AN OVERVIEW OF NONLINEAR OPTIMIZATION

YU-HONG DAI

ABSTRACT

Nonlinear optimization stems from calculus and becomes an independent subject due to the proposition of Karush–Kuhn–Tucker optimality conditions. The ever-growing realm of applications and the explosion in computing power is driving nonlinear optimization research in new and exciting directions. In this article, I shall give a brief overview of nonlinear optimization, mainly on unconstrained optimization, constrained optimization, and optimization with least constraint violation.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 90C30; Secondary 65K05, 90C06

KEYWORDS

unconstrained optimization, constrained optimization, optimization with least constraint violation, gradient method, conjugate gradient method, sequential quadratic programming method, interior-point method, augmented Lagrangian method of multipliers

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 7, pp. 5290–5313
DOI 10.4171/ICM2022/72

1. INTRODUCTION

It is known that nonlinear optimization stems from calculus. Consider the unconstrained optimization problem

$$\min_{x\in\Re^n} \quad f(x), \tag{1.1}$$

where $f:\Re^n \to \Re$ is smooth and its gradient g is available. The calculus invented by Newton and Leibniz in the seventeenth century provided a necessary condition for a point x to be the optimal solution of (1.1), which is $\nabla f(x) = 0$, i.e., the tangent line of f at x is horizontal. For equality constrained optimization, the necessary optimality condition is that the derivatives of the Lagrangian function with respect to the primal and dual variables are equal to zero, which was exposed by Lagrange in the eighteenth century. Nonlinear optimization became an independent subject when Karush [52] and Kuhn and Tucker [53] provided necessary optimality conditions for general optimization subjected to equality and inequality constraints,

$$\begin{aligned} \min \quad & f(x) && (1.2)\\ \text{such that} \quad & h(x) = 0, && (1.3)\\ & g(x) \ge 0, && (1.4) \end{aligned}$$

where $f:\Re^n \to \Re$, $h:\Re^n \to \Re^{m_E}$, $g:\Re^n \to \Re^{m_I}$ are supposed to be twice continuously differentiable functions. The proposition of the Fletcher–Reeves conjugate gradient method [39] and the Davidon–Fletcher–Powell quasi-Newton method [33, 38] greatly promoted the development of nonlinear optimization.

This article shall give a brief overview of nonlinear optimization, mainly on unconstrained optimization, constrained optimization, and optimization with least constraint violation.

2. UNCONSTRAINED OPTIMIZATION

The design and analysis of numerical methods for unconstrained optimization is closely related to the unconstrained quadratic optimization

$$\min_{x\in\Re^n} \quad q(x) := \frac{1}{2}x^T Ax - b^T x, \tag{2.1}$$

where $b \in \Re^n$ and $A \in \Re^{n\times n}$ is symmetric and positive definite with eigenvalues $0 < \lambda_1 \le \cdots \le \lambda_n$. Fundamental methods for unconstrained optimization include gradient methods, conjugate gradient methods, quasi-Newton methods, Newton method, and derivative-free methods. We focus on two classes of first-order methods, gradient methods, and conjugate gradient methods, which are suitable for large-scale problems.

The gradient method can be dated back to Cauchy [9], and the first nonlinear conjugate gradient method is due to Fletcher and Reeves [39]. Driven by practical applications, various variants of the methods have been proposed for convex optimization, nonsmooth optimization, stochastic optimization, etc. For smooth optimization, the two classes of methods are significantly improved by asking their search directions to be close to the Newton or quasi-Newton direction in some sense. Typical examples of the conjugate gradient method are the Dai–Yuan method [27], the Hager–Zhang method [46], and the Dai–Kou method [22]. For the gradient method, one milestone work is the Barzilai–Borwein (nonmonotone) gradient method, while another significant work is the Yuan stepsize [87], which leads to the proposition of the efficient Dai–Yuan (monotone) gradient method [30]. Interestingly enough, Huang et al. [51] found that it is possible to equip the Barzilai–Borwein method with the two-dimensional quadratic termination property.

2.1. Gradient methods

Gradient methods search along the negative gradient and are of the form

$$x_{k+1} = x_k - \alpha_k g_k, \tag{2.2}$$

where $g_k = \nabla f(x_k)$ and $\alpha_k > 0$ is the stepsize. Different choices of the stepsize α_k lead to different gradient methods. The steepest descent (SD) method, which is due to Cauchy [9], determines its stepsize by the exact line search, i.e.,

$$\alpha_k^{\mathrm{SD}} = \arg\min_{\alpha>0} f(x_k - \alpha g_k). \tag{2.3}$$

The SD method is shown to be Q-linearly convergent, but its performance is poor when the problem is ill-conditioned [1]. Specifically, the SD method will asymptotically tend to minimize the function in some two-dimensional subspace and produce zigzags [40]. If the dimension is greater than one, the SD stepsize (2.3) always tends to be long, and some shortened SD methods are proposed in [30].

One milestone work on the gradient method is due to Barzilai and Borwein [3]. Its basic idea is to ask the matrix $\alpha_k^{-1} I$ or $\alpha_k I$ have a certain quasi-Newton property. Then by minimizing $\|s_{k-1} - (\alpha_k^{-1} I) y_{k-1}\|$ or $\|(\alpha_k I) s_{k-1} - y_{k-1}\|$ with respect to α_k, where $s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = g_k - g_{k-1}$ and $\|\cdot\|$ is the two-norm, two stepsizes are derived as

$$\alpha_k^{\mathrm{BB}\,1} = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}, \quad \alpha_k^{\mathrm{BB}\,2} = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}. \tag{2.4}$$

The stepsizes $\alpha_k^{\mathrm{BB}\,1}$ and $\alpha_k^{\mathrm{BB}\,2}$ are called long and short Barzilai–Borwein (BB) stepsizes, respectively, since $\alpha_k^{\mathrm{BB}\,1} \geq \alpha_k^{\mathrm{BB}\,2}$ if $s_{k-1}^T y_{k-1} > 0$. Despite its heavy nonmonotone behavior, the BB method performs significantly better than the SD method in practice; see, e.g., [36]. For unconstrained quadratic optimization, the BB method is proved to be R-superlinearly convergent if the dimension is two [3]. For general dimension, the BB method is globally convergent [73] and the convergence is R-linear [25]. An efficient extension of the BB method for unconstrained optimization is given in [74] by incorporating the Grippo–Lampariello–Lucidi (GLL) nonmonotone line search [45]. Interestingly enough, it is shown in [25] that the BB stepsize can asymptotically be accepted by the GLL nonmonotone line search when the iterate is close to the solution. This property is similar to the fact that the unit stepsize can asymptotically be accepted by Newton or quasi-Newton methods using the Armijo or Wolfe line search. Furthermore, efficient projected gradient methods based on BB-like methods and applications can be found in [4,20,63,90], among many other references. The numerical efficiency of the BB method over the SD method has stimulated many studies on the gradient method.

However, it is intriguing to provide theoretical evidence showing that the BB method performs much better than the SD method for high-dimensional problems. One possible angle is to relate the stepsize in the gradient method to the eigenvalues of the Hessian of the function. To this aim, consider the unconstrained quadratic optimization problem (2.1). In this case, by (2.1) and (2.2), we have that $g_{k+1} = (I - \alpha_k A) g_k$ for all $k \geq 1$. Then we see that the gradient method with constant stepsizes (i.e., $\alpha_k \equiv \alpha$ for some $\alpha > 0$) is equivalent to the shifted power method for computing some eigenvalue of the matrix A since

$$\frac{g_{k+1}}{\|g_{k+1}\|} = \frac{(I - \alpha A)^k g_1}{\|(I - \alpha A)^k g_1\|} = \frac{(A - \alpha^{-1} I)^k g_1}{\|(A - \alpha^{-1} I)^k g_1\|}. \tag{2.5}$$

For example, if $\alpha_k \equiv \frac{1}{2L}$, where L is the gradient Lipschitz constant, which was one choice in the early times [2], $\frac{g_{k+1}}{\|g_{k+1}\|}$ will tend to the eigenvector corresponding to the minimal eigenvalue of A provided the initial gradient g_1 has a nonzero component in this eigenvector. Another fact is the quadratic termination property of the gradient method, which was exposed by Lai [54]. To see this, notice that $g_{k+1} = \prod_{j=1}^{k} (I - \alpha_j A) g_1$ for $k \geq 1$. Then by the Hamilton–Caley theorem, we have that g_{n+1} vanishes if the set of stepsizes $\{\alpha_i : 1 \leq i \leq n\}$ coincides with the set of inverse eigenvalues of A, $\{\lambda_i^{-1} : 1 \leq i \leq n\}$. A natural corollary is as follows.

Lemma 2.1. *Consider the gradient method* (2.2) *for the unconstrained quadratic optimization problem* (2.1). *Assume that the initial gradient g_1 has nonzero components in all eigenvectors of the matrix A. If the gradient method is R-superlinearly convergent, then, for each eigenvalue λ_i $(1 \leq i \leq n)$ of A, there exists a subsequence $\{\alpha_{k_i}\}$ such that* $\lim_{k_i \to \infty} \alpha_{k_i} = \lambda_i^{-1}$.

The above lemma provides us an insight about convergence properties of gradient methods. From the proof of the R-superlinear convergence of the BB method in the two-dimensional setting [3], it is easy to see that there do exist subsequences $\{\alpha_{k_1}\}$ and $\{\alpha_{k_2}\}$ such that they converge to the two inverse eigenvalues of the Hessian, respectively. Dai and Fletcher [19] observed this phenomenon for the BB method in the three-dimensional setting as well and showed that the BB method is likely to be R-superlinearly convergent in this case. It is also shown in [19] that the cyclic steepest descent method is likely to be R-superlinearly convergent for n-dimensional convex quadratic functions provided that $m \geq \frac{n+1}{2}$, where m is the cyclic time of the steepest descent stepsize.

Another significant addition to the gradient method is the Yuan stepsize [87], which is such that, if the previous and later steps use SD stepsizes, the gradient method can give the exact minimizer of a two-dimensional convex quadratic function. A variant of the Yuan stepsize is given by Dai and Yuan [30] as

$$\alpha_k^{\mathrm{DY}} = \frac{2}{\frac{1}{\alpha_{k-1}^{\mathrm{SD}}} + \frac{1}{\alpha_k^{\mathrm{SD}}} + \sqrt{\left(\frac{1}{\alpha_{k-1}^{\mathrm{SD}}} - \frac{1}{\alpha_k^{\mathrm{SD}}}\right)^2 + \frac{4\|g_k\|^2}{(\alpha_{k-1}^{\mathrm{SD}} \|g_{k-1}\|)^2}}}. \tag{2.6}$$

They also suggested the so-called Dai–Yuan gradient method (2.2) with

$$\alpha_k = \begin{cases} \alpha_k^{\mathrm{SD}}, & \text{if } \mathrm{mod}(k,4) = 0, 1, \\ \alpha_k^{\mathrm{DY}}, & \text{if } \mathrm{mod}(k,4) = 2, 3. \end{cases} \tag{2.7}$$

The Dai–Yuan gradient method is monotone since $\alpha_k^{\mathrm{DY}} \le \alpha_k^{\mathrm{SD}}$. This is the first monotone gradient method which can beat the BB nonmonotone gradient method for unconstrained quadratic optimization.

A recent progress in the gradient method is provided by Huang et al. [51], who introduced a new mechanism for the gradient method to achieve the two-dimensional quadratic termination property. Given $\nu_1(k), \nu_2(k) \in \{1, \dots, k\}$ and some suitable functions ψ_1, ψ_2, ψ_3, ψ_4 satisfying $\psi_1(A)\psi_2(A) = \psi_3(A)\psi_4(A)$, they suggested calculating the stepsize α_k by solving the following quadratic equation:

$$\begin{aligned} & g_{\nu_1(k)}^T \psi_1(A)(I - \alpha_k A) g_k \cdot g_{\nu_2(k)}^T \psi_2(A)(I - \alpha_k A) g_k \\ & \quad = g_{\nu_1(k)}^T \psi_3(A)(I - \alpha_k A) g_k \cdot g_{\nu_2(k)}^T \psi_4(A)(I - \alpha_k A) g_k, \end{aligned} \tag{2.8}$$

and proved that the gradient method using any stepsize obtained from (2.8) and α_{k+2} in the form of $\frac{(A^\mu g_{k+i})^T (A^\mu g_{k+i})}{(A^\mu g_{k+i})^T A (A^\mu g_{k+i})}$ with $i = 1$ or 2 and μ being some real number achieves the two-dimensional quadratic termination property. Interestingly, the stepsize α_k^{DY} in (2.6) is a solution of equation (2.8) corresponding to $\nu_1(k) = k - 1$, $\nu_2(k) = k$, $\psi_1(A) = \psi_4(A) = (I - \alpha_{k-1} A)^{-1}$, and $\psi_2(A) = \psi_3(A) = I$.

To equip the BB method with the two-dimensional quadratic termination property, Huang et al. [51] chose $\nu_1(k) = k - 2$, $\nu_2(k) = k - 1$, $\psi_1(A) = (I - \alpha_{k-2} A)^{-1}$, $\psi_2(A) = (I - \alpha_{k-1} A)^{-1}$, $\psi_3(A) = \psi_1(A)\psi_2(A)$, and $\psi_4(A) = I$. Then by (2.8), they obtained the following novel stepsize:

$$\alpha_k^{\mathrm{HDL}} = \frac{2}{\frac{\phi_2}{\phi_3} + \sqrt{\left(\frac{\phi_2}{\phi_3}\right)^2 - 4\frac{\phi_1}{\phi_3}}}, \tag{2.9}$$

where

$$\frac{\phi_1}{\phi_3} = \frac{\alpha_{k-1}^{\mathrm{BB\,2}} - \alpha_k^{\mathrm{BB\,2}}}{\alpha_{k-1}^{\mathrm{BB\,2}} \alpha_k^{\mathrm{BB\,2}} (\alpha_{k-1}^{\mathrm{BB\,1}} - \alpha_k^{\mathrm{BB\,1}})}, \quad \frac{\phi_2}{\phi_3} = \frac{\alpha_{k-1}^{\mathrm{BB\,1}} \alpha_{k-1}^{\mathrm{BB\,2}} - \alpha_k^{\mathrm{BB\,1}} \alpha_k^{\mathrm{BB\,2}}}{\alpha_{k-1}^{\mathrm{BB\,2}} \alpha_k^{\mathrm{BB\,2}} (\alpha_{k-1}^{\mathrm{BB\,1}} - \alpha_k^{\mathrm{BB\,1}})}. \tag{2.10}$$

It is observed in [51] that the use of the stepsize α_k^{HDL} can make both the BB1 and BB2 methods achieve the two-dimensional quadratic termination property. The computation of α_k^{HDL} only involves the BB stepsizes in the previous two iterations and does not require exact line searches or the Hessian computation. Hence it can easily be extended for nonlinear optimization.

Based on the new stepsize α_k^{HDL} and the general framework in [92], an efficient gradient method for solving unconstrained optimization problem (2.1) is suggested in [51]. In particular, the method uses $\alpha_1 = \alpha_1^{\mathrm{SD}}$, $\alpha_2 = \alpha_2^{\mathrm{BB\,1}}$, and, for $k \ge 3$,

$$\alpha_k = \begin{cases} \min\{\alpha_{k-1}^{\mathrm{BB\,2}}, \alpha_k^{\mathrm{BB\,2}}, \alpha_k^{\mathrm{HDL}}\}, & \text{if } \alpha_k^{\mathrm{BB\,2}}/\alpha_k^{\mathrm{BB\,1}} < \tau_k, \\ \alpha_k^{\mathrm{BB\,1}}, & \text{otherwise}, \end{cases} \tag{2.11}$$

where $\tau_k > 0$ is chosen in some way. The method (2.11) appears to be much better than BB, Dai–Yuan, and some other recent gradient methods. With the projection technique, the method (2.11) was also extended in [51] to unconstrained optimization, box-constrained optimization, and singly linearly box-constrained optimization, and good numerical results were obtained.

There are still many questions about the gradient method to be investigated. Is it possible to provide more theoretical evidence showing the efficiency of the BB method for high-dimensional functions? What is the best choice of the stepsize in the gradient method? How to extend the existing efficient gradient methods to many other areas?

2.2. Conjugate gradient methods

Conjugate gradient methods are a class of important methods for solving (1.1). They are of the form

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2.12}$$

where α_k is the stepsize obtained by a line search and d_k is the search direction given by

$$d_k = -g_k + \beta_k d_{k-1}, \tag{2.13}$$

except for $d_1 = -g_1$. The scalar β_k is the so-called conjugate gradient parameter such that the method (2.12)–(2.13) reduces to the linear conjugate gradient method if the objective function is quadratic and the line search is exact.

For nonlinear functions, however, different formulae for the parameter β_k result in different conjugate gradient methods and their properties can be significantly different. To distinguish the linear conjugate gradient method, sometimes we call the conjugate gradient method for unconstrained optimization as the nonlinear conjugate gradient method. The work of Fletcher and Reeves [39] not only opened the door to the nonlinear conjugate gradient field but also greatly stimulated the study of nonlinear optimization. Four well-known formulae for β_k are called the Fletcher–Reeves (FR) [39], Dai–Yuan (DY) [27], Polak–Ribière–Polyak (PRP) [67,68], and Hestenes–Stiefel (HS) [50], which are given by

$$\begin{aligned} \beta_k^{\mathrm{FR}} &= \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad & \beta_k^{\mathrm{DY}} &= \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \\ \beta_k^{\mathrm{PRP}} &= \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \quad & \beta_k^{\mathrm{HS}} &= \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \end{aligned} \tag{2.14}$$

respectively, where $y_{k-1} = g_k - g_{k-1}$ as before.

Since the exact line search is usually expensive and impractical, the strong Wolfe line search is often considered in the early convergence analysis and numerical implementation for nonlinear conjugate gradient methods. The strong Wolfe line search aims to find a stepsize $\alpha_k > 0$ satisfying

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k, \tag{2.15}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \le -\sigma g_k^T d_k, \tag{2.16}$$

where $0 < \delta < \sigma < 1$. However, it was shown in [28] that even with strong Wolfe line searches, none of the FR, PRP, and HS methods can ensure the descent property of the search direction if the parameter σ is not properly chosen. If a descent search direction is not produced, a practical remedy is to restart the method along $-g_k$. This may degrade the efficiency of the method since the second-derivative information achieved along the previous search direction is discarded.

It is known that quasi-Newton methods often use the standard Wolfe line search, which aims to find a stepsize $\alpha_k > 0$ satisfying (2.15) and

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{2.17}$$

where $0 < \delta < \sigma < 1$. Dai and Yuan [27] were able to establish the descent property and global convergence of the DY method with the standard Wolfe line search under weak assumptions on the objective function.

Assumption 2.1. (i) The level set $\mathcal{L} = \{x \in \Re^n : f(x) \leq f(x_1)\}$ is bounded, where x_1 is the starting point; (ii) f is continuously differentiable in some neighborhood of $\mathcal{L}$ and its gradient is Lipschitz continuous.

Theorem 2.1 ([27]). *Suppose that f satisfies Assumption 2.1. Consider the sequence $\{x_k\}$ generated by the DY method (2.12)–(2.13) with $\beta_k = \beta_k^{\text{DY}}$ and the standard Wolfe line search (2.15) and (2.17). Assume that $\|g_k\| \neq 0$ for all k. Then we have that $g_k^T d_k < 0$ for all k. Furthermore, the DY method converges in the sense that* $\liminf_{k\to+\infty} \|g_k\| = 0$.

It is noted in [27] that the DY formula can be rewritten as $\beta_k^{\text{DY}} = \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$. It is remarkable that the DY method has a certain self-adjusting property that is independent of the line search and the function convexity. The DY direction can also be used to restart optimization methods while guaranteeing the global convergence of the method (see [28]). Interestingly enough, Dai [16] provided another nonlinear conjugate gradient method which can ensure the descent property of the search direction without any line searches.

The following theorems provide general convergence results for nonlinear conjugate gradient methods with the strong Wolfe line search and the standard Wolfe line search, respectively. The results are very useful in the convergence analysis of various nonlinear conjugate gradient methods.

Theorem 2.2 ([21]). *Suppose that Assumption 2.1 holds. Consider any method of the form (2.12)–(2.13) with d_k satisfying $g_k^T d_k < 0$ and with the strong Wolfe line search (2.15) and (2.16). Then the method is globally convergent in the sense that* $\liminf_{k\to+\infty} \|g_k\| = 0$ *if $\sum_{k\geq 1} \|d_k\|^{-2} = +\infty$.*

Theorem 2.3 ([17]). *Suppose that Assumption 2.1 holds. Consider any method of the form (2.12)–(2.13) with d_k satisfying $g_k^T d_k < 0$ and with the standard Wolfe line search (2.15) and (2.17). Then the method is globally convergent in the sense that* $\liminf_{k\to+\infty} \|g_k\| = 0$ *if the scalar β_k is such that $\sum_{k\geq 1} \prod_{j=2}^{k} \beta_j^{-2} = +\infty$.*

Powell [71] found that the PRP method can automatically generate a search direction close to the steepest descent direction once a small step occurs, whereas the FR method may produce many tiny steps continuously. This explains why the PRP method sometimes performs much better than the FR method in practice. Nevertheless, Powell [71] showed that, even with exact line searches, the PRP method can cycle indefinitely without approaching a stationary point. To change this unbalanced state, Touati-Ahmed and Storey [79] proposed the hybrid conjugate gradient method, where

$$\beta_k^{\mathrm{FRPRP}} = \max\{0, \min\{\beta_k^{\mathrm{PRP}}, \beta_k^{\mathrm{FR}}\}\}. \tag{2.18}$$

Gilbert and Nocedal [42] modified the PRP method by setting

$$\beta_k^{\mathrm{PRP+}} = \max\left\{\frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, 0\right\}. \tag{2.19}$$

They established the global convergence results for the FRPRP and PRP+ methods, but found that the two methods are not significantly more efficient than the PRP method itself. Nevertheless, Dai and Yuan [29] were able to extend the convergence theorem of the DY method, Theorem 2.1, to the following hybrid conjugate gradient method:

$$\beta_k^{\mathrm{DYHS}} = \max\{0, \min\{\beta_k^{\mathrm{HS}}, \beta_k^{\mathrm{DY}}\}\}, \tag{2.20}$$

and found that the DYHS method with the standard Wolfe line search performs much better than the PRP method using the strong Wolfe line search.

Since $y_{k-1} = As_{k-1} = \alpha_{k-1}Ad_{k-1}$ in case of unconstrained quadratic optimization, an equivalent expression of the conjugacy condition $d_k^T Ad_{k-1} = 0$ is $d_k^T y_{k-1} = 0$. For general functions, however, we have for quasi-Newton methods that $d_k = -B_k^{-1} g_k$, where the approximation matrix B_k satisfies the quasi-Newton equation $B_k s_{k-1} = y_{k-1}$. This hints us at the nonlinear conjugate gradient condition $d_k^T y_{k-1} = (-B_k^{-1} g_k)^T (B_k s_{k-1}) = -g_k^T s_{k-1}$. By introducing a scaling factor t, Dai and Liao [24] considered a nonlinear conjugacy condition $d_k^T y_{k-1} = -t g_k^T s_{k-1}$ and proposed the following family for conjugate gradient methods:

$$\beta_k^{\mathrm{DL}}(t) = \frac{g_k y_{k-1}}{d_{k-1}^T y_{k-1}} - t \frac{g_k s_{k-1}}{d_{k-1}^T y_{k-1}}. \tag{2.21}$$

Although the descent property of the search direction is sufficient for establishing the convergence results, efficient conjugate gradient methods have been proposed such that the sufficient descent condition

$$g_k^T d_k \le -c\|g_k\|^2 \tag{2.22}$$

holds for some constant $c > 0$ and all $k \ge 1$. Specifically, Hager and Zhang [46] proposed a family of conjugate gradient methods, where

$$\beta_{k+1}^{\mathrm{HZ}} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - \lambda_k \frac{\|y_k\|^2}{d_k^T y_k} \frac{g_{k+1}^T d_k}{d_k^T y_k}, \tag{2.23}$$

with $\lambda_k > \bar{\lambda} > \frac{1}{4}$, and they preferred the choice $\lambda_k = 2$. By introducing a suitable truncation of β_k and the approximate Wolfe line search, they built a conjugate gradient software, called

CG_DESCENT [47], which performs better than PRP+ of Gilbert and Nocedal. By observing that the loss of orthogonality in the sequence of gradients caused by numerical error might slow down the convergence of conjugate gradient methods, Hager and Zhang [48] updated CG_ DESCENT to Version 6.8 by combining the limited memory technique.

By projecting the search direction of the self-scaling memoryless BFGS method, which was proposed by Perry [66] and Shanno [77], into the one-dimensional manifold $\mathcal{S}_k = \mathrm{Span}\{-g_k + \beta d_{k-1}\}$, Dai and Kou [22] proposed a family of conjugate gradient methods, where

$$\beta_k(\tau_k) = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \left\{\tau_{k-1} + \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}} - \frac{s_{k-1}^T y_{k-1}}{\|s_{k-1}\|^2}\right\} \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \tag{2.24}$$

Then by choosing $\tau_{k-1} = \frac{s_{k-1}^T y_{k-1}}{\|s_{k-1}\|^2}$, they recommended the formula

$$\beta_k^{\mathrm{DK}} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}} \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}, \tag{2.25}$$

which is such that the sufficient descent condition (2.22) holds with $c = \frac{3}{4}$. The software CGOPT was then developed in [22] based on the Dai–Kou method and an improved Wolfe line search. Furthermore, CGOPT was updated in [60] to Version 2.0, which consists of standard CG iterations and subspace iterations and is a strong competitor of CG_DESCENT.

Despite significant progresses, we feel there is still much more room to seek for the best nonlinear conjugate gradient algorithms. For example, Yuan and Stoer [88] first presented the subspace minimization conjugate gradient method by determining the search direction via the subproblem $\min\{g_k^T d + \frac{1}{2} d^T B_k d : d \in \mathrm{Span}\{g_k, d_{k-1}\}\}$. Following this line, Dai and Kou [23] approximated the term $g_k^T B_k g_k$ by $\frac{3}{2} \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}} \|g_k\|^2$ and presented an efficient Barzilai–Borwein conjugate gradient method.

3. CONSTRAINED OPTIMIZATION

An intuitive way to deal with constrained optimization problems is to transform them into unconstrained optimization problems via penalty functions or indicator functions. Nowadays, there are many classes of numerical methods and software for constrained optimization; see, e.g., [65, 89]. Sequential quadratic programming methods and interior-point methods are two classes of very efficient numerical methods for constrained optimization among many others. In addition, augmented Lagrangian methods of multipliers also received a lot of attention since they form the base of alternating direction method of multipliers (see [5]), which can deal with large-scale structured problems. In this section, we shall briefly review some of our recent contributions to the three classes of methods.

3.1. Sequential quadratic programming methods

The sequential quadratic programming (SQP) method, also called Wilson–Han–Powell method, is one of the most effective methods for constrained optimization and can be viewed as a natural extension of Newton and quasi-Newton methods. Its basic idea is to transform the original problem into a sequence of quadratic program (QP) subproblems. After solving each QP subproblem, we wish the full SQP-step to be a superlinearly convergent step; by combining some criterion, we evaluate whether to accept this full step and introduce some remedy if necessary. Based on the used criterion, SQP methods can roughly be classified into two categories. One is penalty-type methods, whose main feature is to use some penalty function. The other is penalty-free methods, which do not use any penalty parameters, e.g., filter methods [37], the methods without any penalty function or a filter [44].

However, two possible difficulties may arise in SQP methods. One is that the QP subproblem may be inconsistent. The other is how to avoid the Maratos effect [61] since the full SQP-step may lead to an increase in both the objective function and the constraint violation even when the iteration is arbitrarily close to a regular minimizer.

Various techniques are available for dealing with inconsistency of the QP subproblem. Early such works include the scaling technique by Powell [70] and the Sl_1QP method by Fletcher [35]. Spellucct [78] introduced some slack variables for dealing with inconsistent subproblems. Liu and Yuan [58] provided a robust SQP method by solving an unconstrained piecewise quadratic subproblem and a QP subproblem at each iteration. Fabien [34] solved a relaxed, strictly convex, QP subproblem if the constraints are inconsistent.

For the Maratos effect, Chen et al. [13] gave the following formal definition.

Definition 3.1. Let x^* and $v(\|c(x)\|)$ be a solution and a measurement of the constraint violation of (1.2)–(1.4), respectively. Given a sequence $\{x_k\}$ which converges to x^* and a sequence of full SQP-steps $\{d_k\}$, we say that the Maratos effect happens if (i) $\lim_{k\to+\infty} \|x_k + d_k - x^*\|/\|x_k - x^*\| = 0$; (ii) $f(x_k + d_k) > f(x_k)$ and $v(\|c(x_k + d_k)\|) > v(\|c(x_k)\|)$.

When the Maratos effect happens, the full SQP-step may not be accepted since it makes both the objective function and the constraint violation worse. In fact, in this case, we see that the pair $(\|h(x_k)\|, f(x_k))$ dominates the pair $(\|h(x_k + d_k)\|, f(x_k + d_k))$ even if $x_k + d_k$ is much closer to x^* than x_k and hence $x_k + d_k$ will not be accepted by the filter method initially proposed by Fletcher and Leyffer [37]. This is also the case for many other globally convergent penalty-type and penalty-free-type algorithms. For example, if l_1 and l_∞ exact penalty functions are used, the full trial step d_k will be rejected as well since the value $f(x) + \sigma\|h(x)\|_p$ $(\sigma > 0,\ p = 1, \infty)$ becomes worse.

Several approaches have been proposed for avoiding the Maratos effect, including nonmonotone line search strategies [11], second order correction step [41, 62], and the use of differentiable exact penalty functions [72]. The computation of second-order correction steps may be cumbersome, and the nonmonotone framework will complicate the algorithmic implementation. Another approach of avoiding the Maratos effect is to utilize the Lagrangian function value instead of the objective function value. Such an idea can be found in Ulbrich [80], who proposed a trust-region filter-SQP method by introducing the Lagrangian function value in the filter.

For efficiency evidence of using the Lagrangian function value in avoiding the Maratos effect, Chen et al. [13] provided the following basic result (for simplicity, it is assumed that $m_I = 0$).

Theorem 3.1. *Suppose that (x^*, λ^*) is a KKT pair of problem (1.2)–(1.3), at which the second-order sufficient conditions and the linear independence constraint qualification hold. Assume that $v(\|h(x)\|)$ is a measurement of constraint violation of the problem, $\lambda(x)$ is a Lipschitz continuous multipliers function, and $P_k(B_k - \nabla_{xx}^2 L(x_k, \lambda(x_k)))d_k = o(\|d_k\|)$, where $\{x_k\}$ converges to x^*, B_k is the approximation of $\nabla_{xx}^2 L(x_k, \lambda(x_k))$ in the QP subproblem, P_k is an orthogonal projection matrix from $\mathcal{R}^n$ to the null space of A_k^T, and d_k is the full SQP-step. If $v(\|h(x_k + d_k)\|) > v(\|h(x_k)\|)$, then there must exist some constant $b_0 > 0$ such that $L(x_k + d_k, \lambda(x_k + d_k)) \le L(x_k, \lambda(x_k)) - b_0\|d_k\|^2$.*

The above theorem indicates that, when the Maratos effect happens, there must be a sufficient decrease in the Lagrangian function. Thus we see that the Lagrangian function value can play an important role. In this case, we can prove that Fletcher's differentiable exact penalty function is decreasing as well.

Furthermore, Chen et al. [12] proposed a penalty-free trust-region method with the Lagrangian function value without using feasibility restoration phase. Chen et al. [13] presented a line search penalty-free SQP method for equality constrained optimization with the Lagrangian function value. Thus with the use of the Lagrangian function value, one would expect SQP methods to control possible erratic behavior in a better manner and share the rapid convergence of Newton-like methods. More researches are required on the use of Lagrangian function value in SQP methods for general nonlinear optimization.

3.2. Interior-point methods

Interior-point methods have been among the most efficient methods for continuous optimization, see, e.g., Ye [84], Byrd et al. [8], Vanderbei and Shanno [81], Wächter and Biegler [83], Liu and Yuan [59], Curtis [15], and Gould et al. [43]. These methods are iterative and require every iterate to be an interior point. The numerical efficiency and polynomial computational complexity of interior-point methods for linear programming made a lot of researchers to be interested again in interior-point methods for nonlinear optimization.

However, Wächter and Biegler [82] noticed that many line-search interior-point methods for nonlinear optimization may fail to find a feasible point of a single-variable nonlinear and nonconvex problem, even though the problem is well posed. In addition, the algorithmic framework of interior-point methods for nonlinear optimization often includes an inner-loop and an outer-loop, in which the inner-loop finds an approximate solution of a logarithmic barrier subproblem and the outer-loop focuses on the update of the barrier parameter. This framework is distinct from that of interior-point methods for linear programming (which reduces the barrier parameter at each iteration) and is believed to be ineffective sometimes.

Below we shall describe a primal–dual interior-point relaxation method with nice properties. The method was recently introduced in [56].

In order to avoid requiring feasible interior-point iterates, traditional interior-point methods for problem (1.2)–(1.4) introduce slack variables for inequality constraints and solve the logarithmic barrier subproblem

$$\min \quad f(x) - \mu \sum_{i=1}^{m_I} \ln z_i \tag{3.1}$$

$$\text{such that} \quad h(x) = 0, \tag{3.2}$$

$$g(x) - z = 0, \tag{3.3}$$

where $\mu > 0$ is a barrier parameter and $z_i > 0$ is the ith component of z.

Noting that the subproblem (3.1)–(3.3) is an equality constrained optimization problem, we reformulate it as another constrained optimization by using the Hestenes–Powell augmented Lagrangian:

$$\min_{x,z} \quad f(x) - \mu \sum_{i=1}^{m_I} \ln z_i - v^T\big(g(x) - z\big) + \frac{1}{2}\rho\big\|\big(g(x) - z\big)\big\|^2 \tag{3.4}$$

$$\text{such that} \quad h(x) = 0, \tag{3.5}$$

where $v \in \Re^{m_I}$ is an estimate of the Lagrange multipliers associated with the original inequality constraints, $\mu > 0$ is a barrier parameter, and $\rho > 0$ is a penalty parameter. Since the objective function in (3.4) is strictly convex with respect to z, the unique minimizer of z can be derived with the expression

$$z_i = \frac{1}{2\rho}\Big[\sqrt{\big(v_i - \rho g_i(x)\big)^2 + 4\rho\mu} - \big(v_i - \rho g_i(x)\big)\Big], \quad i = 1, \dots, m_I. \tag{3.6}$$

The preceding expression depends on the primal variable vector x and the dual variable vector v, thus can be taken as a function of (x, v). For simplicity, corresponding to (3.6), denote

$$y_i = \frac{1}{2\rho}\Big[\sqrt{\big(v_i - \rho g_i(x)\big)^2 + 4\rho\mu} + \big(v_i - \rho g_i(x)\big)\Big], \quad i = 1, \dots, m_I. \tag{3.7}$$

Substituting (3.6) for z in the objective function in (3.4) and maximizing the derived function with respect to v, since it is a strictly concave function of v, the subproblem (3.4)–(3.5) can be reformulated as

$$\min_x \max_v \quad f(x) - \mu \sum_{i=1}^{m_I} \ln z_i + \frac{1}{2}\rho\|y\|^2 - \frac{1}{2\rho}\|v\|^2 \tag{3.8}$$

$$\text{such that} \quad h(x) = 0, \tag{3.9}$$

where both z and y are real-valued functions of $x \in \Re^n$ and $v \in \Re^{m_I}$ defined by (3.6) and (3.7).

Although the subproblem (3.8)–(3.9) is originated from the logarithmic barrier subproblem (3.1)–(3.3), it is different from the latter in that z_i is not a primal variable but a positive function of primal and dual variables. Thus, the requirement that the primal and dual iterates are interior-points is relieved. Our primal–dual interior-point relaxation method is proposed to solve the subproblem (3.8)–(3.9) approximately. In particular, the barrier and penalty parameters are updated adaptively during the iterative process.

We firstly describe the relation between the logarithmic barrier subproblem (3.1)–(3.3) and its augmented Lagrangian reformulation (3.8)–(3.9).

Theorem 3.2 ([56]). *Suppose $\mu > 0$ and $\rho > 0$. Then $(x^*, v^*) \in \Re^n \times \Re^{m_I}$ is a local solution of the constrained minimax problem (3.8)–(3.9) if and only if x^* is a local solution of the logarithmic-barrier subproblem (3.1)–(3.3) and $g_i(x^*) > 0$, $v_i^* = \mu / g_i(x^*)$ for all $i = 1, \dots, m_I$.*

Then we show the relation between the original problem (1.2)–(1.4) and the augmented Lagrangian reformulation (3.8)–(3.9).

Theorem 3.3 ([56]). *Given $\rho > 0$. Let z be defined by (3.6). The point (x^*, u^*, v^*) is a KKT triple of the original problem (1.2)–(1.4) if and only if (x^*, u^*, v^*) and μ^* satisfy the system*

$$\mu = 0, \tag{3.10}$$

$$\nabla f(x) - \nabla h(x) u - \nabla g(x) v = 0, \tag{3.11}$$

$$h(x) = 0, \tag{3.12}$$

$$g(x) - z = 0. \tag{3.13}$$

It should be noted that equations (3.11)–(3.13) are the KKT conditions of the subproblem (3.8)–(3.9). Moreover, for all $\mu > 0$ and $i = 1, \dots, m_I$, both z_i and y_i are twice continuously differentiable with respect to x and v. Thus the subproblem (3.8)–(3.9) can be thought as a *smoothing* problem of the original problem (1.2)–(1.4) in the sense that the system (3.11)–(3.13) is a smoothing system of the KKT conditions of the original problem. Letting the merit function $\phi_{(\mu,\rho)}(x, u, v)$ be the square of l_2 residuals of the system (3.11)–(3.13), the preceding system (3.10)–(3.13) can be further reformulated as the system

$$\mu + \gamma \phi_{(\mu,\rho)}(x, u, v) = 0, \tag{3.14}$$

$$\nabla f(x) - \nabla h(x) u - \nabla g(x) v = 0, \tag{3.15}$$

$$h(x) = 0, \tag{3.16}$$

$$g(x) - z = 0, \tag{3.17}$$

where $\gamma \in (0, 1)$ is a scalar. The two systems (3.10)–(3.13) and (3.14)–(3.17) are equivalent, but the connection between the parameter μ and the KKT residual $\phi_{(\mu,\rho)}(x, u, v)$ is enhanced in (3.14) which requires that μ vanishes with $\phi_{(\mu,\rho)}(x, u, v)$.

Then by sequentially solving the linearized system of the system (3.14)–(3.17) and using the merit function $\phi_{(\mu,\rho)}(x, u, v)$, an efficient primal–dual interior-point relaxation method was provided in [55]. Under suitable assumptions, the new method is proved to have strong global convergence and rapid local convergence [55, 56]. In particular, [26] shows that some variant of this method is capable of rapidly detecting the infeasibility of nonlinear optimization. Numerical experiments demonstrate that the new method not only is efficient for well-posed feasible problems, but also is applicable for some feasible problems without LICQ or MFCQ and some infeasible problems.

The new method is robust in the following three aspects. Firstly, the new method does not require any primal or dual iterate to be an interior-point but prompts the iterate to be an interior-point, which is quite different from most of the globally convergent interior-point methods in the literature. Secondly, the new method uses a single-loop framework and updates the barrier parameter adaptively, which is similar to that of interior-point methods for linear programming. Thirdly, the new method has strong global convergence and is capable of rapidly detecting the infeasibility.

For convex and linear programming, our primal–dual interior point relaxation method provides an intermediate approach between the simplex method and the interior-point method. In addition, we admit the components of $g(x)$ and v to be zero during the iterative process and thus μ can be zero when the solution is obtained. Based on these observations, we may expect our relaxation method to give a solution with high accuracy and to avoid the ill-conditioning phenomenon of interior-point methods, and improve the performance of interior-point methods for large scale problems. Some future topics include its extension for nonlinear semidefinite programming and its complexity when applied for linear programming. An efficient extension of the method has been given for convex quadratic programming [91]. More researches and software-building are expected to go along this line.

3.3. Augmented Lagrangian method of multipliers

The augmented Lagrangian method of multipliers (ALM) was initially proposed by Hestenes [49] and Powell [69] for solving nonlinear optimization with only equality constraints. The ALM minimizes the Hestenes–Powell augmented Lagrangian approximately and circularly with update of multipliers and has been attracting extensive attention in the community. It was generalized by Rockafellar [76] to solve optimization problems with inequality constraints. Many ALMs have been proposed for various optimization problems.

Consider the general nonlinear optimization problem (1.2)–(1.4). By introducing some slack variables z_i $(i = 1, \dots, m_I)$ for the inequality constraints, the problem can equivalently be transformed to that with general equality constraints and nonnegative constraints

$$
\begin{aligned}
\min \quad & f(x) && (3.18)\\
\text{such that} \quad & h(x) = 0, && (3.19)\\
& g(x) - z = 0, && (3.20)\\
& z \geq 0. && (3.21)
\end{aligned}
$$

Using the augmented Lagrangian on equality constraints, problem (3.18)–(3.21) is reformulated as a nonlinear program with only nonnegative constraints:

$$
\min_{x,z} \quad f(x) - u^T h(x) - v^T\big(g(x) - z\big) + \frac{1}{2}\rho\big(\|h(x)\|^2 + \|g(x) - z\|^2\big) \tag{3.22}
$$

$$
\text{such that} \quad z \geq 0, \tag{3.23}
$$

where $u \in \Re^{m_E}$ and $v \in \Re^{m_I}$ are the estimates of Lagrange multipliers and $\rho > 0$ is the penalty parameter. Thanks to the strict convexity of the objective function with respect to z, we may explicitly get the optimal z, yielding an equivalent unconstrained optimization subproblem of (3.22)–(3.23),

$$\min_{x} \quad f(x) - u^T h(x) + \frac{1}{2}\rho\big\|h(x)\big\|^2 + \sum_{i=1}^{m_I} \phi\big(g_i(x), v_i; \rho\big), \tag{3.24}$$

where $\phi(g_i(x), v_i; \rho)$ equals to $-v_i g_i(x) + \frac{1}{2}\rho g_i(x)^2$ if $g_i(x) \leq v_i/\rho$ and $-\frac{1}{2}v_i^2/\rho$ otherwise. Unfortunately, the function ϕ in (3.24) is in general discontinuous in the second derivative with respect to x. Some other unconstrained reformulation is based on the optimization (3.4)–(3.5) in the form

$$\min_{x,z} \quad f(x) - \mu \sum_{i=1}^{m_I} \ln z_i - u^T h(x) + \frac{1}{2}\rho\big\|h(x)\big\|^2 - v^T\big(g(x) - z\big) + \frac{1}{2}\rho\big\|\big(g(x) - z\big)\big\|^2, \tag{3.25}$$

where both x and z are primal variables, u and v are dual estimates, and z should be an interior-point.

Originated from solving the augmented Lagrangian reformulation of problem (3.8)–(3.9), the new ALM, proposed by Liu et al. [57], solves the following problem approximately and circularly with update of multipliers u and v:

$$\min_{x} \quad f(x) - u^T h(x) + \frac{1}{2}\rho\big\|h(x)\big\|^2 + \sum_{i=1}^{m_I} \psi\big(g_i(x), v_i; \mu, \rho\big), \tag{3.26}$$

where $\psi(g_i(x), v_i; \mu, \rho) = -\mu \ln z_i + \frac{1}{2}\rho y_i^2 - \frac{1}{2\rho}v_i^2$ and z_i and y_i are defined by (3.6) and (3.7). A detailed description of the new ALM is given in Algorithm 1. It is a generalization of the classical Hestenes–Powell augmented Lagrangian and a combination of the augmented Lagrangian and the interior-point technique.

Algorithm 1: A new ALM for problem (1.2)–(1.4) [57]

```
Given $(x_0, u_0, v_0)$, and $\mu_0$, $\rho_0$. Let $k := 0$.
while $\mu_k > \epsilon$ or $\phi_{(\mu_k,\rho_k)}(x_k, u_k, v_k) > \epsilon$ do
    Compute $x_{k+1}$ to be an approximate solution of problem (3.26) with the
      initial point $x_k$.
    Update $u_k$ by $u_{k+1} = u_k - \rho_k h(x_k)$.
    Update $v_k$ by $v_{k+1} = \rho_k y(x_{k+1}, v_k; \mu_k, \rho_k)$.
    Update $\rho_{k+1} \geq 2\rho_k$ if $\|z(x_{k+1}, v_{k+1}; \mu_k, \rho_k) - c(x_{k+1})\|$ is not small.
    Update $\mu_{k+1} \leq 0.5\mu_k$ if $\|z(x_{k+1}, v_{k+1}; \mu_k, \rho_k) - c(x_{k+1})\|$ is small.
    Let $k := k + 1$.
end
```

Liu et al. [57] proved that the new ALM is of strong global convergence, rapid infeasibility detection, and shares the same convergence rate to the KKT point as the Hestenes–Powell augmented Lagrangian for optimization with equality constraints.

Although the subproblem (3.26) is similar to the augmented Lagrangian counterpart (3.24) and the interior-point counterpart (3.25) in appearance that all of them are unconstrained optimization and first-order smooth, but it is essentially distinct from the latter two subproblems in the following aspects.

Firstly, the function ψ in (3.26) has one more parameter μ than ϕ in (3.24) and is always twice continuously differentiable with respect to x provided g is twice continuously differentiable and v holds fixed, while ϕ in (3.24) has discontinuous second derivative with respect to x. The problem (3.25) has the same property with (3.26). Secondly, the subproblems (3.24) and (3.26) are convex if the original problem (1.2)–(1.4) is convex, while the subproblem (3.25) can be nonconvex even though the original problem is convex. Thirdly, unlike subproblem (3.25), the subproblems (3.24) and (3.26) do not require any primal or dual variable to be positive. Moreover, $\psi(g_i(x), v_i; \mu, \rho)$ is well defined for every $x \in \Re^n$ and $v \in \Re^{m_I}$, while (3.25) requests $z > 0$ and $v > 0$.

To summarize, the new ALM can deal with optimization problems with inequality constraints and shares the same convergence rate to the KKT point as the Hestenes–Powell augmented Lagrangian for optimization problems with equality constraints. As the new ALM has nice properties, more researches are expected along this line.

4. OPTIMIZATION WITH LEAST CONSTRAINT VIOLATION

The theory and algorithms for constrained optimization usually assume the feasibility of the optimization problem. If the constraints are inconsistent, several numerical algorithms have been proposed to find infeasible stationary points, which have nothing to do with the objective function; see, e.g., Byrd et al. [7], Burke et al. [6], and Dai et al. [26]. However, there are important optimization problems, which may be either feasible or infeasible and whose objective function is wished to be minimized with the least constraint violation even if they are infeasible. A typical example comes from rocket trajectory optimal control, where the fuel is minimized with the aim of landing at a target point and subjected to other constraints. If landing at the target is not possible, we might wish to minimize the distance between the real landing point and the target and thereafter optimize the required fuel. Hence we are led to optimization problems with least constraint violation.

For optimization with possible inconsistent constraints, we prove that the minimization problem with least constraint violation is equivalent to a Lipschitz equality constrained optimization problem. To this aim, consider the nonlinear optimization problem

$$\begin{aligned} \min \quad & f(x) \\ \text{such that} \quad & Ax = b, \\ & g_i(x) \geq 0, \ i = 1, \ldots, p, \end{aligned} \tag{4.1}$$

where $f : \Re^n \to \Re$ is smooth and g_i $(i = 1, \dots, p)$ are differentiable concave functions. In this case, the optimization problem with the least constraint violation can be expressed as

$$\begin{aligned} &\min && f(x) \\ &\text{such that} && A^T(Ax - b) + \mathcal{J}g(x)^T\big[g(x)\big]_- = 0. \end{aligned} \tag{4.2}$$

Define $H(x, y) = A^T A - \sum_{j=1}^{p} y_j \nabla^2 g_j(x)$. For $y^* = [-g(x^*)]_+$ and $z^* = [g(x^*)]_+$, define $\alpha^* = \{i : y_i^* > 0\}$, $\beta^* = \{i : y_i^* = z_i^* = 0\}$, $\gamma^* = \{i : z_i^* > 0\}$. Then we are able to give an elegant necessary optimality condition from the classical optimality theory of Lipschitz continuous optimization.

Theorem 4.1 ([31]). *Let (x^*, y^*) be a local minimizer of problem (4.2). Suppose that the matrix $H(x^*, y^*) + \mathcal{J}g_{\alpha^*}(x^*)^T \mathcal{J}g_{\alpha^*}(x^*)$ is positive definite. Then there exist $\lambda^* \in \Re^n$ and $[v_b]_{\beta^*} \in \Re^{|\beta^*|}$ satisfying $[v_b]_i \in [0, 1]$, $i \in \beta^*$ such that*

$$\begin{aligned} \nabla f(x^*) + \big[&H(x^*, y^*) + \mathcal{J}g_{\alpha^*}(x^*)^T \mathcal{J}g_{\alpha^*}(x^*) \\ &+ \mathcal{J}g_{\beta^*}(x^*)^T \mathrm{Diag}\big([v_b]_{\beta^*}\big)\mathcal{J}g_{\beta^*}(x^*)\big]\lambda^* = 0. \end{aligned} \tag{4.3}$$

Dai and Zhang [31] found that the penalty method can be used for solving optimization problems with least constraint violation. Chiche and Gilbert [14] proved that the augmented Lagrangian method of multipliers (ALM) can deal with an infeasible convex quadratic optimization problem. Is the ALM still valid for general convex optimization with the least constraint violation?

To this aim, consider the following convex constrained optimization problem:

$$(\mathrm{P}) \qquad \begin{aligned} &\min && f(x) \\ &\text{such that} && g(x) \in \mathcal{K}, \end{aligned} \tag{4.4}$$

where $f : \Re^n \to \Re$, $g : \Re^n \to \mathcal{Y}$, $\mathcal{K} \subset \mathcal{Y}$ is a nonempty closed convex set, and $\mathcal{Y}$ is a finite-dimensional Hilbert space. We analyze the dual of the problem with the least constraint violation. By introducing a vector $y \in \mathcal{Y}$, problem (4.4) is equivalently expressed as

$$\begin{aligned} &\min && f(x) \\ &\text{such that} && g(x) = y, \\ & && y \in \mathcal{K}. \end{aligned} \tag{4.5}$$

For a given $s \in \mathcal{Y}$, the shifted problem is defined as

$$\mathrm{P}(s) \qquad \begin{aligned} &\min && f(x) \\ &\text{such that} && g(x) + s \in \mathcal{K}. \end{aligned} \tag{4.6}$$

Here we call s a shift. The set of feasible shifts, denoted as $\mathcal{S}$, is defined by

$$\mathcal{S} := \big\{s \in \mathcal{Y} : \text{there exists some } x \in \Re^n \text{ such that } g(x) + s \in \mathcal{K}\big\}. \tag{4.7}$$

Define the smallest norm shift by $\bar{s} = \arg\min\{\frac{1}{2}\|s\|^2 : s \in \mathcal{S}\}$. If $\mathcal{S}$ is closed, then $\bar{s}$ can be achieved, i.e., $\bar{s} \in \mathcal{S}$. In this case, the optimization problem with the least constraint violation is expressed as follows:

$$\mathrm{P}(\bar{s}) \qquad \begin{aligned} &\min && f(x) \\ &\text{such that} && g(x) + \bar{s} \in \mathcal{K}. \end{aligned} \tag{4.8}$$

Now we shall present the properties of the ALM for problem (4.5), which was provided by Dai and Zhang [32]. The Lagrangian of problem (4.5), denoted by l, is defined by $l(x,y,\lambda) = f(x) + \lambda^T(g(x) - y)$. The augmented Lagrangian function of problem (4.5), denoted by l_r, is defined by

$$l_r(x,y,\lambda) = f(x) + \lambda^T\big(g(x) - y\big) + \frac{r}{2}\big\|g(x) - y\big\|^2. \tag{4.9}$$

The dual function $\theta : \mathcal{Y} \to \overline{\Re}$ associated with problem (4.5) is

$$\theta(\lambda) := -\inf_{x \in \Re^n, y \in K} l(x,y,\lambda). \tag{4.10}$$

Denote by D and D(s) the conjugate dual problems of P and P(s), respectively. Then problems D and D(s) are expressed as follows:

$$(\mathrm{D}) \quad \max_{\lambda}\big[-\theta(\lambda)\big], \quad \big(\mathrm{D}(s)\big) \quad \max_{\lambda}\big[s^T\lambda - \theta(\lambda)\big]. \tag{4.11}$$

The following proposition reveals that the solution set of the dual problem, if nonempty, is unbounded when $\bar{s} \neq 0$.

Proposition 4.1 ([32]). *Assume that* $\bar{s} \neq 0$, $\operatorname{val} \mathrm{P}(\bar{s}) \in \Re$, v *is lower semicontinuous at* $\bar{s}$ *and* $\operatorname{Sol} \mathrm{D}(\bar{s}) \neq \emptyset$. *Then* $\operatorname{Sol} \mathrm{D}(\bar{s})$ *is unbounded with* $-\bar{s} \in [\operatorname{Sol} \mathrm{D}(\bar{s})]^\infty$.

For the sequence $\{(x^k, y^k, \lambda^k)\}$ generated by the ALM for solving problem (4.5), defining $s^k = y^k - g(x^k)$, we are able to prove the following theorem.

Theorem 4.2 ([32]). *Assume that* $\bar{s} \neq 0$, $\operatorname{val} \mathrm{P}(\bar{s}) \in \Re$, v *is lower semicontinuous at* $\bar{s}$ *and* $\operatorname{Sol} \mathrm{D}(\bar{s}) \neq \emptyset$. *Assume also that* $\{r_k\}$ *has a positive lower bound and* $\{(x^k, y^k)\}$ *has an accumulation point. Then we have that* (i) $s^k \to \bar{s}$; (ii) $\{\lambda^k\}$ *diverges;* (iii) *for every* $\varepsilon > 0$, *there exists an index* k *large enough such that* (x^k, y^k) *satisfies* ε*-approximate optimality conditions of problem* P$(\bar{s})$ *in terms of the augmented Lagrangian.*

The above theorem shows that the ALM can deal with convex optimization with least constraint violation. Studies on the theory and algorithms for optimization with least constraint violation are clearly required.

5. SOME DISCUSSIONS

Due to limited space, this article only reviewed some numerical methods for general nonlinear optimization. An early good review on unconstrained optimization is given by Nocedal [64], where two open questions about quasi-Newton methods were summarized. One is whether the DFP method with the Wolfe line search converges for uniformly convex functions. The other is whether the BFGS method with the Wolfe line search converges for general nonlinear functions. A negative answer of the second open question has been known (see, e.g., Dai [18]). Although Yuan [86] made a significant progress on the first open question, we do not know its answer, yet. The infimum of the Q-order of the convergence of quasi-Newton methods is only one [85]. The work of Rodomanov and Nesterov [75] stimulated research interests on this topic again.

Previous studies for constrained optimization usually assumed the feasibility of the optimization problem. This article described optimality conditions for optimization with least constraint violation and the ability of the ALM method to deal with such a problem. Independently of the work [31], Censor et al. [10] proposed a data-compatibility approach for the problem and presented some theoretical analysis. However, more researches are clearly required along this line.

The development of nonlinear optimization influences many research directions in optimization such as matrix optimization, sparse optimization, and nonsmooth optimization. There is still much to do in extending nonlinear optimization methods for minimax optimization, which arises from both modern machine learning and tradition research areas.

ACKNOWLEDGMENTS

The author is very grateful to Professor Ya-xiang Yuan for his long-time guidance and help, and to his collaborators for their valuable discussions and cooperations.

FUNDING

This work was partially supported by the NSFC grants (nos. 12021001, 11991021, 11991020, 11631013, and 11971372) and the Strategic Priority Research Program of Chinese Academy of Sciences (no. XDA27000000).

REFERENCES

[1] H. Akaike, On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method. *Ann. Inst. Statist. Math.* **11** (1959), no. 1, 1–16.

[2] L. Armijo, Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific J. Math.* **16** (1966), no. 1, 1–3.

[3] J. Barzilai and J. M. Borwein, Two-point step size gradient methods. *IMA J. Numer. Anal.* **8** (1988), no. 1, 141–148.

[4] E. G. Birgin, J. M. Martínez, and M. Raydan, Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optim.* **10** (2000), no. 4, 1196–1211.

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3** (2011), no. 1, 1–122.

[6] J. V. Burke, F. E. Curtis, and H. Wang, A sequential quadratic optimization algorithm with rapid infeasibility detection. *SIAM J. Optim.* **24** (2014), no. 2, 839–872.

[7] R. H. Byrd, F. E. Curtis, and J. Nocedal, Infeasibility detection and SQP methods for nonlinear optimization. *SIAM J. Optim.* **20** (2010), no. 5, 2281–2299.

[8] R. H. Byrd, M. E. Hribar, and J. Nocedal, An interior point algorithm for large-scale nonlinear programming. *SIAM J. Optim.* **9** (1999), no. 4, 877–900.

[9] A. Cauchy, Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris* **25** (1847), 536–538.

[10] Y. Censor, M. Zaknoon, and A. J. Zaslavski, Data-compatibility of algorithms. 2019, arXiv:1911.11389.

[11] R. Chamberlain, M. J. D. Powell, C. Lemarechal, and H. Pedersen, The watchdog technique for forcing convergence in algorithms for constrained optimization. In *Algorithms for constrained minimization of smooth nonlinear functions*, edited by A. G. Buckley and J. Goffin, pp. 1–17, Springer, 1982.

[12] Z. W. Chen, Y. H. Dai, and J. Y. Liu, A penalty-free method with superlinear convergence for equality constrained optimization. *Comput. Optim. Appl.* **76** (2020), no. 3, 801–833.

[13] Z. W. Chen, Y. H. Dai, and T. Y. Zhang, *A line search penalty-free SQP method for equality constrained optimization without Maratos effect*. Tech. rep., AMSS, Chinese Academy of Sciences, Beijing, China, 2021.

[14] A. Chiche and J. C. Gilbert, How the augmented Lagrangian algorithm can deal with an infeasible convex quadratic optimization problem. *J. Convex Anal.* **23** (2016), no. 2, 425–459.

[15] F. E. Curtis, A penalty-interior-point algorithm for nonlinear constrained optimization. *Math. Program. Comput.* **4** (2012), no. 2, 181–209.

[16] Y. H. Dai, A nonmonotone conjugate gradient algorithm for unconstrained optimization. *J. Syst. Sci. Complex.* **15** (2002), 139–145.

[17] Y. H. Dai, Convergence analysis of nonlinear conjugate gradient methods. In *Optimization and regularization for computational inverse problems and applications*, edited by Y. Wang, C. Yang, and A. G. Yagola, pp. 157–181, Springer, 2010.

[18] Y. H. Dai, A perfect example for the BFGS method. *Math. Program.* **138** (2013), no. 1, 501–530.

[19] Y. H. Dai and R. Fletcher, On the asymptotic behaviour of some new gradient methods. *Math. Program.* **13** (2005), no. 3, 541–559.

[20] Y. H. Dai and R. Fletcher, New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds. *Math. Program.* **106** (2006), no. 3, 403–421.

[21] Y. H. Dai, J. Han, G. Liu, D. Sun, H. Yin, and Y. X. Yuan, Convergence properties of nonlinear conjugate gradient methods. *SIAM J. Optim.* **10** (1999), no. 2, 345–358.

[22] Y. H. Dai and C. X. Kou, A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search. *SIAM J. Optim.* **23** (2013), no. 1, 296–320.

[23] Y. H. Dai and C. X. Kou, A Barzilai–Borwein conjugate gradient method. *Sci. China Math.* **59** (2016), no. 8, 1511–1524.

[24] Y. H. Dai and L. Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods. *Appl. Math. Optim.* **43** (2001), no. 1, 87–101.

[25] Y. H. Dai and L. Z. Liao, R-linear convergence of the Barzilai and Borwein gradient method. *IMA J. Numer. Anal.* **22** (2002), no. 1, 1–10.

[26] Y. H. Dai, X. W. Liu, and J. Sun, A primal–dual interior-point method capable of rapidly detecting infeasibility for nonlinear programs. *J. Ind. Manag. Optim.* **16** (2020), no. 2, 1009–1035.

[27] Y. H. Dai and Y. X. Yuan, A nonlinear conjugate gradient with a strong global convergence property. *SIAM J. Optim.* **10** (1999), no. 1, 177–182.

[28] Y. H. Dai and Y. X. Yuan, *Nonlinear conjugate gradient methods*. Shanghai Scientific & Technical Publishers, Shanghai, 2000 (in Chinese).

[29] Y. H. Dai and Y. X. Yuan, An efficient hybrid conjugate gradient method for unconstrained optimization. *Ann. Oper. Res.* **103** (2001), no. 1, 33–47.

[30] Y. H. Dai and Y. X. Yuan, Analysis of monotone gradient methods. *J. Ind. Manag. Optim.* **1** (2005), no. 2, 181–192.

[31] Y. H. Dai and L. W. Zhang, Optimization with least constraint violation. *CSIAM Trans. Appl. Math.* **2** (2021), no. 3, 551–584.

[32] Y. H. Dai and L. W. Zhang, *The augmented Lagrangian method can approximately solve convex optimization with least constraint violation*. Tech. rep., AMSS, Chinese Academy of Sciences, Beijing, China, 2021.

[33] W. C. Davidon, Variable metric methods for minimization. *SIAM J. Optim.* **1** (1991), no. 1, 1–17.

[34] B. C. Fabien, Implementation of a robust SQP algorithm. *Optim. Methods Softw.* **23** (2008), no. 6, 827–846.

[35] R. Fletcher, *Practical methods of optimization*, 2nd edition. John Wiley, Chichester, 1987.

[36] R. Fletcher, On the Barzilai–Borwein method. In *Optimization and control with applications*, edited by L. Qi, K. Teo, and X. Yang, pp. 235–256, Springer, 2005.

[37] R. Fletcher and S. Leyffer, Nonlinear programming without a penalty function. *Math. Program.* **91** (2002), no. 2, 239–269.

[38] R. Fletcher and M. J. D. Powell, A rapidly convergent descent method for minimization. *Comput. J.* **6** (1963), no. 2, 163–168.

[39] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients. *Comput. J.* **7** (1964), no. 2, 149–154.

[40] G. E. Forsythe, On the asymptotic directions of the s-dimensional optimum gradient method. *Numer. Math.* **11** (1968), no. 1, 57–76.

[41] M. Fukushima, A successive quadratic programming algorithm with global and superlinear convergence properties. *Math. Program.* **35** (1986), no. 3, 253–264.

[42] J. C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.* **2** (1992), no. 1, 21–42.

[43] N. I. Gould, D. Orban and P. L. Toint, An interior-point ℓ_1-penalty method for nonlinear optimization. In *Numerical analysis and optimization*, edited by M. Al-Baali, L. Grandinetti, A. Purnama, Springer Proc. Math. Stat. 134, Springer, Berlin, Cham, 2015.

[44] N. I. Gould and P. L. Toint, Nonlinear programming without a penalty function or a filter. *Math. Program.* **122** (2010), no. 1, 155–196.

[45] L. Grippo, F. Lampariello, and S. Lucidi, A nonmonotone line search technique for Newton’s method. *SIAM J. Numer. Anal.* **23** (1986), no. 1, 707–716.

[46] W. W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM J. Optim.* **16** (2005), no. 1, 170–192.

[47] W. W. Hager and H. C. Zhang, Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent. *ACM Trans. Math. Software* **32** (2006), no. 1, 113–137.

[48] W. W. Hager and H. C. Zhang, The limited memory conjugate gradient method. *SIAM J. Optim.* **23** (2013), no. 4, 2150–2168.

[49] M. R. Hestenes, Multiplier and gradient methods. *J. Optim. Theory Appl.* **4** (1969), no. 5, 303–320.

[50] M. R. Hestenes and E. L. Stiefel, Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.* **49** (1952), no. 6, 409–436.

[51] Y. K. Huang, Y. H. Dai, and X. W. Liu, Equipping the Barzilai–Borwein method with the two dimensional quadratic termination property. *SIAM J. Optim.* **31** (2021), no. 4, 3068–3096.

[52] W. Karush, *Minima of functions of several variables with inequalities as side constraints*. M. Sc. thesis, University of Chicago, 1939.

[53] H. W. Kuhn and A. Tucker, Nonlinear programming. In *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, edited by J. Neyman, pp. 481–492, University of California Press, Berkeley, CA, 1951.

[54] Y. L. Lai, Some properties of the steepest descent method. *Acta Math. Appl. Sin.* **4** (1981), no. 2, 106–116 (in Chinese).

[55] X. W. Liu and Y. H. Dai, A globally convergent primal-dual interior-point relaxation method for nonlinear programs. *Math. Comp.* **89** (2020), no. 323, 1301–1329.

[56] X. W. Liu, Y. H. Dai, and Y. K. Huang, A primal–dual interior-point relaxation method with adaptively updating barrier for nonlinear programs. 2020, arXiv:2007.10803.

[57] X. W. Liu, Y. H. Dai, Y. K. Huang, and J. Sun, A novel augmented Lagrangian method of multipliers for optimization with general inequality constraints. 2021, arXiv:2106.15044.

[58] X. W. Liu and Y. X. Yuan, A robust algorithm for optimization with general equality and inequality constraints. *SIAM J. Sci. Comput.* **22** (2000), no. 2, 517–534.

[59] X. W. Liu and Y. X. Yuan, A null-space primal-dual interior-point algorithm for nonlinear optimization with nice convergence properties. *Math. Program.* **123** (2010), no. 1, 163–193.

[60] Z. X. Liu, H. W. Liu, and Y. H. Dai, An improved Dai–Kou conjugate gradient algorithm for unconstrained optimization. *Comput. Optim. Appl.* **75** (2020), no. 1, 145–167.

[61] N. Maratos, *Exact penalty function algorithms for finite dimensional and control optimization problems*. PhD Thesis, University of London, 1978.

[62] D. Q. Mayne and E. Polak, A superlinearly convergent algorithm for constrained optimization problems. In *Algorithms for constrained minimization of smooth nonlinear functions*, edited by A. G. Buckley and J. L. Goffin, pp. 45–61, Springer, 1982.

[63] M. Mu, H. Xu, and W. Duan, A kind of initial errors related to “spring predictability barrier” for El Ninõ events in Zebiak–Cane mode. *Geophys. Res. Lett.* **34** (2007), L03709.

[64] J. Nocedal, Theory of algorithms for unconstrained optimization. *Acta Numer.* **1** (1991), 199–242.

[65] J. Nocedal and S. Wright, *Numerical optimization*. Springer, New York, 2006.

[66] A. Perry, *A class of conjugate gradient algorithms with a two-step variable metric memory*. Tech. rep., Northwestern University, Center for Mathematical Studies in Economics and Management Science, 1977.

[67] E. Polak and G. Ribiere, Note sur la convergence de méthodes de directions conjuguées. *ESAIM Math. Model. Numer. Anal.* **3** (1969), no. R1, 35–43.

[68] B. T. Polyak, The conjugate gradient method in extreme problems. *USSR Comput. Math. Math. Phys.* **9** (1969), no. 4, 94–112.

[69] M. J. D. Powell, A method for nonlinear constraints in minimization problems. In *Optimization*, edited by R. Fletcher, pp. 283–298, Academic Press, London, 1969.

[70] M. J. D. Powell, A fast algorithm for nonlinearly constrained optimization calculations. In *Numerical analysis*, edited by G. A. Watson, pp. 144–157, Springer, Berlin, 1977.

[71] M. J. D. Powell, Nonconvex minimization calculations and the conjugate gradient method. In *Numerical analysis*, edited by D. F. Griffiths, pp. 122–141, Lecture Notes in Math. 1066, Springer, Berlin, Heidelberg, 1984.

[72] M. J. D. Powell and Y. X. Yuan, A trust region algorithm for equality constrained optimization. *Math. Program.* **49** (1991), no. 1, 1189–211.

[73] M. Raydan, On the Barzilai and Borwein choice of steplength for the gradient method. *IMA J. Numer. Anal.* **13** (1993), no. 3, 321–326.

[74] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM J. Optim.* **7** (1997), no. 1, 26–33.

[75] A. Rodomanov and Y. Nesterov, Greedy quasi-Newton methods with explicit superlinear convergence. *SIAM J. Optim.* **31** (2021), no. 1, 785–811.

[76] R. T. Rockafellar, A dual approach to solving nonlinear programming problems by unconstrained optimization. *Math. Program.* **5** (1973), no. 1, 354–373.

[77] D. F. Shanno, On the convergence of a new conjugate gradient algorithm. *SIAM J. Numer. Anal.* **15** (1978), no. 6, 1247–1257.

[78] P. Spellucct, A new technique for inconsistent QP problems in the SQP method. *Math. Methods Oper. Res.* **47** (1998), no. 3, 355–400.

[79] D. Touati-Ahmed and C. Storey, Efficient hybrid conjugate gradient techniques. *J. Optim. Theory Appl.* **64** (1990), no. 2, 379–397.

[80] S. Ulbrich, On the superlinear local convergence of a filter-SQP method. *Math. Program.* **100** (2004), no. 1, 217–245.

[81] R. J. Vanderbei and D. F. Shanno, An interior-point algorithm for nonconvex nonlinear programming. *Comput. Optim. Appl.* **13** (1999), no. 1, 231–252.

[82] A. Wächter and L. T. Biegler, Failure of global convergence for a class of interior point methods for nonlinear programming. *Math. Program.* **88** (2000), no. 3, 565–574.

[83] A. Wächter and L. T. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.* **106** (2006), no. 1, 25–57.

[84] Y. Ye, *Interior-point algorithm: theory and analysis*. Wiley & Sons, New York, 1997.

[85] Y. X. Yuan, On the least Q-order of convergence of variable metric algorithms. *IMA J. Numer. Anal.* **4** (1984), no. 2, 233–239.

[86] Y. X. Yuan, Convergence of DFP algorithm. *Sci. China Ser. A* **38** (1995), no. 11, 1281–1294.

[87] Y. X. Yuan, A new stepsize for the steepest descent method. *J. Comput. Math.* **24** (2006), no. 2, 149–156.

[88] Y. X. Yuan and J. Stoer, A subspace study on conjugate gradient algorithms. *Z. Angew. Math. Mech.* **75** (1995), no. 1, 69–77.

[89] Y. X. Yuan and W. Y. Sun, *Optimization theories and methods*. Science Press, Beijing, 1997 (in Chinese).

[90] L. Zanni, T. Serafini, G. Zanghirati, K. P. Bennett, and E. Parrado-Hernández, Parallel software for training large scale support vector machines on multiprocessor systems. *J. Mach. Learn. Res.* **7** (2006), no. 54, 1467–1492.

[91] R. Zhang, X. W. Liu, and Y. H. Dai, *Solving convex quadratic programming by a primal-dual interior-point relaxation method*. Tech. rep., AMSS, Chinese Academy of Sciences, Beijing, China, 2021.

[92] B. Zhou, L. Gao, and Y. H. Dai, Gradient methods with adaptive step-sizes. *Comput. Optim. Appl.* **35** (2006), no. 1, 69–86.

YU-HONG DAI

AMSS, Chinese Academy of Sciences, Beijing 100190, China, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China, dyh@lsec.cc.ac.cn