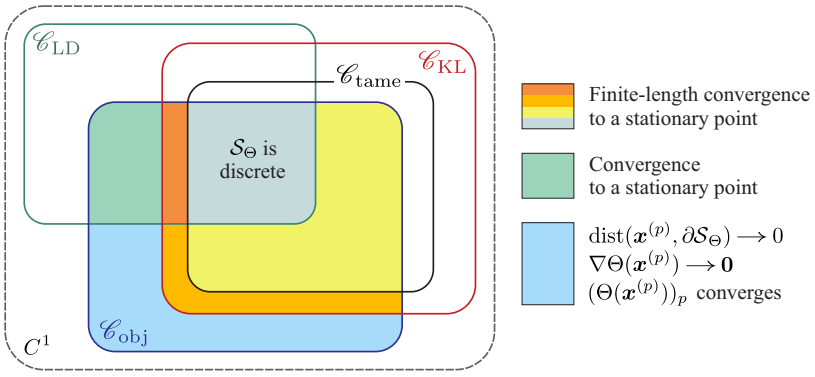


FIGURE 4. Summary of the global convergence properties of an inexact FHQ sequence $(\boldsymbol{x}^{(p)})_p$ starting in a generalized cup of the objective Θ (Theorems 4.7(iv), 4.9, 5.5, and 6.3).

the case for $\mathscr{C}_{\mathrm{obj}} \cap \mathscr{C}_{\mathrm{KL}} \cap \mathscr{C}_{\mathrm{LD}} \setminus \mathscr{C}_{\mathrm{tame}}$ (orange) and $\mathscr{C}_{\mathrm{obj}} \cap \mathscr{C}_{\mathrm{KL}} \setminus \mathscr{C}_{\mathrm{LD}} \setminus \mathscr{C}_{\mathrm{tame}}$ (light orange). Note that the set of stationary points $\mathcal{S}_\Theta$ is necessarily discrete if $\Theta \in \mathscr{C}_{\mathrm{obj}} \cap \mathscr{C}_{\mathrm{tame}} \cap \mathscr{C}_{\mathrm{LD}}$. Indeed, if $\Theta \in \mathscr{C}_{\mathrm{obj}} \cap \mathscr{C}_{\mathrm{tame}}$ and $\mathcal{S}_\Theta$ is nondiscrete, then it follows from Proposition 6.5 that $\mathcal{S}_\Theta$ contains a flat nondegenerate continuum and hence is not level-discrete. In other words, there is no distinction between discrete and level-discrete sets of stationary points for tame objectives.

Example 6 (Nontame KL objectives). Let θ_1 and θ_2 be the admissible potentials defined as follows:

$$\theta_1(t) := \begin{cases} t^2(\frac{1}{2} - \frac{1}{3}t) & \text{if } t \in [0,1], \\ \frac{1}{6} & \text{if } t > 1, \end{cases}$$

$$\theta_2(t) := \int_0^{\min\{t,1\}} u\eta(1-u)\,\mathrm{d}u, \tag{6.2}$$

where $\eta : (0,1] \longrightarrow \mathbb{R}$ is given by

$$\eta(t) := t\big(1 - t + t\sin(t^{-1})\big)^{-1}$$

(θ_2 is admissible because η increases from $\lim_{t\to 0^+}\eta(t) = 0$ to $\eta(1) = (\sin 1)^{-1}$). Consider the function $\Theta : \mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$\Theta(x) := \theta_1(|x-1|) + \theta_2(|x|).$$

The set of stationary points of Θ is

$$\mathcal{S}_\Theta = (-\infty, -1] \cup \{0, 1\} \cup \{1 - ((k+1)\pi)^{-1} : k \in \mathbb{N}\} \cup [2, +\infty),$$

and Θ is strictly decreasing on $[-1, 0]$ and strictly increasing on $[0, 2]$. Since $\mathcal{S}_\Theta \cap (-1, 1)$ has infinitely many connected components (all of which are singletons), it follows from Proposition 6.5(i) that $\Theta|_{(-1,1)}$ is nondefinable; so Θ is not tame. On the other hand, Θ is a KL function: it is analytic on $(-1, 1)\setminus\{0\}$ and semialgebraic on $(1, 2)$, it has a local maximum at $x = -1$, and it satisfies the KL inequality at $x = 0$ and $x = 1$ with $\varphi(t) \propto \sqrt{t}$.

Since θ_1 and θ_2 are constant on $[1, +\infty)$, the function Θ does not belong to $\mathscr{C}_{\mathrm{obj}}$ (Condition (C4) is not satisfied). To obtain a function in $\mathscr{C}_{\mathrm{obj}}$ that behaves like Θ

on $[-1,2]$, we can add to Θ the semialgebraic objective $\Theta_{[a,b]}$ defined in Example 3 with $a \leqslant -1$ and $b \geqslant 2$. The sum $\Theta+\Theta_{[a,b]}$ then belongs to $\mathscr{C}_{\mathrm{obj}} \cap \mathscr{C}_{\mathrm{KL}} \setminus \mathscr{C}_{\mathrm{tame}}$ and its set of stationary points is $\mathcal{S}_\Theta \cap [a,b]$. This set is level-discrete if $a=-1$ and $b=2$, while it contains a flat nondegenerate continuum if $a<-1$ or $b>2$. Therefore $\Theta+\Theta_{[-1,2]} \in \mathscr{C}_{\mathrm{obj}} \cap \mathscr{C}_{\mathrm{KL}} \cap \mathscr{C}_{\mathrm{LD}} \setminus \mathscr{C}_{\mathrm{tame}}$, while $\Theta+\Theta_{[a,b]} \in \mathscr{C}_{\mathrm{obj}} \cap \mathscr{C}_{\mathrm{KL}} \setminus \mathscr{C}_{\mathrm{LD}} \setminus \mathscr{C}_{\mathrm{tame}}$ if $a<-1$ or $b>2$.

6.2. **The special case of piecewise analytic potentials.** Here we consider the usual case where the potentials are piecewise-defined by analytic functions restricted to closed intervals. This assumption ensures that the objective has the classical Łojasiewicz property (that is, for every $\boldsymbol{x} \in \mathcal{S}_\Theta$ there is a positive integer s such that $|\Theta-\Theta(\boldsymbol{x})|^{1-1/s}\|\nabla\Theta\|^{-1}$ is bounded around $\boldsymbol{x}$), allowing us to relate the convergence rate to the geometric properties of the objective around its stationary points.

Definition 6.6 (Restricted analytic function)**.** A function $f : \mathcal{A} \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ is said to be *restricted analytic* if there exist an open set $\mathcal{O} \subseteq \mathbb{R}^n$ and an analytic function $g : \mathcal{O} \longrightarrow \mathbb{R}$ such that $\overline{\mathcal{A}} \subset \mathcal{O}$ and $g|_{\mathcal{A}} = f$. The class of restricted analytic functions on $\mathcal{A}$ is denoted by $\mathbf{an}(\mathcal{A})$.

Definition 6.7 (Piecewise analytic function of a real variable)**.** Let $\mathcal{I}$ be a nondegenerate interval. A function $\theta : \mathcal{I} \longrightarrow \mathbb{R}$ is said to be *piecewise analytic* if there exists a finite partition of $\mathcal{I}$ into subintervals $\mathcal{I}_1, \dots, \mathcal{I}_l$ such that $\theta|_{\mathcal{I}_i} \in \mathbf{an}(\mathcal{I}_i)$ for all $i \in [1\,..\,l]$.

Theorem 6.8 (Piecewise analyticity and KL inequality)**.** *Let $\Theta \in C^1(\mathbb{R}^n)$ be of the form* (1.3) *and suppose the potentials $\theta_1, \dots, \theta_K$ are piecewise analytic. Then Θ is tame in $\mathbb{R}_{\mathrm{an}}$ and for every $\boldsymbol{x} \in \mathcal{S}_\Theta$ there is an integer $s \geqslant 2$ such that Θ satisfies the KL inequality at $\boldsymbol{x}$ with $\varphi(t) \propto t^{1/s}$.*

Proof. Suppose $\theta_1, \dots, \theta_K$ are piecewise analytic. For every $k \in [1\,..\,K]$ and $\rho > 0$, the graph of $\theta_k|_{[0,\rho)}$ is a finite union of graphs of restricted analytic functions on bounded intervals, and hence is $\mathbb{R}_{\mathrm{an}}$-definable. So the potentials are tame in $\mathbb{R}_{\mathrm{an}}$, and it follows from Theorem 6.3 that Θ is tame in $\mathbb{R}_{\mathrm{an}}$. By definition, an $\mathbb{R}_{\mathrm{an}}$-definable set is the image of a bounded subanalytic set under an analytic isomorphism (namely, the inverse of (B.1)) and is thus subanalytic; therefore $\Theta|_{\mathscr{B}(\mathbf{0},r)}$ is subanalytic for all $r > 0$. Let $\boldsymbol{x} \in \mathcal{S}_\Theta$. Applying Theorem B.7 to the function $\boldsymbol{y} \in \mathscr{B}(\mathbf{0}, \|\boldsymbol{x}\|+1) \longmapsto \Theta(\boldsymbol{y}) - \Theta(\boldsymbol{x})$, we obtain that there exist positive constants c and α and an integer $s \geqslant 2$ such that

$$\|\nabla\Theta(\boldsymbol{y})\| \geqslant c|\Theta(\boldsymbol{y}) - \Theta(\boldsymbol{x})|^{1-1/s} \quad \text{for all } \boldsymbol{y} \in \mathscr{B}(\boldsymbol{x},1) \cap \{\Theta - \Theta(\boldsymbol{x}) \in (0,\alpha)\},$$

which completes the proof. □

The rational number $1/s \in (0, \frac{1}{2}]$ is called the *Łojasiewicz exponent* of $\boldsymbol{x}$. This exponent can be interpreted as a measure of flatness around $\boldsymbol{x}$: the smaller $1/s$, the flatter Θ. Theorem 6.9 below relates the convergence rate of an inexact FHQ sequence to the Łojasiewicz exponent of its limit, confirming the intuition that the flatter the objective around $\boldsymbol{x} \in \mathcal{S}_\Theta$, the slower the convergence to $\boldsymbol{x}$. (The case $s=2$ arises when Θ is quadratic positive definite in a neighborhood of $\boldsymbol{x}$.)

Theorem 6.9 (Convergence rate and Łojasiewicz exponent)**.** *Let $\Theta \in \mathscr{C}_{\mathrm{obj}}(\mathbb{R}^n)$ and suppose the potentials $\theta_1, \dots, \theta_K$ are piecewise analytic. Let $\mathcal{X} := (\boldsymbol{x}^{(p)})_p$ be an inexact FHQ sequence starting in a generalized cup of Θ, and let $1/s$ be the Łojasiewicz exponent of the limit $\boldsymbol{x}$ of $\mathcal{X}$.*

(i) *If $s = 2$, then there is a constant $\xi \in (0,1)$ such that $\|\boldsymbol{x}^{(p)} - \boldsymbol{x}\| = O(\xi^p)$.*
(ii) *If $s \geqslant 3$, then $\|\boldsymbol{x}^{(p)} - \boldsymbol{x}\| = O(p^{-1/(s-2)})$.*

Proof. Suppose $\theta_1, \dots, \theta_K$ are piecewise analytic. It follows immediately from Theorems 5.5 and 6.8 that $\mathcal{X}$ converges to a stationary point $\boldsymbol{x}$ with Łojasiewicz exponent $1/s$, $s \in [2\,..\,\infty)$. By the triangle inequality, we have

$$\|\boldsymbol{x}^{(p)} - \boldsymbol{x}\| \leqslant \sum_{q \in [p\,..\,\infty)} \|\boldsymbol{x}^{(q+1)} - \boldsymbol{x}^{(q)}\| =: Z_p \qquad \text{for all } p \in \mathbb{N},$$

so it suffices to show that the stated convergence bounds hold with $\|\boldsymbol{x}^{(p)} - \boldsymbol{x}\|$ replaced by Z_p.

Suppose that no point $\boldsymbol{x}^{(p)}$ belongs to $\mathcal{S}_\Theta$ (for otherwise $\|\boldsymbol{x}^{(p)} - \boldsymbol{x}\|$ is eventually zero and the theorem is trivial) and set $t_p := \Theta(\boldsymbol{x}^{(p)}) - \Theta(\boldsymbol{x})$ for all p. The sequence $(t_p)_p$ is strictly decreasing to zero (by Theorem 4.7(iv)), and it follows from the KL inequality at $\boldsymbol{x}$ that there exist $c_0 > 0$ and $p_0 \in \mathbb{N}$ such that

$$\|\nabla\Theta(\boldsymbol{x}^{(p)})\| \geqslant c_0 t_p^{1-1/s} \quad \text{for all } p \geqslant p_0. \tag{6.3}$$

We assume from now on that $p \geqslant p_0$. Since the function $t \in \mathbb{R}_+ \longmapsto t^{1/s}$ is concave,

$$s\big(t_p^{1/s} - t_{p+1}^{1/s}\big) \geqslant t_p^{1/s-1}(t_p - t_{p+1}) \geqslant c_0 \|\nabla\Theta(\boldsymbol{x}^{(p)})\|^{-1}(t_p - t_{p+1}).$$

By Lemma 5.3, there are positive constants c_1 and c_2 independent of p such that

$$c_1\|\nabla\Theta(\boldsymbol{x}^{(p)})\| \leqslant \|\boldsymbol{x}^{(p+1)} - \boldsymbol{x}^{(p)}\| \quad \text{and} \quad c_2\|\boldsymbol{x}^{(p+1)} - \boldsymbol{x}^{(p)}\|^2 \leqslant t_p - t_{p+1}. \tag{6.4}$$

Consequently, there exists $\gamma_0 > 0$ such that

$$\gamma_0\big(t_p^{1/s} - t_{p+1}^{1/s}\big) \geqslant \|\boldsymbol{x}^{(p+1)} - \boldsymbol{x}^{(p)}\|,$$

and hence

$$Z_p \leqslant \gamma_0 \sum_{q \in [p\,..\,\infty)} \big(t_q^{1/s} - t_{q+1}^{1/s}\big) = \gamma_0 t_p^{1/s}. \tag{6.5}$$

From (6.3) and the first inequality in (6.4), we also have

$$t_p^{1-1/s} \leqslant (c_0 c_1)^{-1}\|\boldsymbol{x}^{(p+1)} - \boldsymbol{x}^{(p)}\| = (c_0 c_1)^{-1}(Z_p - Z_{p+1}). \tag{6.6}$$

Combining (6.5) and (6.6), we obtain that there exists $\gamma_1 > 0$ such that

$$Z_p^{s-1} \leqslant \gamma_1(Z_p - Z_{p+1}). \tag{6.7}$$

Consider the case $s = 2$. Since $Z_{p+1} < Z_p$, it follows from (6.7) that

$$Z_{p+1} < \xi Z_p, \qquad \xi := \gamma_1(1+\gamma_1)^{-1}.$$

Therefore $Z_{p_0+q} < \xi^{p_0+q}(Z_{p_0}\xi^{-p_0})$ for all $q \geqslant 1$, and so $Z_p = O(\xi^p)$.

Now suppose that $s \geqslant 3$. Using (6.7) again, we have

$$0 < \gamma_1^{-1} \leqslant (Z_p - Z_{p+1})Z_p^{1-s} \leqslant \int_{Z_{p+1}}^{Z_p} t^{1-s}\,\mathrm{d}t \leqslant (s-2)^{-1}\big(Z_{p+1}^{2-s} - Z_p^{2-s}\big).$$

Hence there exists $\gamma_2 > 0$ such that

$$\gamma_2(p+1-p_0) \leqslant \sum_{q \in [p_0\,..\,p]} \big(Z_{q+1}^{2-s} - Z_q^{2-s}\big) = Z_{p+1}^{2-s} - Z_{p_0}^{2-s},$$

and so

$$Z_{p+1} \leqslant \big(\gamma_2(p+1-p_0) + Z_{p_0}^{2-s}\big)^{1/(2-s)} \sim \big(\gamma_2(p+1)\big)^{1/(2-s)}.$$

Therefore $Z_p = O(p^{1/(2-s)})$. □

To our knowledge, the potentials used in the computer vision and robust statistics literature are piecewise analytic (common admissible examples are given in Appendix A), so Theorem 6.9 covers the current applications of FHQ optimization. Still, a practical model describing a larger class of tame potentials can be useful; this is subject of the next section.

6.3. **Construction rules for tame objectives.** We first provide a simple set of rules for constructing potentials that are tame in $\mathbb{R}_{\text{an,exp}}$. Then we extend this description to include infinite branches of certain special functions. The resulting potentials are tame in expansions of $\mathbb{R}_{\text{an,exp}}$ that contain the graph of either the Riemann zeta function or the gamma function and are closed under integration. We conclude with a summary of the investigated hierarchy of potentials.

6.3.1. *Analytic-exponential tameness.* Piecewise analytic potentials are tame in $\mathbb{R}_{\text{an}}$ (as the proof of Theorem 6.8 shows), but there exist admissible o-minimal potentials that are not. An example is the function

$$\theta_{\text{e}}(t) := \begin{cases} t^2 \exp(t^{-1})\big(t + \exp(t^{-1})\big)^{-1} & \text{if } t > 0, \\ 0 & \text{if } t = 0, \end{cases} \tag{6.8}$$

where exp is the natural exponential function. If $\theta_{\text{e}}|_{(0,r)}$ is $\mathbb{R}_{\text{an}}$-definable for some $r > 0$, then so must be $\exp|_{(r^{-1},+\infty)}$ (because $\exp(u) = u\theta_{\text{e}}(u^{-1})(1 - u^2\theta_{\text{e}}(u^{-1}))$ for all $u > 0$), contradicting the fact that $\mathbb{R}_{\text{an}}$ is polynomially bounded. Therefore none of the restrictions $\theta_{\text{e}}|_{(0,r)}$ is $\mathbb{R}_{\text{an}}$-definable, and so θ_{e} is not tame in $\mathbb{R}_{\text{an}}$. More generally, a potential θ is not tame in $\mathbb{R}_{\text{an}}$ if it *confines* the asymptotics of exp, meaning that there is a finite interval on which the definition of θ requires an infinite branch of the graph of exp. This leads us to consider o-minimal expansions of $\mathbb{R}_{\text{an}}$ in which exp is definable.

Consider the collection of real functions

$$\mathscr{A} := \mathbf{an}([-1,1]) \cup \mathbf{af}(\mathbb{R}) \cup \{\exp\}, \tag{6.9}$$

where $\mathbf{an}([-1,1])$ is the class of restricted analytic functions on $[-1,1]$ and $\mathbf{af}(\mathbb{R})$ is the class of affine functions on $\mathbb{R}$. We denote by $\mathscr{F}$ the class of functions which contains $\mathscr{A}$ and whose members are obtained by applying the following rules finitely many times, where f and g are functions in $\mathscr{F}$ with respective domains $\mathcal{I}_f$ and $\mathcal{I}_g$ in $\mathbb{R}$.

(R1) $f|_{\mathcal{I}} \in \mathscr{F}$ for any nondegenerate interval $\mathcal{I} \subset \mathcal{I}_f$.

(R2) If $\mathcal{I}_f = \mathcal{I}_g$, then the sum $f + g$ and the product fg belong to $\mathscr{F}$.

(R3) If $0 \notin \mathcal{I}_f$, then the reciprocal $1/f$ belongs to $\mathscr{F}$.

(R4) If $f(\mathcal{I}_f) \subseteq \mathcal{I}_g$, then the composite $g \circ f$ belongs to $\mathscr{F}$.

(R5) If f is strictly monotonic, then the inverse f^{-1} belongs to $\mathscr{F}$.

Every function in $\mathscr{F}$ is a real $\mathbb{R}_{\text{an,exp}}$-definable function defined on some interval $\mathcal{I}$ and analytic on $\mathcal{I}^\circ$. In particular, $\mathscr{F}$ contains the real polynomial functions, the natural logarithm function, the power functions $t \in (0,+\infty) \longmapsto t^\upsilon$, $\upsilon \in \mathbb{R}$, the hyperbolic sine and cosine functions and their inverses, the restriction of the tangent function to $(-\frac{\pi}{2}, \frac{\pi}{2})$, and the restrictions of the sine and cosine functions to finite intervals. Note that since $\mathscr{F}$ is a class of o-minimal functions, it does not contain any function f that oscillates infinitely often in the sense that f itself or

one of its successive derivatives has infinitely many isolated zeros (see Remark 10 after Theorem B.5). This is why the sine and cosine functions on infinite intervals are not members of $\mathscr{F}$.

Definition 6.10 (Piecewise $\mathscr{F}$-definable function)**.** Let $\mathcal{I}$ be a nondegenerate interval. A function $\theta : \mathcal{I} \longrightarrow \mathbb{R}$ is said to be *piecewise $\mathscr{F}$-definable* if there exists a finite partition of $\mathcal{I}$ into subintervals $\mathcal{I}_1, \dots, \mathcal{I}_l$ such that $\theta|_{(\mathcal{I}_i)^\circ} \in \mathscr{F}$ for all $i \in [1\,..\,l]$.

Theorem 6.11 (Piecewise $\mathscr{F}$-definable potentials yield KL objectives)**.** *Let $\Theta \in C^1(\mathbb{R}^n)$ be of the form (1.3). Suppose that for every $k \in [1\,..\,K]$, the potential θ_k is piecewise $\mathscr{F}$-definable or there exists $\rho \in (0, +\infty)$ such that $\theta_k|_{(0,\rho)}$ is piecewise $\mathscr{F}$-definable and $\theta_k|_{[\rho,+\infty)}$ is restricted analytic. Then Θ is tame in $\mathbb{R}_{\mathrm{an,exp}}$ and hence is a KL function.*

Proof. First note that piecewise $\mathscr{F}$-definable functions are $\mathbb{R}_{\mathrm{an,exp}}$-definable. Let θ_k be as stated, with the convention that $\rho = +\infty$ if θ_k is piecewise $\mathscr{F}$-definable. Let $r > 0$. If $r \leqslant \rho$, then $\theta_k|_{(0,r)}$ is $\mathbb{R}_{\mathrm{an,exp}}$-definable; and if $r > \rho$, then the graph of $\theta_k|_{(0,r)}$ is the union of the graphs of $\theta_k|_{(0,\rho)}$ and $\theta_k|_{[\rho,r)}$, which are respectively $\mathbb{R}_{\mathrm{an,exp}}$- and $\mathbb{R}_{\mathrm{an}}$-definable. So Theorem 6.3 applies with $\mathcal{M} = \mathbb{R}_{\mathrm{an,exp}}$. □

Remark 6. Theorem 6.11 follows from the fact that piecewise $\mathscr{F}$-definable functions are $\mathbb{R}_{\mathrm{an,exp}}$-definable. So $\mathscr{F}$ can be extended by adding real analytic $\mathbb{R}_{\mathrm{an,exp}}$-definable functions to $\mathscr{A}$.

Remark 7. The case where θ_k is eventually analytic is included in Theorem 6.11 because piecewise analytic potentials are not necessarily piecewise $\mathscr{F}$-definable (indeed, piecewise analyticity implies piecewise $\mathscr{F}$-definability only on finite intervals). Consider, for example, the admissible potential

$$\theta(t) := \begin{cases} (1 + 4t^2)^{1/2} - t^{-1} \sin t & \text{if } t > 0, \\ 0 & \text{if } t = 0. \end{cases} \tag{6.10}$$

While $\theta \in \mathbf{an}(\mathbb{R}_+)$, the restrictions $\theta|_{(\rho,+\infty)}$, $\rho > 0$, are nondefinable (for otherwise $\sin|_{(\rho,+\infty)}$ would be o-minimal) and hence do not belong to $\mathscr{F}$.

Theorem 6.11 generalizes Theorem 6.8 in that piecewise $\mathscr{F}$-definability allows to confine the asymptotics of the exponential function. We say for short that exp is *$\mathscr{F}$-confinable*; clearly, the natural logarithm function, the real power functions, and the hyperbolic functions are also $\mathscr{F}$-confinable.

6.3.2. *Confining special functions.* The limitations of Theorem 6.11 appear when needed to confine special functions such as the Gauss error function

$$\mathrm{erf} : t \in \mathbb{R} \longmapsto 2\pi^{-1/2} \int_0^t \exp(-u^2)\,\mathrm{d}u, \tag{6.11}$$

the Riemann zeta function

$$\zeta : t \in (1, +\infty) \longmapsto \sum_{n \in [1\,..\,\infty)} n^{-t}, \tag{6.12}$$

or the gamma function

$$\Gamma : t \in (0, +\infty) \longmapsto \int_0^{+\infty} u^{t-1} \exp(-u)\,\mathrm{d}u. \tag{6.13}$$

For example, let θ_{erf}, θ_ζ, and θ_Γ be defined on $(0,+\infty)$ as follows:

$$\theta_{\mathrm{erf}}(t) := t^2 \operatorname{erf}(t^{-1}), \tag{6.14}$$

$$\theta_\zeta(t) := t^2\big(t+\zeta(1+t^{-1})\big)^{-1}, \tag{6.15}$$

$$\theta_\Gamma(t) := t^2\Gamma(2+t^{-1})\big(t+\Gamma(2+t^{-1})\big)^{-1}. \tag{6.16}$$

These functions are analytic, and their continuous extensions to $\mathbb{R}_+$ (which are zero at $t=0$) are admissible potentials; but they are not covered by Theorem 6.11 because erf, ζ, and Γ are not $\mathbb{R}_{\mathrm{an,exp}}$-definable [26] and hence not $\mathscr{F}$-confinable.

Remark 8. For every $r>0$, we have $\operatorname{erf}|_{[-r,r]} \in \mathbf{an}([-r,r]) \subset \mathscr{F}$. So the error function can be used to define piecewise $\mathscr{F}$-definable potentials as long as its asymptotics is not confined. We can also use the restrictions of the zeta and gamma functions to any finite interval, though for a different reason: $\zeta|_{(1,1+r)}$ and $\Gamma|_{(0,r)}$ are $\mathbb{R}_{\mathrm{an}}$-definable for all $r>0$ [26] and thus can be added to $\mathscr{A}$ for extending $\mathscr{F}$ (see Remark 6).

The asymptotics of erf, ζ, and Γ could be used freely if these functions were definable in a same o-minimal expansion of $\mathbb{R}_{\mathrm{an,exp}}$, but there is currently no evidence that such a structure exists. The largest currently known o-minimal structures are the so-called *Pfaffian closures* [27–29] of the expansions $\mathbb{R}_{\mathrm{an}^*}$ and $\mathbb{R}_{\mathcal{G}}$ of $\mathbb{R}_{\mathrm{alg}}$ by convergent generalized power series [30] and multisummable series [31]. Here is what we need to know about these structures:

- $\mathbb{R}_{\mathrm{an}^*}$ and $\mathbb{R}_{\mathcal{G}}$ are o-minimal,
- $\langle \mathbb{R}_{\mathrm{an}^*}, \{\exp\}\rangle$ and $\langle \mathbb{R}_{\mathcal{G}}, \{\exp\}\rangle$ are o-minimal expansions of $\mathbb{R}_{\mathrm{an,exp}}$,
- ζ and Γ are, respectively, $\langle \mathbb{R}_{\mathrm{an}^*}, \{\exp\}\rangle$- and $\langle \mathbb{R}_{\mathcal{G}}, \{\exp\}\rangle$-definable,
- the Pfaffian closure of an o-minimal structure $\mathcal{M}$, denoted by $\mathrm{Pf}(\mathcal{M})$, is an o-minimal expansion of $\mathcal{M}$ which is closed under integration of continuous functions of one variable,
- $\mathrm{Pf}(\mathbb{R}_{\mathrm{an}^*})$ and $\mathrm{Pf}(\mathbb{R}_{\mathcal{G}})$ are expansions of $\langle \mathbb{R}_{\mathrm{an}^*}, \{\exp\}\rangle$ and $\langle \mathbb{R}_{\mathcal{G}}, \{\exp\}\rangle$, respectively.

In particular, ζ is $\mathrm{Pf}(\mathbb{R}_{\mathrm{an}^*})$-definable, Γ is $\mathrm{Pf}(\mathbb{R}_{\mathcal{G}})$-definable, and erf is definable in both $\mathrm{Pf}(\mathbb{R}_{\mathrm{an}^*})$ and $\mathrm{Pf}(\mathbb{R}_{\mathcal{G}})$. So the best we can do is to extend $\mathscr{F}$ to classes of o-minimal functions in which erf and either ζ or Γ are confinable.

Given $\sigma \in \{\zeta, \Gamma\}$, we let $\mathscr{F}_\sigma$ be the class of functions whose definition is obtained from that of $\mathscr{F}$ by (i) adding σ to the collection $\mathscr{A}$, (ii) replacing the symbol $\mathscr{F}$ by $\mathscr{F}_\sigma$, and (iii) supplementing (R1)–(R5) with the following rule:

(R6) for any $u \in (\mathcal{I}_f)^\circ$, the antiderivative $t \in (\mathcal{I}_f)^\circ \longmapsto \displaystyle\int_u^t f$ belongs to $\mathscr{F}_\sigma$.

Piecewise $\mathscr{F}_\sigma$-definable functions are defined as in Definition 6.10, with $\mathscr{F}$ replaced by $\mathscr{F}_\sigma$. Since the functions in $\mathscr{F}$ are $\mathbb{R}_{\mathrm{an,exp}}$-definable, piecewise $\mathscr{F}_\zeta$- and $\mathscr{F}_\Gamma$-definable functions are respectively $\mathrm{Pf}(\mathbb{R}_{\mathrm{an}^*})$- and $\mathrm{Pf}(\mathbb{R}_{\mathcal{G}})$-definable. Therefore the proof of Theorem 6.11 remains valid if we replace $\mathscr{F}$ and $\mathbb{R}_{\mathrm{an,exp}}$ by $\mathscr{F}_\zeta$ and $\mathrm{Pf}(\mathbb{R}_{\mathrm{an}^*})$ or by $\mathscr{F}_\Gamma$ and $\mathrm{Pf}(\mathbb{R}_{\mathcal{G}})$. This yields the following result.

Theorem 6.12 (Expansion to special functions)**.** *Let $\sigma \in \{\zeta, \Gamma\}$. In Theorem* 6.11*, the piecewise $\mathscr{F}$-definability condition can be relaxed to piecewise $\mathscr{F}_\sigma$-definability. The objective Θ is then tame in $\mathrm{Pf}(\mathbb{R}_{\mathrm{an}^*})$ if $\sigma=\zeta$ or in $\mathrm{Pf}(\mathbb{R}_{\mathcal{G}})$ if $\sigma=\Gamma$.*

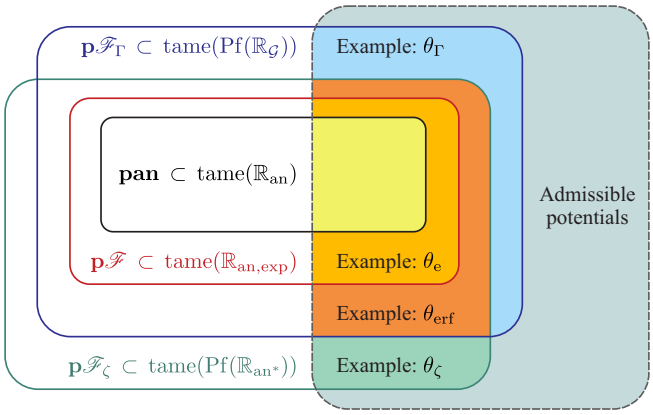


FIGURE 5. Hierarchy of tame potentials. The class **pan** contains all the potentials listed in Appendix A; the examples θ_{e}, θ_{erf}, θ_ζ, and θ_Γ are defined in (6.8) and (6.14)–(6.16). A C^1 objective of the form (1.3) is tame if all its potentials belong to a same class—namely either $\mathbf{p}\mathscr{F}_\zeta$ or $\mathbf{p}\mathscr{F}_\Gamma$.

6.3.3. *Summary.* The combination of Theorems 5.5 and 6.11 guarantees finite-length convergence to a stationary point for a wide class of admissible potentials. This class consists of tame functions defined by simple rules and can be extended to include special functions (Theorem 6.12). Two mild restrictions must be kept in mind. First, as there is no known o-minimal structure expanding both $\mathrm{Pf}(\mathbb{R}_{\mathrm{an}^*})$ and $\mathrm{Pf}(\mathbb{R}_\mathcal{G})$, infinite branches of o-minimal special functions cannot be used freely; in particular, tameness is not guaranteed if the asymptotics of ζ and Γ at $+\infty$ are confined simultaneously (this rules out, for example, objectives containing both the potential functions θ_ζ and θ_Γ defined in (6.15) and (6.16)). Second, tameness precludes potentials whose definition involves—either directly or after differentiation(s)—a function with a countably infinite set of zeros in a finite interval (examples of such potentials are given by (5.13) and (6.2)). Nontame potentials, however, are unlikely to be needed in practice; but if this happens, we can always fall back on the global convergence properties stated in Theorem 4.9.

To conclude this section, Figure 5 shows the hierarchy of tame potentials and their associated o-minimal structures. The notation $\mathrm{tame}(\mathcal{M})$ stands for the class of functions on $\mathbb{R}_+$ that are tame in $\mathcal{M}$ (Definition 6.1), **pan** is the class of piecewise analytic functions on $\mathbb{R}_+$ (Definition 6.7), and $\mathbf{p}\mathscr{F}$ denotes the class of functions that are piecewise $\mathscr{F}$-definable on $(0,\rho)$ for some $\rho \in (0,+\infty]$ (Definition 6.10) and restricted analytic on $[\rho,+\infty)$ if $\rho < +\infty$.

7. Implementation using the conjugate gradient method

There is freedom in implementing the inexact FHQ algorithm, since the accuracy parameter μ can take any value in $(0,1)$ (recall that the smaller μ, the closer to the behavior of the exact FHQ algorithm). Nevertheless, we must ensure that a generated sequence $(\boldsymbol{x}^{(p)})_p$ is a true inexact FHQ sequence, namely, that it satisfies

$$\sup_{p\in\mathbb{N}} \inf\left\{\mu \in (0,1) : \boldsymbol{x}^{(p+1)} \in \boldsymbol{\Phi}_\mu(\boldsymbol{x}^{(p)})\right\} < 1. \tag{7.1}$$

We first show in Section 7.1 that inexact FHQ sequences can be generated by using the CG method to compute each iterate. Then, in Section 7.2, we give a version of the CG method tailored to inexact FHQ optimization, and we discuss two CG termination strategies. We conclude in Section 7.3 with the main stopping criterion.

7.1. **Why use CG?** We begin with an alternative definition of $\boldsymbol{\Phi}_\mu$ in terms of an upper bound on the energy norm of the error for a linear system.

Proposition 7.1 (Reformulation of the inexact FHQ map)**.** *Let $\Theta \in \mathscr{C}_{\text{obj}}(\mathbb{R}^n)$ and $\mu \in (0,1)$. For all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,*

$$\boldsymbol{y} \in \boldsymbol{\Phi}_\mu(\boldsymbol{x}) \iff \|\boldsymbol{y} - \boldsymbol{y}^*\|_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}} \leqslant \sqrt{\mu}\|\boldsymbol{x} - \boldsymbol{y}^*\|_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}}, \tag{7.2}$$

where $\boldsymbol{y}^* := \boldsymbol{\Phi}_0(\boldsymbol{x})$ is the unique solution of the linear system

$$\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}\boldsymbol{z} = \boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{a}. \tag{7.3}$$

Proof. Let $\boldsymbol{x} \in \mathbb{R}^n$. By definition, $\boldsymbol{\Phi}_0(\boldsymbol{x})$ is the global minimizer of the positive definite quadratic function $\overline{\Theta}(\boldsymbol{x}, \cdot)$ (see (3.4)), whose gradient is

$$\nabla\overline{\Theta}(\boldsymbol{x}, \cdot)(\boldsymbol{z}) = \boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})(\boldsymbol{A}\boldsymbol{z} - \boldsymbol{a}).$$

So $\boldsymbol{\Phi}_0(\boldsymbol{x})$ is the unique solution of the linear system (7.3). Let $\boldsymbol{y} \in \mathbb{R}^n$. Using (4.2) and (4.3), we have

$$\begin{aligned}
\|\boldsymbol{y} - \boldsymbol{y}^*\|^2_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}} &= \|\boldsymbol{y} - \boldsymbol{x} + (\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A})^{-1}\nabla\Theta(\boldsymbol{x})\|^2_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}} \\
&= \|\boldsymbol{y} - \boldsymbol{x}\|^2_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}} + 2(\boldsymbol{y} - \boldsymbol{x})^T\nabla\Theta(\boldsymbol{x}) + \|\nabla\Theta(\boldsymbol{x})\|^2_{(\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A})^{-1}} \\
&= 2\big(\overline{\Theta}(\boldsymbol{x}, \boldsymbol{y}) - \Theta(\boldsymbol{x})\big) + 2\big(\Theta(\boldsymbol{x}) - \min\overline{\Theta}(\boldsymbol{x}, \cdot)\big) \\
&= 2\big(\overline{\Theta}(\boldsymbol{x}, \boldsymbol{y}) - \min\overline{\Theta}(\boldsymbol{x}, \cdot)\big).
\end{aligned}$$

Hence

$$\begin{aligned}
\|\boldsymbol{y} - \boldsymbol{y}^*\|^2_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}} &\leqslant \mu\|\boldsymbol{x} - \boldsymbol{y}^*\|^2_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}} \\
&\iff \overline{\Theta}(\boldsymbol{x}, \boldsymbol{y}) - \min\overline{\Theta}(\boldsymbol{x}, \cdot) \leqslant \mu\big(\overline{\Theta}(\boldsymbol{x}, \boldsymbol{x}) - \min\overline{\Theta}(\boldsymbol{x}, \cdot)\big) \\
&\iff \overline{\Theta}(\boldsymbol{x}, \boldsymbol{y}) - \min\overline{\Theta}(\boldsymbol{x}, \cdot) \leqslant h_\mu(\boldsymbol{x}) \\
&\iff \boldsymbol{y} \in \boldsymbol{\Phi}_\mu(\boldsymbol{x}),
\end{aligned}$$

which completes the proof. □

From Proposition 7.1, the $(p+1)$th iteration of the inexact FHQ algorithm amounts to solving the following inner optimization problem: given $\boldsymbol{x} := \boldsymbol{x}^{(p)}$, find an approximate solution $\boldsymbol{y}$ of the system (7.3) such that the energy norm of the error is bounded above as in (7.2), where μ is independent of p. Since the $n \times n$ matrix $\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}$ is symmetric positive definite, it is natural to use the CG method when n is large, as is typically the case for inverse imaging problems.

We assume symmetric preconditioning: given a nonsingular matrix $\boldsymbol{U}$ and letting $\boldsymbol{z}' := \boldsymbol{U}\boldsymbol{z}$, the system (7.3) is equivalent to

$$\underbrace{\boldsymbol{U}^{-T}\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}\boldsymbol{U}^{-1}}_{=:\,\boldsymbol{B}(\boldsymbol{x})}\boldsymbol{z}' = \boldsymbol{U}^{-T}\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{a}. \tag{7.4}$$

Let $\boldsymbol{y}^{(j)}$ be the jth iterate of the preconditioned CG algorithm for (7.3) starting from $\boldsymbol{y}^{(0)} = \boldsymbol{x}$ (so $\boldsymbol{U}\boldsymbol{y}^{(j)}$ is the jth iterate of unpreconditioned CG for (7.4) starting from $\boldsymbol{U}\boldsymbol{x}$). Let $\kappa(\boldsymbol{x})$ denote the Euclidean condition number of the preconditioned

system matrix $\boldsymbol{B}(\boldsymbol{x})$, that is, $\kappa(\boldsymbol{x}) := \lambda'_n(\boldsymbol{x})(\lambda'_1(\boldsymbol{x}))^{-1}$, where $\lambda'_n(\boldsymbol{x})$ and $\lambda'_1(\boldsymbol{x})$ are the largest and smallest eigenvalues of $\boldsymbol{B}(\boldsymbol{x})$, respectively. Using the fact that $\|\boldsymbol{U}\boldsymbol{z}\|_{\boldsymbol{B}(\boldsymbol{x})} = \|\boldsymbol{z}\|_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}}$, we have the following bound on the energy norm of the error (see, e.g., [32, Theorem 9.4.12]): for every $j \in \mathbb{N}$,

$$\|\boldsymbol{y}^{(j)} - \boldsymbol{y}^*\|_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}} \leqslant \omega_j \circ \Lambda(\boldsymbol{x})\, \|\boldsymbol{x} - \boldsymbol{y}^*\|_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}}, \tag{7.5a}$$

where the functions $\Lambda : \mathbb{R}^n \longrightarrow [0,1)$ and $\omega_j : [0,1) \longrightarrow [0,1)$ are defined by

$$\Lambda(\boldsymbol{x}) := \big(\sqrt{\kappa(\boldsymbol{x})} - 1\big)\big(\sqrt{\kappa(\boldsymbol{x})} + 1\big)^{-1} \quad \text{and} \quad \omega_j(t) := 2t^j(1+t^{2j})^{-1}. \tag{7.5b}$$

The quantity $\omega_j \circ \Lambda(\boldsymbol{x})$ increases with increasing $\Lambda(\boldsymbol{x})$ and decreases with increasing j. Therefore, if $\kappa(\boldsymbol{x})$ is bounded, inexact FHQ sequences can be generated by solving the inner optimization problems using CG with a fixed minimum number of iterations. This leads to the following theorem.

Theorem 7.2 (Inexact FHQ sequences via truncated CG)**.** *Let $\Theta \in \mathscr{C}_{\mathrm{obj}}(\mathbb{R}^n)$, let J be a positive integer, and let $\mathcal{X} := (\boldsymbol{x}^{(p)})_{p\in\mathbb{N}}$ be constructed as follows:*

(i) *the starting point $\boldsymbol{x}^{(0)}$ is in a generalized cup of Θ,*

(ii) *for every $p \in \mathbb{N}$, the iterate $\boldsymbol{x}^{(p+1)}$ is an approximate solution of the linear system $\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x}^{(p)})\boldsymbol{A}\boldsymbol{z} = \boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x}^{(p)})\boldsymbol{a}$ obtained in J or more CG iterations starting from $\boldsymbol{x}^{(p)}$.*

Then there exists $\mu \in (0,1)$ such that $\mathcal{X}$ is generated by $\boldsymbol{\Phi}_\mu$.

Proof. Let $\mathcal{G}$ be a generalized cup of Θ containing $\boldsymbol{x}^{(0)}$. Since $\overline{\mathcal{G}}$ is compact and $\lambda'_1(\boldsymbol{x}) > 0$ for all $\boldsymbol{x}$, the condition number $\kappa(\boldsymbol{x})$ is bounded on $\overline{\mathcal{G}}$, and hence

$$\Lambda(\overline{\mathcal{G}}) := \sup_{\boldsymbol{x}\in\overline{\mathcal{G}}} \Lambda(\boldsymbol{x}) < 1.$$

Let $\boldsymbol{x} \in \overline{\mathcal{G}}$ and $j \in [J\,..\,\infty)$. From (7.5) we deduce that

$$\|\boldsymbol{y}^{(j)} - \boldsymbol{y}^*\|_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}} \leqslant \omega_J(\Lambda(\overline{\mathcal{G}}))\, \|\boldsymbol{x} - \boldsymbol{y}^*\|_{\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}},$$

and so, in view of Proposition 7.1,

$$\boldsymbol{y}^{(j)} \in \boldsymbol{\Phi}_{\mu_J}(\boldsymbol{x}), \qquad \mu_J := (\omega_J(\Lambda(\overline{\mathcal{G}})))^2. \tag{7.6}$$

Since $\boldsymbol{\Phi}_{\mu_J}(\mathcal{G}) \subseteq \mathcal{G}$ (see the proof of Lemma 4.5(i)), it follows by induction that $\boldsymbol{x}^{(p+1)} \in \boldsymbol{\Phi}_{\mu_J}(\boldsymbol{x}^{(p)})$ for all $p \in \mathbb{N}$. □

7.2. **The inner CG algorithm.** Let $x \in \mathbb{R}^n$ and define

$$\boldsymbol{A}(\boldsymbol{x}) := \boldsymbol{E}(\boldsymbol{x})^{1/2}\boldsymbol{A} \quad \text{and} \quad \boldsymbol{a}(\boldsymbol{x}) := \boldsymbol{E}(\boldsymbol{x})^{1/2}\boldsymbol{a}. \tag{7.7}$$

The system (7.3) can be written as

$$\boldsymbol{A}(\boldsymbol{x})^T\boldsymbol{A}(\boldsymbol{x})\boldsymbol{z} = \boldsymbol{A}(\boldsymbol{x})^T\boldsymbol{a}(\boldsymbol{x}), \tag{7.8}$$

which are the normal equations for minimizing $\|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{z} - \boldsymbol{a}(\boldsymbol{x})\|$, $\boldsymbol{z} \in \mathbb{R}^n$. This suggests to use the CG method for least squares problems, or CGLS [33]; we recall below the preconditioned version of this algorithm.

Algorithm. Symmetrically preconditioned CGLS applied to (7.8)

Given $\boldsymbol{y}^{(0)} \in \mathbb{R}^n$, set $\boldsymbol{r}^{(0)} := \boldsymbol{A}(\boldsymbol{x})^T(\boldsymbol{a}(\boldsymbol{x}) - \boldsymbol{A}(\boldsymbol{x})\boldsymbol{y}^{(0)})$,
$\boldsymbol{s}^{(0)} := \boldsymbol{U}^{-1}\boldsymbol{U}^{-T}\boldsymbol{r}^{(0)}$, $\boldsymbol{p}^{(0)} := \boldsymbol{s}^{(0)}$, $\tau_0 := (\boldsymbol{r}^{(0)})^T\boldsymbol{s}^{(0)}$
for $j = 0, 1, 2, \ldots$ **do**
  $\boldsymbol{q}^{(j)} := \boldsymbol{A}(\boldsymbol{x})\boldsymbol{p}^{(j)}$
  $\alpha_j := \tau_j / \|\boldsymbol{q}^{(j)}\|^2$
  $\boldsymbol{y}^{(j+1)} := \boldsymbol{y}^{(j)} + \alpha_j \boldsymbol{p}^{(j)}$
  $\boldsymbol{r}^{(j+1)} := \boldsymbol{r}^{(j)} - \alpha_j \boldsymbol{A}(\boldsymbol{x})^T\boldsymbol{q}^{(j)}$
  $\boldsymbol{s}^{(j+1)} := \boldsymbol{U}^{-1}\boldsymbol{U}^{-T}\boldsymbol{r}^{(j+1)}$
  $\tau_{j+1} := (\boldsymbol{r}^{(j+1)})^T\boldsymbol{s}^{(j+1)}$
  $\beta_j := \tau_{j+1}/\tau_j$
  $\boldsymbol{p}^{(j+1)} := \boldsymbol{s}^{(j+1)} + \beta_j \boldsymbol{p}^{(j)}$
end for

Theorem 7.2 leaves us the freedom to choose when to stop the CG algorithm beyond the minimum number of iterations J. We describe below two strategies to control, respectively, the backward error of the computed solution and the accuracy of the inexact FHQ map. We then briefly discuss the behavior in finite precision arithmetic.

7.2.1. *Control of the backward error.* Backward error measures are often used in stopping criteria of iterative linear system solvers [34, 35]. In the case of CG, the preference is for the *normwise relative backward error* [36] whose definition follows. Let $\boldsymbol{y}$ be an approximate solution to a square, nonsingular linear system $\boldsymbol{B}\boldsymbol{z} = \boldsymbol{b}$. The normwise relative backward error of $\boldsymbol{y}$ is

$$\begin{aligned}\nu(\boldsymbol{y}) := \min\big\{ c \in \mathbb{R}_+ :\ & (\boldsymbol{B} + \Delta\boldsymbol{B})\boldsymbol{y} = \boldsymbol{b} + \Delta\boldsymbol{b}, \\ & \|\Delta\boldsymbol{B}\| \leqslant c\|\boldsymbol{B}\|,\ \|\Delta\boldsymbol{b}\| \leqslant c\|\boldsymbol{b}\| \big\},\end{aligned} \tag{7.9}$$

where the vector and matrix norms are compatible in the sense that $\|\boldsymbol{M}\boldsymbol{z}\| \leqslant \|\boldsymbol{M}\|\,\|\boldsymbol{z}\|$ for all matrices $\boldsymbol{M}$ and vectors $\boldsymbol{z}$ (the usual choice is the ℓ_2-norm, which is compatible with the Frobenius and spectral norms). In other words, $\nu(\boldsymbol{y})$ is the minimum relative amount by which $\boldsymbol{B}$ and $\boldsymbol{b}$ must be changed so that $\boldsymbol{y}$ is the exact solution of the perturbed system $(\boldsymbol{B} + \Delta\boldsymbol{B})\boldsymbol{z} = \boldsymbol{b} + \Delta\boldsymbol{b}$.

As shown in [37],

$$\nu(\boldsymbol{y}) \;=\; \|\boldsymbol{B}\boldsymbol{y} - \boldsymbol{b}\|\big(\|\boldsymbol{B}\|\,\|\boldsymbol{y}\| + \|\boldsymbol{b}\|\big)^{-1}. \tag{7.10}$$

Hence, by setting an upper threshold ϵ on $\nu(\boldsymbol{y})$ and identifying $\boldsymbol{B}\boldsymbol{z} = \boldsymbol{b}$ with (7.8) and $\boldsymbol{y}$ with the jth CG iterate, we obtain the stopping criterion

$$\|\boldsymbol{r}^{(j)}\| \;\leqslant\; \epsilon\big(\|\boldsymbol{A}(\boldsymbol{x})^T\boldsymbol{A}(\boldsymbol{x})\|\,\|\boldsymbol{y}^{(j)}\| + \|\boldsymbol{A}(\boldsymbol{x})^T\boldsymbol{a}(\boldsymbol{x})\|\big), \tag{7.11}$$

where $\boldsymbol{r}^{(j)} := \boldsymbol{A}(\boldsymbol{x})^T(\boldsymbol{a}(\boldsymbol{x}) - \boldsymbol{A}(\boldsymbol{x})\boldsymbol{y}^{(j)})$ is the residual at the jth iteration. The implementation of this criterion requires the additional computation of $\|\boldsymbol{A}(\boldsymbol{x})^T\boldsymbol{A}(\boldsymbol{x})\|$ and $\|\boldsymbol{A}(\boldsymbol{x})^T\boldsymbol{a}(\boldsymbol{x})\|$ in the initialization step and of $\|\boldsymbol{y}^{(j)}\|$ at each iteration.

7.2.2. *Control of the accuracy.* Under the assumptions of Theorem 7.2, the accuracy μ is smaller than the quantity μ_J defined in (7.6), but this bound is overly pessimistic and cannot even be estimated in practice. We propose here a stopping

criterion which, unlike (7.11), allows to control the value of μ and involves only quantities that are available in the CG algorithm.

For every integers j_1 and j_2 such that $0 \leqslant j_1 < j_2$, define $\mathfrak{E}_{j_1,j_2}$ by

$$\mathfrak{E}_{j_1,j_2}^{\,2} := \sum_{i \in [j_1 \,..\, j_2-1]} \alpha_i \tau_i \,, \tag{7.12}$$

where α_i is the step length in the ith conjugate direction and $\tau_i := (\boldsymbol{r}^{(i)})^T \boldsymbol{s}^{(i)} = \|\boldsymbol{U}^{-T}\boldsymbol{r}^{(i)}\|^2$. This quantity is of particular interest because it gives the difference between the squares of the j_1th and j_2th error energy norms [38, Section 3.1]:

$$\mathfrak{E}_{j_1,j_2}^{\,2} \,=\, \mathfrak{E}_{j_1}^{\,2} - \mathfrak{E}_{j_2}^{\,2}, \qquad \mathfrak{E}_j \,:=\, \|\boldsymbol{A}(\boldsymbol{x})(\boldsymbol{y}^{(j)} - \boldsymbol{y}^*)\|. \tag{7.13}$$

Suppose the CG algorithm starts from $\boldsymbol{y}^{(0)} = \boldsymbol{x}$. Setting $\boldsymbol{y} = \boldsymbol{y}^{(j)}$ in (7.2) and $(j_1, j_2) = (0, j)$ in (7.13), we deduce that

$$\boldsymbol{y}^{(j)} \in \boldsymbol{\Phi}_\mu(\boldsymbol{x}) \iff \mathfrak{E}_j^{\,2} \leqslant \mu(1-\mu)^{-1}\,\mathfrak{E}_{0,j}^{\,2}\,. \tag{7.14}$$

Let $d \in [1\,..\,\infty)$. If the error energy norm decreases sufficiently between iterations j and $j+d$, so that $\mathfrak{E}_{j+d}^{\,2} \ll \mathfrak{E}_j^{\,2}$, then it follows from (7.13) that $\mathfrak{E}_{j,j+d}$ is a tight lower bound on $\mathfrak{E}_j$. Using this bound as an approximation to $\mathfrak{E}_j$ in (7.14), we obtain the stopping criterion

$$\begin{aligned} \mathfrak{E}_{j,j+d}^{\,2} \;&\leqslant\; \mu\big(\mathfrak{E}_{0,j}^{\,2} + \mathfrak{E}_{j,j+d}^{\,2}\big) \\ &\iff \sum_{i \in [j \,..\, j+d-1]} \alpha_i \tau_i \;\leqslant\; \mu \sum_{i \in [0 \,..\, j+d-1]} \alpha_i \tau_i \,. \end{aligned} \tag{7.15}$$

If d is large enough, this simple test ensures that any sequence constructed as in Theorem 7.2 is generated by $\boldsymbol{\Phi}_{\mu'}$ with $\mu' \approx \mu$. Moreover, since the evaluation of the sums in (7.15) requires running the CG algorithm for $j+d$ iterations, it is natural to set the outer iterate $\boldsymbol{x}^{(p+1)}$ to $\boldsymbol{y}^{(j+d)}$ (rather than $\boldsymbol{y}^{(j)}$) when exiting the CG loop, so the assertion that $\mathcal{X}$ is generated with an accuracy close to μ is all the more true. Then, since (7.15) is not satisfied for $j = 0$, the minimum number of iterations J is implicitly set to $d+1$ (so $j \in [0\,..\,d]$ in the CG loop). Alternatively, we can set J independently of d to act as a safeguard when d is too small for $\mathfrak{E}_{j,j+d}$ to be a good estimate of $\mathfrak{E}_j$.

Remark 9 (Gauss-Radau quadrature). The relation (7.13)—originally proposed by Hestenes and Stiefel in their seminal paper on the CG method [39, Equation (6:2)]—can be recovered from the connection between CG and the Gauss quadrature approximation of an underlying Riemann-Stieltjes integral (see, e.g., [40]). This connection can also be exploited to derive an upper bound on $\mathfrak{E}_j$ using the so-called Gauss-Radau quadrature rule [40, 41], which is useful if needed to ensure that $\mathcal{X}$ is generated by $\boldsymbol{\Phi}_\mu$ for a precise value of μ. However, we did not pursue this direction for two reasons. First, it is sufficient in practice to guarantee that $\mathcal{X}$ is generated with an accuracy reasonably close to a prescribed value. Second, the Gauss-Radau rule is computationally expensive because it requires a tight lower bound on the smallest eigenvalue of the preconditioned system matrix at every outer iteration.

7.2.3. *Behavior in finite precision arithmetic.* It is shown in [42] and verified experimentally in [43] that finite precision CG for solving a linear system $\boldsymbol{B}\boldsymbol{z} = \boldsymbol{b}$ is equivalent to exact precision CG applied to a larger system whose eigenvalues lie in small intervals around those of $\boldsymbol{B}$. It follows that the exact arithmetic error bound (7.5) (which is defined solely in terms of the extreme eigenvalues of the preconditioned system matrix) holds to a close approximation for finite precision computations, implying that rounding errors do not affect Theorem 7.2. The same goes for the proposed inner termination strategies. Indeed, the backward error (7.10) can be computed with negligible rounding error, and the estimate $\mathfrak{E}_{j,j+d}$ (see (7.12)) of the jth error energy norm is stable in finite precision arithmetic [38]. So both the stopping criteria (7.11) and (7.15) can be used safely.

7.3. **Termination of the outer iterations.** Three standard criteria can be used to test the convergence of the inexact FHQ algorithm: the iterate-based criterion

$$\|\boldsymbol{x}^{(p+1)} - \boldsymbol{x}^{(p)}\| \leqslant \breve{\epsilon} \max\big\{1, \|\boldsymbol{x}^{(p)}\|\big\}, \tag{7.16}$$

the objective-based criterion

$$\Theta(\boldsymbol{x}^{(p)}) - \Theta(\boldsymbol{x}^{(p+1)}) \leqslant \breve{\epsilon} \max\big\{1, |\Theta(\boldsymbol{x}^{(p)})|\big\}, \tag{7.17}$$

and the gradient-based criterion

$$\|\nabla\Theta(\boldsymbol{x}^{(p)})\| \leqslant \breve{\epsilon} \max\big\{1, |\Theta(\boldsymbol{x}^{(p)})|\big\}, \tag{7.18}$$

where $\breve{\epsilon}$ is a given tolerance and the $\max\{1, \cdot\}$ function prevents the right-hand side from becoming too small in case $\|\boldsymbol{x}^{(p)}\|$ or $\Theta(\boldsymbol{x}^{(p)})$ gets close to zero. (The use of the objective function in the right-hand side of (7.18) ensures scale independence, as multiplying Θ by a constant does not change the inequality $\|\nabla\Theta\| \leqslant \breve{\epsilon}|\Theta|$.) All three criteria are appropriate in light of Theorems 4.7(iii)–(iv) and 4.9(iii). But since we are primarily interested in finding stationary points, the gradient-based criterion is the wisest choice.

8. Experiments

In this section we illustrate the behavior of the CG-based implementation of the inexact FHQ algorithm on three inverse problems. In each case, the observation model is of the form (1.1) and we look for minimizers of a nonconvex objective $\Theta \in \mathscr{C}_{\mathrm{obj}}$ with piecewise analytic potentials. We distinguish two versions of the algorithm: the *backward error* FHQ algorithm, which uses the stopping criterion (7.11) to control the normwise relative backward error of the CG iterates, and the *Gauss quadrature* FHQ algorithm, which controls the accuracy μ via the criterion (7.15) (the reason we use the term "Gauss quadrature" is given in Remark 9). Both generate inexact FHQ sequences as in Theorem 7.2 and thus converge to a stationary point of Θ by Theorems 5.5 and 6.8.

After some preliminary remarks, we begin in Section 8.2 by examining the stability of the Gauss quadrature algorithm, first with the gradient-based outer termination criterion (7.18) and then in the long run. For this purpose, we consider a limited-data tomography problem with a standard objective combining a quadratic data-fidelity term with a regularizer operating on the spatial gradient. The behavior of the Gauss quadrature algorithm is further investigated in Section 8.3, where we look at the effect of the outer termination tolerance and of preconditioning; here the focus is on image deblurring under mixed Gaussian-impulse noise, and

the objective involves a non-quadratic data-fidelity term and a tight-frame regularizer. Finally, we compare the backward error and Gauss quadrature algorithms in Section 8.4, where we consider an image inpainting problem and a regularizer operating in the domain of a patch-based sparsifying transform.

8.1. **Preliminary remarks.** The backward error and Gauss quadrature algorithms always start from $\boldsymbol{x}^{(0)} = \boldsymbol{0}$ (the zero image), and the default choice for preconditioning the inner systems is the Jacobi preconditioner: the matrix $\boldsymbol{U}$ in (7.4) is

$$(\mathrm{diag}(\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}))^{1/2} \;=\; \mathrm{diag}\big(\|\boldsymbol{A}(:,1)\|_{\boldsymbol{E}(\boldsymbol{x})}, \ldots, \|\boldsymbol{A}(:,n)\|_{\boldsymbol{E}(\boldsymbol{x})}\big),$$

where $\boldsymbol{A}(:,k)$ denotes the kth column of $\boldsymbol{A}$. In Section 8.3, we also use incomplete Cholesky (IC) preconditioning, for which $\boldsymbol{U}$ is the upper triangular IC factor of the inner system matrix.

The quality of the solutions is assessed with the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index. The PSNR (in dB) of a computed solution $\boldsymbol{x}$ is defined by

$$\mathrm{PSNR} \;:=\; 20\log_{10}\!\left(\sqrt{n}\,\frac{\max_k[\boldsymbol{x}^\bullet]_k - \min_k[\boldsymbol{x}^\bullet]_k}{\|\boldsymbol{x}-\boldsymbol{x}^\bullet\|}\right),$$

where $\boldsymbol{x}^\bullet$ denotes the original image, and the SSIM index is defined in [44] (we use the MATLAB implementation available at `http://ece.uwaterloo.ca/~z70wang/research/ssim` with default settings).

In the long run, we aim to reach a *stationary point to machine precision*, that is, a point $\boldsymbol{x}$ satisfying $\|\nabla\Theta(\boldsymbol{x})\| \leqslant \|\nabla\Theta(\boldsymbol{x}) - \nabla\Theta(\boldsymbol{y})\|$ for any $\boldsymbol{y}$ with entries $[\boldsymbol{y}]_k$ such that $\{[\boldsymbol{x}]_k, [\boldsymbol{y}]_k\}$ is a two-element set of consecutive double-precision floating-point numbers. For simplicity, we say that machine precision is reached when the number of outer iterations p is large enough so that $\boldsymbol{x}^{(p)}$ and all subsequent iterates are stationary to machine precision.

8.2. **Reconstruction from limited projection data.** We consider the problem of reconstructing the modified 256×256 Shepp-Logan phantom shown in Figure 6(a) from the limited parallel-beam projection data displayed in Figure 6(b). These data consist of 90 evenly spaced projections over 180°, each with four gaps of width twice the detector size and poisson noise corresponding to 2×10^4 incident photons per detector. The observation matrix $\boldsymbol{D}$ is of size $27000\times n$, $n = 256^2$, and its density (namely, the percentage ratio of its number of nonzero entries to its total number of entries) is 0.69%.

We look for solutions that minimize the objective $\Theta : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by

$$\Theta(\boldsymbol{x}) \;:=\; \|\boldsymbol{D}\boldsymbol{x}-\boldsymbol{d}\|^2 \;+\; \gamma\sum_l \vartheta_\delta\big(\|\boldsymbol{G}_l\boldsymbol{x}\|\big), \qquad \vartheta_\delta(t) \;:=\; (\delta^2+t^2)^{1/4} - \sqrt{\delta}, \tag{8.1}$$

where $\gamma > 0$ controls the regularization strength, $\delta > 0$ controls how well the admissible potential ϑ_δ approximates the square root function (the smaller the value of δ, the better the approximation), and $\{\boldsymbol{G}_l \in \mathbb{R}^{2\times n}\}_l$ is the usual spatial-gradient operator: l indexes the pixels in $\boldsymbol{x}$ and the two components of $\boldsymbol{G}_l\boldsymbol{x}$ are the horizontal and vertical differences at the lth pixel location (see, e.g., [2, Equation (5.2)]). The sum $\sum_l \vartheta_\delta(\|\boldsymbol{G}_l\boldsymbol{x}\|)$ approximates the square root of the $\ell_{1/2}$-norm of the vector with entries $\|\boldsymbol{G}_l\boldsymbol{x}\|$ and hence promotes sparsity of the spatial gradient. Looked at another way, since ϑ_δ is strictly concave on $(\sqrt{2}\delta, +\infty)$, the regularizer preserves spatial-gradient magnitudes greater than $\sqrt{2}\delta\,\mathrm{cm}^{-1}$. We set $\delta = 10^{-3}$,

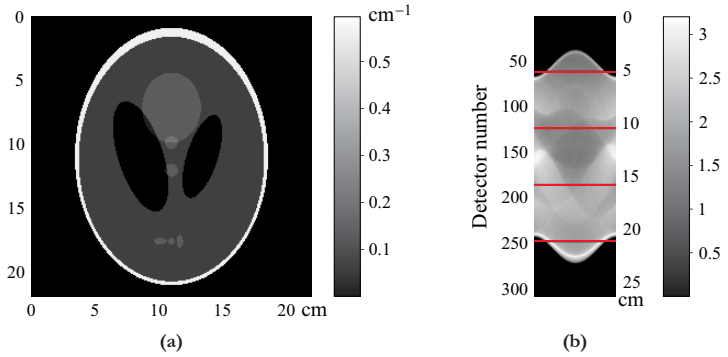


FIGURE 6. Reconstruction problem: (a) modified 256×256 Shepp-Logan phantom; (b) limited parallel-beam projection data (the detector size matches that of the pixels in the phantom and the red lines indicate gaps due to defective detectors).

so $\sqrt{2}\delta$ is much smaller than the smallest nonzero value of the spatial-gradient magnitude in the phantom:

$$\sqrt{2}\delta \approx 1.41 \times 10^{-3}\,\mathrm{cm}^{-1} \ll \min_{l\,:\,\boldsymbol{G}_l \boldsymbol{x}^\bullet \neq \boldsymbol{0}} \|\boldsymbol{G}_l \boldsymbol{x}^\bullet\| = 5.55 \times 10^{-2}\,\mathrm{cm}^{-1}.$$

The choice of the strength γ is beyond the scope of this paper; we simply adjust it to obtain the highest PSNR (about 48 dB, achieved for $\gamma = 0.016$).

The intersection of the null spaces of the matrices $\boldsymbol{G}_l$ is the set of constant images, and the only constant image in the null space of $\boldsymbol{D}$ is the zero image. Therefore Θ is coercive and Theorem 5.5 holds independently of the starting point (see Remarks 2 and 3). However, to avoid being trapped in a shallow basin of attraction, we guide the first iterations of the inexact FHQ algorithm by replacing ϑ_δ with admissible potentials $\vartheta^{(p)}$ such that the optimization difficulty increases with p. We limit this continuation process to a fixed number of iterations so as not to question our convergence results (see [1] for motivation and details). Let id_δ be the admissible approximation to the identity function defined by

$$\mathrm{id}_\delta(t) := (\delta^2 + t^2)^{1/2} - \delta.$$

We use a continuation sequence $(\vartheta^{(p)})_{p \in [0..50]}$ of the form

$$\vartheta^{(p)} := \kappa_p \vartheta_\delta + (1 - \kappa_p)\mathrm{id}_{\delta_p}, \tag{8.2}$$

where $(\kappa_p)_{p \in [0..50]}$ increases linearly from 0 to 1 and $(\delta_p)_{p \in [0..50]}$ decreases linearly from 100δ to δ. Examples of these potentials are shown in Figure 7.

We first examine the behavior of the Gauss quadrature FHQ algorithm combined with the gradient-based outer termination criterion (7.18). Table 1 gives the total number of outer iterations (denoted by $\mathfrak{N}$), the total number of CG iterations (denoted by $\mathfrak{N}_{\mathrm{CG}}$), and the PSNR of the computed solutions for practical values of the inner control parameters μ and d and for the outer termination tolerance $\breve{\epsilon} = 10^{-6}$. The number $\mathfrak{N}_{\mathrm{CG}}$ is given by

$$\mathfrak{N}_{\mathrm{CG}} := \sum_{p \in [1..\mathfrak{N}]} (d + j_p), \tag{8.3}$$

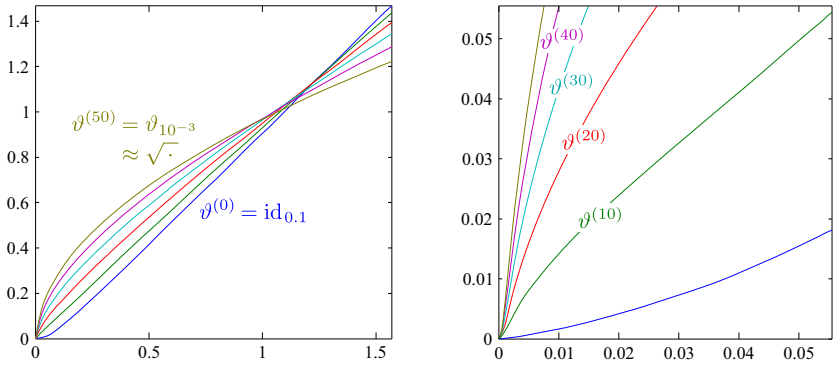


FIGURE 7. Continuation sequence (8.2) for the square root approximation $\vartheta_{10^{-3}}$ (the maximum abscissa of the left plot is twice the maximum value of the spatial-gradient magnitude in the phantom).

where j_p is the smallest integer $j \geqslant 1$ satisfying the CG stopping criterion (7.15) at the pth outer iteration. The algorithm consistently produces high-quality reconstructions: the PSNR is always greater than 48 dB and is constant to five significant digits for $\mu \leqslant 10^{-5}$. As an example, Figure 8 shows the reconstruction obtained for $(\mu, d) = (10^{-3}, 4)$ together with the highest-PSNR solution found using TV regularization (that is, by replacing ϑ_δ with id_δ in (8.1)). The PSNR of the convex reconstruction is 11 dB smaller than that of the nonconvex reconstruction, the differences between the two being particularly pronounced at edge locations, where the magnitude of the spatial gradient is large. Another indicator of good behavior is the stability of the total number of outer iterations when the inner solutions are sufficiently accurate: $\mathfrak{N}$ varies little for $\mu \leqslant 10^{-3}$ and is constant for $\mu \leqslant 10^{-6}$. We do not report the computation time, which is machine-dependent, but there is a nearly affine relationship between the total run time and the total number of CG iterations; moreover, $\mathfrak{N}_{\mathrm{CG}}$ is eventually affine in both $-\log_{10}\mu$ and d (we will call the quantity $-\log_{10}\mu$ the *log-accuracy*). Note also that the Gauss quadrature FHQ algorithm performs well even for "unwise" choices of μ and d. For example, for $(\mu, d) = (0.9, 1)$, it terminates in 886 outer iterations at a PSNR of 47.136 dB.

We now examine the long-run convergence. Figure 9 plots the norm of the gradient $\|\nabla\Theta(\boldsymbol{x}^{(p)})\|$, the objective $\Theta(\boldsymbol{x}^{(p)})$, and the PSNR over 1000 outer iterations for $d = 4$ and log-accuracies between 1 and 7. For each accuracy level, the norm of the gradient eventually decreases to a steady value slightly below 10^{-12}, a plateau that occurs when $\boldsymbol{x}^{(p)}$ is a stationary point to machine precision. Because of the nonconvexity of the regularizer, the objective is riddled with shallow local minima[2]. This explains the slight differences between the computed solutions, as well as why two inexact FHQ sequences with the same starting point but different accuracies can follow different paths in the objective landscape. That said, the trajectory of the iterates hardly changes when μ is decreased below 10^{-2}, so the magenta curve corresponding to $\mu = 10^{-7}$ represents the behavior of the exact FHQ algorithm for the 550 or so first iterations (before machine precision is reached). The objective

[2]Note, however, that nonconvex regularization does not necessarily yields nonconvex objectives, even when the data-fidelity term is quadratic (see, e.g., [2, Appendix B]).

TABLE 1. Behavior of the Gauss-quadrature FHQ algorithm as a function of μ and d for an outer termination tolerance of 10^{-6}. (The PSNR is rounded to five significant digits.)

		μ					
		10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
		Total number of outer iterations ($\mathfrak{N}$)					
	2	653	233	290	290	290	289
	4	383	287	291	290	289	289
d	8	242	290	291	298	289	289
	16	297	298	289	289	289	289
	32	289	289	289	289	289	289
		Total number of CG iterations ($\mathfrak{N}_{\mathrm{CG}}$)					
	2	2749	2285	4787	7268	9925	12607
	4	2499	3532	5819	8435	11087	13752
d	8	2607	4680	7230	10168	12592	15247
	16	5376	7053	9599	12352	15032	17700
	32	9852	11498	14192	16996	19684	22335
		PSNR in dB					
	2	48.008	48.031	48.058	48.058	48.058	48.058
	4	48.002	48.058	48.058	48.058	48.058	48.058
d	8	48.044	48.058	48.058	48.066	48.058	48.058
	16	48.066	48.066	48.058	48.058	48.058	48.058
	32	48.058	48.058	48.058	48.058	48.058	48.058

always decreases with p—even when $\|\nabla\Theta(\boldsymbol{x}^{(p)})\|$ increases—and is nearly constant between consecutive gradient peaks starting from $p = 200$. In other words, in the close vicinity of a stationary point, the trajectory of the iterates switches between nearly flat regions of decreasing height in the objective landscape. Each gradient peak also coincides with a rise in PSNR; however, after 200 iterations, the maximum relative difference of the PSNR lies between 2.7×10^{-4} for $\mu = 10^{-7}$ and 4.6×10^{-4} for $\mu = 0.1$, so there is no need to reach machine precision from the standpoint of this quality measure. Finally, we note that the outer termination criterion (7.18) (represented by the gray dashed curve in the upper plot) seems to amount to a constant threshold on the norm of the gradient. Indeed, after 50 iterations, the objective lies in the interval $[25.085, 25.138]$ and thus (7.18) is approximately equivalent to $\|\nabla\Theta(\boldsymbol{x}^{(p)})\| \leqslant 25\,\check{\epsilon}$.

8.3. **Deblurring under mixed Gaussian-impulse noise.** Here we focus on the problem of restoring the 254×254 "peppers" image shown in Figure 10(a) from the blurred and noisy data displayed in Figure 10(b). These data were generated by first blurring with a 7×7 rotationally symmetric Gaussian kernel with a standard deviation of one pixel, then adding white Gaussian noise at 30 dB SNR (this decibel level corresponds to a noise standard deviation of 1.94), and finally degrading the

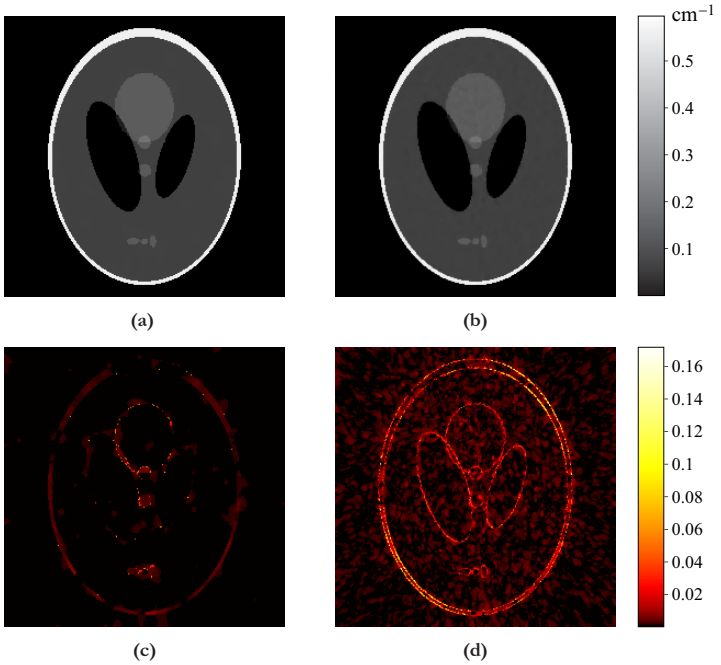


FIGURE 8. Reconstructions of the Shepp-Logan phantom using spatial-gradient regularization with (a) the square root approximation $\vartheta_{10^{-3}}$ (PSNR = 48.058 dB) and (b) the identity approximation $\mathrm{id}_{10^{-3}}$ (PSNR = 37.006 dB). The images shown in (c) and (d) are the corresponding absolute differences with the phantom.

result by 20% random-valued impulse noise. The observation matrix $\boldsymbol{D}$ is of size $248^2 \times n$, $n = 254^2$, and has a density of 0.076%.

Borrowing from robust statistics [45], we deal with mixed Gaussian-impulse noise by using the Huber function ϑ_{Hu} defined in (A.8). The objective is of the form

$$\Theta(\boldsymbol{x}) := \sum_k \vartheta_{\mathrm{Hu}}\big(|[\boldsymbol{D}\boldsymbol{x} - \boldsymbol{d}]_k|/\varsigma\big) + \Theta_{\mathrm{Frame}}(\boldsymbol{x}) + \Theta_{\mathrm{TV}}(\boldsymbol{x}),$$

where ς is the standard deviation of the Gaussian noise component, Θ_{Frame} is a nonconvex tight-frame regularizer, and Θ_{TV} is a first-order isotropic TV regularizer that ensures coercivity and penalizes isolated outliers. The tight-frame regularizer is given by

$$\Theta_{\mathrm{Frame}}(\boldsymbol{x}) := \gamma_1 \sum_l \vartheta_{\mathrm{LE}}\big(|[\boldsymbol{H}\boldsymbol{x}]_l|\big), \tag{8.4}$$

where ϑ_{LE} is the Lorentzian error function defined in (A.9), and $\boldsymbol{H} \in \mathbb{R}^{16n \times n}$ is the vertical concatenation of the high-pass operators of a two-level tight-frame system generated from piecewise linear filters (we refer to [46] for details). Note that there is no need to scale the argument of ϑ_{LE} to preserve tight-frame coefficients of large magnitude, since ϑ_{LE} is strictly concave on $(1, +\infty)$ and the standard deviation of the tight-frame coefficients of $\boldsymbol{x}^\bullet$ (the original "peppers" image) is greater than 5.

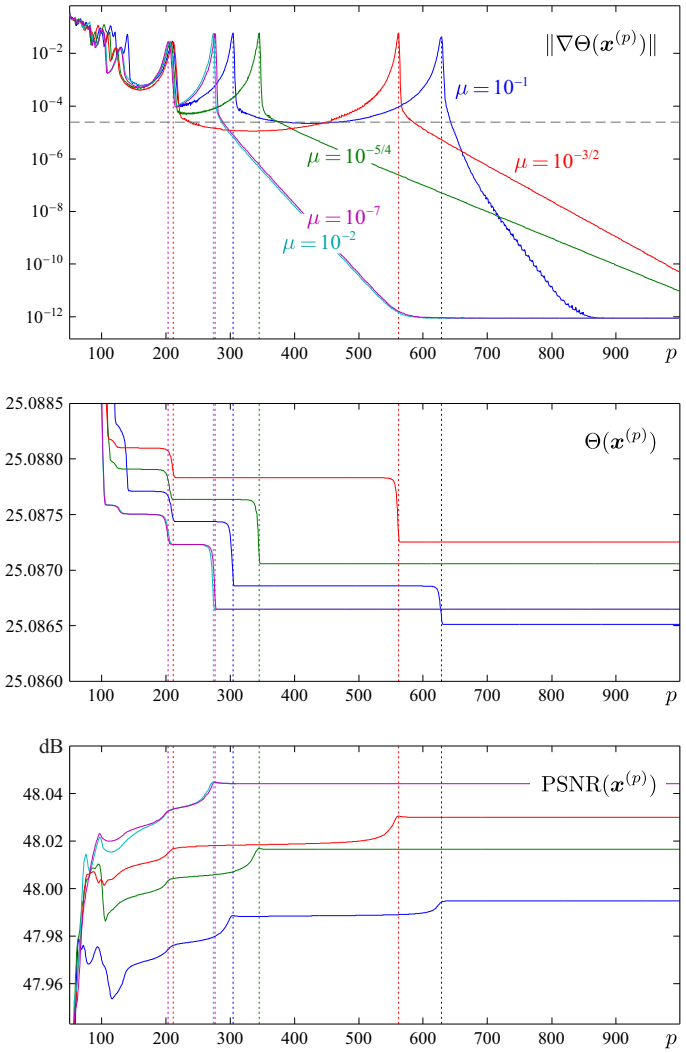


FIGURE 9. From top to bottom: norm of the gradient, objective, and PSNR versus number of outer iterations for $d = 4$ and different accuracy levels. The vertical dotted lines indicate the locations of the main gradient peaks, and the gray dashed curve in the upper plot represents the right-hand side of the termination criterion (7.18) for a tolerance of 10^{-6}.

The TV regularizer has the form

$$\Theta_{\mathrm{TV}}(\boldsymbol{x}) \;:=\; \gamma_2 \sum_l \mathrm{id}_\delta\big(\|\boldsymbol{G}_l\boldsymbol{x}\|\big),$$

where id_δ and $\{\boldsymbol{G}_l\}_l$ are as in Section 8.2. We set $\delta = 0.1$, which is much smaller than the standard deviation of the spatial-gradient magnitudes in $\boldsymbol{x}^\bullet$ (about 18.0). As for the regularization strengths, we adjust γ_1 to obtain the highest SSIM without TV regularization (about 0.946, achieved for $\gamma_1 = 0.11$), and then we choose γ_2

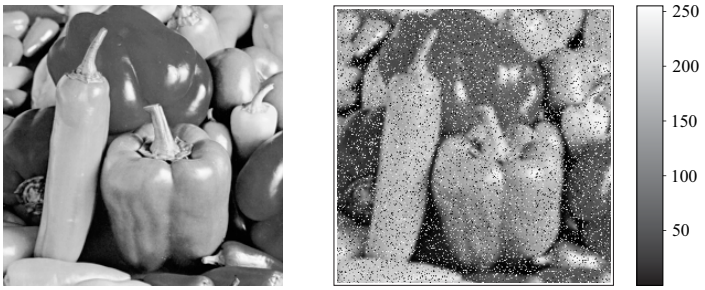


FIGURE 10. Deblurring problem: (a) original 254×254 "peppers" image; (b) 248×248 degraded observation (7×7 Gaussian blur, 30 dB white Gaussian noise, and 20% random-valued impulse noise).

small enough to not interfere with the tight-frame regularizer except for preventing outliers ($\gamma_2 = 0.05$).

We guide the first steps of the optimization process using the continuation sequence $(\vartheta^{(p)})_{p\in[0..25]}$ illustrated in Figure 11 and defined by

$$\vartheta^{(p)} := \kappa_p \vartheta_{\mathrm{LE}} + (1-\kappa_p)\vartheta_{\mathrm{MS}}, \tag{8.5}$$

where $(\kappa_p)_{p\in[0..25]}$ increases linearly from 0 to 1 and ϑ_{MS} is the function of minimal surfaces defined in (A.2). Furthermore, to avoid unnecessary CG computations during the continuation phase, we replace the potential in the TV regularizer by id_{δ_p} with $(\delta_p)_{p\in[0..25]}$ decreasing linearly from 100δ to δ. Figure 12 shows the restoration obtained for $(\mu, d) = (10^{-3}, 4)$ and an outer termination tolerance of 10^{-6} (we keep using the Gauss quadrature FHQ algorithm with the gradient-based termination criterion (7.18)). Nonconvex regularization outperforms convex regularization: the displayed solution has an SSIM of 0.94601 and a PSNR of 33.143 dB, while the best solution obtained with ϑ_{LE} replaced by $\mathrm{id}_{10^{-2}}$ (so that $\Theta_{\mathrm{Frame}}/\gamma_1$ approximates the ℓ_1-norm of the tight-frame coefficients) has an SSIM of 0.94491 and a PSNR of 32.875 dB.

We now examine the impact of the outer termination tolerance and of preconditioning the inner systems. We do not seek to compare possible preconditioning strategies, but it is important to note that computationally expensive preconditioners should be discarded because the inner system matrix $\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x}^{(p)})\boldsymbol{A}$ changes at each outer iteration. As an alternative to Jacobi preconditioning, we consider IC preconditioning with an absolute dropping strategy, which is feasible because $\boldsymbol{A}^T\boldsymbol{A}$ is only 0.26% dense (as opposed to 71% for the tomographic reconstruction problem). The drop tolerance is set to 10^{-3}, the largest integer power of 10 for which there is no pivot breakdown.

Table 2 summarizes the effects of decreasing the termination tolerance $\check{\epsilon}$ and switching from Jacobi to IC preconditioning; the reported values are the minimum, the maximum relative difference, and the digit accuracy of the SSIM and PSNR of the solutions obtained for $\mu \leqslant 0.1$ and $d \geqslant 2$. (The digit accuracy of a positive function f on a set $\mathcal{A}$ is the largest integer $\mathfrak{d} \geqslant 1$ such that the difference $\sup_{\mathcal{A}} f - \inf_{\mathcal{A}} f$ is less than half a unit in the $\mathfrak{d}$th significant digit of $\inf_{\mathcal{A}} f$.) We see that

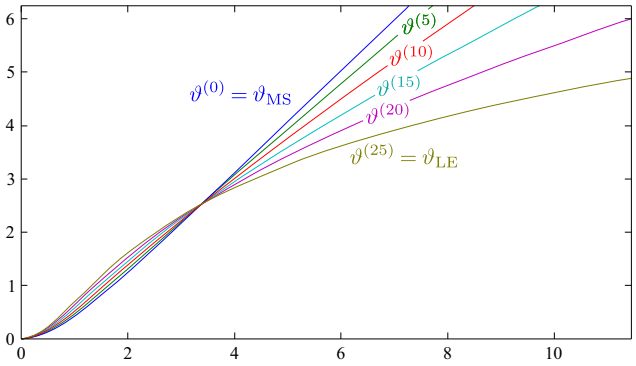
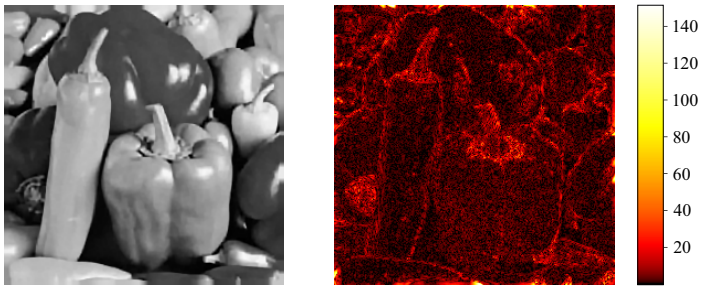


FIGURE 11. Continuation sequence (8.5) for the Lorentzian error potential (the maximum abscissa is twice the standard deviation of the tight-frame coefficients of the original "peppers" image).



(a) (b)

FIGURE 12. (a) Restoration of the "peppers" image using tight-frame regularization with the Lorentzian error potential (SSIM = 0.94601, PSNR = 33.143 dB); (b) absolute difference with the original image.

the performance and stability of the Gauss quadrature FHQ algorithm — which are already good for $\breve{\epsilon} = 10^{-6}$ and Jacobi preconditioning — are significantly improved by decreasing $\breve{\epsilon}$ and using IC preconditioning. Furthermore, IC preconditioning brings the trajectories closer together: for $\breve{\epsilon} = 10^{-12}$ and Jacobi preconditioning, the total number of outer iterations $\mathfrak{N}$ is in the range $[724\,..\,1416]$ and stabilizes to 798 when $\mu \leqslant 10^{-5}$, whereas with IC preconditioning $\mathfrak{N} = 798$ for all $\mu \leqslant 0.1$.

Figure 13 shows the total number of CG iterations, $\mathfrak{N}_{\mathrm{CG}}$, as a function of the inner control parameters for the two preconditioners and $\breve{\epsilon} = 10^{-12}$. Compared with Jacobi preconditioning, IC preconditioning substantially reduces $\mathfrak{N}_{\mathrm{CG}}$ and its rate of increase with the log-accuracy. For Jacobi preconditioning, $\mathfrak{N}_{\mathrm{CG}}$ is eventually affine in the log-accuracy with a slope of about $4.8 \times \mathfrak{N}$, whereas for IC preconditioning $\mathfrak{N}_{\mathrm{CG}}$ increases approximately by $\mathfrak{N}$ every several units in log-accuracy: $\mathfrak{N}_{\mathrm{CG}} \approx (d+1)\mathfrak{N}$ if $-\log_{10}\mu \leqslant 3$, $(d+2)\mathfrak{N}$ if $-\log_{10}\mu \in [4\,..\,7]$, and $(d+3)\mathfrak{N}$ if $-\log_{10}\mu \in \{8, 9\}$. This piecewise constant-like behavior is explained as follows. At low accuracy, up to $-\log_{10}\mu = 3$, the CG stopping criterion (7.15) is usually satisfied for $j = 1$, its evaluation then requiring $d+1$ CG iterations. If the

TABLE 2. Effects of decreasing the outer termination tolerance and using IC preconditioning: minimum, maximum relative difference, and digit accuracy of the SSIM and PSNR for $\mu \leqslant 0.1$ and $d \geqslant 2$.

	$\breve{\epsilon}$, preconditioner		
	10^{-6}, Jacobi	10^{-12}, Jacobi	10^{-12}, IC
	SSIM		
min.	0.946026	0.946194	0.946196
max. rel. diff.	1.13×10^{-4}	4.65×10^{-6}	1.31×10^{-12}
digit acc.	3	5	11
	PSNR in dB		
min.	33.1375	33.1740	33.1748
max. rel. diff	5.13×10^{-4}	1.18×10^{-4}	6.65×10^{-12}
digit acc.	3	4	11

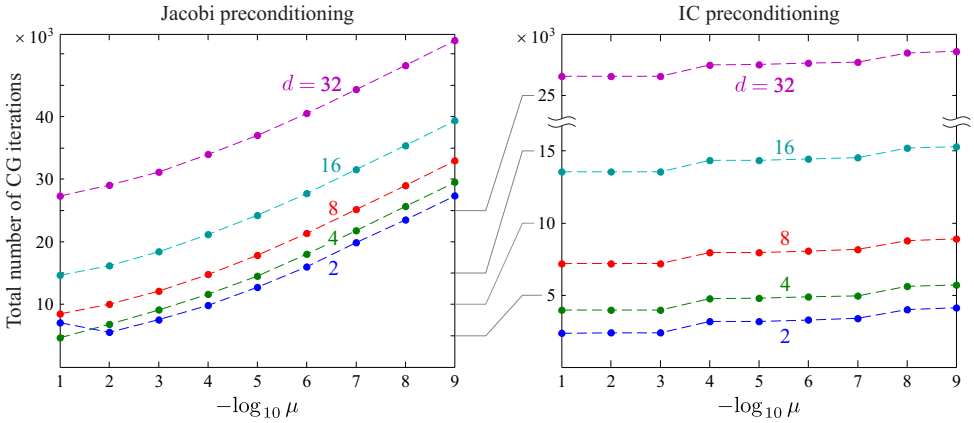


FIGURE 13. Total number of CG iterations ($\mathfrak{N}_{\mathrm{CG}}$) as a function of μ and d for Jacobi and IC preconditioning and for an outer termination tolerance of 10^{-12}.

log-accuracy is increased to 4, then (7.15) generally holds for $j = 2$, thus adding one CG iteration per outer iteration. Furthermore, this extra CG iteration increases the log-accuracy by several units, resulting in $\mathfrak{N}_{\mathrm{CG}}$ remaining approximately constant up to $-\log_{10}\mu = 7$. Similar arguments can be made when increasing the log-accuracy from 7 to 8.

Figure 14 plots the norm of the gradient versus the cumulative number of iterations (that is, $\|\nabla\Theta(\boldsymbol{x}^{(p)})\|$ versus $\sum_{q\in[1..p]}(d+j_q)$, where j_q is defined in the same way as for (8.3)) and the number of CG iterations per outer iterations (that is, $d+j_p$ versus p) for the two preconditioners and for $(\mu, d) = (10^{-2}, 4)$ and $(10^{-8}, 4)$. The gradient curves are nearly scaled versions of each other (the dotted curves in Figure 14(b) are those of Figure 14(a)), so the iterates $\boldsymbol{x}^{(p)}$ follow similar paths for the four cases considered; indeed, the trajectory of the iterates is stable when $\mu \leqslant 10^{-2}$, like for the reconstruction problem. As expected, the ratio of the total

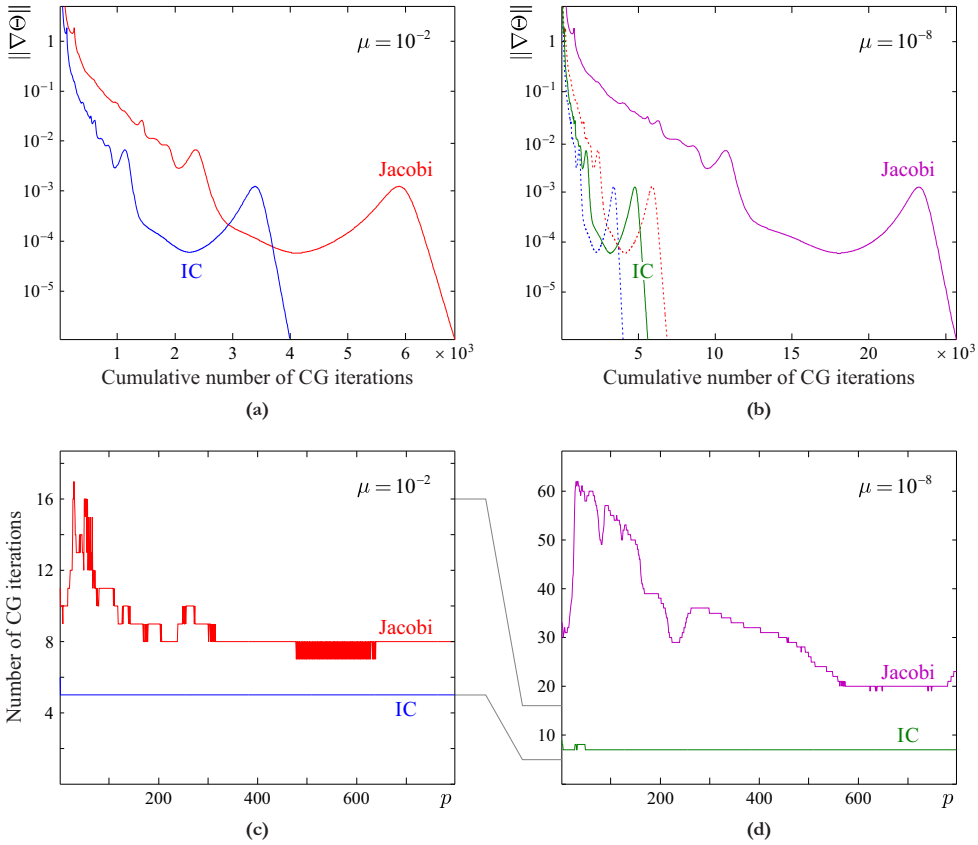


FIGURE 14. Jacobi versus IC preconditioning at low and high accuracy: (a)(b) norm of the gradient as a function of the cumulative number of CG iterations; (c)(d) number of CG iterations per outer iteration. (The inner control parameter d is kept equal to 4.)

number of CG iterations for Jacobi preconditioning to that for IC preconditioning increases with the log-accuracy, from 1.72 for $\mu = 10^{-2}$ to 4.58 for $\mu = 10^{-8}$. We see from Figures 14(c) and (d) that with IC preconditioning, the number of CG iterations $d + j_p$ is equal to 5 (from the second iteration on) for $\mu = 10^{-2}$ and quickly stabilizes at 7 for $\mu = 10^{-8}$; also, the increase in j_p resulting from increasing the log-accuracy is much smaller for IC than for Jacobi preconditioning. These last observations agree with those from Figure 13.

8.4. **Patch-based inpainting.** We complete our experiments with the restoration of the 256×256 "Barbara" face image shown in Figure 15(a) from the moderately sparse data displayed in Figure 15(b). The original image is the upper right quadrant of the full 512×512 "Barbara" image, and the data consist of 20% of the original pixels chosen uniformly at random. The observation matrix $\boldsymbol{D}$ is an $m \times n$ binary matrix with $m = 13108$ and $n = 256^2$; its nonzero entries are $(1, l_1), \ldots, (m, l_m)$, where $l_1, \ldots, l_m \in [1\,..\,n]$ are the indices of the nonmissing pixels.

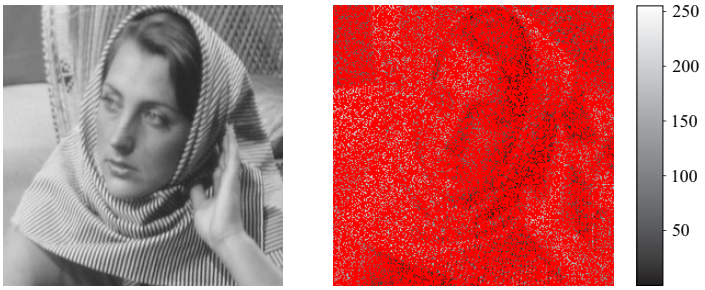


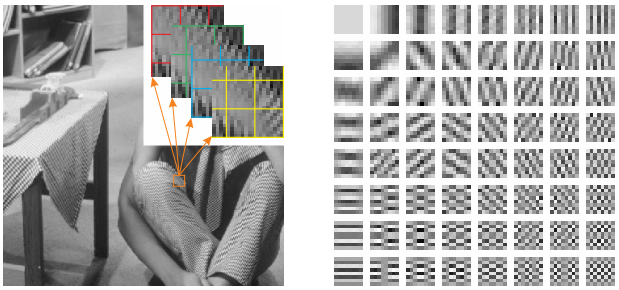
FIGURE 15. Inpainting problem: (a) original 256×256 "Barbara" face image; (b) data obtained by randomly selecting 20% of the pixels (the missing pixels are displayed in red).

We seek to minimize the sum of a quadratic data-fidelity term and a regularization term based on the local sparse model introduced in [47]. This model assumes that small patches of size, say, $\sqrt{w} \times \sqrt{w}$ in the original image $\boldsymbol{x}^\bullet$ are *approximately sparsifiable* by a single transform $\boldsymbol{W} \in \mathbb{R}^{w \times w}$ (we suppose that w is the square of an integer); more precisely, the columnwise representation $\boldsymbol{y} \in \mathbb{R}^w$ of any patch in $\boldsymbol{x}^\bullet$ satisfies $\boldsymbol{W}\boldsymbol{y} = \boldsymbol{z} + \boldsymbol{e}$, where $\boldsymbol{z}$ is sparse and $\boldsymbol{e}$ is a small residual error in the transform domain. The objective is defined by

$$\Theta(\boldsymbol{x}) := \|\boldsymbol{D}\boldsymbol{x} - \boldsymbol{d}\|^2 + \gamma \sum_i \sum_j \vartheta_{\mathrm{GM}}\big(|[\boldsymbol{W}\boldsymbol{P}_i\boldsymbol{x}]_j|/\delta\big),$$

where ϑ_{GM} is the Geman and McClure function defined in (A.10), and $\boldsymbol{P}_i \in \mathbb{R}^{w \times n}$ is a binary matrix that extracts the patch at the ith pixel location ($\boldsymbol{P}_i\boldsymbol{x}$ is the columnwise representation of the pixels in the image subdomain $[i_1 \,..\, i_1 + \sqrt{w} - 1] \times [i_2 \,..\, i_2 + \sqrt{w} - 1]$, where i_1 and i_2 are the ith pixel coordinates). The outer sum in the regularizer is over evenly-spaced pixel locations, and the inner sum is over the components of the vector $\boldsymbol{W}\boldsymbol{P}_i\boldsymbol{x} \in \mathbb{R}^w$.

We consider patches of size 8×8 (that is, $w = 64$) and we use the *doubly-sparse* transform (DST) proposed in [48]: $\boldsymbol{W}$ is the product $\boldsymbol{S}\boldsymbol{V}$ of a learnt sparse transform $\boldsymbol{S} \in \mathbb{R}^{64 \times 64}$ with the 2-D discrete cosine transform (DCT) $\boldsymbol{V} \in \mathbb{R}^{64 \times 64}$ (defined as the Kronecker product of the unitary, 8×8, 1-D DCT matrix with itself). The matrix $\boldsymbol{W}$ is called doubly sparse because it is a sparsifying transform and its left factor $\boldsymbol{S}$ is sparse. As illustrated in Figure 16(a), the training patches for learning $\boldsymbol{S}$ are extracted from the the upper left, lower left, and lower right quadrants of the full "Barbara" image, with 50% overlap between any two successive patches in the horizontal or vertical direction. The learning is performed using the MATLAB code available from `http://transformlearning.csl.illinois.edu` with default settings—a patch representation of the resulting DST is shown in Figure 16(b). The regularizer operates on a patch cover with 75% horizontal and vertical overlap; in other words, the regularization operator is the patch-based DST formed by stacking the matrices $\boldsymbol{W}\boldsymbol{P}_i$ with i indexing the pixels whose coordinates are both odd (this concatenated transform is of size $10^6 \times n$ and about 0.1% dense). Noting that ϑ_{GM} is strictly concave on $(1/\sqrt{3}, +\infty)$, we set $\delta = 5$ so as to preserve the DST coefficients with magnitude greater than $5/\sqrt{3}$, as is the case for nearly one



(a) (b)

FIGURE 16. DST learning: (a) the 8×8 training patches are extracted from the full 512×512 "Barbara" image with the upper right quadrant removed (the sliding distance is 4 pixels, both horizontally and vertically); (b) rows of the learnt DST $\boldsymbol{W} \in \mathbb{R}^{64\times 64}$ represented as 8×8 patches.

third of the original DST coefficients $[\boldsymbol{W}\boldsymbol{P}_i\boldsymbol{x}^\bullet]_j$. Again, the regularization strength γ is adjusted to obtain the highest SSIM (about 0.873, achieved for $\gamma = 7.9\times 10^{-6}$).

Since $\operatorname{null}(\boldsymbol{D}) \neq \{\boldsymbol{0}\}$ and ϑ_{GM} is bounded, Θ is not coercive. However, Condition (C4), which becomes $\operatorname{null}(\boldsymbol{D}) \cap \big(\bigcap_i \operatorname{null}(\boldsymbol{W}\boldsymbol{P}_i)\big) = \{\boldsymbol{0}\}$, is satisfied. Indeed, $\operatorname{null}(\boldsymbol{W}\boldsymbol{P}_i)$ is the set of images whose ith patch is zero (because $\boldsymbol{W}$ is invertible) and the patches extracted via the matrices $\boldsymbol{P}_i$ cover the whole pixel grid; so $\bigcap_i \operatorname{null}(\boldsymbol{W}\boldsymbol{P}_i) = \{\boldsymbol{0}\}$. We use a continuation sequence similar to (8.5) but with ϑ_{LE} replaced by ϑ_{GM}. This sequence is illustrated in Figure 17; its role is to guide the first iterates towards a generalized cup of Θ so that Theorem 5.5 holds. Figure 18(a) shows the restoration obtained for $(\mu, d) = (10^{-3}, 5)$ after 85 iterations, the minimum number needed to satisfy the gradient-based termination criterion (7.18) with a tolerance of 10^{-6}. For comparison, Figure 18(b) shows the cubic interpolation computed using the MATLAB function "griddata". Cubic interpolation properly recovers smooth regions but fails to restore the strip texture of the shawl; the corresponding SSIM and PSNR are, respectively, about 13% and 4.2 dB smaller than for the DST-regularized solution.

Below we compare the Gauss quadrature FHQ algorithm considered so far to the backward error FHQ algorithm. The latter uses the CG stopping criterion (7.11), whose parameter ϵ is not to be confused with the tolerance $\check{\epsilon}$ of the outer termination criterion. We recall that the minimum number J of CG iteration per outer iteration is implicitly set to $d + 1$ for the Gauss quadrature algorithm, and we set $J = 1$ for the backward error algorithm. So the numbers of CG iterations at the pth outer iteration of the Gauss quadrature and backward error algorithms are $d + j_p$ and j'_p, respectively, where j_p is the smallest integer $j \geqslant 1$ satisfying (7.15) with $\boldsymbol{x} = \boldsymbol{x}^{(p)}$ and j'_p is defined similarly for (7.11).

Figure 19 plots the norm of the gradient and the number of CG iterations over 8000 outer iterations for the Gauss quadrature algorithm with $(\mu, d) = (10^{-2}, 4)$ and $(10^{-3}, 5)$ and for the backward error algorithm with $\epsilon = 10^{-14}$, 10^{-15}, and $2^{-53} \approx 1.11\times 10^{-16}$ (the latter value is the unit roundoff in IEEE standard double-precision arithmetic). The number of outer iterations to reach machine precision is significantly smaller for the Gauss quadrature algorithm (about 2500) than for

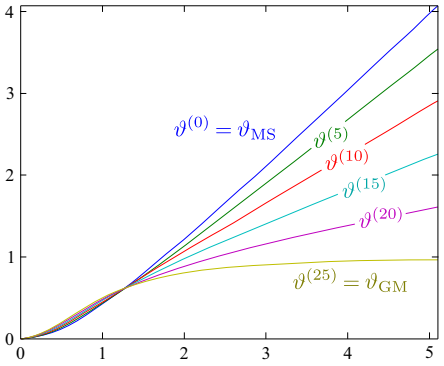


FIGURE 17. Continuation sequence for the Geman and McClure potential (the maximum abscissa is twice the standard deviation of the scaled original DST coefficients $[\boldsymbol{W}\boldsymbol{P}_i\boldsymbol{x}^\bullet]_j/\delta$).

(a) (b)

FIGURE 18. Restoration of the "Barbara" face image: (a) using patch-based DST regularization with the Geman and McClure potential (SSIM = 0.87376, PSNR = 25.847 dB); (b) using cubic interpolation (SSIM = 0.75858, PSNR = 21.631 dB).

the backward error algorithm (between 3500 and 8000). Beyond 2500 outer iterations, the cumulative number of CG iterations is larger for the lower-accuracy Gauss quadrature algorithm (with $(\mu, d) = (10^{-2}, 4)$) than for the higher-accuracy backward error algorithm (with $\epsilon = 2^{-53}$). However, the backward error algorithm is computationally less efficient because of the extra computations needed to evaluate (7.11). For example, over 5000 outer iterations, the cumulative number of CG iterations is about five times larger for the higher-accuracy Gauss quadrature algorithm (with $(\mu, d) = (10^{-3}, 5)$) than for the lower-accuracy backward error algorithm (with $\epsilon = 10^{-14}$), whereas the former is 36% faster than the latter.

In the case of the Gauss quadrature algorithm, the number of CG iterations per outer iteration is stable from the end of the continuation phase until machine precision is reached (for all $p \in [25\,..\,2500]$, $d + j_p$ is in $[9\,..\,12]$ for $(\mu, d) = (10^{-2}, 4)$ and in $[12\,..\,16]$ for $(\mu, d) = (10^{-3}, 5)$), after which j_p drops to 1. The number of CG iterations j'_p controlled by the backward error criterion is less stable than j_p. We also observe that $j'_p = 1$ whenever the norm of the gradient is below some threshold. Indeed, as shown in Appendix C, there is a constant $\epsilon' \in (0, \epsilon)$ such

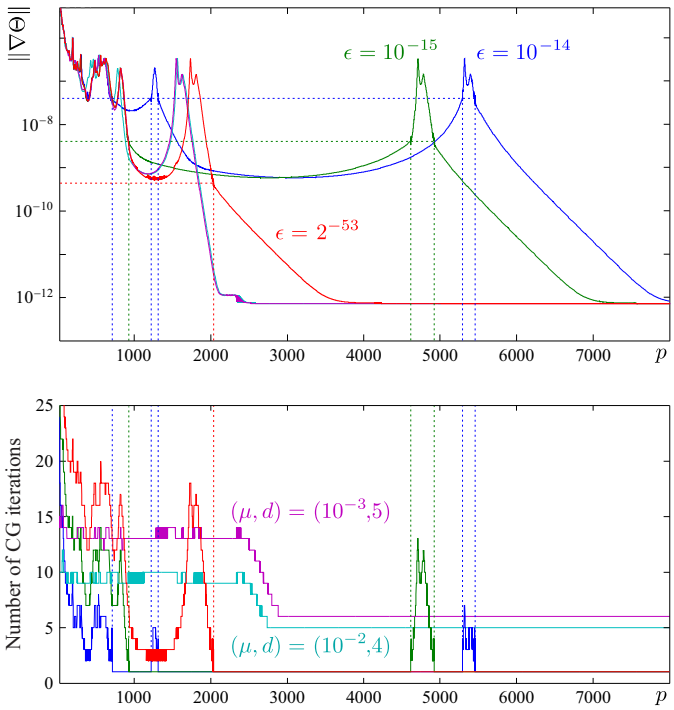


FIGURE 19. Norm of the gradient (top) and number of CG iterations (bottom) versus number of outer iterations for the Gauss quadrature FHQ algorithm (cyan and magenta) and the backward error FHQ algorithm (red, green and blue). The horizontal dotted lines represent the thresholds below which the backward error criterion (7.11) is satisfied from the start.

that for sufficiently large p, the backward error criterion is satisfied from the start if $\|\nabla\Theta(\boldsymbol{x}^{(p)})\| \leqslant \epsilon'\mathfrak{T}^*$, where $\mathfrak{T}^*$ is a constant depending on the limit stationary point. For the three values of ϵ considered here, the backward error algorithm converge to the same stationary point with $\mathfrak{T}^* \approx 4 \times 10^6$, and the gradient threshold delimiting the intervals in which $j'_p = 1$ is very close to $\epsilon\mathfrak{T}^*$, suggesting that $\|\boldsymbol{r}^{(1,p)}\| \leqslant \|\boldsymbol{r}^{(0,p)}\|$ for sufficiently large p (see Proposition C.2).

It is important to note that the observations from Figure 19 concern small behavior changes in the long run. In fact, for $\check{\epsilon} = 10^{-6}$ (the standard value of the outer termination tolerance), the two algorithms terminate in 84 or 85 outer iterations and perform similarly in terms of objective, PSNR, and SSIM: beyond 85 outer iterations, the objective and the PSNR are constant to five significant digits and vary by less than 0.08% and 0.22%, respectively, and the SSIM is constant to four significant digits and varies by less than 0.09%.

9. Conclusion

We showed that the inexact FHQ method converges globally to a stationary point of the objective if the conditions (C1)–(C4) hold and the potentials are tame. These minimal assumptions cover all current applications of minimizing C^1 functions of

the general form (1.3), which are prevalent in regularized image reconstruction and restoration. Moreover, even in the unlikely situation where some potentials are not tame, the inexact FHQ method converges either to a stationary point or to the boundary of the set of stationary points.

We proposed an efficient and numerically stable implementation based on truncated CG and coming in two versions: the backward error algorithm, which uses a standard CG stopping criterion with no control on the accuracy of the outer iterates, and the Gauss quadrature algorithm, which uses a less conventional criterion that monitors the error energy norm to control the accuracy. The Gauss quadrature algorithm is the more computationally efficient of the two; it is also robust to the choice of its inner parameters μ and d. In practice, a good compromise between solution quality and computation time is achieved by setting $(\mu, d) = (10^{-3}, 5)$ and using the gradient-based outer termination criterion (7.18) with a tolerance of 10^{-6}. When the objective in nonconvex, we recommend using a continuation sequence at the beginning of the optimization process. This inexpensive technique leads to deeper minima without altering the convergence properties of the inexact FHQ method (it also guides the first iterates in case the starting point is not in a generalized cup).

We investigated the behavior of the inexact FHQ method in the contexts of tomographic reconstruction, deblurring, and inpainting. Our experiments showed that it performs well in various nonconvex scenarios and consistently converges to stationary points to machine precision, illustrating the fact that inexact FHQ optimization is a fast reliable answer to numerous inverse problems.

Appendix A. Common potential functions

A potential function θ is usually defined by scaling the argument and the value of a *mother potential function* ϑ, that is,

$$\theta(t) := \gamma\vartheta(t/\delta), \tag{A.1}$$

where γ and δ are free positive parameters. We list below some common mother potentials from the computer vision and robust statistics literature, beginning with convex functions with affine behavior at infinity (definitions (A.2)–(A.8)) and continuing with nonconvex functions (definitions (A.9)–(A.14)). Each of them is admissible, that is, ϑ is increasing on $\mathbb{R}_+$, C^1 on $(0, +\infty)$, continuous at zero, and such that $t^{-1}\vartheta'(t)$ is decreasing and bounded on $(0, +\infty)$.

- Function of minimal surfaces [49]:

$$\vartheta_{\mathrm{MS}}(t) := (1 + t^2)^{1/2} - 1. \tag{A.2}$$

- Green function [50]:

$$\vartheta_{\mathrm{Gr}}(t) := \ln(\cosh t). \tag{A.3}$$

- Lange functions [51]:

$$\vartheta_{\mathrm{La},1}(t) := t^2(1 + t)^{-1}, \tag{A.4}$$

$$\vartheta_{\mathrm{La},2}(t) := t + \exp(-t) - 1, \tag{A.5}$$

$$\vartheta_{\mathrm{La},3}(t) := t - \ln(1 + t), \tag{A.6}$$

$$\vartheta_{\mathrm{La},4}(t) := 2t \arctan t - \ln(1 + t^2). \tag{A.7}$$

- Huber function [45, 52, 53]:

$$\vartheta_{\mathrm{Hu}}(t) := \begin{cases} \frac{1}{2}t^2 & \text{if } t \leqslant 1, \\ t - \frac{1}{2} & \text{if } t > 1. \end{cases} \tag{A.8}$$

- Lorentzian error function [52, 54]:

$$\vartheta_{\mathrm{LE}}(t) := \ln(1 + t^2). \tag{A.9}$$

- Geman and McClure function [55]:

$$\vartheta_{\mathrm{GM}}(t) := t^2(1 + t^2)^{-1}. \tag{A.10}$$

- Welsch function [56, 57]:

$$\vartheta_{\mathrm{We}}(t) := 1 - \exp(-t^2). \tag{A.11}$$

- Hyperbolic tangent function [20]:

$$\vartheta_{\mathrm{HT}}(t) := \tanh(t^2). \tag{A.12}$$

- Tukey's biweight function [45]:

$$\vartheta_{\mathrm{TB}}(t) := \begin{cases} 1 - \left(1 - \frac{1}{6}t^2\right)^3 & \text{if } t \leqslant \sqrt{6}, \\ 1 & \text{if } t > \sqrt{6}. \end{cases} \tag{A.13}$$

- Andrew's sine function [45]:

$$\vartheta_{\mathrm{AS}}(t) := \begin{cases} (\sin t)^2 & \text{if } t \leqslant \frac{\pi}{2}, \\ 1 & \text{if } t > \frac{\pi}{2}. \end{cases} \tag{A.14}$$

Appendix B. O-minimal structures on the real field

In this appendix, we recall some definitions and results concerning o-minimal structures. The content is limited to what is necessary for the paper to be self-contained; we refer to [14–16] for a comprehensive introduction to the subject.

Definition B.1 (O-minimal structure)**.** A *structure* on the real field $(\mathbb{R}, +, \cdot)$ is a sequence $\mathcal{M} := (\mathcal{M}_n)_{n\in\mathbb{N}}$ such that for all $n \in \mathbb{N}$,

(i) $\mathcal{M}_n$ is a boolean algebra of subsets of $\mathbb{R}^n$ (that is, $\mathcal{M}_n$ is nonempty and closed under complementation and finite union),

(ii) $\mathcal{M}_n$ contains the family of algebraic subsets of $\mathbb{R}^n$ (that is, the sets of the form $\{\boldsymbol{x} \in \mathbb{R}^n : P(\boldsymbol{x}) = 0\}$, where P is a real polynomial function on $\mathbb{R}^n$),

(iii) if $\mathcal{A} \in \mathcal{M}_n$ and $\mathcal{B} \in \mathcal{M}_p$, then $\mathcal{A} \times \mathcal{B} \in \mathcal{M}_{n+p}$,

(iv) if $\mathcal{A} \in \mathcal{M}_{n+1}$, then $\pi(\mathcal{A}) \in \mathcal{M}_n$, where $\pi : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^n$ is the projection map on the first n coordinates.

If, in addition,

(v) the sets in $\mathcal{M}_1$ are exactly the finite unions of intervals and points,

then $\mathcal{M}$ is said to be *o-minimal*.

Definition B.2 (Definability)**.** Let $\mathcal{M} := (\mathcal{M}_n)_{n\in\mathbb{N}}$ be a structure on $(\mathbb{R}, +, \cdot)$. A set $\mathcal{A}$ is said to be *definable in* $\mathcal{M}$, or $\mathcal{M}$*-definable*, if $\mathcal{A} \in \mathcal{M}_n$ for some $n \in \mathbb{N}$. Given such a set, a function $f : \mathcal{A} \longrightarrow \mathbb{R}$ is called $\mathcal{M}$-definable if its graph $\{(\boldsymbol{x}, h) \in \mathcal{A} \times \mathbb{R} : f(\boldsymbol{x}) = h\}$ belongs to $\mathcal{M}_{n+1}$.

Sets and functions that are not definable in any structure are said to be *nondefinable*. A function definable in an o-minimal structure is simply called an *o-minimal function*.

B.1. **Examples of o-minimal structure.** We give here three fundamental examples of o-minimal structures on the real field: the structure $\mathbb{R}_{\mathrm{alg}}$ of semialgebraic sets, the structure $\mathbb{R}_{\mathrm{an}}$ of globally subanalytic sets [58], and the structure $\mathbb{R}_{\mathrm{an,exp}}$ of analytic-exponential sets [59]. The structures $\mathbb{R}_{\mathrm{an}}$ and $\mathbb{R}_{\mathrm{an,exp}}$ are defined by successively expanding $\mathbb{R}_{\mathrm{alg}}$. We first recall the definition of the expansion of a structure.

Definition B.3 (Expansion)**.** Let $\mathcal{M} := (\mathcal{M}_n)_{n\in\mathbb{N}}$ and $\mathcal{M}' := (\mathcal{M}'_n)_{n\in\mathbb{N}}$ be two structures on $(\mathbb{R}, +, \cdot)$. If $\mathcal{M}_n \subseteq \mathcal{M}'_n$ for all $n \in \mathbb{N}$, then $\mathcal{M}'$ is called an *expansion* of $\mathcal{M}$. Given a collection $\mathscr{F}$ of real functions, each with domain in $\mathbb{R}^n$ for some $n \in \mathbb{N}$, we denote by $\langle \mathcal{M}, \mathscr{F} \rangle$ the smallest expansion of $\mathcal{M}$ in which all the functions in $\mathscr{F}$ are definable.

Example 7 (Semialgebraic sets). Let $\mathbb{R}_{\mathrm{alg}} := (\mathbf{alg}(\mathbb{R}^n))_{n\in\mathbb{N}}$, where $\mathbf{alg}(\mathbb{R}^n)$ denotes the class of semialgebraic subsets of $\mathbb{R}^n$, that is, the finite unions of sets of the form (5.11). The sequence $\mathbb{R}_{\mathrm{alg}}$ clearly satisfies conditions (i)–(iii) and (v) in Definition B.1, and condition (iv) follows from the Tarski-Seidenberg principle [23, Theorem 2.2.1]. Furthermore, it is easy to see that any o-minimal structure on $(\mathbb{R}, +, \cdot)$ is an expansion of $\mathbb{R}_{\mathrm{alg}}$. In other words, $\mathbb{R}_{\mathrm{alg}}$ is the smallest o-minimal structure on the real field (in contrast, there is no largest o-minimal structure [60]). The potential functions ϑ_{MS}, $\vartheta_{\mathrm{La},1}$, ϑ_{Hu}, ϑ_{GM}, and ϑ_{TB} defined in Appendix A are $\mathbb{R}_{\mathrm{alg}}$-definable.

Example 8 (Finitely subanalytic sets). Let $\mathbb{R}_{\mathrm{an}} := \langle \mathbb{R}_{\mathrm{alg}}, \bigcup_{n\in\mathbb{N}} \mathbf{an}([-1,1]^n) \rangle$, where $\mathbf{an}([-1,1]^n)$ is the class of restricted analytic functions on $[-1,1]^n$ (see Definition 6.6). The structure $\mathbb{R}_{\mathrm{an}}$ is o-minimal [58] and consists of the so-called *finitely subanalytic sets*: a set $\mathcal{A} \subseteq \mathbb{R}^n$ is $\mathbb{R}_{\mathrm{an}}$-definable if and only if the image of $\mathcal{A}$ under the isomorphism

$$\boldsymbol{x} \in \mathbb{R}^n \longmapsto \big(x_1(1+x_1^2)^{-1/2}, \ldots, x_n(1+x_n^2)^{-1/2}\big) \in (-1,1)^n \tag{B.1}$$

is subanalytic [14, Section 3] (we refer to [61] for the definition and properties of subanalytic sets). Finitely subanalytic sets can also be defined in a way similar to semialgebraic sets [16, Example 1.6]: an $\mathbb{R}_{\mathrm{an}}$-definable set in $\mathbb{R}^n$ is the union of finitely many sets of the form

$$\bigcap_{i\in[1\,..\,l]} \big\{ \boldsymbol{x} \in \mathbb{R}^n : f_i(\boldsymbol{x}) = 0,\ g_i(\boldsymbol{x}) > 0 \big\}, \tag{B.2}$$

where the functions $f_i, g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ are obtained by composition from

- the constant functions, the coordinate functions $\boldsymbol{x} \longmapsto x_m$, $m \in [1\,..\,n]$, addition, and multiplication,
- the functions $f : \mathbb{R}^m \longrightarrow \mathbb{R}$, $m \in \mathbb{N}$, whose restriction to $[-1,1]^m$ belongs to $\mathbf{an}([-1,1]^m)$ and which are identically zero outside $[-1,1]^m$,
- the extended reciprocal function $t \in \mathbb{R} \longmapsto t^{-1}$ if $t \neq 0$, 0 if $t = 0$.

The structure $\mathbb{R}_{\mathrm{an}}$ is *polynomially bounded* [62], meaning that for every $\mathbb{R}_{\mathrm{an}}$-definable function $f : \mathbb{R} \longrightarrow \mathbb{R}$, there exists $p \in \mathbb{N}$ such that $|f(t)| < t^p$ for all t sufficiently

large (in particular, exponential functions are not $\mathbb{R}_{\text{an}}$-definable). Among the potentials listed in Appendix A, the Andrew's sine function ϑ_{AS} is the only one that is $\mathbb{R}_{\text{an}}$- but not $\mathbb{R}_{\text{alg}}$-definable.

Example 9 (Analytic-exponential sets). Let $\mathbb{R}_{\text{an,exp}} := \langle \mathbb{R}_{\text{an}}, \{\exp\} \rangle$, where exp is the natural exponential function on $\mathbb{R}$. The structure $\mathbb{R}_{\text{an,exp}}$ is o-minimal [59]. Like $\mathbb{R}_{\text{an}}$, it can be described in the style of semialgebraic sets [16, Example 1.8]: the $\mathbb{R}_{\text{an,exp}}$-definable sets are those of the form (B.2) with the functions f_i and g_i obtained by composition from

- the same functions as for finitely subanalytic sets,
- the natural exponential function,
- the extended natural logarithm function $t \in \mathbb{R} \longmapsto \ln t$ if $t > 0$, 0 if $t \leqslant 0$.

All the potentials listed in Appendix A are $\mathbb{R}_{\text{an,exp}}$-definable.

B.2. **Some properties of definable sets and functions.** Definable sets and functions have the following basic properties (see, e.g., [14]).

- The interior and the closure of an $\mathcal{M}$-definable set are $\mathcal{M}$-definable.
- Let $f, g : \mathcal{A} \subseteq \mathbb{R}^m \longrightarrow \mathbb{R}$ be $\mathcal{M}$-definable. The sum $f + g$, the product fg, and the reciprocal $1/f$ (defined on $\mathcal{A} \setminus \{f = 0\}$) are $\mathcal{M}$-definable.
- A function $f = (f_1, \ldots, f_n) : \mathcal{A} \subseteq \mathbb{R}^m \longrightarrow \mathbb{R}^n$ is $\mathcal{M}$-definable if and only if its coordinate functions $f_1, \ldots, f_n : \mathcal{A} \longrightarrow \mathbb{R}$ are $\mathcal{M}$-definable.

Let $f : \mathcal{A} \subseteq \mathbb{R}^m \longrightarrow \mathbb{R}^n$ be $\mathcal{M}$-definable.

- If $\mathcal{A}' \subseteq \mathcal{A}$ is $\mathcal{M}$-definable, then the image $f(\mathcal{A}')$ and the restriction $f|_{\mathcal{A}'}$ are $\mathcal{M}$-definable.
- If $\mathcal{B} \subseteq \mathbb{R}^n$ is $\mathcal{M}$-definable, then the preimage $f^{-1}(\mathcal{B})$ is $\mathcal{M}$-definable.
- Let $g : \mathcal{B} \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^p$ be $\mathcal{M}$-definable. The composite $g \circ f : f^{-1}(\mathcal{B}) \longrightarrow \mathbb{R}^p$ is $\mathcal{M}$-definable.
- The set $\mathcal{A}' := \{\boldsymbol{x} \in \mathcal{A}^\circ : f \text{ is differentiable at } \boldsymbol{x}\}$ is $\mathcal{M}$-definable and the gradient $\nabla f : \mathcal{A}' \longrightarrow \mathbb{R}^m$ is $\mathcal{M}$-definable.

We recall below some important results on the connectedness of definable sets, the smoothness and monotonicity of definable functions on $\mathbb{R}$, and the local behavior of definable functions on $\mathbb{R}^n$.

Theorem B.4 (Connectedness [16, Corollary 3.2])**.** *Let $\mathcal{A} \subseteq \mathbb{R}^n$ be definable in an o-minimal structure $\mathcal{M}$ on $(\mathbb{R}, +, \cdot)$. Then $\mathcal{A}$ has finitely many connected components, each of which is $\mathcal{M}$-definable and path-connected.*

Remark 10 (Nondefinable functions). It follows from Theorem B.4 that the sine and cosine functions on $\mathbb{R}$ are nondefinable, as the intersections of their graphs with $\mathbb{R} \times \{0\}$ have infinitely many connected components (however, the restrictions of sin and cos to finite intervals are $\mathbb{R}_{\text{an}}$-definable). More generally, a function of class C^p on an open interval is nondefinable if any of its derivatives up to order p has infinitely many isolated zeros.

Theorem B.5 (Smoothness and monotonicity [16, Section 2])**.** *Let $\mathcal{M}$ be an o-minimal structure on $(\mathbb{R}, +, \cdot)$ and let $f : \mathcal{I} \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ be an $\mathcal{M}$-definable function, where $\mathcal{I}$ is a nondegenerate interval. For every $p \in \mathbb{N}$, there exists a finite partition of $\mathcal{I}$ into subintervals $\mathcal{I}_1, \ldots, \mathcal{I}_l$ such that for each $i \in [1\,..\,l]$, the restriction $f|_{(\mathcal{I}_i)^\circ}$ is of class C^p and either strictly monotone or constant.*

Theorem B.6 (KL inequality [12, Theorem 1]). *Let $f : \mathcal{O} \longrightarrow \mathbb{R}$ be a C^1 function, where $\mathcal{O}$ is a bounded open set in $\mathbb{R}^n$. Suppose f is definable in an o-minimal structure $\mathcal{M}$ on $(\mathbb{R}, +, \cdot)$ and $f(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathcal{O}$. Then there exist a positive constant α and a strictly increasing, $\mathcal{M}$-definable C^1 function $\varphi : (0, +\infty) \longrightarrow (0, +\infty)$ such that*

$$\|\nabla(\varphi \circ f)(\boldsymbol{x})\| \geqslant 1 \quad \text{for all } \boldsymbol{x} \in \mathcal{O} \cap \{f < \alpha\}. \tag{B.3}$$

If f is subanalytic, which is the case when f is $\mathbb{R}_{\text{an}}$-definable, then the following theorem shows that (B.3) holds with $\varphi(t) \propto t^{1/s}$ for some positive integer s.

Theorem B.7 (Łojasiewicz inequality [12, Theorem (ŁI)]). *Let $f : \mathcal{O} \longrightarrow \mathbb{R}$ be a subanalytic C^1 function, where $\mathcal{O}$ is a bounded open set in $\mathbb{R}^n$. Then there exist positive constants c and α and an integer $s \geqslant 2$ such that*

$$\|\nabla f(\boldsymbol{x})\| \geqslant c|f(\boldsymbol{x})|^{1-1/s} \quad \text{for all } \boldsymbol{x} \in \mathcal{O} \cap \{|f| \in (0, \alpha)\}. \tag{B.4}$$

Appendix C. Asymptotic behavior of the backward error criterion

Consider an inexact FHQ sequence $(\boldsymbol{x}^{(p)})_{p\in\mathbb{N}}$ constructed as in Theorem 7.2, and denote by $\boldsymbol{r}^{(j,p)}$ and $\boldsymbol{y}^{(j,p)}$ the jth iterate and residual, respectively, of the inner CG algorithm at the $(p+1)$th outer iteration. The backward error criterion (7.11) is then

$$\epsilon^{-1}\|\boldsymbol{r}^{(j,p)}\| \leqslant \|\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x}^{(p)})\boldsymbol{A}\|\,\|\boldsymbol{y}^{(j,p)}\| + \|\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x}^{(p)})\boldsymbol{a}\| =: \mathfrak{T}(j,p), \tag{C.1}$$

and the residual can be written as

$$\boldsymbol{r}^{(j,p)} = -\nabla\Theta(\boldsymbol{x}^{(p)}) - \boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x}^{(p)})\boldsymbol{A}(\boldsymbol{y}^{(j,p)} - \boldsymbol{x}^{(p)}). \tag{C.2}$$

Note that (C.1) is checked after the fixed minimum number of CG iterations has been performed (that is, when $j \geqslant J$), and that $\boldsymbol{r}^{(0,p)} = -\nabla\Theta(\boldsymbol{x}^{(p)})$ (since $\boldsymbol{y}^{(0,p)} = \boldsymbol{x}^{(p)}$ by construction). Also keep in mind that the norm of the residual does not necessarily decrease with j (see, e.g., [63, Section 2.3]).

The following propositions show that as p increases, the threshold $\mathfrak{T}(j,p)$ in (C.1) becomes increasingly independent of both j and p, and that consequently the backward error criterion is satisfied when the norm of the gradient of $\boldsymbol{x}^{(p)}$ is small enough. We assume that the set of stationary points of the objective Θ is level-discrete or that Θ is a KL function, so $(\boldsymbol{x}^{(p)})_p$ converges to a stationary point $\boldsymbol{x}^*$ by Theorem 4.9(i) or Theorem 5.5.

Proposition C.1 (Limit threshold of the backward error criterion). *Let*

$$\mathfrak{T}^* := \|\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x}^*)\boldsymbol{A}\|\,\|\boldsymbol{x}^*\| + \|\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x}^*)\boldsymbol{a}\|. \tag{C.3}$$

As $p \to \infty$, the threshold $\mathfrak{T}(j,p)$ in (C.1) *tends to $\mathfrak{T}^*$ independently of j, that is,*

$$\lim_p \sup_{j\geqslant J} |\mathfrak{T}(j,p) - \mathfrak{T}^*| = 0. \tag{C.4}$$

Proof. Since the maps $\boldsymbol{x} \longmapsto \|\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{A}\|$ and $\boldsymbol{x} \longmapsto \|\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x})\boldsymbol{a}\|$ are continuous, we need only show that

$$\lim_p \inf_{j\geqslant J} \|\boldsymbol{y}^{(j,p)}\| = \lim_p \sup_{j\geqslant J} \|\boldsymbol{y}^{(j,p)}\| = \|\boldsymbol{x}^*\|.$$

Let $p \in \mathbb{N}$ and $j \in [J\,..\,\infty)$. Since

$$\|\boldsymbol{y}^{(j,p)} - \boldsymbol{\Phi}_0(\boldsymbol{x}^{(p)})\| \leqslant \|\boldsymbol{y}^{(0,p)} - \boldsymbol{\Phi}_0(\boldsymbol{x}^{(p)})\|$$

(see [39, Theorem 6.3]), we have

$$\begin{aligned}\|\boldsymbol{y}^{(j,p)} - \boldsymbol{x}^{(p)}\| &\leqslant \|\boldsymbol{y}^{(0,p)} - \boldsymbol{\Phi}_0(\boldsymbol{x}^{(p)})\| + \|\boldsymbol{x}^{(p)} - \boldsymbol{\Phi}_0(\boldsymbol{x}^{(p)})\| \\ &= 2\|\boldsymbol{x}^{(p)} - \boldsymbol{\Phi}_0(\boldsymbol{x}^{(p)})\|.\end{aligned} \tag{C.5}$$

Hence, from Lemma 5.3, there is a constant $c > 0$ such that

$$\|\boldsymbol{y}^{(j,p)} - \boldsymbol{x}^{(p)}\| \leqslant c\big(\Theta(\boldsymbol{x}^{(p)}) - \Theta(\boldsymbol{\Phi}_0(\boldsymbol{x}^{(p)}))\big)^{1/2}.$$

Therefore

$$\lim_p \sup_{j \geqslant J} \|\boldsymbol{y}^{(j,p)} - \boldsymbol{x}^{(p)}\| \leqslant c\big(\Theta(\boldsymbol{x}^*) - \Theta(\lim_p \boldsymbol{\Phi}_0(\boldsymbol{x}^{(p)}))\big)^{1/2} = 0,$$

where the last equality follows from (4.2). Also,

$$\begin{aligned}\|\boldsymbol{x}^{(p)}\| - \sup_{j \geqslant J} \|\boldsymbol{y}^{(j,p)} - \boldsymbol{x}^{(p)}\| &\leqslant \inf_{j \geqslant J} \|\boldsymbol{y}^{(j,p)}\| \\ &\leqslant \sup_{j \geqslant J} \|\boldsymbol{y}^{(j,p)}\| \leqslant \|\boldsymbol{x}^{(p)}\| + \sup_{j \geqslant J} \|\boldsymbol{y}^{(j,p)} - \boldsymbol{x}^{(p)}\|.\end{aligned}$$

Taking the limit as $p \to \infty$ completes the proof. □

Proposition C.2 (Backward error control is eventually superfluous)**.** *There is a constant $\epsilon' \in (0, \epsilon)$ such that for p sufficiently large,* (C.1) *is satisfied for $j = J$ if*

$$\|\nabla\Theta(\boldsymbol{x}^{(p)})\| \leqslant \epsilon' \mathfrak{T}^*. \tag{C.6}$$

This result holds for any $\epsilon' \in (0, \epsilon)$ if $\|\boldsymbol{r}^{(J,p)}\| \leqslant \|\boldsymbol{r}^{(0,p)}\|$ for p sufficiently large.

Proof. From (C.2) and (C.5) we have

$$\|\boldsymbol{r}^{(J,p)}\| \leqslant \|\nabla\Theta(\boldsymbol{x}^{(p)})\| + 2\|\boldsymbol{A}^T\boldsymbol{E}(\boldsymbol{x}^{(p)})\boldsymbol{A}\|\,\|\boldsymbol{x}^{(p)} - \boldsymbol{\Phi}_0(\boldsymbol{x}^{(p)})\|,$$

and from the proof of Lemma 5.3 there is a constant $\tilde{c} > 0$ such that

$$\|\boldsymbol{x}^{(p)} - \boldsymbol{\Phi}_0(\boldsymbol{x}^{(p)})\| \leqslant \tilde{c}\|\nabla\Theta(\boldsymbol{x}^{(p)})\|.$$

So there is a constant $\tilde{c}' > 1$ independent of p such that

$$\|\boldsymbol{r}^{(J,p)}\| \leqslant \tilde{c}'\|\nabla\Theta(\boldsymbol{x}^{(p)})\|. \tag{C.7}$$

Let $\epsilon' \in (0, \epsilon/\tilde{c}')$. By Proposition C.1, we can assume that p is large enough so that $\tilde{c}'\epsilon'\mathfrak{T}^* \leqslant \epsilon\mathfrak{T}(J,p)$. Then, if (C.6) holds, we obtain from (C.7) that $\|\boldsymbol{r}^{(J,p)}\| \leqslant \epsilon\mathfrak{T}(J,p)$.

Now consider the case where $\|\boldsymbol{r}^{(J,p)}\| \leqslant \|\boldsymbol{r}^{(0,p)}\|$ for p sufficiently large. Let $\epsilon' \in (0, \epsilon)$. Using Proposition C.1 again, we have that $\epsilon\mathfrak{T}(J,p)$ is eventually greater than $\epsilon'\mathfrak{T}^*$. So if p is large enough and (C.6) holds, then

$$\|\boldsymbol{r}^{(J,p)}\| \leqslant \|\boldsymbol{r}^{(0,p)}\| = \|\nabla\Theta(\boldsymbol{x}^{(p)})\| \leqslant \epsilon'\mathfrak{T}^* \leqslant \epsilon\mathfrak{T}(J,p).$$

This completes the proof. □

References

[1] M. Robini, Y. Zhu, and J. Luo, *Edge-preserving reconstruction with contour-line smoothing and non-quadratic data-fidelity*, Inverse Probl. Imaging **7** (2013), no. 4, 1331–1366.

[2] M. Robini and Y. Zhu, *Generic half-quadratic optimization for image reconstruction*, SIAM J. Imaging Sci. **8** (2015), no. 3, 1752–1797.

[3] S. Li, *Markov random field modeling in computer vision*, Springer, 1995.

[4] R. He, W.-S. Zheng, T. Tan, and Z. Sun, *Half-quadratic-based iterative minimization for robust sparse representation*, IEEE Trans. Pattern Anal. Machine Intell. **36** (2014), no. 2, 261–275.

[5] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, *Deterministic edge-preserving regularization in computed imaging*, IEEE Trans. Image Process. **6** (1997), no. 2, 298–311.

[6] A. Delaney and Y. Bresler, *Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography*, IEEE Trans. Image Process. **7** (1998), no. 2, 204–221.

[7] J. Idier, *Convex half-quadratic criteria and interacting auxiliary variables for image restoration*, IEEE Trans. Image Process. **10** (2001), no. 7, 1001–1009.

[8] M. Nikolova and M. Ng, *Analysis of half-quadratic minimization methods for signal and image recovery*, SIAM J. Sci. Comput. **27** (2005), no. 3, 937–966.

[9] M. Allain, J. Idier, and Y. Goussard, *On global and local convergence of half-quadratic algorithms*, IEEE Trans. Image Process. **15** (2006), no. 5, 1130–1142.

[10] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, *On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision*, SIAM J. Imaging Sci. **8** (2015), no. 1, 331–372.

[11] R. Meyer, *Sufficient conditions for the convergence of monotonic mathematical programming algorithms*, J. Comput. System Sci. **12** (1976), no. 1, 108–121.

[12] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Ann. Inst. Fourier **48** (1998), no. 3, 769–783.

[13] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality*, Math. Oper. Res. **35** (2010), no. 2, 438–457.

[14] L. van den Dries and C. Miller, *Geometric categories and o-minimal structures*, Duke Math. J. **84** (1996), no. 2, 497–540.

[15] L. van den Dries, *Tame topology and o-minimal structures*, Cambridge University Press, 1998.

[16] ———, *O-minimal structures and real analytic geometry*, Current developments in mathematics, 1998, pp. 105–152.

[17] D. Hunter and K. Lange, *A tutorial on MM algorithms*, Amer. Statist. **58** (2004), no. 1, 30–37.

[18] T. Wu and K. Lange, *The MM alternative to EM*, Statist. Sci. **25** (2010), no. 4, 492–505.

[19] M. Jacobson and J. Fessler, *An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms*, IEEE Trans. Image Process. **16** (2007), no. 10, 2411–2422.

[20] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, *A majorize-minimize subspace approach for ℓ_2-ℓ_0 image regularization*, SIAM J. Imaging Sci. **6** (2013), no. 1, 563–591.

[21] H. Attouch, J. Bolte, and B. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Math. Program. Ser. A **137** (2013), no. 1, 91–129.

[22] J. Bolte, A. Daniilidis, and A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim. **17** (2007), no. 4, 1205–1223.

[23] J. Bochnak, M. Coste, and M.-F. Roy, *Real algebraic geometry*, Springer, 1998.

[24] J. Bolte, A. Daniilidis, and A. Lewis, *Tame functions are semismooth*, Math. Program. Ser. B **117** (2009), no. 1, 5–19.

[25] H. Whitney, *A function not constant on a connected set of critical points*, Duke Math. J. **1** (1935), no. 4, 514–517.

[26] L. van den Dries, A. Macintyre, and D. Marker, *Logarithmic-exponential power series*, J. Lond. Math. Soc. **56** (1997), no. 3, 417–434.

[27] P. Speissegger, *The Pfaffian closure on an o-minimal structure*, J. Reine Angew. Math. **508** (1999), 198–211.

[28] ———, *Pfaffian sets and o-minimality*, Lecture notes on o-minimal structures and real analytic geometry, 2012, pp. 179–217.
[29] G. Jones and P. Speissegger, *Generating the Pfaffian closure with total Pfaffian functions*, J. Logic Anal. **4** (2012).
[30] L. van den Dries and P. Speissegger, *The real field with convergent generalized power series*, Trans. Amer. Math. Soc. **350** (1998), no. 11, 4377–4421.
[31] ———, *The field of reals with multisummable series and the exponential function*, Proc. Lond. Math. Soc. **81** (2000), no. 3, 513–565.
[32] W. Hackbusch, *Iterative solution of large sparse systems of equations*, Applied Mathematical Sciences, vol. 95, Springer, 1994.
[33] Å. Björck, T. Elfving, and Z. Strakoš, *Stability of conjugate gradient and Lanczos methods for linear least squares problems*, SIAM J. Matrix Anal. Appl. **19** (1998), no. 3, 720–736.
[34] R. Barrett, M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst, *Templates for the solution of linear systems: building blocks for iterative methods*, Society for Industrial and Applied Mathematics, 1994.
[35] N. Higham, *Accuracy and stability of numerical algorithms*, 2nd ed., Society for Industrial and Applied Mathematics, 2002.
[36] Z. Strakoš and J. Liesen, *On numerical stability in large scale linear algebraic computations*, Z. Angew. Math. Mech. **85** (2005), no. 5, 307–325.
[37] J. Rigal and J. Gaches, *On the compatibility of a given solution with the data of a linear system*, J. Assoc. Comput. Mach. **14** (1967), no. 3, 543–548.
[38] Z. Strakoš and P. Tichý, *Error estimation in preconditioned conjugate gradients*, BIT **45** (2005), no. 4, 789–817.
[39] M. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Natl. Bur. Stand. **49** (1952), no. 6, 409–436.
[40] G. Golub and G. Meurant, *Matrices, moments and quadrature with applications*, Princeton University Press, 2010.
[41] G. Meurant and P. Tichý, *On computing quadrature-based bounds for the A-norm of the error in conjugate gradients*, Numer. Algorithms **62** (2013), no. 2, 163–191.
[42] A. Greenbaum, *Behavior of slightly perturbed Lanczos and conjugate-gradient recurrences*, Linear Algebra Appl. **113** (1989), 7–63.
[43] A. Greenbaum and Z. Strakoš, *Predicting the behavior of finite precision Lanczos and conjugate gradient computations*, SIAM J. Matrix Anal. Appl. **13** (1992), no. 1, 121–137.
[44] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, *Image quality assessment: from error visibility to structural similarity*, IEEE Trans. Image Process. **13** (2004), no. 4, 600–612.
[45] M. Black and A. Rangarajan, *On the unification of line processes, outlier rejection, and robust statistics with applications in early vision*, Int. J. Comput. Vis. **19** (1996), no. 1, 57–91.
[46] J.-F. Cai, R. Chan, L. Shen, and Z. Shen, *Convergence analysis of tight framelet approach for missing data recovery*, Adv. Comput. Math. **31** (2009), no. 1, 87–113.
[47] S. Ravishankar and Y. Bresler, *Learning sparsifying transforms*, IEEE Trans. Signal Processing **61** (2013), no. 5, 1072–1086.
[48] ———, *Learning doubly sparse transforms for images*, IEEE Trans. Image Process. **22** (2013), no. 12, 4598–4612.
[49] R. Acar and C. Vogel, *Analysis of bounded variation penalty method for ill-posed problems*, Inverse Problems **10** (1994), no. 6, 1217–1229.
[50] P. Green, *Bayesian reconstructions from emission tomography data using a modified EM algorithm*, IEEE Trans. Med. Imag. **9** (1990), no. 1, 84–93.
[51] K. Lange, *Convergence of EM image reconstruction algorithms with Gibbs priors*, IEEE Trans. Med. Imag. **9** (1990), no. 4, 439–446.
[52] M. Black, G. Sapiro, D. Marimont, and D. Heeger, *Robust anisotropic diffusion*, IEEE Trans. Image Process. **7** (1998), no. 3, 421–432.
[53] W. Li and J. Swetits, *The linear ℓ_1 estimator and the Huber M-estimator*, SIAM J. Optim. **8** (1998), no. 2, 457–475.
[54] T. Hebert and R. Leahy, *A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors*, IEEE Trans. Med. Imag. **8** (1989), no. 2, 194–202.

[55] S. Geman and D. McClure, *Bayesian image analysis: an application to single photon emission tomography*, Proc. stat. comput. section: annual meeting of the amer. stat. assoc., 1985Aug., pp. 12–18.
[56] J. Dennis and R. Welsch, *Techniques for nonlinear least squares and robust regression*, Commun. Stat. Simul. Comput. **7** (1978), no. 4, 345–359.
[57] Y. Leclerc, *Constructing stable descriptions for image partitioning*, Int. J. Comput. Vis. **3** (1989), 73–102.
[58] L. van den Dries, *A generalization of the Tarski-Seidenberg theorem, and some nondefinability results*, Bull. Amer. Math. Soc. (N.S.) **15** (1986), no. 2, 189–193.
[59] L. van den Dries and C. Miller, *On the real exponential field with restricted analytic functions*, Israel J. Math. **85** (1994), no. 1-3, 19–56.
[60] J.-P. Rolin, P. Speissegger, and A. Wilkie, *Quasianalytic Denjoy-Carleman classes and o-minimality*, J. Amer. Math. Soc. **16** (2003), no. 4, 751–777.
[61] E. Bierstone and P. Milman, *Semianalytic and subanalytic sets*, Publ. Math. Inst. Hautes Études Sci. **67** (1988), 5–42.
[62] C. Miller, *Expansions of the real field with power functions*, Ann. Pure Appl. Logic **68** (1994), no. 1, 79–94.
[63] G. Meurant, *The Lanczos and conjugate gradient algorithms: from theory to finite precision computations*, Society for Industrial and Applied Mathematics, 2006.

M. Robini and Y. Zhu are with the CREATIS laboratory (CNRS research unit UMR5220 and INSERM research unit U1206), INSA Lyon, 69621 Villeurbanne, France
E-mail address: marc.robini@creatis.insa-lyon.fr

F. Yang is with the School of Computer and Information Technology, Beijing Jiaotong University, no. 3 Shangyuancun, Haidian District, Beijing 100044, China
E-mail address: fengyang@bjtu.edu.cn