SCIENCE CHINA
Mathematics

• ARTICLES •

May 2016 Vol. 59 No. 5: 999–1014
doi: 10.1007/s11425-015-5116-z

A smoothing trust region filter algorithm for nonsmooth least squares problems

CHEN XiaoJun[1,*], DU ShouQiang[2] & ZHOU Yang[1]

[1]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China;
[2]College of Mathematics, Qingdao University, Qingdao 266071, China
Email: maxjchen@polyu.edu.hk, qddsq1@163.com, andres.zhou@connect.polyu.hk

Received March 13, 2015; accepted September 18, 2015; published online December 30, 2015

Abstract We propose a smoothing trust region filter algorithm for nonsmooth nonconvex least squares problems. We present convergence theorems of the proposed algorithm to a Clarke stationary point or a global minimizer of the objective function under certain conditions. Preliminary numerical experiments show the efficiency of the proposed algorithm for finding zeros of a system of polynomial equations with high degrees on the sphere and solving differential variational inequalities.

Keywords smoothing approximation, trust region method, filter technique

MSC(2010) 90C26, 65K05

Citation: Chen X J, Du S Q, Zhou Y. A smoothing trust region filter algorithm for nonsmooth least squares problems. Sci China Math, 2016, 59: 999–1014, doi: 10.1007/s11425-015-5116-z

1 Introduction

This paper considers the nonsmooth nonconvex least squares problem

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|r(x)\|^2, \tag{1.1}$$

where $r:\mathbb{R}^n\to\mathbb{R}^m$ is a locally Lipschitz continuous function but not necessarily differentiable and $\|\cdot\|$ is the Euclidean norm. This problem has many important applications in engineering and economics, which includes constrained smooth nonlinear equations and nonsmooth equations as special cases.

Denote the objective function of (1.1) by f, i.e., $f(x)=\frac{1}{2}\|r(x)\|^2$. In general, $f:\mathbb{R}^n\to\mathbb{R}_+$ is nonconvex and nonsmooth. In the presence of nonsmoothness and noncovexity, most optimization methods only guarantee convergence to a Clarke stationary point of the objective function f [7, 8, 15, 21, 29].

The trust region method [16, 25, 28] is a classic and widely used numerical method for optimization problems and filter techniques are proposed in [20, 22] as a globalization strategy. In this paper, we propose a smoothing trust region filter (STRF) algorithm to find a global minimizer of (1.1) when r is nonsmooth and there is x^* such that $r(x^*)=0$. This algorithm combines trust region methods [16, 25], filter techniques [20, 22] and smoothing approximations [5, 8, 10]. Using a smoothing function $\tilde{r}$ of r, we can define a smoothing function $\tilde{f}$ of f and construct a good quadratic approximation of f in a certain region at each iteration. In the proposed STRF algorithm, the trust region method [16, 25] is used to find a low value of smoothing function $\tilde{f}$, while the filter technique [20, 22] is used to build a filter by using

*Corresponding author

the original nonsmooth function r. A new point is generated based on the new value of the smoothing function and the new filter at the current step. To guarantee the convergence of the STRF algorithm to a Clarke stationary point or a global minimizer, a new scheme is introduced to update the smoothing parameter by using both the nonsmooth function f and the gradient of the smoothing function $\nabla\tilde{f}$. Note that the proposed STRF is different from the smoothing trust region method in [10] and the filter method in [22]. The smoothing trust region method [10] can reduce the objective values and guarantee convergence to a Clarke stationary point, but has no convergence results to a global minimizer. The filter method [20,22] is a technique for finding a global minimizer of a twice continuously differentiable function under certain conditions, but application to a nonsmooth nonconvex minimization problem has not been investigated. The proposed STRF algorithm is a novel combination of these optimization techniques for nonsmooth and nonconvex least squares problems.

To verify the efficiency of the proposal STRF algorithm for finding global minimizers of least squares problems, we compare the STRF algorithm with several codes in Matlab on the following two challenging problems.

Spherical t_ϵ-designs. A set X_N of N points on the unit sphere is called a spherical t-design if the average value of any polynomial of degree at most t over X_N is equal to the average value of the polynomial over the sphere. A spherical t-design provides an equal positive weight integration rule which is the exact integral for any polynomial of degree at most t. Spherical t-designs have many important applications in geophysics and bioengineering, and provide many challenging problems in computational mathematics [2, 3, 9, 12, 27]. It is shown in [12] that finding a spherical t-design can be reformulated as a system of polynomial equations. In this paper, we define a spherical t_ϵ-design which provides an integration rule with a set X_N^ϵ of N points on the unit sphere and positive weights satisfying

$$(1-\epsilon)^2 \leqslant \frac{\min \text{weight}}{\max \text{weight}} \leqslant 1.$$

The integration rule also gives the exact integral for any polynomial of degree at most t. When $\epsilon = 0$, the spherical t_ϵ-design reduces to the spherical t-design. Due to the flexibility of choice for the weights, the number of points in the integration rule can be less for making the exact integral for any polynomial of degree at most t. We show that finding a spherical t_ϵ-design can be reformulated as a system of polynomial equations with box constraints. Using the projection operator, the system can be written as a nonsmooth nonconvex least squares problem (1.1) with zero residual.

Differential variational inequalities (DVI). The DVI is a powerful mathematical paradigm for the increasing number of engineering and economics problems that involve dynamics and equilibrium problems [1, 11, 13, 14, 23, 26]. The time-stepping method is widely used for solving the DVI, at each step of which, a standard variational inequality problem (VIP) has to be solved efficiently. It is known that a standard VIP can be reformulated as a system of nonsmooth equations [17, 18], and thus a nonsmooth nonconvex least-squares problem (1.1) with zero residual. We use a time-stepping method with the STRF algorithm to solve several DVI. Moreover, we use the least-norm time-stepping method with the STRF algorithm to solve a nonlinear complementarity system where the complementarity problem has a unbounded solution set and a least-norm solution was required for the stability of the system. Preliminary numerical experiments show that the STRF algorithm with regularization is robust in finding a global and stable minimizer of (1.1) at each time in the dynamic system.

This paper is organized as follows. In Section 2, we introduce the STRF algorithm and show that the STRF algorithm converges to a Clarke stationary point or a global minimizer of the objective function in (1.1) under certain conditions. In Section 3, we present numerical results of the STRF algorithm for finding spherical t_ϵ-designs which is equivalent to finding zeros of a system of polynomial equations with high degrees on the sphere, and solving differential variational inequalities. Comparing with several algorithms and codes including `fmincon`, `lsqnonlin`, `fsolve` in Matlab, the STRF is more efficient for solving nonsmooth nonconvex least squares problems. Finally, Section 4 concludes the paper.

Throughout the paper, $\|\cdot\|$ represents the Euclidean norm, $\mathbb{R}_+ = \{\alpha \in \mathbb{R} \mid \alpha \geqslant 0\}$ and $\mathbb{R}_{++} = \{\alpha \in \mathbb{R} \mid \alpha > 0\}$.

2 A smoothing trust region filter (STRF) algorithm

We use the ideas in [22] to construct the filter, which partition $r(x)$ into p sets $\{r_i(x)\}_{i\in I_j}, j=1,\dots,p$, with $\{1,\dots,m\}=I_1\cup\cdots\cup I_p$. For readability and simplicity, we explain how to construct the filter with a disjoint partition. Let

$$r(x)=\begin{pmatrix} r_{I_1}(x)\\ \vdots \\ r_{I_p}(x)\end{pmatrix},\quad \theta_j(x)=\|r_{I_j}(x)\|,\quad j=1,\dots,p,\quad \theta(x)=\begin{pmatrix} \theta_1(x)\\ \vdots \\ \theta_p(x)\end{pmatrix},$$

where $r_{I_j}:\mathbb{R}^n\to\mathbb{R}^{m_j}$ and $\sum_{j=1}^p m_j=m$.

Obviously, a vector x is a solution of (1.1) with $f(x)=0$ if and only if $\theta(x)=0$.

At the k-th iteration, the filter $\mathcal{F}$ is a subset of $\{\theta(x_0),\theta(x_1),\dots,\theta(x_k)\}$. A new trial point x_k^+ is acceptable for the filter $\mathcal{F}$ if and only if for any $\theta(x_\ell)\in\mathcal{F}$ there is $j\in\{1,\dots,p\}$ such that

$$\theta_j(x_k^+)<\theta_j(x_\ell)-\gamma\min\{\|\theta(x_k^+)\|,\|\theta(x_\ell)\|\},\tag{2.1}$$

where $\gamma\in(0,1/\sqrt{p})$ is a positive constant.

We remove $\theta(x_\ell)$ from the filter $\mathcal{F}$ if

$$\exists\,\theta(x_j)\in\mathcal{F},\quad \text{such that}\quad \theta(x_\ell)-\gamma\|\theta(x_\ell)\|e\geqslant\theta(x_j),\tag{2.2}$$

where $e=(1,\dots,1)^{\mathrm{T}}$.

We say that a vector x dominates a vector y whenever $\theta(x)<\theta(y)$. The inequality in (2.2) implies that x_j dominates x_ℓ. From the construction of the filter, if x_ℓ is removed from the filter at the k-th iteration, x_ℓ will not be added back to the filter after the k-th iteration.

To overcome the nonsmoothness of r, we use a smoothing function $\tilde r(\cdot,\mu)$ of r.

Definition 2.1. Let $r:\mathbb{R}^n\to\mathbb{R}^m$ be a locally Lipschitz continuous function. We call $\tilde r:\mathbb{R}^n\times\mathbb{R}_{++}\to\mathbb{R}^m$ a smoothing function of r, if for any fixed $\mu\in\mathbb{R}_{++}$, $\tilde r(\cdot,\mu)$ is continuously differentiable in $\mathbb{R}^n$ and for any fixed $\hat x\in\mathbb{R}^n$,

$$\lim_{x\to\hat x,\mu\downarrow 0}\tilde r(x,\mu)=r(\hat x).$$

Using a smoothing function $\tilde r$, we can define a smoothing function $\tilde f$ of f by

$$\tilde f(x,\mu)=\frac12\|\tilde r(x,\mu)\|^2.$$

By Definition 2.1, for any fixed $\mu>0$, $\tilde f(\cdot,\mu)$ is continuously differentiable in $\mathbb{R}^n$ and for any fixed $\hat x\in\mathbb{R}^n$,

$$\lim_{x\to\hat x,\mu\downarrow 0}\tilde f(x,\mu)=f(\hat x).$$

In this paper, we assume that the smoothing function $\tilde r$ satisfies the following condition:

$$|\tilde r_i(x,\mu)-r_i(x)|\leqslant\kappa(\mu),\quad i=1,\dots,m,\tag{2.3}$$

where $\kappa:\mathbb{R}_{++}\to\mathbb{R}_+$ satisfies $\kappa(\mu_1)\leqslant\kappa(\mu_2)$ for $\mu_1\leqslant\mu_2$, and $\kappa(\mu)\to0$ as $\mu\to0$. Many smoothing functions satisfy condition (2.3) (see [8]). In Section 3, we give examples of the smoothing function $\tilde r$ satisfying (2.3).

Using the smoothing function $\tilde r$, we can define the gradient of the objective function $\tilde f$ as follows:

$$g(x,\mu)=\nabla_x\tilde f(x,\mu)=J(x,\mu)^{\mathrm{T}}\tilde r(x,\mu),\quad\text{where}\quad J(x,\mu)=\nabla_x\tilde r(x,\mu).$$

The smoothing trust region method computes a trial point $x_k^+=x_k+d_k$ for some step d_k by a quadratic approximation function

$$q_k(d)=\tilde f(x_k,\mu_k)+g(x_k,\mu_k)^{\mathrm{T}}d+\frac12 d^{\mathrm{T}}B_kd\tag{2.4}$$

of $\tilde{f}(x,\mu)$ in a trust region $\{x_k + d \mid \|d\| \leqslant \Delta_k\}$, where Δ_k is the radius of the trust region and $B_k = J(x_k,\mu_k)^{\mathrm{T}} J(x_k,\mu_k) + \sqrt{\mu_k} I$.

Since B_k is a symmetric positive definite matrix, q_k is strongly convex and d_k in Step 1 is uniquely defined. The term $\sqrt{\mu_k} I$ in B_k plays a regularization role and ensures the nonsingularity of B_k, which yields the strong convexity of q_k. Using the smoothing function $\tilde{r}$, we can easily compute the matrix B_k and the function q_k, and find the unique solution d_k in Step 1 of the STRF algorithm. Hence, the STRF algorithm is well-defined. When both smoothing and regularization techniques are used in an algorithm, it is recommended to let the smoothing parameter go to zero faster than the regularization parameter for good numerical performance [11].

Smoothing trust region filter (STRF) algorithm.

Step 0. Initialization. Given constants $0 < \bar{\Delta} < \infty$, $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_1 < 1 < \gamma_2$, $0 < \sigma < 1$, $0 < \gamma < 1/\sqrt{p}$, $0 < \beta < \infty$, an initial vector $x_0 \in \mathbb{R}^n$, the radius of a trust region $\Delta_0 \in (0, \bar{\Delta})$, the smoothing parameter $\mu_0 > 0$, and filter $\mathcal{F} = \{\theta(x_0)\}$.

Step 1. Define a trial point. Compute $d_k = \operatorname{argmin}_{\|d\| \leqslant \Delta_k} q_k(d)$ and set $x_k^+ = x_k + d_k$.

Step 2. Evaluate the reduction at the trial step. If $d_k = 0$, set $x_{k+1} = x_k$, $\Delta_{k+1} = \Delta_k$, and go to Step 5. Otherwise, compute

$$\rho_k = \frac{\tilde{f}(x_k,\mu_k) - \tilde{f}(x_k^+,\mu_k)}{q_k(0) - q_k(d_k)}.$$

Step 3. Update the trust-region radius. Set

$$\Delta_{k+1} = \begin{cases} \min\{\gamma_2 \Delta_k, \bar{\Delta}\}, & \text{if } \rho_k \geqslant \eta_2, \quad \|d_k\| = \Delta_k, \\ \gamma_1 \Delta_k, & \text{if } \rho_k \leqslant \eta_1, \\ \Delta_k, & \text{otherwise.} \end{cases}$$

Step 4. Test to accept the trial step.

- x_k^+ is acceptable for the current filter by (2.1): Set $x_{k+1} = x_k^+$ and add $\theta(x_k^+)$ to the filter if $\rho_k < \eta_1$. Update $\mathcal{F}$ by (2.2).
- x_k^+ is not acceptable for the current filter: If $\rho_k \geqslant \eta_1$, set $x_{k+1} = x_k^+$. Otherwise, set $x_{k+1} = x_k$.

Step 5. Update the smoothing parameter. If $\min\{f(x_k), \|\nabla_x \tilde{f}(x_k,\mu_k)\|\} \leqslant \beta\mu_k$, set $\mu_{k+1} = \sigma\mu_k$. Otherwise, set $\mu_{k+1} = \mu_k$. Go to Step 1.

The STRF algorithm is constructed based on the idea of the trust region filter algorithm in [22]. However, the two algorithms have essential differences. The algorithm in [22] is applied to smooth function r and has a decrease of the objective function $f(x_{k+1}) < f(x_k)$ when $x_{k+1} = x_k^+$ and $\theta(x_k^+)$ is not included in the filter. This is a key property for the convergence of the algorithm in [22]. However, the STRF algorithm is applied to nonsmooth function r and has a decrease of the smoothing function $\tilde{f}(x_{k+1},\mu_k) < \tilde{f}(x_k,\mu_k)$ when $x_{k+1} = x_k^+$ and $\theta(x_k^+)$ is not included in the filter. A decrease of the objective function is not guaranteed. To prove the convergence of $\{f(x_k)\}$ generated by the STRF algorithm, an innovative proof is needed.

Now we investigate the convergence of the STRF algorithm. We first consider the case that infinitely many values are added to the filter in the STRF algorithm.

Theorem 2.2. *Assume that $\tilde{r}$ satisfies* (2.3). *If infinitely many values of $\theta(x_k)$ are added to the filter by the STRF algorithm, then*

$$\lim_{k\to\infty} \|\theta(x_k)\| = \lim_{k\to\infty} f(x_k) = 0.$$

Proof. Let $\theta_k = \theta(x_k)$, $\theta_k^+ = \theta(x_k^+)$ and $\theta_{j,k} = \theta_j(x_k), j = 1,\ldots,p$.

Let $\{k_i\}$ index the subsequence of iterations at which $\theta_{k_i} = \theta_{k_i-1}^+$ is added to the filter. Assume on contradiction that there exists a subsequence $\{k_\nu\} \subseteq \{k_i\}$ such that $\|\theta_{k_\nu}\| \geqslant \epsilon$ for some $\epsilon > 0$. By Step 4 and the construction of a filter (2.1) and (2.2), $\{\theta_{k_\nu}\}$ is bounded. Hence, there exists a further subsequence $\{k_\tau\} \subseteq \{k_\nu\}$ such that

$$\lim_{\tau\to\infty} \theta_{k_\tau} = \bar{\theta}. \tag{2.5}$$

Since $\{k_\tau\} \subseteq \{k_\nu\} \subseteq \{k_i\}$ and $\|\theta_{k_\nu}\| \geqslant \epsilon$ for all ν, we know that for all τ, $\min\{\|\theta_{k_{\tau-1}}\|, \|\theta_{k_\tau}\|\} \geqslant \epsilon$ and θ_{k_τ} is acceptable for the filter. Hence for each τ, there exists a $j \in \{1, \dots, p\}$ such that

$$\theta_{j,k_\tau} - \theta_{j,k_{\tau-1}} < -\gamma \min\{\|\theta_{k_{\tau-1}}\|, \|\theta_{k_\tau}\|\} \leqslant -\gamma\epsilon. \tag{2.6}$$

However, by (2.5), we get $\theta_{j,k_\tau} - \theta_{j,k_{\tau-1}} \to 0$, as $\tau \to \infty$. This is a contradiction. Hence, we obtain

$$\lim_{i\to\infty} \|\theta_{k_i}\| = 0, \tag{2.7}$$

which implies

$$\lim_{i\to\infty} f(x_{k_i}) = 0. \tag{2.8}$$

Now, we prove the convergence of the whole sequence $\{\|\theta_k\|\}$ to zero. From $f(x_k) = \frac{1}{2}\|\theta_k\|^2$, it is to prove that the sequence $\{f(x_k)\}$ converges to zero.

We consider any $\ell \notin \{k_i\}$ and let $k_{i(\ell)}$ be the last iteration before ℓ such that $\theta_{k_{i(\ell)}}$ was added to the filter. By the definition of $\{k_{i(\ell)}\}$ and (2.7), we have

$$\lim_{\ell\to\infty} f(x_{k_{i(\ell)}}) = 0. \tag{2.9}$$

Moreover, we have $\mu_{k_{i(\ell)}} \to 0$ as $\ell \to \infty$ by Step 5 of the STRF algorithm. Hence, using $\mu_{k+1} \leqslant \mu_k$, we obtain $\mu_k \to 0$ as $k \to \infty$.

From the condition on the smoothing function (2.3), we derive

$$\begin{aligned}
|\tilde{f}(x_{k_{i(\ell)}}, \mu_{k_{i(\ell)}}) - f(x_{k_{i(\ell)}})| &= \frac{1}{2}|\|\tilde{r}(x_{k_{i(\ell)}}, \mu_{k_{i(\ell)}})\|^2 - \|r(x_{k_{i(\ell)}})\|^2| \\
&= \frac{1}{2}\bigg|\sum_{j=1}^{m}(\tilde{r}_j^2(x_{k_{i(\ell)}}, \mu_{k_{i(\ell)}}) - r_j^2(x_{k_{i(\ell)}}))\bigg| \\
&\leqslant \frac{1}{2}\sum_{j=1}^{m}|\tilde{r}_j(x_{k_{i(\ell)}}, \mu_{k_{i(\ell)}}) - r_j(x_{k_{i(\ell)}})| \cdot |\tilde{r}_j(x_{k_{i(\ell)}}, \mu_{k_{i(\ell)}}) + r_j(x_{k_{i(\ell)}})| \\
&\leqslant \frac{1}{2}\sum_{j=1}^{m}\kappa(\mu_{k_{i(\ell)}})|\tilde{r}_j(x_{k_{i(\ell)}}, \mu_{k_{i(\ell)}}) + r_j(x_{k_{i(\ell)}})| \\
&\leqslant \frac{1}{2}\sum_{j=1}^{m}\kappa(\mu_{k_{i(\ell)}})(\kappa(\mu_{k_{i(\ell)}}) + 2|r_j(x_{k_{i(\ell)}})|) \\
&\leqslant \frac{m}{2}\kappa^2(\mu_{k_{i(\ell)}}) + \kappa(\mu_{k_{i(\ell)}})\|r(x_{k_{i(\ell)}})\|_1 \\
&\leqslant \frac{m}{2}\kappa^2(\mu_{k_{i(\ell)}}) + \kappa(\mu_{k_{i(\ell)}})\sqrt{m}\|r(x_{k_{i(\ell)}})\|_2 \\
&\leqslant \frac{m}{2}\kappa^2(\mu_{k_{i(\ell)}}) + \kappa(\mu_{k_{i(\ell)}})\sqrt{2mf(x_{k_{i(\ell)}})}.
\end{aligned} \tag{2.10}$$

Hence from (2.9) and $\mu_k \to 0$, we obtain

$$\lim_{\ell\to\infty} \tilde{f}(x_{k_{i(\ell)}}, \mu_{k_{i(\ell)}}) = 0. \tag{2.11}$$

By Steps 2 and 4 of the STRF algorithm, if $\theta(x_{k_{i(\ell)}+1})$ is not included in the filter, then we have

$$\tilde{f}(x_{k_{i(\ell)}}, \mu_{k_{i(\ell)}}) - \tilde{f}(x_{k_{i(\ell)}+1}, \mu_{k_{i(\ell)}}) \geqslant 0,$$

which, together with (2.11), implies

$$\lim_{\ell\to\infty} \tilde{f}(x_{k_{i(\ell)}+1}, \mu_{k_{i(\ell)}}) = 0. \tag{2.12}$$

Using the similar argument in (2.10), we can show

$$|\tilde{f}(x_{k_{i(\ell)}+1},\mu_{k_{i(\ell)}})-f(x_{k_{i(\ell)}+1})|\leqslant\frac{m}{2}\kappa^2(\mu_{k_{i(\ell)}})+\kappa(\mu_{k_{i(\ell)}})\sqrt{2m\tilde{f}(x_{k_{i(\ell)}+1},\mu_{k_{i(\ell)}})},\tag{2.13}$$

which, together with (2.12) and

$$\lim_{\ell\to\infty}|\tilde{f}(x_{k_{i(\ell)}+1},\mu_{k_{i(\ell)}})-f(x_{k_{i(\ell)}+1})|\leqslant\lim_{\ell\to\infty}\left(\frac{m}{2}\kappa^2(\mu_{k_{i(\ell)}})+\kappa(\mu_{k_{i(\ell)}})\sqrt{2m\tilde{f}(x_{k_{i(\ell)}+1},\mu_{k_{i(\ell)}})}\right)=0,$$

we obtain

$$\lim_{\ell\to\infty}f(x_{k_{i(\ell)}+1})=0.\tag{2.14}$$

Since $\ell\notin\{k_i\}$ is arbitrarily chosen, from (2.8), (2.9) and (2.14), we can get

$$\lim_{i\to\infty}f(x_{k_i+1})=0.$$

Letting $\ell\notin\{k_i\}\cup\{k_i+1\}$, using the same argument above, we can show

$$\lim_{i\to\infty}f(x_{k_i+2})=0.$$

Hence, repeating the argument, we can derive

$$\lim_{k\to\infty}f(x_k)=0\quad\text{and}\quad\lim_{k\to\infty}\|\theta(x_k)\|=0.\tag{2.15}$$

We complete the proof. □

Now, we study the convergence of the STRF algorithm without assuming that infinitely many values of $\theta(x_k)$ are added to the filter.

We say that f has bounded level sets, if for any $\alpha\geqslant 0$, the level set $\{x\,|\,f(x)\leqslant\alpha\}$ is bounded.

If f has bounded level sets and Condition (2.3) holds, then the smoothing function $\tilde{f}$ has bounded level sets for any fixed $\mu>0$. In fact, using the argument in (2.10) and (2.13) with condition (2.3) and $\mu\leqslant\mu_0$, for any $\alpha>0$, the following holds,

$$\begin{aligned}\{x\,|\,\tilde{f}(x,\mu)\leqslant\alpha\}&\subseteq\left\{x\,|\,f(x)\leqslant\alpha+\frac{m}{2}\kappa^2(\mu)+\kappa(\mu)\sqrt{2m\alpha}\right\}\\&\subseteq\left\{x\,|\,f(x)\leqslant\alpha+\frac{m}{2}\kappa^2(\mu_0)+\kappa(\mu_0)\sqrt{2m\alpha}\right\}.\end{aligned}\tag{2.16}$$

Lemma 2.3. *Suppose that f has bounded level sets and $\nabla\tilde{f}(\cdot,\mu)$ is Lipschitz continuous for any fixed $\mu>0$, then the sequence $\{\mu_k\}$ generated by the STRF algorithm satisfies*

$$\lim_{k\to\infty}\mu_k=0.\tag{2.17}$$

Proof. Let K contain all iterations at which $\mu_{k+1}=\sigma\mu_k$, namely,

$$K=\{k\,|\,\min\{f(x_k),\|\nabla_x\tilde{f}(x_k,\mu_k)\|\}\leqslant\beta\mu_k\}.\tag{2.18}$$

If K is an infinite set, then $\lim_{k\to\infty}\mu_k=0$. Moreover, from Theorem 2.2, if infinitely many values of θ_k are added to the filter, then $\lim_{k\to\infty}\mu_k=0$. Hence, in the following, we will prove that K is an infinite set in the case when only finitely many values of θ_k are added to the filter.

Assume by contradiction that K is finite and only finitely values of θ_k are added to the filter. Then there exists a nonnegative integer $\hat{k}$, such that for all nonnegative integers j, $\theta(x^+_{\hat{k}+j})$ are not added to the filter and $\mu_{\hat{k}+j}=\mu_{\hat{k}}$. This means

$$\tilde{f}(x_{\hat{k}+j},\mu_{\hat{k}})-\tilde{f}(x_{\hat{k}+j+1},\mu_{\hat{k}})\geqslant 0,\quad\text{for}\quad j\geqslant 0\tag{2.19}$$

and

$$\min\{f(x_{\hat{k}+j}), \|\nabla_x \tilde{f}(x_{\hat{k}+j}, \mu_{\hat{k}})\|\} > \beta\mu_{\hat{k}}, \quad \text{for } j \geqslant 0. \tag{2.20}$$

By (2.16) and the assumption that f has bounded level sets, we know that $\tilde{f}(\cdot, \mu_{\hat{k}})$ has bounded level sets. Hence, in such case, the STRF algorithm reduces to [25, Algorithm 4.1] for solving the smooth optimization problem with the objective $\tilde{f}(\cdot, \mu_{\hat{k}})$. From the assumption of Lemma 2.3, $\nabla\tilde{f}(\cdot, \mu_{\hat{k}})$ is Lipschitz continuous, and thus B_k is bounded. Note that d_k is the exact solution of the minimization problem in Step 1 of the STRF algorithm. All conditions of [25, Theorem 4.6] hold. Similar to the proof of [25, Theorem 4.6], we can show

$$\lim_{j\to\infty} \|\nabla_x \tilde{f}(x_{\hat{k}+j}, \mu_{\hat{k}})\| = 0. \tag{2.21}$$

This contradicts (2.20). Hence (2.17) holds. □

Since r is locally Lipschitz continuous, f is locally Lipschitz continuous and almost everywhere differentiable. The Clarke subdifferential of f at $x \in \mathbb{R}^n$ can be defined by

$$\partial f(x) = \text{co}\{v \mid \nabla f(z) \to v,\ f \text{ is differentiable at } z,\ z \to x\},$$

where "co" denotes the convex hull. A vector x is called a Clarke stationary point of f if $0 \in \partial f(x)$. To show that any accumulation point of $\{x_k\}$ generated by the STRF algorithm is a Clarke stationary point of f, we need functions r_i, $i = 1, \ldots, m$ to be regular and their smoothing functions $\tilde{r}_i$ to satisfy the gradient consistency.

Definition 2.4 (See [15]). A function $h : \mathbb{R}^n \to \mathbb{R}$ is said to be regular at $x \in \mathbb{R}^n$ if for all $v \in \mathbb{R}^n$, the directional derivative exists and

$$h(x; v) = \lim_{t\downarrow 0} \frac{h(x + tv) - h(x)}{t} = \limsup_{y\to x, t\downarrow 0} \frac{h(y + tv) - h(y)}{t}.$$

If h is regular at all $x \in \mathbb{R}^n$, h is said to be regular.

Definition 2.5 (See [8]). A smoothing function $\tilde{h}$ of $h : \mathbb{R}^n \to \mathbb{R}$ is said to satisfy the gradient consistency if

$$\text{co}\{v \mid \nabla_x \tilde{h}(x_k, \mu_k) \to v, \text{ for } x_k \to x,\ \mu_k \downarrow 0\} = \partial h(x), \quad \forall\, x \in \mathbb{R}^n.$$

Theorem 2.6. *Assume that $\tilde{r}_i$ satisfies (2.3) and the gradient consistency, for $i = 1, \ldots, m$, f has bounded level sets and $\nabla\tilde{f}(\cdot, \mu)$ is Lipschitz continuous for any fixed $\mu > 0$. Then the sequences $\{x_k\}$ and $\{\mu_k\}$ generated by the STRF algorithm satisfy*

$$\liminf_{k\to\infty} \|\nabla_x \tilde{f}(x_k, \mu_k)\| = 0. \tag{2.22}$$

In addition, if r_i is regular for $i = 1, \ldots, m$, then any accumulation point of $\{x_k\}$ is a Clarke stationary point of f.

Proof. We consider two cases.

Case I. $\liminf_{k\to\infty} f(x_k) = 0$.

In this case, we have

$$\liminf_{k\to\infty} \|r(x_k)\|^2 = \liminf_{k\to\infty} \sum_{j=1}^{m} r_j^2(x_k) = 0.$$

From (2.3) and Lemma 2.3, we get $\mu_k \to 0$, and

$$0 \leqslant \liminf_{k\to\infty} |\tilde{r}_j(x_k, \mu_k)| \leqslant \liminf_{k\to\infty} (|r_j(x_k)| + \kappa(\mu_k)) = 0, \quad \text{for} \quad j = 1, \ldots, m.$$

Since r_i is Lipschitz continuous, the Clarke subdifferential ∂r_i is bounded. Hence from the gradient consistency of r_i, we can get $\|\nabla_x \tilde{r}_i(x_k, \mu_k)\|$ is bounded and

$$\liminf_{k\to\infty} \|\nabla_x \tilde{f}(x_k, \mu_k)\| = \liminf_{k\to\infty} \|\nabla_x \tilde{r}(x_k, \mu_k)^{\mathrm{T}} \tilde{r}(x_k, \mu_k)\| = 0.$$

Case II. $\liminf_{k\to\infty} f(x_k) > 0$.

In this case, there exist $\bar{k}$ and $\epsilon > 0$, such that for $k > \bar{k}$, $f(x_k) \geqslant \epsilon$. By Lemma 2.3, $\mu_k \to 0$. Thus from $\min\{f(x_k), \|\nabla_x \tilde{f}(x_k,\mu_k)\|\} \leqslant \beta\mu_k$, we have

$$\liminf_{k\to\infty} \|\nabla_x \tilde{f}(x_k,\mu_k)\| = 0.$$

Hence we complete the proof for (2.22).

If r_i is regular, then by [4, Proposition 2.1], $\tilde{r}_i^2$ is a smoothing function of r_i^2 and satisfies the gradient consistency. Since $f(x) = \frac{1}{2}\sum_{i=1}^m r_i^2(x)$ is a convex composite function of $r_i^2(x)$, $\tilde{f}(x,\mu) = \frac{1}{2}\sum_{i=1}^m \tilde{r}_i^2(x,\mu)$ is a smoothing function of f and satisfies the gradient consistency, which means

$$\begin{aligned}&\mathrm{co}\{v \mid \nabla f(x_k) \to v, f \text{ is differentiable at } x_k, x_k \to x\}\\ &= \mathrm{co}\{v \mid \nabla_x \tilde{f}(x_k,\mu_k) \to v,\ x_k \to x, \mu_k \downarrow 0\}.\end{aligned}$$

Hence, from (2.22), any accumulation point of $\{x_k\}$ is a Clarke stationary point of f. □

Example 1. To explain the smoothing approximation and gradient consistency, we consider the following example. Let

$$r(x) = Mx + \max(0,x) + q, \quad \text{where} \quad M = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad q = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

At $\bar{x} = (0,0)^{\mathrm{T}}$, $r(x)$ and $f(x)$ are not differentiable. Since r_1 and r_2 are convex, by [15, Proposition 2.3.6], they are regular. By [15, Corollary 3], the Clarke gradient of $f(x)$ at $\bar{x}$ is

$$\begin{aligned}\partial f(\bar{x}) &= \frac{1}{2}(\partial r_1^2(x) + \partial r_2^2(x))\\ &= \mathrm{co}\{v \mid \nabla r_1(x) r_1(x) + \nabla r_2(x) r_2(x) \to v, x_1 \neq 0, x_2 \neq 0, x \to \bar{x}\}\\ &= \left\{\begin{pmatrix} \alpha_1 & 1 \\ 1 & \alpha_2 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix},\ \alpha_1,\alpha_2 \in [1,2]\right\}.\end{aligned}$$

Since $0 \in \partial f(\bar{x})$, $\bar{x}$ is a stationary point.

We use the smoothing function

$$\varphi(t,\mu) = \begin{cases} \max(0,t), & \text{if } |t| > \dfrac{\mu}{2}, \\ \dfrac{t^2}{2\mu} + \dfrac{t}{2} + \dfrac{\mu}{8}, & \text{otherwise} \end{cases}$$

for $\max(0,t)$, and

$$\tilde{r}(x) = Mx + \Phi(x,\mu) + q$$

for $r(x)$, where $\Phi(x,\mu) = (\varphi(x_1,\mu), \varphi(x_2,\mu))^{\mathrm{T}}$. It is easy to see that $0 \leqslant \varphi'(t,\mu) \leqslant 1$. In particular, $\varphi'(-\frac{\mu}{2},\mu) = 0$ and $\varphi'(\frac{\mu}{2},\mu) = 1$. Hence, we find that f satisfies the gradient consistency, i.e.,

$$\mathrm{co}\{v \mid \nabla_x \tilde{f}(x,\mu) = \nabla \tilde{r}(x,\mu)^{\mathrm{T}} \tilde{r}(x,\mu) \to v,\ x \to \bar{x}, \mu \downarrow 0\} = \partial f(\bar{x}).$$

Moreover, we have

$$|\tilde{r}_i(x,\mu) - r_i(x)| = |\varphi(x_i,\mu) - \max(0,x_i)| \leqslant \frac{\mu}{8}, \quad i = 1,2.$$

Hence the smoothing function $\tilde{r}$ satisfies (2.3).

More examples and results on the smoothing approximation, regularity and gradient consistency can be found in [5,6,8].

3 Numerical results

In this section, we report numerical results of the STRF algorithm for solving nonsmooth nonconvex least squares problems (1.1) with zero residual arising from spherical t_ϵ-designs and differential variational inequalities which are described in Section 1. Both problems have many stationary points at which the residual is not zero. We show that all conditions used in last section for convergence of the STRF algorithm hold for these two problems. Numerical results show that the STRF algorithm is efficient and robust for finding global minimizers of the problems.

We implemented the STRF algorithm in MATLAB 2012b on a Lenovo Thinkcenter PC equipped with Intel Core i7-3770 3.4G Hz CPU, 8 GB RAM running Windows 7. The values of parameters in the STRF algorithm are chosen as follows: $\Delta_0 = 10^{-1}$, $\overline{\Delta} = 10^{12}$, $\eta_1 = 0.2$, $\eta_2 = 0.8$, $\gamma_1 = 0.8$, $\gamma_2 = 1.25$, $\sigma = 0.95$, $\mu_0 = 0.5$, $\gamma = 0.01, \beta = 10$. We terminate the STRF algorithm when $\min\{f(x_k), \|\nabla \tilde{f}(x_k, \mu_k)\|\} \leqslant 10^{-10}$.

Example 2. Spherical t_ϵ-design.

Let $\mathbb{P}_t$ be the linear space of restriction of polynomials of degree $\leqslant t$ in 3 variables to the unit sphere $\mathbb{S}^2 = \{\boldsymbol{x} \in \mathbb{R}^3 \,|\, x_1^2 + x_2^2 + x_3^2 = 1\}$.

A spherical t_ϵ-design with $0 \leqslant \epsilon < 1$ on $\mathbb{S}^2$ is a set of points $X_N^\epsilon := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{S}^2$ such that the cubature rule with weights $w = (w_1, \ldots, w_N)^{\mathrm{T}}$ satisfying

$$\frac{4\pi}{N}(1-\epsilon) \leqslant w_i \leqslant \frac{4\pi}{N}(1-\epsilon)^{-1}, \quad i = 1, \ldots, N, \tag{3.1}$$

is exact for all spherical polynomials of degree at most t, i.e.,

$$\sum_{i=1}^{N} w_i p(\boldsymbol{x}_i) = \int_{\mathbb{S}^2} p(\boldsymbol{x}) d\omega(\boldsymbol{x}), \quad \forall\, p \in \mathbb{P}_t. \tag{3.2}$$

When $\epsilon = 0$, the spherical t_ϵ-design reduces to the spherical t-design that is an equally weighted ($w_i = \frac{4\pi}{N}$) cubature rule [12, 27]. Finding spherical t-designs provides many open and challenging problems which attract considerable attention from pure and applied mathematicians.

Now we reformulate the problem finding a spherical t_ϵ-design, that is to find X_N^ϵ and w such that (3.1)–(3.2) hold, as a nonlinear least squares problem (1.1).

Let $\{Y_{\ell,k}, k = 1, \ldots, 2\ell + 1, \ell = 0, \ldots, t\}$ be a set of L_2-orthonormal basis functions of $\mathbb{P}_t$, where $Y_{\ell,k}$ is a spherical harmonic of degree ℓ. The dimension of $\mathbb{P}_t$ is $d_t = (t + 1)^2$. Define $\boldsymbol{Y}(X_N) \in \mathbb{R}^{N\times d_t}$ with elements

$$\boldsymbol{Y}_{i,\ell^2+k}(X_N) = Y_{\ell,k}(\boldsymbol{x}_i), \quad i = 1, \ldots, N, \quad k = 1, \ldots, 2\ell + 1, \quad \ell = 0, \ldots, t.$$

Let $a = \frac{4\pi(1-\epsilon)}{N}\boldsymbol{e}$ and $b = \frac{4\pi(1-\epsilon)^{-1}}{N}\boldsymbol{e}$, where $\boldsymbol{e} = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^N$.

Proposition 3.1. *The set $X_N^\epsilon := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{S}^2$ is a spherical t_ϵ-design if and only if*

$$\boldsymbol{Y}(X_N^\epsilon)^{\mathrm{T}} w - \sqrt{4\pi}\boldsymbol{e}_0 = 0 \quad \textit{and} \quad w - \mathrm{mid}(a, w, b) = 0, \tag{3.3}$$

where $\boldsymbol{e}_0 = (1, 0, \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^{(t+1)^2}$ and

$$(\mathrm{mid}(a, w, b))_i = \mathrm{mid}(a_i, w_i, b_i) = \begin{cases} a_i, & w_i < a_i, \\ w_i, & a_i \leqslant w_i \leqslant b_i, \\ b_i, & w_i > b_i, \end{cases} \quad i = 1, \ldots, N.$$

Proof. It is easy to see that $w - \mathrm{mid}(a, w, b) = 0$ if and only if $a \leqslant w \leqslant b$. Hence, we only need to prove the equivalence between (3.2) and the first equality in (3.3).

Assume (3.2) holds. Since $Y_{0,1}(\boldsymbol{x})$ is a spherical harmonic of degree 0, $\int_{\mathbb{S}^2} Y_{0,1}(\boldsymbol{x})^2 d\omega(\boldsymbol{x}) = 1$ and $\int_{\mathbb{S}^2} d\omega(\boldsymbol{x}) = 4\pi$, we have $Y_{0,1}(\boldsymbol{x}) \equiv 1/\sqrt{4\pi}$ and

$$\sum_{i=1}^{N} w_i Y_{0,1}(\boldsymbol{x}_i) = \int_{\mathbb{S}^2} Y_{0,1}(\boldsymbol{x}) d\omega(\boldsymbol{x}) = Y_{0,1}(\boldsymbol{x}) \int_{\mathbb{S}^2} d\omega(\boldsymbol{x}) = \sqrt{4\pi}.$$

Moreover, from that $\{Y_{\ell,k}, k=1,\dots,2\ell+1, \ell=0,\dots,t\}$ is a set of L_2-orthonormal basis functions of $\mathbb{P}_t$, we obtain

$$\sum_{i=1}^{N} w_i Y_{\ell,k}(\boldsymbol{x}_i) = \int_{\mathbb{S}^2} Y_{\ell,k}(\boldsymbol{x}) d\omega(\boldsymbol{x}) = \sqrt{4\pi} \int_{\mathbb{S}^2} Y_{\ell,k}(\boldsymbol{x}) Y_{0,1}(\boldsymbol{x}) d\omega(\boldsymbol{x}) = 0$$

for $k = 1,\dots,2\ell+1$, and $1 \leqslant \ell \leqslant t$. This implies the first equality in (3.3).

Now we assume that the first equality in (3.3) holds. Then we obtain that

$$\sum_{i=1}^{N} w_i Y_{0,1}(\boldsymbol{x}_i) = \sqrt{4\pi} = \int_{\mathbb{S}^2} Y_{0,1}(\boldsymbol{x}) d\omega(\boldsymbol{x}),$$

and

$$\sum_{i=1}^{N} w_i Y_{\ell,k}(\boldsymbol{x}_i) = 0 = \int_{\mathbb{S}^2} Y_{\ell,k}(\boldsymbol{x}) d\omega(\boldsymbol{x}), \quad \text{for } \ell = 1,\dots,t, \quad k = 1,\dots,2\ell+1.$$

Moreover, for any $p \in \mathbb{P}_t$, there exists a unique group of numbers $p_{\ell,k}$ satisfying

$$p = \sum_{\ell=0}^{t} \sum_{k=1}^{2l+1} p_{\ell,k} Y_{\ell,k}.$$

Hence, (3.2) is derived as the following:

$$\begin{aligned} \int_{\mathbb{S}^2} p(\boldsymbol{x}) d\omega(\boldsymbol{x}) &= \sum_{\ell=0}^{t} \sum_{k=1}^{2\ell+1} p_{\ell,k} \int_{\mathbb{S}^2} Y_{\ell,k}(\boldsymbol{x}) d\omega(\boldsymbol{x}) \\ &= \sum_{\ell=0}^{t} \sum_{k=1}^{2\ell+1} p_{\ell,k} \sum_{i=1}^{N} w_i Y_{\ell,k}(\boldsymbol{x}_i) \\ &= \sum_{i=1}^{N} w_i \sum_{\ell=0}^{t} \sum_{k=1}^{2\ell+1} p_{\ell,k} Y_{\ell,k}(\boldsymbol{x}_i) = \sum_{i=1}^{N} w_i p(\boldsymbol{x}_i). \end{aligned}$$ □

We represent the points $\boldsymbol{x}_i \in \mathbb{S}^2$ using spherical coordinates with angles θ_i, φ_i. Since (3.3) is rotationally invariant with respect to X_N^{ϵ}, we fix $\boldsymbol{x}_1$ at the north pole and $\boldsymbol{x}_2$ on the zero meridian as [12],

$$\boldsymbol{x}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \boldsymbol{x}_2 = \begin{pmatrix} \sin(\theta_2) \\ 0 \\ \cos(\theta_2) \end{pmatrix}, \quad \boldsymbol{x}_i = \begin{pmatrix} \sin(\theta_i)\cos(\varphi_i) \\ \sin(\theta_i)\sin(\varphi_i) \\ \cos(\theta_i) \end{pmatrix}, \quad i = 3,\dots,N.$$

Let $x_\theta = (\theta_2,\dots,\theta_N)^{\mathrm{T}}, x_\varphi = (\varphi_3,\dots,\varphi_N)^{\mathrm{T}}, x = (x_\theta^{\mathrm{T}}, x_\varphi^{\mathrm{T}}, w^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{3N-3}$ and

$$r(x) = \begin{pmatrix} r_{I_1}(x) \\ r_{I_2}(w) \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}^{\mathrm{T}}(x_\theta, x_\varphi) w - \sqrt{4\pi}\boldsymbol{e}_0 \\ w - \mathrm{mid}(a, w, b) \end{pmatrix}. \tag{3.4}$$

A solution of $r(x) = 0$ defines a spherical t_ϵ-design. To use the STRF algorithm, we need a smoothing function $\tilde{r}$ of r and the Jacobian of $\tilde{r}$. Since $r_{I_1} : \mathbb{R}^{3N-3} \to \mathbb{R}^{(t+1)^2}$ is continuously differentiable, we only define a smoothing function of $r_{I_2} : \mathbb{R}^N \to \mathbb{R}^N$ as follows:

$$(\tilde{r}_{I_2}(w,\mu))_i = \begin{cases} w_i - a_i, \quad w_i < a_i - \mu, \\ w_i - \dfrac{1}{4\mu}(w_i - a_i)^2 - \dfrac{1}{2}(w_i - a_i) - \dfrac{\mu}{4} - a_i, \quad a_i - \mu < w_i < a_i + \mu, \\ 0, \quad a_i + \mu \leqslant w_i \leqslant b_i - \mu, \\ w_i + \dfrac{1}{4\mu}(w_i - b_i)^2 - \dfrac{1}{2}(w_i - b_i) + \dfrac{\mu}{4} - b_i, \quad b_i - \mu < w_i < b_i + \mu, \\ w_i - b_i, \quad w_i > b_i + \mu. \end{cases}$$

It is easy to verify that

$$|\tilde{r}_i(x,\mu) - r_i(x)| \leqslant \frac{\mu}{4}, \quad i = 1, \dots, N.$$

Hence the smoothing function $\tilde{r}$ satisfies condition (2.3). Moreover, the function r_{I_2} is Lipschitz continuous and regular, which implies the smoothing function $\tilde{r}_{I_2}$ satisfies the gradient consistency. Since $\|r_{I_2}\|^2$ is continuously differentiable and has bounded level sets, the objective function $f(x) = \frac{1}{2}\|r(x)\|^2$ is continuously differentiable and has bounded level sets. Hence all conditions on r and f in the last section hold. It is worth noting that r is not differentiable, we cannot have a simple and explicit derivative of f. Using the smoothing function $\tilde{r}$, we have

$$\nabla \tilde{f}(x,\mu) = \nabla_x \tilde{r}(x,\mu)^{\mathrm{T}} \tilde{r}(x).$$

Thus we can easily construct the quadratic function (2.4) and compute the minimizer d_k.

The function f is nonconvex with many stationary points. It is hard to find a global minimizer of f by using most existing methods. We use this example to test the STRF algorithm and compare it with the smoothing trust region (STR) algorithm [10] and `fmincon`, `lsqnonlin`, `fsolve` codes in Matlab. To guarantee the fairness of the comparison, we use same paramete in the STR algorithm and the STRF algorithm, and same initial points for all algorithms and codes.

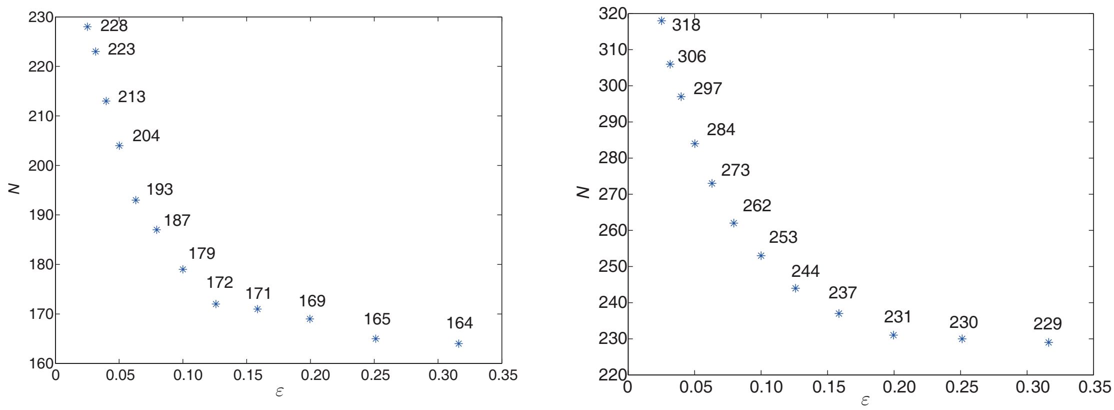
First we generate N points distributed evenly on the whole sphere. The points are generated by "The recursive zonal equal area (EQ) sphere partitioning toolbox" proposed by Leopardi, which could be downloaded from http://sourceforge.net/projects/eqsp/. Next, we add a small random perturbation on the points to create more initial point sets with the same cardinalities. All the perturbation obeys a uniform distribution with expectation as 0.1. We choose initial weights $w_i^0 = \frac{4\pi}{N}$, $i = 1, \dots, N$.

In Table 1, we show numerical results for finding spherical $t_{0.1}$-designs with different t and N points on the sphere. The final value of the objective function $f(x)$ and the CPU time (CPUtime) are reported in the table. Compared with other methods, the STRF algorithm can find a good numerical global minimizer efficiently.

Note that there is no theoretical result which proves the existence of a spherical t-design with $N \leqslant (t+1)^2$ points for arbitrary t. In [9], using a computational algorithm based on interval arithmetic, Chen et al. proved the existence of a spherical t-design with $N = (t+1)^2$ points on the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ for $t = 1, 2, \dots, 100$. In [27], Sloan and Womersley conjectured the existence of a spherical t-design with $N = \lceil (t+1)^2/2 \rceil + 1$ points on the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ for some small t, where $\lceil \cdot \rceil$ denotes rounding up to the nearest integer. We believe that with the flexibility of choice for the weights, the number of points for a spherical t_ϵ-design can be less than $(t+1)^2/2$. To see the minimum number of points for a spherical t_ϵ-design, we solve the least squares problem with $r(x)$ defined in (3.4) for $\lceil (t+1)^2/3 \rceil + 1 \leqslant N \leqslant \lceil (t+2)^2/2 \rceil + 1$ with different ϵ and t. Figure 1 shows the minimal values N such that $f(x_k) \leqslant 10^{-10}$ with $t = 21, 25$ and $\epsilon = 10^{-\alpha}$, $\alpha = 0.5 + i \times 0.1, i = 0, 1, \dots, 11$. From Figure 1, we see that the bigger value of ϵ we choose, the smaller number of points for a spherical t_ϵ-design we need.

Table 1 Values of $r(x)$ (CPU time) for spherical t_ϵ-design with $\epsilon = 0.1$

t, N	fmincon	lsqnonlin	fsolve	STR	STRF
4, 12	1.41e−07 (1.390)	1.91e−05 (0.281)	3.28e−15 (0.185)	2.64e−03 (1.940)	7.78e−11 (0.038)
9, 45	8.54e−07 (10.100)	2.00e−04 (2.290)	3.96e−06 (1.890)	6.81e−03 (6.260)	9.39e−11 (0.350)
12, 80	1.16e−06 (52.500)	3.19e−04 (13.800)	3.95e−06 (15.800)	1.01e−02 (12.100)	7.12e−11 (0.888)
14, 105	1.61e−06 (107.000)	4.99e−03 (66.100)	3.68e−06 (46.800)	1.06e−03 (22.300)	9.68e−11 (2.070)
19, 190	7.66e−06 (492.000)	1.1e−02 (189.000)	2.78e−07 (207.000)	3.06e−04 (70.500)	9.79e−11 (12.100)
21, 235	1.91e−06 (856.000)	1.18e−04 (193.000)	3.98e−08 (310.000)	1.89e−03 (115.000)	9.35e−11 (98.000)
24, 305	2.30e−05 (2064.000)	6.13e−04 (382.000)	8.66e−07 (689.000)	1.56e−03 (220.000)	9.05e−11 (36.000)



(a) $t=21$ (b) $t=25$

Figure 1 Possible minimal number N of points for spherical t_ϵ-designs

Example 3. Differential variational inequalities (DVI).

Given $a \in \mathbb{R}^\ell \cup \{-\infty\}^\ell$ and $b \in \mathbb{R}^\ell \cup \{+\infty\}^\ell$, $A \in \mathbb{R}^{\nu\times\nu}$, $B \in \mathbb{R}^{\nu\times\ell}$, $c(t) \in \mathbb{R}^\nu$ and a continuously differentiable function $F: \mathbb{R}^\ell \times \mathbb{R}^\nu \to \mathbb{R}^\ell$, we consider the following DVI:

$$\begin{cases} \dot{x}(t) = Ax(t) + By(t) + c(t), & t \in [0,T], \\ y(t) \in \mathrm{SOL}(x(t)), & t \in [0,T], \\ x(0) = x^0 \in \mathbb{R}^\nu, \end{cases} \tag{3.5}$$

where $\mathrm{SOL}(x(t))$ is the solution set of the variational inequality, which contains $y(t) \in [a,b]$ such that

$$(v - y(t))^{\mathrm{T}} F(y(t), x(t)) \geqslant 0, \quad \text{for all } v \in [a,\, b].$$

It is easy to verify that $y(t) \in \mathrm{SOL}(x(t))$ if and only if

$$r(y(t)) = y(t) - \mathrm{mid}(a, y(t) - F(y(t), x(t)), b) = 0. \tag{3.6}$$

For a fixed t and $x(t)$, $f(y(t)) = \frac{1}{2}\|r(y(t))\|^2$ is a nonsmooth nonconvex function. We can use the smoothing function of the "mid" function in Example 2 to define a smoothing function $\tilde{r}(y(t),\mu)$ of $r(y(t))$, and a smoothing function $\tilde{f}(y(t),\mu)$ of $f(y(t))$.

The time-stepping method [26] with the STRF algorithm for solving the DVI begins with the division of the time interval $[0,T]$ into N_h subintervals

$$0 = t_{h,0} < t_{h,1} < \cdots < t_{h,N_h} = T,$$

where $t_{h,i+1} - t_{h,i} = h = T/N_h$, $i = 0,\ldots,N_h - 1$. Starting from a given vector $x^{h,0} = x^0 \in \mathbb{R}^\nu$, we compute $y^{h,0} \in \mathrm{SOL}(x^{h,0})$ by the STRF algorithm and two finite families of vectors

$$\{x^{h,1}, x^{h,2}, \ldots, x^{h,N_h}\} \subset \mathbb{R}^\nu \quad \text{and} \quad \{y^{h,1}, y^{h,2}, \ldots, y^{h,N_h}\} \subset \mathbb{R}^\ell$$

by the recursion: for $i = 0,1,\ldots,N_h - 1$,

$$\begin{aligned} & x^{h,i+1} = x^{h,i} + h\{A(\theta x^{h,i} + (1-\theta)x^{h,i+1}) + By^{h,i+1} + c(t_{h,i+1})\}, \\ & y^{h,i+1} \in \mathrm{SOL}(x^{h,i+1}), \end{aligned} \tag{3.7}$$

where $\theta \in [0,1]$ is a scalar.

When $\mathrm{SOL}(x^{h,i+1})$ contains multiple solutions, especially it is unbounded, choosing a least norm solution is necessary for the convergence [14, 23]. The following is a least norm time-stepping method,

$$\begin{aligned} & x^{h,i+1} = x^{h,i} + h\{A(\theta x^{h,i} + (1-\theta)x^{h,i+1}) + By^{h,i+1} + c(t_{h,i+1})\}, \\ & y^{h,i+1} = \mathrm{argmin}\{\|y\| \mid y \in \mathrm{SOL}(x^{h,i+1})\}. \end{aligned} \tag{3.8}$$

Using the residual function (3.6), the minimization problem in (3.8) can be equivalently written as

$$y^{h,i+1} = \text{argmin}\{\|y\| \mid r(y) = 0\}.$$

This is a mathematical programming with equilibrium constraints [24]. The usual mathematical programming constraint qualification such as Mangasarian-Fromovitz constraint qualification does not hold at any feasible solution [24]. To find a stable solution $y^{h,i+1}$, we use the STRF method to solve

$$\min_y \|r(y)\|^2 + \lambda_k \|y\|^2 \tag{3.9}$$

for some positive numbers $\lambda_k > 0$ and λ_k is gradually reduced to zero.

In the numerical experiments, we consider a nonlinear complementarity system, which is a special case of the DVI with $[a,b] = \mathbb{R}^n_+$, and $F(x(t), y(t)) = G(y(t)) + q(x(t))$. The solution set of the nonlinear complementarity subproblem, denoted as $\text{SOL}(x(t))$, contains $y(t) \in \mathbb{R}^\ell$ such that

$$y(t) \geqslant 0, \quad F(x(t), y(t)) = G(y(t)) + q(x(t)) \geqslant 0 \quad \text{and} \quad y^{\mathrm{T}}(t) F(x(t), y(t)) = 0,$$

where $G(\cdot) : \mathbb{R}^\ell \to \mathbb{R}^\ell$ is a nonlinear continuously differentiable function. It is easy to verify that $y(t) \in \text{SOL}(x(t))$ if and only if

$$r(y(t)) = y(t) - \max(0, y(t) - F(x(t))) = 0. \tag{3.10}$$

Taking $a_i = 0$ and $b_i = \{+\infty\}$ in the smoothing function of the "mid" function in Example 2, we can define a smoothing function $\tilde{r}(y(t), \mu)$ of $r(y(t))$ as follows:

$$\tilde{r}_i(y(t)) = \begin{cases} r_i(y(t)), & |y(t) - F(x(t))| \geqslant \mu, \\ y_i(t) - \dfrac{1}{4\mu}(y_i(t) - F_i(x(t)))^2 - \dfrac{1}{2}(y_i(t) - F_i(x(t))) - \dfrac{\mu}{4}, & \text{otherwise}. \end{cases}$$

In the numerical experiments, we set $T = 2$, $\ell = 7$, $\theta = 0$, $A = \left(\begin{smallmatrix} 2 & -2 \\ 1 & -1 \end{smallmatrix}\right)$, $B = \frac{1}{10\ell}\left(\begin{smallmatrix} e^{\mathrm{T}} \\ e^{\mathrm{T}} \end{smallmatrix}\right)$, $c(t) = (8\sin(20\pi t), 8\cos(20\pi t))^{\mathrm{T}}$, $G(y(t)) = \Lambda^{\mathrm{T}} H(y(t))$ and $q(x(t)) = \Lambda^{\mathrm{T}}(Qx(t) - 5e)$ with

$$\Lambda = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & -1 \\ 1 & 1 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 & 1 & -1 & -1 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 & -1 & 0 \end{pmatrix}, \quad H_i(y(t)) = 1 + 0.1\left(\frac{(\Lambda y(t))_i}{5}\right)^3,$$

for $i = 1, \ldots, 8$, and $Q = (e, e)$. Such function F can be found from a multi-commodity formulation for the traffic assignment based on Wardrop equilibrium [19].

The solution set $\text{SOL}(x(t))$ of this problem is unbounded, for any $t \in [0, T]$. To see it, suppose that $y_t^* \in \text{SOL}(x(t))$. Note that $\text{rank}(\Lambda) = 6$ and $\Lambda y_0 = 0$ with $y_0 = (1, 2, 1, 1, 6, 1, 2)^{\mathrm{T}} > 0$. Thus for any $\delta > 0$, we have $y_t^* + \delta y_0 \geqslant y_t^* \geqslant 0$,

$$F(x(t), y_t^* + \delta y_0) = \Lambda^{\mathrm{T}} H(y_t^* + \delta y_0) + q(x(t)) = \Lambda^{\mathrm{T}} H(y_t^*) + q(x(t)) = F(x(t), y_t^*) \geqslant 0,$$

and

$$\begin{aligned} (y_t^* + \delta y_0)^{\mathrm{T}} F(x(t), y_t^* + \delta y_0) &= (y_t^*)^{\mathrm{T}} F(x(t), y_t^*) + \delta y_0^{\mathrm{T}} F(x(t), y_t^*) \\ &= \delta y_0^{\mathrm{T}} F(x(t), y_t^*) = \delta y_0^{\mathrm{T}} \Lambda^{\mathrm{T}} H(y_t^*) + \delta y_0^{\mathrm{T}} \Lambda^{\mathrm{T}}(Qx(t) - 5e) = 0. \end{aligned}$$

Hence $y_t^* + \delta y_0$ is a solution for any $\delta > 0$, and $\|y_t^* + \delta y_0\| \geqslant \delta \to \infty$ as $\delta \to \infty$.

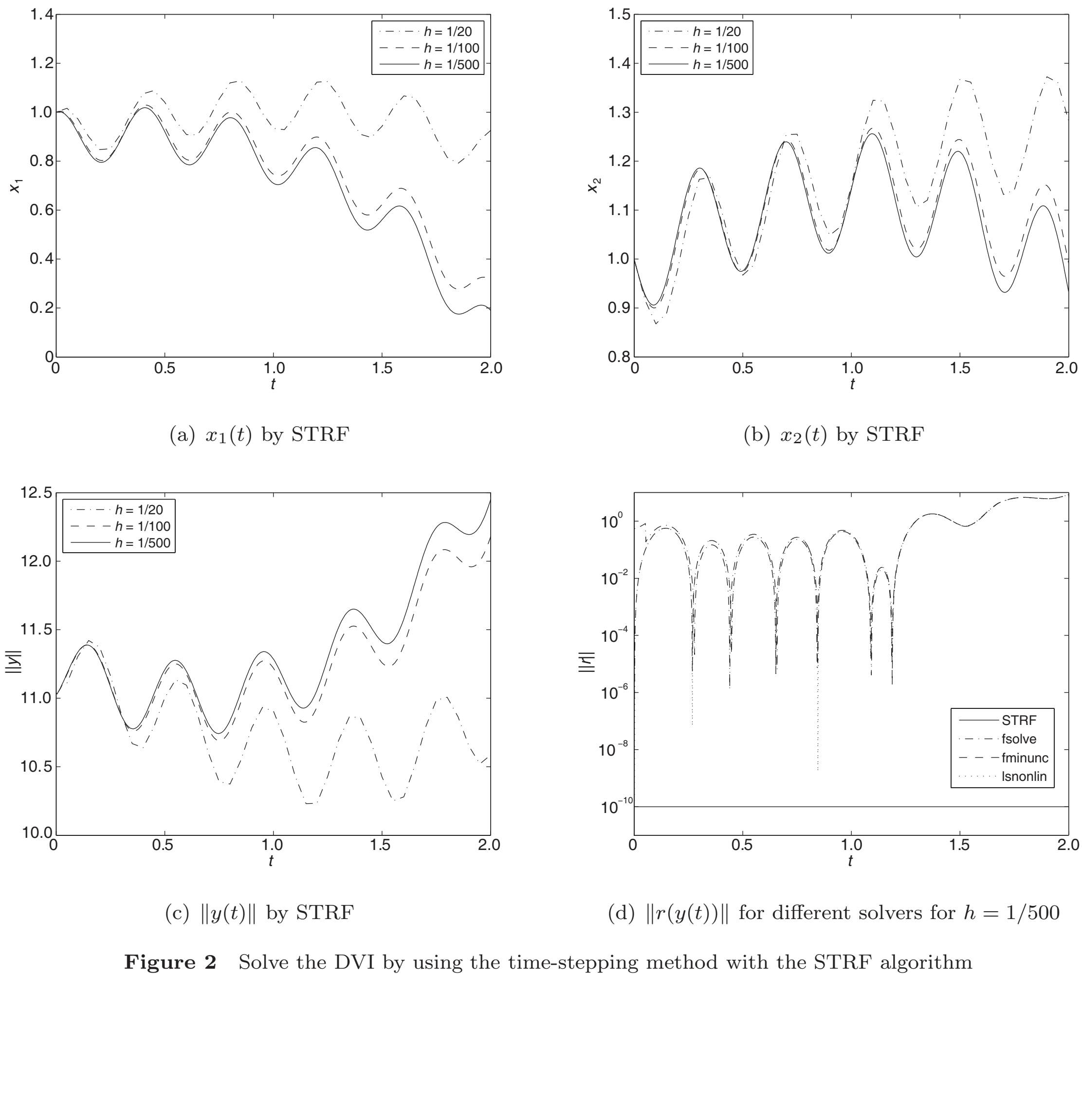
To find a stable solution $y(t)$ of the DVI (3.5), we solve the regularized problem (3.9) at each step of the STRF algorithm with $\lambda^k = 0.5^k$.

Let $x(0) = (1, 1)^{\mathrm{T}}$ and we select different time step sizes as $h = 1/20,\ 1/100,\ 1/500$. For each time step $t_{h,i+1}$, $i = 0, \ldots, N_h - 1$, we solve the least squares problem for $y(t_{h,i+1})$ by solving (3.5) with

$$r(y) = \min\{y\ ,\ G(y) + \Lambda^{\mathrm{T}} Q(I - hA)^{-1}(x^{h,i} + hBy + hc(t_{h,i+1})) - 5\Lambda^{\mathrm{T}} e\}$$

using the STRF algorithm with the initial vector $y^{h,i}$.

Figure 2 shows the numerical results of solving (3.5) by using time-stepping method with the STRF algorithm as the inner solver for the nonlinear complementarity subproblem with the three different time step sizes. Figures 2(a) and 2(b) are the shapes of $x(t)$ obtained with different time steps h. Figure 2(c) shows the changes of $\|y(t)\|$ with different time steps h. We also use the Matlab solvers "fsolve", "fminunc" and "lsqnonlin" to solve (3.5) with $h = 1/500$ and report the residuals $\|r(y)\|$ in Figure 2(d). For the convenience of plotting the figure, we let the residuals equal to 10^{-10} when the value of the residual is smaller than 10^{-10}. Seen from Figures 2(c) and 2(d), $\|y(t)\|$ stays bounded in each time step, and the residuals obtained by the STRF algorithm can reach down to 10^{-10} for $i = -1, \ldots, N_h - 1$, which means that it succeeds in finding global minimizers of (3.5) at all time steps. In contrast, Figure 2(d) also shows that we can not get $r(y)$ small enough at all the steps by using the three Matlab solvers as the inner solver of the time-stepping method. Compared with those solvers, the STRF algorithm is more suitable and efficient to solve the DVI as an inner solver.



(a) $x_1(t)$ by STRF (b) $x_2(t)$ by STRF

(c) $\|y(t)\|$ by STRF (d) $\|r(y(t))\|$ for different solvers for $h = 1/500$

Figure 2 Solve the DVI by using the time-stepping method with the STRF algorithm

4 Conclusions

In this paper, we proposed the STRF algorithm for solving nonsmooth nonconvex least squares problems, and proved that the STRF algorithm converges to a Clarke stationary point or a global minimizer of the objective function under certain conditions. Moreover, we showed that the STRF algorithm is efficient in finding spherical t_ϵ-designs and solutions of differential variational inequalities. How to improve the conditions for finding a global minimizer by the STRF algorithm is worth studying in the future.

Acknowledgements This work was supported by Hong Kong Research Grant Council (Grant No. PolyU5001/12p), and National Natural Science Foundation of China (Grant No. 11101231). The authors are grateful to Sven Leyffer for his useful comments on this paper.

References

1 Acary V, Brogliato B. Numerical Methods for Nonsmooth Dynamical Systems: Applications in Mechanics and Electronics. New York: Springer-Verlag, 2008

2 An C, Chen X, Sloan I H, et al. Well conditioned spherical designs for integration and interpolation on the two-sphere. SIAM J Numer Anal, 2010, 48: 2135–2157

3 Bannai E, Bannai E. A survey on spherical designs and algebraic combinatorics on spheres. European J Combin, 2009, 30: 1392–1425

4 Bian W, Chen X. Neural network for nonsmooth, nonconvex constrained minimization via smooth approximation. IEEE Trans Neural Netw Learn Syst, 2014, 25: 545–556

5 Burke J V, Hoheisel T. Epi-convergent smoothing with applications to convex composite functions. SIAM J Optim, 2013, 23: 1457–1479

6 Burke J V, Hoheisel T, Kanzow C. Gradient consistency for integral-convolution smoothing functions. Set-Valued Var Anal, 2013, 21: 359–376

7 Cartis C, Gould N I M, Toint P L. How Much Patience do You Have: A Worst-case Perspective on Smooth Noncovex Optimization. Swindon: Science and Technology Facilities Council, 2012

8 Chen X. Smoothing methods for nonsmooth, nonconvex minimization. Math Program, 2012, 134: 71–99

9 Chen X, Frommer A, Lang B. Computational existence proofs for spherical t-designs. Numer Math, 2011, 117: 289–305

10 Chen X, Niu L, Yuan Y. Optimality conditions and smoothing trust region Newton method for non-Lipschitz optimization. SIAM J Optim, 2013, 23: 1528–1552

11 Chen X, Wang Z. Convergence of regularized time-stepping methods for differential variational inequalities. SIAM J Optim, 2013, 23: 1647–1671

12 Chen X, Womersley R. Existence of solutions to systems of underdetermined equations and spherical designs. SIAM J Numer Anal, 2006, 44: 2326–2341

13 Chen X, Xiang S. Implicit solution function of P_0 and Z matrix linear complementarity constraints. Math Program, 2011, 128: 1–18

14 Chen X, Xiang S. Newton iterations in implicit time-stepping scheme for differential linear complementarity systems. Math Program, 2013, 138: 579–606

15 Clarke F H. Optimization and Nonsmooth Analysis. New York: John Wiley, 1983

16 Conn A R, Gould N I M, Toint P L. Trust Region Methods. Philadelphia: SIAM, 2000

17 Cottle R W, Pang J S, Stone R E. The Linear Complementarity Problem. Boston: Academic Press, 1992

18 Facchinei F, Pang J S. Finite-Dimensional Variational Inequalities and Complementarity Problems. New York: Springer-Verlag, 2003

19 Ferris M C, Pang J S. Engineering and economic applications of complementarity problems. SIAM Rev, 1997, 39: 669-713

20 Fletcher R, Leyffer S, Toint P L. On the global convergence of a filter-SQP algorithm. SIAM J Optim, 2002, 13: 44–59

21 Garmanjani R, Vicente L N. Smoothing and worst case complexity for direct-search methods in nonsmooth optimization. IMA J Numer Anal, 2013, 33: 1008–1028

22 Gould N I M, Leyffer S, Toint P L. A multidimensional filter algorithm for nonlinear equations and nonlinear least-squares. SIAM J Optim, 2004, 15: 17–38

23 Han L, Tiwari A, Camlibel M K, et al. Convergence of time-stepping schemes for passive and extended linear complementarity systems. SIAM J Numer Anal, 2009, 47: 3768–3796

24 Luo Z Q, Pang J S, Ralph D. Mathematical Programs with Equilibrium Constraints. Cambridge: Cambridge University Press, 1996

25 Nocedal J, Wright S J. Numerical Optimization, 2nd ed. New York: Springer, 2006
26 Pang J S, Stewart D E. Differential variational inequalities. Math Program, 2008, 113: 345–424
27 Sloan I H, Womersley R S. A variational characterisation of spherical designs. J Approx Theory, 2009, 159: 308–318
28 Yuan Y. Recent advances in trust region algorithms. Math Program, 2015, 151: 249–281
29 Zhou W, Chen X. Global convergence of a new hybrid Gauss-Newton structured BFGS method for nonlinear least squares problems. SIAM J Optim, 2010, 20: 2422–2441