COMMUN. MATH. SCI.
Vol. 15, No. 1, pp. 197–211

LITTLE-o CONVERGENCE RATES FOR SEVERAL ALTERNATING MINIMIZATION METHODS*

TAO SUN† AND LIZHI CHENG‡

Abstract. Alternating minimization is an efficient method for solving convex minimization problems whose objective function is a sum of a differentiable function and a separable nonsmooth function. Variants and extensions of the alternating minimization method have been developed in recent years. In this paper, we consider the convergence rate of several existing alternating minimization schemes. We improve the proven big-O convergence rate of these algorithms to little-o under an error bound condition which is actually quite common in many applications. We also investigate the convergence of a variant of alternating minimization proposed in this paper.

Keywords. Alternating minimizations, little-o convergence rate, error bound.

AMS subject classifications. 90C30, 90C26, 47N10.

1. Introduction

Let H, f, and g all be closed, proper, and convex functions. This paper considers the following minimization problem

$$\min_{y\in\mathcal{R}^{N_1},z\in\mathcal{R}^{N_2}} \Phi(y,z)=f(y)+H(y,z)+g(z), \tag{1.1}$$

where H is a differentiable function but f and g may or may not be. Such a model has many applications in signal and image processing, machine learning research, statistical problem, to name a few. The authors in [19] provide plenty of examples.

An old but classical and effective method for problem (1.1) is the Alternating Minimization scheme (AM) [16]. In other literature, this algorithm is also called as coordinate descent method or Gauss–Seidel method [7]. A recent survey on the coordinate descent method is shared by [21]. This method fixes one of y and z in each iteration, and then minimizes the other one. The author in [2] first investigates the convergence rate of the AM method for problem (1.1). A sublinear convergence rate is given in [2] under some trivial assumptions on f, g, and H. In a latter paper [18], the authors point out the drawbacks of AM method for problem (1.1): it solves a minimization problem in each iteration, the stopping criterion is hard to determine, and error accumulates. In view of this, paper [18] proposes the Proximal Alternating Linearized Minimization (PALM). This algorithm solves a linear approximation which is actually the proximal map of f or g; and then, the subproblem in each iteration can be exactly solved if the proximal maps of f and g are easily calculated. The authors in [18] also prove the sublinear convergence of PALM.

It is natural to think of combining AM and PALM, i.e., in each iteration, we use the linearized methodology for one of y and z rather than both. Such an idea is inspired

*Received: November 20, 2015; accepted (in revised form): May 23, 2016. Communicated by Jack Xin.
This research was supported by the National Natural Science Foundation of Hunan Province, China (13JJ2001), and the Science Project of National University of Defense Technology (JC120201), and National Science Foundation of China (No. 61402495).

†College of Science, National University of Defense Technology, Changsha, 410073, Hunan, P.R. China (nudtsuntao@163.com).

‡College of Science & The State Key Laboratory for High Performance Computation, National University of Defense Technology, Changsha, 410073, Hunan, P.R. China (clzcheng@nudt.edu.cn).

by the following problem which reads as

$$\min_{y\in\mathcal{R}^{N_1},z\in\mathcal{R}^{N_2}} \lambda_1\|y\|_1+\|Ay-z\|_2^2+\lambda_2\|z\|_{1,2}, \tag{1.2}$$

where $\lambda_1,\lambda_2>0$, and $\|\cdot\|_1$ and $\|\cdot\|_{1,2}$ are the ℓ_1 norm and $\ell_{1,2}$ norm, respectively. When z is fixed, minimizing the function above needs iteration; therefore, we use the PALM strategy. While when y is fixed, the minimization of the function is quite simple. We call such algorithm as Half Proximal Alternating Linearized Minimization (HPALM) algorithm.

In the paper, we investigate the convergence rates of the three schemes above when function satisfies some error bound. Such an error bound is quite common for the polyhedral minimization. With the summable sequence techniques proposed in [10–12], we improve the proved big-O rate AM and PALM to little-o rate. We also prove the little-o rate of HPALM.

2. Assumptions and preliminaries

We adopt the notation described here throughout the paper. Let x^1 and x^2 be two elements of the space $\mathcal{R}^N$. Let x_i be ith component of x. The inner product of x^1 and x^2 is defined as $\langle x^1,x^2\rangle:=\sum_{i=1}^N x_i^1x_i^2$. The length of x is denoted by $\|x\|_2:=\sqrt{\langle x,x\rangle}$. For a convex set $Q\subseteq\mathcal{R}^N$, the projection of vector x to Q is denoted as $\mathrm{Proj}_Q(x):=\arg\min_{w\in Q}\|x-w\|_2$. Let $\mathrm{dom}(J)$ be the domain of J, i.e., $\mathrm{dom}(J):=\{x\in\mathcal{R}^N|J(x)>-\infty\}$. Through the paper, we follow the convention used in [2] that $x\in R^{N_1+N_2}$ which is generated by

$$x=(y,z). \tag{2.1}$$

All partial derivatives corresponding to y will be denoted as $\nabla_y J(x)$, and so is $\nabla_z J(x)$. Then, we have that

$$\nabla J(x)=(\nabla_y J(x),\nabla_z J(x)).$$

2.1. Mathematical preliminaries.

Here, we collect some basic definitions in the following.

DEFINITION 2.1 (See [15]). *If a function $J(x)$ is convex and differentiable, we call J has continuous Lipschitz gradient $L=L(J)$, provided*

$$\|\nabla J(u)-\nabla J(v)\|_2\le L(J)\|u-v\|_2, u,v\in\mathrm{dom}(J). \tag{2.2}$$

LEMMA 2.1 (See [15]). *If J is convex and differentiable and has continuous Lipschitz gradient L, we have that*

$$J(u)\le J(v)+\langle\nabla J(v),u-v\rangle+\frac{L}{2}\|u-v\|_2^2. \tag{2.3}$$

DEFINITION 2.2 (See [9,17]). *If a function $J(x)$ is convex, the proximal map of $J(x)$ is defined as*

$$\boldsymbol{Prox}_J(x):=\arg\min_w\{J(w)+\frac{1}{2}\|x-w\|_2^2\}. \tag{2.4}$$

It is easy to see that $\boldsymbol{Prox}_J(x)=(I+\partial J)^{-1}(x)$.

LEMMA 2.2 (See [2]). *Let J be a proper, convex and closed function. Then, $w = \boldsymbol{Prox}_{\frac{J}{M}}(x)$ if and only if for any $u \in \mathrm{dom}(J)$*

$$J(u) \geq J(w) + M\langle x-w, u-w\rangle,$$

where $M>0$.

LEMMA 2.3 (See [2,6]). *Let $\{\alpha_k\}_{k\geq 1}$ be nonnegative sequence of real numbers satisfying*

$$\alpha_k - \alpha_{k+1} \geq \gamma\alpha_{k+1}^2. \tag{2.5}$$

Then, for any $k\geq 2$, we have that

$$\alpha_k \leq \max\{\left(\frac{1}{2}\right)^{\frac{k-1}{2}}\alpha_0, \frac{4}{\gamma(k-1)}\}. \tag{2.6}$$

LEMMA 2.4 (Summable sequence convergence rate [12]). *Suppose that nonnegative scalar sequences $\{\xi_j\}_{j=1,2,\ldots}$ and $\{a_j\}_{j=1,2,\ldots}$ satisfy $\sum_j \xi_j a_j < +\infty$. Let $\Sigma_k := \sum_{j=0}^k \xi_j$, $k\geq 1$. If $\{\xi_j\}_{j=1,2,\ldots}$ be monotonically nonincreasing, then*

$$a_k = o(\frac{1}{\Sigma_k - \Sigma_{\lceil \frac{k}{2}\rceil}}). \tag{2.7}$$

In particular, if $\xi_j \geq \epsilon j^p$ for some $p>0$ and $\epsilon>0$, then, $a_k = o(\frac{1}{k^{p+1}})$.

The summable sequence convergence rate (2.7) has been proved by Davis and Yin in [12]. Here, we use it to prove the particular case. The proof is simple and presented here just for completeness's sake.

Proof. Note that

$$\begin{aligned}
\Sigma_k - \Sigma_{\lceil \frac{k}{2}\rceil} &\geq \epsilon \sum_{j=\lceil \frac{k}{2}\rceil}^{k} j^p \\
&\geq \epsilon \sum_{j=\lceil \frac{k}{2}\rceil}^{k} \int_{j-1}^{j} t^p dt \\
&= \epsilon \sum_{j=\lceil \frac{k}{2}\rceil}^{k} \frac{j^{p+1}-(j-1)^{p+1}}{p+1} \\
&= \Omega(k^{p+1}).
\end{aligned}$$

Then, we have

$$0 \leq a_k\Omega(k^{p+1}) \leq a_k(\Sigma_k - \Sigma_{\lceil \frac{k}{2}\rceil}) \to 0. \tag{2.8}$$

That is actually $a_k = o(\frac{1}{k^{p+1}})$. ∎

REMARK 2.1. The summable sequence convergence technique is introduced by Davis and Yin for analyzing several famous splitting schemes [12]; amazingly, the authors present some new results for such well-studied classical schemes in general case. In latter papers [10, 11], the technique is applied to some other methods. Recently, the authors in [1] presents a little-o convergence rate for the FISTA [4]; in fact, they also used such a technique.

2.2. Assumptions and conditions. We collect several assumptions on f, g and H as follows:

A.1: The functions $f:\mathcal{R}^{N_1}\mapsto(-\infty,+\infty]$ and $g:\mathcal{R}^{N_2}\mapsto(-\infty,+\infty]$ are closed and proper convex, and the function H is continuously differentiable convex over $\mathrm{dom}(f)\times\mathrm{dom}(g)$.

A.2: For any $x^0\in\mathrm{dom}(f)\times\mathrm{dom}(g)$, the level set

$$\mathcal{L}_\Phi(x^0):=\{x|\Phi(x)\leq\Phi(x^0)\} \tag{2.9}$$

is compact.

A.3: There exists $L_1\in(0,+\infty)$ such that

$$\|\nabla_y H(y_1,z)-\nabla_y H(y_2,z)\|_2\leq L_1\|y_1-y_2\|_2,$$

for any $y_1,y_2\in\mathrm{dom}(f)$ and $z\in\mathrm{dom}(g)$.

A.4: There exists $L_2\in(0,+\infty)$ such that

$$\|\nabla_z H(y,z_1)-\nabla_z H(y,z_2)\|_2\leq L_2\|z_1-z_2\|_2,$$

for any $y\in\mathrm{dom}(f)$ and $z_1,z_2\in\mathrm{dom}(g)$.

From A.3 and A.4, for any $x^1=(y^1,z^1),x^2=(y^2,z^2)\in\mathrm{dom}(f)\times\mathrm{dom}(g)$, we have that

$$\begin{aligned}\|\nabla H(x^1)-\nabla H(x^2)\|_2&\leq\|\nabla_y H(x^1)-\nabla_y H(x^2)\|_2+\|\nabla_z H(x^1)-\nabla_z H(x^2)\|_2\\&\leq\|\nabla_y H(x^1)-\nabla_y H(y^1,z^2)+\nabla_y H(y^1,z^2)-\nabla_y H(x^2)\|_2\\&\quad+\|\nabla_z H(x^1)-\nabla_z H(y^1,z^2)+\nabla_z H(y^1,z^2)-\nabla_z H(x^2)\|_2\\&\leq(L_1+L_2)\|y^1-y^2\|_2+(L_1+L_2)\|z^1-z^2\|_2\\&\leq\sqrt{2}(L_1+L_2)\|x^1-x^2\|_2.\end{aligned}$$

That means H has a continuous Lipschitz gradient $\sqrt{2}(L_1+L_2)$ over $\mathrm{dom}(f)\times\mathrm{dom}(g)$. We use

$$L=\sqrt{2}(L_1+L_2) \tag{2.10}$$

through the paper.

Q-Sufficient Descent Condition

DEFINITION 2.3. *Let $Q\subseteq\mathcal{R}^N$ be a closed convex set. We call that function J satisfies Q-Sufficient Descent condition (Q-SD), if, for any $x\in Q$,*

$$J(x)-J^*\geq\nu\|x-Proj_{\chi^*}(x)\|_2^2, \tag{2.11}$$

where χ^* is the solution set of $\min_x J(x)$, and J^* is the minimum of J, and $\nu>0$. Further, if $Q=\mathcal{R}^N$, we it as Global Sufficient Descent condition (GSD).

The Q-SD can be regarded as a modification of the error bound which was proposed in [14]. Although not explicitly proposing the definition, the authors in [3] have already employed GSD to prove the linear convergence of Away-Step conditional gradient method. In fact, the GSD has a deep relationship with the Restricted Strongly Convexity (RSC) [13]; the RSC implies GSD(we will provide a brief proof of this in the appendix). In the following, we provide two examples to demonstrate that a wide class of function which owns the same form as the objective function in problem (1.1) satisfies Q-SD/GSD.

Example 1. If κ is a strongly convex function, and S and D are two polyhedral sets, and $\lambda_1,\lambda_2 \geq 0$. Then,

$$\Phi(y,z) = \kappa(Ay + Bz) + \lambda_1 \delta_S(y) + \lambda_2 \delta_D(z)$$

satisfies the GSD condition, where $\delta_S(y)$ and $\delta_D(z)$ are the well-known indicator functions. We can immediately obtain this by using [3, Lemma 2.5].

Example 2. If κ_1 and κ_2 are strongly convex functions and $\lambda_1,\lambda_2 \geq 0$. Then,

$$\Phi(y,z) = \kappa_1(Ay + Bz) + \kappa_2(Cy + Dz) + \lambda_1 \|y\|_1 + \lambda_2 \|z\|_1$$

satisfies the Q-SD condition, where $Q = \{x \mid \|x\|_2 = \|(y,z)\|_2 \leq R\}$ and R is any positive sufficiently large constant. We will provide a proof for this example in the appendix. If Φ satisfies assumption A.2, R can be chosen large such that $\mathcal{L}_\Phi(x^0) \subseteq \mathbf{B}(0,R)$. Therefore, the function satisfies $\mathcal{L}_\Phi(x^0)$-SD in this case.

Remark 2.2. Example 1 and Example 2 have many applications in engineering. In [2], the author proposes using Example 2 to solve the following problem

$$\min_z \|Tz\|_1 + \|b - Az\|_2^2. \tag{2.12}$$

Actually the author focuses on the penalty version of Equation (2.12) rather than the original one, i.e.,

$$\min_{y,z} \rho \|Tz - y\|_2^2 + \|b - Az\|_2^2 + \|y\|_1, \tag{2.13}$$

where $\rho > 0$ is the penalty parameter.

2.3. Schemes. In this subsection, we present the specific schemes of the algorithms mentioned in Section 1. From any starting point $x^0 = (y^0, z^0) \in \mathrm{dom}(f) \times \mathrm{dom}(g)$, the AM method updates $x^{k+1} = (y^{k+1}, z^{k+1})$ as

$$y^{k+1} \in \arg\min_y H(y, z^k) + f(y), \tag{2.14a}$$

$$z^{k+1} \in \arg\min_z H(y^{k+1}, z) + g(z). \tag{2.14b}$$

The KKT condition for the second relation gives that

$$\mathbf{0} \in \nabla_z H(y^{k+1}, z^{k+1}) + \partial g(z^{k+1}).$$

Thus, for any $\gamma > 0$, it holds that

$$z^{k+1} = (I + \partial g/\gamma)^{-1}(z^{k+1} - \nabla_z H(y^{k+1}, z^{k+1})/\gamma). \tag{2.15}$$

The PALM algorithm solves the linearized approximation of the problems above, i.e.,

$$y^{k+1} \in \arg\min_y \langle \nabla_y H(y^k, z^k), y - y^k \rangle + \frac{\gamma_k}{2} \|y - y^k\|_2^2 + f(y), \tag{2.16a}$$

$$z^{k+1} \in \arg\min_z \langle \nabla_z H(y^{k+1}, z^k), z - z^k \rangle + \frac{\lambda_k}{2} \|z - z^k\|_2^2 + g(z). \tag{2.16b}$$

For PALM, the KKT condition gives that

$$\mathbf{0}\in\nabla_y H(y^k,z^k)+\gamma_k(y^{k+1}-y^k)+\partial f(y^{k+1}), \tag{2.17a}$$
$$\mathbf{0}\in\nabla_z H(y^{k+1},z^k)+\lambda_k(z^{k+1}-z^k)+\partial g(z^{k+1}). \tag{2.17b}$$

Therefore, we can derive that

$$y^{k+1}=(I+\frac{\partial f}{\gamma_k})^{-1}(y^k-\frac{1}{\gamma_k}\nabla_y H(y^k,z^k)), \tag{2.18a}$$
$$z^{k+1}=(I+\frac{\partial g}{\lambda_k})^{-1}(z^k-\frac{1}{\lambda_k}\nabla_z H(y^{k+1},z^k)). \tag{2.18b}$$

The HPALM combines the two algorithms above; it has the following two forms

$$y^{k+1}\in\arg\min_y\langle\nabla_y H(y^k,z^k),y-y^k\rangle+\frac{\gamma_k}{2}\|y-y^k\|_2^2+f(y), \tag{2.19a}$$
$$z^{k+1}\in\arg\min_z H(y^{k+1},z)+g(z), \tag{2.19b}$$

and

$$y^{k+1}\in\arg\min_y H(y,z^k)+f(y), \tag{2.20a}$$
$$z^{k+1}\in\arg\min_z\langle\nabla_z H(y^{k+1},z^k),z-z^k\rangle+\frac{\lambda_k}{2}\|z-z^k\|_2^2+g(z). \tag{2.20b}$$

We call the schemes above as HPALM-I and HPALM-II, respectively.

3. Main results

In this part, we present the convergence results of AM, PALM, and HPALM, respectively. We first present a theorem as follows.

THEOREM 3.1. *Assume that x^0 is the starting point and Φ given in problem (1.1) satisfies the $\mathcal{L}_\Phi(x^0)$-SD condition. Let $\{x^k\}_{k\geq 1}$ satisfy*

$$\Phi(x^k)-\Phi(x^{k+1})\geq\frac{\pi}{\mathrm{dist}^2(x^{k+\tau},\chi^*)}(\Phi(x^{k+1})-\Phi^*)^2, \tag{3.1}$$

where χ^* is the solution set of problem (1.1), Φ^* is the minimum of Φ, τ is an integer, and $\pi>0$. Then,

$$\Phi(x^k)-\Phi^*=o(\frac{1}{k}). \tag{3.2}$$

Proof. It is easy to see that x^k and $\mathrm{Proj}_{\chi^*}(x^k)$ all belong to the level set $\mathcal{L}_\Phi(x^0)$. Let

$$R=\max_{x\in\mathcal{L}_\Phi(x^0)}\{\|x-\mathrm{Proj}_{\chi^*}(x)\|\}.$$

From assumption A.2, $R<+\infty$. Thus, Equation (3.1) turns to

$$\Phi(x^k)-\Phi^*-[\Phi(x^{k+1})-\Phi^*]\geq\frac{\pi}{R^2}(\Phi(x^{k+1})-\Phi^*)^2.$$

Note that $\Phi(x^k)-\Phi^*\geq 0$, from Lemma 2.3, we derive that $\Phi(x^k)-\Phi^*\leq O(\frac{1}{k})$. If Φ satisfies the $\mathcal{L}_\Phi(x^0)$-SD condition, we can obtain

$$\Phi(x^k)-\Phi^*\geq\nu\cdot\mathrm{dist}^2(x^k,\chi^*) \tag{3.3}$$

due to the fact that $\{\Phi(x^k)\}_{k=0,1,2,\dots}$ is decreasing. Therefore, we have $\text{dist}^2(x^k,\chi^*)\leq O(\frac{1}{k})$. For a fixed τ, we can further obtain that $\frac{1}{\text{dist}^2(x^{k+\tau},\chi^*)}\geq \mu k$ for some $\mu>0$. Obviously, $\Phi(x^k)-\Phi^*$ is monotonically nonincreasing and

$$\sum_{k=1}^{\ell}\frac{1}{\text{dist}^2(x^{k+\tau},\chi^*)}(\Phi(x^{k+1})-\Phi^*)^2\leq \frac{1}{\pi}(\Phi(x^{\ell+1})-\Phi(x^1))<+\infty. \tag{3.4}$$

Then, from Lemma 2.4, we have that

$$(\Phi(x^k)-\Phi^*)^2=o(\frac{1}{k^2}). \tag{3.5}$$

□

In the following part, we only need to prove that the sequences generated by AM, PALM, and HPALM all satisfy Equation (3.1). The proof is motivated by the methodology presented in [2] and [18], which contains following two main steps:

1. Find a $\rho_1>0$ and function $r(x^{k+i_1},x^{k+i_2},\dots,x^{k+i_\alpha})\geq 0$ such that

$$\Phi(x^{k+1})-\Phi(x^*)\leq \rho_1 r(x^{k+i_1},x^{k+i_2},\dots,x^{k+i_\alpha})\cdot\text{dist}(x^{k+\tau},x^*),\forall k=1,2\dots,$$

   where $\alpha\in\mathbb{N}$ and $i_1,i_2,\dots,i_\alpha\in\mathbb{N}$.
2. For $k=0,1,\dots$, find another positive constant ρ_2 and $r(x^{k+i_1},x^{k+i_2},\dots,x^{k+i_\alpha})\geq 0$ such that

$$\Phi(x^k)-\Phi(x^{k+1})\geq \rho_2\cdot r(x^{k+i_1},x^{k+i_2},\dots,x^{k+i_\alpha})^2.$$

Combining the two relations above, we can obtain that

$$\Phi(x^k)-\Phi(x^{k+1})\geq \frac{\rho_2}{\rho_1^2}\frac{1}{\text{dist}^2(x^{k+\tau},x^*)}(\Phi(x^{k+1})-\Phi(x^*))^2. \tag{3.6}$$

Through the proofs in the following, we use the convention

$$\overline{x^k}:=\text{Proj}_{\chi^*}(x^k)\ \text{ and }\ \overline{x^k}=(\overline{y^k},\overline{z^k}). \tag{3.7}$$

3.1. AM scheme.

LEMMA 3.1. *Let $\{x^k=(y^k,z^k)\}_{k=1,2,\dots}$ be generated by the AM method for problem (1.1). Then, for any $k\in\mathbb{N}$,*

$$\Phi(x^{k+1})-\Phi^*\leq L\|y^k-\textbf{\textit{Prox}}_{f/L}[y^k-\frac{1}{L}\nabla_y H(y^k,z^k)]\|_2\cdot\text{dist}(x^k,\chi^*), \tag{3.8}$$

where L is defined in Equation (2.10).

Proof. Denote that $h(x)=f(y)+g(z)$. Considering the point $(p_1^k,p_2^k)=p^k:=\mathbf{Prox}_{h/L}[x^k-\frac{1}{L}\nabla H(x^k)]$. Then, we have that $p_1^k=\mathbf{Prox}_{f/L}(y^k-\frac{1}{L}\nabla_y H(x^k))$ and $p_2^k=\mathbf{Prox}_{g/L}(z^k-\frac{1}{L}\nabla_z H(x^k))$. From Equation (2.15), we see that

$$p_2^k=z^k. \tag{3.9}$$

Then, the scheme of AM method then gives that

$$H(p_1^k,z^k)+f(p_1^k)\geq H(y^{k+1},z^k)+f(y^{k+1}). \tag{3.10}$$

Then, we have that

$$\begin{aligned}\Phi(p^k) &= H(p_1^k, p_2^k) + f(p_1^k) + g(p_2^k)\\ &= H(p_1^k, z^k) + f(p_1^k) + g(z^k)\\ &\geq H(y^{k+1}, z^k) + f(y^{k+1}) + g(z^k) = \Phi(y^{k+1}, z^k).\end{aligned} \tag{3.11}$$

From the scheme of AM method, we also see that

$$\Phi^* \leq \Phi(y^{k+1}, z^{k+1}) \leq \Phi(y^{k+1}, z^k) \leq \Phi(y^k, z^k). \tag{3.12}$$

Thus, we have that

$$\Phi(x^{k+1}) - \Phi^* \leq \Phi(y^{k+1}, z^k) - \Phi^* \leq \Phi(p^k) - \Phi^*. \tag{3.13}$$

In Lemma 2.2, letting $x = x^k - \frac{1}{L}\nabla H(x^k)$, $u = \overline{x^k}$, $J = h$, and $M = L$, we then obtain that

$$h(\overline{x^k}) \geq h(p^k) + L\langle x^k - \frac{1}{L}\nabla H(x^k) - p^k, \overline{x^k} - p^k\rangle. \tag{3.14}$$

From Lemma 2.1, we derive that

$$H(p^k) - H(\overline{x^k}) \leq H(x^k) + \langle \nabla H(x^k), p^k - x^k\rangle + \frac{L}{2}\|p^k - x^k\|_2^2 - H(\overline{x^k}). \tag{3.15}$$

Combining Equations (3.13), (3.14), and (3.15), we have that

$$\begin{aligned}\Phi(x^{k+1}) - \Phi^* &\leq \Phi(p^k) - \Phi^*\\ &= H(p^k) + h(p^k) - \Phi^*\\ &\leq H(x^k) + h(\overline{x^k}) + L\langle x^k - p^k, p^k - \overline{x^k}\rangle + \langle \nabla H(x^k), \overline{x^k} - x^k\rangle\\ &\quad + \frac{L}{2}\|p^k - x^k\|_2^2 - \Phi^*\\ &\leq H(\overline{x^k}) + h(\overline{x^k}) - \Phi^* + L\langle x^k - p^k, p^k - \overline{x^k}\rangle + \frac{L}{2}\|p^k - x^k\|_2^2\\ &= L\langle x^k - p^k, x^k - \overline{x^k}\rangle - \frac{L}{2}\|p^k - x^k\|_2^2\\ &\leq L\langle x^k - p^k, x^k - \overline{x^k}\rangle\\ &\leq L\|x^k - p^k\|_2 \cdot \|x^k - \overline{x^k}\|_2 = L\|x^k - p^k\|_2 \cdot \mathrm{dist}(x^k, \chi^*).\end{aligned} \tag{3.16}$$

The third inequality is based on the that

$$\langle \nabla H(x^k), \overline{x^k} - x^k\rangle \leq H(\overline{x^k}) - H(x^k). \tag{3.17}$$

From Equation (3.9), we have $\|x^k - p^k\|_2 = \|y^k - p_1^k\|_2$. ∎

LEMMA 3.2. *Let $\{x^k = (y^k, z^k)\}_{k=1,2,\ldots}$ be generated by the AM method for problem (1.1). Then, for any $k \in \mathbb{N}$,*

$$\Phi(x^k) - \Phi(x^{k+1}) \geq \frac{L}{2}\|y^k - \boldsymbol{Prox}_{f/L}[y^k - \frac{1}{L}\nabla_y H(y^k, z^k)]\|_2. \tag{3.18}$$

Proof. From Equation (3.12), we have that

$$\Phi(x^k) - \Phi(x^{k+1}) \geq \Phi(y^k, z^k) - \Phi(y^{k+1}, z^k)$$

$$\geq H(y^k,z^k)+f(y^k)-H(y^{k+1},z^k)-f(y^{k+1}).$$

We using the same notations employed in last lemma. Noth that y^{k+1} minimizes $H(y,z^k)+f(y)$, then, we have that

$$\begin{aligned}&H(y^k,z^k)+f(y^k)-H(y^{k+1},z^k)-f(y^{k+1})\\ \geq\, &H(y^k,z^k)+f(y^k)-H(p_1^k,z^k)-f(p_1^k)\geq \frac{L}{2}\|y^k-p_1^k\|_2^2.\end{aligned} \tag{3.19}$$

The second inequality depends on [5, Lemma 2.3]. □

3.2. PALM scheme.

LEMMA 3.3. *Assume that* $\max_k\{\gamma_k,\lambda_k\}<+\infty$. *Let* $\{x^k=(y^k,z^k)\}_{k=1,2,\ldots}$ *be generated by the PALM method for problem (1.1). Then, for any* $k\in\mathbb{N}$,

$$\Phi(x^{k+1})-\Phi^*\leq \rho_1\|x^{k+1}-x^k\|_2\cdot \mathrm{dist}(x^{k+1},\chi^*), \tag{3.20}$$

where $\rho_1=\sqrt{2}(\max_k\{\gamma_k,\lambda_k\}+L)$.

Proof. Applying Lemma 2.2 to Equation (2.17) with $u=\overline{y^{k+1}}$ and $u=\overline{z^{k+1}}$, we obtain that

$$f(\overline{y^{k+1}})\geq f(y^{k+1})+\gamma_k\langle y^k-y^{k+1},\overline{y^{k+1}}-y^{k+1}\rangle+\langle \nabla_y H(y^k,z^k),y^{k+1}-\overline{y^{k+1}}\rangle \tag{3.21}$$

and

$$g(\overline{z^{k+1}})\geq g(z^{k+1})+\lambda_k\langle z^k-z^{k+1},\overline{z^{k+1}}-z^{k+1}\rangle+\langle \nabla_z H(y^{k+1},z^k),z^{k+1}-\overline{z^{k+1}}\rangle. \tag{3.22}$$

The convexity of H gives that

$$H(x^{k+1})-H(\overline{x^{k+1}})\leq \langle \nabla H(x^{k+1}),x^{k+1}-\overline{x^{k+1}}\rangle. \tag{3.23}$$

Summing the inequalities above, we obtain that

$$\begin{aligned}\Phi(x^{k+1})-\Phi(\overline{x^{k+1}})\leq\, &\gamma_k\langle y^{k+1}-y^k,\overline{y^{k+1}}-y^{k+1}\rangle+\langle \nabla_y H(x^{k+1})-\nabla_y H(y^k,z^k),\\ &y^{k+1}-\overline{y^{k+1}}\rangle+\lambda_k\langle z^{k+1}-z^k,\overline{z^{k+1}}-z^{k+1}\rangle\\ &+\langle \nabla_z H(x^{k+1})-\nabla_z H(y^{k+1},z^k),z^{k+1}-\overline{z^{k+1}}\rangle.\end{aligned} \tag{3.24}$$

With the Schwartz's inequality, we have

$$\begin{aligned}\gamma_k\langle y^{k+1}-y^k,\overline{y^{k+1}}-y^{k+1}\rangle &\leq \gamma_k\|y^{k+1}-y^k\|_2\cdot\|\overline{y^{k+1}}-y^{k+1}\|_2\\ &\leq \max_k\{\gamma_k\}\cdot\|x^{k+1}-x^k\|_2\cdot\|\overline{y^{k+1}}-y^{k+1}\|_2.\end{aligned} \tag{3.25}$$

Similarly, we also have

$$\lambda_k\langle z^{k+1}-z^k,\overline{z^{k+1}}-z^{k+1}\rangle\leq \max_k\{\lambda_k\}\cdot\|x^{k+1}-x^k\|_2\cdot\|\overline{z^{k+1}}-z^{k+1}\|_2. \tag{3.26}$$

We also have that

$$\begin{aligned}&\langle \nabla_y H(x^{k+1})-\nabla_y H(y^k,z^k),y^{k+1}-\overline{y^{k+1}}\rangle\\ \leq\, &\|\nabla_y H(x^{k+1})-\nabla_y H(y^k,z^k)\|_2\cdot\|y^{k+1}-\overline{y^{k+1}}\|_2\\ \leq\, &\|\nabla H(x^{k+1})-\nabla H(y^k,z^k)\|_2\cdot\|y^{k+1}-\overline{y^{k+1}}\|_2\end{aligned}$$

$$\leq L\|x^{k+1}-x^k\|_2\cdot\|y^{k+1}-\overline{y^{k+1}}\|_2 \tag{3.27}$$

and

$$\begin{aligned}&\langle\nabla_z H(x^{k+1})-\nabla_z H(y^{k+1},z^k),z^{k+1}-\overline{z^{k+1}}\rangle\\ \leq&\|\nabla_z H(x^{k+1})-\nabla_z H(y^{k+1},z^k)\|_2\cdot\|z^{k+1}-\overline{z^{k+1}}\|_2\\ \leq& L_2\|z^{k+1}-z^k\|_2\cdot\|z^{k+1}-\overline{z^{k+1}}\|_2\\ \leq& L\|x^{k+1}-x^k\|_2\cdot\|z^{k+1}-\overline{z^{k+1}}\|_2. \end{aligned}\tag{3.28}$$

Substituting Equations (3.25), (3.26), (3.27), and (3.28) into (3.24), we obtain that

$$\Phi(x^{k+1})-\Phi(\overline{x^{k+1}})\leq\sqrt{2}(\max_k\{\gamma_k,\lambda_k\}+L)\|x^{k+1}-x^k\|_2\cdot\mathrm{dist}(x^{k+1},\chi^*). \tag{3.29}$$

∎

LEMMA 3.4. *Assume that* $\min_k\{\gamma_k,\lambda_k\}>L_2+\frac{L}{2}$. *Let* $\{x^k=(y^k,z^k)\}_{k=1,2,\dots}$ *be generated by the PALM method for problem (1.1). Then, for any* $k\in\mathbb{N}$,

$$\Phi(x^k)-\Phi(x^{k+1})\geq\rho_2\|x^{k+1}-x^k\|_2^2, \tag{3.30}$$

where $\rho_2=\min_k\{\gamma_k,\lambda_k\}-(L_2+\frac{L}{2})$.

Proof. Applying Lemma 2.2 to Equation (2.17) with $u=y^k$ and $u=z^k$, we obtain that

$$f(y^k)\geq f(y^{k+1})+\gamma_k\langle y^k-y^{k+1},y^k-y^{k+1}\rangle+\langle\nabla_y H(y^k,z^k),y^{k+1}-y^k\rangle \tag{3.31}$$

and

$$g(z^k)\geq g(z^{k+1})+\lambda_k\langle z^k-z^{k+1},z^k-z^{k+1}\rangle+\langle\nabla_z H(y^{k+1},z^k),z^{k+1}-z^k\rangle. \tag{3.32}$$

Applying Lemma 2.1 to H at x^k, we have

$$H(x^{k+1})-H(x^k)\leq\langle\nabla H(x^k),x^{k+1}-x^k\rangle+\frac{L}{2}\|x^{k+1}-x^k\|_2^2. \tag{3.33}$$

Summing the inequalities above, we have that

$$\begin{aligned}\Phi(x^k)-\Phi(x^{k+1})\geq&(\min_k\{\gamma_k,\lambda_k\}-\frac{L}{2})\|x^k-x^{k+1}\|_2^2\\ &+\langle\nabla_z H(y^{k+1},z^k)-\nabla_z H(y^k,z^k),z^{k+1}-z^k\rangle.\end{aligned}\tag{3.34}$$

Note that

$$\begin{aligned}|\langle\nabla_z H(y^{k+1},z^k)-\nabla_z H(y^k,z^k),z^{k+1}-z^k\rangle|&\leq L_2\|y^{k+1}-y^k\|_2\cdot\|z^{k+1}-z^k\|_2\\ &\leq L_2\|x^{k+1}-x^k\|_2^2.\end{aligned}\tag{3.35}$$

Thus, we have that

$$\Phi(x^k)-\Phi(x^{k+1})\geq(\min_k\{\gamma_k,\lambda_k\}-\frac{L}{2}-L_2)\|x^k-x^{k+1}\|_2^2. \tag{3.36}$$

∎

3.3. HPALM scheme. In this part, we investigate the convergence of both HPALM-I and HPALM-II.

3.3.1. HPALM-I.

LEMMA 3.5. *Assume that* $\max_k\{\gamma_k\}<+\infty$. *Let* $\{x^k=(y^k,z^k)\}_{k=1,2,\dots}$ *be generated by the PALM-*I *method for problem (1.1). If* $\gamma_k\geq\frac{L_1}{2}$, *then, for any* $k\in\mathbb{N}$,

$$\Phi(x^{k+1})-\Phi^*\leq\rho_1\|y^{k+1}-y^k\|_2\cdot\mathrm{dist}(x^k,\chi^*),\tag{3.37}$$

where $\rho_1=\max_k\{\gamma_k\}$.

Proof. Applying Lemma 2.2 to Equation (2.19) with $u=\overline{y^k}$, we obtain that

$$f(\overline{y^k})\geq f(y^{k+1})+\gamma_k\langle y^k-y^{k+1},\overline{y^k}-y^{k+1}\rangle+\langle\nabla_y H(y^k,z^k),y^{k+1}-\overline{y^k}\rangle.\tag{3.38}$$

And from the scheme and the KKT condition, we easily have that

$$\mathbf{0}\in\nabla_z H(y^k,z^k)+\partial g(z^k).$$

The convexity of g then gives that

$$g(\overline{z^k})\geq g(z^k)+\langle\nabla_z H(y^k,z^k),z^k-\overline{z^k}\rangle.\tag{3.39}$$

From Lemma 2.1, we have

$$\begin{aligned}&H(y^{k+1},z^k)-H(\overline{x^k})\\&\leq H(x^k)+\langle\nabla_y H(y^k,z^k),y^{k+1}-y^k\rangle+\frac{L_1}{2}\|y^{k+1}-y^k\|_2^2-H(\overline{x^k}).\end{aligned}\tag{3.40}$$

With the convexity of H, we have that

$$H(x^k)-H(\overline{x^k})\leq\langle\nabla_y H(y^k,z^k),y^k-\overline{y^k}\rangle+\langle\nabla_z H(y^k,z^k),z^k-\overline{z^k}\rangle.\tag{3.41}$$

Combining Equations (3.40) and (3.41), we have

$$\begin{aligned}&H(y^{k+1},z^k)-H(\overline{x^k})\\&\leq\langle\nabla_y H(y^k,z^k),y^{k+1}-\overline{y^k}\rangle+\langle\nabla_z H(y^k,z^k),z^k-\overline{z^k}\rangle+\frac{L_1}{2}\|y^{k+1}-y^k\|_2^2.\end{aligned}\tag{3.42}$$

Note that $\Phi(x^{k+1})\leq\Phi(y^{k+1},z^k)$, summing up Equations (3.38), (3.39), and (3.42), we obtain that

$$\begin{aligned}\Phi(x^{k+1})-\Phi^*&\leq\Phi(y^{k+1},z^k)-\Phi(\overline{y^k},\overline{z^k})\\&\leq\gamma_k\langle y^{k+1}-y^k,\overline{y^k}-y^{k+1}\rangle+\frac{L_1}{2}\|y^{k+1}-y^k\|_2^2\\&\leq\gamma_k\langle y^{k+1}-y^k,\overline{y^k}-y^{k+1}\rangle+\gamma_k\|y^{k+1}-y^k\|_2^2\\&\leq\gamma_k\langle\overline{y^k}-y^k,y^{k+1}-y^k\rangle.\end{aligned}\tag{3.43}$$

It is easy to obtain that

$$\begin{aligned}\gamma_k\langle\overline{y^k}-y^k,y^{k+1}-y^k\rangle&\leq\gamma_k\|\overline{y^k}-y^k\|_2\cdot\|y^{k+1}-y^k\|_2\\&\leq\gamma_k\|\overline{x^k}-x^k\|_2\cdot\|y^{k+1}-y^k\|_2.\end{aligned}\tag{3.44}$$

Substituting Equation (3.44) into Equation (3.43), we have

$$\Phi(x^{k+1})-\Phi^*\leq(\max_k\{\gamma_k\})\|y^{k+1}-y^k\|_2\cdot\mathrm{dist}(x^k,\chi^*).\tag{3.45}$$

□

LEMMA 3.6. *Assume that* $\min_k\{\gamma_k\} > \frac{L_1}{2}$. *Let* $\{x^k = (y^k, z^k)\}_{k=1,2,\dots}$ *be generated by the PALM-*I *method for problem (1.1). Then, for any* $k \in \mathbb{N}$,

$$\Phi(x^k) - \Phi(x^{k+1}) \geq \rho_2 \|y^{k+1} - y^k\|_2, \tag{3.46}$$

where $\rho_2 = \min_k\{\gamma_k\} - \frac{L_1}{2}$.

Proof. Applying Lemma 2.2 to Equation (2.19) with $u = y^k$, we obtain that

$$f(y^k) \geq f(y^{k+1}) + \gamma_k \langle y^k - y^{k+1}, y^k - y^{k+1}\rangle + \langle \nabla_y H(y^k, z^k), y^{k+1} - y^k\rangle. \tag{3.47}$$

Applying Lemma 2.1 to $H(y, z^k)$, we have that

$$H(y^{k+1}, z^k) \leq H(y^k, z^k) + \langle \nabla_y H(y^k, z^k), y^{k+1} - y^k\rangle + \frac{L_1}{2}\|y^{k+1} - y^k\|_2^2. \tag{3.48}$$

Summing them together, we derive that

$$H(y^k, z^k) + f(y^k) - H(y^{k+1}, z^k) - f(y^{k+1}) \geq (\gamma_k - \frac{L_1}{2})\|y^{k+1} - y^k\|_2^2. \tag{3.49}$$

Then, we have

$$\begin{aligned}\Phi(x^k) - \Phi(x^{k+1}) &\geq \Phi(y^k, z^k) - \Phi(y^{k+1}, z^k)\\ &\geq (\gamma_k - \frac{L_1}{2})\|y^{k+1} - y^k\|_2^2\\ &\geq (\min_k\{\gamma_k\} - \frac{L_1}{2})\|y^{k+1} - y^k\|_2^2. \end{aligned} \tag{3.50}$$

□

3.3.2. HPALM-II.

LEMMA 3.7. *Assume that* $\max_k\{\gamma_k\} < +\infty$ *and* $\min_k\{\lambda_k\} \geq \frac{L_2}{2}$. *Let* $\{x^k = (y^k, z^k)\}_{k=1,2,\dots}$ *be generated by the PALM-*II *method for problem (1.1). Then, for any* $k \in \mathbb{N}$,

$$\Phi(x^{k+1}) - \Phi^* \leq \rho_1 \|z^{k+1} - z^k\|_2 \cdot \mathrm{dist}(x^{k+1}, \chi^*), \tag{3.51}$$

where $\rho_1 = \max_k\{\gamma_k\}$.

Proof. Applying Lemma 2.2 to Equation (2.20) with $u = \overline{z^k}$, we obtain that

$$g(\overline{z^k}) \geq g(z^{k+1}) + \gamma_k \langle z^k - z^{k+1}, \overline{z^k} - z^{k+1}\rangle + \langle \nabla_z H(y^{k+1}, z^k), z^{k+1} - \overline{z^k}\rangle. \tag{3.52}$$

And from the scheme and the KKT condition, we easily have that

$$\mathbf{0} \in \nabla_y H(y^{k+1}, z^k) + \partial f(y^{k+1}).$$

The convexity of f then gives

$$f(\overline{y^k}) \geq f(y^{k+1}) + \langle \nabla_y H(y^{k+1}, z^k), y^{k+1} - \overline{y^k}\rangle. \tag{3.53}$$

From Lemma 2.1, we have that

$$H(x^{k+1}) - H(\overline{x^k}) \leq H(y^{k+1}, z^k) + \langle \nabla_z H(y^{k+1}, z^k), z^{k+1} - z^k\rangle$$

$$+\frac{L_2}{2}\|z^{k+1}-z^k\|_2^2-H(\overline{x^k}). \tag{3.54}$$

With the convexity of H, we have that

$$H(y^{k+1},z^k)-H(\overline{x^k})\leq\langle\nabla_y H(y^{k+1},z^k),y^{k+1}-\overline{y^k}\rangle+\langle\nabla_z H(y^{k+1},z^k),z^k-\overline{z^k}\rangle. \tag{3.55}$$

Combining Equations (3.54) and (3.55), we have that

$$\begin{aligned}H(x^{k+1})-H(\overline{x^k})\leq&\langle\nabla_y H(y^{k+1},z^k),y^{k+1}-\overline{y^k}\rangle\\&+\langle\nabla_z H(y^{k+1},z^k),z^{k+1}-\overline{z^k}\rangle+\frac{L_2}{2}\|z^{k+1}-z^k\|_2^2.\end{aligned} \tag{3.56}$$

Summing Equations (3.52), (3.53), and (3.56), we obtain that

$$\begin{aligned}\Phi(x^{k+1})-\Phi^*&\leq\lambda_k\langle z^{k+1}-z^k,\overline{z^k}-z^{k+1}\rangle+\frac{L_2}{2}\|z^{k+1}-z^k\|_2^2\\&\leq\lambda_k\langle z^{k+1}-z^k,\overline{z^k}-z^{k+1}\rangle+\lambda_k\|z^{k+1}-z^k\|_2^2\\&\leq\lambda_k\langle z^{k+1}-z^k,\overline{z^k}-z^k\rangle.\end{aligned} \tag{3.57}$$

It is easy to obtain that

$$\begin{aligned}\lambda_k\langle z^{k+1}-z^k,\overline{z^k}-z^k\rangle&\leq\lambda_k\|z^{k+1}-z^k\|_2\cdot\|\overline{z^k}-z^k\|_2\\&\leq\lambda_k\|z^{k+1}-z^k\|_2\cdot\|\overline{x^k}-x^k\|_2.\end{aligned} \tag{3.58}$$

Substituting Equation (3.58) into Equation (3.57), we have

$$\Phi(x^{k+1})-\Phi^*\leq(\max_k\{\lambda_k\})\|z^{k+1}-z^k\|_2\cdot\mathrm{dist}(x^k,\chi^*). \tag{3.59}$$

□

LEMMA 3.8. *Let $\{x^k=(y^k,z^k)\}_{k=1,2,\dots}$ be generated by the PALM-*II *method for problem (1.1). Then, for any $k\in\mathbb{N}$,*

$$\Phi(x^k)-\Phi(x^{k+1})\geq\rho_2\|z^{k+1}-z^k\|_2, \tag{3.60}$$

where $\rho_2=\min_k\{\lambda_k\}-\frac{L_2}{2}$.

Proof. Applying Lemma 2.2 to Equation (2.20) with $u=z^k$, we obtain that

$$g(z^k)\geq g(z^{k+1})+\lambda_k\langle z^k-z^{k+1},z^k-z^{k+1}\rangle+\langle\nabla_z H(y^{k+1},z^k),z^{k+1}-z^k\rangle. \tag{3.61}$$

Applying Lemma 2.1 to $H(y^{k+1},z)$, we have that

$$H(y^{k+1},z^{k+1})\leq H(y^{k+1},z^k)+\langle\nabla_z H(y^{k+1},z^k),z^{k+1}-z^k\rangle+\frac{L_2}{2}\|z^{k+1}-z^k\|_2^2. \tag{3.62}$$

Summing them together, we derive that

$$H(y^{k+1},z^k)+g(z^k)-H(y^{k+1},z^{k+1})-g(z^{k+1})\geq(\gamma_k-\frac{L_2}{2})\|z^{k+1}-z^k\|_2^2. \tag{3.63}$$

Therefore, we have

$$\begin{aligned}\Phi(x^k)-\Phi(x^{k+1})&\geq\Phi(y^{k+1},z^k)-\Phi(x^{k+1})\\&\geq(\gamma_k-\frac{L_2}{2})\|z^{k+1}-z^k\|_2^2\\&\geq(\min_k\{\gamma_k\}-\frac{L_2}{2})\|z^{k+1}-z^k\|_2^2.\end{aligned} \tag{3.64}$$

□

4. Conclusion

In this paper, we investigate the convergence rates of several alternating minimization schemes. With the Q-SD condition, we improve the big-O rate of AM and PALM to little-o. The proofs employ a technique proposed by Davis and Yin. We also combine AM and PALM, then propose HPALM. The little-o convergence rates of two subvariants of HPALM are also studied.

Acknowledgments. The authors are indebted to the editors and anonymous referees for their useful suggestions.

We are grateful for the support from the National Natural Science Foundation of Hunan Province, China (13JJ2001), and the Science Project of National University of Defense Technology (JC120201), and National Science Foundation of China (No. 61402495), and National Science Foundation of China (No.61571008).

Appendix A.

PROPOSITION A.1. *If a closed convex proper function J satisfies RSC. Then, it satisfies GSD.*

Proof. Let χ^* be the solution set of $\min_x J(x)$. From the definition of RSC [13], we have that

$$\langle \nabla J(x), x-\mathrm{Proj}_{\chi^*}(x)\rangle \geq \nu \|x-\mathrm{Proj}_{\chi^*}(x)\|_2^2, \tag{A.1}$$

for some $\nu>0$. For a fixed x, note that $tx+(1-t)\mathrm{Proj}_{\chi^*}(x)$ also projects $\mathrm{Proj}_{\chi^*}(x)$ onto χ^*. Therefore, we have that

$$J(x)-J(\mathrm{Proj}_{\chi^*}(x)) = \int_0^1 \langle \nabla J[\mathrm{Proj}_{\chi^*}(x)+t(x-\mathrm{Proj}_{\chi^*}(x))], x-\mathrm{Proj}_{\chi^*}(x)\rangle dt. \tag{A.2}$$

Let $y_t = \mathrm{Proj}_{\chi^*}(x)+t(x-\mathrm{Proj}_{\chi^*}(x))$, we have that $\mathrm{Proj}_{\chi^*}(y_t)=\mathrm{Proj}_{\chi^*}(x)$. Thus, $x-\mathrm{Proj}_{\chi^*}(x)=\frac{1}{t}(y_t-\mathrm{Proj}_{\chi^*}(y_t))$. Then, Equation (A.2) turns into

$$\begin{aligned} J(x)-J(\mathrm{Proj}_{\chi^*}(x)) &= \int_0^1 \frac{1}{t}\langle \nabla J(y_t), y_t-\mathrm{Proj}_{\chi^*}(y_t)\rangle \\ &\geq \nu \int_0^1 \frac{1}{t}\|y_t-\mathrm{Proj}_{\chi^*}(y_t)\|_2^2 dt \\ &= \nu \int_0^1 t\|x-\mathrm{Proj}_{\chi^*}(x)\|_2^2 dt = \frac{\nu}{2}\|x-\mathrm{Proj}_{\chi^*}(x)\|_2^2. \end{aligned} \tag{A.3}$$

∎

PROPOSITION A.2. *For any sufficiently large positive constant R, the function in Example 2 satisfies Q-SD, where $Q=\{x \mid \|x\|_2=\|(y,z)\|_2\leq R\}$.*

Proof. Let

$$\kappa(x_1,x_2):=\kappa_1(x_1)+\kappa_2(x_2), \quad W=\begin{pmatrix} a & b \\ c & d\end{pmatrix}, \quad \widetilde{\lambda}:=\begin{pmatrix} \lambda_1 i_{n_1} & \\ & \lambda_2 i_{n_2}\end{pmatrix}.$$

Then, κ is strongly convex. And we have that

$$\Phi(x)=\kappa(Wx)+\|\widetilde{\lambda}x\|_1. \tag{A.4}$$

From [8, Lemma 10], $\Phi(x)$ satisfies Q-SD condition, where $Q=\{x \mid \|x\|_2=\|(y,z)\|_2\leq R\}$ and R is any positive sufficiently large constant. ∎

REFERENCES

[1] H. Attouch and J. Peypouquet, *The rate of convergence of Nesterov's accelerated forward-backward method is actually* $o(\frac{1}{k^2})$, arXiv:1510.08740v1, 2015.

[2] A. Beck, *On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes*, SIAM Journal on Optimization, 25(1), 185–209, 2015.

[3] A. Beck and S. Shtern, *Linearly Convergent Away-Step Conditional Gradient for Non-strongly Convex Functions*, arXiv:1504.05002, 2015.

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

[5] A. Beck and M. Teboulle, *Gradient-based algorithms with applications to signal recovery*, Convex Optimization in Signal Processing and Communications, 2010.

[6] A. Beck and L. Tetruashvili, *On the convergence of block coordinate descent type methods*, SIAM Journal on Optimization, 23(4), 2037–2060, 2013.

[7] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Inc., 1989.

[8] J. Bolte, T.P. Nguyen, J. Peypouquet, and B.W. Suter, *From error bounds to the complexity of first-order descent methods for convex functions*, arXiv:1510.08234v1, 2015.

[9] P.L. Combettes and J.C. Pesquet, *Proximal splitting methods in signal processing*, in Fixed-Point Algorithms for Inverse Problems Science and Engineering Springer Optimization and Its Applications, Springer, New York, 49, 185–212, 2011.

[10] D. Davis, *Convergence rate analysis of the Forward-Douglas–Rachford splitting scheme*, SIAM Journal on Optimization, 25(3), 1760–1786, 2015.

[11] D. Davis, *Convergence rate analysis of primal-dual splitting schemes*, SIAM Journal on Optimization, 25(3), 1912–1943, 2015.

[12] D. Davis and W. Yin, *Convergence rate analysis of several splitting schemes*, arXiv:1406.4834, 2014.

[13] M.J. Lai and W. Yin, *Augmented* ℓ_1 *and nuclear-norm models with a globally linearly convergent algorithm*, SIAM Journal on Imaging Sciences, 6(2), 1059–1091, 2013.

[14] Z.Q. Luo and P. Tseng, *On the linear convergence of descent methods for convex essentially smooth minimization*, SIAM J. Control Optim., 30, 408–425, 1992.

[15] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, 2004.

[16] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer Series in Operations Research, Springer, Second Edition, 2006.

[17] N. Parikh and S. Boyd, *Proximal algorithm*, Foundations and Trends in optimization, 1–96, 2013.

[18] R. Shefi and M. Teboulle, *On the rate of convergence of the proximal alternating linearized minimization algorithm for convex problems*, EURO Journal on Computational Optimization, 1–20, 2015.

[19] S. Sra, S. Nowozin, and S.J. Wright, *Optimization for Machine Learning*, Mit Press, 2012.

[20] P. Tseng and S. Yun, *A coordinate gradient descent method for nonsmooth separable minimization*, Math. Program., 117, 387–423, 2009.

[21] S.J. Wright, *Coordinate descent algorithms*, Mathematical Programming, 151(1), 3–34, 2015.