Linear convergence of CQ algorithms and applications in gene regulatory network inference

Jinhua Wang* Yaohua Hu† Chong Li‡ Jen-Chih Yao§

July 16, 2017

Abstract

In the present paper, we consider the varying stepsize CQ algorithm for solving the split feasibility problem in Hilbert spaces, investigate the linear convergence issue and explore an application in systems biology. In particular, we introduce a notion of bounded linear regularity property for the split feasibility problem, and use it to establish the linear convergence property for the varying stepsize CQ algorithm when using some suitable types of stepsizes, which covers most types of stepsizes used in the literature of CQ algorithms. We also provide some mild sufficient conditions for ensuring this bounded linear regularity property, and then conclude the linear convergence rate of the varying stepsize CQ algorithm for many application cases. To the best of our knowledge, this is the first work to study the linear convergence rate of CQ algorithms. In the aspect of application, we consider the gene regulatory network inference arising in systems biology, which is formulated as a group Dantzig selector and then cast into a split feasibility problem. The numerical study on gene expression data of mouse embryonic stem cell shows that the varying stepsize CQ algorithm is applicable to gene regulatory network inference in the sense that it obtains a reliable solution matching with biological standards.

*Department of Mathematics, Zhejiang University of Technology, Hangzhou 310032, China (wjh@zjut.edu.cn). This author's work was supported in part by the National Natural Science Foundation of China (grant 11371325).

†Corresponding author. College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China (mayhhu@szu.edu.cn). This author's work was supported in part by the National Natural Science Foundation of China (grant 11601343), Natural Science Foundation of Guangdong (grant 2016A030310038), and Foundation for Distinguished Young Talents in Higher Education of Guangdong (grant 2015KQNCX145).

‡Department of Mathematics, Zhejiang University, Hangzhou 310027, China (cli@zju.edu.cn). This author's work was supported in part by the National Natural Science Foundation of China (grants 11571308, 91432302).

§Center for General Education, China Medical University, Taichung 40402, Taiwan (yaojc@maill.cmu.edu.tw). This author's work was supported in part by the National Science Council of Taiwan under grant MOST 102-2115-M-039-003-MY3.

1 Introduction

Let H_1 and H_2 be (finite- or infinite-dimensional) Hilbert spaces, and let A be a bounded linear operator from H_1 to H_2. Let C and Q be nonempty closed convex subsets of H_1 and H_2, respectively. The split feasibility problem (SFP for short) is to find a point x such that

$$x \in C \quad \text{and} \quad Ax \in Q. \tag{1.1}$$

The SFP was introduced by Censor and Elfving [13] to solve the phase retrieval problems, and it provides a unified framework for many inverse problems. Recently, the SFP has received a great amount of attention due to its wide applications in signal processing [7], image reconstruction [20, 31] and intensity-modulated radiation therapy [12, 14].

The development of numerical algorithms for solving the SFP (1.1) has attracted much attention. One of the most popular and practical algorithms for solving the SFP is the CQ algorithm, which was proposed by Byrne [6, 7] and has the following iterative form:

$$x_{n+1} = P_C(x_n - \beta A^*(I - P_Q)Ax_n),$$

where A^* is the adjoint of A, $\beta > 0$ is the stepsize, while P_C and P_Q denote the metric projections onto C and Q, respectively.

As remarked in [6] (also see [30, 41]), one of the main advantages of the CQ algorithm is that it involves only the computations of the metric projections onto C and Q, which are usually easily calculated in many applications (e.g., when C and Q are the closed balls or half-spaces); and it avoids the difficulty of calculating the matrix inverses at each iteration in the algorithm proposed in [13]. Benefitting from this advantage, the CQ algorithm becomes a popular tool for solving the SFP, and various variants of CQ algorithms by using different types of stepsizes have been widely studied in the literature; one can refer to a recent book of Byrne [8]. Some works in the finite-dimensional spaces and infinite-dimensional spaces can be found in [14,43,47] and [30,35,40,41], respectively. In particular, Xu [41] studied the weak convergence of the CQ algorithm for solving the SFP in infinite-dimensional Hilbert spaces by virtue of fixed point theory. López et al. [30] introduced the following dynamic stepsize CQ algorithm and established its weak convergence in the infinite-dimensional setting:

$$x_{n+1} = P_C(x_n - \beta_n A^*(I - P_Q)Ax_n),$$

where

$$\beta_n := \begin{cases} 0, & \text{if } x_n \in C \cap A^{-1}Q, \\ \frac{\rho_n \|(I-P_Q)Ax\|^2}{\|A^*(I-P_Q)Ax\|^2}, & \text{otherwise,} \end{cases} \quad \text{and} \quad \{\rho_n\} \subseteq (0, 2). \tag{1.2}$$

The advantage of the dynamic stepsize CQ algorithm is that it does not require any prior knowledge about the norm of operator (matrix) A. It was pointed out that the CQ algorithm and the dynamic stepsize CQ algorithm may fail to converge strongly in infinite-dimensional Hilbert spaces; see [41, Example 3.7] and [30, pp. 7]. Therefore, modifications of CQ algorithm and dynamic stepsize CQ algorithm, together with their strong convergence property, were proposed in [41] and [30], respectively; and the modification of dynamic stepsize CQ algorithm was further extended in [36] to solve the split common fixed-point problem involving a solution set of equilibrium problem. Under some additional assumptions, the

strong convergence property of the CQ algorithms was studied in [11,16,46] as special cases of some generalized CQ-type algorithms; see Remark 2.9 for more details. However, to the best of our knowledge, there is still no paper devoted to establishing the convergence rate for the CQ algorithm or the dynamic stepsize CQ algorithm.

The aim of the present paper is to continue the convergence study of the CQ algorithm, but from a new perspective, in (not necessarily finite-dimensional) Hilbert spaces; in particular, we contribute to investigating the linear convergence issue for the following varying stepsize CQ algorithm (V-CQ algorithm for short):

Algorithm 1.1. Let $x_0 \in H$ be given. Having $x_0, x_1, \cdots, x_n$, we choose a stepsize $\beta_n > 0$ and determine x_{n+1} as follows:

$$x_{n+1} = P_C(x_n - \beta_n A^*(I - P_Q)Ax_n).$$

Clearly, Algorithm 1.1 includes the CQ algorithm and the dynamic stepsize CQ algorithm as special cases; e.g., when $\beta_n \equiv \beta > 0$ or $\{\beta_n\}$ is given by (1.2).

The main contribution of the present paper is to establish the linear convergence property for the V-CQ algorithm. For this purpose, we introduce a notion of bounded linear regularity property for the SFP, and use it to prove the linear convergence property, meaning that the generated sequence converges linearly to a feasible solution of the SFP, of the V-CQ algorithm by using some suitable types of stepsizes. The established convergence results cover the linear convergence property of the CQ algorithm and that of the dynamic stepsize CQ algorithm. In order to popularize the applications of the established convergence results, we provide some mild sufficient conditions for ensuring this regularity property for the SFP. Some of these sufficient conditions are satisfied for several applications, such as the (group) Dantzig selector [29], which is popular in the fields of compressive sensing, statistics and machine learning. As far as we know, this is the first work to study the linear convergence rate of the CQ algorithm for solving the SFP (1.1). Furthermore, two examples are provided to show the case where the strong convergence of the V-CQ algorithm may fail in the infinite-dimensional Hilbert space, and the case where the linear convergence of the V-CQ algorithm may fail in the Euclidean space if the SFP does not satisfy the bounded linear regularity property, respectively.

Another contribution of the present paper is to infer gene regulatory network of mouse embryonic stem cell (mESC) from gene expression data. Gene regulatory network inference is vital in systems biology to understand complex biological processes, which is to identify the regulatory relationship among the transcription factor complexes (TF complexes) and the target genes from expression data at whole genome level. Due to the molecular biology process, gene regulatory network inference can be understood as a group feature selection problem based on the dependencies between the expression data of TF complexes and that of target genes; see section 3 for the explanation. More specifically, let $D \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times s}$ denote the expression data of the potential TF complexes and that of the target genes of mESC in biological experiments, respectively, and let $Z \in \mathbb{R}^{n\times s}$ denote the regulatory relationship of all the TF-target gene pairs, which is to be predicted. Then, the regulatory relationship between TF complexes and target genes at whole genome level can be represented approximately by a linear system

$$DZ = B + \varepsilon.$$

Equivalently, gene regulatory network inference of mESC at whole genome level consists of a series of inverse problems, each of which is to infer regulatory network for a certain target gene:

$$DZ_{\cdot j} = B_{\cdot j} + \varepsilon_{\cdot j}, \quad j = 1, 2, \ldots, s, \tag{1.3}$$

where $Z_{\cdot j}$ is the j-th column of Z, denoting the regulatory relationship between TF complexes and the j-th target gene, and $B_{\cdot j}$ is the j-th column of B, designating the expression profile of the j-th target gene in the biological experiments. Fixing $j \in \{1, 2, \ldots, s\}$ and using the $\ell_{\infty,1}$ norm to measure the group sparsity of TF complexes, inferring regulatory network for the j-th target gene can be formulated as a group Dantzig selector:

$$\begin{array}{ll} \min & \|Z_{\cdot j}\|_{\infty,1} \\ \text{s.t.} & \|D^\top (DZ_{\cdot j} - B_{\cdot j})\|_\infty \leq \epsilon_j. \end{array} \tag{1.4}$$

By letting $\tilde{D} := D^\top D$ and $\tilde{B}_{\cdot j} := D^\top B_{\cdot j}$, and selecting suitable parameter δ_j, problem (1.4) could be approached by solving the following SFP

$$Z_{\cdot j} \in C_j := \{Z_{\cdot j} : \|Z_{\cdot j}\|_{\infty,1} \leq \delta_j\} \quad \text{and} \quad \tilde{D}Z_{\cdot j} \in Q_j := \overline{\mathbf{Box}}(\tilde{B}_{\cdot j}, \epsilon_j), \tag{1.5}$$

where $\overline{\mathbf{Box}}(x, r)$ denotes a closed box of radius r centered at x. Therefore, due to (1.3), gene regulatory network inference of mESC can be formulated as a series of SFPs ((1.5) with $j = 1, 2, \ldots, s$).

We collect the expression data of mESC, and apply the V-CQ algorithm to sequentially solve a series of SFPs (1.5) resulting from gene regulatory network inference at whole genome level, and compare with some state-of-the-art algorithms. The numerical results exhibit that exploiting the group structure of TF complexes can improve the accuracy of the gene regulation network forecasting, and that the V-CQ algorithm is applicable to gene regulatory network inference in the sense that it obtains more biologically accurate solutions than the existing methods do (improving at least 30% on the AUC value), which may facilitate biologists to study the gene regulation mechanism of higher model organisms in a genome-wide scale.

The paper is organized as follows. In section 2, we establish the linear convergence property for the V-CQ algorithm when using different types of stepsizes and under the assumption of the bounded linear regularity property for the SFP. Applications to gene regulatory network inference and numerical experiments on gene expression data of mESC are demonstrated in section 3. A conclusion is presented in section 4. The proof of some sufficient conditions ensuring the regularity property is provided in Appendix.

2 Linear convergence of CQ algorithms

The notation used in the present paper is standard. Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. For a set $\Omega \subseteq H$, we denote the closure, interior, relative interior and conical hull of Ω by $\mathrm{cl}\Omega$, $\mathrm{int}\Omega$, $\mathrm{ri}\Omega$ and $\mathrm{cone}\Omega$, respectively. For $x \in H$ and $r > 0$, we use $\mathbf{B}(x, r)$ and $\overline{\mathbf{B}(x, r)}$ to denote the open metric ball and the closed metric ball at x with radius r, respectively, that is,

$$\mathbf{B}(x, r) := \{y \in H : \|x - y\| < r\} \quad \text{and} \quad \overline{\mathbf{B}(x, r)} := \{y \in H : \|x - y\| \leq r\}.$$

In particular, we use $\mathbb{B}$ and $\overline{\mathbb{B}}$ to denote the unit open metric ball and the unit closed metric ball at the origin, respectively. For a point x and a set $\Omega \subseteq H$, the classical metric projection of x onto Ω and the distance of x from Ω, denoted by $P_\Omega(x)$ and $\mathrm{d}_\Omega(x)$, are respectively defined by

$$P_\Omega(x) := \operatorname{argmin}\{\|x-y\| : \ y \in \Omega\} \quad \text{and} \quad \mathrm{d}_\Omega(x) := \inf\{\|x-y\| : \ y \in \Omega\}.$$

The following proposition is about some well-known properties of the projection operator, in which (i) is taken from [4, Theorem 3.14]; (ii) and (iii) from [4, Proposition 4.8]; while (iv) and (v) are known in [4, Corollary 4.10].

Proposition 2.1. *Let Ω be a nonempty closed convex set in H, $x, y \in H$ and $z \in \Omega$. Then the following assertions hold:*

(i) $\langle P_\Omega(x) - x, z - P_\Omega(x)\rangle \geq 0$.

(ii) $\|P_\Omega(x) - P_\Omega(y)\|^2 \leq \langle P_\Omega(x) - P_\Omega(y), x - y\rangle$.

(iii) $\|P_\Omega(x) - z\|^2 \leq \|x - z\|^2 - \|P_\Omega(x) - x\|^2$.

(iv) $\langle (I - P_\Omega)x - (I - P_\Omega)y, x - y\rangle \geq \|(I - P_\Omega)x - (I - P_\Omega)y\|^2$.

(v) $\|(I - P_\Omega)x - (I - P_\Omega)y\| \leq \|x - y\|$.

The objective of this section is to investigate the linear convergence of the V-CQ algorithm (Algorithm 1.1) for solving the SFP (1.1). Throughout this section, we always assume that the solution set S of the SFP is nonempty, that is,

$$S := C \cap A^{-1}Q \neq \emptyset.$$

Then the following equivalence holds for any $z \in C$:

$$[z \in S] \quad \Leftrightarrow \quad [(I - P_Q)Az = 0]. \tag{2.1}$$

Regularity conditions have been widely used to analyze the convergence rates of many algorithms; see [9,23] and references therein. In order to establish the linear convergence of the V-CQ algorithm, we introduce a notion of bounded linear regularity property for the SFP.

Definition 2.2. The SFP (1.1) is said to satisfy the bounded linear regularity property if, for any $r > 0$ such that $S \cap \overline{\mathbf{B}(0,r)} \neq \emptyset$, there exists $\gamma_r > 0$ such that

$$\gamma_r \mathrm{d}_S(x) \leq \mathrm{d}_Q(Ax) \quad \text{for any } x \in C \cap \overline{\mathbf{B}(0,r)}. \tag{2.2}$$

Now, under the assumption of bounded linear regularity property, we establish the convergence rate of the V-CQ algorithm using different types of stepsizes.

Theorem 2.3. *Suppose that the SFP* (1.1) *satisfies the bounded linear regularity property. Let* $\{x_n\}$ *be a sequence generated by Algorithm* 1.1 *with* $\{\beta_n\} \subseteq (0, +\infty)$. *Then* $\{x_n\}$ *converges to a solution* x^* *of the SFP* (1.1) *satisfying*

$$\|x_n - x^*\| \leq c\, q^{\sum_{k=1}^{n} \beta_k} \quad \text{for any } n \in \mathbb{N}, \tag{2.3}$$

for two constants $c \geq 1$ *and* $0 < q < 1$, *provided that one of the following conditions is assumed:*

(a) $\{\beta_n\}$ *satisfies*

$$0 < \liminf_{n\to+\infty} \beta_n \leq \limsup_{n\to+\infty} \beta_n < \frac{2}{\|A\|^2}. \tag{2.4}$$

(b) $\{\beta_n\}$ *is given by* (1.2) *with* $\{\rho_n\}$ *satisfying*

$$0 < \liminf_{n\to+\infty} \rho_n \leq \limsup_{n\to+\infty} \rho_n < 2. \tag{2.5}$$

(c) $\{\beta_n\}$ *satisfies*

$$\lim_{n\to+\infty} \beta_n = 0 \quad \text{and} \quad \sum_{n=1}^{+\infty} \beta_n = +\infty. \tag{2.6}$$

Consequently, $\{x_n\}$ *converges to* x^* *linearly in the case when* (a) *or* (b) *is assumed.*

Proof. Without loss of generality, we assume that $x_n \notin S$ for any $n \geq 0$ (otherwise, the V-CQ algorithm terminates in finite iterations and then the conclusions follow trivially). Then, in view of Algorithm 1.1, one sees that $Ax_n \notin Q$ for any $n \geq 0$. Fix $z \in S$ and $n \in \mathbb{N}$. For simplicity, we write

$$\nabla_{x_n} := A^*(I - P_Q)Ax_n. \tag{2.7}$$

Then, one checks that

$$\|\nabla_{x_n}\| \leq \|A\| \mathrm{d}_Q(Ax_n) \quad \text{and} \quad \langle x_n - z, \nabla_{x_n}\rangle \geq \mathrm{d}_Q^2(Ax_n)^1. \tag{2.8}$$

In fact, the first inequality is trivial; while the second one holds because, by Proposition 2.1(iv) and (2.1) (and so $(I - P_Q)Az = 0$),

$$\langle x_n - z, \nabla_{x_n}\rangle = \langle A(x_n - z), (I - P_Q)Ax_n\rangle \geq \|(I - P_Q)Ax_n\|^2 = \mathrm{d}_Q^2(Ax_n). \tag{2.9}$$

Since P_C is nonexpansive, we have

$$\begin{aligned} \|x_{n+1} - z\|^2 &= \|P_C(x_n - \beta_n \nabla_{x_n}) - z\|^2 \\ &\leq \|x_n - \beta_n \nabla_{x_n} - z\|^2 \\ &= \|x_n - z\|^2 - 2\beta_n \langle x_n - z, \nabla_{x_n}\rangle + \beta_n^2 \|\nabla_{x_n}\|^2. \end{aligned}$$

Hence, by (2.9), one has that

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \beta_n \left(2 - \beta_n \frac{\|\nabla_{x_n}\|^2}{\mathrm{d}_Q^2(Ax_n)}\right) \mathrm{d}_Q^2(Ax_n). \tag{2.10}$$

[1] This inequality could be also concluded from [38, Lemma 3.1].

This in particular implies the following implication:

$$[\beta_n \|\nabla_{x_n}\|^2 \leq 2\mathrm{d}_Q^2(Ax_n)] \quad \Rightarrow \quad [\|x_{n+1} - z\| \leq \|x_n - z\|]. \tag{2.11}$$

Below, we show that $\{x_n\}$ converges to a solution x^* of the SFP and (2.3) holds. To do this, suppose that one of (a), (b) and (c) holds. Then we have the following assertions:

(i) If (a) or (c) holds, then there exist $\eta > 0$ and $M \in \mathbb{N}$ such that

$$\beta_n \leq \eta < \frac{2}{\|A\|^2} \quad \text{for any } n \geq M. \tag{2.12}$$

(ii) If (a) or (b) holds, then

$$\liminf_{n\to\infty} \beta_n > 0. \tag{2.13}$$

Indeed, assertions (i) and (ii) for (a) are trivial; assertion (i) for (c) follows from (2.6); while, for (b), assertion (ii) holds by (2.5) because $\beta_n = \frac{\rho_n \mathrm{d}_Q^2(Ax_n)}{\|\nabla_{x_n}\|^2} \geq \frac{\rho_n}{\|A\|^2}$ by (1.2) and (2.8).

Note that there exists $M \in \mathbb{N}$ such that

$$\beta_n \|\nabla_{x_n}\|^2 \leq 2\mathrm{d}_Q^2(Ax_n) \quad \text{for any } n \geq M, \tag{2.14}$$

(this fact follows from (1.2) and (2.7) if (b) is assumed and from (2.12) and (2.8) otherwise). Therefore, $\{\|x_n - z\|\}_{n\geq M}$ is monotone decreasing by (2.11), and so the sequence $\{\|x_n - z\|\}$ is bounded. Hence, there exists $r > 0$ such that $\{x_n\} \subseteq C \cap \overline{\mathbf{B}(0,r)}$. By assumption that the SFP satisfies the bounded linear regularity property, it follows from Definition 2.2 that there exists $\gamma_r > 0$ such that

$$\mathrm{d}_Q(Ax_n) \geq \gamma_r \mathrm{d}_S(x_n) \quad \text{for any } n \geq 0.$$

Then it follows from (2.10) and (2.14) that

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \gamma_r^2 \beta_n \left(2 - \beta_n \frac{\|\nabla_{x_n}\|^2}{\mathrm{d}_Q^2(Ax_n)}\right) \mathrm{d}_S^2(x_n)$$

holds for each $z \in S$; hence

$$\mathrm{d}_S^2(x_{n+1}) \leq \left(1 - \gamma_r^2\, \beta_n \left(2 - \beta_n \frac{\|\nabla_{x_n}\|^2}{\mathrm{d}_Q^2(Ax_n)}\right)\right) \mathrm{d}_S^2(x_n) \quad \text{for any } n \geq M. \tag{2.15}$$

We claim that

$$\liminf_{n\to+\infty} \left(2 - \beta_n \frac{\|\nabla_{x_n}\|^2}{\mathrm{d}_Q^2(Ax_n)}\right) > 0. \tag{2.16}$$

Note by (2.8) that

$$2 - \beta_n \frac{\|\nabla_{x_n}\|^2}{\mathrm{d}_Q^2(Ax_n)} \geq 2 - \beta_n \|A\|^2.$$

Thus, (2.16) is true by assertion (i) (see (2.12)) in the case of (a) or (c), and by (2.5) in the case of (b) because, by (1.2),

$$2 - \beta_n \frac{\|\nabla_{x_n}\|^2}{\mathrm{d}_Q^2(Ax_n)} = 2 - \rho_n \quad \text{for any } n \geq 0$$

(noting that each $Ax_n \notin Q$ by (2.1) as $x_n \notin S$ by the earlier assumption). Therefore, (2.16) is checked; hence

$$\liminf_{n\to+\infty}\left\{\gamma_r^2\left(2-\beta_n\frac{\|\nabla_{x_n}\|^2}{\mathrm{d}_Q^2(Ax_n)}\right)\right\}>0.$$

Then there exists $N\geq M$ such that

$$\delta:=\inf_{n\geq N}\left\{\gamma_r^2\left(2-\beta_n\frac{\|\nabla_{x_n}\|^2}{\mathrm{d}_Q^2(Ax_n)}\right)\right\}>0.$$

From (2.15), it follows that

$$\mathrm{d}_S^2(x_{n+1})\leq(1-\delta\beta_n)\mathrm{d}_S^2(x_n)\leq\mathrm{d}_S^2(x_N)\prod_{k=N+1}^{n}(1-\delta\beta_k)\quad\text{for any } n\geq N. \tag{2.17}$$

Now fix $n\geq N$. Recalling that $\{\|x_m-P_S(x_{n+1})\|\}_{m\geq n}$ is monotone decreasing, we conclude, for any $m>n$, that

$$\|x_m-x_{n+1}\|\leq\|x_m-P_S(x_{n+1})\|+\|x_{n+1}-P_S(x_{n+1})\|\leq 2\|x_{n+1}-P_S(x_{n+1})\|=2\mathrm{d}_S(x_{n+1}).$$

This, together with (2.17), implies that

$$\|x_m-x_{n+1}\|\leq 2\mathrm{d}_S(x_N)\prod_{k=N+1}^{n}\sqrt{1-\delta\beta_k}\quad\text{for any } m>n>N. \tag{2.18}$$

Note that $\ln(1-t)\leq -t$ for any $t\in[0,1)$. It follows, for any $n>N$, that

$$\prod_{k=N+1}^{n}\sqrt{1-\delta\beta_k}=\exp\left(\frac{1}{2}\sum_{k=N+1}^{n}\ln(1-\delta\beta_k)\right)\leq q^{\sum_{k=N+1}^{n}\beta_k},$$

where $q:=e^{-\frac{\delta}{2}}\in(0,1)$. This, together with (2.18), yields that

$$\|x_m-x_{n+1}\|\leq 2\mathrm{d}_S(x_N)\,q^{\sum_{k=N+1}^{n}\beta_k}\quad\text{for any } m>n>N.$$

Since $\sum_{n=1}^{+\infty}\beta_n=+\infty$ by (2.6) and assertion (ii), it follows that $\{x_n\}$ is a Cauchy sequence and converges to a solution x^* of the SFP satisfying

$$\|x_{n+1}-x^*\|\leq 2\mathrm{d}_S(x_N)\,q^{\sum_{k=N+1}^{n}\beta_k}\quad\text{for any } n>N.$$

Then (2.3) holds with c is given by

$$c:=\max\left\{2\mathrm{d}_S(x_N)\,q^{-\sum_{k=1}^{N}\beta_k},\max_{i=1,\ldots,N}\|x_i-x^*\|q^{-\sum_{k=1}^{i}\beta_k}\right\}>0.$$

Consequently, by (2.13), one has that $\{x_n\}$ converges to x^* linearly, if (a) or (b) is assumed. The proof is complete. □

Remark 2.4. The sequence $\{\beta_n\}$ satisfying (2.6) is also called the diminishing stepsize in the literature of gradient methods; see, e.g., [5,24,25]. In particular, if we choose $\{\beta_n\} := \{\frac{1}{n^\alpha}\}$ with $\alpha \in (0,1)$, which satisfies (2.6), then any sequence $\{x_n\}$ generated by Algorithm 1.1 converges to a solution x^* of the SFP at a rate of $1-\alpha$, that is,

$$\|x_n - x^*\| \leq c\, q^{n^{1-\alpha}} \quad \text{for any } n \in \mathbb{N}.$$

Indeed,

$$\sum_{k=1}^{n} \beta_k = \sum_{k=1}^{n} \frac{1}{k^\alpha} = \sum_{k=1}^{n} \int_k^{k+1} \frac{1}{k^\alpha} \mathrm{d}x \geq \sum_{k=1}^{n} \int_k^{k+1} \frac{1}{x^\alpha} \mathrm{d}x = \frac{1}{1-\alpha}\left((n+1)^{1-\alpha} - 1\right).$$

To popularize the applications of the established convergence results, we provide some sufficient conditions ensuring the regularity property for the SFP (1.1) in the following proposition. For the convenience of readers, we give the proof of the proposition in Appendix.

Proposition 2.5. *The SFP* (1.1) *satisfies the bounded linear regularity property provided one of the following conditions holds:*

(i) *C and Q are polyhedrons.*

(ii) *$AC \cap \mathrm{int}Q \neq \emptyset$.*

(iii) *$A(\mathrm{ri}C) \cap Q \neq \emptyset$ and Q is a polyhedron.*

(iv) *$AC \cap \mathrm{ri}Q \neq \emptyset$, C is a polyhedron and Q is finite-codimensional.*

(v) *$A(\mathrm{ri}C) \cap \mathrm{ri}Q \neq \emptyset$, and Q is finite-codimensional.*

The following corollary, which seems new to the best of our knowledge, is a direct consequence of Proposition 2.5 and Theorem 2.3.

Corollary 2.6. *Suppose that one of statements* (i)-(v) *of Proposition 2.5 holds. Let $\{x_n\}$ be a sequence generated by Algorithm* 1.1*. Then $\{x_n\}$ converges to a solution x^* of the SFP satisfying* (2.3)*, provided that one of conditions* (a)*,* (b) *and* (c) *in Theorem 2.3. In particular, in the case of* (a) *or* (b)*, $\{x_k\}$ converges to x^* linearly.*

Examples 2.7 and 2.8 below illustrate that the bounded linear regularity property may fail if none of conditions (i)-(v) in Proposition 2.5 is satisfied. In particular, Example 2.8 in the Euclidean space shows further that the V-CQ algorithm may fail to linearly converge (even though it converges), if the SFP does not satisfy the bounded linear regularity property. Recall that Example 2.7 is taken from [26].

Example 2.7. *Let $\{e_i\}_{i=1}^{\infty}$ denote the orthonormal basis of the Hilbert space l^2. Consider the SFP* (1.1) *with*

$$C := \{x \in l^2 : \langle x, e_1\rangle \leq 0\}, \quad Q := \mathrm{cl}\,(\mathrm{cone}\{f(x) : x \in \mathbb{R}_+\}) \quad \text{and} \quad A := I,$$

where $f : \mathbb{R} \to l^2$ *is defined by*

$$f(x) := e_{\lceil x \rceil + 2} \cos\left(\frac{\pi}{2}(x - \lceil x \rceil)\right) + e_{\lceil x \rceil + 3} \sin\left(\frac{\pi}{2}(x - \lceil x \rceil)\right) + e_1 \exp(-100x^3),$$

and $\lceil x \rceil$ *denotes the largest integer not greater than* x*. Note that* $C \cap Q = \{0\}$*. Chosen* $x_0 \neq 0$ *and* $\beta_n \equiv 1$*, the sequence generated by Algorithm* 1.1 *is formulated by*

$$x_n = P_C(P_Q(x_{n-1})) = (P_C P_Q)^n x_0.$$

It was reported by [26, Theorem 1] that this sequence $\{x_n\}$ *fails to converge strongly to 0. Thus, by Theorem 2.3, one sees that the bounded linear regularity property cannot be satisfied for this problem.*

Example 2.8. *Consider the SFP* (1.1) *with*

$$C := \mathbb{R} \times \{1\}, \quad Q := \overline{\mathbb{B}} \quad \textit{and} \quad A := I.$$

It is clear that the solution set of the SFP is $S = \{(0, 1)\}$*. Let* $x := (u, 1)$*. Then we have*

$$\lim_{u \to 0} \frac{\mathrm{d}_Q(Ax)}{\mathrm{d}_S(x)} = \lim_{u \to 0} \frac{\sqrt{1 + u^2} - 1}{|u|} = 0.$$

Thus there does not exist $\gamma_r > 0$ *such that* (2.2) *holds, and so the bounded linear regularity property is not satisfied for this problem.*

Let $x_0 = (u_0, 1)$ *with* $u_0 > 0$*,* $\beta_n = \beta \in (0, 2)$ *(as* $\|A\| = 1$*), and let* $\{x_n\}$ *be a sequence generated by Algorithm* 1.1*. We have by [6, Theorem 2.1] that* $\{x_n\}$ *converges to* $(0, 1)$*, the unique solution of the SFP. Below, we show that* $\{x_n\}$ *does not linearly converge to the solution of the SFP. In view of Algorithm* 1.1*, we have that* $x_n = (u_n, 1)$*,*

$$P_Q(x_n) = \left(\frac{u_n}{\sqrt{1 + u_n^2}}, \frac{1}{\sqrt{1 + u_n^2}}\right),$$

and

$$x_{n+1} = P_C(x_n - \beta(x_n - P_Q(x_n))) = \left(u_n \left(1 - \beta + \frac{\beta}{\sqrt{1 + u_n^2}}\right), 1\right).$$

Recalling that $\{x_n\}$ *converges to* $(0, 1)$*, we have that* $\lim_{n \to \infty} u_n = 0$*, and then*

$$\lim_{n \to \infty} \frac{\mathrm{d}_S(x_{n+1})}{\mathrm{d}_S(x_n)} = \lim_{n \to \infty} \left(1 - \beta + \frac{\beta}{\sqrt{1 + u_n^2}}\right) = 1.$$

That is, $\{x_n\}$ *does not linearly converge to* $(0, 1)$*, the unique solution of the SFP.*

We end this section with a remark to mention some relevant works on the strong convergence issue of CQ algorithms.

Remark 2.9. Several generalized CQ-type algorithms, together with their strong convergence, were proposed and studied recently for solving more general problems than the SFP, such as the multiple-set split feasibility problem [16], and the variational inequality problem on a solution set of a split common fixed-point problem [11, 46]; these generalized CQ-type algorithms particularly include the CQ algorithm or dynamic stepsize CQ algorithm (i.e., the V-CQ algorithm with stepsizes satisfying (1.2)) as special cases. Thus, as direct consequences, the strong convergence result of the CQ algorithm, under the semicompact assumption and the demiclosed assumption, was obtained by [16, Theorem 3.1] and [46, Theorem 2.1] respectively; while the strong convergence result of the dynamic stepsize CQ algorithm was obtained by [11, Theorem 6.1] under the assumption of the bounded regularity property there. Although the bounded linear regularity property assumed in Theorem 2.3 is stronger than the bounded regularity, demiclosed or semicompact condition, Theorem 2.3 provides a clear convergence rate for the V-CQ algorithm; in particular, the linear convergence result is established for CQ algorithm and dynamic stepsize CQ algorithm, which partially improve the convergence results mentioned above for the SPF.

3 Gene regulatory network inference

Inferring gene regulatory networks from gene expression data at whole genome level is an arduous challenge in systems biology, especially for higher organisms in which the number of genes is large but the number of experimental samples is small. Gene transcriptional regulation network describes the regulatory relationship among transcription factors (TFs) and target genes in systems biology. It is reported in [21, 39] that TFs often act in combination to form TF complexes and control the transcription of target genes collaboratively. Hence this kind of networks intrinsically has a group structure, that is, a TF complex consists of several collaborative TFs, which are all active for the transcription of a certain target gene. Gene regulatory network inference is the process to search a small number of TF complexes (or TFs) from a pool of thousands of TF complexes (or TFs) for the transcription of each target gene. In mathematics, the solution of gene regulatory network has a natural grouping of its components such that the components (i.e., TFs) within each group (i.e., TF complex) are likely to be either all zeros or all nonzeros, and this solution has only a small number of nonzero groups (i.e., active TF complexes). Therefore, gene regulatory network inference can be understood as a group feature selection problem based on the dependencies between the expression of TF complexes (or TFs) and that of target gene.

3.1 Description of real data

The aim of our numerical study is to predict the gene regulatory network in mouse embryonic stem cell (mESC) by considering the TF complex (group) information. Expression of TFs and target genes have been measured in a genome-wide scale under multiple biological experiments on mESC, and the expression data of mESC are downloaded from http://jjwanglab.org/LpRGNI/. In particular, the matrix $A \in \mathbb{R}^{245\times 939}$ includes the log2 transformed gene expression fold changes between control and TF perturbation samples of 939 TFs in 245 experiments. Each row of A is the expression profile of 939 TFs in each experiment, and each column of A is the expression profile of each TF in 245 experiments.

The matrix $B \in \mathbb{R}^{245\times 12488}$ includes the log2 transformed gene expression fold changes between control and TF perturbation samples of 12488 target genes in 245 experiments. Each row of B is the expression profile of 12488 target genes in each experiment, and each column of B is the expression profile of each target gene in 245 experiments (Figure 1A). Let the matrix $X \in \mathbb{R}^{939\times 12488}$ denote the regulatory relationship of all the TF-target gene pairs, which is to be predicted. Then, the regulatory relationship between TFs and target genes can be represented approximately by a linear system

$$AX = B + \varepsilon. \tag{3.1}$$

To guide the search of biologically meaningful solution, ChIP-seq (Chromatin immunoprecipitation followed by sequencing) data are converted into an initial matrix X^0 (cf. [32]) by calculating a prior value for each TF-target gene pair, according to the presence or absence of a binding site of each TF within the promoter of each target gene. The data of X^0 are also downloaded from http://jjwanglab.org/LpRGNI/.

Two independent golden standards, named low-throughput golden standard and high-throughput golden standard, are used to evaluate the accuracy of the inferred gene regulatory network. They are downloaded from iScMiD[2] and ChIP-Array[3], respectively. Low-throughput golden standard includes 97 TF-target gene interactions, which has been verified by biological experiments. High-throughput golden standard contains 40006 TF-target gene interactions evidenced by a *in vivo* binding site of the TF on the target gene's promoter and the expression change of the target gene under the perturbation of the TF (Figure 1C); cf. [33, 37].

3.2 Group structure in gene regulatory networks

Several bioinformatics methods and optimization algorithms have been developed for inferring gene regulatory networks, such as NARROMI [45], ISTA [17], YALL1 [42], etc. However, most of the existing methods usually consider each TF separately and only select TFs at the individual feature level (see, e.g., [32, 45]), and their performance is not satisfactory. It was demonstrated in Figure 2 (i.e., [32, Figure 2]) that they are of poor performance in a genome-wide scale inference, in which the AUCs[4] are close to that of a random prediction on both evaluation data. The poor performance of the existing methods might stem from the fact that the group structure of TF complexes is not exploited in gene regulatory network inference; see, e.g., [21, 39].

It is well-known in the machine learning literature that exploiting the group structure can reduce the degrees of freedom in the solution, thereby leading to better performance; see [1, 22, 44] and references therein. Inspired by this idea, we add the group structure of TF complexes into the linear system (3.1) to improve the accuracy of inference as follows. As remarked in [19], the TF complex information can be inferred from the ChIP-seq data. Following the method proposed in [19], we infer the group structure (TF complexes) from

[2]The iScMiD data is available at http://amp.pharm.mssm.edu/iscmid/.

[3]The ChIP-Array data is available at http://jjwanglab.org/chip-array/.

[4]The area under the curve (AUC) of a receiver operating characteristic (ROC) curve is widely recognized as an important index of the overall classification performance of an algorithm; see [18]. In general, the larger the AUC, the better the performance.

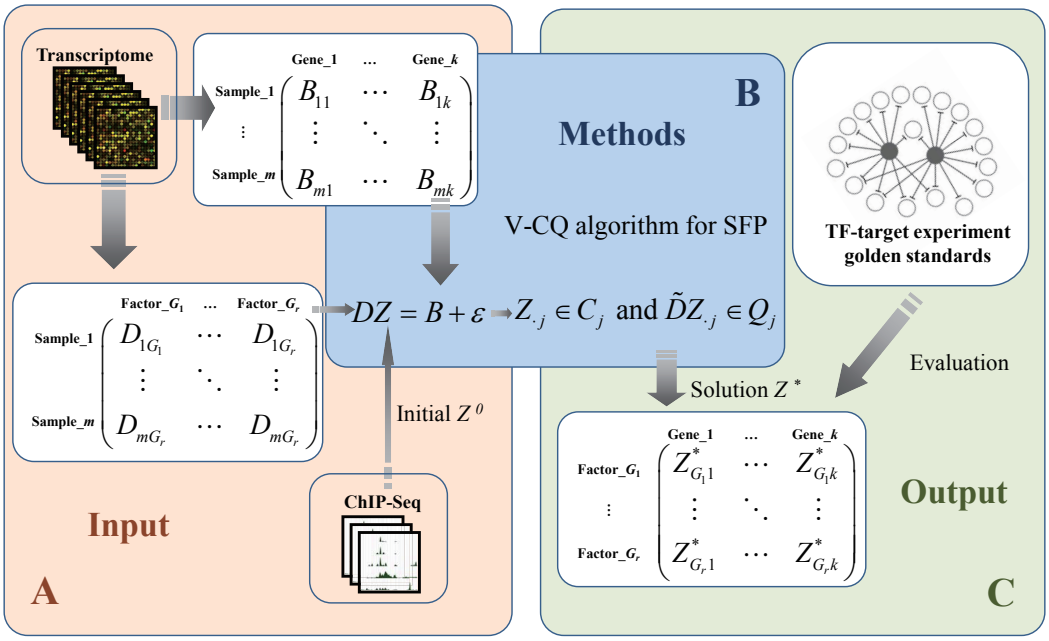
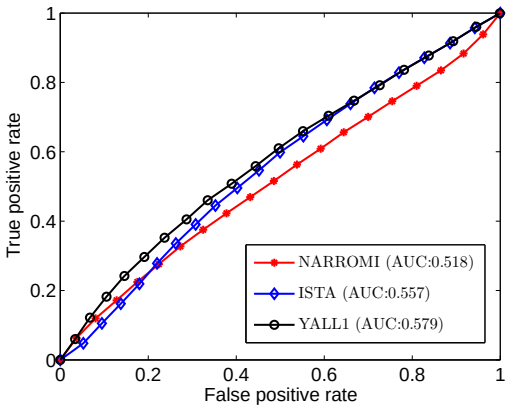
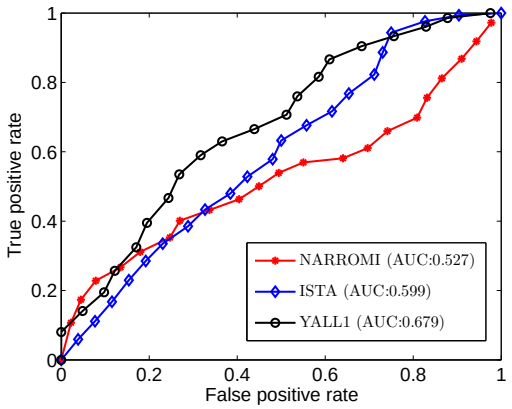


Figure 1: Workflow of gene regulatory network inference via SFP.



(a) Evaluation with high-throughput golden standard.



(b) Evaluation with literature-based low-throughput golden standard.

Figure 2: ROC curves and AUCs of existing methods on gene regulatory network inference.

ChIP-seq data of mESC. In detail, note that multiple TFs that bind at the same position in the genome are regarded as a TF complex. By this principle, we collect the binding sites of TFs from ChIP-seq/chip data of mESCs (see [32, Table 2] and [37, Table S3]), and define a TF complex by consisting of the collaborative TFs, whose binding sites are overlapped in the genome. Following this, we obtain 500 candidates of TF complexes in mESC as shown in Figure 3, where each row denotes a TF complex and each column symbolizes a TF component. In Figure 3, a pixel is marked as *blue* if the TF component is included

in the TF complex; otherwise, it is marked as *white*; we observe from Figure 3 that there may be overlaps in different TF complexes. Therefore, after adding the group structure of TF complexes, the regulatory relationship between the TF complexes and target genes is denoted by a matrix $Z \in \mathbb{R}^{2257\times 12488}$, in which each row denotes the regulatory relationship between a TF in TF complex and target genes. Let the group structure of mESC be denoted by an indicator matrix $W \in \mathbb{R}^{2257\times 939}$, in which $W_{ij} = 1$ if the i-th TF in Z (i.e., i-th row of Z) is the j-th TF in X, and equals to zero otherwise. Then it follows that

$$Z = WX, \tag{3.2}$$

which converts the gene regulatory network between TFs and genes to those between TF complexes and genes. We further use W^+ to denote the Moore-Penrose pseudoinverse of W, and let $D := AW^+$ and $Z^0 := WX^0$. Then the linear system (3.1) can be converted into

$$DZ = B + \varepsilon, \tag{3.3}$$

where $D \in \mathbb{R}^{245\times 2257}$ denotes expression profiles of TF complexes (Figure 1A).

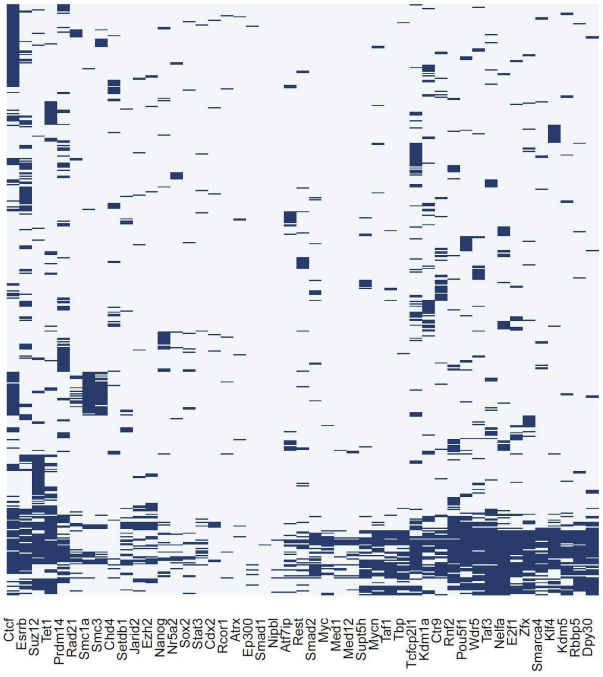


Figure 3: TF complexes in mESC. Each row is a TF complex, and each column is a TF component (marked with a symbol). Blue denotes that the TF component is included in the TF complex, and white denotes that it is absent.

Since the expression data of mESC consists of 12488 target genes, system (3.3) consists of 12488 linear inverse problems, each of which is to infer regulatory network for a certain target gene:

$$DZ_{\cdot j} = B_{\cdot j} + \varepsilon_{\cdot j}, \quad j = 1, 2, \ldots, 12488, \tag{3.4}$$

where $Z_{\cdot j} \in \mathbb{R}^{2257}$ is the j-th column of Z, denoting the regulatory relationship between TF complexes and the j-th target gene, and $B_{\cdot j} \in \mathbb{R}^{245}$ is the j-th column of B, designating the expression profile of the j-th target gene in the experiments. Hence, gene regulatory network inference of mESC is to approach the group sparse solutions of a series of problem (3.4), according to the expression data of TF complexes and that of target genes. Fixing $j \in \{1, 2, \ldots, 12488\}$, inferring regulatory network of j-th target gene (3.4) can be formulated as a group Dantzig selector [29] with the prior knowledge of TF complexes being the predefined group structure, which is to solve the following problem

$$\begin{array}{ll} \min & \|Z_{\cdot j}\|_{\infty,1} \\ \text{s.t.} & \|D^\top (DZ_{\cdot j} - B_{\cdot j})\|_\infty \leq \epsilon_j, \end{array} \tag{3.5}$$

where $\|x\|_{p,1} := \sum_{i=1}^{r} \|\omega_i\|_\infty$ is a group sparsity promoting norm (with $x := (\omega_1^\top, \ldots, \omega_r^\top)^\top$ standing for the group structure of x). By letting $\tilde{D} := D^\top D$ and $\tilde{B}_{\cdot j} := D^\top B_{\cdot j}$, and selecting suitable parameter δ_j, the group Dantzig selector (3.5) could be approached by solving the following SFP

$$Z_{\cdot j} \in C_j := \{Z_{\cdot j} : \|Z_{\cdot j}\|_{\infty,1} \leq \delta_j\} \quad \text{and} \quad \tilde{D}Z_{\cdot j} \in Q_j := \overline{\mathbf{Box}(\tilde{B}_{\cdot j}, \epsilon_j)}. \tag{3.6}$$

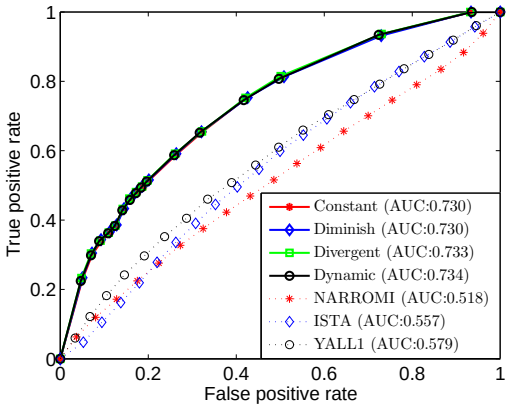
Therefore, due to (3.4), gene regulatory network inference of mESC can be formulated as a series of SFPs (Figure 1B).

We apply the V-CQ algorithm to sequentially solve SFPs ((3.6) with $j = 1, 2, \ldots, 12488$), and hence to infer the gene regulatory network of mESC. All numerical experiments are implemented in MATLAB R2009a and executed on a personal desktop (Intel Core Duo E8500, 3.16 GHz, 4.00 GB of RAM). It is worth noting that we use the $\ell_{\infty,1}$ norm to characterize the group sparsity in the group Dantzig selector (3.5) so as to achieve the high efficiency of the V-CQ algorithm. Indeed, the feasible sets C_j and Q_j in (3.6) are polyhedrons, and thus, it follows from Corollary 2.6 that the V-CQ algorithm linearly converges to a feasible solution of (3.6). The projection onto C_j, an $\ell_{\infty,1}$ ball, is implemented by an efficient algorithm from [34]. In the following subsection, we will show the numerical results of applying the V-CQ algorithm to solve the SFPs and validating by the biological golden standards.

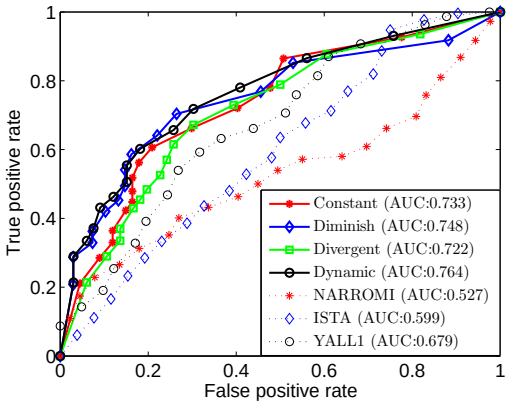
3.3 Numerical results

In this numerical study, we conduct several numerical experiments to show the performance of the V-CQ algorithm in inferring gene regulatory networks of mESC, compare with some state-of-the-art algorithms, and demonstrate the sensitivity analysis of the V-CQ algorithm on initial points and stepsizes, by evaluation with the biological golden standards. Note that these two biological golden standards only measure the regulations between TFs and genes; however, the biological golden standards on regulations between TF complexes and genes are not available at this moment (because, in most biological experiments, only regulations between single TF and its target genes are investigated). Hence, to evaluate the obtained results, we convert the inferred regulations between TF complexes and genes to those between TFs and genes, that is, $X = W^+ Z$ (by (3.2)), and then use the available golden standards to evaluate our results. To calculate the AUC of this X, a score $\text{Score}_{ij} := |X_{ij}|$ is adopted as the predictor for TF i on target gene j.

The first experiment is to show the performance of the V-CQ algorithm to solve SFPs ((3.6) with $j = 1, 2, \ldots, 12488$), by using four types of stepsizes[5] and starting from the initial matrix Z^0, and compare the numerical results with that of the existing methods via matching with the biological golden standards. Matching with the high-throughput and low-throughput golden standards, we draw in Figure 4 the ROC curves of the V-CQ algorithm with the four types of stepsizes, as well as that of the existing methods, to evaluate their accuracy. It is illustrated from Figure 4 that the V-CQ algorithm (plotted in solid) using the four types of stepsizes perform almost the same (as indicated by the almost same AUC value), and significantly outperforms the existing methods (plotted in dots) in the sense that there is an increase of at least 30% on the AUC value. Note that the golden standards we adopt here are obtained from biological experiments, which are well-accepted as true TF-target gene regulations. The higher the AUC, the more biologically accurate the obtained gene regulatory network is. Hence, the numerical result in Figure 4 shows that exploiting the group structure of TF complexes can improve the accuracy of the gene regulation network forecasting, and that the V-CQ algorithm is applicable to gene regulatory network inference in the sense that it obtains more biologically accurate solutions.



(a) Evaluation with high-throughput golden standard.



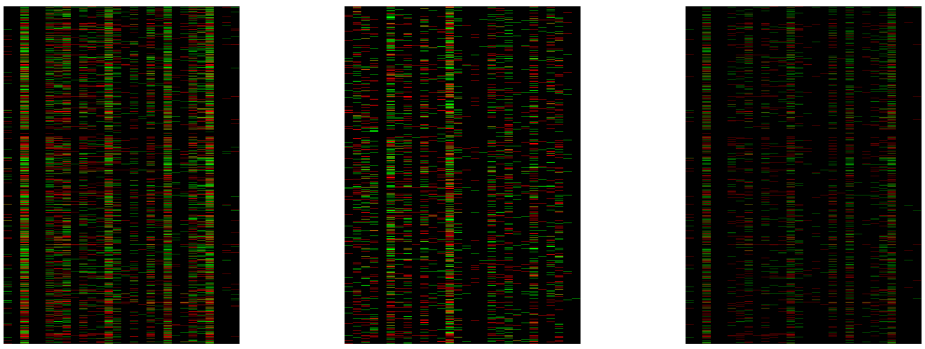
(b) Evaluation with literature-based low-throughput golden standard.

Figure 4: ROC curves and AUCs of the V-CQ algorithm on gene regulatory network inference of mESC.

To illustrate the obtained results, we display the heat maps of the golden stardard, initial X^0 and solution X of the V-CQ algorithm respectively in Figure 5, and the heat maps of initial Z^0 and solution Z of the V-CQ algorithm in Figure 6. It was reported in [32, Table 3] that the high-throughput golden standard is mainly constructed for 28 TFs, including Cdx2, Ctr9, Esrrb, Jarid2, Kdm1a, Klf4, Myc, Mycn, Nacc1, Nanog, Nr0b1, Nr5a2, Pou5f1, Rest, Sall4, Sfpi1, Smad1, Sox2, Stat3, Suz12, Tbx3, Tcf3, Tcfcp2l1, Trim28, Wdr5, Whsc2, Zfp281, Zfp42. Hence, the heat maps of these 28 TFs are plotted in Figure 5 and the ones of

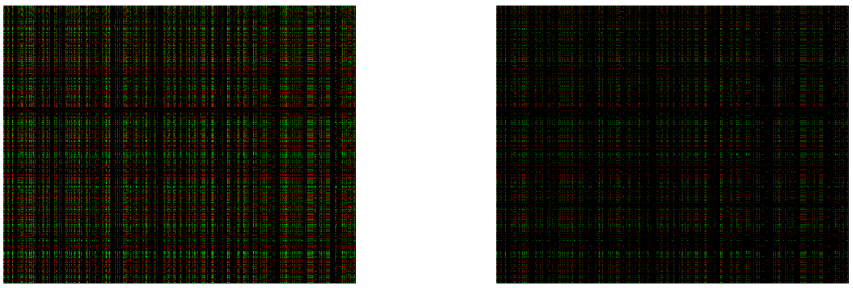
[5]Constant stepsize $\beta_n = 1$; Diminishing stepsize $\beta_n = n^{-0.1}$; Divergent stepsize $\beta_n = \frac{1}{1+n}$; Dynamic stepsize (1.2) with $\rho_n = 1$. Note that both diminishing stepsize and divergent stepsize satisfy condition (2.6).

TF complexes that include at least one of the 28 TFs are plotted in Figure 6, and note that the heat maps plot the transpositions of the relevant matrices. We can observe that X^0 has much more false positives than X, so the gene regulatory network gets higher accuracy after calculation. The AUCs of X and X^0 are listed in Table 3 later, in which it is shown that X is better than X^0 matching with the high-throughput biological standard.



(a) Initial X^0. (b) Golden standard. (c) Solution X of V-CQ.

Figure 5: Heat maps of gene regulatory networks of key factors.



(a) Initial Z^0. (b) Solution Z of V-CQ.

Figure 6: Heat maps of gene regulatory networks of key TF complexes.

We also present the convergence rate of the V-CQ algorithm when inferring gene regulatory networks of mESC. Note that the biological golden standard consists of only part of gene regulatory network, and that the true solution of problem (3.6) is unknown. Alternatively, we adopt the violation of the SFPs (3.6), denoted by $\mathrm{d}_{Q_j}(\tilde{D}Z_{\cdot j})$, to evaluate the convergence behavior of the V-CQ algorithm. The error bar along with the number of iterations is plotted in Figure 7. It is demonstrated from Figure 7 the V-CQ algorithm for inferring gene regulatory networks of mESC linearly converges to a feasible solution when using the four types of stepsizes, which is consistent with the theoretical results provided in Corollary 2.6.

The second experiment is to compare the V-CQ algorithm with the group Lasso [44] and

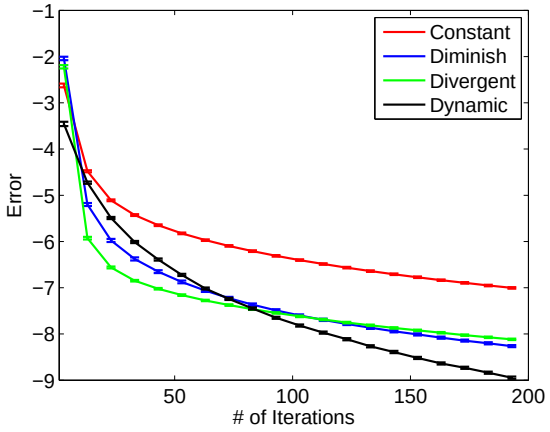


Figure 7: The convergence rate of the V-CQ algorithm with four types of stepsizes.

group Dantzig selector [29] for inferring gene regulatory networks of mESC. The computation results are displayed in Table 1 with the best one of each method being marked as *red*. In this table, the columns of Method and Parameter represent the selected method and the different parameters used in the method, in which the parameter in the V-CQ algorithm is $\delta_j = \eta \|Z^0_{\cdot j}\|_{\infty,1}$ (see (3.6)), parameter λ in the group Lasso is the regularization parameter (see [44]), and parameter ϵ in group Dantzig selector is the one in (3.5); the columns of AUCs (high) and AUCs (low) represent the AUC evaluated with the high-throughput and low-throughput golden standards, respectively; and the column of CPU time denotes the averaged CPU time (hours) cost to accomplish the gene regulatory networks inference of the whole genome. From Table 1, it is observed that the V-CQ algorithm outperforms the group Lasso and group Dantzig selector in the sense that it obtains more reliable solution matching with biological standards, since its solution has a larger AUC value. It is also shown that the V-CQ algorithm costs more CPU time than the group Lasso does, and much less than the group Dantzig selector does. This is because the group Lasso is to solve an unconstrained optimization problem, which avoids the computation of projection; while the group Dantzig selector requires to compute the projection onto its constraint set at each iteration, which costs more time than calculating the projection onto an $\ell_{\infty,1}$ ball does.

The third experiment is to analyze the sensitivity of the V-CQ algorithm on stepsizes or initial points, in terms of AUC values. Table 2 shows the variation of AUCs for different types of stepsizes. Three observations are indicated from Table 2: (i) the convergence of the V-CQ algorithm may fail when the stepsize is too large, which is consistent with Theorem 2.3(a); (ii) the V-CQ algorithm with diminishing stepsize or divergent stepsize converges to a feasible solution of the SFP, which verifies Theorem 2.3(c), since both diminishing stepsize and divergent stepsize satisfy condition (2.6); (iii) the dynamic stepsize V-CQ algorithm converges to a feasible solution of the SFP, which is consistent with Theorem 2.3(b), and the best dynamic stepsize is the one satifying (1.2) with $\rho_n = 1$. Table 3 lists AUCs of the V-CQ algorithm with different initial points, as well as their AUCs. The initial points "$X10$k" (i.e., X^0 used in the preceding experiments), "$X200$" and "$X50$k" are taken from [32]. In particular, "$X200$" contains less false regulations but the less true regulations; "$X50$k"

Table 1: Comparison of the V-CQ algorithm and group Lasso/Dantzig selector.

Method	Parameter	AUCs (high)	AUCs (low)	CPU time
V-CQ algorithm	$\eta = 1$	0.734	0.764	10.4 hours
	$\eta = 1.1$	0.660	0.755	
	$\eta = 1.2$	0.658	0.753	
group Lasso	$\lambda =$ e-2	0.504	0.586	7.1 hours
	$\lambda =$ e-3	0.583	0.680	
	$\lambda =$ e-4	0.701	0.748	
	$\lambda =$ e-5	0.645	0.751	
	$\lambda =$ e-6	0.636	0.750	
group Dantzig selector	$\epsilon =$ e-1	0.576	0.619	30.1 hours
	$\epsilon =$ e-2	0.625	0.627	
	$\epsilon =$ e-3	0.675	0.726	
	$\epsilon =$ e-4	0.597	0.603	
	$\epsilon =$ e-5	0.584	0.673	

covers more true regulations but more false regulations; while "X10k" has the specificity and sensitivity between the above two, whose AUC is 0.534 (see [32, Table 5]). The initial point "Random" denotes an i.i.d. Gaussian ensemble, and "Zeros" denotes a vector of zeros. Table 3 illustrates that the V-CQ algorithm may converge to different feasible solutions of the SFP when starting from different initial points, and it obtains a solution of large AUC value when starting from an initial point that is of biological sense.

Table 2: Sensitivity of the V-CQ algorithm on stepsizes.

Type of stepsize	Stepsize	AUCs (high)	AUCs (low)
Constant stepsize	$\beta_n = 0.5$	0.719	0.725
	$\beta_n = 1$	0.730	0.733
	$\beta_n = 5$	0.588	0.567
Diminishing stepsize	$\beta_n = n^{-0.1}$	0.730	0.748
	$\beta_n = n^{-0.3}$	0.728	0.722
	$\beta_n = n^{-0.5}$	0.714	0.691
Divergent stepsize	$\beta_n = \frac{1}{1+n}$	0.733	0.728
	$\beta_n = \frac{1}{1+2n}$	0.706	0.686
	$\beta_n = \frac{1}{1+3n}$	0.597	0.579
Dynamic stepsize (1.2)	ρ_n=0.5	0.731	0.747
	ρ_n=1	0.734	0.764
	ρ_n=1.5	0.732	0.715

4 Conclusion

We introduced a notion of bounded linear regularity property for the SFP and, under the assumption of this property, established the convergence rate for the V-CQ algorithm with

Table 3: Sensitivity of the V-CQ algorithm on initial points.

Initial point	AUCs of V-CQ (high)	AUCs of V-CQ (low)	AUCs of initials
$X10k$	0.734	0.764	0.534
$X200$	0.622	0.646	0.531
$X50k$	0.681	0.747	0.533
Random	0.511	0.551	-
Zeros	0.449	0.327	-

the bounded, diminishing or dynamic stepsizes; in particular, the linear convergence rate for the V-CQ algorithm with the stepsizes satisfying (2.4) or (2.5) was obtained. Some sufficient conditions ensuring the bounded linear regularity property were provided, which are satisfied in many application problems, especially covering the mathematics problem arising from gene regulatory network inference discussed in this paper. As an application in systems biology, the gene regulatory network inference was formulated as a series of SFPs, by virtue of the group structure of TF complexes. The numerical study on gene expression data of mESC showed that the V-CQ algorithm is applicable to gene regulatory network inference, and that it outperforms the group Lasso and group Dantzig selector in the sense that it obtains a more reliable solution matching with biological standards. This study may facilitate biologists to study the gene regulation mechanism of higher model organisms in a genome-wide scale.

Acknowledgment. The authors are grateful to the anonymous reviewers for their valuable suggestions and remarks that helped to improve the quality of the paper. The authors are also thankful to Professor Junwen Wang (Mayo Clinic) and Dr. Jing Qin (The Chinese University of Hong Kong) for providing the gene expression data and biological golden standards for inferring gene regulatory network of mouse embryonic stem cell.

A Appendix: Proof of Proposition 2.5

To furniture the proof of Proposition 2.5, we recall some known facts in the following proposition: assertion (i) is taken from [3, Corollary 5.26]; while assertion (ii) is known in [2, Propositions 4.6.1 and 4.6.2] for the finite-dimensional space setting and in [48, Proposition 3.6] for the infinite-dimensional space setting.

Proposition A.1. *Let C and D be two closed convex subsets of Hilbert space H. Then the following assertions hold:*

(i) *If C and D are polyhedrons, then there exists $\gamma > 0$ such that*

$$\mathrm{d}_{C\cap D}(x) \leq \gamma \max\{\mathrm{d}_C(x), \mathrm{d}_D(x)\} \tag{A.1}$$

holds for all $x \in H$.

(ii) *If either $C \cap \mathrm{ri}D \neq \emptyset$ and C is a polyhedron, or $\mathrm{ri}C \cap \mathrm{ri}D \neq \emptyset$ and D is finite-codimensional, then for any $r > 0$, there exists $\gamma_r > 0$ such that* (A.1) *holds for any $x \in r\mathbb{B}$ with $\gamma_r > 0$ in place of γ.*

Now, we provide a complete proof of Proposition 2.5.

Proof of Proposition 2.5. (i) Suppose that C and Q are polyhedrons. Then C and $A^{-1}Q$ are also polyhedrons. It follows from Proposition A.1(i) that there exists $\gamma > 0$ such that

$$\mathrm{d}_S(x) \leq \gamma \mathrm{d}_{A^{-1}Q}(x) \quad \text{for any } x \in C. \tag{A.2}$$

We will show the following implication:

$$Q \text{ is a polyhedron} \Rightarrow \text{there exists } \alpha > 0 \text{ such that } \mathrm{d}_{A^{-1}Q}(x) \leq \alpha \mathrm{d}_Q(Ax), \forall x \in H_1. \tag{A.3}$$

To do this, since Q is a polyhedron, Q can be represented as $Q = \cap_{i=1}^m Q_i$ with Q_i being given by

$$Q_i := \{y \in H_2 : \langle s_i, y\rangle \leq r_i\},$$

where $s_i \in H_2$ and $r_i \in \mathbb{R}$ for each $i \in \{1, \ldots, m\}$. Hence $A^{-1}Q = \bigcap_{i=1}^m A^{-1}Q_i$. Fix $i \in \{1, \ldots, m\}$, and note that

$$A^{-1}Q_i = \{x : \ \langle s_i, Ax\rangle \leq r_i\} = \{x : \ \langle A^* s_i, x\rangle \leq r_i\}.$$

Without loss of generality, we may assume that $A^* s_i \neq 0$ (and so $s_i \neq 0$). Then

$$\mathrm{d}_{A^{-1}Q_i}(x) = \frac{\|s_i\|}{\|A^* s_i\|} \mathrm{d}_{Q_i}(Ax) \quad \text{for any } x \in H_1. \tag{A.4}$$

Indeed, it is trivial if $x \in A^{-1}Q_i$; otherwise, one has that

$$\mathrm{d}_{A^{-1}Q_i}(x) = \frac{|\langle A^* s_i, x\rangle - r_i|}{\|A^* s_i\|} = \frac{\|s_i\|}{\|A^* s_i\|} \frac{|\langle s_i, Ax\rangle - r_i|}{\|s_i\|} = \frac{\|s_i\|}{\|A^* s_i\|} \mathrm{d}_{Q_i}(Ax).$$

Moreover, applying inductively Proposition A.1(i), one concludes that there exists a constant $\tilde{\gamma} > 0$ such that

$$\mathrm{d}_{A^{-1}Q}(x) \leq \tilde{\gamma} \max_{1 \leq i \leq m} \mathrm{d}_{A^{-1}Q_i}(x) \quad \text{for any } x \in H_1.$$

This, together with (A.4), implies that

$$\mathrm{d}_{A^{-1}Q}(x) \leq \tilde{\gamma} \max_{1 \leq i \leq m} \frac{\|s_i\|}{\|A^* s_i\|} \mathrm{d}_{Q_i}(Ax) \leq \tilde{\gamma} \max_{1 \leq i \leq m} \frac{\|s_i\|}{\|A^* s_i\|} \mathrm{d}_Q(Ax).$$

Thus, we established the implication (A.3). This, together with (A.2), entails (2.2), and thus the SFP (1.1) satisfies the bounded linear regularity property.

(ii) Suppose that $AC \cap \mathrm{int} Q \neq \emptyset$. Let $x_0 \in C$ be such that $Ax_0 \in \mathrm{int} Q$. Then there exists $\delta > 0$ such that

$$\overline{\mathbf{B}(Ax_0, \delta)} \subseteq Q. \tag{A.5}$$

Let $x \in C$, and write

$$y := P_{Q \cap AC}(Ax) \quad \text{and} \quad z := \frac{\|Ax - y\|}{\|Ax - y\| + \delta} x_0 + \frac{\delta}{\|Ax - y\| + \delta} x;$$

$$\hat{y} := P_Q(Ax) \quad \text{and} \quad w := \frac{\|Ax-\hat{y}\|}{\|Ax-\hat{y}\|+\delta}Ax_0 + \frac{\delta}{\|Ax-\hat{y}\|+\delta}Ax.$$

Consequently, $z \in C$ and $w \in AC$. We further have that

$$z \in S \quad \text{and} \quad w \in AC \cap Q.$$

To show this, we set $\bar{x} \in A^{-1}(y)$, and write $\bar{z} := x_0 + \frac{\delta}{\|Ax-y\|}(x-\bar{x})$. Then it follows that

$$\|A\bar{z} - Ax_0\| = \delta \quad \text{and} \quad z = \frac{\|Ax-y\|}{\|Ax-y\|+\delta}\bar{z} + \frac{\delta}{\|Ax-y\|+\delta}\bar{x}.$$

Hence, $A\bar{z} \in Q$ by (A.5) and so

$$Az = \frac{\|Ax-y\|}{\|Ax-y\|+\delta}A\bar{z} + \frac{\delta}{\|Ax-y\|+\delta}A\bar{x} \in Q.$$

This yields that $z \in S$ because $z \in C$ as noted earlier. Thus, it follows that

$$\mathrm{d}_S(x) \le \|x-z\| = \frac{\|Ax-y\|}{\|Ax-y\|+\delta}\|x-x_0\| = \frac{\|x-x_0\|}{\|Ax-y\|+\delta}\mathrm{d}_{Q\cap AC}(Ax). \tag{A.6}$$

Write $\bar{w} := Ax_0 + \frac{\delta}{\|Ax-\hat{y}\|}(Ax-\hat{y})$. Then, $\|\bar{w} - Ax_0\| = \delta$ and so $\bar{w} \in Q$. It follows that

$$w = \frac{\|Ax-\hat{y}\|}{\|Ax-\hat{y}\|+\delta}\bar{w} + \frac{\delta}{\|Ax-\hat{y}\|+\delta}\hat{y} \in Q.$$

This shows that $w \in AC \cap Q$, since $w \in AC$ as mentioned earlier. Hence we obtain that

$$\mathrm{d}_{Q\cap AC}(Ax) \le \|Ax-w\| = \frac{\|Ax-\hat{y}\|}{\|Ax-\hat{y}\|+\delta}\|Ax-Ax_0\| \le \frac{\|A\|\|x-x_0\|}{\|Ax-\hat{y}\|+\delta}\mathrm{d}_Q(Ax).$$

This, together with (A.6), yields that

$$\mathrm{d}_S(x) \le \frac{\|x-x_0\|}{\|Ax-y\|+\delta}\frac{\|A\|\|x-x_0\|}{\|Ax-\hat{y}\|+\delta}\mathrm{d}_Q(Ax).$$

This implies, for any $r > 0$, that

$$\mathrm{d}_S(x) \le \gamma_r \mathrm{d}_Q(Ax) \quad \text{for any } x \in C \cap \overline{\mathbf{B}(x_0, r)},$$

where $\gamma_r := \frac{\|A\|r^2}{\delta^2}$. This shows that the SFP (1.1) satisfies the bounded linear regularity property.

(iii) Suppose that $A(\mathrm{ri}C) \cap Q \neq \emptyset$ and Q is a polyhedron. Then $\mathrm{ri}C \cap (A^{-1}Q) \neq \emptyset$. Let $r > 0$. It follows from Proposition A.1(ii) that there exists a constant $\gamma_r > 0$ such that

$$\mathrm{d}_S(x) \le \gamma_r \mathrm{d}_{A^{-1}Q}(x) \quad \text{for any } x \in C \cap r\mathbb{B}. \tag{A.7}$$

It follows from (A.3) that there exists $\alpha > 0$ such that $\mathrm{d}_{A^{-1}Q}(x) \le \alpha \mathrm{d}_Q(Ax)$ for all $x \in H_1$. This, together with (A.7), implies that

$$\mathrm{d}_S(x) \le \gamma_r \alpha \mathrm{d}_Q(Ax) \quad \text{for any } x \in C \cap r\mathbb{B}.$$

Hence, the SFP (1.1) satisfies the bounded linear regularity property.

(iv) Suppose that $AC \cap \mathrm{ri}Q \neq \emptyset$, C is a polyhedron and Q is finite-codimensional. We claim that

$$C \cap \mathrm{ri}(A^{-1}Q) \neq \emptyset. \tag{A.8}$$

Indeed, by assumption, we can choose $x_0 \in C$ such that $Ax_0 \in \mathrm{ri}Q$. This means that there exists $\delta > 0$ such that $\mathbf{B}(Ax_0, \delta) \cap \mathrm{aff}Q \subseteq Q$. Consequently,

$$A^{-1}(\mathbf{B}(Ax_0, \delta)) \cap A^{-1}(\mathrm{aff}Q) = A^{-1}(\mathbf{B}(Ax_0, \delta) \cap \mathrm{aff}Q) \subseteq A^{-1}(Q).$$

Since by definition that $\mathbf{B}(x_0, \delta/\|A\|) \subseteq A^{-1}(\mathbf{B}(Ax_0, \delta))$ and that $\mathrm{aff}(A^{-1}Q) \subseteq A^{-1}(\mathrm{aff}Q)$, it follows that

$$\mathbf{B}(x_0, \delta/\|A\|) \cap \mathrm{aff}(A^{-1}(Q)) \subseteq A^{-1}(\mathbf{B}(Ax_0, \delta)) \cap A^{-1}(\mathrm{aff}Q) \subseteq A^{-1}(Q).$$

This shows that $x_0 \in \mathrm{ri}(A^{-1}Q)$ and so (A.8) is proved. Let $r > 0$. Proposition A.1(ii) is applicable to concluding that there exists $\gamma_r > 0$ such that (A.7) holds. Thus we only need to prove that there exists $\alpha_r > 0$ such that

$$\mathrm{d}_{A^{-1}Q}(x) \leq \alpha_r \mathrm{d}_Q(Ax) \quad \text{for all } x \in r\mathbb{B}. \tag{A.9}$$

To do this, write $Z := \mathrm{aff}\mathrm{Q}$. Then, Z is a polyhedron and, by (A.3), there exists $\alpha > 0$ such that

$$\mathrm{d}_{A^{-1}Z}(x) \leq \alpha \mathrm{d}_Z(Ax) \leq \alpha \mathrm{d}_Q(Ax) \quad \text{for any } x \in H_1. \tag{A.10}$$

Since $\mathrm{ri}Q \neq \emptyset$ by assumption, it follows from [27, Lemma 3.1] that there exists a closed convex subset $\tilde{Q} \subseteq H_2$ such that

$$Ax_0 \in \mathrm{int}\tilde{Q}, \quad \mathrm{int}\tilde{Q} \neq \emptyset \quad \text{and} \quad \tilde{Q} \cap Z = Q \subseteq \tilde{Q}. \tag{A.11}$$

This clearly implies that $A^{-1}Q = A^{-1}Z \cap A^{-1}\tilde{Q}$ and $A(A^{-1}Z) \cap \mathrm{int}\tilde{Q} \neq \emptyset$. Thus the conclusion under assumption (ii) is application (to $A^{-1}Z$ in place of C), and we conclude that there exists $\tilde{\gamma}_r > 0$ such that

$$\mathrm{d}_{A^{-1}Q}(x) \leq \mathrm{d}_S(x) \leq \tilde{\gamma}_r \mathrm{d}_Q(Ax) \quad \text{for any } x \in (A^{-1}Z) \cap r\mathbb{B}.$$

Now fix $x \in r\mathbb{B}$ and set $\bar{x} := P_{A^{-1}Z \cap r\mathbb{B}}(x)$. Then $\|x - \bar{x}\| = \mathrm{d}_{A^{-1}Z \cap r\mathbb{B}}(x)$, and

$$\mathrm{d}_{A^{-1}Q}(x) \leq \|x - \bar{x}\| + \mathrm{d}_{A^{-1}Q}(\bar{x}) \leq \|x - \bar{x}\| + \tilde{\gamma}_r \mathrm{d}_Q(A\bar{x}) = \mathrm{d}_{A^{-1}Z \cap r\mathbb{B}}(x) + \tilde{\gamma}_r \mathrm{d}_Q(A\bar{x}). \tag{A.12}$$

Since $\|Ax - A\bar{x}\| \leq \|A\|\|x - \bar{x}\| = \|A\|\mathrm{d}_{A^{-1}Z \cap r\mathbb{B}}(x)$, it follows that

$$\mathrm{d}_Q(A\bar{x}) \leq \|Ax - A\bar{x}\| + \mathrm{d}_Q(Ax) \leq \|A\|\mathrm{d}_{A^{-1}Z \cap r\mathbb{B}}(x) + \mathrm{d}_Q(Ax). \tag{A.13}$$

Furthermore, by [28, Lemma 4.10], one has that $\mathrm{d}_{A^{-1}Z \cap r\mathbb{B}}(x) \leq 4\mathrm{d}_{A^{-1}Z}(x)$ (noting that $x \in r\mathbb{B}$). This, together with (A.12) and (A.13), implies that

$$\mathrm{d}_{A^{-1}Q}(x) \leq 4(1 + \|A\|\tilde{\gamma}_r)\mathrm{d}_{A^{-1}Z}(x) + \tilde{\gamma}_r \mathrm{d}_Q(Ax). \tag{A.14}$$

Thus, by (A.10), one sees that (A.9) hods with $\alpha_r := 4(1 + \|A\|\tilde{\gamma}_r)\alpha + \tilde{\gamma}_r$, and thus the proof for (iv) is complete.

(v) Suppose that $A(\mathrm{ri}C) \cap \mathrm{ri}Q \neq \emptyset$, and Q is finite-codimensional. Write $Z := \mathrm{aff}Q$ and let $r > 0$. Then, by the conclusion under assumption (iii) (applied to Z in place of Q), it follows that there exists $\gamma_r > 0$ such that

$$\mathrm{d}_{C\cap A^{-1}Z}(x) \leq \gamma_r \mathrm{d}_Z(Ax) \leq \gamma_r \mathrm{d}_Q(Ax) \quad \text{for any } x \in C \cap r\mathbb{B}. \tag{A.15}$$

Similar to the proof of (iv), since $\mathrm{ri}Q \neq \emptyset$, there exists a closed convex subset $\tilde{Q} \subseteq H_2$ satisfying (A.11). Consequently, $S = (C \cap A^{-1}Z) \cap A^{-1}\tilde{Q}$ and $A(C \cap A^{-1}Z) \cap \mathrm{int}\tilde{Q} \neq \emptyset$. Hence the conclusion under assumption (ii) is applicable (to $(C \cap A^{-1}Z)$ and $\tilde{Q}$ in place of C and Q), and we have that there exists $\tilde{\gamma}_r > 0$ such that

$$\mathrm{d}_S(x) \leq \tilde{\gamma}_r \mathrm{d}_{\tilde{Q}}(Ax) \leq \tilde{\gamma}_r \mathrm{d}_Q(Ax) \quad \text{for all } x \in r\mathbb{B} \cap (C \cap A^{-1}Z).$$

Fix $x \in r\mathbb{B} \cap C$. Then, replacing $A^{-1}Z$ by $(C \cap A^{-1}Z)$ and using the same arguments we did for (A.14), we have that

$$\mathrm{d}_S(x) \leq 4(1 + \|A\|\tilde{\gamma}_r)\mathrm{d}_{C\cap A^{-1}Z}(x) + \tilde{\gamma}_r \mathrm{d}_Q(Ax).$$

This, together with (A.15), shows that the SFP (1.1) satisfies the bounded linear regularity property. The proof is complete. □

References

[1] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Structured sparsity through convex optimization. *Statistics Science*, 27(4):450–468, 2012.

[2] H. H. Bauschke. *Projection Algorithms and Monotone Operators.* Ph. D. dissertation, Simon Fraser University, Canada, 1996.

[3] H. H. Bauschke and J. M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Rev.*, 38(3):367–426, 1996.

[4] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, New York, 2011.

[5] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Cambridge, 1999.

[6] C. Byrne. Iterative oblique projection onto convex sets and the split feasibility problem. *Inverse Problems*, 18(2):441–453, 2002.

[7] C. Byrne. A unified treatment of some iterative algorithms in signal processing and image reconstruction. *Inverse Problems*, 20(1):103–120, 2004.

[8] C. Byrne. *Iterative Optimization in Inverse Problems.* CRC Press, New York, 2014.

[9] J. V. Burke and M. C. Ferris. A Gauss-Newton method for convex composite optimization. *Mathematical Programming* 71:179–194, 1995.

[10] E. Candès and T. Tao. The Dantzig selector: Statistical estimation when p is much larger than n. *Annals of Statistics*, 35:2313–2351, 2007.

[11] A. Cegielski and F. Al-Musallam. Strong convergence of a hybrid steepest descent method for the split common fixed point problem. *Optimization*, 65(7):1463–1476, 2016.

[12] Y. Censor, T. Bortfeld, B. Martin, and A. Trofimov. A unified approach for inversion problems in intensity-modulated radiation therapy. *Physics in Medicine and Biology*, 51(10):2353–2365, 2006.

[13] Y. Censor and T. Elfving. A multiprojection algorithm using Bregman projections in a product space. *Numerical Algorithms*, 8(2):221–239, 1994.

[14] Y. Censor, T. Elfving, N. Kopf, and T. Bortfeld. The multiple-sets split feasibility problem and its applications for inverse problems. *Inverse Problems*, 21(6):2071–2084, 2005.

[15] Y. Censor and A. Segal. The split common fixed point problem for directed operators. *Journal of Convex Analysis*, 16(2):587–600, 2009.

[16] S. Chang, J. K. KimEmail, Y. J. Cho and J. Y. Sim. Weak- and strong-convergence theorems of solutions to split feasibility problem for nonspreading type mapping in Hilbert spaces. *Fixed Point Theory and Applications*, 1:11 pages, 2014.

[17] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57:1413–1457, 2004.

[18] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[19] E. G. Giannopoulou and O. Elemento. Inferring chromatin-bound protein complexes from genome-wide binding assays. *Genome Research*, 23(8):1295–1306, 2013.

[20] H. He, C. Ling, and H.-K. Xu. An implementable splitting algorithm for the ℓ_1-norm regularized split feasibility problem. *Journal of Scientific Computing*, pages 1–18, 2015.

[21] D. Hnisz *et al.* Super-enhancers in the control of cell identity and disease. *Cell*, 155(4):934–947, 2013.

[22] Y. Hu, C. Li, K. Meng, J. Qin, and X. Yang. Group sparse optimization via $\ell_{p,q}$ regularization. *Journal of Machine Learning Research*, to appear, 2017.

[23] Y. Hu, C. Li, and X. Yang. On convergence rates of linearized proximal algorithms for convex composite optimization with applications. *SIAM Journal on Optimization*, 26(2):1207–1235, 2016.

[24] Y. Hu, C.-K. Sim, and X. Yang. A subgradient method based on gradient sampling for solving convex optimization problems. *Numerical Functional Analysis and Optimization*, 36(12):1559–1584, 2015.

[25] Y. Hu, C. K. W. Yu and C. Li. Stochastic subgradient method for quasi-convex optimization problems. *Journal of Nonlinear and Convex Analysis*, 17(4):711–724, 2016.

[26] H. S. Hundal. An alternating projection that does not converge in norm. *Nonlinear Analysis:Theory, Methods & Applications*, 57(1):35–61, 2004.

[27] C. Li and K. F. Ng. Strong CHIP for infinite system of closed convex sets in normed linear spaces. *SIAM Journal on Optimization*, 16(2):311–340, 2005.

[28] C. Li, K. F. Ng, and T. K. Pong. The SECQ, linear regularity, and the strong CHIP for an infinite system of closed convex sets in normed linear spaces. *SIAM Journal on Optimization*, 18(2):643–665, 2007.

[29] H. Liu, J. Zhang, X. Jiang, and J. Liu. The group Dantzig selector. *Journal of Machine Learning Research-Proceedings Track*, 9:461–468, 2010.

[30] G. López, V. Martín-Márquez, F. Wang, and H.-K. Xu. Solving the split feasibility problem without prior knowledge of matrix norms. *Inverse Problems*, 28(8):085004, 2012.

[31] D. A. Lorenz, F. Schöpfer, and S. Wenger. The linearized Bregman method via split feasibility problems: Analysis and generalizations. *SIAM Journal on Imaging Sciences*, 7(2):1237–1262, 2014.

[32] J. Qin, Y. Hu, F. Xu, H. K. Yalamanchili, and J. Wang. Inferring gene regulatory networks by integrating ChIP-seq/chip and transcriptome data via LASSO-type regularization methods. *Methods*, 67(3):294–303, 2014.

[33] J. Qin, M. J. Li, P. Wang, M. Q. Zhang, and J. Wang. ChIP-Array: Combinatory analysis of ChIP-seq/chip and microarray gene expression data to discover direct/indirect targets of a transcription factor. *Nucleic Acids Research*, pages W430–436, 2011.

[34] A. Quattoni, X. Carreras, M. Collins, and T. Darrell. An efficient projection for $\ell_{1,\infty}$ regularization. *In International Conference on Machine Learning*, 2009.

[35] F. Schöpfer, T. Schuster, and A. K. Louis. An iterative regularization method for the solution of the split feasibility problem in banach spaces. *Inverse Problems*, 24(5):055008, 2008.

[36] J. Tang, S. Chang and F. Yuan. A strong convergence theorem for equilibrium problems and split feasibility problems in Hilbert spaces. *Fixed Point Theory and Applications*, 36:16 pages, 2014.

[37] P. Wang, J. Qin, Y. Qin, Y. Zhu, L. Y. Wang, M. J. Li, M. Q. Zhang, and J. Wang. ChIP-Array 2: Integrating multiple omics data to construct gene regulatory networks. *Nucleic Acids Research*, pages W264–9, 2015.

[38] F. Wang and H.-K. Xu. Cyclic algorithms for split feasibility problems in Hilbert spaces. *Nonlinear Analysis:Theory, Methods & Applications*, 74(12):4105–4111, 2011.

[39] D. Xie, A. P. Boyle, L. Wu, J. Zhai, T. Kawli, and M. Snyder. Dynamic trans-acting factor colocalization in human cells. *Cell*, 155(3):713–724, 2013.

[40] H.-K. Xu. A variable Krasnosel'skiĭ-Mann algorithm and the multiple-set split feasibility problem. *Inverse Problems*, 22(6):2021–2034, 2006.

[41] H.-K. Xu. Iterative methods for the split feasibility problem in infinite-dimensional hilbert spaces. *Inverse Problems*, 26(10):105018, 2010.

[42] J. Yang and Y. Zhang. Alternating direction algorithms for ℓ_1-problems in compressive sensing. *SIAM Journal on Scientific Computing*, 33(1):250–278, 2011.

[43] Q. Yang. The relaxed CQ algorithm solving the split feasibility problem. *Inverse Problems*, 20(4):1261–1266, 2004.

[44] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of The Royal Statistical Society, series B*, 68:49–67, 2006.

[45] X. Zhang *et al.* NARROMI: a noise and redundancy reduction technique improves accuracy of gene regulatory network inference. *Bioinformatics*, 29(1):106–113, 2013.

[46] J. Zhao and S. He. Strong convergence of the viscosity approximation process for the split common fixed-point problem of quasi-nonexpansive mappings. *Journal of Applied Mathematics*, Article ID 438023, 12 pages, 2012.

[47] J. Zhao and Q. Yang. Several solution methods for the split feasibility problem. *Inverse Problems*, 21(5):1791–1799, 2005.

[48] X. P. Zhao, K. F. Ng, C. Li and J.-C. Yao. Linear regularity and linear convergence of projection-based methods for solving convex feasibility problems. *Appl. Math. Opt.*, to appear.