Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2010, Article ID 976529, 19 pages
doi:10.1155/2010/976529

Research Article

Two-Phase Generalized Reduced Gradient Method for Constrained Global Optimization

Abdelkrim El Mouatasim

Department of Mathematics, Faculty of Science, Jazan University, P.O. Box 2097, Jazan, Saudi Arabia

Correspondence should be addressed to Abdelkrim El Mouatasim, a.mouatasim@gmail.com

Received 1 June 2010; Revised 17 September 2010; Accepted 31 October 2010

Academic Editor: Ying U. Hu

The random perturbation of generalized reduced gradient method for optimization under nonlinear differentiable constraints is proposed. Generally speaking, a particular iteration of this method proceeds in two phases. In the Restoration Phase, feasibility is restored by means of the resolution of an auxiliary nonlinear problem, a generally nonlinear system of equations. In the Optimization Phase, optimality is improved by means of the consideration of the objective function, on the tangent subspace to the constraints. In this paper, optimal assumptions are stated on the Restoration Phase and the Optimization Phase that establish the global convergence of the algorithm. Some numerical examples are also given by mixture problem and octagon problem.

1. Introduction

We consider the problem

$$\begin{aligned} &\text{Maximize } : \mathbf{f}(\mathbf{X}) \\ &\text{subject to } : \mathbf{h}(\mathbf{X}) = 0, \\ &\mathbf{X} \in \Omega, \end{aligned} \tag{1.1}$$

where $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable, and $\Omega \subset \mathbb{R}^n$ is closed and convex set (e.g., $\Omega = \prod_{i=1}^{n} [\mathbf{a}_i, \mathbf{b}_i]$).

Notice that any nonlinear programming can be put into standard form (1.1), by introduction of nonnegative *slack variables*, if there are inequalities (other than bounds on the variables) among the constraints, and by allowing some of the bounds to be $+\infty$ or $-\infty$ if necessary. The standard form is adopted here for ease in notations and discussion.

Feasible methods for optimization with constraints like generalized reduced gradient play an important role in practicing and are still widely used in technological applications. The main techniques proposed for solving constrained optimization problems in this work are generalized reduced gradient method via random perturbation. We are mainly interested in the situation where, on one hand, $\mathbf{f}$ is not concave and, on the other hand, the constraints are in general not linear. It is worth noting that some variant of the generalized gradient method reduces, in the case where all the constraints are linear, to the reduced gradient method [1], and some other variant, in the case of linear programming, to the Dantzig simplex method.

The problem (1.1) can be numerically approached by using sequential quadratic programming in [2], other methods for nonlinear programming in [3, 4], and generalized reduced gradient method [5], which generates a sequence $\{\mathbf{X}^k\}_{k\geq 0}$, where $\mathbf{X}^0$ is an initial feasible point and, for each $k > 0$, a new feasible point $\mathbf{X}^{k+1}$ is generated from $\mathbf{X}^k$ by using an operator $\mathbf{Q}_k$ (see Section 3). Thus the iterations are given by

$$k \geq 0 : \mathbf{X}^{k+1} = \mathbf{Q}_k\left(\mathbf{X}^k\right). \tag{1.2}$$

A fundamental difficulty arises due to the lack of concavity: the convergence of the sequence $\{\mathbf{X}^k\}_{k\geq 0}$ to a global maximum point is not ensured for the general situation considered. In order to prevent from converging to local maximum, various modifications of these basic methods have been introduced in the literature. For instance, we can find in the literature modifications of the basic descent methods [6–9], stochastic methods combined to penalty functions [10], evolutionary methods [11], and simulated annealing [12]. We introduce in this paper a different approach, inspired from the method of random perturbations introduced in [13] for unconstrained minimization of continuously differentiable functions and adapted to linearly constrained problems in [14].

In such a method, the sequence $\{\mathbf{X}^k\}_{k\geq 0}$ is replaced by a random vectors sequence $\{\widehat{\mathbf{X}}^k\}_{k\geq 0}$ and the iterations are modified as follows:

$$k \geq 0 : \widehat{\mathbf{X}}^{k+1} = \mathbf{Q}_k\left(\widehat{\mathbf{X}}^k\right) + \rho_k, \tag{1.3}$$

where ρ_k is a suitable random variable, called the stochastic perturbation. The sequence $\{\rho_k\}_{k\geq 0}$ goes to zero slowly enough in order to prevent convergence to a local maximum (see Section 4), the generalized reduced gradient method is recalled in Section 3, the notations are introduced in Section 2, and the results of some numerical experiments are given in Section 5.

2. Notations and Assumptions

We have the following.

(i) $\mathbf{E} = \mathbb{R}^n$ is the n-dimensional real Euclidean space.

(ii) $\mathbf{X}$ stands for $(\mathbf{x}_1, \ldots, \mathbf{x}_n)^T \in \mathbf{E}$.

(iii) $\mathbf{h}(\mathbf{X})$ is the column vector whose components are $\mathbf{h}_1(\mathbf{X}), \ldots, \mathbf{h}_m(\mathbf{X})$.

(iv) $\|\mathbf{X}\| = (\mathbf{x}_1^2 + \cdots + \mathbf{x}_n^2)^{1/2}$ is the Euclidean norm of $\mathbf{X}$:

$$\|\mathbf{A}\|_\infty = \sup\{\|\mathbf{AX}\| : \|\mathbf{X}\| = 1\}, \tag{2.1}$$

(v) $\mathbf{A}^T$ is the transpose matrix associated to $\mathbf{A}$.

Definition 2.1. A point $\mathbf{X}$ *is said to be feasible if* $|\mathbf{h}_i(\mathbf{X})| \leq \varepsilon_0$, *for* $i = 1, \ldots, m$, *where* ε_0 *is same preassigned small positive constant.*

Assume that we know some feasible point $\mathbf{X}^0$. We make the following nonedegeneracy assumption. The vector $\mathbf{X}$ can be split into two components: an m-dimensional component $\mathbf{y}$ (the basic part), and $\mathbf{x}$, a component of dimension $n - m$ (the nonbasic part) such that the following two properties hold:

(H1) $\mathbf{y}^0$ is strictly between bounds;

(H2) the square $m \times m$ matrix $\partial\mathbf{h}/\partial\mathbf{y}$, computed at $\mathbf{X}^0$, is nonsingular.

If the property (H2) does not hold or $n < m$, we add *artificial variables* to the constraints, so the property (H2) holds and $n \geq m$ in associate problem (1.1).

Let

$$\mathbf{C} = \{\mathbf{X} \in \mathbf{E} \mid \mathbf{h}(\mathbf{X}) = 0, \mathbf{X} \in \Omega\}. \tag{2.2}$$

The objective function is $\mathbf{f} : \mathbf{E} \to \mathbb{R}$, and its upper bound on $\mathbf{C}$ is denoted by $\mathbf{l}^*$:

$$\mathbf{l}^* = \max_{\mathbf{C}} \mathbf{f}. \tag{2.3}$$

Let us introduce

$$\mathbf{C}_\alpha = \mathbf{S}_\alpha \cap \mathbf{C}; \quad \text{where} \quad \mathbf{S}_\alpha = \{\mathbf{X} \in \mathbf{E} \mid \mathbf{f}(\mathbf{X}) \geq \alpha\}. \tag{2.4}$$

We assume that

$$\mathbf{f} \text{ is twice continuously differentiable on } \mathbf{E}, \tag{2.5}$$

$$\forall \alpha < \mathbf{l}^* : \mathbf{C}_\alpha \text{ is not empty, closed and bounded,} \tag{2.6}$$

$$\forall \alpha < \mathbf{l}^\star : \operatorname{meas}(\mathbf{C}_\alpha) > 0, \tag{2.7}$$

where $\operatorname{meas}(\mathbf{C}_\alpha)$ is the measure of $\mathbf{C}_\alpha$.

Since $\mathbf{E}$ is a finite dimensional space, assumption (2.6) is verified when $\mathbf{C}$ is bounded or $\mathbf{f}$ is coercive, that is, $\lim_{\|\mathbf{X}\| \to +\infty} \mathbf{f}(\mathbf{X}) = -\infty$. Assumption (2.6) is verified when $\mathbf{C}$ contains a sequence of neighborhoods of a point of optimum $\mathbf{X}^\star$ having strictly positive measure, that is, when $\mathbf{X}^\star$ can be approximated by a sequence of points of the interior of $\mathbf{C}$. We observe that the assumptions (2.5)-(2.6) yield that

$$\mathbf{C} = \bigcup_{\alpha < \mathbf{l}^\star} \mathbf{C}_\alpha, \quad \text{that is, for all } \mathbf{X} \in \mathbf{C} : \exists \alpha < \mathbf{l}^* \text{ such that } \mathbf{X} \in \mathbf{C}_\alpha. \tag{2.8}$$

From (2.5)-(2.6),

$$\gamma_1 = \sup\{\|\nabla \mathbf{f}(\mathbf{X})\| : \mathbf{X} \in \mathbf{C}_\alpha\} < +\infty. \tag{2.9}$$

Then

$$\gamma_2 = \sup\{\|\mathbf{d}\| : \mathbf{X} \in \mathbf{C}_\alpha\} < +\infty, \tag{2.10}$$

where $\mathbf{d}$ is the direction determined on the basis of the gradient $\nabla \mathbf{f}(\mathbf{X})$.

Thus,

$$\beta(\alpha, \varepsilon) = \sup\{\|\mathbf{Y} - (\mathbf{X} + \eta \mathbf{d})\| : (\mathbf{X}, \mathbf{Y}) \in \mathbf{C}_\alpha \times \mathbf{C}_\alpha, 0 \le \eta \le \varepsilon\} < +\infty, \tag{2.11}$$

where ε, η are positive real numbers.

3. Generalized Reduced Gradient Method

By the implicit function theorem, there exists, in some neighborhood $\mathbf{V}$ of $\mathbf{x}^0$, a unique continuous function (mapping), say $\mathbf{y}(\mathbf{x})$, such that $\mathbf{f}(\mathbf{x}, \mathbf{y}(\mathbf{x}))$ is identically zero in $\mathbf{V}$. In addition, $\mathbf{y}(\mathbf{x})$ has a continuous derivative $\mathbf{dy}/\mathbf{dx}$, which can be computed by the chain rule:

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}} + \left(\frac{\partial \mathbf{h}}{\partial \mathbf{y}}\right)\left(\frac{\mathbf{dy}}{\mathbf{dx}}\right) = 0, \tag{3.1}$$

or, more conveniently,

$$\frac{\mathbf{dy}}{\mathbf{dx}} = -\left(\frac{\partial \mathbf{h}}{\partial \mathbf{y}}\right)^{-1}\left(\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right). \tag{3.2}$$

In what follows, we call $\mathbf{A}$ the Jacobian of $\mathbf{h}(\mathbf{X})$ computed at $\mathbf{X}^0$. Similarly, we set

$$\mathbf{B} = \frac{\partial \mathbf{h}}{\partial \mathbf{y}}, \qquad \mathbf{N} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \left(\text{computed at } \mathbf{X}^0\right). \tag{3.3}$$

Substituting $\mathbf{y}(\mathbf{x})$ into the objective function $\mathbf{f}(\mathbf{x}, \mathbf{y})$, we obtain the reduced function:

$$\phi(\mathbf{x}) = \mathbf{f}(\mathbf{x}, \mathbf{y}(\mathbf{x})), \tag{3.4}$$

the gradient of which at $\mathbf{x}^0$ is, by the chain rule again,

$$\mathbf{g} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} + \left(\frac{\partial \mathbf{f}}{\partial \mathbf{y}}\right)\left(\frac{\mathbf{dy}}{\mathbf{dx}}\right) \tag{3.5}$$

(all derivatives computed at $\mathbf{X}^0$). Setting

$$\mathbf{v} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}, \qquad \mathbf{w} = \frac{\partial \mathbf{f}}{\partial \mathbf{y}} \left(\text{computed at } \mathbf{X}^0\right), \tag{3.6}$$

we have the following formula for the *reduced gradient* g:

$$\mathbf{g} = \mathbf{v} - \mathbf{w}\mathbf{B}^{-1}\mathbf{N}. \tag{3.7}$$

The generalized reduced gradient method tries to extend the methods of linear optimization to the nonlinear case. These methods are close to, or equivalent to, projected gradient methods [15]; just the presentation of the methods is frequently quite different.

Let us define the projected reduced gradient $\mathbf{p}$ (in the space of $\mathbf{g}$) by its components:

(i) $\mathbf{p}_j = 0$ if $\mathbf{x}_j^0 = \mathbf{a}_j$ and $\mathbf{g}_j < 0$;

(ii) $\mathbf{p}_j = 0$ if $\mathbf{x}_j^0 = \mathbf{b}_j$ and $\mathbf{g}_j > 0$;

(iii) $\mathbf{p}_j = \mathbf{g}_j$, otherwise.

It is convenient to set

$$\mathbf{u} = \mathbf{w}\mathbf{B}^{-1}, \tag{3.8}$$

and hence

$$\mathbf{w} - \mathbf{u}\mathbf{B} = 0, \qquad \mathbf{v} - \mathbf{u}\mathbf{N} = \mathbf{g}. \tag{3.9}$$

It is a simple matter to verify that Kuhn-Tucker condition for the problem in standard form reduces to $\mathbf{p} = 0$, and that $\mathbf{u}$ is the row-vector of multipliers corresponding to the equations $\mathbf{h}(\mathbf{X}) = 0$. We assume from now on that $\mathbf{p} \neq 0$.

The following relations

$$\mathbf{x}_j = \mathbf{a}_j \text{ if } \mathbf{x}_j^0 = \mathbf{a}_j \text{ and } \mathbf{g}_j < 0,$$

$$\mathbf{x}_j = \mathbf{b}_j \text{ if } \mathbf{x}_j^0 = \mathbf{b}_j \text{ and } \mathbf{g}_j > 0$$

define what we call the *face* (at $\mathbf{X}^0$), denoted by $\mathbf{F}$. The row-vector $\mathbf{p}$ (the projected reduced gradient) is also named the projection of $\mathbf{g}$ onto $\mathbf{F}$.

Let $\mathbf{d}$ be any nonzero vector column-vector in $\mathbf{F}$ such that $\mathbf{p}\mathbf{d} > 0$; the vector $\mathbf{d}$ is an ascent direction for the reduced function $\phi(\mathbf{x})$.

There is a striking analogy with what is usually done in linear programming, where $\mathbf{y}(\mathbf{x})$ can be computed in close form. This is generally not the case if the constraints $\mathbf{h}(\mathbf{X})$ are nonlinear. Even if close form is available, actual substitution may very well be undesirable.

The generalized reduced gradient algorithm consists of the following steps.

Step 1. Assume that some that feasible $\mathbf{X}^0$ is known. Set $k = 0$ and go to the next step.

Step 2. Step 2 is conveniently divided into substeps.

(1.1) Compute the Jacobian $\mathbf{A}$ and the gradient of the objective function.

(1.2) Determine a splitting of $\mathbf{X}$ into $(\mathbf{x}, \mathbf{y})$ and corresponding of $\mathbf{A}$ into $(\mathbf{N}, \mathbf{B})$, such that $\mathbf{y}^k$ is strictly between bound and $\mathbf{B}$ is nonsingular;

invert $\mathbf{B}$.

(1.3) Compute the Lagrange multipliers $\mathbf{u}$ and the reduced gradient $\mathbf{g}$.

(1.4) Determine the face $\mathbf{F}$ and the projection $\mathbf{p}$ of $\mathbf{g}$ onto $\mathbf{F}$ [16].

(1.5) If $\mathbf{p}$ is zero (or almost zero in some sense), then terminate. $\mathbf{X}^k$ is a KKT point [17, 18]. Otherwise, go to the next step.

Step 3. Choose ascent direction $\mathbf{d} = \mathbf{p}^t$ [18], that is,

$$d_j = \begin{cases} 0 & \text{if } g_j > 0,\ u_j = a_j, \\ 0 & \text{if } g_j < 0,\ u_j = b_j, \\ g_j, & \text{otherwise.} \end{cases} \tag{3.10}$$

Step 4. Choose a first stepsize η_1.

Step 5. Maximize, with respect to η, the function

$$\psi(\eta) = \mathbf{f}\left(\mathbf{x}^k + \eta\mathbf{d}, \mathbf{y}\left(\mathbf{x}^k + \eta\mathbf{d}\right)\right) = \phi\left(\mathbf{x}^k + \eta\mathbf{d}\right) \tag{3.11}$$

with more or less accuracy (the linear search). For each value of η under consideration, this step requires solving the following system of m equations:

$$\mathbf{h}\left(\mathbf{x}^k + \eta\mathbf{d}, \mathbf{y}\right) = 0, \tag{3.12}$$

where $\mathbf{y}$ is the m-dimensional vector of unknowns.

Step 6. Assuming that Step 5 succeeds, an improved feasible point is obtained, which replaced $\mathbf{X}^k$. Replace k by $k + 1$ and go back to Step 2.

Let us recall briefly the essential points of the generalized reduced gradient method: an initial feasible guess $\mathbf{X}^0 \in \mathbf{C}$ is given and a sequence $\{\mathbf{X}^k\}_{k\geq 0} \subset \mathbf{C}$ is generated by using iterations of the general form:

$$\forall k \geq 0 \quad \mathbf{X}^{k+1} = \mathbf{Q}_k\left(\mathbf{X}^k\right) = \mathbf{X}^k + \eta_k\mathbf{d}^k. \tag{3.13}$$

Remark 3.1. The theoretical convergence of generalized reduced gradient method has been proved by [17, 19, 20].

4. Two-Phase Generalized Reduced Gradient (TPGRG) Method

The main difficulty remains the lack of concaveity: if $\mathbf{f}$ is not concave, the Kuhn-Tucker points may not correspond to a global maximum (see e.g., [7, 8]). In the next, this point is improved by using an appropriate random perturbation.

The sequence of real numbers $\{\mathbf{X}^k\}_{k\geq 0}$ is replaced by a sequence of random variables $\{\widehat{\mathbf{X}}^k\}_{k\geq 0}$ involving a random perturbation ρ_k of the deterministic iteration (3.13); then we have $\widehat{\mathbf{X}}^0 = \mathbf{X}^0$;

$$\forall k \geq 0 \quad \widehat{\mathbf{X}}^{k+1} = \mathrm{Q}_k\left(\widehat{\mathbf{X}}^k\right) + \rho_k = \widehat{\mathbf{X}}^k + \eta_k \mathbf{d}^k + \rho_k = \widehat{\mathbf{X}}^k + \eta_k\left(\mathbf{d}^k + \frac{\rho_k}{\eta_k}\right), \tag{4.1}$$

where $\eta_k \neq 0$ satisfied the Step 5 in GRG algorithm, and

$$\forall k \geq 1 : \rho_k \text{ is independent from } \left(\widehat{\mathbf{X}}^{k-1}, \ldots, \widehat{\mathbf{X}}^0\right), \tag{4.2}$$

see [9], and

$$\widehat{\mathbf{X}} \in \mathbf{C} \Longrightarrow \mathbf{Q}_k\left(\widehat{\mathbf{X}}\right) + \rho_k \in \mathbf{C}. \tag{4.3}$$

Equation (4.1) can be viewed as perturbation of the ascent direction $\mathbf{d}^k$, which is replaced by a new direction $\mathbf{D}_k = \mathbf{d}^k + \rho_k/\eta_k$, and the iterations (4.1) become

$$\widehat{\mathbf{X}}^{k+1} = \widehat{\mathbf{X}}^k + \eta_k \mathbf{D}_k. \tag{4.4}$$

General properties defining convenient sequences of perturbation $\{\rho_k\}_{k\geq O}$ can be found in the literature [13, 14]: usually, sequence of Gaussian laws may be used in order to produce elements satisfying these properties.

We introduce a random vector $\mathfrak{Z}_k$, and we denote by Φ_k and ϕ_k the cumulative distribution function and the probability density of $\mathfrak{Z}_k$, respectively.

We denote by $\mathbf{F}_{k+1}(\mathbf{Y} \mid \widehat{\mathbf{X}}^k = \mathbf{X})$ the conditional cumulative distribution function:

$$\mathbf{F}_{k+1}\left(\mathbf{Y} \mid \widehat{\mathbf{X}}^k = \mathbf{X}\right) = P\left(\widehat{\mathbf{X}}^{k+1} < \mathbf{Y} \mid \widehat{\mathbf{X}}^k = \mathbf{X}\right), \tag{4.5}$$

and the condition probability density of $\widehat{\mathbf{X}}^{k+1}$ is denoted by $\mathbf{f}_{k+1}$.

Let us introduce a sequence of n-dimensional random vectors $\{\mathbf{Z}_k\}_{k\geq 0} \in \mathbf{C}$. We consider also $\{\xi_k\}_{k\geq 0}$, a suitable decreasing sequence of strictly positive real numbers converging to 0 and such that $\xi_0 \leq 1$.

The optimal choice for η_k is determined by Step 5. Let $\rho_k = \xi_k \mathbf{Z}_k$:

$$\mathbf{F}_{k+1}\left(\mathbf{Y} \mid \widehat{\mathbf{X}}^k = \mathbf{X}\right) = P\left(\widehat{\mathbf{X}}^{k+1} < \mathbf{Y} \mid \widehat{\mathbf{X}}^k = \mathbf{X}\right). \tag{4.6}$$

It follows that

$$\mathbf{F}_{k+1}\left(\mathbf{Y} \mid \widehat{\mathbf{X}}^k = \mathbf{X}\right) = P\left(\mathbf{Z}_k < \frac{\mathbf{Y} - \mathbf{Q}_k(\mathbf{X})}{\xi_k}\right) = \Phi_k\left(\frac{\mathbf{Y} - \mathbf{Q}_k(\mathbf{X})}{\xi_k}\right). \tag{4.7}$$

So, we have

$$\mathbf{f}_{k+1}\left(\mathbf{Y} \mid \widehat{\mathbf{X}}^k = \mathbf{X}\right) = \frac{1}{\xi_k^n}\phi_k\left(\frac{\mathbf{Y} - \mathbf{Q}_k(\mathbf{X})}{\xi_k}\right) \quad \mathbf{Y} \in \mathbf{C}. \tag{4.8}$$

The relation (2.11) shows that

$$\|\mathbf{Y} - \mathbf{Q}_k(\mathbf{X})\| \leq \beta(\alpha, \varepsilon) \quad \text{for } (\mathbf{X}, \mathbf{Y}) \in \mathbf{C}_\alpha \times \mathbf{C}_\alpha. \tag{4.9}$$

We assume that there exists a *decreasing function* $t \mapsto \mathbf{g}_k(t)$, $\mathbf{g}_k(t) > 0$ on $\mathbb{R}^+$ such that

$$\mathbf{Y} \in \mathbf{C}_\alpha \Longrightarrow \phi_k\left(\frac{\mathbf{Y} - \mathbf{Q}_k(\mathbf{X})}{\xi_k}\right) \geq \mathbf{g}_k\left(\frac{\beta(\alpha, \varepsilon)}{\xi_k}\right). \tag{4.10}$$

For simplicity, let

$$\mathbf{Z}_k = 1_{\mathbf{C}}(\mathfrak{Z}_k)\mathfrak{Z}_k, \tag{4.11}$$

where $\mathfrak{Z}$ is a random variable; for simplicity let $\mathfrak{Z} \sim \mathbf{N}(\mathbf{0}, 1)$.

The procedure generates a sequence $\mathbf{U}_k = f(\widehat{\mathbf{X}}^k)$. By construction this sequence is increasing and upper bounded by $\mathbf{l}^*$:

$$\forall k \geq 0 : \mathbf{l}^* \geq \mathbf{U}_{k+1} \geq \mathbf{U}_k. \tag{4.12}$$

Thus, there exists $\mathbf{U} \leq \mathbf{l}^*$ such that

$$\mathbf{U}_k \longrightarrow \mathbf{U} \quad \text{for } k \longrightarrow +\infty. \tag{4.13}$$

Lemma 4.1. *Let* $\mathcal{D}_k = \xi_k \mathbf{Z}_k$ *and* $\gamma = \mathbf{f}(\mathbf{X}^0)$ *if* $\mathbf{Z}_k$ *is given by* (4.11)*; then there exists* $\nu > 0$ *such that*

$$P(\mathbf{U}_{k+1} > \theta \mid \mathbf{U}_k \leq \theta) \geq \frac{\operatorname{meas}(\mathbf{C}_\gamma - \mathbf{C}_\theta)}{\xi_k^n}\mathbf{g}_k\left(\frac{\beta(\gamma, \varepsilon)}{\xi_k}\right) > 0 \quad \forall \theta \in [\mathbf{l}^* - \nu, \mathbf{l}^*), \tag{4.14}$$

where $\mathbf{n} = \dim(\mathbf{E})$.

Proof. Let $\widehat{\mathbf{C}}_{\boldsymbol{\theta}} = \{\mathbf{X} \in \mathbf{C} \mid \mathbf{f}(\mathbf{X}) > \theta\}$, for $\theta \in [\mathbf{l}^* - \nu, \mathbf{l}^*)$. □

Since $\mathbf{C}_\alpha \subset \widehat{\mathbf{C}}_{\boldsymbol{\theta}}$, $\mathbf{l}^* > \alpha > \theta$, it follows from (2.7) that $\widehat{\mathbf{C}}_\theta$ is not empty and has a strictly positive measure.

If $\text{meas}(\mathbf{C} - \widehat{\mathbf{C}}_\theta) = 0$ for any $\theta \in [\mathbf{l}^* - \nu, \mathbf{l}^*)$, the result is immediate, since we have $f(\mathbf{X}) = \mathbf{l}^*$ on $\mathbf{C}$.

Let us assume that there exists $\varepsilon > 0$ such that $\text{meas}(\mathbf{C} - \widehat{\mathbf{C}}_\varepsilon) > 0$. For $\theta \in [\mathbf{l}^* - \varepsilon, \mathbf{l}^*)$, we have $\widehat{\mathbf{C}}_{\boldsymbol{\theta}} \subset \widehat{\mathbf{C}}_\varepsilon$ and $\text{meas}(\mathbf{C} - \widehat{\mathbf{C}}_{\boldsymbol{\theta}}) > 0$.

$P(\widehat{\mathbf{X}}^k \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}) = P(\widehat{\mathbf{X}}^k \in \mathbf{C} - \widehat{\mathbf{C}}_{\boldsymbol{\theta}}) = \int_{\mathbf{C}-\widehat{\mathbf{C}}_\theta} P(\widehat{\mathbf{X}}^k \in \mathbf{dX}) > 0$ for any $\theta \in [\mathbf{l}^* - \varepsilon, \mathbf{l}^*)$, since the sequence $\{\mathbf{U}_i\}_{i\geq 0}$ is increasing, and we have also

$$\left\{\widehat{\mathbf{X}}^i\right\}_{i\geq 0} \subset \mathbf{C}_\gamma. \tag{4.15}$$

Thus

$$P\left(\widehat{\mathbf{X}}^k \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right) = P\left(\widehat{\mathbf{X}}^k \in \mathbf{C}_\gamma - \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right) = \int_{\mathbf{C}_\gamma - \widehat{\mathbf{C}}_{\boldsymbol{\theta}}} P\left(\widehat{\mathbf{X}}^k \in \mathbf{dX}\right) > 0 \quad \text{for any } \theta \in [\mathbf{l}^* - \varepsilon, \mathbf{l}^*). \tag{4.16}$$

Let $\theta \in [\mathbf{l}^* - \varepsilon, \mathbf{l}^*)$; we have from (4.12)

$$P(\mathbf{U}_{k+1} > \theta \mid \mathbf{U}_k \leq \theta) = P\left(\widehat{\mathbf{X}}^{k+1} \in \widehat{\mathbf{C}}_{\boldsymbol{\theta}} \mid \widehat{\mathbf{X}}^i \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}},\ i = 0, \ldots, k\right). \tag{4.17}$$

But Markov chain yields that

$$P\left(\widehat{\mathbf{X}}^{k+1} \in \widehat{\mathbf{C}}_{\boldsymbol{\theta}} \mid \widehat{\mathbf{X}}^i \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}},\ i = 0, \ldots, k\right) = P\left(\widehat{\mathbf{X}}^{k+1} \in \widehat{\mathbf{C}}_{\boldsymbol{\theta}} \mid \widehat{\mathbf{X}}^k \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right). \tag{4.18}$$

By the conditional probability rule

$$P\left(\widehat{\mathbf{X}}^{k+1} \in \widehat{\mathbf{C}}_{\boldsymbol{\theta}} \mid \widehat{\mathbf{X}}^k \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right) = \frac{P\left(\widehat{\mathbf{X}}^{k+1} \in \widehat{\mathbf{C}}_{\boldsymbol{\theta}}, \widehat{\mathbf{X}}^k \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right)}{P\left(\widehat{\mathbf{X}}^k \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right)}. \tag{4.19}$$

Moreover

$$P\left(\widehat{\mathbf{X}}^{k+1} \in \widehat{\mathbf{C}}_{\boldsymbol{\theta}}, \widehat{\mathbf{X}}^k \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right) = \int_{\mathbf{C}-\widehat{\mathbf{C}}_\theta} P\left(\widehat{\mathbf{X}}^k \in \mathbf{dX}\right) \int_{\widehat{\mathbf{C}}_\theta} \mathbf{f}_{k+1}\left(\mathbf{Y} \mid \widehat{\mathbf{X}}^k = \mathbf{X}\right)\mathbf{dY}. \tag{4.20}$$

From (4.15), we have

$$P\left(\widehat{\mathbf{X}}^{k+1} \in \widehat{\mathbf{C}}_{\boldsymbol{\theta}}, \widehat{\mathbf{X}}^k \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right) = \int_{\mathbf{C}_\gamma-\widehat{\mathbf{C}}_\theta} P\left(\widehat{\mathbf{X}}^k \in \mathbf{dX}\right) \int_{\widehat{\mathbf{C}}_\theta} \mathbf{f}_{k+1}\left(\mathbf{Y} \mid \widehat{\mathbf{X}}^k = \mathbf{X}\right)\mathbf{dY}, \tag{4.21}$$

$$P\left(\widehat{\mathbf{X}}^{k+1} \in \widehat{\mathbf{C}}_{\boldsymbol{\theta}}, \widehat{\mathbf{X}}^k \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right) \geq \inf_{\mathbf{X}\in\mathbf{C}_\gamma-\widehat{\mathbf{C}}_\theta} \left\{\int_{\widehat{\mathbf{C}}_\theta} \mathbf{f}_{k+1}\left(\mathbf{Y} \mid \widehat{\mathbf{X}}^k = \mathbf{X}\right)\mathbf{dY}\right\} \int_{\mathbf{C}_\gamma-\widehat{\mathbf{C}}_\theta} P\left(\widehat{\mathbf{X}}^k \in \mathbf{dX}\right). \tag{4.22}$$

Thus

$$P\left(\widehat{\mathbf{X}}^{k+1} \in \widehat{\mathbf{C}}_{\boldsymbol{\theta}} \mid \widehat{\mathbf{X}}^{k} \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right) \geq \inf_{\mathbf{X} \in \mathbf{C}_{\gamma} - \widehat{\mathbf{C}}_{\theta}} \left\{ \int_{\widehat{\mathbf{C}}_{\theta}} \mathbf{f}_{k+1}\left(\mathbf{Y} \mid \widehat{\mathbf{X}}^{k} = \mathbf{X}\right) \mathbf{d}\mathbf{Y} \right\}. \tag{4.23}$$

Taking (4.8) into account, we have

$$P\left(\widehat{\mathbf{X}}^{k+1} \in \widehat{\mathbf{C}}_{\boldsymbol{\theta}} \mid \widehat{\mathbf{X}}^{k} \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right) \geq \frac{1}{\xi_k^n} \inf_{\mathbf{X} \in \mathbf{C}_{\gamma} - \widehat{\mathbf{C}}_{\theta}} \left\{ \int_{\widehat{\mathbf{C}}_{\theta}} \phi_k\left(\frac{\mathbf{Y} - \mathbf{Q}_k(\mathbf{X})}{\xi_k}\right) \mathbf{d}\mathbf{Y} \right\}. \tag{4.24}$$

The relation (2.11) shows that

$$\|\mathbf{Y} - \mathbf{Q}_k(\mathbf{X})\| \leq \beta(\gamma, \varepsilon), \tag{4.25}$$

and (4.10) yields that

$$\phi_k\left(\frac{\mathbf{Y} - \mathbf{Q}_k(\mathbf{X})}{\xi_k}\right) \geq \mathbf{g}_k\left(\frac{\beta(\gamma, \varepsilon)}{\xi_k}\right). \tag{4.26}$$

Hence

$$\begin{aligned} P\left(\widehat{\mathbf{X}}^{k+1} \in \widehat{\mathbf{C}}_{\boldsymbol{\theta}} \mid \widehat{\mathbf{X}}^{k} \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right) &\geq \frac{1}{\xi_k^n} \inf_{\mathbf{X} \in \mathbf{C}_{\gamma} - \widehat{\mathbf{C}}_{\theta}} \int_{\widehat{\mathbf{C}}_{\theta}} \mathbf{g}_k\left(\left(\frac{\beta(\gamma, \varepsilon)}{\xi_k}\right)\right) \mathbf{d}\mathbf{Y}, \\ P\left(\widehat{\mathbf{X}}^{k+1} \in \widehat{\mathbf{C}}_{\boldsymbol{\theta}} \mid \widehat{\mathbf{X}}^{k} \notin \widehat{\mathbf{C}}_{\boldsymbol{\theta}}\right) &\geq \frac{\operatorname{meas}\left(\mathbf{C}_{\gamma} - \mathbf{C}_{\theta}\right)}{\xi_k^n} \mathbf{g}_k\left(\frac{\beta(\gamma, \varepsilon)}{\xi_k}\right). \end{aligned} \tag{4.27}$$

4.1. Global Convergence

The global convergence is a consequence of the following result, which yields from the Borel-Catelli's lemma (e.g., see [13]):

Lemma 4.2. *Let $\{\mathbf{U}_k\}_{k\geq 0}$ be an increasing sequence, upper bounded by $\mathbf{l}^*$. Then, there exists $\mathbf{U}$ such that $\mathbf{U}_k \to \mathbf{U}$ for $k \to +\infty$. Assume that there exists $\nu > 0$ such that for any $\theta \in [\mathbf{l}^* - \nu, \mathbf{l}^*)$, there is a sequence of strictly positive real numbers $\{\mathbf{c}_k(\theta)\}_{k\geq 0}$ such that*

$$\forall k \geq 0 : P(\mathbf{U}_{k+1} > \theta \mid \mathbf{U}_k \leq \theta) \geq \mathbf{c}_k(\theta) > 0, \qquad \sum_{k=0}^{+\infty} \mathbf{c}_k(\theta) = +\infty. \tag{4.28}$$

Then $\mathbf{U} = \mathbf{l}^*$ almost surely.

Proof. For instance, see [13, 21]. □

Theorem 4.3. *Let $\gamma = \mathbf{f}(\mathbf{X}^0)$; assume that $\mathbf{X}^0 \in \mathbf{C}$, the sequence ξ_k is nonincreasing, and*

$$\sum_{k=0}^{+\infty} \mathbf{g}_k\left(\frac{\beta(\gamma,\varepsilon)}{\xi_k}\right) = +\infty. \tag{4.29}$$

Then $\mathbf{U} = \mathbf{l}^*$ almost surely.

Proof. Let

$$\mathbf{c}_k(\theta) = \frac{\operatorname{meas}(\mathbf{C}_\gamma - \mathbf{C}_\theta)}{\xi_k^n}\mathbf{g}_k\left(\frac{\beta(\gamma,\varepsilon)}{\xi_k}\right) > 0. \tag{4.30}$$

Since the sequence $\{\xi_k\}_{k\geq 0}$ is nonincreasing,

$$\mathbf{c}_k(\theta) \geq \frac{\operatorname{meas}(\mathbf{C}_\gamma - \mathbf{C}_\theta)}{\xi_0^n}\mathbf{g}_k\left(\frac{\beta(\gamma,\varepsilon)}{\xi_k}\right) > 0. \tag{4.31}$$

Thus, (4.29) shows that

$$\sum_{k=0}^{+\infty} \mathbf{c}_k(\theta) \geq \frac{\operatorname{meas}(\mathbf{C}_\gamma - \mathbf{C}_\theta)}{\xi_0^n}\sum_{k=0}^{+\infty} \mathbf{g}_k\left(\frac{\beta(\gamma,\varepsilon)}{\xi_k}\right) = +\infty. \tag{4.32}$$

Using Lemmas 4.1 and 4.2 we have $\mathbf{U} = \mathbf{l}^*$ almost surely. □

Theorem 4.4. *Let $\mathbf{Z}_k$ be defined by* (4.11)*, and let*

$$\xi_k = \sqrt{\frac{\widehat{\mathbf{a}}}{\log\left(k + \widehat{\mathbf{d}}\right)}}, \tag{4.33}$$

where $\widehat{\mathbf{a}} > 0$, $\widehat{\mathbf{d}} > 0$, and k is the iteration number. If $\mathbf{X}^0 \in C$, then, for $\widehat{\mathbf{a}}$ large enough, $\mathbf{U} = l^*$ almost surely.

Proof . We have

$$\phi_k(\mathbf{Z}) = \frac{1}{\left(\sqrt{2\pi}\right)^n}\exp\left(-\frac{1}{2}\|\mathbf{Z}\|^2\right) = \mathbf{g}_k(\|\mathbf{Z}\|) > 0. \tag{4.34}$$

So,

$$\mathbf{g}_k\left(\frac{\beta(\gamma,\varepsilon)}{\xi_k}\right) = \frac{1}{\left(\sqrt{2\pi}\right)^n\left(k + \widehat{\mathbf{d}}\right)^{\beta(\gamma,\varepsilon)^2/(2\widehat{\mathbf{a}})}}. \tag{4.35}$$

For $\hat{\mathbf{a}}$ such that

$$0 < \frac{\beta(\gamma,\varepsilon)^2}{2\hat{\mathbf{a}}} < 1, \tag{4.36}$$

we have

$$\sum_{k=0}^{\infty} \mathbf{g}_k\left(\frac{\beta(\gamma,\varepsilon)}{\xi_k}\right) = +\infty, \tag{4.37}$$

and, from Theorem 4.4, we have $\mathbf{U} = l^*$ almost surely. □

4.2. Practical Implementation

The above results suggest the following numerical algorithm.

(1) An initial guess $\mathbf{X}^0 \in C$ is given.

(2) At the iteration number $k \geq 0$, $\mathbf{X}^k$ is known and $\mathbf{X}^{k+1}$ is determined by performing the following three substeps.

(2.1) Unperturbed ascent: we determine the descent direction $\mathbf{d}^k$ and the step η^k using ascent method (3.13). This generates the first trial point:

$$\mathbf{T}_{k+1}^0 = \mathbf{Q}_k\left(\mathbf{X}^k\right). \tag{4.38}$$

(2.2) Perturbation: we determine a sample $(\rho_k^1, \ldots, \rho_k^{k_{\text{sto}}})$ of k_{sto} new trial points:

$$\mathbf{T}_{k+1}^i = \mathbf{T}_{k+1}^0 + \rho_k^i, \quad i = 1, \ldots, k_{\text{sto}}. \tag{4.39}$$

(2.3) Dynamics: we determine $\mathbf{X}^{k+1}$ by selecting it from the set of available points:

$$\mathcal{A}_k = \left\{\mathbf{X}^k, \mathbf{T}_{k+1}^0, \ldots, \mathbf{T}_{k+1}^{k_{\text{sto}}}\right\}. \tag{4.40}$$

The computation of $\mathbf{X}^{k+1}$ is performed in two phases.

First phase: we determine a trial point (unperturbed ascent step and perturbation step).

Second phase: we determine $\mathbf{X}^{k+1}$ by selection it from $\mathcal{A}_k$ (dynamics step).

As was shown in Theorem 4.4, Substep (2.2) may use $\rho_k^i = \xi_k \mathbf{Z}_k$, where $\mathcal{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_{k+1})$ is a sample of $N(0,1)$ and ξ_k is given by (4.33).

For instance, we can consider elitistic dynamics:

$$\mathbf{X}^{k+1} = \arg\min_{\mathbf{X}\in\mathcal{A}_k} \mathbf{f}(\mathbf{X}). \tag{4.41}$$

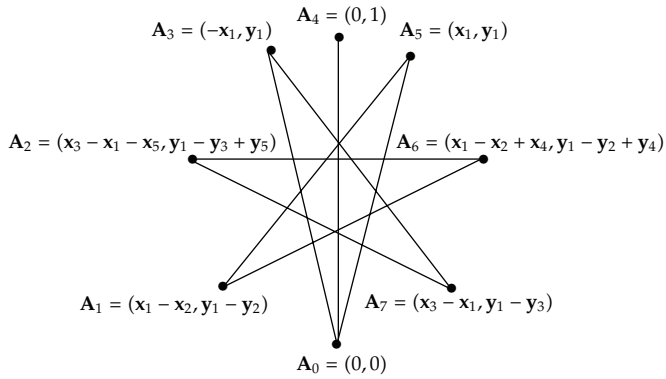

Figure 1: Definition of variables for the configuration conjectured by Graham to have maximum area.

5. Numerical Results

In order to apply the method presented in (4.41), we start at the initial value $\widehat{\mathbf{X}}^0 = \mathbf{X}^0 \in C$. At step $k \geq 0$, $\widehat{\mathbf{X}}^k$ is known and $\widehat{\mathbf{X}}^{k+1}$ is determined.

We generate k_{sto} the number of perturbation; the iterations are stopped when $\mathbf{X}^k$ is a Kuhn-Tucker point. We denote by k_{end} the value of k when the iterations are stopped (it corresponds to the number of evaluations of the gradient of $\mathbf{f}$). The optimal value and optimal point are $\mathbf{f}_{\text{opt}}$ and $\mathbf{X}^{\text{opt}}$, respectively. The perturbation is normally distributed and samples are generated by using the log-trigonometric generator and the standard random number generator of the FORTRAN library. We use $\xi_k = \sqrt{\widehat{\mathbf{a}}/\log(k+2)}$, where $\widehat{\mathbf{a}} > 0$.

Concern experiments performed on a workstation HP Intel(R) Celeron(R) M processor 1.30 GHz, 224 Mo RAM. The row cpu gives the mean CPU time in seconds for one run.

5.1. Octagon Problem

Consider polygon in the plane with 5 sides (5-gons for short) and unit diameter. Which of them have maximum area? (see e.g., [22]).

Graham's conjecture states that the optimal octagon can be illustrated as in Figure 1, in which a solid line between two vertices indicates that the distance between these points is one.

This question can be formulated to the quadratically constrained quadratic optimization problem defining this configuration that appears as follows:

$$\text{Maximize } \frac{1}{2}\{(\mathbf{x}_2 + \mathbf{x}_3 - 4\mathbf{x}_1)\mathbf{y}_1 + (3\mathbf{x}_1 - 2\mathbf{x}_3 + \mathbf{x}_5)\mathbf{y}_2 + (3\mathbf{x}_1 - 2\mathbf{x}_2 + \mathbf{x}_4)\mathbf{y}_3 + (\mathbf{x}_3 - 2\mathbf{x}_1)\mathbf{y}_4 + (\mathbf{x}_2 - 2\mathbf{x}_1)\mathbf{y}_5\} - \mathbf{x}_1$$

Subject to

$$\|\mathbf{A}_0 - \mathbf{A}_1\| \leq 1 : (\mathbf{x}_1 - \mathbf{x}_2)^2 + (\mathbf{y}_1 - \mathbf{y}_2)^2 + \mathbf{z}_1 = 1,$$

$$\|\mathbf{A}_0 - \mathbf{A}_2\| \leq 1 : (-\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{x}_5)^2 + (\mathbf{y}_1 - \mathbf{y}_3 + \mathbf{y}_5)^2 + \mathbf{z}_2 = 1,$$

$$
\begin{aligned}
&\|\mathbf{A}_0-\mathbf{A}_6\|\le 1 : (\mathbf{x}_1-\mathbf{x}_2+\mathbf{x}_4)^2+(\mathbf{y}_1-\mathbf{y}_2+\mathbf{y}_4)^2+\mathbf{z}_3=1,\\
&\|\mathbf{A}_0-\mathbf{A}_7\|\le 1 : (-\mathbf{x}_1+\mathbf{x}_3)^2+(\mathbf{y}_1-\mathbf{y}_3)^2+\mathbf{z}_4=1,\\
&\|\mathbf{A}_1-\mathbf{A}_2\|\le 1 : (2\mathbf{x}_1-\mathbf{x}_2-\mathbf{x}_3+\mathbf{x}_5)^2+(\mathbf{y}_2-\mathbf{y}_3+\mathbf{y}_5)^2+\mathbf{z}_5=1,\\
&\|\mathbf{A}_1-\mathbf{A}_3\|\le 1 : (2\mathbf{x}_1+\mathbf{x}_2)^2+\mathbf{y}_2^2+\mathbf{z}_6=1,\\
&\|\mathbf{A}_1-\mathbf{A}_4\|\le 1 : (\mathbf{x}_1-\mathbf{x}_2)^2+(\mathbf{y}_1-\mathbf{y}_2-1)^2+\mathbf{z}_7=1,\\
&\|\mathbf{A}_1-\mathbf{A}_7\|\le 1 : (2\mathbf{x}_1-\mathbf{x}_2-\mathbf{x}_3)^2+(-\mathbf{y}_2+\mathbf{y}_3)^2+\mathbf{z}_8=1,\\
&\|\mathbf{A}_2-\mathbf{A}_3\|\le 1 : (\mathbf{x}_3-\mathbf{x}_5)^2+(-\mathbf{y}_3+\mathbf{y}_5)^2+\mathbf{z}_9=1,\\
&\|\mathbf{A}_2-\mathbf{A}_4\|\le 1 : (-\mathbf{x}_1+\mathbf{x}_3-\mathbf{x}_5)^2+(\mathbf{y}_1-\mathbf{y}_3+\mathbf{y}_5-1)^2+\mathbf{z}_{10}=1,\\
&\|\mathbf{A}_2-\mathbf{A}_5\|\le 1 : (2\mathbf{x}_1+\mathbf{x}_3-\mathbf{x}_5)^2+(-\mathbf{y}_3+\mathbf{y}_5)^2+\mathbf{z}_{11}=1,\\
&\|\mathbf{A}_2-\mathbf{A}_6\|= 1 : (2\mathbf{x}_1-\mathbf{x}_2-\mathbf{x}_3+\mathbf{x}_4+\mathbf{x}_5)^2+(-\mathbf{y}_2+\mathbf{y}_3+\mathbf{y}_4-\mathbf{y}_5)^2=1,\\
&\|\mathbf{A}_3-\mathbf{A}_6\|\le 1 : (-2\mathbf{x}_1+\mathbf{x}_2-\mathbf{x}_4)^2+(\mathbf{y}_2-\mathbf{y}_4)^2+\mathbf{z}_{12}=1,\\
&\|\mathbf{A}_4-\mathbf{A}_6\|\le 1 : (\mathbf{x}_1-\mathbf{x}_2+\mathbf{x}_4)^2+(\mathbf{y}_1-\mathbf{y}_2+\mathbf{y}_4-1)^2+\mathbf{z}_{13}=1,\\
&\|\mathbf{A}_4-\mathbf{A}_7\|\le 1 : (\mathbf{x}_1-\mathbf{x}_3)^2+(1-\mathbf{y}_1+\mathbf{y}_3)^2+\mathbf{z}_{14}=1,\\
&\|\mathbf{A}_5-\mathbf{A}_6\|\le 1 : (\mathbf{x}_2-\mathbf{x}_4)^2+(\mathbf{y}_2-\mathbf{y}_4)^2+\mathbf{z}_{15}=1,\\
&\|\mathbf{A}_5-\mathbf{A}_7\|\le 1 : (2\mathbf{x}_1-\mathbf{x}_3)^2+\mathbf{y}_3^2+\mathbf{z}_{16}=1,\\
&\|\mathbf{A}_6-\mathbf{A}_1\|\le 1 : (2\mathbf{x}_1-\mathbf{x}_2-\mathbf{x}_3+\mathbf{x}_4)^2+(-\mathbf{y}_2+\mathbf{y}_3+\mathbf{y}_4)^2+\mathbf{z}_{17}=1,\\
&\qquad \mathbf{x}_2-\mathbf{x}_3-\mathbf{z}_{18}=0,\\
&\qquad \mathbf{x}_i^2+\mathbf{y}_i^2=1,\quad i=1,2,3,4,5,\\
&\qquad \mathbf{x}_1+\mathbf{z}_{19}=0.5,\\
&\qquad \mathbf{x}_i+\mathbf{z}_{18+i}=1,\quad i=2,3,4,5,\\
&\qquad 0\le \mathbf{x}_i,\mathbf{y}_i\quad i=1,2,\ldots,5,\quad 0\le \mathbf{z}_i\quad i=1,2,\ldots,23.
\end{aligned}
\tag{5.1}
$$

There are 33 variables and 34 constraints.

We use $\widehat{\mathbf{a}} = 10d+6$ and $k_{\text{sto}} = 15000$ and used regular octagon as a starting point (area of regular octagon ≈ 0.7071 see Figure 2), and if we let the error less than 10^{-5}.

The Fortran code of TPGRG furnishs the following optimal solutions, $f_{\text{opt}} \approx 0.72687$:

$$
\begin{aligned}
&\mathbf{X}^{\text{opt}} = (0.25949, 0.66664, 0.66664, 0.90703, 0.90730),\\
&\mathbf{y}_i^{\text{opt}} = \sqrt{1-\mathbf{x}_i^{\text{opt}}},\quad k_{\text{end}} = 626,\quad \text{CPU time} = 4684\,\text{s (1.5 hours)}.
\end{aligned}
\tag{5.2}
$$

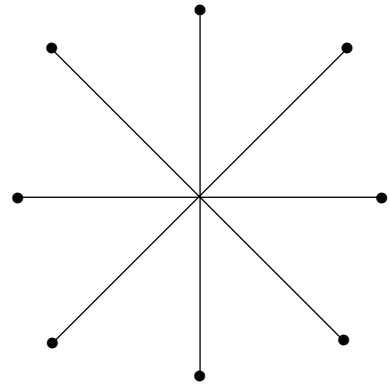

Figure 2: Regular octagon.

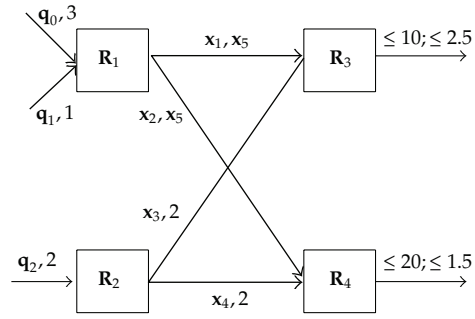

Figure 3: A Problem of mixture.

Branch and cut method [23] was solving the octagon problem with using regular octagon as a starting point, the maximum area found is ≈ 0.72687, but the *cpu* time is more than 30 hours.

5.2. Mixture Problem

In this example of petrochemical mixture, we have four reservoirs: $\mathbf{R}_1$, $\mathbf{R}_2$, $\mathbf{R}_3$, and $\mathbf{R}_4$. The two first receive three distinct source products, and then their content is combined in the two other in order to create the wanted miscellanies. The question is to determine the quantity of every product to buy in order to maximize the profits (see e.g., [24, 25]).

The $\mathbf{R}_1$ reservoir receives two quality products 3 and 1 in quantities $\mathbf{q}_0$ and $\mathbf{q}_1$, respectively. The $\mathbf{R}_2$ reservoir contains a product of the third source, of quality 2 in quantity $\mathbf{q}_2$. One wants to get in the reservoirs $\mathbf{R}_3$ and $\mathbf{R}_4$ of capacities 10 and 20, respectively.

Figure 3, where the variable $\mathbf{x}_1$ to $\mathbf{x}_2$ represents some quantities, illustrates this situation. The unit prices of the products bought are, respectively, 60, 160, and 100; those of the products finished are 90 and 150. Therefore the difference between the costs of purchase

and sale is

$$60\mathbf{q}_0 + 160\mathbf{q}_1 + 100\mathbf{q}_2 - 90(\mathbf{x}_1 + \mathbf{x}_3) - 150(\mathbf{x}_2 - \mathbf{x}_4). \tag{5.3}$$

The quality of the mixture done contained in $\mathbf{R}_1$ is

$$\mathbf{x}_5 = \frac{3\mathbf{q}_0 + \mathbf{q}_1}{\mathbf{q}_0 + \mathbf{q}_1} \in [1, 3]. \tag{5.4}$$

The qualities of the final miscellanies are

$$\frac{\mathbf{x}_1\mathbf{x}_5 + 2\mathbf{x}_3}{\mathbf{x}_1 + \mathbf{x}_3} \leq 2.5, \qquad \frac{\mathbf{x}_2\mathbf{x}_5 + 2\mathbf{x}_4}{\mathbf{x}_2 + \mathbf{x}_4} \leq 1.5. \tag{5.5}$$

The addition of the constraints of volume conservation $\mathbf{q}_0 + \mathbf{q}_1 = \mathbf{x}_1 + \mathbf{x}_2$ and $\mathbf{q}_2 = \mathbf{x}_3 + \mathbf{x}_4$ does permit the elimination of the variables $\mathbf{q}_0, \mathbf{q}_1$, and $\mathbf{q}_2$:

$$\mathbf{q}_0 = \frac{(\mathbf{x}_5 - 1)(\mathbf{x}_1 + \mathbf{x}_2)}{2}, \quad \mathbf{q}_1 = \frac{(3 - \mathbf{x}_5)(\mathbf{x}_1 + \mathbf{x}_2)}{2}, \quad \mathbf{q}_2 = \mathbf{x}_3 + \mathbf{x}_4. \tag{5.6}$$

Remark 5.1. One has

$$1 \leq \mathbf{x}_5 \leq 1.5. \tag{5.7}$$

By the relation (5.5),

$$\mathbf{x}_2\mathbf{x}_5 \leq 1.5\mathbf{x}_2. \tag{5.8}$$

Suppose that $\mathbf{x}_5 > 1.5$, then $\mathbf{x}_2\mathbf{x}_5 > 1.5\mathbf{x}_2$, and so $\mathbf{x}_2\mathbf{x}_5 > 1.5\mathbf{x}_2 - 0.5\mathbf{x}_4$ that contradict with inequality (5.8).

The mathematical model transformed is the following:

$$\begin{aligned}
&\text{Maximize} - 120\mathbf{x}_1 - 60\mathbf{x}_2 - 10\mathbf{x}_3 + 50\mathbf{x}_4 + 50\mathbf{x}_1\mathbf{x}_5 + 50\mathbf{x}_2\mathbf{x}_5\\
&\text{Subject to}\\
&\qquad 2.5\mathbf{x}_1 + 0.5\mathbf{x}_3 - \mathbf{x}_1\mathbf{x}_5 - \mathbf{z}_1 = 0,\\
&\qquad 1.5\mathbf{x}_2 - 0.5\mathbf{x}_4 - \mathbf{x}_2\mathbf{x}_5 - \mathbf{z}_2 = 0,\\
&\qquad \mathbf{x}_1 + \mathbf{x}_3 + \mathbf{z}_3 = 10,\\
&\qquad \mathbf{x}_2 + \mathbf{x}_4 + \mathbf{z}_4 = 20,\\
&\qquad \mathbf{x}_5 + \mathbf{z}_5 = 1.5,\\
&\qquad \mathbf{x}_5 - \mathbf{z}_6 = 1,\\
&\qquad 0 \le \mathbf{x}_i \quad i = 1, \ldots, 5; \quad 0 \le \mathbf{z}_i \quad i = 1, \ldots, 6.
\end{aligned} \tag{5.9}$$

There are 11 variables and 10 constraints.

We use $\widehat{\mathbf{a}} = 10d10$, $k_{\text{sto}} = 20000$, and

$$\mathbf{X}^0 = (2, 9, 0, 8, 1), \qquad \mathbf{f}\left(\mathbf{X}^0\right) = 170. \tag{5.10}$$

The Fortran code of TPGRG furnishes the following optimal solutions: $cpu = 100$ second, and $k_{\text{end}} = 1390$,

$$\mathbf{X}_{\text{opt}} = (0, 10, 0, 10, 1), \qquad \mathbf{f}_{\text{opt}} = 400, \tag{5.11}$$

since the mixture problem is known by this global solution.

6. Concluding Remarks

A two-phase generalized reduced gradient method is presented for nonlinear constraints, involving the adjunction of a stochastic perturbation. This approach leads to a stochastic ascent method where the deterministic sequence generated by the two-phase generalized gradient method is replaced by a sequence of random variables.

TPGRG method converges to global maximum for all differential objective function, but GRG method converges to local maximum.

The numerical experiments show that the method is effective to calculate for global optimization problems. Here yet, we observe that the adjunction of the stochastic perturbation improves the result, with a larger number of evaluations of the objective function. The main difficulty in the practical use of the stochastic perturbation is connected to the tuning of the parameters $\widehat{\mathbf{a}}$ and k_{sto}.

References

[1] A. El Mouatasim and A. Al-Hossain, "Reduced gradient method for minimax estimation of a bounded Poisson mean," *Journal of Statistics: Advances in Theory and Applications*, vol. 2, no. 2, pp. 183–197, 2009.

[2] J. Mo, K. Zhang, and Z. Wei, "A variant of SQP method for inequality constrained optimization and its global convergence," *Journal of Computational and Applied Mathematics*, vol. 197, no. 1, pp. 270–281, 2006.

[3] H. Jiao, Y. Guo, and P. Shen, "Global optimization of generalized linear fractional programming with nonlinear constraints," *Applied Mathematics and Computation*, vol. 183, no. 2, pp. 717–728, 2006.

[4] Y. C. Park, M. H. Chang, and T.-Y. Lee, "A new deterministic global optimization method for general twice-differentiable constrained nonlinear programming problems," *Engineering Optimization*, vol. 39, no. 4, pp. 397–411, 2007.

[5] A. A. Haggag, "A variant of the generalized reduced gradient algorithm for nonlinear programming and its applications," *European Journal of Operational Research*, vol. 7, no. 2, pp. 161–168, 1981.

[6] C. C. Y. Dorea, "Stopping rules for a random optimization method," *SIAM Journal on Control and Optimization*, vol. 28, no. 4, pp. 841–850, 1990.

[7] A. El Mouatasim, R. Ellaia, and J. E. Souza de Cursi, "Random perturbation of the variable metric method for unconstrained nonsmooth nonconvex optimization," *International Journal of Applied Mathematics and Computer Science*, vol. 16, no. 4, pp. 463–474, 2006.

[8] J. E. Souza de Cursi, R. Ellaia, and M. Bouhadi, "Global optimization under nonlinear restrictions by using stochastic perturbations of the projected gradient," in *Frontiers in Global Optimization*, vol. 74 of *Nonconvex Optimization and Its Applications*, pp. 541–563, Kluwer Academic Publishers, Boston, Mass, USA, 2004.

[9] P. Sadegh and J. C. Spall, "Optimal random perturbations for stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Transactions on Automatic Control*, vol. 43, no. 10, pp. 1480–1484, 1998.

[10] W. L. Price, "Global optimization by controlled random search," *Journal of Optimization Theory and Applications*, vol. 40, no. 3, pp. 333–348, 1983.

[11] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Mass, USA, 1989.

[12] A. Corana, M. Marchesi, C. Martini, and S. Ridella, "Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm," *ACM Transactions on Mathematical Software*, vol. 13, no. 3, pp. 262–280, 1987.

[13] M. Pogu and J. E. Souza de Cursi, "Global optimization by random perturbation of the gradient method with a fixed parameter," *Journal of Global Optimization*, vol. 5, no. 2, pp. 159–180, 1994.

[14] J. E. Souza de Cursi, R. Ellaia, and M. Bouhadi, "Stochastic perturbation methods for affine restrictions," in *Advances in Convex Analysis and Global Optimization*, vol. 54 of *Nonconvex Optimization and Its Applications*, pp. 487–499, Kluwer Academic Publishers, Boston, Mass, USA, 2001.

[15] P. Beck, L. Lasdon, and M. Engquist, "A reduced gradient algorithm for nonlinear network problems," *ACM Transactions on Mathematical Software*, vol. 9, no. 1, pp. 57–70, 1983.

[16] J. Abadie, "The GRG method for non-linear programming," in *Design and Implementation of Optimization Software*, H. J. Greenberg, Ed., Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands, 1978.

[17] D. Gabay and D. G. Luenberger, "Efficiently converging minimization methods based on the reduced gradient," *SIAM Journal on Control and Optimization*, vol. 14, no. 1, pp. 42–61, 1976.

[18] E. P. de Carvalho, A. dos Santos Jr., and T. F. Ma, "Reduced gradient method combined with augmented Lagrangian and barrier for the optimal power flow problem," *Applied Mathematics and Computation*, vol. 200, no. 2, pp. 529–536, 2008.

[19] H. Mokhtar-Kharroubi, "Sur la convergence théorique de la méthode du gradient réduit généralisé," *Numerische Mathematik*, vol. 34, no. 1, pp. 73–85, 1980.

[20] Y. Smeers, "Generalized reduced gradient method as an extension of feasible direction methods," *Journal of Optimization Theory and Applications*, vol. 22, no. 2, pp. 209–226, 1977.

[21] P. L'Ecuyer and R. Touzin, "On the Deng-Lin random number generators and related methods," *Statistics and Computing*, vol. 14, no. 1, pp. 5–9, 2004.

[22] R. L. Graham, "The largest small hexagon," *Journal of Combinatorial Theory. Series A*, vol. 18, pp. 165–170, 1975.

[23] C. Audet, P. Hansen, F. Messine, and J. Xiong, "The largest small octagon," *Journal of Combinatorial Theory. Series A*, vol. 98, no. 1, pp. 46–59, 2002.

[24] L. R. Foulds, D. Haugland, and K. Jörnsten, "A bilinear approach to the pooling problem," *Optimization*, vol. 24, no. 1-2, pp. 165–180, 1992.

[25] C. A. Haverly, "Studies of the behaviour of recursion for the pooling problem," *ACM SIGMAP Bulletin*, vol. 25, pp. 19–28, 1996.