Acta Mathematica Sinica, English Series
Sep., 201x, Vol. x, No. x, pp. 1–11
Published online: August 15, 201x
DOI: 0000000000000000
Http://www.ActaMath.com

Analysis on a superlinearly convergent augmented Lagrangian method

Ya-xiang YUAN
State Key Laboratory of Scientific/Engineering Computing,
Institute of Computational Mathematics and Scientific/Engineering Computing,
Academy of Mathematics and Systems Science, Chinese Academy of Sciences,
Zhong Guan Cun Donglu 55, Beijing, 100190, P.R. China,
E-mail: yyx@lsec.cc.ac.cn

Abstract The augmented Lagrangian method is a classical method for solving constrained optimization. Recently, the augmented Lagrangian method attracts much attention due to its applications to sparse optimization in compressive sensing and low rank matrix optimization problems. However, most Lagrangian methods use first order information to update the Lagrange multipliers, which lead to only linear convergence. In this paper, we study a update technique based on second order information and prove that superlinear convergence can be obtained. Theoretical properties of the update formula are given and some implementation issues regarding the new update are also discussed.

Keywords nonlinearly constrained optimization, augmented Lagrange function, Lagrange multiplier, convergence.

MR(2010) Subject Classification 90C30, 65N05

1 Introduction

The invention of the augmented Lagrange function is one of the most important milestones in the development of nonlinear optimization. Consider the general nonlinear optimization problem which has the following form

$$\min_{x\in\Re^n} \quad f(x) \tag{1.1}$$
$$\text{s. t.} \quad c_i(x) = 0,\ i = 1, ..., m_e; \tag{1.2}$$
$$c_i(x) \geq 0, i = m_e + 1, ..., m, \tag{1.3}$$

where $f(x)$, $c_i(x)(i = 1, ..., m)$ are continuous function defined in $\Re^n$ and $m \geq m_e$ are two non-negative integers. Pioneer works on the augmented Lagrange function were given by Hestenes[1], Powell[2], Rockafellar[3] and Fletcher[4]. The augmented Lagrange function for general constrained optimization problem (1.1)-(1.3) is defined by

$$P(x, \lambda, \sigma) \quad = \quad f(x) - \sum_{i=1}^{m}\left[\lambda^{(i)}c_i(x) - \frac{1}{2}\sigma^{(i)}(c_i(x))^2\right]$$

Received x x, 201x, accepted x x, 201x
this work was partially supported by NSFC grants 10831006, 11021101 and by CAS grant kjcx-yw-s7

$$-\sum_{i=m_e+1}^{m}\begin{cases}\lambda^{(i)}c_i(x)-\frac{1}{2}\sigma^{(i)}(c_i(x))^2, & \text{if } c_i(x)<\lambda^{(i)}/\sigma^{(i)},\\ \frac{1}{2}(\lambda^{(i)})^2/\sigma^{(i)}, & \text{otherwise.}\end{cases} \tag{1.4}$$

Here, $\sigma^{(i)} > 0 (i = 1, ..., m)$ are the penalty parameters, and $\lambda = (\lambda^{(1)}, ..., \lambda^{(m)})^T$ are the Lagrange multipliers satisfying $\lambda^{(i)} \geq 0$ for $i > m_e$.

One good property of the augmented Lagrange function is that it is an exact penalty function if λ is the Lagrange multiplier at the solution and if the penalty parameters are sufficiently large. Due to this nice property, the augmented Lagrangian method generates a sequence of points $\{x_k,\ k = 1, 2, ...\}$, each of them is an approximate solution of

$$\min_{x\in\Re^n} P(x, \lambda_k, \sigma_k). \tag{1.5}$$

Define the vector $c^{(-)}$ by

$$c_i^{(-)}(x)=\begin{cases}c_i(x), & i=1,...,m_e\,,\\ \min[c_i(x),\,0] & i=m_e+1,...,m\,.\end{cases} \tag{1.6}$$

It is easy to see that the minimizer $\bar{x}_k$ of the unconstrained optimization problem (1.5) is a KKT point of the original constrained optimization problem (1.1)-(1.3), if $\bar{x}_k$ is a feasible point of (1.2)-(1.3), namely $c^{(-)}(\bar{x}_k) = 0$. Thus, the augmented Lagrangian method tries to update λ_k and σ_k to reduce $||c^{(-)}(\bar{x}_k)||$ to certain error level. The augmented Lagrangian method can be stated as follows[5].

Algorithm 1.1 (Augmented Lagrangian Method)

Step 1 *Given starting point* $x_1 \in \Re^n$, $\lambda_1 \in \Re^m$ *with* $\lambda_1^{(i)} \geq 0 (i > m_e)$; $\sigma_1^{(i)} > 0 (i = 1, ..., m)$; $\epsilon \geq 0$, $k := 1$.

Step 2 *Find an approximate solution* x_{k+1} *of (1.5).*
If $||c^{(-)}(x_{k+1})||_\infty \leq \epsilon$, *then stop.*

Step 3 *For i=1,...,m, set*

$$\sigma_{k+1}^{(i)}=\begin{cases}\sigma_k^{(i)}\,, & \textit{if } |c_i^{(-)}(x_{k+1})|\leq|c_i^{(-)}(x_k)|/4,\\ \max[10\sigma_k^{(i)},k^2]\,, & \textit{otherwise.}\end{cases} \tag{1.7}$$

Step 4 *Obtain* λ_{k+1} *by*

$$\lambda_{k+1}^{(i)} = \lambda_k^{(i)} - \sigma_k^{(i)} c_i(x_{k+1}), \qquad i = 1, \cdots, m_e, \tag{1.8}$$

$$\lambda_{k+1}^{(i)} = \max\{\lambda_k^{(i)} - \sigma_k^{(i)} c_i(x_{k+1}), 0\}, \quad i = m_e + 1, \cdots, m\,, \tag{1.9}$$

$k := k + 1$, *go to Step 2.*

For more details of the augmented Lagrangian method and its properties, please see Conn, Gould and Toint[6] and Sun and Yuan[5].

The software package LANCELOT, based on the augmented Lagrangian method, is one of most efficient nonlinear optimization solvers. In 1994, Andrew Conn, Nick Gould and Philippe Toint won the Beale-Orchard-Hays prize from the Mathematical Programming Society for their work on the LANCELOT package. Recently, the augmented Lagrange function method attracts much attention due to the fact that it is widely used in L_1 minimization for compressive sensing problems (for example, see [7–9]), semi-definite programming (for example, see [10, 11]) and low rank matrix optimization problems)(for example, see [12, 13]).

The key parts of an augmented Lagrangian method are finding an approximate minimizer of the augmented Lagrange function and updating the Lagrange multipliers and the penalty parameters. We notice that in all the augmented Lagrangian methods, the update rule for the Lagrange multiplier λ is (1.8)-(1.9). The aim of the paper is to investigate the properties of the update techniques and to explore whether more efficient updates are possible.

The paper is organized as follows. In the next section, we re-derive an update formula for λ which was discovered by Buys[14]. In Section 3, some properties of Buys's update are given. Finally, brief discussions are given in Section 4.

2 New Derivation of an Old Update Formulae for λ

In this section, we re-derive an old update formula for Lagrange multipliers. First we will derive the update formula by considering the simple case when all constraints are equality constraints. Then we will extend our results to the general case when there are both equality and inequality constraints.

2.1 Equality Constraints

For simplicity, first we consider only equality constraints. Let x^* be a local minimizer of equality constrained optimization problem

$$\min_{x\in\Re^n} \quad f(x) \tag{2.1}$$

$$\text{s. t.} \quad c_i(x) = 0, \quad i = 1, ..., m, \tag{2.2}$$

and λ^* be the corresponding Lagrange multiplier λ^* satisfying

$$\nabla f(x^*) - \sum_{i=1}^{m} (\lambda^*)^{(i)} \nabla c_i(x^*) \;=\; 0\,. \tag{2.3}$$

If the second order sufficient condition holds at x^*, there exists a positive number σ^* such that $P(x, \lambda^*, \sigma)$ is an exact penalty function for any σ as long as

$$\sigma^{(i)} \geq \sigma^* \quad \text{for all } \; i = 1, ..., m. \tag{2.4}$$

Since the aim of this paper is to investigate efficient techniques for updating λ, we assume that $\sigma_k^{(i)} \geq \sigma^*$ for all k.

Let $\bar{x}_k$ be a minimizer of $P(x, \lambda_k, \sigma_k)$, the optimality condition implies that

$$\nabla f(\bar{x}_k) - \sum_{i=1}^{m} [\lambda_k^{(i)} - \sigma_k^{(i)} c_i(\bar{x}_k)] \nabla c_i(\bar{x}_k) \;=\; 0\,. \tag{2.5}$$

Comparing (2.3) and (2.5), we can easily see that, $\lambda_k^{(i)} - \sigma_k^{(i)} c_i(\bar{x}_k)$ are approximate Lagrange multipliers. In fact, if $\nabla c_i(\bar{x}_k)(i = 1, ..., m)$ are linearly independent, (2.5) indicates that $\lambda_k^{(i)} - \sigma_k^{(i)} c_i(\bar{x}_k)(i = 1, ..., m)$ is the unique solution of the linear system $A(\bar{x}_k)\lambda - \nabla f(\bar{x}_k) = 0$, hence it is also exactly the same multiplier of Fletcher's differentiable exact penalty function in which λ is the least square solution of $\min_\lambda ||A(x)\lambda - \nabla f(x)||_2^2$.

In a practical augmented Lagrangian method, the iterate point x_{k+1} is an approximation to $\bar{x}_k$, hence it is reasonable to set the next Lagrange multipliers by

$$\lambda_{k+1}^{(i)} = \lambda_k^{(i)} - \sigma_k^{(i)} c_i(x_{k+1}) \,. \tag{2.6}$$

which is exactly (1.8) when all the constraints are equalities.

The update formula (2.6) is used by most augmented Lagrangian methods, for example, see [6] and [5]. Now we consider the convergence property of the update formular (2.6). Consider the ideal case when $x_{k+1} = \bar{x}_k = argmin P(x, \lambda_k, \sigma_k)$. Define the diagonal matrix $\Sigma_k = diag[\sigma_k^{(1)}, ..., \sigma_k^{(m)}]$, the optimality conditions (2.3) and (2.5) give that

$$[\nabla^2_{xx} L(x^*, \lambda^*) + A(x^*)\Sigma_k A(x^*)^T](x_{k+1} - x^*) + O(||x_{k+1} - x^*||_2^2) = A(x_{k+1})(\lambda_k - \lambda^*) \,, \tag{2.7}$$

where $L(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i c_i(x)$ and $A(x)$ is defined by

$$A(x) = [\nabla c_1(x), \nabla c_2(x), ..., \nabla c_m(x)] \,. \tag{2.8}$$

Denote $W(x^*, \lambda^*, \sigma_k) = \nabla^2_{xx} L(x^*, \lambda^*) + A(x^*)\Sigma_k A(x^*)^T$, which is positive definite due to our assumption that $\sigma_k^{(i)} \geq \sigma^*$ for all k. Thus,

$$\begin{aligned} \lambda_{k+1} - \lambda^* &= \lambda_k - \lambda^* - \Sigma_k A(x^*)^T (x_{k+1} - x^*) + O(||x_{k+1} - x^*||_2^2) \\ &= \left[I - \Sigma_k A(x^*)^T [W(x^*, \lambda^*, \sigma)]^{-1} A(x^*)\right] (\lambda_k - \lambda^*) + o(||\lambda_k - \lambda^*||) \,. \end{aligned} \tag{2.9}$$

Therefore, the above relation indicates that, unless

$$A(x^*)^T \nabla^2_{xx} L(x^*, \lambda^*) A(x^*) = 0 \,,$$

$||\lambda_k - \lambda^*||$ converges to zero only Q-linearly. Consequently, it follows from (2.7) that x_k converges to x^* only Q-linearly. Indeed, Sun and Yuan[5] shows that

$$\left[\Sigma A(x^*)^T + A(x^*)^+ \nabla^2_{xx} L(x^*, \lambda^*)\right] (x_{k+1} - x^*) \approx A(x^*)^+ \nabla^2_{xx} L(x^*, \lambda^*)(x_k - x^*) \,. \tag{2.10}$$

Consider the following toy problem:

$$\min_{(x_1, x_2) \in \Re^2} \quad -\frac{1}{2}(x_1^2 - x_2^2) \tag{2.11}$$

$$\text{s. t.} \quad x_1 = 0. \tag{2.12}$$

This is easy to see that the solution of (2.11)-(2.12) is $x^* = (x_1^*, x_2^*)^T = (0, 0)^T$ and the corresponding Lagrange multiplier at x^* is $\lambda^* = 0$. For any $\sigma > 1$, the augmented Lagrange function $-\frac{1}{2}(x_1^2 - x_2^2) - \lambda^* x_1 + \frac{1}{2}\sigma(x_1)^2$ is an exact penalty function. We consider the case when we choose $\sigma = 2.01$. It turns out that for any given λ_k, the minimizer of the augmented Lagrange function $-\frac{1}{2}(x_1^2 - x_2^2) - \lambda_k x_1 + \frac{1}{2}\sigma(x_1)^2$ is $x_{k+1}^{(1)} = \frac{100}{101}\lambda_k$ and $x_{k+1}^{(2)} = 0$. Thus,

$$\lambda_{k+1} = \lambda_k - 2.01 * \frac{100}{101}\lambda_k = -\frac{100}{101}\lambda_k \,.$$

This example and our above theoretical analysis indicate that the update technique (2.6) leads to only linearly convergence. In general, (2.9) indicates that $||\lambda_{k+1} - \lambda^*|| = O(\frac{1}{\sigma_k})||\lambda_k - \lambda^*||)$ if σ_k is very large. Thus, when the penalty parameters σ_k remain bounded, update formulae (1.8)-(1.9) converge only linearly.

Now we derive an update formula for the Lagrange multipliers that ensures superlinearly convergence. Consider λ is an approximation to the exact Lagrange multiplier λ^*. Let $x(\lambda)$ be the minimizer of $P(x, \lambda, \sigma)$. The optimality condition implies that

$$\nabla_x f(x(\lambda)) - A(x(\lambda))\lambda + A(x(\lambda))\Sigma c(x(\lambda)) = 0\,. \tag{2.13}$$

We try to obtain an update formula by applying Newton's method to

$$c(x(\lambda)) = 0\,, \tag{2.14}$$

which can be written as

$$c(x(\lambda_k)) + A(x(\lambda_k))^T \left(\frac{\partial x}{\partial \lambda}\right)^T_{\lambda=\lambda_k} [\lambda_{k+1} - \lambda_k] = 0\,, \tag{2.15}$$

where

$$\frac{\partial x}{\partial \lambda} = [\nabla_\lambda x_1(\lambda), \nabla_\lambda x_2(\lambda), ..., \nabla_\lambda x_n(\lambda)]\ . \tag{2.16}$$

Differentiating (2.13) with respect to λ gives

$$\left[\nabla^2 f(x) - \sum_{i=1}^{m}[\lambda^{(i)} - \sigma_i c_i(x)]\nabla c_i^2(x) + A(x)\Sigma A(x)^T\right]\left(\frac{\partial x}{\partial \lambda}\right)^T = A(x)\,, \tag{2.17}$$

where $x = x(\lambda)$. The above equation is the detailed form of the following relation

$$\nabla^2_{xx} P(x, \lambda, \sigma)\left(\frac{\partial x}{\partial \lambda}\right)^T = A(x)\,. \tag{2.18}$$

Thus, Newton's method for (2.14) has the following form

$$c(x(\lambda_k)) + A(x(\lambda_k))^T(\nabla^2_{xx} P(x(\lambda_k), \lambda_k, \sigma_k)^{-1} A(x(\lambda_k))[\lambda_{k+1} - \lambda_k] = 0\,, \tag{2.19}$$

which gives

$$\lambda_{k+1} = \lambda_k - \left[A(x_{k+1})^T \nabla^2_{xx} P(x_{k+1}, \lambda_k, \sigma_k)^{-1} A(x_{k+1})\right]^{-1} c(x_{k+1})\,, \tag{2.20}$$

if we let $x_{k+1} = x(\lambda_k)$. Unfortunately, the above update (2.20) is not new, it was first discovered by Buys[14], and later discussed in [15]. Indeed, Fontecilla, Steihaug and Tapia[15] also considered other possible updates for the Lagrangian multiplier λ. However, it should be pointed out that our technique for deriving the update formular (2.20) is different from that of Buys[14]. We are trying to adjust the Lagrange multiplier in order to satisfy the feasibility condition $c(x(\lambda)) = 0$ while Buys[14] derive the update (2.20) thorough Newton's method for the dual problem.

Now let us look at the toy example (2.11)-(2.12) again. For any λ_k and $\sigma_k > 1$, we have that $x^{(1)}_{k+1} = \frac{\lambda_k}{\sigma_k - 1}$ and $x^{(2)}_{k+1} = 0$. Thus, $c(x_{k+1}) = \frac{\lambda_k}{\sigma_k - 1}$, $A(x_{k+1}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and

$$\nabla^2_{xx} P(x_{k+1}, \lambda_k, \sigma_k) = \begin{bmatrix} \sigma_k - 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Update formula (2.20) will gives

$$\lambda_{k+1} = \lambda_k - (\sigma_k - 1)\frac{\lambda_k}{\sigma_k - 1} = 0 = \lambda^*\,.$$

This indicates that the Augmented Lagrange function method with updating technique (2.20) will find the exact solution of the toy problem after two iterations for any initial λ_1 and any $\sigma_1 > 1$.

For general nonlinear equality constrained problems, we have the following result.

Theorem 2.1 *Let x^* be a KKT point of (2.1)-(2.2). If the gradient of the constraints $\nabla c_i(x^*)(i = 1, ..., m)$ are linearly independent, and the second order sufficient condition holds at x^*, if the penalty parameters $\sigma_k^{(i)}$ are large enough, then the local convergence rate of update formula (2.20) is Q-quadratic in the sense that*

$$||\lambda_{k+1} - \lambda^*|| = O(||\lambda_k - \lambda^*||_2^2) . \tag{2.21}$$

Proof. The assumptions of our theorem imply that $A(x^*)$ is full column rank and the matrix $\nabla^2_{xx} P(x^*, \lambda^*, \sigma^*)$ is positive definite for some fixed vector σ^*. Therefore, the matrix

$$\left[A(x^*)^T \nabla^2_{xx} P(x^*, \lambda^*, \bar{\sigma})^{-1} A(x^*)\right]$$

is positive definite for any fixed vector $\bar{\sigma} \geq \sigma^*$.

Define $x(\lambda)$ to be the minimizer of $P(x, \lambda, \bar{\sigma})$ and $\psi(\lambda) = c(x(\lambda))$. Thus, our update formula for λ (2.20) is exactly Newton's method for $\psi(\lambda) = 0$. Our assumption indicates that $\psi(\lambda^*) = 0$ and $\nabla \psi(\lambda^*)$ is non-singular. Thus, when λ_k is sufficiently close to λ^*, we know that (2.21) holds. □

A direct consequence of the above result is that $||x_k - x^*||$ converges to zero R-quadratically, due to the fact that $||x_k - x^*|| = O(||\lambda_k - \lambda^*||)$.

2.2 Inequality constraints

Now we consider the general case when there are both equality and inequality constraints. Let x^* be a KKT point of (1.1)-(1.3). Denote $x(\lambda, \sigma)$ be the minimizer of the augmented Lagrange function $P(x, \lambda, \sigma)$ defined by (1.4).

Define the sets $\mathcal{E} = \{1, ..., m_e\}$, $\mathcal{I} = \{m_e + 1, ..., m\}$, and

$$\mathcal{I}(x(\lambda, \sigma)) = \{i \mid i \in \mathcal{I},\ c_i(x(\lambda, \sigma)) < \lambda^{(i)}/\sigma^{(i)}\} .$$

The optimality condition for $x(\lambda, \sigma)$ implies that

$$\nabla f(x(\lambda, \sigma)) - \sum_{i \in \mathcal{E} \cup \mathcal{I}(x(\lambda, \sigma))} \lambda^{(i)} \nabla c_i(x(\lambda, \sigma)) = 0 . \tag{2.22}$$

We assume that second order sufficient condition holds at $x(\lambda, \sigma)$, which means that

$$\begin{aligned} W(\lambda, \sigma) \quad = \quad & \nabla^2 f(x(\lambda, \sigma)) - \sum_{i \in \mathcal{E} \cup \mathcal{I}(x(\lambda, \sigma))} [\lambda^{(i)} - \sigma^{(i)} c_i(\lambda, \sigma)] \nabla^2 c_i(x(\lambda, \sigma)) \\ + \quad & \sum_{i \in \mathcal{E} \cup \mathcal{I}(x(\lambda, \sigma))} \sigma^{(i)} \nabla c_i(x(\lambda, \sigma)) [\nabla c_i(x(\lambda, \sigma))]^T \end{aligned} \tag{2.23}$$

is positive definite.

Now, for fixed σ, our updating technique for λ is based on Newton's method for the following system of equations

$$c_i(x(\lambda, \sigma)) = 0, \qquad i \in \mathcal{E} \cup \mathcal{I}(x(\lambda, \sigma)) , \tag{2.24}$$

$$\lambda^{(i)} = 0, \qquad \text{otherwise.} \tag{2.25}$$

For given λ_k and σ_k, denote $x_{k+1} = x(\lambda_k, \sigma_k)$, $\mathcal{J}_k = \mathcal{E} \cup \mathcal{I}(x_{k+1})$, and

$$\bar{c}_{k+1} = [c_i(x_{k+1}),\, i \in \mathcal{J}_k]^T \tag{2.26}$$

$$\bar{A}_{k+1} = [\nabla c_i(x_{k+1}),\, i \in \mathcal{J}_k] \tag{2.27}$$

$$\bar{\lambda} = [\lambda^{(i)}\,,\, i \in \mathcal{J}_k]^T\,. \tag{2.28}$$

Our results in the previous section tell us that Newton's method for (2.24)-(2.25) has the following form:

$$\bar{\lambda}_{k+1} = \bar{\lambda}_k - [\bar{A}_{k+1}^T \bar{W}(x_{k+1}, \lambda_k, \sigma_k)^{-1} \bar{A}(x_{k+1})]^{-1} \bar{c}_{k+1}\,, \tag{2.29}$$

$$\lambda_{k+1}^{(i)} = 0,\, i \notin \mathcal{J}_k\,, \tag{2.30}$$

where

$$\bar{W}(x_{k+1}, \lambda_k, \sigma_k) = \nabla^2 f(x_{k+1}) - \sum_{i \in \mathcal{J}_k} [\lambda_k^{(i)} - \sigma_k^{(i)} c_i(x_{k+1})] \nabla^2 c_i(x_{k+1}) + \sum_{i \in \mathcal{J}_k} \sigma_k^{(i)} \nabla c_i(x_{k+1}) (\nabla c_i(x_{k+1}))^T\,. \tag{2.31}$$

Thus, we have derived update formulae (2.29)-(2.30) for updating Lagrange multipliers in augmented Lagrangian method applied to general nonlinear constrained optimization problems (1.1)-(1.3). In the following, we will call (2.29)-(2.30) as Buys's update.

3 Properties of Buys's Update

In order to study the relations between Buys's update (2.29)-(2.30) and the classic technique (1.8)-(1.9), we need some elementary results on real matrices.

Lemma 3.1 *Let the block matrix*

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \in \Re^{n \times n} \tag{3.1}$$

be nonsingular, where B_{11}, B_{11}, B_{21} and B_{22} are matrices in $\Re^{m\times m}$, $\Re^{m\times(n-m)}$, $\Re^{(n-m)\times m}$ and $\Re^{(n-m)\times(n-m)}$ respectively. If B_{22} is also nonsingular, then the matrix $(B_{11} - B_{12}B_{22}^{-1}B_{21})$ is also nonsingular, and moreover, the block representation of B^{-1} can be written as

$$B^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}^{-1} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \tag{3.2}$$

such that

$$H_{11} = (B_{11} - B_{12}B_{22}^{-1}B_{21})^{-1}\,. \tag{3.3}$$

Proof. This lemma can be proved by using the Schur complement (for example, see exercise 20.3 in [16]). Here we give a direct proof. The representation (3.2) requires that

$$B_{11}H_{11} + B_{12}H_{21} = I\,, \tag{3.4}$$

$$B_{21}H_{11} + B_{22}H_{21} = 0\,. \tag{3.5}$$

It follows from (3.5) that $H_{21} = -B_{22}^{-1}B_{21}H_{11}$. Substituting this relation into (3.4), we see that (3.3) holds. □

Theorem 3.2 *Let $A \in \Re^{n\times m}$ be full column rank, $B \in \Re^{n\times n}$ be symmetric positive definite, and $H \in \Re^{m\times m}$ be symmetric positive definite. Then*

$$\left[A^T(B + AHA^T)^{-1}A\right]^{-1} = H + (A^TB^{-1}A)^{-1} . \tag{3.6}$$

Proof. Since A is full column rank, there exist a unitary matrix $Q \in \Re^{n\times n}$ and a nonsingular upper triangular matrix $R \in \Re^{m\times m}$ such that

$$A = Q \begin{bmatrix} R \\ 0 \end{bmatrix} . \tag{3.7}$$

Let $Q = [Q_1, Q_2]$, where $Q_1 \in \Re^{n\times m}$ and $Q_2 \in \Re^{n\times(n-m)}$. The left hand side of (3.6) can be rewritten as

$$\left([R^T\,0] \left[\begin{pmatrix} Q_1^TBQ_1 & Q_1^TBQ_2 \\ Q_2^TBQ_1 & Q_2^TBQ_2 \end{pmatrix} + \begin{pmatrix} RHR^T & 0 \\ 0 & 0 \end{pmatrix} \right]^{-1} \begin{pmatrix} R \\ 0 \end{pmatrix} \right)^{-1} .$$

By applying Lemma 3.1, the above term can be simplified as

$$\begin{aligned} &(R^T[RHR^T + Q_1^TBQ_1 - Q_1^TBQ_2(Q_2^TBQ_2)^{-1}Q_2^TBQ_1]^{-1}R)^{-1} \\ &= H + R^{-1}[Q_1^TBQ_1 - Q_1^TBQ_2(Q_2^TBQ_2)^{-1}Q_2^TBQ_1]R^{-T} . \end{aligned} \tag{3.8}$$

Now, we look at the right hand of (3.6), which can be written as

$$H + (R^TQ_1^TB^{-1}Q_1R)^{-1} = H + R^{-1}(Q_1^TB^{-1}Q_1)^{-1}R^T . \tag{3.9}$$

Noticing the fact that

$$\begin{pmatrix} Q_1^TBQ_1 & Q_1^TBQ_2 \\ Q_2^TBQ_1 & Q_2^TBQ_2 \end{pmatrix}^{-1} = \begin{pmatrix} Q_1^TB^{-1}Q_1 & Q_1^TB^{-1}Q_2 \\ Q_2^TB^{-1}Q_1 & Q_2^TB^{-1}Q_2 \end{pmatrix} , \tag{3.10}$$

Lemma 3.1 implies that

$$(Q_1^TB^{-1}Q_1)^{-1} = Q_1^TBQ_1 - Q_1^TBQ_2(Q_2^TBQ_2)^{-1}Q_2^TBQ_1 . \tag{3.11}$$

This relation, (3.8) and (3.9) show that (3.6) holds. □

A direct consequence of the above the theorem is the following result.

Corollary 3.3 *Let $A \in \Re^{n\times m}$ be full column rank, $B \in \Re^{n\times n}$ be symmetric positive definite, and $H \in \Re^{m\times m}$ be symmetric positive definite. Then*

$$\lim_{\epsilon\to 0_+} A^T[\epsilon B + AHA^T]^{-1}A = H^{-1} . \tag{3.12}$$

The above results on real matrices help us to explore the relations between Buys's update and the classic update technique for the Lagrange multipliers.

First, we consider under what condition Buys's update reduced to the class one. From (2.23), we can write

$$W(\lambda_k, \sigma_k) = \nabla^2 f(x_{k+1}) - \sum_{i\in\mathcal{J}_k} (\lambda_k^{(i)} - \sigma_k^{(i)} c_i(x_{k+1}))\nabla^2 c_i(x_{k+1}) \;+\; \bar{A}_{k+1}\bar{\Sigma}_k\bar{A}_{k+1}^T ,$$

where $\bar{\Sigma}_k = Diag(\sigma_k^{(i)}\,,\, i \in \mathcal{J}_k)$. If we replace

$$\nabla^2 f(x_{k+1}) - \sum_{i\in\mathcal{J}_k} (\lambda_k^{(i)} - \sigma_k^{(i)} c_i(x_{k+1}))\nabla^2 c_i(x_{k+1})$$

by ϵB, and let $\epsilon \to 0_+$, the above lemma shows that update (2.29)-(2.30) would be changed to

$$\lambda_{k+1}^{(i)} = \lambda_k^{(i)} - \sigma_k^{(i)} c_i(x_{k+1}), \qquad i \in \mathcal{J}_k, \tag{3.13}$$

$$\lambda_{k+1}^{(i)} = 0, \qquad \text{otherwise}, \tag{3.14}$$

which are exactly the same as the classical update formulae (1.8)-(1.9). Therefore, the class update formulae (1.8)-(1.9) can be viewed in some sense as a special case of Buys's update where the matrix $\nabla^2 f(x_{k+1}) - \sum_{i\in\mathcal{J}_k}(\lambda_k^{(i)} - \sigma_k^{(i)} c_i(x_{k+1}))\nabla^2 c_i(x_{k+1})$ is replaced by a zero matrix.

In general, we denote

$$B_k = \nabla^2 f(x_{k+1}) - \sum_{i\in\mathcal{J}_k}(\lambda_k^{(i)} - \sigma_k^{(i)} c_i(x_{k+1}))\nabla^2 c_i(x_{k+1}).$$

If B_k is positive definite, Theorem 3.2 implies that Buys's update formulae (2.29)-(2.30) can be written as

$$\bar{\lambda}_{k+1} = \bar{\lambda}_k - \bar{\Sigma}_k \bar{c}(x_{k+1}) - (\bar{A}_{k+1}^T B_k^{-1} \bar{A}_{k+1})^{-1}\bar{c}(x_{k+1}), \tag{3.15}$$

$$\lambda_{k+1}^{(i)} = 0, \quad i \notin \mathcal{J}_k. \tag{3.16}$$

In a practical implementation of Buys's formulae, we need to enforce

$$\lambda_{k+1}^{(i)} \ge 0 \quad \text{for all } i = m_e+1, ..., m. \tag{3.17}$$

Our superlinear convergence results in the previous section suggest that (3.17) should be true when x_k is close to a local minimizer x^* and when λ_k is sufficiently close to the true Lagrange multiplier at x^*. However, in general case we are not able to prove that our new update (3.15) can ensure the non-negativity properties of the multipliers for inequality constraints. Thus, in practice, if a negative $\lambda_{k+1}^{(i)}(j \in \mathcal{I}(x_{k+1}))$ is computed by (3.15), we simply replace it by zero.

4 Discussions and Conclusion

In this paper, we re-derive a formula of Buys[14] for updating the Lagrange multipliers in the augmented Lagrangian method. The update formula is obtained by applying Newton's method for feasibility conditions. Under certain conditions, we prove that Buys's update technique leads to the Q-quadratic convergence of the multipliers.

The theoretical results in this paper are obtained under the assumption that x_{k+1} is the exact minimizer of the augmented Lagrange function. It would be interesting to analyze the case when x_{k+1} is an approximate solution of (1.5), as some practical implementations of augmented Lagrangian method do not solve (1.5) exactly. Whether there exists an efficient (fast convergent) update formula for the Lagrange multipliers without requiring the exact solution of (1.5) is an open question which is of importance both theoretically and practically.

The improvement we made on the update is achieved with the price that we need to use second order information. If we apply an unconstrained optimization algorithm that uses second order derivatives to solve problem (1.5), the new update formula can be easily obtained. However, when the unconstrained optimization algorithm used for minimizing the augmented Lagrange function does not compute any second order derivatives, additional computational cost of calculating the second order derivatives is required in order to update the multipliers by the new formula. Thus, it is interesting to investigate whether it is possible to obtain high accurate approximation of

$$(\bar{A}_{k+1}^T B_k^{-1} \bar{A}_{k+1})^{-1} \bar{c}(x_{k+1}) \tag{4.1}$$

without using any second order information. This might be possible, as (4.1) is a vector though its definition depends on the second order derivative of the Lagrange function. Now, let us consider a different but similar approximation problem. Define $\hat{\lambda}_k = \bar{\lambda}_k - \bar{\Sigma}_k \bar{c}(x_{k+1})$ and

$$\tilde{A}_k = (\nabla c_i(x_k)\,,\ i \in \mathcal{I}(x_{k+1}))\,.$$

Consider the Lagrangian function

$$\phi(x) = L_k(x, \hat{\lambda}_k) = f(x) - \sum_{i \in \mathcal{I}(x_{k+1})} \hat{\lambda}_k^{(i)} c_i(x)\,. \tag{4.2}$$

The definition of x_{k+1} implies that

$$\nabla \phi(x_{k+1}) = 0 \ \text{ and } \ \nabla^2 \phi(x_{k+1}) = B_k\,. \tag{4.3}$$

If B_k is positive definite, we have the following relation

$$B_k^{-1} \nabla \phi(x_k) = (\nabla^2 \phi(x_{k+1}))^{-1} [\nabla \phi(x_k) - \nabla \phi(x_{k+1})] \approx x_k - x_{k+1} \tag{4.4}$$

which can be written as

$$B_k^{-1} (\nabla f(x_k) - \tilde{A}_k \hat{\lambda}_k) \approx x_k - x_{k+1}\,. \tag{4.5}$$

This shows that the vector $B_k^{-1}(\nabla f(x_k) - \tilde{A}_k \hat{\lambda}_k)$, whose definition depends on second order information, can be approximated by $x_k - x_{k+1}$, which is independent of any second order derivatives. It is not clear whether we can also obtain good approximation to (4.1) without using any second order derivatives.

Another interesting problem is whether the analysis in the paper can be extended to non-smooth optimization, as many applications lead to non-smooth problems. For example, L_1 minimization appeared in image processing is a very special non-smooth optimization, and neuclear norm minimization in matrix optimization is another special non-smooth optimization.

References

[1] M.R. Hestenes, "Multiplier and gradient methods", *J. Opt. Theory Applns.*, 4(1969), 303-320.

[2] M.J.D. Powell, "A method for nonlinear constraints in minimization problems", in: R. Fletcher, ed., *Optimization* Academic Press, London, 1969, pp. 283-298.

[3] R.T. Rockafellar, "Augmented Lagrange multiplier functions and duality in non-convex programming", *SIAM J. Contrrol* 12(1974) 268-285.

[4] R. Fletcher, "An ideal penalty function for constrained optimization", *J. Inst. Maths. Applns.* 15(1975), 319-342.

[5] W.Y. Sun and Y.X. Yuan, *Optimization Theory and Methods: Nonlinear Programming* Springer, New York, 2006.

[6] A. Conn, N.Could and Ph. Toint, *LANCELOT: a Fortran Package for Large-Scale Nonlinear Optimization*, Springer, 1992.

[7] J.F. Cai, S. Osher and Z.W. Shen, "Convergence of the linearized Bregman iteration for l_1-norm minimization", *Math. Comp.*, 78(2009) 2127-2136.

[8] A.Y. Yang, S.S. Sastry, A. Ganesh and Y. Ma, "Fast l_1-minimization algorithms and an application in robust face recognition: A review" *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP)*, 2010, pp. 1849-1852.

[9] W.T. Yin, S. Osher, D. Goldfarb and J. Darbon, "Bregman Iterative Algorithms for l_1-Minimization with Applications to Compressed Sensing", *SIAM Imaging Sciences*, 1(2008) 143-168.

[10] Z.W. Wen, D. Goldfarb and W.T. Yin "Alternating direction augmented Lagrangian methods for semidefinite programming", *Math. Program. Comp.* 2(2010), 203-230.

[11] X. Zhao, D. Sun, K. Toh, "A Newton-cg augmented Lagrangian method for semi-definite programming", *SIAM J. Optim.* 20(2010) 1737-1765.

[12] Z.C. Lin, M.M. Chen and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices", *Arxiv preprint:* arXiv:1009.5055, 2010.

[13] X.M. Yuan and J.F. Yang, "Sparse and low-rank matrix decomposition via alternating direction methods", to appear in *Pacific Journal of Optimization*.

[14] J.D. Buys, "Dual algorithms for constrained optimization problems", *Ph. D. Thesis*, University of Leiden, the Netherlands, 1972.

[15] R. Fontecilla, T. Steihaug and R. Tapia, "A convergence theory for a class of quasi-Newton method for constrained optimization", *SIAM J. Numer. Anal.* 24(1987) 1133-1151.

[16] L.N. Trefethen and D. Bao, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.