RESEARCH   **Open Access**

Two spectral gradient projection methods for constrained equations and their linear convergence rate

Jing Liu[1†] and Yongrui Duan[2*†]

*Correspondence: yrduan@163.com
[2]School of Economics and Management, Tongji University, Siping Street, Shanghai, 200092, P.R. China
[†]Equal contributors
Full list of author information is available at the end of the article

Abstract

Due to its simplicity and numerical efficiency for unconstrained optimization problems, the spectral gradient method has received more and more attention in recent years. In this paper, two spectral gradient projection methods for constrained equations are proposed, which are combinations of the well-known spectral gradient method and the hyperplane projection method. The new methods are not only derivative-free, but also completely matrix-free, and consequently they can be applied to solve large-scale constrained equations. Under the condition that the underlying mapping of the constrained equations is Lipschitz continuous or strongly monotone, we establish the global convergence of the new methods. Compared with the existing gradient methods for solving such problems, the new methods possess a linear convergence rate under some error bound conditions. Furthermore, a relax factor γ is attached in the update step to accelerate convergence. Preliminary numerical results show that they are efficient and promising in practice.

Keywords: constrained equations; spectral gradient method; projection method; global convergence

1 Introduction

In this paper, we consider the problems of finding a solution of the following constrained equations, denoted by CES(F, C),

$$F(x^*) = 0 \quad \text{subject to} \quad x^* \in C, \tag{1}$$

where $F : C \to R^n$ is a given continuous nonlinear mapping and C is a nonempty closed convex set of R^n. Obviously, when $C = R^n$, (1) reduces to the nonlinear equations, which is intensively studied by many scholars. The constrained system of equations (1) appears in wide variety of problems in applied mathematics, and some important problems, such as economic equilibrium problems [1], power flow equations [2], and chemical equilibrium systems [3], can be reformulated as a problem of the kind (1).

Among various numerical methods for solving CES(F, C) [4–8], the gradient projection methods (GPMs) are the most efficient, especially when the projection onto the feasible set C is easy to implement. For example, when C is the nonnegative orthant, or a box, or a ball, GPMs require the lowest computational cost. In addition, the GPMs are also

the simplest, because they do not need to store any matrix during the iteration process. Therefore, they are completely matrix-free, and consequently, they can be applied to solve large-scale CES(F, C).

It is well known that the spectral gradient method [9, 10] and the conjugate gradient method [11] are two efficient methods for solving large-scale unconstrained optimization problems due to their simplicity and low storage. Recently, combined with the projection technique, they are extended to solve constrained equations CES(F, C) by some scholars [6, 7]. In [6], Yu *et al.* proposed a spectral gradient projection method for solving monotone CES(F, C), which can be applied to nonsmooth constrained equation, and works quite well even for large-scale CES(F, C). Quite recently, Liu *et al.* [7] developed two unified frameworks of some sufficient descent conjugate gradient projection methods for solving monotone CES(F, C), which are also applied to solve large-scale nonsmooth constrained equations. However, the convergence rate issue of the methods in [6, 7] is not investigated. Therefore, whether they have a linear convergence rate is an open problem. Can we design a spectral/conjugate gradient projection method with a linear convergence rate for CES(F, C)? In this paper, we answer this question positively for spectral gradient projection method. Note that, in [12], Dai and Liao proved a nice conclusion for the spectral gradient method. In fact, they established the R-linear convergence of the spectral gradient method for strongly convex quadratics of any number of dimensions, and they also proved the locally R-linear convergence for the general objective function. Obviously, the general minimization problem discussed in [12] is equivalent to the system of nonlinear equations under some mild conditions. However, for the system of constrained nonlinear equations, we shall establish the locally R-linear convergence of the spectral gradient method in this paper. Therefore, our result extends the conclusion in [12] in some sense.

In fact, in this paper, motivated by the projection methods in [13, 14] and the spectral gradient method in [6], we propose two spectral gradient projection methods for solving nonsmooth constrained equations, which can be viewed as combinations of the well-known spectral gradient method and the famous hyperplane projection method, and they possess a linear convergence rate under some error bound conditions. The remainder of this paper is organized as follows. In the next section, we describe the new methods and present their global convergence analysis. The linear convergence rates of the new methods are established in Section 3. Numerical results are reported in Section 4. Finally, some final remarks are included in Section 5.

2 Algorithm and convergence analysis

First, we denote $\|x\| = \sqrt{x^\top x}$ as the Euclidean-norm. Let C^* denote the solution set of CES(F, C). Throughout this paper, we assume that:

(A1) The solution set C^* is nonempty.

(A2) The mapping $F(\cdot)$ is monotone on C, *i.e.*,

$$\langle F(x) - F(y), x - y \rangle \geq 0, \quad \text{for all } x, y \in C.$$

(A3) The mapping $F(\cdot)$ is Lipschitz continuous on C, *i.e.*, there is a positive constant L such that

$$\|F(x) - F(y)\| \leq L\|x - y\|, \quad \text{for all } x, y \in C.$$

(A4) The mapping $F(\cdot)$ is strongly monotone on C, *i.e.*, there is a positive constant η such that

$$\langle F(x) - F(y), x - y \rangle \geq \eta \|x - y\|^2, \quad \text{for all } x, y \in C. \tag{2}$$

Obviously, (A4) implies (A2), and from (2) and the Cauchy-Schwartz inequality, we have

$$\|F(x) - F(y)\| \geq \eta \|x - y\|, \quad \text{for all } x, y \in C. \tag{3}$$

Then let $P_C(\cdot)$ denote the projection mapping from R^n onto the convex set C, *i.e.*,

$$P_C(x) = \operatorname{argmin}\{\|x - y\| \mid y \in C\},$$

which has the following nonexpansive property:

$$\|P_C(x) - P_C(y)\| \leq \|x - y\|, \quad \forall x, y \in R^n. \tag{4}$$

Now, we review the spectral gradient method for the unconstrained minimization problem:

$$\min f(x), \quad x \in R^n, \tag{5}$$

where $f : R^n \to R$ is smooth and its gradient is available. The spectral gradient for solving (5) is an iterative method of the form

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k),$$

where α_k is a step size defined by (see [9])

$$\alpha_k^{\mathrm{I}} = \frac{s_{k-1}^\top y_{k-1}}{y_{k-1}^\top y_{k-1}} \quad \text{or} \quad \alpha_k^{\mathrm{II}} = \frac{s_{k-1}^\top s_{k-1}}{s_{k-1}^\top y_{k-1}}, \tag{6}$$

in which $s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$. The step sizes (6) are called Barzilai-Borwein (BB) step sizes, and the corresponding gradient methods are spectral gradient methods. The spectral gradient with step size α_k^{II} has been extended to solve the constrained equations (1) by Yu *et al.* [6], however, as discussed in the Introduction, we do not know whether the method in [6] possesses the linear convergence rate. In the following, we will extend the spectral gradient with step size α_k^{I} and α_k^{II} to solve constrained equations (1) by some new type Armijo line searches, and we propose two spectral gradient projection methods, which are not only globally convergent, but also have a linear convergence rate.

The spectral gradient projection methods are stated as follows.

Algorithm 2.1

Step 0. Set an arbitrary initial point $x_0 \in C$, the parameters $0 < \rho < 1$, $0 < \sigma < r < 1$, $0 < \gamma < 2$, and $0 < \beta_{\min} < \beta_{\max}$. Set the initial step size $\beta_0 = 1$ and set $k := 0$.

Step 1. If $F(x_k) = 0$, then stop; otherwise, go to Step 2.

Step 2. Compute d_k by

$$d_k = \begin{cases} -F(x_k), & \text{if } k = 0, \\ -\theta_k F(x_k), & \text{if } k \geq 1, \end{cases} \tag{7}$$

where

$$\theta_k = \frac{s_{k-1}^\top y_{k-1}}{y_{k-1}^\top y_{k-1}}, \tag{8}$$

which is similar to α_k^{I} defined in (6), $y_{k-1} = F(x_k) - F(x_{k-1})$, but s_{k-1} is defined by

$$s_{k-1} = x_k - x_{k-1} + r y_{k-1},$$

which is different from the standard definition of s_{k-1}. Stop if $d_k = 0$; otherwise, go to Step 3.

Step 3. Find the trial point $z_k = x_k + \alpha_k d_k$, where $\alpha_k = \beta_k \rho^{m_k}$ with m_k being the smallest nonnegative integer m such that

$$-\langle F(x_k + \alpha_k d_k), d_k \rangle \geq \sigma \|F(x_k)\|^2. \tag{9}$$

Step 4. Compute

$$x_{k+1} = P_C[x_k - \gamma \xi_k F(z_k)], \tag{10}$$

where

$$\xi_k = \frac{\langle F(z_k), x_k - z_k \rangle}{\|F(z_k)\|^2}. \tag{11}$$

Choose an initial step size β_{k+1} such that $\beta_{k+1} \in [\beta_{\min}, \beta_{\max}]$. Set $k := k + 1$ and go to Step 1.

Algorithm 2.2

Step 0. Set an arbitrary initial point $x_0 \in C$, compute L, the Lipschitz constant of $F(\cdot)$, choose the parameters $0 < \rho < 1$, $0 < r < 1$, $0 < \sigma < r^2/(L + r)$, $0 < \gamma < 2$, and $0 < \beta_{\min} < \beta_{\max}$. Set the initial step size $\beta_0 = 1$ and set $k := 0$.

Step 1. If $F(x_k) = 0$, then stop; otherwise, go to Step 2.

Step 2. Compute d_k by

$$d_k = \begin{cases} -F(x_k), & \text{if } k = 0, \\ -\vartheta_k F(x_k), & \text{if } k \geq 1, \end{cases}$$

where

$$\vartheta_k = \frac{s_{k-1}^\top s_{k-1}}{s_{k-1}^\top y_{k-1}},$$

which is similar to α_k^{II} defined in (6), $s_{k-1} = x_k - x_{k-1}$, but y_{k-1} is defined by

$$y_{k-1} = F(x_k) - F(x_{k-1}) + r s_{k-1},$$

which is different from the standard definition of s_{k-1}. Stop if $d_k = 0$; otherwise, go to Step 3.

Step 3. Find the trial point $z_k = x_k + \alpha_k d_k$, where $\alpha_k = \beta_k \rho^{m_k}$ with m_k being the smallest nonnegative integer m such that

$$-\langle F(x_k + \alpha_k d_k), d_k \rangle \geq \sigma \|d_k\|^2. \tag{12}$$

Step 4. See Step 4 of Algorithm 2.1.

The discussions of the global convergence and linear convergence rate of Algorithm 2.2 are similar to those of Algorithm 2.1. Therefore, in the following, we discuss Algorithm 2.2 in detail, and we only give the corresponding results of Algorithm 2.2.

Remark 2.1 For Algorithm 2.1, by (3), we have

$$
\begin{aligned}
s_{k-1}^\top y_{k-1} &= \langle x_k - x_{k-1} + r y_{k-1}, y_{k-1} \rangle \\
&\leq \frac{1}{\eta} \|y_{k-1}\|^2 + r \|y_{k-1}\|^2 \\
&= \left(\frac{1}{\eta} + r \right) \|y_{k-1}\|^2.
\end{aligned}
$$

In addition, by the monotonicity of $F(\cdot)$, we also have

$$s_{k-1}^\top y_{k-1} \geq r \|y_{k-1}\|^2.$$

So we have from the above two inequalities and (7)

$$r \|F(x_k)\| \leq \|d_k\| \leq \left(\frac{1}{\eta} + r \right) \|F(x_k)\|, \tag{13}$$

from which we can get $\|F(x_k)\| = 0$ if $\|d_k\| = 0$, which means x_k is a solution of CES(F, C). Thus, Algorithm 2.1 can also terminate when $\|d_k\| = 0$. Similarly, for Algorithm 2.2, by the Lipschitz continuity and monotonicity of $F(\cdot)$, we can deduce that

$$\frac{\|F(x_k)\|}{L + r} \leq \|d_k\| \leq \frac{\|F(x_k)\|}{r}.$$

In what follows, we assume that $\|F(x_k)\| \neq 0$ and $\|d_k\| \neq 0$, for all k, *i.e.*, Algorithm 2.1 or Algorithm 2.2 generates an infinite sequence $\{x_k\}$.

Remark 2.2 In (10), we attach a relax factor $\gamma \in (0, 2)$ to $F(z_k)$ based on numerical experiences.

Remark 2.3 The line search (9) is different from that of [6, 7], which is well defined by the following lemma.

Lemma 2.1 *For all $k \geq 0$, there exists a nonnegative number m_k satisfying* (9).

Proof For the sake of contradiction, we suppose that there exists $k_0 \geq 0$ such that (9) is not satisfied for any nonnegative integer m, *i.e.*,

$$-\langle F(x_{k_0} + \beta_{k_0}\rho^m d_{k_0}), d_{k_0}\rangle < \sigma \|F(x_{k_0})\|^2, \quad \forall m \geq 1.$$

Letting $m \to \infty$ and using the continuity of $F(\cdot)$ yield

$$-\langle F(x_{k_0}), d_{k_0}\rangle \leq \sigma \|F(x_{k_0})\|^2. \tag{14}$$

On the other hand, by (7) and (13), we obtain

$$-\langle F(x_0), d_0\rangle = \|F(x_0)\|^2 > r\|F(x_0)\|^2$$

and

$$-\langle F(x_k), d_k\rangle = \theta_k \|F(x_k)\|^2 \geq r\|F(x_k)\|^2, \quad \forall k \geq 1,$$

which together with (14) means that $\sigma \geq r$, however, this contradicts the fact that $\sigma < r$. Therefore the assertion of Lemma 2.1 holds. This completes the proof. □

For the line search (12), we have a similar result, in the following lemma.

Lemma 2.2 *For all $k \geq 0$, there exists a nonnegative number m_k satisfying* (12).

Proof The lemma can be proved by contradiction as that of Lemma 2.1, and we omit the proof for concision. This completes the proof. □

The step length α_k and the norm of the function $F(x_k)$ satisfy the following property, which is an important result for proving the global convergence of Algorithm 2.1.

Lemma 2.3 *Suppose that $F(\cdot)$ is strongly monotone and let $\{x_k\}$ and $\{z_k\}$ be the sequences generated by Algorithm 2.1, then $\{x_k\}$ and $\{z_k\}$ are both bounded. Furthermore, we have*

$$\lim_{k \to \infty} \alpha_k \|F(x_k)\|^2 = 0. \tag{15}$$

Proof From (9), we have

$$\langle F(z_k), x_k - z_k\rangle \geq \sigma \alpha_k \|F(x_k)\|^2 > 0. \tag{16}$$

For any $x^* \in C^*$, from (4), we have

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 \\
= \|P_C[x_k - \gamma\xi_k F(z_k)] - x^*\|^2 \\
\leq \|x_k - \gamma\xi_k F(z_k) - x^*\|^2 \\
= \|x_k - x^*\|^2 - 2\gamma\xi_k\langle F(z_k), x_k - x^*\rangle + \gamma^2\xi_k^2\|F(z_k)\|^2.
\end{aligned} \tag{17}$$

By the monotonicity of the mapping $F(\cdot)$, we have

$$
\begin{aligned}
\langle F(z_k), x_k - x^* \rangle &= \langle F(z_k), x_k - z_k \rangle + \langle F(z_k), z_k - x^* \rangle \\
&\geq \langle F(z_k), x_k - z_k \rangle + \langle F(x^*), z_k - x^* \rangle \\
&= \langle F(z_k), x_k - z_k \rangle.
\end{aligned}
\tag{18}
$$

Substituting (16) and (18) into (17), we have

$$
\begin{aligned}
\| x_{k+1} - x^* \|^2 &\leq \| x_k - x^* \|^2 - 2\gamma \xi_k \langle F(z_k), x_k - z_k \rangle + \gamma^2 \xi_k^2 \| F(z_k) \|^2 \\
&= \| x_k - x^* \|^2 - \gamma(2-\gamma) \frac{\langle F(z_k), x_k - z_k \rangle^2}{\| F(z_k) \|^2} \\
&\leq \| x_k - x^* \|^2 - \gamma(2-\gamma) \frac{\sigma^2 \alpha_k^2 \| F(x_k) \|^4}{\| F(z_k) \|^2},
\end{aligned}
\tag{19}
$$

which together with $\gamma \in (0, 2)$ indicates that, for all k,

$$
\| x_{k+1} - x^* \| \leq \| x_k - x^* \|,
\tag{20}
$$

which shows that the sequence $\{x_k\}$ is bounded. By (13), $\{d_k\}$ is bounded and so is $\{z_k\}$. Then, by the continuity of $F(\cdot)$, there exists a constant $M > 0$ such that $\| F(z_k) \| \leq M$, for all k. Therefore it follows from (19) that

$$
\gamma(2-\gamma) \frac{\sigma^2}{M^2} \sum_{k=0}^{\infty} \alpha_k^2 \| F(x_k) \|^4 \leq \sum_{k=0}^{\infty} \left(\| x_k - x^* \|^2 - \| x_{k+1} - x^* \|^2 \right) < \infty,
$$

which implies that the assertion (15) holds. The proof is completed. □

Lemma 2.4 *Suppose that $F(\cdot)$ is monotone and Lipschitz continuous and let $\{x_k\}$ and $\{z_k\}$ be the sequences generated by Algorithm 2.2, then $\{x_k\}$ and $\{z_k\}$ are both bounded. Furthermore, we have*

$$
\lim_{k \to \infty} \alpha_k \| d_k \|^2 = 0.
$$

Proof The conclusion is a little different from (15), which results from the difference of the right hands of the line searches (9) and (12). In fact, this conclusion can be proved as that of Lemma 2.3, and we also omit it for concision. This completes the proof. □

Now, we establish the global convergence theorems for Algorithm 2.1 and Algorithm 2.2.

Theorem 2.1 *Suppose that the conditions in Lemma 2.3 hold. Then the sequence $\{x_k\}$ generated by Algorithm 2.1 globally converges to a solution of* CES(F, C).

Proof We consider the following two possible cases.

Case 1: $\liminf_{k\to\infty}\|F(x_k)\| = 0$, which together with the continuity of $F(\cdot)$ implies that the sequence $\{x_k\}$ has some accumulation point $\bar{x}$ such that $F(\bar{x}) = 0$. From (20), $\{\|x_k - \bar{x}\|\}$ converges, and since $\bar{x}$ is an accumulation point of $\{x_k\}$, $\{x_k\}$ must converge to $\bar{x}$.

Case 2: $\liminf_{k\to\infty}\|F(x_k)\| > 0$. Then by (15), it follows that $\lim_{k\to\infty}\alpha_k = 0$. Therefore, from the line search (9), for sufficiently large k, we have

$$-\langle F(x_k + \beta_k \rho^{m_k-1} d_k), d_k\rangle < \sigma\|F(x_k)\|^2. \tag{21}$$

Since $\{x_k\}$, $\{d_k\}$ are both bounded, we can choose a sequence and letting $k \to \infty$ in (21), we can obtain

$$-\langle F(\bar{x}), \bar{d}\rangle \leq \sigma\|F(\bar{x})\|^2, \tag{22}$$

where $\bar{x}, \bar{d}$ are limit points of corresponding subsequences. On the other hand, by (13), we obtain

$$-\langle F(x_k), d_k\rangle = \theta_k\|F(x_k)\|^2 \geq r\|F(x_k)\|^2, \quad \forall k \geq 1.$$

Letting $k \to \infty$ in the above inequality, we obtain

$$-\langle F(\bar{x}), \bar{d}\rangle \geq r\|F(\bar{x})\|^2. \tag{23}$$

Thus, by (22) and (23), we get $r \leq \sigma$, and this contradicts the fact that $r > \sigma$. Therefore $\liminf_{k\to\infty}\|F(x_k)\| > 0$ does not hold. This completes the proof. $\qquad\square$

For Algorithm 2.2, we also have the following global convergence.

Theorem 2.2 *Suppose that the conditions in Lemma* 2.4 *hold. Then the sequence* $\{x_k\}$ *generated by Algorithm* 2.2 *globally converges to a solution of* $\mathrm{CES}(F, C)$.

Proof Following a process similar to the proof for Theorem 2.1, we can get the desired conclusion. This completes the proof. $\qquad\square$

3 Convergence rate

By Theorem 2.1 and Theorem 2.2, we know that the sequence $\{x_k\}$ generated by Algorithm 2.1 or Algorithm 2.2 converges to a solution of $\mathrm{CES}(F, C)$. In what follows, we always assume that $x_k \to x^*$ as $k \to \infty$, where $x^* \in C^*$. To establish the local convergence rate of the sequence generated by Algorithm 2.1 or Algorithm 2.2, we need the following assumption.

Assumption 3.1 For $x^* \in C^*$, there exist three positive constants δ, c, and L such that

$$c\,\mathrm{dist}(x, C^*) \leq \|F(x)\|, \quad \forall x \in N(x^*, \delta) \tag{24}$$

and

$$\|F(x) - F(y)\| \leq L\|x - y\|, \quad \forall x, y \in N(x^*, \delta), \tag{25}$$

where dist(x, C^*) denotes the distance from x to the solution set C^*, and

$$N(x^*, \delta) = \{x \in R^n \mid \|x - x^*\| \le \delta\}.$$

Obviously, (A3) in Section 2 implies (25). Here, we set the constant c so that

$$0 < \frac{\gamma(2 - \gamma)\sigma\alpha c^2\eta^2}{L^2(\beta_{\max}L(1 + r\eta) + \eta)^2} < 1. \tag{26}$$

Now, we analyze the convergence rate of the sequence $\{x_k\}$ generated by Algorithm 2.1 or Algorithm 2.2 under the conditions (24) and (25).

Lemma 3.1 *If* (A4) *and the conditions in Assumption* 3.1 *hold, then the sequence* $\{\alpha_k\}$ *generated by the line search* (9) *has a positive bound from below.*

Proof We only need to prove that for sufficiently large k, α_k has a positive bound from below. If $\alpha_k \le \beta_k$, then by the construction of α_k, we have

$$-\langle F(x_k + \beta_k\alpha_k\rho^{-1}d_k), d_k\rangle < \sigma\|F(x_k)\|^2.$$

In addition, by (7), we have

$$-\langle F(x_k), d_k\rangle = \theta_k\|F(x_k)\|^2 \ge r\|F(x_k)\|^2.$$

Then, by the above two inequalities, we can obtain

$$\langle F(x_k + \beta_k\alpha_k\rho^{-1}d_k) - F(x_k), d_k\rangle \ge (r - \sigma)\|F(x_k)\|^2. \tag{27}$$

On the other hand, from (13) and (25), we have

$$\langle F(x_k + \beta_k\alpha_k\rho^{-1}d_k) - F(x_k), d_k\rangle \le \frac{L\beta_k\alpha_k}{\rho}\|d_k\|^2 \le \frac{L\beta_k\alpha_k(1 + r\eta)^2}{\rho\eta^2}\|F(x_k)\|^2. \tag{28}$$

By (27) and (28), for k sufficiently large we obtain

$$\alpha_k \ge \frac{\rho(r - \sigma)\eta^2}{L\beta_k(1 + r\eta)^2} \ge \frac{\rho(r - \sigma)\eta^2}{L\beta_{\max}(1 + r\eta)^2}.$$

Therefore, there is a positive constant α, such that

$$\alpha_k \ge \alpha, \tag{29}$$

for all k. The proof is completed.   □

Lemma 3.2 *If* (A2), (A3), *and the conditions in Assumption* 3.1 *hold, then the sequence* $\{\alpha_k\}$ *generated by the line search* (12) *has a positive bound from below.*

Proof The proof is similar to that of Lemma 3.1, and we omit it for concision. This completes the proof.   □

Theorem 3.1 *In addition to the assumptions in Theorem* 2.1, *if conditions* (24) *and* (25) *hold, then the sequence* $\{\mathrm{dist}(x_k, C^*)\}$ *generated by Algorithm* 2.1 *converges locally to* 0 *at the Q-linear rate, hence the sequence* $\{x_k\}$ *converges locally to* x^* *at the R-linear rate.*

Proof Let $v_k \in C^*$ be the closest solution to x_k. That is, $\|x_k - v_k\| = \mathrm{dist}(x_k, C^*)$. By (19), we have

$$\|x_{k+1} - v_k\|^2 \leq \|x_k - v_k\|^2 - \gamma(2 - \gamma)\frac{\langle F(z_k), x_k - z_k\rangle^2}{\|F(z_k)\|^2}. \tag{30}$$

For sufficiently large k, it follows from (13) and (25) that

$$
\begin{aligned}
\left\|F(z_k)\right\| &= \left\|F(z_k) - F(v_k)\right\| \\
&\leq L\|z_k - v_k\| \\
&\leq L\big(\|x_k - y_k\| + \|x_k - v_k\|\big) \\
&\leq L\big(\beta_{\max}\|d_k\| + \|x_k - v_k\|\big) \\
&\leq L\left(\frac{\beta_{\max}(1 + r\eta)\|F(x_k)\|}{\eta} + \|x_k - v_k\|\right) \\
&= L\left(\frac{\beta_{\max}(1 + r\eta)\|F(x_k) - F(v_k)\|}{\eta} + \|x_k - v_k\|\right) \\
&\leq L\left(\frac{\beta_{\max}L(1 + r\eta)}{\eta} + 1\right)\|x_k - v_k\| \\
&= L\left(\frac{\beta_{\max}L(1 + r\eta)}{\eta} + 1\right)\mathrm{dist}(x_k, C^*).
\end{aligned}
$$

Thus, from (9), (24), and (29), for sufficiently large k, we have

$$\langle F(z_k), x_k - z_k\rangle \geq \sigma\alpha_k\left\|F(x_k)\right\|^2 \geq \sigma\alpha\left\|F(x_k)\right\|^2 \geq \sigma\alpha c^2\,\mathrm{dist}^2(x_k, C^*).$$

Substituting the above two inequalities into (30) and from (26), we have

$$\mathrm{dist}^2(x_{k+1}, C^*) \leq \|x_{k+1} - v_k\|^2 \leq \left(1 - \frac{\gamma(2 - \gamma)\sigma\alpha c^2\eta^2}{L^2(\beta_{\max}L(1 + r\eta) + \eta)^2}\right)\mathrm{dist}^2(x_k, C^*),$$

which implies that the sequence $\{\mathrm{dist}(x_k, C^*)\}$ converges locally to 0 at the *Q*-linear rate. Therefore, the sequence $\{x_k\}$ converges locally to x^* at the *R*-linear rate. The proof is completed. □

Theorem 3.2 *In addition to the assumptions in Theorem* 2.2, *if conditions* (24) *and* (25) *hold, then the sequence* $\{\mathrm{dist}(x_k, C^*)\}$ *generated by Algorithm* 2.1 *converges locally to* 0 *at the Q-linear rate, hence the sequence* $\{x_k\}$ *converges locally to* x^* *at an R-linear rate.*

Proof The proof is similar to that of Theorem 3.1, and we also omit it for concision. This completes the proof. □

4 Numerical results

In this section, we test Algorithm 2.1 and Algorithm 2.2, and compare them with the spectral gradient projection method in [15]. We give the following three simple problems to test the efficiency of the three methods.

Problem 1 The mapping $F(\cdot)$ is taken as $F(x) = (f_1(x), f_2(x), \ldots, f_n(x))^\top$, where

$$f_i(x) = e^{x_i} - 1, \quad \text{for } i = 1, 2, \ldots, n$$

and $C = R_+^n$. Obviously, this problem has a unique solution $x^* = (0, 0, \ldots, 0)^\top$.

Problem 2 The mapping $F(\cdot)$ is taken as $F(x) = (f_1(x), f_2(x), \ldots, f_n(x))^\top$, where

$$f_i(x) = x_i - \sin|x_i - 1|, \quad \text{for } i = 1, 2, \ldots, n$$

and $C = \{x \in R_+^n \mid \sum_{i=1}^n x_i \leq n, x_i \geq 0, i = 1, 2, \ldots, n\}$. Obviously, Problem 2 is nonsmooth at $x = (1, 1, \ldots, 1)^\top$.

Problem 3 The problem is adapted from [13]. The mapping $F(\cdot)$ is taken as $F(x) = D(x) + Mx$, where $D(x)$ and Mx are the nonlinear part and linear part of $F(x)$, respectively. Here, the components of $D(x)$ is defined by $D_j(x) = a_j \arctan(x_j)$, where a_j is a random variable in $(0, 100)$, and the matrix $M = A^\top A + B$, where A is an $n \times n$ matrix whose entries are randomly generated in the interval $(-1, 1)$ and a skew-symmetric matrix B is generated in the same way. In addition, $C = R_+^n$.

The codes are written in Mablab 7.0 and run on a personal computer with 2.0 GHz CPU processor. The parameters used in Algorithm 2.1 and Algorithm 2.2 are set as $\rho = 0.6$, $r = 10^{-3}$, $\sigma = 10^{-4}$, and $\gamma = 1.8$ for Problem 1 and $\gamma = 1$ for Problems 2 and 3. The initial step size in Step 2 of Algorithm 2.1 or Algorithm 2.2 is set to be $\beta_k = 1$. We stop the iteration if the iteration number exceeds 1,000 or the inequality $\|F(x_k)\| \leq 10^{-5}$ is satisfied. The method in [15] (denoted by CGD) is implemented with the following parameters: $\rho = 0.1$, $r = 0.01$, $\sigma = 10^{-4}$, and $\xi = 1$.

For Problems 1 and 2, the initial point is set as $x_0 = \text{ones}(n, 1)$, and for Problem 3, the initial point is set as $x_0 = \text{rand}(n, 1)$. Tables 1-3 give the numerical results by Algorithm 2.1, Algorithm 2.2, and CGD with different dimensions, where Iter. denotes the iteration number, Fn denotes the number of function evaluations, and CPU denotes the CPU time in seconds when the algorithms terminate.

The numerical results given in Tables 1-3 show that: (1) the three methods can solve all the tested problems successfully; (2) for the two easy Problems 1 and 2, Algorithm 2.2 performs a little better than Algorithm 2.1 for the CPU time, and both methods perform better than CGD for the three criteria: Iter., Fn, and CPU; (3) for the difficult Problem 3, Algorithm 2.1 performs best among the three methods, and both Algorithm 2.1 and Algorithm 2.2 perform much better than CGD, especially for the CPU time. From the above analysis, we conclude that Algorithm 2.1 and Algorithm 2.2 are better than CGD.

Table 1 Numerical results with different dimensions of Problem 1

Dimension	Method	Iter.	Fn	CPU
1,000	Algorithm 2.1	1	5	0.02
	Algorithm 2.2	1	5	0.02
	CGD	11	51	0.03
5,000	Algorithm 2.1	1	5	0.03
	Algorithm 2.2	1	5	0.03
	CGD	12	56	0.16
50,000	Algorithm 2.1	1	5	0.14
	Algorithm 2.2	1	5	0.11
	CGD	13	60	1.33
100,000	Algorithm 2.1	1	5	0.25
	Algorithm 2.2	1	5	0.25
	CGD	13	60	2.86

Table 2 Numerical results with different dimensions of Problem 2

Dimension	Method	Iter.	Fn	CPU
1,000	Algorithm 2.1	10	59	0.05
	Algorithm 2.2	8	52	0.05
	CGD	12	59	0.06
5,000	Algorithm 2.1	10	59	0.17
	Algorithm 2.2	8	52	0.16
	CGD	12	59	0.22
50,000	Algorithm 2.1	11	64	1.58
	Algorithm 2.2	10	63	1.47
	CGD	12	59	1.55
100,000	Algorithm 2.1	12	69	3.50
	Algorithm 2.2	10	63	3.19
	CGD	13	69	3.88

Table 3 Numerical results with different dimensions of Problem 3

Dimension	Method	Iter.	Fn	CPU
100	Algorithm 2.1	125	479	0.80
	Algorithm 2.2	141	743	1.39
	CGD	356	5,422	10.00
500	Algorithm 2.1	215	873	19.48
	Algorithm 2.2	239	1,475	32.81
	CGD	270	2,700	63.13

5 Conclusions

Two spectral gradient projection methods for solving constrained equations have been developed, which are not only derivative-free, but also completely matrix-free. Consequently, they can be applied to solve large-scale nonsmooth constrained equations. We established the global convergence without the requirement of differentiability of the equations, and presented the linear convergence rate under standard conditions. We also reported some numerical results to show the efficiency of the proposed methods.

Author details
¹School of Mathematics and Statistics, Zhejiang University of Finance and Economics, Xueyuan Street, Hangzhou, 310018, P.R. China. ²School of Economics and Management, Tongji University, Siping Street, Shanghai, 200092, P.R. China.

References
1. Dirkse, SP, Ferris, MC: MCPLIB: a collection of nonlinear mixed complementarity problems. Optim. Methods Softw. **5**, 319-345 (1995)
2. Wood, AJ, Wollenberg, BF: Power Generation, Operation, and Control. Wiley, New York (1996)
3. Meintjes, K, Morgan, AP: A methodology for solving chemical equilibrium systems. Appl. Math. Comput. **22**, 333-361 (1987)
4. Qi, LQ, Tong, XJ, Li, DH: An active-set projected trust region algorithm for box constrained nonsmooth equations. J. Optim. Theory Appl. **120**, 601-625 (2004)
5. Ortega, JM, Rheinboldt, WC: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York (1970)
6. Yu, ZS, Lin, J, Sun, J, Xiao, YH, Liu, LY, Li, ZH: Spectral gradient projection method for monotone nonlinear equations with convex constraints. Appl. Numer. Math. **59**, 2416-2423 (2009)
7. Liu, SY, Huang, YY, Jiao, HW: Sufficient descent conjugate gradient methods for solving convex constrained nonlinear monotone equations. Abstr. Appl. Anal. **2014**, Article ID 305643 (2014)
8. Sun, M, Liu, J: Three derivative-free projection methods for large-scale nonlinear equations with convex constraints. J. Appl. Math. Comput. (2014). doi:10.1007/s12190-014-0774-5
9. Barzilai, J, Borwein, JM: Two point stepsize gradient methods. IMA J. Numer. Anal. **8**, 141-148 (1988)
10. Birgin, EG, Martinez, JM, Raydan, M: Spectral projected gradient methods: review and perspectives. J. Stat. Softw. **60**, 1-21 (2014)
11. Fletcher, R, Reeves, C: Function minimization by conjugate gradients. Comput. J. **7**, 149-154 (1964)
12. Dai, YH, Liao, LZ: *R*-Linear convergence of the Barzilai and Borwein gradient method. IMA J. Numer. Anal. **22**, 1-10 (2002)
13. Wang, CW, Wang, YJ, Xu, CL: A projection method for a system of nonlinear monotone equations with convex constraints. Math. Methods Oper. Res. **66**, 33-46 (2007)
14. Zheng, L: A new projection algorithm for solving a system of nonlinear equations with convex constraints. Bull. Korean Math. Soc. **50**, 823-832 (2013)
15. Xiao, YH, Zhu, H: A conjugate gradient method to solve convex constrained monotone equations with applications in compressive sensing. J. Math. Anal. Appl. **405**, 310-319 (2013)