LOWER BOUNDS OF NONZERO ENTRIES IN SOLUTIONS

XIAOJUN CHEN[*], LEI GUO[†], ZHAOSONG LU[‡], AND JANE J. YE[§]

Abstract. This is a supplementary material of paper [2].

1. Lower bound for nonzero entries of local minimizers. In [2], we consider the following non-Lipschitz nonlinear programming problem:

$$\begin{aligned}
\min \quad & f(x) + \Phi(x) \\
\text{s.t.} \quad & c(x) = 0, \\
& d(x) \leq 0,
\end{aligned} \qquad (1.1)$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $c : \mathbb{R}^n \to \mathbb{R}^m$ and $d : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions, and $\Phi : \mathbb{R}^n \to \mathbb{R}$ is a continuous function (possibly nonconvex non-Lipschitz).

As observed in the literature and our numerical experiments in [2], the local minimizers of problem (1.1) are often sparse when Φ is nonconvex and non-Lipschitz. This has made problem (1.1) a popular model in seeking a sparse solution (see, e.g., [1, 3]). In this report, we derive a lower bound on nonzero entries of any local minimizer of problem (1.1) under some suitable assumptions on its objective and constraint functions, which shows that the magnitude of all nonzero entries of a local minimizer must be above a certain positive number. This provides a possible interpretation why local minimizers of problem (1.1) are often sparse. To this end, we assume in this report that the function f is twice continuously differentiable, $c(x) := Ax - b$ with $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, and

$$d(x) := \begin{pmatrix} l - x \\ x - u \end{pmatrix} \text{ with } l, u \in \overline{\mathbb{R}}^n \text{ and } l < u. \qquad (1.2)$$

It is clear that $d(x) \leq 0$ is equivalent to $x \in [l, u]$.

The lower bounds on local minimizers have been derived in the literature for some special cases of problem (1.1) with the regularization term $\Phi(x) = \|x\|_q^q$ with $q \in (0, 1)$. In particular, Chen et al. [3] derived lower bounds for ℓ_q-regularized unconstrained convex quadratic programming. Lu [6] studied lower bounds for general ℓ_q-regularized unconstrained nonlinear programming. Recently, Chen et al. [1] developed lower bounds for the case where f is a convex quadratic function, $A = e^T$ or $[e^T - e^T]$ where e is the vector of all ones, $b = 1$, $m = 1$, $l = 0$ and $u = \infty$. For convenience of presentation, we first introduce some notations that will be used in this section only. We set the infimum or minimum over an empty set to be $+\infty$. We let $\mathcal{S}^*$ be the set of local minimizers of problem (1.1) and assume that $L_f := \sup_{x \in \mathcal{S}^*} \|\nabla^2 f(x)\|$ is

finite which clearly holds provided that either the gradient of f is globally Lipschitz continuous over $[l, u]$ or $\mathcal{S}^*$ is bounded.

Given a function $\varphi : \mathbb{R} \to \mathbb{R}$ and $a \in \mathbb{R}$, we define

$$\varphi^{\diamond}(a) := \inf \left\{ |t| \,\middle|\, \varphi(t) \geq a \right\}.$$

Given $\emptyset \neq \mathcal{I} \subseteq \{1, \ldots, n\}$, let $\mathcal{I}^c$ be the complement of $\mathcal{I}$ with respect to $\{1, \ldots, n\}$, and

$$\Gamma := (A_{\mathcal{I}})^+ A_{\mathcal{I}}, \quad \theta(\mathcal{I}) := \left\{ i \in \mathcal{I} \,\middle|\, \sum_{j \in \mathcal{I}} [(e_i - \Gamma_i)^T (e_j - \Gamma_j)]^2 > 0 \right\}, \tag{1.3}$$

$$\sigma(\mathcal{I}) := \left\{ i \in \mathcal{I} \,\middle|\, \Gamma_i = e_i \text{ and } e_i^T (A_{\mathcal{I}})^+ (b - \sum_{j \in \mathcal{I}^c} A_j x_j) \neq 0 \text{ for some } x_j \in \{l_j, 0, u_j\}\ j \in \mathcal{I}^c \right\}, \tag{1.4}$$

$$\bar{\lambda}_i(\mathcal{I}) := -\frac{L_f \sum\limits_{j \in \mathcal{I}} [(e_i - \Gamma_i)^T (e_j - \Gamma_j)]^2}{\|e_i - \Gamma_i\|^4} \qquad i \in \theta(\mathcal{I}) \neq \emptyset,$$

$$\alpha(\mathcal{I}) := \min_{i \in \theta(\mathcal{I})} (\phi_i'')^{\diamond} \left(\bar{\lambda}_i(\mathcal{I}) \right), \tag{1.5}$$

$$\beta(\mathcal{I}) := \min_{i \in \sigma(\mathcal{I})} \left\{ |e_i^T (A_{\mathcal{I}})^+ (b - \sum_{j \in \mathcal{I}^c} A_j x_j)| \,\middle|\, e_i^T (A_{\mathcal{I}})^+ (b - \sum_{j \in \mathcal{I}^c} A_j x_j) \neq 0,\ x_j \in \{l_j, 0, u_j\}\ j \in \mathcal{I}^c \right\}, \tag{1.6}$$

$$\delta := \min \left\{ \min_{\mathcal{I} \subseteq \{1, \ldots, n\}} \{\alpha(\mathcal{I}), \beta(\mathcal{I})\}, \min_{u_i \neq 0} |u_i|, \min_{l_i \neq 0} |l_i| \right\}, \quad \lambda_0 := \min_{\mathcal{I} \subseteq \{1, \ldots, n\}} \min_{i \in \theta(\mathcal{I})} \bar{\lambda}_i(\mathcal{I}). \tag{1.7}$$

Here, $A_{\mathcal{I}}^+$ denotes the Moore–Penrose pseudoinverse of the matrix $A_{\mathcal{I}}$.

It is not hard to observe that δ and λ_0 only depend on the data of problem (1.1). Moreover, we can see that $\lambda_0 < 0$ if $\lambda_0 \neq \infty$.

Before establishing the main result of this section, we make some further assumptions on Φ that will be used in this section only.

ASSUMPTION 1.1. *Assume that $\Phi(x) := \sum_{i=1}^n \phi_i(x_i)$ with $\phi_i : \mathbb{R} \to \mathbb{R}$ lower semi-continuous and for each $i = 1, \ldots, n$, ϕ_i is twice continuously differentiable everywhere except at 0 and*

$$\phi_i''(t) < 0 \quad \forall t \neq 0, \qquad \limsup_{t \to 0} \phi_i''(t) < \lambda_0,$$

where λ_0 is defined in (1.7).

It is easy to verify that the following widely used regularization functions satisfy Assumption 1.1. Note that the bridge penalty function is non-Lipschitz and it is locally Lipschitz everywhere expect at 0 while both the logistic penalty function and the fraction penalty function are locally Lipschitz everywhere.

(i) Bridge penalty [4] with $\phi_i(t) = \lambda|t|^q$ for any $q \in (0, 1)$ and $\lambda > 0$;

(ii) Logistic penalty [5] with $\phi_i(t) = \log(1 + \lambda|t|)$ for any $\lambda \begin{cases} > \sqrt{-\lambda_0} & \text{if } \lambda_0 < 0, \\ \geq 0 & \text{if } \lambda_0 = \infty; \end{cases}$

(iii) Fraction penalty [7] with $\phi_i(t) = \frac{\lambda|t|}{1+\lambda|t|}$ for any $\lambda \begin{cases} > \sqrt{\frac{-\lambda_0}{2}} & \text{if } \lambda_0 < 0, \\ \geq 0 & \text{if } \lambda_0 = \infty. \end{cases}$

Under Assumption 1.1, we can show $\delta > 0$ as follows.

PROPOSITION 1.1. *Let δ be defined in* (1.7). *If Assumption 1.1 holds, then $\delta > 0$.*

Proof. By the definition of δ, one can easily see $\delta \geq 0$. It suffices to show $\delta \neq 0$. Suppose for contradiction that $\delta = 0$. Notice that $\alpha(\mathcal{I}) \geq 0$ and $\beta(\mathcal{I}) > 0$ for any $\mathcal{I} \subseteq \{1, \ldots, n\}$. These together with the definition of δ and the assumption $\delta = 0$ imply that $\alpha(\mathcal{I}) = 0$ for some $\mathcal{I} \subseteq \{1, \ldots, n\}$. It then follows from (1.5) that there exists some $i_0 \in \theta(\mathcal{I})$ such that $(\phi_{i_0}'')^\diamond(\bar{\lambda}_{i_0}(\mathcal{I})) = 0$, which along with the definitions of $(\phi_{i_0}'')^\diamond$ and λ_0 implies that there exists a sequence $\{t_k\}$ converging to 0 such that $\phi_{i_0}''(t_k) \geq \bar{\lambda}_{i_0}(\mathcal{I}) \geq \lambda_0$. This contradicts the second relation in Assumption 1.1. $\square$

We are now ready to establish the main result of this section. The result shows that δ defined in (1.7) is a lower bound for nonzero entries of local minimizers of problem (1.1).

THEOREM 1.1. *Suppose that Assumption 1.1 holds. Let δ be defined in* (1.7) *and x^* a local minimizer of problem* (1.1). *Then for any i, there holds $|x_i^*| \geq \delta > 0$ if $x_i^* \neq 0$.*

Proof. Without loss of generality, we assume for simplicity that $l, u \in \mathbb{R}^n$. Let $\mathcal{I}_l = \{i \mid x_i^* = l_i\}$, $\mathcal{I}_u = \{i \mid x_i^* = u_i\}$ and

$$\mathcal{I} = \big\{i \mid x_i^* \notin \{l_i, 0, u_i\}\big\}, \quad \Theta = (I - \Gamma)[\nabla^2 f(x^*)_{\mathcal{I}\mathcal{I}}](I - \Gamma),$$

where I is the $|\mathcal{I}| \times |\mathcal{I}|$ identity matrix and Γ is defined as in (1.3). In addition, let the associated $\theta(\mathcal{I})$, $\alpha(\mathcal{I})$ and $\beta(\mathcal{I})$ be defined according to (1.3), (1.5) and (1.6), respectively.

Without loss of generality, assume that $x^* = (l_{\mathcal{I}_l}, x_{\mathcal{I}}^*, 0, u_{\mathcal{I}_u})$. Since x^* is a local minimizer of problem (1.1), one can observe that $x_{\mathcal{I}}^*$ is a local minimizer of the problem

$$
\begin{aligned}
\min \quad & f(l_{\mathcal{I}_l}, x_{\mathcal{I}}, 0, u_{\mathcal{I}_u}) + \sum_{i \in \mathcal{I}} \phi_i(x_i) \\
\text{s.t.} \quad & A_{\mathcal{I}} x_{\mathcal{I}} = b - \sum_{j \in \mathcal{I}_l} A_j l_j - \sum_{j \in \mathcal{I}_u} A_j u_j.
\end{aligned}
\tag{1.8}
$$

Notice that the objective function of problem (1.8) is twice continuously differentiable at $x_{\mathcal{I}}^*$. By the second-order optimality condition, we have

$$d^T[\nabla^2 f(x^*)_{\mathcal{I}\mathcal{I}} + \mathrm{Diag}(\phi''(x_{\mathcal{I}}^*))]d \geq 0 \quad \forall d \in \{d \mid A_{\mathcal{I}} d = 0\}, \tag{1.9}$$

where $\phi''(x_{\mathcal{I}}^*) := (\phi_i''(x_i^*) : i \in \mathcal{I})$. Since $I - (A_{\mathcal{I}})^+ A_{\mathcal{I}}$ is the orthogonal projector onto the null space of $A_{\mathcal{I}}$, we have

$$\{d \mid A_{\mathcal{I}} d = 0\} = \{d \mid d = (I - (A_{\mathcal{I}})^+ A_{\mathcal{I}})w, \ w \in \mathbb{R}^{|\mathcal{I}|}\},$$

which together with (1.9) implies

$$w^T \left(I - (A_{\mathcal{I}})^+ A_{\mathcal{I}}\right) [\nabla^2 f(x^*)_{\mathcal{I}\mathcal{I}} + \mathrm{Diag}(\phi''(x_{\mathcal{I}}^*))] \left(I - (A_{\mathcal{I}})^+ A_{\mathcal{I}}\right) w \geq 0 \quad \forall w \in \mathbb{R}^{|\mathcal{I}|}.$$

This inequality can be rewritten as

$$w^T \Theta w + w^T (I - \Gamma)\mathrm{Diag}(\phi''(x_{\mathcal{I}}^*))(I - \Gamma)w \geq 0 \quad \forall w \in \mathbb{R}^{|\mathcal{I}|}.$$

Letting $w = e_i - \Gamma_i$ and using this inequality and $\phi_i''(x_i^*) < 0$ for $i \in \mathcal{I}$, we see

$$(e_i - \Gamma_i)^T \Theta(e_i - \Gamma_i) + \phi_i''(x_i^*)\|e_i - \Gamma_i\|^4 \geq 0 \quad i \in \mathcal{I}. \tag{1.10}$$

3

Suppose that $|x_i^*| > 0$ for some i. If $x_i^* \in \{l_i, u_i\}$, it immediately follows from the definition of δ that $|x_i^*| \geq \delta$ holds. Otherwise, we have $i \in \mathcal{I}$. We next show that $|x_i^*| \geq \delta$ also holds by considering two separate cases as follows.

Case 1): $\Gamma_i \neq e_i$. Noting that $\|\nabla^2 f(x^*)_{\mathcal{II}}\| \leq L_f$, it follows from (1.10) that

$$0 < (e_i - \Gamma_i)^T \Theta (e_i - \Gamma_i) \leq L_f \sum_{j \in \mathcal{I}} [(e_i - \Gamma_i)^T (e_j - \Gamma_j)]^2. \qquad (1.11)$$

This together with (1.3) implies that $i \in \theta(\mathcal{I}) \neq \emptyset$. Moreover, (1.10) and (1.11) imply

$$\phi_i''(x_i^*) \geq -\frac{L_f \sum_{j \in \mathcal{I}} [(e_i - \Gamma_i)^T (e_j - \Gamma_j)]^2}{\|e_i - \Gamma_i\|^4}.$$

By the definitions of $(\phi_i'')^\diamond$ and $\alpha(\mathcal{I})$, we obtain that

$$|x_i^*| \geq (\phi_i'')^\diamond \left(-\frac{L_f \sum_{j \in \mathcal{I}} [(e_i - \Gamma_i)^T (e_j - \Gamma_j)]^2}{\|e_i - \Gamma_i\|^4} \right) \geq \alpha(\mathcal{I}).$$

It then follows from this inequality and the definition of δ that $|x_i^*| \geq \delta$ holds.

Case 2): $\Gamma_i = e_i$. It follows from the relation $Ax^* = b$ that

$$A_{\mathcal{I}} x_{\mathcal{I}}^* = b - \sum_{j \in \mathcal{I}_l} A_j l_j - \sum_{j \in \mathcal{I}_u} A_j u_j.$$

Pre-multiplying this relation by $e_i^T A_{\mathcal{I}}^+$ on both sides and using $\Gamma = (A_{\mathcal{I}})^+ A_{\mathcal{I}}$, we have

$$e_i^T \Gamma x_{\mathcal{I}}^* = e_i^T A_{\mathcal{I}}^+ (b - \sum_{j \in \mathcal{I}_l} A_j l_j - \sum_{j \in \mathcal{I}_u} A_j u_j),$$

which together with $\Gamma_i = e_i$ and $x_i^* \neq 0$ yields

$$x_i^* = e_i^T A_{\mathcal{I}}^+ (b - \sum_{j \in \mathcal{I}_l} A_j l_j - \sum_{j \in \mathcal{I}_u} A_j u_j) \neq 0$$

and hence $i \in \sigma(\mathcal{I}) \neq \emptyset$, where $\sigma(\mathcal{I})$ is defined in (1.4). It follows from these relations and (1.6) that $|x_i^*| \geq \beta(\mathcal{I})$, which together with the definition of δ implies that $|x_i^*| \geq \delta$.

The proof is complete since we observe that $\delta > 0$ by Proposition 1.1. $\square$

REFERENCES

[1] C. CHEN, X. LI, C. TOLMAN, S. WANG AND Y. YE, *Sparse portfolio selection via quasi-norm regularization*, ArXiv preprint, arXiv:1312.6350, 2014.
[2] X. CHEN, L. GUO, Z. LU AND J. YE, *A Feasible augmented Lagrangian method for non-Lipschitz nonconvex programming*, Preprint, 2016.

[3] X. CHEN, F. XU AND Y. YE, *Lower bound theory of nonzero entries in solutions of l_2-l_p minimization*, SIAM J. Sci. Comput., 32 (2010), pp. 2832–2852.

[4] W.J. FU, *Penalized regression: The bridge versus the lasso*, J. Comput. Graph. Stat., 7 (1998), pp. 397–416.

[5] D. GEMAN AND G. REYNOLDS, *Constrained restoration and the recovery of discontinuities*, IEEE Trans. Pattern Anal. Mach. Intell., 14 (1992), pp. 357–383.

[6] Z. LU, *Iterative reweighted minimization methods for l_p regularized unconstrained nonlinear programming*, Math. Program., 147 (2014), pp. 277–307.

[7] M. NIKOLOVA, M.K. NG, S. ZHANG AND W. CHING, *Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization*, SIAM J. Imaging Sci., 1 (2008), 2–25.