# A PRIMAL-DUAL AUGMENTED LAGRANGIAN*

Philip E. GILL† Daniel P. ROBINSON‡

UCSD Department of Mathematics
Technical Report NA-08-02
April 2008

**Abstract**

Nonlinearly constrained optimization problems can be solved by minimizing a sequence of simpler unconstrained or linearly constrained subproblems. In this paper, we discuss the formulation of subproblems in which the objective is a primal-dual generalization of the Hestenes-Powell augmented Lagrangian function. This generalization has the crucial feature that it is minimized with respect to both the primal *and* the dual variables simultaneously. A benefit of this approach is that the quality of the dual variables is monitored explicitly during the solution of the subproblem. Moreover, each subproblem may be regularized by imposing explicit bounds on the dual variables. Two primal-dual variants of conventional primal methods are proposed: a primal-dual bound constrained Lagrangian (pdBCL) method and a primal-dual $\ell_1$ linearly constrained Lagrangian (pd$\ell_1$-LCL) method.

**Key words.** Nonlinear programming, nonlinear inequality constraints, augmented Lagrangian methods, bound constrained Lagrangian methods, linearly constrained Lagrangian methods, primal-dual methods.

**AMS subject classifications.** 49J20, 49J15, 49M37, 49D37, 65F05, 65K05, 90C30

*Research supported by the National Science Foundation grant DMS-0511766.

†Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112 (pgill@ucsd.edu).

‡Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford, OX1 3QD, UK (daniel.robinson@comlab.ox.ac.uk).

## 1. Introduction

This paper considers a primal-dual augmented Lagrangian function that may be used to find a local solution of the 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. Our analysis for this problem format easily carries over to the more general setting with $l \leq x \leq u$. Much of the discussion will focus on the simplified problem

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

This equality constrained problem has all the properties necessary for a description of the new function, while avoiding the complications of dealing with the bound constraints.

Since the early 1960s, the idea of replacing a constrained optimization problem by a sequence of unconstrained problems parameterized by a scalar parameter $\rho$ has played a fundamental role in the formulation of algorithms (for a seminal reference, see Fiacco and McCormick [12,13]). 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 $\rho$ that "penalizes" the sum of the squares of the constraint violations. Under certain conditions (see, e.g. [13, 18, 34, 36]), the minimizers of the penalty function define a differentiable *trajectory* or *central path* of solutions that passes through the solution as $\rho \rightarrow \infty$. Penalty methods approximate this path by minimizing the penalty function for a finite sequence of increasing values of $\rho$. 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 $\rho$. As $\rho \rightarrow \infty$, 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 [25] and Powell [28] 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 $\rho$ to go to infinity for convergence. The price that must be paid for keeping $\rho$ finite is the need to update an estimate of the Lagrange multiplier vector 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 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 [14]; DiPillo and Grippo [11]; 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 methods to sequential linearly constrained methods. In this context, the augmented Lagrangian has been used successfully within a number of different algorithmic frameworks for the solution of problem (NP). The method used in the package LANCELOT [8] 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 [27] employs a variant of Robinson's linearly constrained Lagrangian (LCL) method [30] in which an augmented Lagrangian is minimized subject to the linearized nonlinear constraints. Friedlander and Saunders [19] define a globally convergent version of the LCL method that can treat infeasible constraints or infeasible subproblems. Augmented Lagrangian functions have also been used extensively as a merit function for sequential quadratic programming (SQP) methods (see, e.g., [7, 16, 33, 21, 22, 4, 6]).

The development of path-following interior methods for linear programming in the mid-1980s led to a renewed interest in the treatment of constraints by sequential unconstrained optimization. This new attention not only lead to a new understanding of the computational complexity of existing methods but also gave new insight into how sequential unconstrained methods could be extended and improved. A notable development was the derivation of efficient path-following methods for linear programming based on minimization and zero-finding simultaneously with respect to both the primal and dual variables. These developments lead to a new focus on two computational aspects of penalty- and barrier-function methods. First, the primal-dual formulation made it evident that the inherent ill-conditioning of penalty and barrier methods 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 simultaneously with respect to both the primal variables *and* the dual variables. A benefit of this approach is that the quality of the dual variables is monitored explicitly during the solution of the subproblem. Moreover, each subproblem can be regularized by imposing explicit bounds on the dual variables. Two primal-dual variants of classical primal methods are proposed: a primal-dual bound constrained Lagrangian (pdBCL) method and a primal-dual $\ell_1$ linearly constrained Lagrangian (pd$\ell_1$-LCL) method.

The paper is organized as follows. In Section 2 we review some 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 is arbitrary. In Section 3 we introduce a generalized primal-dual augmented Lagrangian function that may be used to define a *continuum* of methods that include several well-known methods as specific cases. Similarities with the Hestenes-Powell augmented Lagrangian are also discussed. In Section 4 it is shown how artificial bounds on the dual variables may be used to *regularize* the associated subproblem.

Finally, in Section 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.

### 1.1. Notation and Terminology

Unless explicitly indicated otherwise, $\|\cdot\|$ denotes the vector two-norm or its subordinate 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 quantity $a \cdot b$ denotes the vector with $i$th component $a_i b_i$. Finally, $e$ denotes the vector of all ones whose dimension is determined by the context.

Much of our notation is standard. A local solution of (NP) and (NEP) is denoted by $x^*$; $g(x)$ is the gradient of $f(x)$, and $H(x)$ its (symmetric) Hessian; the matrix $H_j(x)$ is the Hessian of $c_j(x)$; $J(x)$ is the $m \times n$ Jacobian matrix of the constraints, with $i$th row $\nabla c_i(x)^T$. The *Lagrangian function* associated with (NEP) is $\mathcal{L}(x, y) = f(x) - y^T c(x)$. The Hessian of the Lagrangian with respect to $x$ is $\nabla^2 \mathcal{L}(x, y) = H(x) - \sum_{j=1}^m y_j H_j(x)$.

## 2. The Hestenes-Powell Augmented Lagrangian

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

$$\mathcal{L}_A(x, y, \rho) = f(x) - c(x)^T y + \frac{\rho}{2}\|c(x)\|_2^2,$$

where $\rho$ is the penalty parameter, and $x$ and $y$ are primal and dual variables respectively. If $y$ is fixed at the optimal multiplier vector $y^*$, then a solution $x^*$ of (NEP) is a stationary point of $\mathcal{L}_A$ regarded as a function of $x$. Moreover, if the second-order sufficient conditions for optimality hold, then there exists a *finite* $\bar{\rho}$ such that $x^*$ is an isolated unconstrained minimizer of $\mathcal{L}_A$ for all $\rho > \bar{\rho}$. Based on this result, Hestenes and Powell proposed that $x^*$ be found by minimizing a sequence of augmented Lagrangians $\mathcal{L}_A(x, y_k, \rho_k)$ in which the choice of multiplier estimate $y_k$ is based on the minimizer of $\mathcal{L}_A(x, y_{k-1}, \rho_{k-1})$.

In order to emphasize the role of the penalty parameter $\rho$ in problem regularization, we will focus on the Hestenes-Powell augmented Lagrangian in the form

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

where $\mu$ is a positive scalar and $y_e$ is an estimate of the Lagrange multipliers $y^*$. Here we use the notation $\mathcal{L}_A(x\,; y, \mu)$ to indicate that $\mathcal{L}_A$ is a function of $x$ parameterized by $y$ and $\mu$. Throughout the discussion we use the quantity $\rho$ to denote the inverse of the parameter $\mu$. In order to emphasize this dependency in expressions that use both $\mu$ and $1/\mu$, we use $\rho_\mu$ to denote $1/\mu$.

The augmented Lagrangian may be minimized by solving a sequence of subproblems of the form

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

If $\nabla^2\mathcal{L}_A(x)$ is positive definite, then $p$ 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

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

where, as discussed above, $\rho_\mu$ denotes the inverse parameter $\rho_\mu = 1/\mu$. With this notation, we have

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

and the Newton equations are simply

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

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

The next result shows that the Newton direction for the augmented Lagrangian function satisfies a "primal-dual" system. This system will be used later.

**Lemma 2.1.** *If $y$ is an arbitrary $m$-vector, then the augmented Lagrangian direction $p$ 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

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

Let $q$ denote the $m$-vector

$$q = -\rho_\mu\big(Jp + (c + \mu(y - y_e))\big). \tag{2.5}$$

Using the definition $\pi = y_e - \rho_\mu c$, equations (2.4) and (2.5) may be combined to give

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

Applying the nonsingular matrix

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

to both sides of this equation yields

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

∎

In particular, if we choose $y = y_e$ in (2.3), then the Newton direction $p$ satisfies the equations

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

which may be considered as a primal-dual variant of primal Newton equations (2.2) analogous to the primal-dual formulation of the quadratic penalty method considered by Gould [23] (for related methods, see Murray [26] and Biggs [3]).

Note that the *nonzero* $(2,2)$ block in the primal-dual matrix above *regularizes* the system; i.e., it is not necessary for $J(x)$ to have full row rank for the Newton equations to be nonsingular. A full-rank assumption is required if the $(2,2)$ block is zero.

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

## 3. The Generalized Primal-Dual Augmented Lagrangian

In Powell's derivation of the conventional augmented Lagrangian method, the solution of problem (NEP) is found by repeatedly minimizing the quadratic penalty function for the "shifted" problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; f(x) \quad \text{subject to} \quad c(x) - \mu y_e = 0, \tag{3.1}$$

where $y_e$ is an estimate of the Lagrange multipliers $y^*$. This method is based on the observation that for sufficiently small $\mu$, $x^*$ is a minimizer of the quadratic penalty function associated with problem (3.1) for the choice $y_e = y^*$.

In this section we consider an augmented Lagrangian that can be minimized with respect to both the primal and the dual variables. Given an approximate Lagrange multiplier vector $y_e$, the proposed function is given by

$$\mathcal{M}(x, y\,; y_e, \mu, \nu) = f(x) - c(x)^T y_e + \frac{1}{2\mu}\|c(x)\|^2 + \frac{\nu}{2\mu}\|c(x) + \mu(y - y_e)\|^2, \tag{3.2}$$

where $\nu$ and $\mu$ are constant scalars with $\mu > 0$. This function may be derived as the Forsgren-Gill primal-dual penalty function associated with the shifted constraints $c(x) - \mu y_e = 0$ (see [17]). Using the $m$-vector $\pi(x) = y_e - \rho_\mu c(x)$ of (2.1), the gradient and Hessian for $\mathcal{M}(x, y\,; y_e, \mu, \nu)$ may be written as

$$\nabla\mathcal{M}(x, y\,; y_e, \mu, \nu) = \begin{pmatrix} g - J^T\big(\pi + \nu(\pi - y)\big) \\ \nu\big(c + \mu(y - y_e)\big) \end{pmatrix} = \begin{pmatrix} g - J^T\big(\pi + \nu(\pi - y)\big) \\ \nu\mu(y - \pi) \end{pmatrix}, \tag{3.3a}$$

and

$$\nabla^2\mathcal{M}(x,y\,;y_e,\mu,\nu)=\begin{pmatrix}H\big(x,\pi+\nu(\pi-y)\big)+\rho_\mu(1+\nu)J^TJ & \nu J^T\\ \nu J & \nu\mu I\end{pmatrix}, \tag{3.3b}$$

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

The next result indicates the potential role of $\mathcal{M}$ 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 $\mathcal{M}(x,y\,;y^*,\mu,\nu)$ for $\mu$ sufficiently small and $\nu>0$.

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

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

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

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

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

**Proof.** We must show that $\nabla\mathcal{M}$ is zero and $\nabla^2\mathcal{M}$ 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 (3.3a) and using assumption (ii), gives $\nabla\mathcal{M}(x^*,y^*\,;y^*,\mu,\nu)=0$ directly.

Similarly, the Hessian (3.3b) is given by

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

where, for simplicity, we have used $\rho=\rho_\mu$, $\nabla^2\mathcal{M}=\nabla^2\mathcal{M}(x^*,y^*\,;y^*,\mu,\nu)$, $J=J(x^*)$, and $H=H(x^*,y^*)$.

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

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

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

$$\text{In}\,\big(L^T\nabla^2\mathcal{M}L\big)=\text{In}\,\big(\nabla^2\mathcal{M}\big)=(m,0,0)+\text{In}(H+\rho J^TJ),$$

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 $J$. Since $Q$ is nonsingular, $H + \rho J^T J$ must have the same inertia as $Q^T(H + \rho J^T J)Q$ from Sylvester's Law of Inertia. Pre- and post-multiplying $H + \rho J^T J$ by $Q^T$ and $Q$ gives

$$Q^T(H + \rho J^T J)Q = Q^T H Q + \rho Q^T J^T J Q = \begin{pmatrix} Z^T H Z & Z^T H Y \\ Y^T H Z & Y^T H Y + \rho I_r \end{pmatrix}.$$

This matrix has the form

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

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

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

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

$$\operatorname{In}\big(\nabla^2 \mathcal{M}\big) = \big(m, 0, 0\big) + \big(n, 0, 0\big) = \big(m + n, 0, 0\big),$$

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

Theorem 3.1 indicates that if an estimate of $y^*$ is known for problem (NEP), then an approximate minimization of $\mathcal{M}$ with respect to both $x$ and $y$ is likely to provide an even better estimate. Since our goal is to develop second-order methods, it is of interest to consider the Newton system for the primal-dual augmented Lagrangian.

Using the derivatives (3.3a) and (3.3b) for $\mathcal{M}$, the Newton direction for the primal-dual augmented Lagrangian satisfies

$$\begin{pmatrix} H\big(x, \pi + \nu(\pi - y)\big) + \rho_\mu(1+\nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = -\begin{pmatrix} g - J^T\big(\pi + \nu(\pi - y)\big) \\ \nu\big(c + \mu(y - y_e)\big) \end{pmatrix}, \tag{3.4}$$

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*

$$\begin{pmatrix} H + \rho_\mu(1+\nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = -\begin{pmatrix} g - J^T\big(\pi + \nu(\pi - y)\big) \\ \nu\big(c + \mu(y - y_e)\big) \end{pmatrix}, \tag{3.5}$$

*and*

$$\begin{pmatrix} H & J^T \\ J & -\mu I \end{pmatrix} \begin{pmatrix} p \\ -q \end{pmatrix} = -\begin{pmatrix} g - J^Ty \\ c + \mu(y - y_e) \end{pmatrix}. \tag{3.6}$$

*are equivalent for all $\nu \neq 0$, i.e., $(p, q)$ is a solution of (3.5) if and only if it is a solution of (3.6).*

**Proof.** Multiplying both sides of (3.5) by the nonsingular matrix

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

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

Several well-known functions may be recovered from the primal-dual merit function with appropriate choices for the parameters $y_e$ and $\nu$.

**The quadratic penalty function ($\nu \equiv 0$, $y_e \equiv 0$).** In this case, the function is given by

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

which is a function in the primal variables only. The primal-dual form of the Newton equations analogous to (2.3) are given by:

$$\begin{pmatrix} H(x,\pi) & J^T \\ J & -\mu I \end{pmatrix} \begin{pmatrix} p \\ -q \end{pmatrix} = -\begin{pmatrix} g - J^Ty \\ c + \mu y \end{pmatrix}, \tag{3.7}$$

which have been studied by Biggs [3] and Gould [23]. Few competitive modern methods are based on the direct minimization of the quadratic penalty function, but several reliable and efficient methods are designed to behave like the quadratic penalty method when the set of optimal multipliers is unbounded (see, e.g., [9], [19], [27], and [29]).

**The proximal-point penalty function ($\nu \equiv -1$, $y_e \equiv 0$).** This function has the form

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

The proximal-point penalty function has been used in the formulation of stabilized SQP methods (see, e.g., Hager [24] and Wright [35]). In this case, the Newton equations 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 with the primal-dual equations (3.7) for the quadratic penalty function. However, the direction $p$ is different because the Lagrangian Hessian is evaluated with different values of the multipliers.

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

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

Lemma 2.1 implies that if $\pi$ is substituted for $y$ in the Hessian associated with the primal-dual augmented Lagrangian system (3.6), then the vector $p$ 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 ($\nu \equiv -1$).** In this case we have

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

The proximal-point Lagrangian function has been studied, for example, by Rockafellar in [31, 32].

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

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

Methods based on the use of the primal-dual penalty function and its barrier function counterpart are discussed by Forsgren and Gill [17] and Gertz and Gill [20].

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

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

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

Table 1 summarizes the six functions discussed above.

Table 1: Functions associated with $\mathcal{M}(x, y\,; y_e, \mu, \nu)$.

| | $\nu$ | $y_e$ |
|---|---|---|
| The quadratic penalty function | 0 | 0 |
| The proximal-point penalty function | $-1$ | 0 |
| The Hestenes-Powell augmented Lagrangian | 0 | $y_e$ |
| The proximal-point Lagrangian | $-1$ | $y_e$ |
| The primal-dual quadratic penalty function | 1 | 0 |
| The primal-dual augmented Lagrangian | 1 | $y_e$ |

## 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 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:

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

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 minimizers. However, if $\gamma$ restricts the subproblem solutions, then a different problem is being solved. This 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. 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 effecting the subproblems near $(x^*, y^*)$. In Section 5 we propose 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 components of $y$ become active it becomes unclear which problem is being solved. The next theorem shows that the solutions are related to those obtained by minimizing an exact penalty function.

**Theorem 4.1.** *Let $\gamma > 0$. If $(\bar{x}, \bar{y}, \bar{w})$ is a solution of*

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

*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 solution to*

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

**Proof.** Define $\bar{\pi} = \bar{y} - 2\rho_\mu \bar{w}$. The diagonal scaling matrix $P \triangleq \operatorname{diag}(\rho_1, \ldots, \rho_m)$ is then defined as

$$\rho_i = \begin{cases} \bar{\pi}_i & \text{if } \bar{w}_i < 0, \\ -\bar{\pi}_i & \text{if } \bar{w}_i > 0, \\ |\bar{\pi}_i| + \epsilon & \text{if } \bar{w}_i = 0, \end{cases} \tag{4.3}$$

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

$$\begin{aligned} \underset{x\in\mathbb{R}^n, u\in\mathbb{R}^m, v\in\mathbb{R}^m}{\text{minimize}} \quad & f(x) + \sum_{i=1}^{m} \rho_i(u_i + v_i) \\ \text{subject to} \quad & c(x) + \mu(\bar{y} - y_e) - u + v = 0, \quad u \ge 0, \quad v \ge 0. \end{aligned} \tag{4.4}$$

Define the following:

$$\bar{u}_i = \begin{cases} 0 & \text{if } \bar{w}_i \le 0, \\ \bar{w}_i & \text{if } \bar{w}_i > 0, \end{cases} \tag{4.5a}$$

$$\bar{v}_i = \begin{cases} 0 & \text{if } \bar{w}_i \ge 0, \\ -\bar{w}_i & \text{if } \bar{w}_i < 0, \end{cases} \tag{4.5b}$$

$$\bar{z}_u = Pe + \bar{\pi}, \tag{4.5c}$$

$$\bar{z}_v = Pe - \bar{\pi}. \tag{4.5d}$$

It will be shown that $(\bar{x}, \bar{u}, \bar{v}, \bar{\pi}, \bar{z}_u, \bar{z}_v)$ is a solution to (4.4), where $\bar{\pi}$ is the Lagrange multiplier vector for the general equality constraint, $\bar{z}_u$ is the Lagrange multiplier vector for $u \ge 0$, and $\bar{z}_v$ is the Lagrange multiplier vector for $v \ge 0$.

The solution $(\bar{x}, \bar{y}, \bar{w})$ satisfies the following optimality condition:

$$J(\bar{x})^T\big(2\pi(\bar{x}) - \bar{y}\big) = g(\bar{x}), \tag{4.6a}$$
$$c(\bar{x}) + \mu(\bar{y} - y_e) = \bar{w}, \tag{4.6b}$$
$$-\gamma e \le \bar{y} \le \gamma e, \tag{4.6c}$$
$$\min(\gamma e - \bar{y}, \bar{y} + \gamma e, |\bar{w}|) = 0, \tag{4.6d}$$
$$\bar{w} \cdot (\gamma e + \bar{y}) \le 0, \tag{4.6e}$$
$$\bar{w} \cdot (-\gamma e + \bar{y}) \le 0, \tag{4.6f}$$

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

C1. $\bar{u} \ge 0, \quad \bar{v} \ge 0, \quad \bar{z}_u \ge 0, \quad \bar{z}_v \ge 0, \quad \bar{u} \cdot \bar{z}_u = 0, \quad \bar{v} \cdot \bar{z}_v = 0;$

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 (4.6b).
(proof of C3): By definition of $\bar{\pi}$ and $\pi(x)$ and use of (4.6b), the following equality holds:

$$\bar{\pi} = \bar{y} - 2\rho_\mu \bar{w} = 2y_e - 2\rho_\mu c(\bar{x}) - \bar{y} = 2\pi(\bar{x}) - \bar{y}. \tag{4.7}$$

C3 follows from this equality and (4.6a).
(proof of C4): Follows by definition (4.5c).
(proof of C5): Follows by definition (4.5d).
(proof of C1): $\bar{u} \ge 0$ and $\bar{v} \ge 0$ by definition.

Next it is shown that $\bar{u} \cdot \bar{z}_u = 0$. The result is trivial if $\bar{u}_i = 0$. So suppose that $\bar{u}_i \ne 0$. This implies that $\bar{w}_i > 0$ and thus $\rho_i = -\bar{\pi}_i$. 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 $[\bar{z}_v]_i \triangleq \rho_i - \bar{\pi}_i = 0$.

Next consider the following three 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 $\bar{\pi}_i = \bar{y}_i - 2\rho_\mu \bar{w}_i = -\gamma - 2\rho_\mu \bar{w}_i < 0$. This implies that $\rho_i = -\bar{\pi}_i > 0$ and that $[\bar{z}_v]_i = \rho_i - \bar{\pi}_i = -2\bar{\pi}_i > 0$. Likewise, $[\bar{z}_u]_i = \rho_i + \bar{\pi}_i = 0$.

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

The proof is complete since in all cases $\bar{z}_u \ge 0$, $\bar{z}_v \ge 0$, and $\rho_i > 0$. ∎

### 4.1. Interpretation of the artificial bounds

Algorithms 5.1 and 5.3 given in Section 5 explicitly bound the dual variables by adding artificial bound constraints. This section gives 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 $(x_\mu, y_\mu) = \big(x(\mu), y(\mu)\big)$ denote the solution to the unconstrained problem

$$\underset{x,y}{\text{minimize}} \;\; \mathcal{M}(x, y; y_e, \mu). \tag{4.8}$$

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

$$\underset{x,y}{\text{minimize}} \;\; \mathcal{M}(x, y) + \sigma\|y^v\|_1, \tag{4.9}$$

where $\sigma$ is a positive penalty parameter and $y^v = \min(0, \gamma - |y|)$ (the definition of $y^v$ should be interpreted component-wise and is a measure of how much $y$ violates its bounds). If $\sigma > \|\bar{w}\|_\infty$, it is well known that solutions of problem (4.9) are solutions of (4.1) (see, e.g., [15]). We consider the quantity $\|\bar{w}\|_\infty$ as the "required penalization". It follows from the optimality conditions for problems (4.1) and (4.8) that $\bar{w} = c(\bar{x}) - c(x_\mu) + \mu(\bar{y} - y_\mu)$, which implies that the required penalization satisfies

$$\|\bar{w}\|_\infty \leq \|c(\bar{x}) - c(x_\mu)\|_\infty + \mu\|\bar{y} - y_\mu\|_\infty. \tag{4.10}$$

This shows that the required penalization is intimately associated with the magnitudes of the quantities $\|c(\bar{x}) - c(x_\mu)\|_\infty$ 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 (4.9). Specifically, the $\mu\|\bar{y}-y_\mu\|_\infty$ term in (4.10) 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, $\|c(\bar{x}) - c(x_\mu)\|_\infty$ term in (4.10) shows that the required penalization is likely to be large if the constraint values differ substantially. We note that for small $\mu$, the minimizers of the merit function will be close to minimizers of the quadratic penalty function. It follows that $\|c(\bar{x})\|_\infty$ 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 component-wise result.

**Theorem 4.2.** *If the point $(\bar{x}, \bar{y}, \bar{w})$ is a solution to*

$$\underset{x,y}{\text{minimize}} \;\; \mathcal{M}(x, y; y_e, \mu) \quad \text{subject to} \quad y_\ell \leq y \leq y_u, \tag{4.11}$$

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

**Proof.** The result follows from the standard results associated with $\ell_1$ penalty functions (see, e.g., [9] and [15]). ∎

## 5. Algorithms

The augmented Lagrangian has been used very successfully within different algorithmic frameworks. In the context of problem (NP), the optimization code LANCELOT [8] approximately minimizes a sequence of bound constrained Lagrangian (BCL) problems. These problems take 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 [9] show that this BCL method is globally convergent, exhibits R-linear convergence, and has a penalty parameter $\mu$ that is uniformly 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 [30], which sequentially minimizes the Lagrangian $\mathcal{L}(x\,;y_e) = f(x) - c(x)^T y_e$, subject to the linearized constraints. Under reasonable assumptions, Robinson proved that his algorithm exhibits R-quadratic convergence when started sufficiently close to a solution satisfying the second-order sufficient conditions. A positive penalty parameter was introduced in the method used by the software package MINOS [27] in order to improve convergence from poor starting points. In other words, MINOS uses the *augmented* Lagrangian function instead of the Lagrangian function. This modification improves the robustness of Robinson's method, but the question of convergence from arbitrary starting points is open. Further improvement was made by Friedlander and Saunders in [19]. Their stabilized LCL (sLCL) method remedies three drawbacks associated with MINOS. First, the sLCL subproblems are always feasible. Second, if the distance from the linearization point to the subproblem solution becomes large, it may be counteracted by decreasing the penalty parameter associated with the linearized constraints. Third, the sLCL method was proved to be globally convergent. These improvements to MINOS resulted from the definition of an elastic subproblem, which is known to be equivalent to an $\ell_1$ penalization of the linearized constraint violation. 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 Robinson's method.

The augmented Lagrangian function may also be used as a *merit* function in sequential quadratic programming (SQP) methods. A merit function is a single function that is used to assess the quality of "trial points". In line-search methods, a pre-determined search direction is typically given as the solution of a quadratic program that models the merit function. The most obvious choice for computing a search direction for the augmented Lagrangian, is to minimize the second-order

Taylor approximation of $\mathcal{L}_A(x\,;y_e,\mu)$. It can be shown that minimizing the second-order Taylor approximation of $\mathcal{L}_A(x\,;y_e,\mu)$ is equivalent to minimizing a certain quadratic objective function subject to linear constraints. For more details see Section 15.3.1 of [10].

The SQP algorithm SNOPT [21] uses the augmented Lagrangian function as a merit function in a different way. Given a current approximation $(x,y)$ to a solution of problem (NEP), SNOPT generates a search direction by solving the quadratic programming subproblem:

$$\underset{p\in\mathbb{R}^n}{\text{minimize}}\;\; g(x)^Tp+\tfrac{1}{2}p^TBp \quad \text{subject to} \quad c(x)+J(x)p=0, \tag{5.2}$$

where $B$ is a symmetric positive-definite approximation of the Hessian of the Lagrangian. A search direction in both the $x$- *and* $y$-variables is then defined by using the minimizer $p$ and the multipliers from the QP subproblem (5.2). In this way, the augmented Lagrangian is a continuous function of both the primal *and* the dual variables.

Now we consider algorithms for problem (NP) that incorporate the primal-dual augmented Lagrangian. In particular, primal-dual variants of the methods discussed in Section 2 will be given: a primal-dual bound constrained Lagrangian method, and a stabilized primal-dual linearly constrained augmented Lagrangian method. The purpose of this section is to illustrate how the primal-dual augmented Lagrangian may be used in different frameworks. This section is *not* intended to give the theoretical developments of those algorithms; this development will be given in a separate paper.

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

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

$$\underset{x\in\mathbb{R}^n,y\in\mathbb{R}^m}{\text{minimize}}\;\; \mathcal{M}(x,y\,;y_k^e,\mu_k) \quad \text{subject to} \quad x\geq 0, \quad -\gamma_k e\leq y\leq \gamma_k e, \tag{5.3}$$

where $\mathcal{M}(x,y\,;y_k^e,\mu_k)$ is the primal-dual augmented Lagrangian of Section 3.1, $\{\gamma_k\}$ is a sequence of constants, $\{y_k^e\}$ is a sequence of estimates of the Lagrange multiplier vector satisfying $y_k^e\in[-\gamma_k e,\gamma_k e]$, and $\{\mu_k\}$ is a sequence of positive penalty parameters. An approximate solution of the subproblem (5.3) is denoted by $(x_k^*,y_k^*,z_k^*,w_k^*)$, where $z_k^*$ and $w_k^*$ are the Lagrange multiplier vectors for the inequality constraints $x\geq 0$ and $-\gamma_k e\leq y\leq\gamma_k e$.

Classical bound-constrained Lagrangian (BCL) methods are known to be locally convergent if the penalty parameter is sufficiently small and if the sequence of subproblems are solved exactly. Bertsekas [1] extends this result by showing that only an approximate solution of each BCL subproblems 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. Algorithm 5.1 below is similar to the algorithm proposed by Conn, Gould, and Toint [9], which is the basis for the LANCELOT code (see [8]).

Based on the degree of primal infeasibility, each iterate is regarded as either "successful" or "unsuccessful". In the successful case, if $y_k^*$ is "close" to the boundary then $\mu_k$ is decreased. This defines a larger artificial bound $\gamma_k$, which encourages the dual variables to be inactive during the next iteration. In the unsuccessful case, the parameter $\mu_k$ is decreased and the artificial bound $\gamma_k$ is increased. Again, large artificial bounds encourage the dual variables in the subproblem to be inactive; when approaching a solution it is highly desirable for the dual variables to be inactive.

As described in [1], convergence of the multiplier method depends critically on the size of $\mu_k$ and $\|y_k^e - y^*\|$. The strategy of decreasing $\mu_k$ in order to force convergence is based on the assumption that if $\mu_k$ is decreased enough, then the iterates will eventually enter a "cone of convergence". Once this cone of convergence has been entered, the penalty parameter will no longer need to be decreased and the algorithm will converge.

In Algorithm 5.1, the vector $\widetilde{y}_k$ is the first-order *primal-dual* multiplier estimate $\widetilde{y}_k = 2\pi(x_k^*) - y_k^*$, where $\pi(x_k^*) = y_k^e - c(x_k^*)/\mu_k$. The fixed parameters associated with the algorithm are listed below.

- $\eta_*$ $(0 < \eta_* \ll 1)$ is the primal convergence tolerance.
- $\omega_*$ $(0 < \omega_* \ll 1)$ is the dual convergence tolerance for the $x$-variables.
- $\tau_*$ $(0 < \tau_* \ll 1)$ is the dual convergence tolerance for the $y$-variables.
- $\mu_0$ $(0 < \mu_0 < 1)$ is the initial penalty parameter.
- $\eta_0$ $(\eta_* \leq \eta_0 < \frac{1}{2})$ is the initial primal feasibility tolerance.
- $\omega_0$ $(\omega_* \leq \omega_0 < \frac{1}{2})$ is the initial dual infeasibility tolerance for the $x$-variables.
- $\mu_c$ $(0 < \mu_c < 1)$ is the contraction factor for $\mu_k$.
- $\tau_f$ $(0 < \tau_f \leq 1)$ is the feasibility tolerance for the $y$-variables.
- $k_\tau$ $(k_\tau > 0)$ is a constant used to update $\tau_k$.
- $\nu$ is a positive constant used in the definition of $\gamma_k$. Given $\mu_0$ and $y_0^e$, $\nu$ is set to $\nu = (\|y_0^e\|_\infty + 1)\mu_0^{\alpha_\gamma}$, which implies that $\gamma_0 = \|y_0^e\|_\infty + 1$. This update guarantees that $y_0^e \in [-\gamma_0 e, \gamma_0 e]$. Moreover, if $(x_0, y_0)$ is optimal, then the algorithm will exit on the first iteration.
- $\alpha_\omega$ is a positive constant used in the update to $\omega_k$ in the "unsuccessful" case. This parameter ensures that $\{\omega_k\}$ converges to zero and that $\omega_k < \omega_0$ for all $k > 0$.
- $\alpha_\eta$ $(0 < \alpha_\eta < \min(1, \alpha_\omega))$ is used to update $\eta_k$ in the unsuccessful case. The condition $\alpha_\eta > 0$ guarantees that $\eta_k < \eta_0$ for all $k > 0$.
- $\beta_\omega$ $(\beta_\omega > 0)$ is used to update $\omega_k$ so that $\omega_{k+1} < \omega_k$ in the successful case.
- $\beta_\eta$ $(0 < \beta_\eta < \min(1, \beta_\omega))$ is used to update $\eta_k$ in the successful case. The condition $\beta_\eta > 0$ ensures that $\eta_{k+1} < \eta_k$.

- $\alpha_\gamma$ $(0 < \alpha_\gamma < 1)$ and $\alpha_\tau$ $(\alpha_\tau > 1)$ are used to update $\gamma_k$ and $\tau_k$.

During each iteration, $(x_k^*, y_k^*, z_k^*, w_k^*)$ is accepted as a solution of problem (5.3) if it satisfies the following conditions:

$$x_k^* \geq 0, \tag{5.4a}$$
$$\|\min(x_k^*, z_k^*)\|_\infty \leq \omega_k, \tag{5.4b}$$
$$-\gamma_k e \leq y_k^* \leq \gamma_k e, \tag{5.4c}$$
$$\|\min(\gamma_k e - y_k^*, y_k^* + \gamma_k e, |w_k^*|)\|_\infty \leq \tau_k, \tag{5.4d}$$
$$\text{If } [\gamma_k e - y_k^*]_j \leq \tau_k, \text{ then } [w_k^*]_j \leq \tau_k, \tag{5.4e}$$
$$\text{If } [y_k^* + \gamma_k e]_j \leq \tau_k, \text{ then } [w_k^*]_j \geq -\tau_k, \tag{5.4f}$$
$$\nabla \mathcal{M}(x_k^*, y_k^*; y_k^e, \mu_k) = \begin{pmatrix} z_k^* \\ w_k^* \end{pmatrix}. \tag{5.4g}$$

Condition (5.4b) is equivalent to $[x_k^*]_i \geq -\omega_k$ and $[z_k^*]_i \geq -\omega_k$ holding in addition to either $[x_k^*]_i \leq \omega_k$ or $[z_k^*]_i \leq \omega_k$ holding for $i = 1:n$. A similar statement holds for condition (5.4d).

---

**Algorithm 5.1.** Primal-Dual BCL Algorithm
INPUT: $(x_0, y_0)$
Set constants $\mu_0$, $\eta_0$, $\omega_0$, $\eta_*$, $\omega_*$, $\tau_*$, $\mu_c$, $\tau_f$, $k_\tau$, $\nu$, $\alpha_\eta$, $\alpha_\omega$, $\beta_\eta$, $\beta_\omega$, $\alpha_\gamma$, and $\alpha_\tau$;
Set $y_0^e = y_0$; $\nu = (\|y_0^e\|_\infty + 1)\mu_0^{\alpha_\gamma}$; $\gamma_0 = \nu\mu_0^{-\alpha_\gamma}$; $\tau_0 = \min(\mu_0\omega_0, k_\tau\mu_0^{\alpha_\tau})$;
*converged* $\leftarrow$ **false**;
**while not** *converged* **do**
  Find $(x_k^*, y_k^*, z_k^*, w_k^*)$, a solution to (5.3) as determined by conditions (5.4).
  **if** $(x_k^*, \widetilde{y}_k, z_k^*)$ satisfies stopping criteria **then** *converged* $\leftarrow$ **true end if**
  Compute $(x_{k+1}^s, y_{k+1}^s)$ to increase the convergence rate.
  **if** $\|c(x_k^*)\| \leq \max(\eta_*, \eta_k)$ **then** [successful]
    **if** $\|y_k^*\|_\infty > \gamma_k - \tau_f$ **then** [approaching boundary]
      $\mu_{k+1} \leftarrow \mu_c\mu_k$; $\gamma_{k+1} \leftarrow \nu\mu_{k+1}^{-\alpha_\gamma}$;
    **end if**
    $\eta_{k+1} \leftarrow \eta_k\mu_{k+1}^{\beta_\eta}$; $\omega_{k+1} \leftarrow \omega_k\mu_{k+1}^{\beta_\omega}$; [decrease $\eta_k$ and $\omega_k$]
    $\tau_{k+1} \leftarrow \min(\mu_{k+1}\omega_{k+1}, k_\tau\mu_{k+1}^{\alpha_\tau})$;
  **else** [unsuccessful]
    $\mu_{k+1} \leftarrow \mu_c\mu_k$; $\gamma_{k+1} \leftarrow \nu\mu_{k+1}^{-\alpha_\gamma}$;
    $\eta_{k+1} \leftarrow \eta_0\mu_{k+1}^{\alpha_\eta}$; $\omega_{k+1} \leftarrow \omega_0\mu_{k+1}^{\alpha_\omega}$; [increase or decrease $\eta_k$ and $\omega_k$]
    $\tau_{k+1} \leftarrow \min(\mu_{k+1}\omega_{k+1}, k_\tau\mu_{k+1}^{\alpha_\tau})$;
  **end if**
  **if** $\|y_{k+1}^s\|_\infty \leq \gamma_{k+1}$ **then**
    $y_{k+1}^e = y_{k+1}^s$; [high-order estimate]
  **else if** $\|\widetilde{y}_k\|_\infty \leq \gamma_{k+1}$ **then**
    $y_{k+1}^e = \widetilde{y}_k$; [first-order estimate]
  **else**

$y^e_{k+1} = y^e_k$;
**end if**
$k \leftarrow k + 1$;
**end while**
OUTPUT: $(x^*, y^*, z^*) \leftarrow (x^*_k, \widetilde{y}_k, z^*_k)$

### 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

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n, y\in\mathbb{R}^m}{\text{minimize}} & \mathcal{M}(x, y\,; y^e_k, \rho_k) \\ \text{subject to} & \bar{c}_k(x) = 0, \\ & x \geq 0, \quad -\gamma_k e \leq y \leq \gamma_k e, \end{array} \tag{5.5}$$

where $\gamma_k$ is a positive constant, $y^e_k$ 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 $\rho_k$ is the $k$th penalty parameter. Notice that we have now switched our definition of penalty parameter from $\mu_k$ to $\rho_k$. This notational switch is used to keep consistent with notation used previously in LCL algorithms. In these algorithms the penalty parameter $\rho_k$ may need to converge to infinity and therefore we may consider the penalty parameters to satisfy the relationship $\rho_k = 1/\mu_k$.

Subproblem (5.5) may suffer from two major deficiencies. First, the constraints $\bar{c}_k(x) = 0$ and $x \geq 0$ may be infeasible, in which case the subproblem has no solution. Second, the distance from the point of linearization to the subproblem solution may be arbitrarily large, i.e., the quantity $\|x^*_k - x_k\|$ may be arbitrarily large. These problems are addressed by Friedlander and Saunders [19], who regularize the standard LCL subproblem by including an $\ell_1$ penalty term of the linearized constraint violations. The analogous approach for the primal-dual augmented Lagrangian gives the so-called *elastic subproblem*:

$$\begin{array}{ll} \underset{x,y,u,v}{\text{minimize}} & \mathcal{M}(x, y\,; y^e_k, \rho_k) + \sigma_k e^T(u + v) \\ \text{subject to} & \bar{c}_k(x) + u - v = 0, \\ & x, u, v \geq 0, \quad -\gamma_k e \leq y \leq \gamma_k e. \end{array} \tag{5.6}$$

The point $(x^*_k,\ y^*_k,\ u^*_k,\ v^*_k,\ \Delta y^*_k,\ z^*_k,\ w^*_k)$ is regarded as an approximate solution of

subproblem (5.6) if it satisfies

$$x_k^* \geq 0, \tag{5.7a}$$
$$\|\min(x_k^*, z_k^*)\|_\infty \leq \omega_k, \tag{5.7b}$$
$$-\gamma_k e \leq y_k^* \leq \gamma_k e, \tag{5.7c}$$
$$\|\min(\gamma_k e - y_k^*, y_k^* + \gamma_k e, |w_k^*|)\|_\infty \leq \tau_k, \tag{5.7d}$$
$$\text{If } [\gamma_k e - y_k^*]_j \leq \tau_k, \text{ then } [w_k^*]_j \leq \tau_k, \tag{5.7e}$$
$$\text{If } [y_k^* + \gamma_k e]_j \leq \tau_k, \text{ then } [w_k^*]_j \geq -\tau_k, \tag{5.7f}$$
$$\begin{pmatrix} \nabla_x \mathcal{M}(x_k^*, y_k^*; y_k^e, \rho_k) - J(x_k)^T \Delta y_k^* \\ \nabla_y \mathcal{M}(x_k^*, y_k^*; y_k^e, \rho_k) \end{pmatrix} = \begin{pmatrix} z_k^* \\ w_k^* \end{pmatrix}, \tag{5.7g}$$
$$\bar{c}_k(x_k^*) + u_k^* - v_k^* = 0, \tag{5.7h}$$
$$\|\min(u_k^*,\ \sigma_k e - \Delta y_k^*)\|_\infty \leq \omega_k, \tag{5.7i}$$
$$\|\min(v_k^*,\ \sigma_k e + \Delta y_k^*)\|_\infty \leq \omega_k, \tag{5.7j}$$

where $z_k^*$ and $w_k^*$ are the multiplier vectors for $x \geq 0$ and $-\gamma_k e \leq y \leq \gamma_k e$, and $\Delta y_k^*$ denotes the Lagrange multiplier vector for the elastic linearized constraint $\bar{c}_k(x) + u - v = 0$. Note that inequalities (5.7i) and (5.7j) imply that

$$\|\Delta y_k^*\|_\infty \leq \sigma_k + \omega_k. \tag{5.8}$$

Algorithm 5.3 given below is very similar to the *stabilized* LCL algorithm proposed by Friedlander and Saunders [19]. The principal differences are: (i) Algorithm 5.3 uses the *primal-dual* augmented Lagrangian instead of the classical augmented Lagrangian; (ii) explicit artificial bounds are imposed on the dual variables of the subproblem; and (iii) an alternative update for $\sigma_k$ may be used in the situation where an iterate is labeled as "successful". An explanation for this alternative update is given below.

Based on the current degree of infeasibility, each pd$\ell_1$-LCL iterate is regarded as either "successful" or "unsuccessful". In the successful case, the solution estimates are updated by using information from the current subproblem solution. Next an optimality check is performed, followed by a decrease in the primal infeasibility parameter $\eta_k$. The penalty parameter $\sigma_k$ may also be decreased if $\Delta y_k^*$ is "too large". Finally, the index $j$, which represents the number of consecutive successful iterations, is incremented by one. In the case of an unsuccessful iteration, the subproblem solutions are discarded and then the penalty parameter $\rho_k$ is increased in an attempt to decrease the primal infeasibility at the next iteration. Next, the artificial bound $\gamma_k$ on the dual variables is increased, $\eta_k$ is reset, and $\sigma_k$ is decreased. Decreasing $\sigma_k$ is appropriate because small values of $\sigma_k$ encourage deviation from the linearized constraints, which may be necessary in order to decrease primal infeasibility.

In Algorithm 5.3, the vector $\widetilde{y}_k$ is a first-order *primal-dual* multiplier estimate, which is given by $\widetilde{y}_k = 2\pi(x_k^*) - y_k^* + \Delta y_k^*$, where $\pi(x_k^*) = y_k^e - c(x_k^*)/\mu_k$. In addition to the parameters $\eta_*$, $\omega_*$ and $\tau_*$ defined for pdBCL, the following parameters are associated with Algorithm 5.3:

- $\rho_0$ ($\rho_0 \geq 0$) is the initial penalty parameter;

- $\sigma_0$ ($\sigma_0 \geq 0$) is the initial $\ell_1$ penalty parameter for the linearized constraints;
- $\eta_0$ ($\eta_* \leq \eta_0$) is the initial primal infeasibility tolerance;
- $\omega_0$ ($\omega_* \leq \omega_0 < 1$) is the initial dual infeasibility tolerance for the $x$-variables;
- $\alpha_\eta$ ($0 < \alpha_\eta < 1$) is the constant used in the update to $\eta_k$ in the unsuccessful case.
- $\beta_\eta$ is the positive constant used in the update to $\eta_k$ in the successful case;
- $\tau_\rho$ ($\tau_\rho > 1$) is the expansion factor for $\rho_k$;
- $\tau_\sigma$ ($\tau_\sigma > 1$) is the scale-factor for $\sigma_k$;
- $\alpha_\gamma$ ($0 < \alpha_\gamma < 1$) is the constant used in the definition of $\gamma_k$.
- $\alpha_\tau$ ($\alpha_\tau < -1$) is the constant used in the definition of $\tau_k$;
- $k_\tau$ ($k_\tau > 0$) is used in the definition of $\tau_k$;
- $\delta$ ($\delta > 0$) is another positive constant;
- $\nu$ ($\nu > 0$) is used in definition of $\gamma_k$. Given $\rho_0$ and $y_0^e$, the parameter $\nu$ is set to $\nu = (\|y_0^e\|_\infty + 1)\rho_0^{-\alpha_\gamma}$. As $\gamma_0 = \nu\alpha_0^{\alpha_\gamma}$, it follows that $\gamma_0 = \|y_0^e\|_\infty + 1$. Therefore, $y_0^e$ will be in the interior of $[-\gamma_0 e, \gamma_0 e]$, which is unfortunate if $\|y_0\| \gg \|y^*\|$. In practice, an *a priori* limit on the norm of $\gamma_0$ should be enforced.

In the successful case, Friedlander and Saunders use the following update for $\sigma_k$ to deal with multiple limit points:

---

**Algorithm 5.2.** Update to $\sigma_k$ in the successful case (multiple limit points).

**if** $\|\Delta y_k^*\|_\infty \leq \delta(\frac{1}{2})^j$ **then**
  $\sigma_{k+1} \leftarrow \min(1 + \|\Delta y_k^*\|_\infty, \bar{\sigma})/(1 + \rho_k)$; [reset $\sigma_k$]
**else**
  $\sigma_{k+1} = \sigma_k/\tau_\sigma$; [decrease $\sigma_k$]
**end if**

---

The integer $j$ represents the number of consecutive successful major iterations and $\delta$ is a positive parameter. With this update, the authors claim global convergence without the single limit point assumption. They also note that any forcing sequence converging to zero may be used in the "if" part, but that requiring only a mild decrease in $\|\Delta y_k^*\|_\infty$ at each iteration should interfere less with the fast local convergence of the method since $\|\Delta y_k^*\|$ may be expected to decrease at a linear rate. The proof involves four cases. However, Case 2 does not appear to treat all the possible situations that may arise. In Case 2, $\{\rho_k\}$ is uniformly bounded and every iterate is successful for $k$ sufficiently large, which implies that $j \to \infty$. The authors argue that if $\Delta y_k^*$ does not satisfy the "if" part of Algorithm 5.2 infinitely often, then $\{\sigma_k\} \to 0$. However, this does not appear to be guaranteed since it is

possible that both the "if" and "else" statements occur infinitely often—implying that $\sigma_k$ alternates between being decreased and being reset.

However, it is true that if $\{x_k^*\}$ contains *finitely many* limit points, then there exists *some* subsequence $\bar{K}$ (say), such that $\lim_{k\in\bar{K}}(x_k^*, \widetilde{y}_k, z_k^*, \Delta y_k^*) = (x_*, y_*, z_*, 0)$ and that $(x_*, y_*, z_*)$ is a KKT point. This can be seen as follows. Let $K_1$, $K_2$, $\ldots$, $K_l$ be disjoint subsequences of the integers such that $\lim_{k\in K_i} x_k^* = x_*^i$, a limit point of $\{x_k^*\}$, for $i = 1, 2, \ldots, l$. Therefore, $\{x_*^i\}$ for $i = 1, 2, \ldots, l$, is the finite set of $l$ limit points of $\{x_k^*\}$. Since $\{\rho_k\}$ is assumed uniformly bounded, then all iterates are successful for $k$ sufficiently large. If the "if" part happens *finitely* often, then $\{\sigma_k\} \to 0$ and Part 2 of Lemma 5.2 by Friedlander and Saunders holds. If the "if" part occurs infinitely often, then it must be true that the "if" part occurs infinitely often on $K_m$ for some $1 \le m \le l$. It could occur infinitely often on more than one subsequence, but one will suffice. This implies that there exists a sub-subsequence $K_{\bar{m}} \subseteq K_m$ such that $\{\Delta y_k^*\}_{K_{\bar{m}}} \to 0$ since $j \to \infty$. Since $\{x_k^*\}_{K_{\bar{m}}} \to x_*^m$, $\{\Delta y_k^*\}_{K_{\bar{m}}} \to 0$, and $c(x_*^m) = 0$, Lemma 5.2 by Friedlander and Saunders implies that $(x_*^m, y_*^m, z_*^m)$ is a KKT point where

$$\lim_{k\in K_{\bar{m}}} \widetilde{y}_k = y_*^m \quad \text{and} \quad z_*^m = g(x_*^m) - J(x_*^m)^T y_*^m. \tag{5.9}$$

Note that this convergence occurs on the sub-subsequence $K_{\bar{m}}$, and not on $K_m$, as stated in the Friedlander-Saunders Lemma 5.2.

So it appears that the update used by the sLCL algorithm to handle *multiple* limit points *may* not work as predicted in certain situations. However, most practical problems have a unique limit point, and even when multiple limit points exist, it seems unlikely that the case described above will occur. It seems more likely that $\sigma_k$ will be driven to zero, which will cause the algorithm to behave like the (globally convergent) BCL method. The possible ramifications of this observation on the sLCL algorithm is unclear at this point. The update used in the stabilized primal-dual LCL algorithm below does not suffer from this potential problem. However, it is unclear at this point if this alternative updating scheme will give $\sigma$-values that are unnecessarily smaller than those of the Friedlander-Saunders algorithm.

---

**Algorithm 5.3.** Primal-Dual $\ell_1$-LCL Algorithm
INPUT: $(x_0, y_0)$
Set $\rho_0$, $\sigma_0$, $\eta_*$, $\eta_0$, $\omega_*$, $\tau_*$, $\alpha_\eta$, $\beta_\eta$, $\tau_\rho$, $\tau_\sigma$, $\alpha_\gamma$, $\alpha_\tau$, $k_\tau$, and $\delta$;
Set $y_0^e = y_0$; $z_0 = g(x_0) - J(x_0)^T y_0$; $w_0 = 0$;
Set $\nu = (\|y_0^e\|_\infty + 1)\rho_0^{-\alpha_\gamma}$; $\gamma_0 = \nu\rho_0^{\alpha_\gamma}$;
Set *converged* $\leftarrow$ **false**; $k \leftarrow 0$; $j \leftarrow 0$;
**while not** *converged* **do**
  Choose $\tau_k$, $\omega_k > 0$ so that $\lim_{k\to\infty}\tau_k \le \tau_*$, $\tau_k \le k_\tau\rho_k^{\alpha_\tau}$, and $\lim_{k\to\infty}\omega_k \le \omega_*$;
  Find $(x_k^*, y_k^*, u_k^*, v_k^*, \Delta y_k^*, z_k^*, w_k^*)$, a solution to (5.6), satisfying (5.7);
  Choose $(x_k^*, y_k^*, \Delta y_k^*, z_k^*, w_k^*)$ closest to $(x_k, y_k, z_k, w_k)$;
  **if** $\|c(x_k^*)\| \le \max(\eta_*, \eta_k)$ **then** [successful]
    $x_{k+1} \leftarrow x_k^*$; $y_{k+1} \leftarrow y_k^*$; $z_{k+1} \leftarrow z_k^*$; $y_{k+1}^e \leftarrow \widetilde{y}_k$;
    **if** $(x_{k+1}, y_{k+1}^e, z_{k+1})$ satisfies stopping criteria **then** *converged* $\leftarrow$ **true**;

$\rho_{k+1} \leftarrow \rho_k; \quad \gamma_{k+1} \leftarrow \gamma_k$
$\eta_{k+1} \leftarrow \eta_k/(1+\rho_{k+1}^{\beta_\eta});$ [decrease $\eta_k$]
**if** $\|\Delta y_k^*\|_\infty \geq \delta(\frac{1}{2})^j$ **then** $\sigma_{k+1} \leftarrow \sigma_k/\tau_\sigma;$ [decrease $\sigma_k$]
$j \leftarrow j+1;$
**else** [unsuccessful]
$x_{k+1} \leftarrow x_k; \quad y_{k+1} \leftarrow y_k; \quad z_{k+1} \leftarrow z_k; \quad y_{k+1}^e \leftarrow y_k^e;$
$\rho_{k+1} \leftarrow \tau_\rho \rho_k; \quad \gamma_{k+1} \leftarrow \nu \rho_{k+1}^{\alpha_\gamma};$ [increase $\rho_k$ and $\gamma_k$]
$\eta_{k+1} \leftarrow \eta_0/(1+\rho_{k+1}^{\alpha_\eta});$ [increase or decrease $\eta_k$]
$\sigma_{k+1} \leftarrow \sigma_k/\tau_\sigma;$ [decrease $\sigma_k$]
$j \leftarrow 0;$
**end if**
$k \leftarrow k+1;$
**end while**
OUTPUT: $(x^*, y^*, z^*) \leftarrow (x_k, y_k^e, z_k);$

### 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 is an appropriate merit function that has the potential of forcing convergence to points satisfying the second-order necessary conditions for optimality. This is a consequence of Theorem 3.1, which shows implies that minimizers of problem (NEP) are also minimizers of the primal-dual augmented Lagrangian function. The formulation and analysis of an SQP method in this context will be the topic of a future 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 simultaneously with respect to both the primal and the dual variables. In its most general form, the function may be considered as one of a continuum of functions that have certain well-known functions as specific cases. One variant of this generalized function—the primal-dual augmented Lagrangian—is proposed as the basis of two primal-dual methods. The first is a primal-dual bound-constrained Lagrangian method (pdBCL) based on a primal method given by Conn, Gould, and Toint [9]. The second method is a primal-dual linearly constrained Lagrangian method (pd$\ell_1$LCL) based on the method of Friedlander and Saunders [19].

## References


[1] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Computer Science and Applied Mathematics. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1982.

[2] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Athena Scientific, Belmont, Massachusetts, 1996.

[3] M. C. Biggs. Constrained minimization using recursive equality quadratic programming. In F. A. Lootsma, editor, *Numerical Methods for Nonlinear Optimization*, pages 411–428. Academic Press, London and New York, 1972.

[4] P. T. Boggs, A. J. Kearsley, and J. W. Tolle. A global convergence analysis of an algorithm for large-scale nonlinear optimization problems. *SIAM J. Optim.*, 9(4):833–862, 1999.

[5] P. T. Boggs and J. W. Tolle. A family of descent functions for constrained optimization. *SIAM J. Numer. Anal.*, 21:1146–1161, 1984.

[6] P. T. Boggs and J. W. Tolle. Sequential quadratic programming. In *Acta Numerica, 1995*, Acta Numer., pages 1–51. Cambridge Univ. Press, Cambridge, 1995.

[7] R. H. Byrd, R. A. Tapia, and Y. Zhang. An SQP augmented Lagrangian BFGS algorithm for constrained optimization. *SIAM J. Optim.*, 20:210–241, 1992.

[8] A. R. Conn, N. I. M. Gould, and P. L. Toint. A comprehensive description of LANCELOT. Technical Report 91/10, Département de Mathématique, Facultés Universitaires de Namur, 1991.

[9] A. R. Conn, N. I. M. Gould, and P. L. Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM J. Numer. Anal.*, 28:545–572, 1991.

[10] A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust-Region Methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[11] G. DiPillo and L. Grippo. A new class of augmented Lagrangians in nonlinear programming. *SIAM J. Control Optim.*, 17:618–628, 1979.

[12] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley and Sons, Inc., New York-London-Sydney, 1968.

[13] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 1990. Reprint of the 1968 original.

[14] R. Fletcher. A class of methods for nonlinear programming with termination and convergence properties. In J. Abadie, editor, *Integer and Nonlinear Programming*, pages 157–175. North-Holland, The Netherlands, 1970.

[15] R. Fletcher. *Practical Methods of Optimization*. Volume 2: Constrained Optimization. John Wiley and Sons, Chichester and New York, 1981.

[16] R. Fontecilla, T. Steihaug, and R. A. Tapia. A convergence theory for a class of quasi-Newton methods for constrained optimization. *SIAM J. Numer. Anal.*, 24:1133–1151, 1987.

[17] A. Forsgren and P. E. Gill. Primal-dual interior methods for nonconvex nonlinear programming. *SIAM J. Optim.*, 8:1132–1152, 1998.

[18] A. Forsgren, P. E. Gill, and M. H. Wright. Interior methods for nonlinear optimization. *SIAM Rev.*, 44:525–597, 2002.

[19] M. P. Friedlander and M. A. Saunders. A globally convergent linearly constrained Lagrangian method for nonlinear optimization. *SIAM J. Optim.*, 15(3):863–897 (electronic), 2005.

[20] E. M. Gertz and P. E. Gill. A primal-dual trust-region algorithm for nonlinear programming. *Math. Program., Ser. B*, 100:49–94, 2004.

[21] P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Rev.*, 47:99–131, 2005.

[22] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. Some theoretical properties of an augmented Lagrangian merit function. In P. M. Pardalos, editor, *Advances in Optimization and Parallel Computing*, pages 101–128. North Holland, North Holland, 1992.

[23] N. I. M. Gould. On the accurate determination of search directions for simple differentiable penalty functions. *IMA J. Numer. Anal.*, 6:357–372, 1986.

[24] W. W. Hager. Stabilized sequential quadratic programming. *Comput. Optim. Appl.*, 12(1-3):253–273, 1999. Computational optimization—a tribute to Olvi Mangasarian, Part I.

[25] M. R. Hestenes. Multiplier and gradient methods. *J. Optim. Theory Appl.*, 4:303–320, 1969.

[26] W. Murray. *Constrained Optimization*. PhD thesis, Department of Computer Science, University of London, 1969.

[27] B. A. Murtagh and M. A. Saunders. MINOS 5.5 User's Guide. Report SOL 83-20R, Department of Operations Research, Stanford University, Stanford, CA, Revised 1998.

[28] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, pages 283–298, London and New York, 1969. Academic Press.

[29] D. P. Robinson. *Primal-Dual Methods for Nonlinear Optimization*. PhD thesis, Department of Mathematics, University of California San Diego, September 2007.

[30] S. M. Robinson. A quadratically-convergent algorithm for general nonlinear programming problems. *Math. Program.*, 3:145–156, 1972.

[31] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1(2):97–116, 1976.

[32] R. T. Rockafellar. Monotone operators and the proximal-point algorithm. *SIAM J. Control Optim.*, 14:877–898, 1976.

[33] R. A. Tapia. Diagonalized multiplier methods and quasi-Newton methods for constrained optimization. *J. Optim. Theory Appl.*, 22:135–194, 1977.

[34] M. H. Wright. Interior methods for constrained optimization. In *Acta Numerica, 1992*, pages 341–407. Cambridge University Press, New York, USA, 1992.

[35] S. J. Wright. An algorithm for degenerate nonlinear programming with rapid local convergence. *SIAM J. Optim.*, 15(3):673–696 (electronic), 2005.

[36] S. J. Wright and D. Orban. Properties of the log-barrier function on degenerate nonlinear programs. *Math. Oper. Res.*, 27(3):585–613, 2002.