Comput Optim Appl (2012) 51:1–25
DOI 10.1007/s10589-010-9339-1

# A primal-dual augmented Lagrangian

**Philip E. Gill · Daniel P. Robinson**

Received: 4 December 2009 / Published online: 3 August 2010


**Abstract** Nonlinearly constrained optimization problems can be solved by minimizing a sequence of simpler unconstrained or linearly constrained subproblems. In this paper, we consider the formulation of subproblems in which the objective function is a generalization of the Hestenes-Powell augmented Lagrangian function. The main feature of the generalized function is that it is minimized with respect to both the primal *and* the dual variables simultaneously. The benefits of this approach include: (i) the ability to control the quality of the dual variables during the solution of the subproblem; (ii) the availability of improved dual estimates on early termination of the subproblem; and (iii) the ability to regularize the subproblem by imposing explicit bounds on the dual variables. We propose two primal-dual variants of conventional primal methods: a primal-dual bound constrained Lagrangian (pdBCL) method and a primal-dual $\ell_1$ linearly constrained Lagrangian (pd$\ell_1$LCL) method. Finally, a new sequential quadratic programming (pdSQP) method is proposed that uses the primal-dual augmented Lagrangian as a merit function.

**Keywords** Nonlinear programming · Nonlinear constraints · Augmented Lagrangian methods · Bound constrained Lagrangian methods · Linearly constrained Lagrangian methods · Sequential quadratic programming · Primal-dual methods

P.E. Gill's research supported in part by National Science Foundation grants DMS-0511766 and DMS-0915220, and by Department of Energy grant DE-SC0002349.
D.P. Robinson's research supported in part by National Science Foundation grant DMS-0511766 and EPSRC grant EP/F005369/1.

P.E. Gill
Department of Mathematics, University of California, La Jolla, San Diego, CA 92093-0112, USA
e-mail: pgill@ucsd.edu

D.P. Robinson (✉)
Mathematical Institute, University of Oxford, 24–29 St Giles', Oxford, OX1 3LB, UK
e-mail: robinson@maths.ox.ac.uk

## 1 Introduction

This paper concerns methods for finding a local solution of the nonlinearly constrained minimization problem:

$$\text{(NP)} \qquad \underset{x\in\mathbb{R}^n}{\text{minimize}}\ f(x) \quad \text{subject to} \quad c(x)=0, \quad x\geq 0,$$

where $c:\mathbb{R}^n\mapsto\mathbb{R}^m$ and $f:\mathbb{R}^n\mapsto\mathbb{R}$ are twice-continuously differentiable. This problem format assumes that all general inequality constraints have been converted to equalities by the use of slack variables. Methods for solving problem (NP) easily carry over to the more general setting with $l\leq x\leq u$. Much of the discussion will focus on the equality-constrained problem

$$\text{(NEP)} \qquad \underset{x\in\mathbb{R}^n}{\text{minimize}}\ f(x) \quad \text{subject to} \quad c(x)=0.$$

This problem has all the properties necessary for a description of the proposed methods, while avoiding the complications of dealing with bound constraints.

The idea of replacing a constrained optimization problem by a sequence of unconstrained problems parameterized by a scalar $\mu$ has played a fundamental role in the formulation of algorithms since the early 1960s (for a seminal reference, see Fiacco and McCormick [14, 15]). One of the best-known methods for solving the equality-constrained problem (NEP) uses an unconstrained function based on the quadratic penalty function, which combines $f$ with a term of order $1/\mu$ that "penalizes" the sum of the squares of the constraint violations. Under certain conditions (see, e.g., [15, 20, 54, 56]), the minimizers of the penalty function define a differentiable *trajectory* or *central path* that approaches the solution as $\mu\to 0$. Penalty methods approximate this path by minimizing the penalty function for a finite sequence of decreasing values of $\mu$. In this form, the methods have a two-level structure of inner and outer iterations: the inner iterations are those of the method used to minimize the penalty function, and the outer iterations test for convergence and adjust the value of $\mu$. As $\mu\to 0$, the Newton equations for minimizing the penalty function are increasingly ill-conditioned, and this ill-conditioning was perceived to be the reason for the poor numerical performance on some problems. In separate papers, Hestenes [34] and Powell [42] proposed the augmented Lagrangian function for (NEP), which is an unconstrained function based on augmenting the Lagrangian function with a quadratic penalty term that does not require $\mu$ to go to zero for convergence. The price that must be paid for keeping $1/\mu$ finite is the need to update estimates of the Lagrange multipliers in each outer iteration.

Since the first appearance of the Hestenes-Powell function, many algorithms have been proposed based on using the augmented Lagrangian as an objective function for sequential unconstrained minimization. Augmented Lagrangian functions have also been proposed that treat the multiplier vector as a continuous function of $x$; some of these ensure global convergence and permit local superlinear convergence (see, e.g., Fletcher [16]; DiPillo and Grippo [13]; Bertsekas [1, 2]; Boggs and Tolle [5]).

As methods for treating linear inequality constraints and bounds became more sophisticated, the emphasis of algorithms shifted from sequential unconstrained min-

imization to sequential linearly constrained minimization. In this context, the augmented Lagrangian has been used successfully within a number of different algorithmic frameworks for problem (NP). The method used in the software package LANCELOT [9] finds the approximate solution of a sequence of bound constrained problems with an augmented Lagrangian objective function. Similarly, the software package MINOS of Murtagh and Saunders [40] employs a variant of Robinson's linearly constrained Lagrangian (LCL) method [46] in which an augmented Lagrangian is minimized subject to the linearized nonlinear constraints. Friedlander and Saunders [22] define a globally convergent version of the LCL method that can treat infeasible constraints and infeasible subproblems. Augmented Lagrangian functions have also been used extensively as a merit function for sequential quadratic programming (SQP) methods (see, e.g., [4, 6, 8, 18, 24, 26, 49–52]).

The development of path-following interior methods for linear programming in the mid-1980s stimulated renewed interest in the treatment of constraints by sequential unconstrained optimization. This new attention not only resulted in a new understanding of the computational complexity of existing methods but also provided the impetus for the development of new approaches. A notable development was the derivation of efficient path-following methods for linear programming based on applying Newton's method with respect to both the primal and dual variables. These new approaches also refocused attention on two computational aspects of penalty- and barrier-function methods for nonlinear optimization. First, the recognition of the formal equivalence between some primal-dual methods and conventional penalty methods indicated that the inherent ill-conditioning of penalty and barrier functions is not necessarily the reason for poor numerical performance. Second, the crucial role of penalty and barrier functions in problem regularization was recognized and better understood.

In this paper we consider some of these developments in the context of a generalization of the Hestenes-Powell augmented Lagrangian that is minimized jointly with respect to the primal and dual variables. The benefits of this approach include: (i) the ability to control the quality of the dual variables during the solution of the subproblem; (ii) the availability of improved dual estimates on early termination of the subproblem; and (iii) the ability to regularize the subproblem by imposing explicit bounds on the dual variables. Three primal-dual variants of conventional primal methods are proposed: a primal-dual bound constrained Lagrangian (pdBCL) method, a primal-dual $\ell_1$ linearly constrained Lagrangian (pd$\ell_1$LCL) method, and a primal-dual sequential quadratic programming (pdSQP) method.

The paper is organized in five sections. Section 2 is a review of some of the basic properties of the Hestenes-Powell augmented Lagrangian function. It is shown that the Newton direction for the unconstrained minimization of the augmented Lagrangian satisfies a certain primal-dual system in which the change in the dual variables may be specified arbitrarily. In Sect. 3, a generalized primal-dual augmented Lagrangian function is introduced that may be used to define a *continuum* of methods that include several well-known methods as specific cases. Similarities with the conventional Hestenes-Powell augmented Lagrangian method are also discussed. In Sect. 4 it is shown how artificial bounds on the dual variables may be used to *regularize* the associated subproblem. Finally, in Sect. 5 we illustrate the use of the

primal-dual augmented function in three methods: a primal-dual bound constrained Lagrangian method; a primal-dual $\ell_1$ linearly constrained Lagrangian method; and a primal-dual sequential quadratic programming method for solving problem (NEP). In order to provide some preliminary information on the effectiveness of the proposed SQP method, we give results from numerical experiments on a subset of equality constrained problems from the CUTEr test collection (see Bongartz et al. [7] and Gould, Orban and Toint [28]).

### 1.1 Notation and terminology

Unless explicitly indicated otherwise, $\|\cdot\|$ denotes the vector two-norm or its induced matrix norm. The inertia of a real symmetric matrix $A$, denoted by $\text{In}(A)$, is the integer triple $(a_+, a_-, a_0)$ giving the number of positive, negative and zero eigenvalues of $A$. Given vectors $a$ and $b$ with the same dimension, the vector with $i$th component $a_i b_i$ is denoted by $a \cdot b$. The symbol $e$ is used to denote a column vector of ones with dimension determined by the context. A local solution of an optimization problem is denoted by $x^*$. The vector $g(x)$ is used to denote $\nabla f(x)$, the gradient of $f(x)$, and $H(x)$ denotes the (symmetric) Hessian matrix $\nabla^2 f(x)$. The matrix $J(x)$ denotes the $m \times n$ constraint Jacobian, which has $i$th row $\nabla c_i(x)^T$, the gradient of the $i$th constraint function $c_i(x)$. The matrix $H_i(x)$ denotes the Hessian of $c_i(x)$. The Lagrangian function associated with (NEP) is $\mathcal{L}(x, y) = f(x) - c(x)^T y$, where $y$ is an $m$-vector of dual variables. The Hessian of the Lagrangian with respect to $x$ is $H(x, y) = H(x) - \sum_{i=1}^{m} y_i H_i(x)$.

## 2 The Hestenes-Powell augmented Lagrangian

In its most commonly-used form, the Hestenes-Powell augmented Lagrangian function for problem (NEP) is given by

$$\mathcal{L}_A(x\,; y_e, \mu) = f(x) - c(x)^T y_e + \frac{1}{2\mu}\|c(x)\|^2,$$

where $\mu$ is a positive penalty parameter, $x$ is the vector of primal variables and $y_e$ is an approximate Lagrange multiplier vector. If $y_e$ is chosen to be a Lagrange multiplier vector $y^*$ of (NEP), then the associated solution $x^*$ is a stationary point of $\mathcal{L}_A$. Moreover, if the second-order sufficient conditions for optimality hold, then there exists a positive $\bar{\mu}$ such that $x^*$ is an isolated unconstrained minimizer of $\mathcal{L}_A$ for all $0 < \mu < \bar{\mu}$. Based on this result, Hestenes and Powell proposed that $x^*$ be found by minimizing a sequence of augmented Lagrangian functions $\mathcal{L}_A(x\,; y_k, \mu_k)$ in which a Lagrange multiplier estimate $y_k$ is defined in terms of a minimizer of $\mathcal{L}_A(x\,; y_{k-1}, \mu_{k-1})$.

For given $\mu$ and $y_e$, the function $\mathcal{L}_A(x) \triangleq \mathcal{L}_A(x\,; y_e, \mu)$ may be minimized using standard trust-region or line-search methods for unconstrained optimization. Here we focus on the properties of line-search methods, but much of the discussion may be extended to the formulation of trust-region methods. (For example, see Gertz and Gill [23] for a primal-dual penalty method in the trust-region setting.) In a typical

line-search method, if $\nabla^2\mathcal{L}_A(x)$ is sufficiently positive definite, a direction $p$ is found by solving the subproblem

$$\underset{p\in\mathbb{R}^n}{\text{minimize}}\ \nabla\mathcal{L}_A(x)^T p + \tfrac{1}{2}p^T\nabla^2\mathcal{L}_A(x)\,p.$$

In this case, $p$ is the unique solution of the Newton equations $\nabla^2\mathcal{L}_A(x)\,p = -\nabla\mathcal{L}_A(x)$. The gradient and Hessian of $\mathcal{L}_A(x)$ may be written in terms of the $m$-vector $\pi(x)$ such that

$$\pi(x) = y_e - \frac{1}{\mu}c(x). \tag{2.1}$$

With this definition of $\pi(x)$, we have

$$\nabla\mathcal{L}_A(x) = g(x) - J(x)^T\pi(x) \quad \text{and} \quad \nabla^2\mathcal{L}_A(x) = H\big(x,\pi(x)\big) + \frac{1}{\mu}J(x)^T J(x),$$

and the Newton equations may be written in the form

$$\left(H\big(x,\pi(x)\big) + \frac{1}{\mu}J(x)^T J(x)\right)p = -\big(g(x) - J(x)^T\pi(x)\big). \tag{2.2}$$

The elements of $\pi(x)$ may be viewed as approximate Lagrange multipliers, and are referred to as *first-order primal multiplier estimates*.

We motivate the derivation of the primal-dual augmented Lagrangian by showing that the Newton direction for minimizing the conventional augmented Lagrangian satisfies a "primal-dual" system in which the change in the dual variables may be specified arbitrarily.

**Lemma 2.1** *Given an* arbitrary *$m$-vector $y$, the Newton direction $p$ associated with minimizing the Hestenes-Powell augmented Lagrangian satisfies the equations*

$$\begin{pmatrix} H\big(x,\pi(x)\big) & J(x)^T \\ J(x) & -\mu I \end{pmatrix}\begin{pmatrix} p \\ -q \end{pmatrix} = -\begin{pmatrix} g(x) - J(x)^T y \\ c(x) + \mu(y - y_e) \end{pmatrix}, \tag{2.3}$$

*where $q$ depends on the value of $y$.*

*Proof* Define $J = J(x)$, $g = g(x)$, $c = c(x)$, $H = H(x,\pi)$, and $\pi = \pi(x)$. Then the Newton equations (2.2) may be written as

$$\left(H + \frac{1}{\mu}J^T J\right)p = -\big(g - J^T\pi\big). \tag{2.4}$$

For a given choice of $y$, consider the $m$-vector $q$ such that

$$q = -\frac{1}{\mu}\left(\big(Jp + \big(c + \mu(y - y_e)\big)\big)\right). \tag{2.5}$$

Equations (2.4) and (2.5) may be combined to give

$$\begin{pmatrix} H + \frac{2}{\mu} J^T J & J^T \\ J & \mu I \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = - \begin{pmatrix} g - J^T y + 2J^T (y - \pi) \\ \mu (y - \pi) \end{pmatrix},$$

where $\pi = y_e - c/\mu$ (see (2.1)). Applying the nonsingular matrix

$$\begin{pmatrix} I_n & -\frac{2}{\mu} J^T \\ 0 & I_m \end{pmatrix}$$

to both sides of this equation yields

$$\begin{pmatrix} H & J^T \\ J & -\mu I \end{pmatrix} \begin{pmatrix} p \\ -q \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ c + \mu (y - y_e) \end{pmatrix}. \qquad \square$$

If $y$ is chosen as an approximate Lagrange multiplier vector, then $y + q$ may be interpreted as the updated Lagrange multiplier estimate associated with $x + p$. In particular, if $\mu = 0$ and $H = H(x, y)$, then $p$ and $q$ are the primal-dual SQP directions at $(x, y)$ defined by one step of Newton's method for a zero of the gradient of the Lagrangian. Alternatively, if we choose $y = y_e$ in (2.3), the Newton direction $p$ satisfies the equations

$$\begin{pmatrix} H(x, \pi(x)) & J(x)^T \\ J(x) & -\mu I \end{pmatrix} \begin{pmatrix} p \\ -q \end{pmatrix} = - \begin{pmatrix} g(x) - J(x)^T y_e \\ c(x) \end{pmatrix},$$

which may be interpreted as a primal-dual variant of the primal Newton equations (2.2) analogous to the primal-dual formulation of the quadratic penalty method considered by Gould [27] (for related methods, see Murray [37, 38] and Biggs [3]). The nonzero $(2, 2)$ block in the primal-dual matrix serves to *regularize* the system; i.e., it is not necessary for $J(x)$ to have full row rank for the Newton equations to be nonsingular.

In conventional implementations of the augmented Lagrangian method, the direction $q$ is not used. The motivation for the generalized primal-dual augmented Lagrangian considered in the next section is the ability to exploit changes in both the primal *and* the dual variables during the unconstrained minimization.

## 3 The generalized primal-dual augmented Lagrangian

In this section we propose an augmented Lagrangian that is minimized with respect to both the primal and the dual variables. The *generalized primal-dual augmented Lagrangian* is

$$\mathcal{M}^v(x, y; y_e, \mu) = f(x) - c(x)^T y_e + \frac{1}{2\mu} \|c(x)\|^2 + \frac{v}{2\mu} \|c(x) + \mu (y - y_e)\|^2, \quad (3.1)$$

where $y_e$ is an approximate Lagrange multiplier vector, $v$ is a fixed scalar and $\mu$ is a positive parameter. The function $\mathcal{M}^v$ is equivalent to the Forsgren-Gill primal-dual

quadratic penalty function [19] defined in terms of the shifted constraints $c(x) - \mu y_e = 0$. (This derivation reflects Powell's observation that for sufficiently small $\mu$, $x^*$ is a minimizer of the quadratic penalty function defined with shifted constraints $c(x) - \mu y^* = 0$.)

Using the $m$-vector $\pi(x) = y_e - c(x)/\mu$ of (2.1), the gradient and Hessian for $\mathcal{M}^\nu(x, y; y_e, \mu)$ may be written as

$$\nabla\mathcal{M}^\nu(x, y; y_e, \mu) = \begin{pmatrix} g - J^T\big(\pi + \nu(\pi - y)\big) \\ \nu\big(c + \mu(y - y_e)\big) \end{pmatrix} = \begin{pmatrix} g - J^T\big(\pi + \nu(\pi - y)\big) \\ \nu\mu(y - \pi) \end{pmatrix}, \tag{3.2a}$$

and

$$\nabla^2\mathcal{M}^\nu(x, y; y_e, \mu) = \begin{pmatrix} H\big(x, \pi + \nu(\pi - y)\big) + \frac{1}{\mu}(1 + \nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix}, \tag{3.2b}$$

where $J$, $g$, $c$, and $\pi$ denote $J(x)$, $g(x)$, $c(x)$, and $\pi(x)$, respectively. Observe that the first-order multipliers $\pi(x) = y_e - c(x)/\mu$ minimize $\mathcal{M}^\nu(x, y; y_e, \mu)$ with respect to $y$ for a fixed value of $x$.

The next result emphasizes the potential role of $\mathcal{M}^\nu$ as the objective function in a sequential unconstrained minimization method for solving constrained problems. It states that a solution $(x^*, y^*)$ of problem (NEP) is a *minimizer* of $\mathcal{M}^\nu(x, y; y^*, \mu)$ for $\mu$ sufficiently small and all positive $\nu$.

**Theorem 3.1** *Assume that $(x^*, y^*)$ satisfies the following optimality conditions associated with problem* (NEP):

(i) $c(x^*) = 0$,
(ii) $g(x^*) - J(x^*)^T y^* = 0$, *and*
(iii) *there exists a positive scalar $\omega$ such that $p^T H(x^*, y^*)p \geq \omega\|p\|^2$ for all $p$ satisfying $J(x^*)p = 0$.*

*Then $(x^*, y^*)$ is a stationary point of the primal-dual function*

$$\mathcal{M}^\nu(x, y; y^*, \mu) = f(x) - c(x)^T y^* + \frac{1}{2\mu}\|c(x)\|^2 + \frac{\nu}{2\mu}\|c(x) + \mu(y - y^*)\|^2.$$

*Moreover, if $\nu > 0$, then there exists a positive scalar $\bar{\mu}$ such that $(x^*, y^*)$ is an isolated unconstrained minimizer of $\mathcal{M}^\nu(x, y; y^*, \mu)$ for all $0 < \mu < \bar{\mu}$.*

*Proof* We must show that $\nabla\mathcal{M}^\nu$ is zero and $\nabla^2\mathcal{M}^\nu$ is positive definite at the primal-dual point $(x, y) = (x^*, y^*)$. Assumption (i) and the definition $\pi(x) = y^* - c(x)/\mu$ implies that $\pi(x^*) = y^*$. Substituting for $\pi$, $x$ and $y$ in the gradient (3.2a) and using assumption (ii), gives $\nabla\mathcal{M}^\nu(x^*, y^*; y^*, \mu) = 0$ directly. Similarly, the Hessian (3.2b) is given by

$$\nabla^2\mathcal{M}^\nu = \begin{pmatrix} H + \frac{1}{\mu}(1 + \nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I_m \end{pmatrix},$$

where $\nabla^2\mathcal{M}^\nu = \nabla^2\mathcal{M}^\nu(x^*, y^*; y^*, \mu)$, $J = J(x^*)$, and $H = H(x^*, y^*)$.

It may be verified by direct multiplication that the matrix $L$ such that

$$L = \begin{pmatrix} I_n & 0 \\ -\frac{1}{\mu}J & I_m \end{pmatrix} \quad \text{gives} \quad L^T \nabla^2 \mathcal{M}^\nu L = \begin{pmatrix} H + \frac{1}{\mu}J^T J & 0 \\ 0 & \nu\mu I_m \end{pmatrix}.$$

As $L$ is nonsingular, we may apply Sylvester's Law of Inertia to infer that

$$\text{In}\left(L^T \nabla^2 \mathcal{M}^\nu L\right) = \text{In}\left(\nabla^2 \mathcal{M}^\nu\right) = (m, 0, 0) + \text{In}\left(H + \frac{1}{\mu}J^T J\right),$$

for all $\nu > 0$.

Let $r$ denote the rank of $J$, so that $r \leq \min(m, n)$. The singular-value decomposition of $J$ can be written as

$$J = USV^T = U \begin{pmatrix} S_r & 0 \\ 0 & 0 \end{pmatrix} V^T,$$

where $U$ and $V$ are orthogonal, and $S_r$ is an $r \times r$ diagonal matrix with positive diagonal entries. If the columns of $U$ and $V$ are partitioned to conform with the zero and nonzero columns of $S$, then $U = (U_r \;\; U_{m-r})$ and $V = (V_r \;\; V_{n-r})$, which gives $J = U_r S_r V_r^T$. The $n \times n$ matrix $Q$ defined such that $Q = (V_{n-r} \;\; V_r S_r^{-1})$ is nonsingular, with $JQ = (0 \;\; U_r)$. If we define $Z = V_{n-r}$ and $Y = V_r S_r^{-1}$, then $Q = (Z \;\; Y)$, with the $n - r$ columns of $Z$ forming a basis for the null-space of $J$. As $Q$ is nonsingular, $H + \frac{1}{\mu}J^T J$ must have the same inertia as $Q^T(H + \frac{1}{\mu}J^T J)Q$ from Sylvester's Law of Inertia. Pre- and post-multiplying $H + \frac{1}{\mu}J^T J$ by $Q^T$ and $Q$ gives

$$Q^T\left(H + \frac{1}{\mu}J^T J\right)Q = Q^T H Q + \frac{1}{\mu}Q^T J^T J Q = \begin{pmatrix} Z^T H Z & Z^T H Y \\ Y^T H Z & Y^T H Y + \frac{1}{\mu}I_r \end{pmatrix}.$$

This matrix has the form

$$\begin{pmatrix} H_{11} & H_{21}^T \\ H_{21} & H_{22} + \frac{1}{\mu}I_r \end{pmatrix},$$

where $H_{11} = Z^T H Z$, $H_{21} = Y^T H Z$ and $H_{22} = Y^T H Y$. As $H_{11}$ is positive definite by assumption (iii), we may write the block $2 \times 2$ matrix as the product

$$\begin{pmatrix} I_{n-r} & 0 \\ H_{21}H_{11}^{-1} & I_r \end{pmatrix} \begin{pmatrix} H_{11} & 0 \\ 0 & H_{22} - H_{21}H_{11}^{-1}H_{21}^T + \frac{1}{\mu}I_r \end{pmatrix} \begin{pmatrix} I_{n-r} & H_{11}^{-1}H_{21}^T \\ 0 & I_r \end{pmatrix}.$$

Repeated use of Sylvester's Law of Inertia then gives the inertia of $H + \frac{1}{\mu}J^T J$ as the inertia of $\text{diag}(H_{11}, H_{22} - H_{21}H_{11}^{-1}H_{21}^T + \frac{1}{\mu}I_r)$. Clearly, this matrix is positive definite for all $0 < \mu < \bar{\mu}$, where $\bar{\mu}$ satisfies $\bar{\mu} = 1/\max\{-\lambda_{\min}, 1\}$ and $\lambda_{\min}$ is the least eigenvalue of $H_{22} - H_{21}H_{11}^{-1}H_{21}^T$. Hence

$$\text{In}\left(\nabla^2 \mathcal{M}^\nu\right) = (m, 0, 0) + (n, 0, 0) = (m + n, 0, 0),$$

which implies that the Hessian $\nabla^2\mathcal{M}^\nu(x^*, y^*; y^*, \mu)$ is positive definite for all $\nu > 0$ and all $0 < \mu < \bar{\mu}$. It follows that $(x^*, y^*)$ is an isolated unconstrained minimizer of $\mathcal{M}^\nu(x, y; y^*, \mu)$. □

Theorem 3.1 suggests that if a sufficiently accurate estimate $y_e \approx y^*$ is known for problem (NEP), then an approximate minimizer of $\mathcal{M}^\nu(x, y; y_e, \mu)$ with respect to both $x$ and $y$ will provide an even better estimate (for a proof of this result, see Robinson [45]). An analogous result holds for the conventional augmented Lagrangian (see Conn, Gould and Toint [10]).

Standard line-search or trust-region methods for unconstrained minimization may be used to find an unconstrained minimizer of $\mathcal{M}^\nu(x, y; y_e, \mu)$. As our goal is to develop second-order methods, it is of interest to consider the Newton equations for the primal-dual augmented Lagrangian. Using the derivatives (3.2a) and (3.2b) for $\mathcal{M}^\nu$, the Newton direction for the primal-dual augmented Lagrangian satisfies

$$\begin{pmatrix} H\big(x, \pi + \nu(\pi - y)\big) + \frac{1}{\mu}(1+\nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = -\begin{pmatrix} g - J^T\big(\pi + \nu(\pi - y)\big) \\ \nu\big(c + \mu(y - y_e)\big) \end{pmatrix},$$

where $p$ and $q$ are the Newton directions in the primal and dual variables.

### 3.1 Relationships between methods for problem (NEP)

The next result shows that the Newton equations above may be transformed into a system similar to the primal-dual equations (2.3) associated with the classical augmented Lagrangian function.

**Lemma 3.1** *Let $H$ denote an arbitrary symmetric matrix. The equations*

$$\begin{pmatrix} H + \frac{1}{\mu}(1+\nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = -\begin{pmatrix} g - J^T\big(\pi + \nu(\pi - y)\big) \\ \nu\big(c + \mu(y - y_e)\big) \end{pmatrix}, \tag{3.3}$$

*and*

$$\begin{pmatrix} H & J^T \\ J & -\mu I \end{pmatrix} \begin{pmatrix} p \\ -q \end{pmatrix} = -\begin{pmatrix} g - J^Ty \\ c + \mu(y - y_e) \end{pmatrix}. \tag{3.4}$$

*are equivalent for all $\nu \neq 0$, i.e., $(p, q)$ is a solution of (3.3) if and only if it is a solution of (3.4).*

*Proof* Multiplying both sides of (3.3) by the nonsingular matrix

$$U = \begin{pmatrix} I & -\frac{(1+\nu)}{\nu\mu}J^T \\ 0 & \frac{1}{\nu}I \end{pmatrix},$$

and scaling the last $m$ columns by $-1$ gives the result. □

If $\mu = 0$ and $H = H(x, y)$, equations (3.4) are identical to those used to define the step for a conventional SQP method. This property will be exploited in Sect. 5.3, where the formulation of a primal-dual SQP method is considered.

Several well-known functions are equivalent to the primal-dual augmented Lagrangian for appropriate choices of the parameters $y_e$ and $\nu$.

*The quadratic penalty function ($\nu \equiv 0,\ y_e \equiv 0$)* In this case, $\mathcal{M}^\nu$ is the same as the quadratic penalty function

$$\mathcal{P}(x\,;\mu) = f(x) + \frac{1}{2\mu}\|c(x)\|^2,$$

which is defined in terms of the primal variables only. The primal-dual form of the Newton equations analogous to (2.3) is given by

$$\begin{pmatrix} H(x,\pi) & J^T \\ J & -\mu I \end{pmatrix} \begin{pmatrix} p \\ -q \end{pmatrix} = -\begin{pmatrix} g - J^T y \\ c + \mu y \end{pmatrix}, \tag{3.5}$$

which have been studied by Biggs [3] and Gould [27]. Few competitive modern methods are based on the direct minimization of the quadratic penalty function, but several reliable and efficient methods are designed to emulate the quadratic penalty method when the set of optimal multipliers is unbounded (see, e.g., [10, 22, 40], and [45]).

*The proximal-point penalty function ($\nu \equiv -1,\ y_e \equiv 0$)* The function $\mathcal{M}$ is equivalent to the proximal-point penalty function

$$\mathcal{P}_P(x, y) = f(x) - c(x)^T y - \frac{\mu}{2}\|y\|^2.$$

This function has been used in the formulation of stabilized SQP methods (see, e.g., Hager [32] and Wright [55]). The Newton equations for a stationary point of $\nabla\mathcal{P}_P(x, y)$ are given by:

$$\begin{pmatrix} H(x,y) & J^T \\ J & -\mu I \end{pmatrix} \begin{pmatrix} p \\ -q \end{pmatrix} = -\begin{pmatrix} g - J^T y \\ c + \mu y \end{pmatrix}.$$

Note the similarities between these equations and the primal-dual equations (3.5) for the quadratic penalty function. However, the Newton direction $p$ is not the same because the Lagrangian Hessian is evaluated with different values of the multipliers.

As $\nu$ is negative, $\nabla^2\mathcal{P}_P(x, y)$ is indefinite and the associated proximal-point penalty function has an unbounded minimizer. In this case, a stationary point $(x^*, y^*)$ solves the min-max problem

$$\min_x \max_y \ f(x) - c(x)^T y - \frac{\mu}{2}\|y\|^2.$$

*The Hestenes-Powell augmented Lagrangian ($\nu \equiv 0$)* This is the conventional augmented Lagrangian

$$\mathcal{L}_A(x\,; y_e, \mu) = f(x) - c(x)^T y_e + \frac{1}{2\mu}\|c(x)\|^2.$$

**Table 1** Functions associated with $\mathcal{M}^v(x, y; y_e, \mu)$

| | $v$ | $y_e$ |
|---|---|---|
| Quadratic penalty function | 0 | 0 |
| Proximal-point penalty function | $-1$ | 0 |
| Hestenes-Powell augmented Lagrangian | 0 | $y_e$ |
| Proximal-point Lagrangian | $-1$ | $y_e$ |
| Primal-dual quadratic penalty function | 1 | 0 |
| Primal-dual augmented Lagrangian | 1 | $y_e$ |

Lemma 2.1 implies that if $\pi$ is substituted for $y$ in the Hessian associated with the primal-dual augmented Lagrangian system (3.4), then the Newton direction $p$ associated with the solution of the resulting modified Newton system

$$\begin{pmatrix} H(x,\pi) & J^T \\ J & -\mu I \end{pmatrix} \begin{pmatrix} p \\ -q \end{pmatrix} = -\begin{pmatrix} g - J^T y \\ c + \mu(y - y_e) \end{pmatrix},$$

is the Hestenes-Powell augmented Lagrangian direction given by (2.3).

*The proximal-point Lagrangian ($v \equiv -1$)* In this case, $\mathcal{M}^v$ is the proximal-point Lagrangian function

$$\mathcal{L}_P(x, y) = f(x) - c(x)^T y - \frac{\mu}{2}\|y - y_e\|^2,$$

which has been studied, for example, by Rockafellar in [47, 48]. Analogous to the proximal-point penalty function, a stationary point $(x^*, y^*)$ solves the min-max problem

$$\min_x \max_y \; f(x) - c(x)^T y - \frac{\mu}{2}\|y - y_e\|^2.$$

*The primal-dual quadratic penalty function ($v \equiv 1,\ y_e \equiv 0$)*

$$\mathcal{P}(x, y; \mu) = f(x) + \frac{1}{2\mu}\|c(x)\|^2 + \frac{1}{2\mu}\|c(x) + \mu y\|^2.$$

Methods based on the use of the primal-dual penalty function and its barrier function counterpart are discussed by Forsgren and Gill [19], and Gertz and Gill [23].

*The primal-dual augmented Lagrangian ($v \equiv 1$).*

$$\mathcal{M}(x, y; y_e, \mu) = f(x) - c(x)^T y_e + \frac{1}{2\mu}\|c(x)\|^2 + \frac{1}{2\mu}\|c(x) + \mu(y - y_e)\|^2.$$

This function is the basis of the primal-dual BCL, sLCL and SQP algorithms proposed in Sect. 5.

Table 1 summarizes the six functions discussed above.

## 4 Regularization by bounding the multipliers

Typically, augmented Lagrangian methods are based on the assumption that the multipliers of the subproblem remain bounded, or do not grow too rapidly relative to the inverse of the penalty parameter. In the primal setting these assumptions are enforced by making appropriate modifications to $\mu$ and $y_e$ after the completion of each subproblem. An attractive feature of the primal-dual augmented Lagrangian function is that bounds on the multipliers may be enforced explicitly during the solution of the subproblem. To develop this idea further, consider an algorithm that minimizes a sequence of problems of the form

$$\underset{x\in\mathbb{R}^n, y\in\mathbb{R}^m}{\text{minimize}}\ \mathcal{M}(x, y; y_e, \mu).$$

As this is an unconstrained problem in both the primal and dual variables, we can impose explicit artificial bounds on the dual variables, i.e., we can solve the subproblem:

$$\underset{x\in\mathbb{R}^n, y\in\mathbb{R}^m}{\text{minimize}}\ \mathcal{M}(x, y; y_e, \mu) \quad \text{subject to} \quad -\gamma e \le y \le \gamma e,$$

for some positive constant $\gamma$. A sequence of these subproblems may be solved for appropriate values of $\mu$ and $y_e$. If all the bounds on $y$ are inactive at the solution of a subproblem, then the minimizer lies on the path of unconstrained minimizers of $\mathcal{M}$. However, if $\gamma$ restricts the subproblem solutions, then a different problem is being solved. Active bound constraints may occur for two reasons. First, the magnitudes of the optimal multipliers $y^*$ may be bounded but larger than the current value of $\gamma$. In this case, a poor choice of $\gamma$ will inhibit the convergence of the subproblem to the point on the path of subproblem solutions. Second, the subproblem multipliers may not exist or may be unbounded—for example, the Mangasarian-Fromovitz constraint qualification may not hold (see, e.g., [35, 41]). In this situation, an explicit bound on the dual variables will prevent the multipliers from diverging to infinity.

The previous discussion makes it clear that if some components of $y$ are active at a subproblem solution, then $\mu$ must be decreased in order to obtain convergence. As $\mu$ approaches zero, the subproblems become similar to those of the quadratic penalty function. The idea is to choose $\mu$ and $\gamma$ so that the artificial bounds will stabilize the method when far from a solution without affecting the subproblems near $(x^*, y^*)$. In Sect. 5 we consider two algorithms that are formulated with these goals in mind.

If the artificial bounds are inactive, then the solution of the subproblem lies on the conventional path of minimizers. However, when some bounds on $y$ are active it is unclear which problem is being solved. The next result shows that constrained solutions are related to those obtained by minimizing an exact penalty function.

**Theorem 4.1** *Let $\gamma$ be a positive scalar. If $(\bar{x}, \bar{y}, \bar{w})$ is a solution of*

$$\underset{x\in\mathbb{R}^n, y\in\mathbb{R}^m}{\text{minimize}}\ \mathcal{M}(x, y; y_e, \mu) \quad \textit{subject to} \quad -\gamma e \le y \le \gamma e, \tag{4.1}$$

*where $\bar{w}$ are the multipliers for the constraints $-\gamma e \le y \le \gamma e$, then there exists a positive diagonal scaling matrix $P$ such that $\bar{x}$ is a minimizer of the nonsmooth un-*

*constrained problem*

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ f(x)+\|P\big(c(x)+\mu(\bar{y}-y_e)\big)\|_1. \tag{4.2}$$

*Proof* Let $\bar{\pi}=\bar{y}-2\bar{w}/\mu$ and define the diagonal scaling matrix $P=\text{diag}(\rho_1,\ \rho_2,\ \ldots,\ \rho_m)$ such that

$$\rho_i=\begin{cases}\bar{\pi}_i & \text{if } \bar{w}_i<0,\\ -\bar{\pi}_i & \text{if } \bar{w}_i>0,\\ |\bar{\pi}_i|+\epsilon & \text{if } \bar{w}_i=0,\end{cases}$$

where $\epsilon$ is any positive real number. It will be shown that the diagonals of $P$ are strictly positive. The nonsmooth unconstrained problem (4.2) is equivalent to the constrained problem

$$\begin{aligned}\underset{x\in\mathbb{R}^n,u\in\mathbb{R}^m,v\in\mathbb{R}^m}{\text{minimize}}\quad & f(x)+\sum_{i=1}^{m}\rho_i(u_i+v_i)\\ \text{subject to}\quad & c(x)+\mu(\bar{y}-y_e)-u+v=0,\quad u\geq 0,\quad v\geq 0.\end{aligned} \tag{4.3}$$

Let $\bar{u}=\max(\bar{w},0)$ and $\bar{v}=-\min(\bar{w},0)$. Also, define

$$\bar{z}_u=Pe+\bar{\pi}, \tag{4.4a}$$

$$\bar{z}_v=Pe-\bar{\pi}. \tag{4.4b}$$

It is now shown that $(\bar{x},\bar{u},\bar{v},\bar{\pi},\bar{z}_u,\bar{z}_v)$ is a solution of (4.3), where $\bar{\pi}$ is the multiplier vector for the equality constraint $c(x)+\mu(\bar{y}-y_e)-u+v=0$, $\bar{z}_u$ is the multiplier vector for $u\geq 0$, and $\bar{z}_v$ is the multiplier vector for $v\geq 0$. A solution $(\bar{x},\bar{y},\bar{w})$ of (4.1) must satisfy the optimality conditions

$$J(\bar{x})^T\big(2\pi(\bar{x})-\bar{y}\big)=g(\bar{x}), \tag{4.5a}$$

$$c(\bar{x})+\mu(\bar{y}-y_e)=\bar{w}, \tag{4.5b}$$

$$-\gamma e\leq\bar{y}\leq\gamma e, \tag{4.5c}$$

$$\min(\gamma e-\bar{y},\bar{y}+\gamma e,|\bar{w}|)=0, \tag{4.5d}$$

$$\bar{w}\cdot(\bar{y}+\gamma e)\leq 0, \tag{4.5e}$$

$$\bar{w}\cdot(\bar{y}-\gamma e)\leq 0, \tag{4.5f}$$

where $\pi(x)=y_e-c(x)/\mu$. The optimality conditions that must be verified for the point $(\bar{x},\bar{u},\bar{v},\bar{\pi},\bar{z}_u,\bar{z}_v)$ are:

C1. $\bar{u}\geq 0,\ \bar{v}\geq 0,\ \bar{z}_u\geq 0,\ \bar{z}_v\geq 0,\ \bar{u}\cdot\bar{z}_u=0,\ \bar{v}\cdot\bar{z}_v=0$;
C2. $c(\bar{x})+\mu(\bar{y}-y_e)-\bar{u}+\bar{v}=0$;
C3. $g(\bar{x})=J(\bar{x})^T\bar{\pi}$;
C4. $Pe=\bar{z}_u-\bar{\pi}$;
C5. $Pe=\bar{z}_v+\bar{\pi}$.

(Proof of C2): Note that $\bar{w} = \bar{u} - \bar{v}$. Thus C2 follows directly from (4.5b).
(Proof of C3): The definitions of $\bar{\pi}$ and $\pi(x)$ together with (4.5b) imply that

$$\bar{\pi} = \bar{y} - 2\bar{w}/\mu = 2y_e - 2c(\bar{x})/\mu - \bar{y} = 2\pi(\bar{x}) - \bar{y}. \tag{4.6}$$

Condition C3 follows from this equality and (4.5a).
(Proof of C4): Follows from definition (4.4a).
(Proof of C5): Follows from definition (4.4b).
(Proof of C1): $\bar{u} \geq 0$ and $\bar{v} \geq 0$ by definition.

Next it is shown that $\bar{u} \cdot \bar{z}_u = 0$. The result is trivial if $\bar{u}_i = 0$. If $\bar{u}_i \neq 0$, it must hold that $\bar{w}_i > 0$ and hence $\rho_i = -\bar{\pi}_i$. It follows that $[\bar{z}_u]_i \triangleq \rho_i + \bar{\pi}_i = 0$.

Now it is shown that $\bar{v} \cdot \bar{z}_v = 0$. The result is trivial if $\bar{v}_i = 0$. So suppose that $\bar{v}_i > 0$. This implies that $\bar{w}_i < 0$ and thus $\rho_i = \bar{\pi}_i$. It follows that $[\bar{z}_v]_i \triangleq \rho_i - \bar{\pi}_i = 0$.

The inequalities $\bar{z}_u \geq 0$, $\bar{z}_v \geq 0$, and $\rho_i > 0$ are established by considering the following cases.

1. Suppose $\bar{w}_i = 0$. Then $\rho_i = |\bar{\pi}_i| + \epsilon > 0$ and $[\bar{z}_v]_i = |\bar{\pi}_i| + \epsilon - \bar{\pi}_i > 0$. Similarly, $[\bar{z}_u]_i = |\bar{\pi}_i| + \epsilon + \bar{\pi}_i > 0$.
2. Suppose $\bar{w}_i > 0$. Then $\bar{y}_i = -\gamma$ and $\bar{\pi}_i = \bar{y}_i - 2\bar{w}_i/\mu = -\gamma - 2\bar{w}_i/\mu < 0$. This implies that $\rho_i = -\bar{\pi}_i > 0$ and that $[\bar{z}_v]_i = \rho_i - \bar{\pi}_i = -2\bar{\pi}_i > 0$. Similarly, $[\bar{z}_u]_i = \rho_i + \bar{\pi}_i = 0$.
3. Suppose $\bar{w}_i < 0$. Then $\bar{y}_i = \gamma$ and $\bar{\pi}_i = \bar{y}_i - 2\bar{w}_i/\mu = \gamma - 2\bar{w}_i/\mu > 0$. This implies that $\rho_i = \bar{\pi}_i > 0$ and that $[\bar{z}_v]_i = \rho_i - \bar{\pi}_i = 0$. Similarly, $[\bar{z}_u]_i = \rho_i + \bar{\pi}_i = 2\bar{\pi}_i > 0$.

The proof is complete since the point $(\bar{x}, \bar{u}, \bar{v}, \bar{\pi}, \bar{z}_u, \bar{z}_v)$ satisfies C1–C5. □

### 4.1 Interpretation of the artificial bounds

In Sect. 5 we consider two subproblems (5.2) and (5.4) that bound the dual variables explicitly using artificial constraints. In this section we give a brief description of one way in which these additional constraints may be interpreted.

Let $(\bar{x}, \bar{y}, \bar{w})$ denote a solution of the bound-constrained problem (4.1), where $\bar{w}$ is the multiplier vector for the simple bounds. Also, let $(x_\mu, y_\mu) = (x(\mu), y(\mu))$ denote the solution to the unconstrained problem

$$\underset{x,y}{\text{minimize}}\ \mathcal{M}(x, y; y_e, \mu). \tag{4.7}$$

If we apply an $\ell_1$ penalty to the bound constraints in problem (4.1), we obtain the equivalent problem

$$\underset{x,y}{\text{minimize}}\ \mathcal{M}(x, y; y_e, \mu) + \sigma \|y^v\|_1, \tag{4.8}$$

where $\sigma$ is a positive penalty parameter and $y^v = \min(0, \gamma - |y|)$ (the definition of $y^v$ should be interpreted componentwise and is a measure of how much $y$ violates its bounds). If $\sigma > \|\bar{w}\|_\infty$, it is well known that solutions of problem (4.8) are solutions

of (4.1) (see, e.g., [17]). The quantity $\|\bar{w}\|_\infty$ may be regarded as the "required penalization". It follows from the optimality conditions for problems (4.1) and (4.7) that $\bar{w} = c(\bar{x}) - c(x_\mu) + \mu(\bar{y} - y_\mu)$, which implies that the required penalization satisfies

$$\|\bar{w}\|_\infty \leq \|c(\bar{x}) - c(x_\mu)\|_\infty + \mu\|\bar{y} - y_\mu\|_\infty. \tag{4.9}$$

This bound implies that the required penalization is closely associated with the magnitudes of $\|c(\bar{x}) - c(x_\mu)\|_\infty$ and $\mu\|\bar{y} - y_\mu\|_\infty$, which are zero if the artificial bounds are inactive.

The discussion above implies that the artificial bounds in problem (4.1) may be interpreted as a second form of regularization—the first being the presence of $\mu I$ in the $(2, 2)$ block of the Newton equations. In this second regularization, $|y|$ is bounded explicitly by problem (4.1) and implicitly by the penalty term in problem (4.8). Specifically, the $\mu\|\bar{y} - y_\mu\|_\infty$ term in (4.9) implies that if the artificial bounds prevent the "natural" solution from being found, then the required penalization is likely to be large. However, the presence of the $\mu$-term makes this implicit penalization diminish as $\mu$ is decreased to zero. Similarly, the presence of the $\|c(\bar{x}) - c(x_\mu)\|_\infty$ term in (4.9) implies that the required penalization is likely to be large if the constraint values differ substantially. For small $\mu$, the minimizers of the merit function will be close to minimizers of the quadratic penalty function. In this case, $\|c(\bar{x})\|_\infty$ and $\|c(x_\mu)\|_\infty$ can be expected to be small (and hence the term $\|c(\bar{x}) - c(x_\mu)\|_\infty$ will be small).

The previous discussion generalizes to the case where each dual variable is given a separate bound in problem (4.1). We have the following componentwise result.

**Theorem 4.2** *If $(\bar{x}, \bar{y}, \bar{w})$ is a primal-dual solution of*

$$\underset{x,y}{\text{minimize}}\ \mathcal{M}(x, y; y_e, \mu) \quad \textit{subject to} \quad y_\ell \leq y \leq y_u,$$

*then $(\bar{x}, \bar{y})$ minimizes $\mathcal{M}(x, y; y_e, \mu) + \|D(\bar{w})y^v\|_1$, where $D(\bar{w}) = \text{diag}(d_1, d_2, \ldots, d_m)$ and $d_i \geq \bar{w}_i$ for all $i = 1, 2, \ldots, m$.*

*Proof* The result follows from the standard properties of the $\ell_1$ penalty function (see, e.g., [10] and [17]). □

## 5 Algorithms

The augmented Lagrangian has been used successfully within a number of different algorithmic frameworks. In the context of problem (NP), the software package LANCELOT [9] finds an approximate minimizer of a sequence of bound constrained Lagrangian (BCL) subproblems of the form

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ \mathcal{L}_A(x\,; y_e, \mu) \quad \text{subject to} \quad x \geq 0. \tag{5.1}$$

After each approximate minimization, the Lagrange multiplier estimate $y_e$ may be updated, while parameters and tolerances are adjusted. Conn, Gould, and Toint [10] show that under certain standard assumptions, the BCL method is globally convergent,

exhibits R-linear convergence, and has a penalty parameter $\mu$ that is bounded away from zero.

Linearly constrained Lagrangian (LCL) methods also make use of the augmented Lagrangian. LCL methods are based on the properties of Robinson's method [46], which sequentially minimizes the Lagrangian $\mathcal{L}(x\,;\,y_e) = f(x) - c(x)^T y_e$, subject to the linearized constraints. Robinson shows that this algorithm exhibits R-quadratic convergence when started sufficiently close to a solution satisfying the second-order sufficient conditions. Two potential drawbacks of the LCL method are that the linearly constrained subproblem may be infeasible and that convergence is guaranteed only in a neighborhood of the solution. Murtagh and Saunders [39, 40] include a penalty term in their LCL software package MINOS in order to encourage convergence from poor starting points (i.e., MINOS uses an *augmented* Lagrangian instead of the Lagrangian). This modification improves the robustness of Robinson's method, but the question of convergence from arbitrary starting points remains open.

The stabilized LCL (sLCL) method of Friedlander and Saunders [22] includes further improvements to the method of MINOS. The sLCL algorithm is provably globally convergent and uses a formulation of the subproblem that is always feasible. These improvements result from the use of a so-called elastic subproblem, which is equivalent to an $\ell_1$ penalization of the linearized constraint violations. Friedlander and Saunders show that the sLCL algorithm constitutes a range of algorithms, with the BCL method at one extreme and Robinson's LCL method at the other. The sLCL algorithm inherits global convergence from the BCL method and R-quadratic convergence from the LCL method.

The augmented Lagrangian function may also be used as a merit function in an SQP method. In a typical line-search SQP method, the search direction is the solution of a quadratic programming subproblem involving a quadratic model of the Lagrangian and the linearized constraints (see, e.g., Han [33] and Powell [43]). A similar SQP approach starts with an unconstrained quadratic model derived from second-order Taylor-series approximation of $\mathcal{L}_A(x; y_e, \mu)$. This unconstrained subproblem is then solved as an equivalent quadratic program (for more details, see [11, Section 15.3.1]). The SQP algorithm in the software package SNOPT [24] uses an augmented Lagrangian merit function in a different way. Given a primal-dual approximate solution, the method of SNOPT solves a convex QP subproblem defined using a quasi-Newton approximation of the Hessian of the Lagrangian. A search direction in both the primal *and* dual variables is then defined from the primal and dual solution of the QP subproblem. The dual variables are included in the line search to make the augmented Lagrangian continuous in both the primal and dual variables. This property is used to force convergence from arbitrary starting points (see Gill, Murray, Saunders and Wright [26]).

In the remainder of this section we illustrate how the primal-dual augmented Lagrangian may be used within various algorithmic frameworks for problem (NP). In the case of the SQP method, some preliminary numerical experiments are used to justify the further development of these methods. A complete theoretical treatment is currently under investigation.

### 5.1 Primal-dual bound-constrained Lagrangian methods

Problem (NP) may be solved as a sequence of bound constrained problems of the form

$$\underset{x\in\mathbb{R}^n,\, y\in\mathbb{R}^m}{\text{minimize}}\ \mathcal{M}(x, y\,; y_k^e, \mu_k) \quad \text{subject to} \quad x \geq 0, \quad -\gamma_k e \leq y \leq \gamma_k e, \tag{5.2}$$

where $\mathcal{M}(x, y\,; y_k^e, \mu_k)$ is the primal-dual augmented Lagrangian of Sect. 3.1, $\{\gamma_k\}$ is a positive sequence of parameters, $\{y_k^e\}$ is a sequence of Lagrange multiplier estimates satisfying $y_k^e \in [-\gamma_k e, \gamma_k e]$, and $\{\mu_k\}$ is a sequence of positive penalty parameters. One possible choice of $\gamma_k$ is $\max(1/\mu_k, \gamma_{\max})$, where $\gamma_{\max}$ is a preassigned anticipated bound on the magnitude of the multipliers, e.g., $\gamma_{\max} = 10^5$. An approximate solution of the subproblem (5.2) is denoted by $(x_k^*, y_k^*, z_k^*, w_k^*)$, where $z_k^*$ and $w_k^*$ are the Lagrange multipliers for the inequality constraints $x \geq 0$ and $-\gamma_k e \leq y \leq \gamma_k e$.

Conventional BCL methods are known to be locally convergent if the penalty parameter is sufficiently small and each subproblem is solved exactly. Bertsekas [1] extends this result by showing that only an approximate solution of each BCL subproblem need be found. In both cases it may be necessary to drive the penalty parameter to zero to guarantee global convergence. In this case, solutions of the BCL subproblems are similar to those of the quadratic penalty method.

The formulation of a pdBCL algorithm based on solving a sequence of subproblems (5.2) requires that both the penalty parameter $\mu_k$ and the Lagrange multiplier estimate $y_k^e$ be updated. The strategy used by Conn, Gould, and Toint [10] for the LANCELOT package [9] is still appropriate in this case. However, in the primal-dual case, the new multiplier estimate is $y_{k+1}^e = 2\pi(x_k^*) - y_k^*$, with $\pi(x_k^*) = y_k^e - c(x_k^*)/\mu_k$. This update is based on the first-order optimality conditions for the *primal-dual* augmented Lagrangian. (See [45, Chapter 4] for further details of the global and local convergence properties of the pdBCL method.)

### 5.2 Stabilized primal-dual LCL methods

Problem (NP) may be solved as a sequence of linearly constrained subproblems. In the primal-dual setting, given an estimate $(x_k, y_k)$ of a solution to problem (NP), the subproblems take the form

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n,\, y\in\mathbb{R}^m}{\text{minimize}} & \mathcal{M}(x, y\,; y_k^e, \mu_k) \\ \text{subject to} & \bar{c}_k(x) = 0, \quad x \geq 0, \quad -\gamma_k e \leq y \leq \gamma_k e, \end{array} \tag{5.3}$$

where $\gamma_k$ is a positive parameter, $y_k^e$ is an estimate of the Lagrange multiplier vector, $\bar{c}_k(x) \triangleq c(x_k) + J(x_k)(x - x_k)$ is a linearization of the constraints, and $\mu_k$ is the $k$th penalty parameter. As in pdBCL, the bounds on the dual variables may be defined as $\gamma_k = \max(1/\mu_k, \gamma_{\max})$.

The subproblem (5.3) may be unsuitable for two reasons. First, the constraints $\bar{c}_k(x) = 0$ and $x \geq 0$ may be infeasible. Second, the distance $\|x_k - x_k^*\|$ from the point of linearization to the solution of the subproblem may be arbitrarily large. These

problems are addressed by Friedlander and Saunders [22], who regularize the standard LCL subproblem by including an $\ell_1$ penalty on the linearized constraint violations. The analogous approach for the primal-dual augmented Lagrangian gives the so-called *elastic subproblem*:

$$\begin{array}{ll} \underset{x,y,u,v}{\text{minimize}} & \mathcal{M}(x, y; y_k^e, \mu_k) + \sigma_k e^T(u+v) \\ \text{subject to} & \bar{c}_k(x) + u - v = 0, \quad x, u, v \geq 0, \quad -\gamma_k e \leq y \leq \gamma_k e. \end{array} \tag{5.4}$$

An approximate solution of this problem is denoted by $(x_k^*,\ y_k^*,\ u_k^*,\ v_k^*,\ \Delta y_k^*,\ z_k^*,\ w_k^*)$, where $\Delta y_k^*$ is the (approximate) multiplier vector for the linearized constraints, and $z_k^*$ is the (approximate) multiplier vector for the bounds $x \geq 0$. The subproblem (5.4) forms the basis of a pd$\ell_1$LCL method that uses the updating strategy for $\mu_k$ and $y_k^e$ proposed by Friedlander and Saunders [22]. As in the pdBCL algorithm, a "primal-dual" update is defined for the new multiplier estimate. In this case, $y_{k+1}^e = 2\pi(x_k^*) - y_k^* + \Delta y_k^*$, where $\pi(x_k^*) = y_k^e - c(x_k^*)/\mu_k$. This definition is based on the first-order optimality conditions for problem (5.4). (See [45, Chapter 5] for further details of the global and local convergence properties of the pd$\ell_1$LCL method.)

### 5.3 Primal-dual SQP methods

Some of the most efficient algorithms for nonlinear optimization are sequential quadratic programming (SQP) methods. This class of methods provides an important application of the primal-dual function considered here. In particular, the primal-dual augmented Lagrangian may be used as a merit function to force convergence to points satisfying the second-order necessary conditions for optimality. This is a consequence of Theorem 3.1, which shows that minimizers of problem (NEP) are also minimizers of the primal-dual augmented Lagrangian function. In what remains of this section, we propose an SQP method for (NEP) that uses $\mathcal{M}(x, y; y_e, \mu)$ as a merit function.

If $\mu = 0$ and $H = H(x, y)$, equations (3.4) are identical to the conventional equations for the SQP step. This provides the motivation for using different penalty parameters for the step computation and the merit function. In particular, given an iterate $v_k = (x_k, y_k)$ and Lagrange multiplier estimate $y_k^e$, the search direction $\Delta v_k = (p_k, q_k)$ is computed from the equations

$$\begin{pmatrix} \bar{H}_k & J_k^T \\ J_k & -\mu_R I \end{pmatrix} \begin{pmatrix} p_k \\ -q_k \end{pmatrix} = -\begin{pmatrix} g_k - J_k^T y_k \\ c_k + \mu_R(y_k - y_k^e) \end{pmatrix}, \tag{5.5}$$

where $\bar{H}_k$ is an approximate Lagrangian Hessian and $\mu_R$ is a small *fixed* parameter. In this context, $\mu_R$ plays the role of a *regularization* parameter rather than a *penalty* parameter, thereby providing an $O(\mu_R)$ estimate of the conventional SQP direction. In the numerical experiments given below, $\bar{H}_k = H(x_k, y_k) + E_k$, where $E_k$ is a positive-semidefinite modification chosen to ensure that the inertia of the regularized equations (5.5) is $(n, m, 0)$. If the inertia is correct, then $E_k = 0$; otherwise $E_k$ is defined implicitly by modifying the eigenvalues associated with the spectral decomposition of $H(x_k, y_k)$ (see Greenstadt [31]). Other, more practical approaches

include: (i) modifying an inertia-controlling factorization of the KKT matrix [19, 21]; (ii) using a positive-definite quasi-Newton approximation to $H(x_k, y_k)$ [24, 29, 30, 44]; and (iii) adding increasing positive multiples of the identity matrix to $H(x_k, y_k)$ until the inertia is correct [53].

Once the search direction $\Delta v_k$ has been determined, a penalty parameter $\mu_k$ is defined such that $\mu_k \in [\mu_R, \mu_{k-1}]$ and $\Delta v_k$ is a descent direction for $\mathcal{M}(x, y; y_k^e, \mu_k)$ at $(x_k, y_k)$. Specifically, $\mu_k = \max(10^{-\ell}\mu_{k-1}, \mu_R)$, where $\ell \geq 0$ is the smallest integer such that

$$\Delta v_k^T \nabla \mathcal{M}(v_k; y_k^e, \mu_k) \leq \max\left(\frac{1}{4}\Delta v_k^T \nabla \mathcal{M}(v_k; y_k^e, \mu_R), -10^{-3}\|\Delta v_k\|^2\right), \tag{5.6}$$

and $\mathcal{M}(v; y_e, \mu)$ denotes the merit function evaluated at $v = (x, y)$. As the inertia of the matrix in (5.5) is correct (as described above), the primal-dual direction satisfies $\Delta v_k^T \nabla \mathcal{M}(v_k; y_k^e, \mu_R) \leq 0$, which implies that the right-hand side of (5.6) is nonpositive. Condition (5.6) holds for $\mu_k = \mu_R$ and so it is always possible to find a $\mu_k$ that satisfies the descent condition (5.6). Once $\mu_k$ has been determined, a "flexible" backtracking line search is performed on the primal-dual augmented Lagrangian. A conventional backtracking line search defines $v_{k+1} = v_k + \alpha_k \Delta v_k$, where $\alpha_k = 2^{-j}$ and $j$ is the smallest nonnegative integer such that

$$\mathcal{M}(v_k + \alpha_k \Delta v_k; y_k^e, \mu_k) \leq \mathcal{M}(v_k; y_k^e, \mu_k) + \alpha_k \eta_S \Delta v_k^T \nabla \mathcal{M}(v_k; y_k^e, \mu_k)$$

for a given scalar $\eta_S \in (0, 1)$. However, this approach would suffer from the Maratos effect [36] simply because the penalty parameter $\mu_k$ and the regularization parameter $\mu_R$ generally have different values. Thus we use a "flexible penalty function" approach proposed by Curtis and Nocedal [12] and define $\alpha_k = 2^{-j}$, where $j$ is the smallest nonnegative integer such that

$$\mathcal{M}(v_k + \alpha_k \Delta v_k; y_k^e, \sigma) \leq \mathcal{M}(v_k; y_k^e, \sigma) + \alpha_k \eta_S \Delta v_k^T \nabla \mathcal{M}(v_k; y_k^e, \sigma) \tag{5.7}$$

for some value $\sigma \in [\mu_R, \mu_k]$. Once an appropriate value for $\alpha_k$ is found, the new primal-dual solution estimate is given by

$$x_{k+1} = x_k + \alpha_k p_k, \qquad y_{k+1} = y_k + \alpha_k q_k, \quad \text{and} \quad y_{k+1}^e = y_{k+1}.$$

The iteration is concluded by increasing the iteration counter $k$.

Numerical results from a simple MATLAB implementation of pdSQP were obtained for the nonlinear equality constrained problems from the CUTEr test collection (see Bongartz et al. [7] and Gould, Orban and Toint [28]). The MATLAB implementation was defined with the parameter values $\mu_R = 10^{-8}$, $\mu_{-1} = 1$, and $\eta_S = 10^{-1}$. The sequence $\{(x_k, y_k)\}$ was judged to have converged if

$$\max\left(\|c(x_k)\|, \|g(x_k) - J(x_k)^T y_k\|\right) < 10^{-6}. \tag{5.8}$$

If the inertia of (5.5) was not correct, the eigenvalues $\{\bar{\lambda}_j\}$ associated with the spectral decomposition of the $(1, 1)$-block $\bar{H}(x_k, y_k)$ in (5.5) were defined as $\bar{\lambda}_j =$

**Algorithm 5.1** Primal-dual equality constrained SQP algorithm (pdSQP)

Input: $(x_0, y_0)$;
Set parameters $\mu_R > 0$, $\mu_{-1} \in [\mu_R, \infty)$, and $\eta_S \in (0, 1)$;
Set $y_0^e = y_0$ and $k = 0$;
**while not** *converged* **do**
  Compute $f(x_k)$, $c(x_k)$, $g(x_k)$, $J(x_k)$, and $H(x_k, y_k)$;
  If necessary, modify $H(x_k, y_k)$ to ensure that the inertia of (5.5) is $(n, m, 0)$;
  Solve (5.5) for the search direction $\Delta v_k = (p_k, q_k)$;
  Choose $\mu_k$ to satisfy (5.6);
  Use a *flexible* backtracking line search to find an $\alpha_k$ satisfying (5.7);
  Update the primal-dual estimate $x_{k+1} = x_k + \alpha_k p_k, \quad y_{k+1} = y_k + \alpha_k q_k$;
  Update the multiplier estimate : $y_{k+1}^e = y_{k+1}; \quad k \leftarrow k + 1$;
**end do**

$\max\{\gamma_L, \lambda_j\}$, where $\{\lambda_j\}$ are the eigenvalues of $H(x_k, y_k)$. The lower bound $\gamma_L$ was $\gamma_L = \max\{1, \gamma_U/\texttt{condmax}\}$, where $\gamma_U = \max_{1 \le j \le n} \{|\lambda_j|, 1\}$ and `condmax` is a preassigned upper bound on the condition number of $\bar{H}(x_k, y_k)$. The value of `condmax` was $10^3$ in all cases.

In order to provide some context for the pdSQP results, the SQP package SNOPT [24] was applied to the same test set. The default SNOPT parameter values were used throughout. These include a termination criterion comparable (but not identical) to the pdSQP condition (5.8) above. However, it must be emphasized that the results should not be used in an attempt to assess the relative efficiency of pdSQP and SNOPT. The implementation of pdSQP is simply a verbatim MATLAB version of Algorithm 5.1, whereas SNOPT is an established package that incorporates many years of development. On the other hand, SNOPT uses only first derivatives and is implemented with different termination criteria (for more information on the implementation of SNOPT, see Gill, Murray and Saunders [25]).

A total of 80 equality-constrained CUTEr problems were identified: `arglcle`, `bt1`–`bt12`, `byrdsphr`, `coolhans`, `dixchlng`, `eigena2`, `eigenaco`, `eigenb2`, `eigenbco`, `eigenc2`, `eigencco`, `elec`, `gridnete`, `gridneth`, `hs6`–`hs9`, `hs26`–`hs28`, `hs39`–`hs40`, `hs42`, `hs46`–`hs52`, `hs56`, `hs61`, `hs77`–`hs79`, `hs100lnp`, `hs111lnp`, `lch`, `lukvle1`–`lukvle4`, `lukvle6`–`lukvle18`, `maratos`, `mss1`, `mwright`, `orthrdm2`, `orthrds2`, `orthrega`, `orthregb`, `orthregc`, `orthregd`, `orthrgdm`, `orthrgds`, `s316-322`, and `woodsne`.

Of these 80 problems, 9 were excluded from the test set: `arglcle`, `lukvle2`, `lukvle4`, `lukvle11`, `lukvle12`, `lukvle15`, `lukvle17`, `lukvle18`, and `woodsne`. Problems `arglcle` and `woodsne` have infeasible constraints. Problem `lukvle2` has an unbounded solution. Problems `lukvle4`, `lukvle11`, `lukvle12`, `lukvle15`, `lukvle17`, and `lukvle18` were excluded because they could not be solved by either SNOPT or pdSQP within 500 iterations.

Tables 2–4 give the details of runs on the remaining 71 problems. For each test problem we list the number of equality constraints ("$m$") and variables ("$n$"). In addition, we give the number of function evaluations ("fe") and iterations ("itns") needed by SNOPT and pdSQP. Both methods found the same local solution for the

**Table 2** Nonlinear equality constrained CUTEr problems (A–E)

| Problem | $m$ | $n$ | SNOPT | | pdSQP | |
|---|---|---|---|---|---|---|
| | | | fe | itns | fe | itns |
| bt1 | 1 | 2 | 21 | 10 | 9 | 6 |
| bt2 | 1 | 3 | 16 | 15 | 14 | 13 |
| bt3 | 3 | 5 | 7 | 6 | 2 | 1 |
| bt4 | 2 | 3 | 10 | 7 | 14 | 7 |
| bt5 | 2 | 3 | 11 | 8 | 7 | 5 |
| bt6 | 2 | 5 | 16 | 14 | 11 | 9 |
| bt7 | 3 | 5 | 36 | 19 | 57 | 41 |
| bt8 | 2 | 5 | 13 | 11 | 20 | 19 |
| bt9 | 2 | 4 | 30 | 18 | 23 | 14 |
| bt10 | 2 | 2 | 23 | 13 | 7 | 6 |
| bt11 | 3 | 5 | 14 | 11 | 8 | 7 |
| bt12 | 3 | 5 | 9 | 8 | 5 | 4 |
| byrdsphr | 2 | 3 | 14 | 10 | 59 | 30 |
| coolhans | 9 | 9 | 28 | 19 | 14 | 13 |
| dixchlng | 5 | 10 | 30 | 29 | 43 | 40 |
| eigena2 | 55 | 110 | 4 | 3 | 5 | 4 |
| eigenaco | 55 | 110 | 4 | 3 | 10 | 8 |
| eigenb2 | 55 | 110 | 4 | 3 | 37 | 22 |
| eigenbco | 55 | 110 | 4 | 3 | 125 | 74 |
| eigenc2 | 231 | 462 | 290 | 243 | 83 | 44 |
| eigencco | 231 | 462 | 253 | 208 | 93 | 48 |
| elec | 200 | 600 | 403 | 359 | 115 | 59 |

problems with more than one local solution. A table entry "$t$" indicates that the run was terminated after 500 iterations.

Algorithm pdSQP was unable to solve the 8 cases mss1, lukvle6, lukvle8, lukvle13, lukvle14, lukvle16, orthrds2, and orthrega within the allotted 500 iterations. However, pdSQP can solve orthrds2 in 200 function evaluations and 198 iterations if the convergence tolerance in (5.8) is increased from $10^{-6}$ to $10^{-5}$. If all of the difficult lukvle problems are omitted from the test set, pdSQP was able to solve all but three of the remaining 61 problems, many of which are nontrivial.

As is to be expected, the more mature package SNOPT was slightly more robust than pdSQP, with only 4 problems unsolved (lukvle9, lukvle10, lch, and orthrgds). However, pdSQP usually required fewer iterations than SNOPT on the problems for which both methods converged successfully. In particular, pdSQP required fewer iterations than SNOPT in 72% of the cases. Moreover, the number of function evaluations is typically not much greater than the number of iterations, which implies that the regularized SQP step was usually accepted by the backtracking line search. The performance of pdSQP on problem maratos is particularly encouraging because this problem is a known source of difficulty for SQP methods that suffer from the Maratos effect.

**Table 3** Nonlinear equality constrained CUTEr problems (F–H)

| Problem | $m$ | $n$ | SNOPT | | pdSQP | |
|---|---|---|---|---|---|---|
| | | | fe | itns | fe | itns |
| gridnete | 36 | 60 | 38 | 37 | 4 | 3 |
| gridneth | 36 | 60 | 73 | 72 | 6 | 5 |
| hs6 | 1 | 2 | 7 | 6 | 26 | 11 |
| hs7 | 1 | 2 | 30 | 17 | 11 | 8 |
| hs8 | 2 | 2 | 6 | 5 | 6 | 4 |
| hs9 | 1 | 2 | 8 | 6 | 5 | 4 |
| hs26 | 1 | 3 | 24 | 23 | 19 | 18 |
| hs27 | 1 | 3 | 23 | 20 | 15 | 10 |
| hs28 | 1 | 3 | 11 | 10 | 2 | 1 |
| hs39 | 2 | 4 | 30 | 18 | 23 | 14 |
| hs40 | 3 | 4 | 7 | 6 | 5 | 4 |
| hs42 | 2 | 4 | 8 | 6 | 6 | 5 |
| hs46 | 2 | 5 | 26 | 25 | 20 | 18 |
| hs47 | 3 | 5 | 31 | 22 | 22 | 17 |
| hs48 | 2 | 5 | 8 | 7 | 2 | 1 |
| hs49 | 2 | 5 | 32 | 31 | 17 | 16 |
| hs50 | 3 | 5 | 21 | 19 | 10 | 9 |
| hs51 | 3 | 5 | 8 | 7 | 2 | 1 |
| hs52 | 3 | 5 | 8 | 6 | 2 | 1 |
| hs56 | 4 | 7 | 14 | 10 | 6 | 5 |
| hs61 | 2 | 3 | 174 | 68 | 19 | 13 |
| hs77 | 2 | 5 | 14 | 12 | 11 | 9 |
| hs78 | 3 | 5 | 7 | 6 | 5 | 4 |
| hs79 | 3 | 5 | 14 | 11 | 5 | 4 |
| hs100lnp | 2 | 7 | 19 | 14 | 22 | 18 |
| hs111lnp | 3 | 10 | 104 | 49 | 34 | 18 |

We emphasize that the formulation of pdSQP proposed here has been chosen to illustrate the *local* rather than *global* properties of the merit function. Clearly, a more sophisticated updating strategy for $y_k^e$ is needed to ensure that pdSQP is globally convergent. Moreover, many of the cases that did not converge involved a substantial number of iterations in which the Hessian was modified. More efficient updates for $y_k^e$, and matrix modification schemes involving a trust-region strategy are currently under investigation and are beyond the scope of this paper.

## 6 Conclusion

Merit functions have played an important role in the formulation and analysis of methods for solving constrained optimization problems. In this paper we have introduced a generalized primal-dual augmented Lagrangian that may be minimized

**Table 4** Nonlinear equality constrained CUTEr problems (J–Z)

| Problem | $m$ | $n$ | SNOPT | | pdSQP | |
|---|---|---|---|---|---|---|
| | | | fe | itns | fe | itns |
| lukvle1 | 98 | 100 | 13 | 11 | 11 | 10 |
| lukvle3 | 2 | 100 | 36 | 35 | 10 | 9 |
| lukvle6 | 49 | 99 | 29 | 28 | *t* | *t* |
| lukvle7 | 4 | 100 | 34 | 33 | 12 | 8 |
| lukvle8 | 98 | 100 | 24 | 18 | *t* | *t* |
| lukvle9 | 6 | 100 | *t* | *t* | 129 | 74 |
| lukvle10 | 98 | 100 | *t* | *t* | 13 | 10 |
| lukvle13 | 64 | 98 | 76 | 68 | *t* | *t* |
| lukvle14 | 64 | 98 | 38 | 33 | *t* | *t* |
| lukvle16 | 72 | 97 | 342 | 160 | *t* | *t* |
| lch | 1 | 300 | *t* | *t* | 19 | 17 |
| mss1 | 73 | 90 | 69 | 59 | *t* | *t* |
| maratos | 1 | 2 | 13 | 7 | 4 | 3 |
| mwright | 3 | 5 | 10 | 9 | 9 | 7 |
| orthrdm2 | 100 | 203 | 10 | 7 | 8 | 6 |
| orthrds2 | 100 | 203 | 161 | 80 | *t* | *t* |
| orthrega | 64 | 133 | 21 | 19 | *t* | *t* |
| orthregb | 6 | 27 | 8 | 6 | 5 | 4 |
| orthregc | 10 | 25 | 16 | 14 | 8 | 7 |
| orthregd | 10 | 23 | 1462 | 442 | 8 | 6 |
| orthrgdm | 10 | 23 | 14 | 11 | 8 | 6 |
| orthrgds | 76 | 155 | *t* | *t* | 17 | 11 |
| s316-322 | 1 | 2 | 10 | 6 | 25 | 24 |

jointly with respect to the primal and dual variables. In its most general form, the function may be considered as one of a continuum of functions that have some well-known functions as specific cases. One variant of this generalized function—the primal-dual augmented Lagrangian—is proposed as the basis of three primal-dual methods. The first is a primal-dual bound-constrained Lagrangian method based on a primal method given by Conn, Gould, and Toint [10]; the second is a primal-dual linearly constrained Lagrangian method based on the method of Friedlander and Saunders [22]; and the third is a new primal-dual SQP method. Preliminary numerical results from nonlinearly constrained problems with equality constraints indicate that the primal-dual SQP approach generates trial steps that are well suited to a primal-dual augmented Lagrangian merit function.

**Acknowledgements** The authors would like to thank the referees for constructive comments that significantly improved the presentation.

## References


1. Bertsekas, D.P.: Constrained optimization and Lagrange multiplier methods. In: Computer Science and Applied Mathematics. Academic Press, New York (1982)
2. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Athena Scientific, Belmont (1996)
3. Biggs, M.C.: Constrained minimization using recursive equality quadratic programming. In: Lootsma, F.A. (ed.) Numerical Methods for Nonlinear Optimization, pp. 411–428. Academic Press, London (1972)
4. Boggs, P.T., Kearsley, A.J., Tolle, J.W.: A global convergence analysis of an algorithm for large-scale nonlinear optimization problems. SIAM J. Optim. **9**(4), 833–862 (1999)
5. Boggs, P.T., Tolle, J.W.: A family of descent functions for constrained optimization. SIAM J. Numer. Anal. **21**, 1146–1161 (1984)
6. Boggs, P.T., Tolle, J.W.: Sequential quadratic programming. Acta Numer. **4**, 1–51 (1995)
7. Bongartz, I., Conn, A.R., Gould, N.I.M., Toint, Ph.L.: CUTE: Constrained and unconstrained testing environment. ACM Trans. Math. Software **21**(1), 123–160 (1995)
8. Byrd, R.H., Tapia, R.A., Zhang, Y.: An SQP augmented Lagrangian BFGS algorithm for constrained optimization. SIAM J. Optim. **20**, 210–241 (1992)
9. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: A comprehensive description of LANCELOT. Technical Report 91/10, Département de Mathématique, Facultés Universitaires de Namur (1991)
10. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. SIAM J. Numer. Anal. **28**, 545–572 (1991)
11. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: Trust-Region Methods. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2000)
12. Curtis, F.E., Nocedal, J.: Flexible penalty functions for nonlinear constrained optimization. IMA J. Numer. Anal. **28**, 749–769 (2008)
13. DiPillo, G., Grippo, L.: A new class of augmented Lagrangians in nonlinear programming. SIAM J. Control Optim. **17**, 618–628 (1979)
14. Fiacco, A.V., McCormick, G.P.: Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Wiley, New York (1968)
15. Fiacco, A.V., McCormick, G.P.: Nonlinear programming. In: Classics in Applied Mathematics, 2nd edn. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1990). Reprint of the 1968 original
16. Fletcher, R.: A class of methods for nonlinear programming with termination and convergence properties. In: Abadie, J. (ed.) Integer and Nonlinear Programming, pp. 157–175. North-Holland, Amsterdam (1970)
17. Fletcher, R.: Practical Methods of Optimization. Vol. 2: Constrained Optimization. Wiley, Chichester (1981)
18. Fontecilla, R., Steihaug, T., Tapia, R.A.: A convergence theory for a class of quasi-Newton methods for constrained optimization. SIAM J. Numer. Anal. **24**, 1133–1151 (1987)
19. Forsgren, A., Gill, P.E.: Primal-dual interior methods for nonconvex nonlinear programming. SIAM J. Optim. **8**, 1132–1152 (1998)
20. Forsgren, A., Gill, P.E., Wright, M.H.: Interior methods for nonlinear optimization. SIAM Rev. **44**, 525–597 (2002)
21. Forsgren, A., Murray, W.: Newton methods for large-scale linear equality-constrained minimization. SIAM J. Matrix Anal. Appl. **14**, 560–587 (1993)
22. Friedlander, M.P., Saunders, M.A.: A globally convergent linearly constrained Lagrangian method for nonlinear optimization. SIAM J. Optim. **15**(3), 863–897 (2005)
23. Gertz, E.M., Gill, P.E.: A primal-dual trust-region algorithm for nonlinear programming. Math. Program., Ser. B **100**, 49–94 (2004)
24. Gill, P.E., Murray, W., Saunders, M.A.: SNOPT: An SQP algorithm for large-scale constrained optimization. SIAM Rev. **47**, 99–131 (2005)
25. Gill, P.E., Murray, W., Saunders, M.A.: User's guide for SNOPT version 7: Software for large-scale nonlinear programming. Numerical Analysis Report 06-2, Department of Mathematics, University of California, San Diego, La Jolla, CA (2006)
26. Gill, P.E., Murray, W., Saunders, M.A., Wright, M.H.: Some theoretical properties of an augmented Lagrangian merit function. In: Pardalos, P.M. (ed.) Advances in Optimization and Parallel Computing, pp. 101–128. North Holland, Amsterdam (1992)

27. Gould, N.I.M.: On the accurate determination of search directions for simple differentiable penalty functions. IMA J. Numer. Anal. **6**, 357–372 (1986)
28. Gould, N.I.M., Orban, D., Toint, Ph.L.: CUTEr and SifDec: A constrained and unconstrained testing environment, revisited. ACM Trans. Math. Softw. **29**(4), 373–394 (2003)
29. Gould, N.I.M., Robinson, D.P.: A second derivative SQP method: Global convergence. SIAM J. Optim. **20**(4), 2023–2048 (2010)
30. Gould, N.I.M., Robinson, D.P.: A second derivative SQP method: Local convergence and practical issues. SIAM J. Optim. **20**(4), 2049–2079 (2010)
31. Greenstadt, J.: On the relative efficiencies of gradient methods. Math. Comp. **21**, 360–367 (1967)
32. Hager, W.W.: Stabilized sequential quadratic programming. Comput. Optim. Appl. **12**(1–3), 253–273 (1999). Computational optimization—a tribute to Olvi Mangasarian, Part I
33. Han, S.P.: A globally convergent method for nonlinear programming. J. Optim. Theory Appl. **22**, 297–309 (1977)
34. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)
35. Mangasarian, O.L., Fromovitz, S.: The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. J. Math. Anal. Appl. **17**, 37–47 (1967)
36. Maratos, N.: Exact penalty function algorithms for finite-dimensional and control optimization problems. PhD thesis, Department of Computing and Control, University of London (1978)
37. Murray, W.: An algorithm for constrained minimization. In: Optimization, Sympos., Univ. Keele, Keele, 1968, pp. 247–258. Academic Press, London (1969)
38. Murray, W.: Constrained optimization. PhD thesis, Department of Computer Science, University of London (1969)
39. Murtagh, B.A., Saunders, M.A.: A projected Lagrangian algorithm and its implementation for sparse nonlinear constraints. Math. Program. **16**, 84–117 (1982)
40. Murtagh, B.A., Saunders, M.A.: MINOS 5.5 User's Guide. Report SOL 83-20R, Department of Operations Research, Stanford University, Stanford, CA, (Revised 1998)
41. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (1999)
42. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization, pp. 283–298. Academic Press, London (1969)
43. Powell, M.J.D.: A fast algorithm for nonlinearly constrained optimization calculations. Technical Report 77/NA 2, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England (1977)
44. Powell, M.J.D.: The convergence of variable metric methods for nonlinearly constrained optimization calculations. In: Nonlinear Programming, 3, Proc. Sympos., Special Interest Group Math. Programming, Univ. Wisconsin, Madison, Wis., 1977, pp. 27–63. Academic Press, New York (1978)
45. Robinson, D.P.: Primal-Dual Methods for Nonlinear Optimization. PhD thesis, Department of Mathematics, University of California, San Diego (September 2007)
46. Robinson, S.M.: A quadratically-convergent algorithm for general nonlinear programming problems. Math. Program. **3**, 145–156 (1972)
47. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**(2), 97–116 (1976)
48. Rockafellar, R.T.: Monotone operators and the proximal-point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
49. Schittkowski, K.: The nonlinear programming method of Wilson, Han, and Powell with an augmented Lagrangian type line search function. I. Convergence analysis. Numer. Math. **38**(1), 83–114 (1981/82)
50. Schittkowski, K.: The nonlinear programming method of Wilson, Han, and Powell with an augmented Lagrangian type line search function. II. An efficient implementation with linear least squares subproblems. Numer. Math. **38**(1), 115–127 (1981/82)
51. Schittkowski, K.: On the convergence of a sequential quadratic programming method with an augmented Lagrangian line search function. Math. Oper.forsch Stat. Ser. Optim. **14**(2), 197–216 (1983)
52. Tapia, R.A.: Diagonalized multiplier methods and quasi-Newton methods for constrained optimization. J. Optim. Theory Appl. **22**, 135–194 (1977)
53. Wächter, A., Biegler, L.T., Lang, Y.-D., Raghunathan, A.: IPOPT: An interior point algorithm for large-scale nonlinear optimization (2002). http://www.coin-or.org
54. Wright, M.H.: Interior methods for constrained optimization. Acta Numer. **1992**, 341–407 (1992)
55. Wright, S.J.: An algorithm for degenerate nonlinear programming with rapid local convergence. SIAM J. Optim. **15**(3), 673–696 (2005)
56. Wright, S.J., Orban, D.: Properties of the log-barrier function on degenerate nonlinear programs. Math. Oper. Res. **27**(3), 585–613 (2002)