Upper-Bounding the Regularization Constant for Convex Sparse Signal Reconstruction

Renliang Gu
Iowa State University

Aleksandar Dogandžić
Iowa State University, ald@iastate.edu

Follow this and additional works at: https://lib.dr.iastate.edu/ece_pubs
Part of the Signal Processing Commons

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/ece_pubs/187. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Upper-Bounding the Regularization Constant for Convex Sparse Signal Reconstruction

Abstract
Consider reconstructing a signal $x$ by minimizing a weighted sum of a convex differentiable negative log-likelihood (NLL) (data-fidelity) term and a convex regularization term that imposes a convex-set constraint on $x$ and enforces its sparsity using $\ell_1$-norm analysis regularization. We compute upper bounds on the regularization tuning constant beyond which the regularization term overwhelmingly dominates the NLL term so that the set of minimum points of the objective function does not change. Necessary and sufficient conditions for irrelevance of sparse signal regularization and a condition for the existence of finite upper bounds are established. We formulate an optimization problem for finding these bounds when the regularization term can be globally minimized by a feasible $x$ and also develop an alternating direction method of multipliers (ADMM) type method for their computation. Simulation examples show that the derived and empirical bounds match.

Disciplines
Electrical and Computer Engineering | Signal Processing

Comments
This is the pre-print of the article Gu, Renliang, and Aleksandar Dogandžić. "Upper-Bounding the Regularization Constant for Convex Sparse Signal Reconstruction." arXiv preprint arXiv:1702.07930 (2017).

This article is available at Iowa State University Digital Repository: https://lib.dr.iastate.edu/ece_pubs/187
Upper-Bounding the Regularization Constant for Convex Sparse Signal Reconstruction

Renliang Gu and Aleksandar Dogandžić

Abstract—Consider reconstructing a signal $x$ by minimizing a weighted sum of a convex differentiable negative log-likelihood (NLL) (data-fidelity) term and a convex regularization term that imposes a convex-set constraint on $x$ and enforces its sparsity using $\ell_1$-norm analysis regularization. We compute upper bounds on the regularization tuning constant beyond which the regularization term overwhelmingly dominates the NLL term so that the set of minimum points of the objective function does not change. Necessary and sufficient conditions for irrelevance of sparse signal regularization and a condition for the existence of finite upper bounds are established. We formulate an optimization problem for finding these bounds when the regularization term can be globally minimized by a feasible $x$ and also develop an alternating direction method of multipliers (ADMM) type method for their computation. Simulation examples show that the derived and empirical bounds match.

Consider a convex NLL $\mathcal{L}(x)$ and a regularization term

$$r(x) = \|C(x) + \Psi^H x\|_1$$

(2)

that imposes a convex-set constraint on $x$, $x \in C \subseteq \mathbb{R}^p$, and sparsity of an appropriate linearly transformed $x$, where $\Psi \in \mathbb{C}^{p \times p}$ is a known sparsifying dictionary matrix. Assume that the NLL $\mathcal{L}(x)$ is differentiable and lower bounded within the closed convex set $C$, and satisfies

$$\text{dom} \mathcal{L}(x) \supseteq C$$

(3)

which ensures that $\mathcal{L}(x)$ is computable for all $x \in C$. Define the convex sets of solutions to $\min_{x} f_u(x)$, $\min_{x} r(x)$, and $\min_{x \in Q} \mathcal{L}(x)$:

$$X_u \triangleq \{ x \mid f_u(x) = \min_{x} f_u(x) \}$$

(4a)

$$Q \triangleq \{ x \mid r(x) = \min_{x} r(x) \}$$

(4b)

$$X^\circ \triangleq \{ x \mid x \in Q \mid \mathcal{L}(x) = \min_{x \in Q} \mathcal{L}(x) \} \neq \emptyset$$

(4c)

where the existence of $X^\circ$ is ensured by the assumption that $\mathcal{L}(x)$ is lower bounded in $C$.

We review the notation: $\ast^*$, $\ast^t$, $\ast^H$, $\ast^+$, $\| \cdot \|_p$, $\| \cdot \|_1$, $\ominus$, $\leq$, $\geq$, $I_N$, $1_{N \times 1}$, and $0_{N \times 1}$ denote complex conjugation, transpose, Hermitian transpose, Moore-Penrose matrix inverse, $\ell_p$-norm over the complex vector space $\mathbb{C}^N$ defined by $\| z \|_p = \sum_{i=1}^N |z_i|^p$ for $z = (z_i) \in \mathbb{C}^N$, absolute value, Kronecker product, elementwise versions of $\geq$ and $\leq$, the identity matrix of size $N$ and the $N \times 1$ vectors of ones and zeros, respectively (replaced by $I$, $1$, and $0$ when the dimensions can be inferred). $I_C(a) = \begin{cases} 0, & a \in C \\ +\infty, & \text{otherwise} \end{cases}$, $P_C(a) = \arg\min_{x \in C} \| x - a \|_2$, and $\exp_{a} \alpha$ denote the indicator function, projection onto $C$, and the elementwise exponential function: $[\exp_{a} \alpha]_i = \exp_{\alpha_i}$.

Denote by $\mathcal{N}(A)$ and $\mathcal{R}(A)$ the null space and range (column space) of a matrix $A$. These vector spaces are real or complex depending on whether $A$ is a real- or complex-valued matrix. For a set $S$ of complex vectors of size $p$, define $\text{Re} S \triangleq \{ s \in \mathbb{R}^p \mid s + j t \in S \text{ for some } t \in \mathbb{R}^p \}$ and $S \cap \mathbb{R}^p \triangleq \{ s \in \mathbb{R}^p \mid s + j 0 \in S \}$, where $j = \sqrt{-1}$. For $A \in \mathbb{C}^{M \times N}$,

$$\mathcal{N}(A^H) \cap \mathbb{R}^M = \mathcal{N}(A^T), \quad \text{Re}(\mathcal{R}(A)) = \mathcal{R}(A)$$

(5)

The authors are with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: \{renliang,ald\}@iastate.edu). This work was supported by the National Science Foundation under Grant CCF-1421480.
are the real null space and range of $A^T$ and $A$, respectively, where
\[ A \triangleq [\Re A \ \Im A] \in \mathbb{R}^{M \times 2N}. \] (6)
If $A$ in (6) has full row rank, we can define
\[ A^\dagger \triangleq A^H [\Re(AA^H)]^{-1} \] (7)
which reduces to $A^+$ for real-valued $A$. The following are equivalent: $\Re(\mathcal{R}(\Psi)) = \mathbb{R}^p$, $N(\Psi^H) \cap \mathbb{R}^p = \{0\}$, and $d = p$, where
\[ d \triangleq \dim(\Re(\mathcal{R}(\Psi))) \leq \min(p, 2p'). \] (8)

We can decompose $\Psi$ as
\[ \Psi = FZ \] (9)
where $F \in \mathbb{R}^{p \times d}$ and $Z \in \mathbb{C}^{d \times p'}$ with rank $F = d$ and rank $Z = d$; $Z = [\Re Z \ \Im Z] \in \mathbb{R}^{d \times 2p'}$, consistent with the notation in (6). Here, $\mathcal{R}(F)$ denotes the real range of the real-valued matrix $F$. Clearly, $d \geq 1$ is of interest; otherwise $\Psi = 0$. Observe that (see (7))
\[ \Re(\Psi Z^T) = F \] (10a)
\[ \mathcal{R}(F) = \Re(\mathcal{R}(\Psi)). \] (10b)

The subdifferential of the indicator function $N_C(x) = \partial \mathbb{L}_C(x)$ is the normal cone to $C$ at $x$ [7, Sec. 5.4] and, by the definition of a cone, satisfies
\[ N_C(x) = aN_C(x), \quad \text{for any } a > 0. \] (11)

Define
\[ G(s) \triangleq \begin{cases} \{s/|s|\}, & s \neq 0 \\ \{w \in \mathbb{C} \mid |w| \leq 1\}, & s = 0 \end{cases} \] (12)
and its elementwise extension $G(s)$ for vector arguments $s$, which can be interpreted as twice the Wirtinger subdifferential of $|s|_1$ with respect to $s$ [8]. Note that $s^H G(s) = \{|s|_1\}$, and, when $s$ is a real vector, $\Re(G(s))$ is the subdifferential of $|s|_1$ with respect to $s$ [9, Sec. 11.3.4].

**Lemma 1:** For $\Psi \in \mathbb{C}^{p \times p}$ and $x \in \mathbb{R}^p$, the subdifferential of $\|\Psi^H x\|_1$ with respect to $x$ is
\[ \partial_x \|\Psi^H x\|_1 = \Re(\Psi G(\Psi^H x)). \] (13)

**Proof:** (13) follows from
\[ \partial_x \|\Psi_j^H x\|_1 = \Re(\Psi_j G(\Psi_j^H x)) \] (14)
where $\psi_j$ is the $j$th column of $\Psi$. We obtain (14) by replacing the linear transform matrix in [10, Prop. 2.1] with $[\Re \psi_j \Im \psi_j]^T$.

We now use Lemma 1 to formulate the necessary and sufficient conditions for $x \in X_u$:
\[ 0 \in u \Re(\Psi G(\Psi^H x)) + \nabla \mathcal{L}(x) + N_C(x) \] (15a)
and $x \in Q$:
\[ 0 \in \Re(\Psi G(\Psi^H x)) + N_C(x) \] (15b)
respectively.

When the signal vector $x = \text{vec} \ X$ corresponds to an image $X \in \mathbb{R}^{J \times K}$, its isotropic and anisotropic total-variation (TV) regularizations correspond to [11, Sec. 2.1]
\[ \Psi = \Psi_v + j\Psi_h \in \mathbb{C}^{JK \times JK} \quad (\text{isotropic}) \] (16a)
\[ \Psi = [\Psi_v, \Psi_h] \in \mathbb{R}^{JK \times 2JK} \quad (\text{anisotropic}) \] (16b)
respectively, where $\Psi_v = I_K \otimes D^T(J)$ and $\Psi_h = D^T(K) \otimes I_J$ are the vertical and horizontal difference matrices (similar to those in [12, Sec. 15.3.3]), and
\[ D(L) \triangleq \begin{bmatrix} 1 & -1 & & & \\ -1 & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & \ddots \\ & & \ddots & \ddots & -1 \\ 0 & 0 & \ldots & 0 & 0 \end{bmatrix} \in \mathbb{R}^{L \times L} \] (17)
obtained by appending an all-zero row from below to the $(L-1) \times L$ upper-trapezoidal matrix with first row $[1, -1, 0, \ldots, 0]$; note that $D(1) = 0$. Here, $d = JK -1$ and
\[ N(\Psi^H) = \mathcal{R}(1) \] (18)
for both the isotropic and anisotropic TV regularizations.

The scenario where
\[ N(\Psi^H) \cap C \neq \emptyset \] (19)
holds is of practical interest: then $Q = N(\Psi^H) \cap C$ and $x^\circ \in \mathcal{X}^o$ globally minimize the regularization term: $r(x^\circ) = 0$. If (19) holds and $x^\circ \in \mathcal{X}^o$, then $G(\Psi^H x^\circ) = H$, where
\[ H \triangleq \{w \in \mathbb{C}^{p' \times 1} \mid \|w\|_\infty \leq 1\}. \] (20)

If, in addition to (19),
- $d = p$, then $\mathcal{X}^o = Q = \{0\}$;
- $N(\Psi^H) \cap \mathbb{R}^p = \mathcal{R}(1)$, then $Q = \mathcal{R}(1) \cap C$ and $x^\circ \in \mathcal{X}^o$ are constant signals of the form $x^\circ = 1x_0^\circ$, $x_0^\circ \in \mathbb{R}$.

In Section II, we define and explain an upper bound $U$ on useful regularization constants $u$ and establish conditions under which signal sparsity regularization is irrelevant and finite $U$ does not exist. We then present an optimization problem for finding $U$ when (19) holds (Section III), develop a general numerical method for computing bounds $U$ (Section IV), present numerical examples (Section V), and make concluding remarks (Section VI).

**II. UPPER BOUND DEFINITION AND PROPERTIES**

Define
\[ U \triangleq \inf \{ u \geq 0 \mid X_u \cap Q \neq \emptyset \}. \] (21)
If $X_u \cap Q = \emptyset$ for all $u$, then finite $U$ does not exist, which we denote by $U = +\infty$.

We now show that, if $u \geq U$, then the the set of minimum points $X_u$ of the objective function does not change.

**Remark 1:**
- (a) For any $u$, $X_u \cap Q = \mathcal{X}^o$ if and only if $X_u \cap Q \neq \emptyset$.
- (b) Assuming $X_u \cap Q \neq \emptyset$ for some $U \geq 0$, $X_u = \mathcal{X}^o$ for $u > U$. 

Proof: We first prove (a). Necessity follows by the existence of $\mathcal{X}^o$; see (4c). We argue sufficiency by contradiction. Consider any $x_u \in X_u \cap Q$; i.e., $x_u$ minimizes both $f_u(x)$ and $r(x)$. If $x_u \notin \mathcal{X}^o$, there exists a $y \in \mathcal{X}^o$ with $L(y) < L(x_u)$ that, by the definition of $\mathcal{X}^o$, also minimizes $r(x)$. Therefore, $f_u(y) = L(y) + ur(y) < f_u(x_u)$, which contradicts the assumption $x_u \in X_u$. Therefore, $X_u \cap Q \subseteq \mathcal{X}^o$. If there exists a $z \in \mathcal{X}^o \subseteq Q$ such that $z \notin X_u$, then $f_u(z) > f_u(x_u)$ which, since both $z$ and $x_u$ are in $Q$, implies that $L(z) > L(x_u)$ and contradicts the definition of $\mathcal{X}^o$. Therefore, $\mathcal{X}^o \subseteq X_u$.

We now prove (b). By (a), $X_U \cap Q = \mathcal{X}^o$, which confirms (b) for $u = U$. Consider now $u > U$, a $y \in X_U \cap Q = \mathcal{X}^o$, and any $x \in X_u$. Then,
\[
L(x) + ur(x) \geq L(y) + ur(y) \tag{22a}
\]
\[
L(y) + ur(y) \geq L(x) + ur(x) \tag{22b}
\]
By summing the two inequalities in (22) and rearranging, we obtain $r(y) \geq r(x)$. Since $y \in Q$, $x$ is also in $Q$; i.e., $X_u \subseteq Q$, which implies $X_u \subseteq \mathcal{X}^o$ by (a).

As $u$ increases, $X_u$ moves gradually towards $Q$ and, according to the definition (21), $X_u$ and $Q$ do not intersect when $u < U$. Once $u = U$, the intersection of the two sets is $\mathcal{X}^o$, and, by Remark 1(b), $X_u = \mathcal{X}^o$ for all $u > U$.

A. Irrelevant Signal Sparsity Regularization

Remark 2: The following claims are equivalent:

(a) $\mathcal{X}^o \cap X_0 \neq \emptyset$; i.e., there exists an $x^o \in \mathcal{X}^o$ such that
\[
0 \in \nabla L(x^o) + N_C(x^o); \tag{23}
\]

(b) $\mathcal{X}^o \subseteq X_0$; and

(c) $U = 0$; i.e., $X_0 \cap Q \neq \emptyset$.

Proof: (c) follows from (a) because $\mathcal{X}^o \subseteq Q$. (b) follows from (c) by applying Remark 1(a) to obtain $X_0 \cap Q = \mathcal{X}^o$, which implies (b). Finally, (b) implies (a).

Having $\nabla L(x^o) = 0$ for at least one $x^o \in \mathcal{X}^o$ implies (23) and is therefore a stronger condition than (23).

Example 1: Consider $L(x) = \|x\|_2^2$ and $C = \{x \in \mathbb{R}^2 \mid \|x - 1_{2 \times 1}\|_2 \leq 1\}$. (Here, $\mathcal{L}(x)$ could correspond to the Gaussian measurement model with measurements equal to zero.) Since $C$ is a circle within $\mathbb{R}^2_+$, the objective function is
\[
f_u(x) = \|x\|_2^2 + u|x_1 - x_2| + l_C(x) \tag{29a}
\]
with $X_u = \{x_u \in \mathbb{R}^2 \mid \|x - 1_{2 \times 1}\|_2 \leq 1\}$, $l_C(x) = \sqrt{1 + 4u^2 + 8u^2}$, and $\mathcal{X}^o = Q = \{1_{2 \times 1}\}$, which implies $U = +\infty$. Since (19) holds in this example, (25) is necessary and sufficient for $U = +\infty$. Since $-\nabla \mathcal{L}(x^o) = -4$ and $N_C(x^o) = \{a1 \mid a \leq 0\}$, (25) holds.

1) Two cases of finite $U$: If $d = p$ and (19) holds, then $U$ must be finite: in this case, condition (25) in Remark 3 cannot hold, which is easy to confirm by substituting $\mathcal{R}(\Psi) = \mathbb{R}^p$ into (25).

$U$ must also be finite if $\mathcal{X}^o \cap \text{int} C \neq \emptyset$. (30)

Indeed, (30) implies (19) and that for $x^o \in \mathcal{X}^o \cap \text{int} C$,
\[
N_C(x^o) = \{0\} \tag{31a}
\]
and hence (25) cannot hold if substituting (31a) and (31b). Here, (31b) follows from $0 \in \nabla \mathcal{L}(x^o) + N_C(x^o)$, the condition for optimality of the optimization problem $\min_{x \in Q} L(x)$ that defines $\mathcal{X}^o$, by using the fact that $N_C(x^o) = \mathcal{R}(\Psi)$ when $x^o \in \mathcal{X}^o \cap \text{int} C$.

If (30) holds then, by Remark 2, $U = 0$ if and only if
\[
\nabla \mathcal{L}(x^o) = 0. \tag{31b}
\]
III. BOUNDS WHEN (19) HOLDS

We now present an optimization problem for finding $U$ when (19) holds.

**Theorem 1:** Assume that (19) holds and that the convex NLL $L(x)$ is differentiable within $X^\circ$. Consider the following optimization problem:

\[
\begin{align*}
(P_0): \quad & U_0(x^\circ) = \min_{\alpha \in \mathbb{R}_+} ||p(x^\circ, \alpha, t)||_\infty \\
\text{subject to} \quad & \alpha \in N_C(x^\circ) \\
& \nabla L(x^\circ) + a \in R(F)
\end{align*}
\]

with

\[
p(x, \alpha, t) \triangleq t + Z^\dagger \{ F^+ [\nabla L(x) + a] - \text{Re}(Zt) \}.
\]

Then, $U_0(x^\circ) = U$ for all $x^\circ \in X^\circ$ and $U$ in (21).

We now present an optimization problem for finding $U$. We first prove that

\[
\text{Re}(\Psi^H x^\circ) = H \quad \text{for all} \quad x^\circ \in X^\circ
\]

and the condition

\[
\text{Re}(\Psi^H x^\circ) = H \quad \text{for all} \quad x^\circ \in X^\circ
\]

simplifies to

\[
U = \min_{\alpha \in N_C(x^\circ), t \in \mathbb{C}^r} \left| t + \text{Re}(\Psi^H x^\circ) - \text{Re}(Zt) \right|_\infty.
\]

**Proof:** (32a) and (32c) are equivalent to

\[
\text{Re}(\Psi^H x^\circ) = H \quad \text{for all} \quad x^\circ \in X^\circ
\]

and the condition

\[
\text{Re}(\Psi^H x^\circ) = H \quad \text{for all} \quad x^\circ \in X^\circ
\]

can be computed as

\[
U = \min_{\alpha \in N_C(x^\circ), t \in \mathbb{C}^r} \left| t + \text{Re}(\Psi^H x^\circ) - \text{Re}(Zt) \right|_\infty
\]

with any $x^\circ \in X^\circ \cap \text{int} C$.

**Proof:** Thanks to (30), (19) and (31a)–(31b) are satisfied, Theorem 1 applies, $U$ must be finite, and $\alpha = 0$ by (31a). By using these facts, we simplify (32) to obtain (41).

If $d = p$ and $0 \in \text{int} C$, then both Corollaries 1 and 2 apply and the bound $U$ can be obtained by setting $\alpha = 0$ and $N_C(0) = \{0\}$ in (40) or by setting $x^\circ = 0$ and $F = I$ in (41).

**Example 4:** Consider a real invertible $\Psi \in \mathbb{R}^{p \times p}$.

(a) If $C = \mathbb{R}_+^p$, Corollary 1 applies and (40) becomes

\[
U = \min_{\alpha \in N_C(0), t \in \mathbb{C}^r} \left| \text{Re}(\Psi^H x^\circ) + a \right|_\infty.
\]

(b) If $0 \in \text{int} C$, Corollaries 1 and 2 apply and the bound $U$ simplifies to

\[
U = \left| \text{Re}(\Psi^H x^\circ) \right|_\infty.
\]
Example 5 (One-dimensional TV regularization): Consider 1D TV regularization with $\Psi = D^T(p) \in \mathbb{R}^{p \times p}$ obtained by setting $K = 1, J = p$ in (16a); note that $d = p - 1$. Consider a constant signal $x^c = x_0^c \in X^c$. Then Theorem 1 applies and yields

$$U = \min_{a \in N_C(1)} \max_{1 \leq j < p} \left\{ \frac{1}{p} \left[ j \right] \cdot \left( \nabla L(x_0^c) + a \right) \right\},$$

(43a)

where we have used the factorization (9) with $F$ obtained by the block partitioning $F = [F_{0p \times 1} \ 0 \ \cdots \ 0_{(p-1) \times 1}]$, and the fact that $F^T$ is equal to the $(p - 1) \times p$ lower-triangular matrix of ones. When (30) holds, $1x_0^c \in X^c \cap \text{int} \mathcal{C}$, Corollary 2 applies, $a = 0$ (see (31a)), and (43a) reduces to:

$$U = \max_{1 \leq j < p} \left\{ \frac{1}{p} \cdot \left[ j \right] \cdot \left( \nabla L(x_0^c) \right) \right\}.$$  

(43b)

The bounds obtained by solving (P0) are often simple but restricted to the scenario where (19) holds. In the following section, we remove assumption (19) and develop a general numerical method for finding $U$ in (21).

IV. ADMM ALGORITHM FOR COMPUTING $U$

We focus on the nontrivial scenario where (23) does not hold and assume $u > 0$. We also assume that an $x^c \in X^c$ is available, which will be sufficient to obtain the $U$ in (21). We use the duality of norms [14, App. A.1.6]:

$$\|\Psi^H x\|_1 = \max_{\|w\|_{\infty} \leq 1} \text{Re}(w^H \Psi^H x)$$

(44)

to rewrite the minimization of (1) as the following min-max problem (see also (20)):

$$\min_{x \in \mathbb{C}^N} \max_{w \in \mathbb{C}^{N \times p}} \left\{ -v + \text{Re}(w^H \Psi^H x) + I_C(x) - I_H(w) \right\}.$$  

(45)

Since the objective function in (45) is convex with respect to $x$ and concave with respect to $w$, the optimal $(x, w) = (x_u, w_u)$ is at the saddle point of (45) and satisfies

$$0 \in \nabla L(x_u) + u \text{Re}(\Psi w_u) + N_C(x_u),$$

(46a)

$$w_u \in G(\Psi^H x_u).$$

(46b)

Now, select $U$ as the smallest $u$ for which (46a)–(46b) hold with $x_u = x^c$:

$$U = \frac{1}{v^c} \|\nabla L(x^c)\|_2.$$  

(47)

where $(v^c, w^c, t^c)$ is the solution to the following constrained linear programming problem:

$$\begin{array}{ll}
(P_1): & \min_{v, w, t} -v + I_G(\Psi^H x^c)(w) + I_N(x^c)(t) \\
& \text{subject to} \quad v g + \text{Re}(\Psi w) + t = 0
\end{array}$$

(48a)

obtained from (46a)–(46b) with $x_u$ and $w_u$ replaced by $x^c$ and $w$. Here,

$$g \triangleq \nabla L(x^c) / \|\nabla L(x^c)\|_2$$

(49)

is the normalized gradient (for numerical stability) of the NLL at $x^c$; $\nabla L(x^c) \neq 0$ because (23) does not hold. Due to (15b), $v = 0$ is a feasible point that satisfies the constraints (48b), which implies that $v^c \geq 0$. When (25) holds, $v$ has to be zero, implying $U = +\infty$.

To solve (P1) and find $v^c$, we apply an iterative algorithm based on alternating direction method of multipliers (ADMM) [15, 16]

$$u^{(i+1)} = \arg \min_{w \in G(\Psi^H x^c)} \|v^{(i)} g + \text{Re}(\Psi w) + t^{(i)} + z^{(i)}\|_2^2$$

(50a)

$$\begin{array}{ll}
|v^{(i)} - \rho g^T [\text{Re}(\Psi w^{(i+1)}) + v^{(i)}] + z^{(i)}| \\
\begin{array}{ll}
|t^{(i+1)} = P_{N_C(x^c)} (-v^{(i+1)} g - \text{Re}(\Psi w^{(i+1)}) - z^{(i)}) \\
\begin{array}{ll}
z^{(i+1)} = z^{(i)} + \text{Re}(\Psi w^{(i+1)}) + v^{(i)} g + t^{(i+1)}
\end{array}
\end{array}
\end{array}$$

(50c)

(50d)

where $\rho > 0$ is a tuning parameter for the ADMM iteration and we solve (50a) using the Broyden-Fletcher-Goldfarb-Shanno optimization algorithm with box constraints [17] and projected Nesterov’s proximal-gradient (PNPG) algorithm [18] for real and complex $\Psi$, respectively. We initialize the iteration (50) with $v^{(0)} = 1$, $t^{(0)} = 0$, $z^{(0)} = 0$, and $\rho = 1$, where $\rho$ is adaptively adjusted thereafter using the scheme in [15, Sec. 3.4.1].

In special cases, (50) simplifies. If (19) holds, then $\Psi^H x^c = 0$ and the constraint in (50a) simplifies to $|w|_{\infty} \leq 1$; see (20). If $\text{Re}(\Psi^H) = i \mathbb{C}$, $c > 0$, and $\Psi \in \mathbb{R}^{p \times p}$ or $\Psi \in \mathbb{C}^{p \times p}$, (50a) has the following analytical solution:

$$u^{(i+1)} = P_{G(\Psi^H x^c)} \left( -\frac{1}{c} \Psi^H (v^{(i)} g + t^{(i)} + z^{(i)}) \right).$$

(51)

When (30) holds, (50c) reduces to $t^{(i)} = 0$ for all $i$, thanks to (31a).

When $\Psi$ is real, the constraints imposed by $I_C(\Psi^H x^c)(w)$ become linear and (P1) becomes a linear programming problem with linear constraints.

V. NUMERICAL EXAMPLES

Matlab implementations of the presented examples are available at https://github.com/issucsp/ImgRecSrc/uboundEx. In all numerical examples, the empirical upper bounds $U$ were obtained by a grid search over $u$ with $X_u = \{x_u\}$ obtained using the PNPG method [18].

A. Signal reconstruction for Gaussian linear model

We adopt the linear measurement model with white Gaussian noise and scaled NLL $L(x) = 0.5\|y - \Phi x\|_2^2$, where the elements of the sensing matrix $\Phi \in \mathbb{R}^{N \times p}$ are independent, identically distributed (i.i.d.) and drawn from the uniform distribution on a unit sphere. We reconstruct the nonnegative “skyline” signal $x_{true} \in \mathbb{R}_+^{1024 \times 1}$ in [18, Sec. V-B] from noisy linear measurements $y$ using the discrete wavelet transform (DWT) and 1D TV regularizations, where the DWT matrix $\Psi$ is orthogonal ($\Psi^T = \Psi$), constructed using the Daubechies-4 wavelet with three decomposition levels. Define the signal-to-noise ratio (SNR) as

$$\text{SNR} (dB) = 10 \log_{10} \frac{\|\Phi x_{true}\|_2^2}{N \sigma^2}$$

(52)

where $\sigma^2$ is the variance of the Gaussian noise added to $\Phi x_{true}$ to create the noisy measurement vector $y$. 

TABLE I: Theoretical and empirical bounds $U$ for the linear Gaussian model.

| SNR/dB | $C = \mathbb{R}_+^d$, DWT theoretical | empirical | $C = \mathbb{R}^p$, DWT theoretical | empirical | $C = \mathbb{R}_+^d$, TV theoretical | empirical | $C = \mathbb{R}^p$, TV theoretical | empirical |
|--------|--------------------------------|------------|--------------------------------|------------|--------------------------------|------------|--------------------------------|------------|
| 30     | 8.87                          | 8.87       | 9.43                          | 9.43       | 101.55                         | 101.54     | 100.21                         | 100.21     |
| 20     | 8.91                          | 8.91       | 9.47                          | 9.47       | 96.47                          | 96.47      | 96.47                          | 96.47      |
| 10     | 9.03                          | 9.03       | 9.59                          | 9.59       | 92.49                          | 92.49      | 93.46                          | 93.46      |
| 0      | 9.43                          | 9.43       | 9.98                          | 9.98       | 87.49                          | 87.49      | 87.94                          | 87.94      |
| -10    | 11.88                         | 11.89      | 14.03                         | 14.02      | 152.07                         | 152.07     | same as                        |            |
| -20    | 27.77                         | 27.78      | 43.28                         | 43.28      | 361.56                         | 361.56     | same as                        |            |
| -30    | 88.78                         | 88.82      | 139.67                        | 139.66     | 1024.04                        | 1024.04    | 909.50                         | 909.48     |
| -30    | 77.29                         | 77.31      | 123.91                        | 123.90     | 683.43                         | 683.43     | 999.50                         | 999.48     |

TABLE II: Theoretical and empirical bounds $U$ for the PET example.

For $C = \mathbb{R}_+^d$ and $C = \mathbb{R}^p$ with DWT regularization, $\mathcal{X}^o = \{0\}$ and Example 4 applies and yields the upper bounds (42a) and (42b), respectively.

For TV regularization, we apply the result in Example 5. For $C = \mathbb{R}^p$ and $C = \mathbb{R}_+^d$, we have $\mathcal{X}^o = \{1 x_0\}$ and $\mathcal{X}^o = \{1 \max(x_0, 0)\}$, respectively, where

$$
x_0 \defas \arg \min_{x \in \mathbb{R}} \mathcal{L}(x) = 1^T \Phi^T y / \| \Phi 1 \|_2^2.
$$

If $x_0 \in \mathcal{X}^o$, which holds when $C = \mathbb{R}_+$ or when $C = \mathbb{R}^p$ and $x_0 = 0$, then the bound $U$ is given by (43b). For $C = \mathbb{R}_+$ and if $x_0 \leq 0$, then $\mathcal{X}^o = \{0\}$ and (43a) applies. In this case, $U = 0$ if $\nabla \mathcal{L}(0)_i \geq 0$ for $i = 1, \ldots, p - 1$, which occurs only when $\nabla \mathcal{L}(0) = 0$ for all $i$.

Table I shows the theoretical and empirical bounds for DWT and TV regularizations and $C = \mathbb{R}_+^d$ and $C = \mathbb{R}^p$; we decrease the SNR from 30dB to -30dB with independent noise realizations for different SNRs. The theoretical bounds in Sections III and IV coincide. For DWT regularization, $\mathcal{X}^o$ is the same for both convex sets $C$ and thus the upper bound $U$ for $C = \mathbb{R}_+^d$ is always smaller than its counterpart for $C = \mathbb{R}^p$, thanks to being optimized over variable $a$ in (42a). For TV regularization, when $x_0 > 0$, the upper bounds $U$ coincide for both $C$ because, in this case, $\mathcal{X}^o$ is the same for both $C$ and $\mathcal{X}^o \in C$. In the last row of Table I we show the case where $x_0 \leq 0$; then, $\mathcal{X}^o$ differs for the two convex sets $C$, and the upper bound $U$ for $C = \mathbb{R}_+$ is smaller than its counterpart for $C = \mathbb{R}^p$, thanks to being optimized over variable $a$ in (43a); compare (43a) with (43b).

B. PET image reconstruction from Poisson measurements

Consider positron emission tomography (PET) reconstruction of the $128 \times 128$ concentration map $x_{\text{true}}$ in [18, Fig. 3a], which represents simulated radiotracer activity in a human chest, from independent noisy Poisson-distributed measurements $y = (y_n)$ with means $\{\Phi x_{\text{true}} + b\}_n$. The choices of parameters in the PET system setup and concentration map $x_{\text{true}}$ have been taken from the Image Reconstruction Toolbox (IRT) [19, emission/em_test_setup.m]. Here,

$$
\mathcal{L}(x) = 1^T (\Phi x + b - y) + \sum_{n, y_n \neq 0} y_n \ln \frac{y_n}{\Phi x + b}_n
$$

and

$$
\Phi = w \text{diag}(\exp(-s \kappa + c)) S \in \mathbb{R}_+^{N \times p}
$$

is the known sensitizing matrix; $\kappa$ is the density map needed to model the attenuation of the gamma rays [20]; $b = (b_1)$ is the known intercept term accounting for background radiation, scattering effect, and accidental coincidence; $c$ is a known vector that models the detector efficiency variation; and $w > 0$ is a known scaling constant, which we use to control the expected total number of detected photons due to electron-positron annihilation, $1^T \mathbb{E}(y - b) = 1^T \Phi x_{\text{true}}$, an SNR measure. We collect the photons from 90 equally spaced directions over 180°, with 128 radial samples at each direction. Here, we adopt the parallel strip-integral matrix $S$ [21, Ch. 25.2] and use its implementation in the IRT [19].

We now consider the nonnegative convex set $C = \mathbb{R}_+^p$, which ensures that (3) holds, and 2D isotropic and anisotropic TV and DWT regularizations, where the 2D DWT matrix $\Psi$ is constructed using the Daubeches-6 wavelet with six decomposition levels.

For TV regularization, $\mathcal{X}^o = \{1 \max(0, x_0)\}$, where $x_0 = \arg \min_{x \in \mathbb{R}} \mathcal{L}(x)$, computed using the bisection method that finds the zero of $\partial \mathcal{L}(x)/\partial x$, which is an increasing function of $x \in \mathbb{R}_+$. Here, no search for $x_0$ is needed when $\partial \mathcal{L}(x)/\partial x|_{x=0} > 0$, because in this case $x_0 < 0$.

We computed the theoretical bounds using the ADMM-type algorithm in Section IV.

Table II shows the theoretical and empirical bounds for DWT and TV regularizations and the SNR $1^T \Phi x_{\text{true}}$ varying from $10^1$ to $10^9$, with independent measurement realizations for different SNRs.

$^2$ The elements of the intercept term have been set to a constant equal to 10% of the sample mean of $\Phi x_{\text{true}}$: $b = (1^T \Phi x_{\text{true}}/(10N)) 1$. 
Denote the isotropic and anisotropic 2D TV bounds by $U_{iso}$ and $U_{ani}$, respectively. Then, it is easy to show that when (19) holds, $U_{ani} \leq U_{iso} \leq \sqrt{2}U_{ani}$, which follows by using the inequalities $\sqrt{a^2 + b^2} \geq |a| + |b| \geq \sqrt{a^2 + b^2}$ and is confirmed in Table II.

VI. CONCLUDING REMARKS

Future work will include obtaining simple expressions for upper bounds $U$ for isotropic 2D TV regularization, based on Theorem 1.

REFERENCES

[1] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC: Winston, 1977.
[2] C. R. Vogel, *Computational Methods for Inverse Problems*. Philadelphia, PA: SIAM, 2002.
[3] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, “An interior-point method for large-scale $\ell_1$-regularized least squares,” *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, 2007.
[4] E. Hale, W. Yin, and Y. Zhang, “Fixed-point continuation for $\ell_1$-minimization: Methodology and convergence,” *SIAM J. Optim.*, vol. 19, no. 3, pp. 1107–1130, 2008.
[5] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, “Sparse reconstruction by separable approximation,” *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, 2009.
[6] E. Allgower and K. Georg, *Introduction to Numerical Continuation Methods*. Philadelphia, PA: SIAM, 2003.
[7] D. P. Bertsekas, *Convex Optimization Theory*. Belmont, MA: Athena Scientific, 2009.
[8] P. Bouboulis, K. Slavakis, and S. Theodoridis, “Adaptive learning in complex reproducing kernel Hilbert spaces employing Wirtinger’s subgradients,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 3, pp. 425–438, 2012.
[9] M. Kabanava and H. Raouhut, “Cosparsity in compressed sensing,” in *Compressed Sensing and Its Applications*, H. Boche, R. Calderbank, G. Kutyniok, and J. Vybiral, Eds. New York: Springer, 2015, pp. 315–339.
[10] Y. Wang, J. Yang, W. Yin, and Y. Zhang, “A new alternating minimization algorithm for total variation image reconstruction,” *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, 2008.
[11] A. Chambolle and T. Pock, “An introduction to continuous optimization for imaging,” *Acta Numer.*, vol. 25, pp. 161–319, May 2016.
[12] S. Boyd and L. Vandenberghe, *Vectors, Matrices, and Least Squares*, 2016. [Online]. Available: http://stanford.edu/class/ee103/mma.pdf (visited on 08/23/2016).
[13] M. Grant and S. Boyd. (Mar. 2014). CVX: Matlab software for disciplined convex programming, version 2.1. [Online]. Available: http://cvxr.com/cvx.
[14] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge Univ. Press, 2004.
[15] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
[16] M. Hong and Z.-Q. Luo. (Mar. 2013). On the linear convergence of the alternating direction method of multipliers. arXiv: 1208.3922 [math.OC].
[17] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, “A limited memory algorithm for bound constrained optimization,” *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, 1995.
[18] R. Gu and A. Dogandžić. (Oct. 2016). Projected Nesterov’s proximal-gradient algorithm for sparse signal reconstruction with a convex constraint. version 6. arXiv: 1502.02613 [stat.CO].
[19] J. A. Fessler, *Image reconstruction toolbox*, 2016. [Online]. Available: http://www.eecs.umich.edu/~fessler/code (visited on 08/23/2016).
[20] J. M. Ollinger and J. A. Fessler, “Positron-emission tomography,” *IEEE Signal Process. Mag.*, vol. 14, no. 1, pp. 43–55, 1997.
[21] J. A. Fessler, *Image Reconstruction*, 2009. [Online]. Available: http://web.eecs.umich.edu/~fessler/book/a-geom.pdf (visited on 08/23/2016).