Characterizations of the Augmented Lagrangian Method for Convex Variational Problems

Jongho Park[0000–0002–0496–649X]

1 Introduction

We consider the following general convex variational problem with a linear constraint:

$$\min_{v \in V} F(v) \quad \text{subject to} \quad Bv = g, \tag{1}$$

where V and W are finite-dimensional real vector spaces equipped with inner products $(\cdot, \cdot)$ and associated norms $\|\cdot\|$, $B\colon V \to W$ is a surjective linear operator, $F\colon V \to \mathbb{R}$ is a differentiable convex functional, and $g \in W$. Let $u \in V$ denote a solution to (1). Problems of the form (1) are ubiquitous in science and engineering; for examples arising from partial differential equations (PDEs), see [2, 12, 15].

A widely used approach for solving (1) is the augmented Lagrangian method [5], which is presented in Algorithm 1.

Algorithm 1 Augmented Lagrangian method for solving (1)

Given $\epsilon > 0$:
Choose $p^{(0)} \in W$.
for $n = 0, 1, 2, \dots$ **do**

$$u^{(n+1)} = \arg\min_{v \in V} \left\{ F(v) + (p^{(n)}, Bv - g) + \frac{1}{2\epsilon} \|Bv - g\|^2 \right\}, \tag{2a}$$

$$p^{(n+1)} = p^{(n)} + \epsilon^{-1} (Bu^{(n+1)} - g). \tag{2b}$$

end for

Jongho Park
Applied Mathematics and Computational Sciences Program, Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia, e-mail: jongho.park@kaust.edu.sa

In the quadratic case, that is, when the functional F is given by

$$F(v) = \frac{1}{2}(Av, v) - (f, v), \quad v \in V, \tag{3}$$

for some symmetric positive definite (SPD) linear operator $A\colon V \to V$ and $f \in V$, it was shown in [8, 13] that the convergence rate of Algorithm 1 becomes arbitrarily fast as $\epsilon \to 0$. In this sense, the constrained optimization problem (1) reduces to the unconstrained one (2a), which becomes nearly singular when $\epsilon \ll 1$. Since the systematic construction of ϵ-robust subspace correction methods for nearly singular problems was proposed in [13], there has been significant progress in solving (1) using ϵ-robust domain decomposition and multigrid methods applied to the nearly singular problem (2a) [12, 19].

More recently, in [15], the arbitrarily fast convergence of Algorithm 1 was established for the general convex optimization problem (1) under an assumption that $F^* \circ (-B^t)$ is strongly convex, where $F^*\colon V \to \overline{\mathbb{R}}$ denotes the Legendre–Fenchel conjugate of F. This result suggests a new strategy for designing domain decomposition and multigrid methods for solving (1) via the nearly semicoercive problem (2a), since ϵ-robust subspace correction methods for nearly semicoercive problems, which generalize the results of [13, 19] to the convex case, were developed in [11].

In this paper, we study the augmented Lagrangian method (Algorithm 1) in more detail. While the arbitrarily fast convergence of Algorithm 1 was established in [15] by observing its equivalence with the proximal point method [18] applied to a dual problem [6] of (1), we show here that Algorithm 1 is also equivalent to the gradient descent method applied to a dual problem of an augmented formulation of (1). This provides an alternative analysis for its arbitrarily fast convergence rate. For completeness, we also review the relevant existing results in detail.

The rest of this paper is organized as follows. In Section 2, we present the quadratic case considered in [14]. In Section 3, we address the general convex case by studying its equivalence with the proximal point method, discussed in [15]. In Section 4, we provide an alternative analysis based on equivalence with the gradient descent method. In Section 5, we discuss the numerical solution of the primal subproblems of the augmented Lagrangian method and the connection to domain decomposition and multigrid methods.

2 Quadratic case

We first consider the special case of quadratic optimization, i.e., the functional F in (1) is given by (3). It is straightforward to observe that (1) is equivalent to the following linear saddle-point problem:

$$\begin{bmatrix} A & B^t \\ B & 0 \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}, \tag{4}$$

where $B^t: W \to V$ denotes the adjoint of B. We define the Schur complement $S: W \to W$ as

$$S = BA^{-1}B^t. \tag{5}$$

The augmented Lagrangian method given in Algorithm 1 reduces to the following form when applied to (4):

$$\begin{aligned} u^{(n+1)} &= (A + \epsilon^{-1}B^tB)^{-1}(f + \epsilon^{-1}B^tg - B^tp^{(n)}), \\ p^{(n+1)} &= p^{(n)} - \epsilon^{-1}(g - Bu^{(n+1)}), \end{aligned} \qquad n \geq 0.$$

This iteration is equivalent to the Richardson method with step size ϵ^{-1} applied to the linear system (see, e.g., [3])

$$S_\epsilon p = d_\epsilon, \quad \text{where} \quad \begin{aligned} S_\epsilon &= B(A + \epsilon^{-1}B^tB)^{-1}B^t, \\ d_\epsilon &= B(A + \epsilon^{-1}B^tB)^{-1}(f + \epsilon^{-1}B^tg) - g. \end{aligned} \tag{6}$$

Therefore, it suffices to estimate the error propagation operator $I - \epsilon^{-1}S_\epsilon$ for the Richardson iteration associated with (6) [16].

In what follows, we summarize the argument presented in [14]. For any $q \in W$,

$$\begin{aligned} (S_\epsilon^{-1}q, q) &= \inf_{v \in V,\ Bv = q} ((A + \epsilon^{-1}B^tB)v, v) \\ &= \inf_{v \in V,\ Bv = q} (Av, v) + \epsilon^{-1}(q, q) = (S^{-1}q, q) + \epsilon^{-1}(q, q), \end{aligned} \tag{7}$$

where the first and last equalities follow from the auxiliary space lemma [16]. Hence,

$$S_\epsilon^{-1} = S^{-1} + \epsilon^{-1}I.$$

We note that a similar result appeared in [9]. Consequently, we obtain the following sharp estimate for the error propagation operator $I - \epsilon^{-1}S_\epsilon$:

$$\|I - \epsilon^{-1}S_\epsilon\| = \lambda_{\max}(I - \epsilon^{-1}S_\epsilon) = 1 - \epsilon^{-1}\lambda_{\min}(S_\epsilon) = \frac{\epsilon}{\lambda_{\min}(S) + \epsilon}.$$

Therefore, the convergence rate of the augmented Lagrangian method becomes arbitrarily fast as ϵ tends to zero.

3 Characterization as a proximal point method

In [15], the arbitrarily fast convergence of the augmented Lagrangian method for solving (1), which generalizes the result in Section 2, was established by observing its equivalence with the proximal point method [18]. While this result follows directly from the dualization framework introduced in [6], we include the details here for completeness.

We begin by recalling the key features of Fenchel–Rockafellar duality [17], which establishes a duality relation between two convex optimization problems. Consider the minimization problem

$$\min_{v\in V} \{F(v) + G(Bv)\}, \tag{8}$$

where $B\colon V \to W$ is a linear operator and $F\colon V \to \overline{\mathbb{R}}$ and $G\colon W \to \overline{\mathbb{R}}$ are convex functionals. We also consider the following minimization problem:

$$\min_{q\in W} \left\{F^*(-B^t q) + G^*(q)\right\}. \tag{9}$$

We refer to (8) as the *primal problem* and to (9) as the *dual problem*. Under mild assumptions (see, e.g., [6]), for instance, if

$$H(q) := \inf_{v\in V}\{F(v) + G(Bv + q)\}$$

is subdifferentiable at $q = 0$ (see also [6, Remark 2.6]), then, if $u \in V$ solves (8) and $p \in W$ solves (9), the following primal–dual relations hold:

$$-B^t p \in \partial F(u), \qquad Bu \in \partial G^*(p). \tag{10}$$

One may refer to [17, Corollary 31.2.1] for a detailed derivation of (10).

Now, we regard the constrained problem (1) as the primal problem (8). More precisely, we rewrite (1) in the equivalent form

$$\min_{v\in V}\{F(v) + \chi_{\{g\}}(Bv)\},$$

where $\chi_{\{g\}}\colon W \to \overline{\mathbb{R}}$ is the indicator function that takes the value 0 when its argument equals g and ∞ otherwise. Thus, the problem fits into the framework of (8) with $G = \chi_{\{g\}}$. The corresponding dual problem (9) then becomes

$$\min_{q\in W} \left\{F^*(-B^t q) + (g, q)\right\}. \tag{11}$$

The proximal point method [18] for solving (11) is described in Algorithm 2.

Algorithm 2 Proximal point method for solving (11)

Given $\epsilon > 0$:
Choose $p^{(0)} \in W$.
for $n = 0, 1, 2, \dots$ **do**

$$p^{(n+1)} = \operatorname*{arg\,min}_{q\in W} \left\{F^*(-B^t q) + (g, q) + \frac{\epsilon}{2}\|p - p^{(n)}\|^2\right\} \tag{12}$$

end for

If we interpret the subproblem (12) in Algorithm 2 as the dual problem (9), then the corresponding primal problem is given by

$$\min_{v \in V} \left\{ F(v) + (Bv - q, p^{(n)}) + \frac{1}{2\epsilon} \|Bv - q\|^2 \right\},$$

which coincides with (2a). If we denote the solution of this problem by $u^{(n+1)}$, then the primal–dual relation (10) yields (2b). This establishes the equivalence between the augmented Lagrangian method and the proximal point method, as summarized in Lemma 1.

Lemma 1 *The augmented Lagrangian method presented in Algorithm 1 is equivalent to the proximal point method presented in Algorithm 2 in the sense that they generate the same sequence* $\{p^{(n)}\}$.

The convergence analysis of the proximal point method is well known [1, Example 23.40]. Combining this result with Lemma 1, we obtain the following convergence theorem for the augmented Lagrangian method. We note that the assumption on the strong convexity of $F^* \circ (-B^t)$ is satisfied, for example, when F is smooth and B is surjective; see [4, Section 3.3] for the relationship between the smoothness of F and the strong convexity of F^*.

Theorem 1 *In* (1)*, suppose that* $F^* \circ (-B^t)$ *is* μ*-strongly convex. Then, for the augmented Lagrangian method presented in Algorithm 1, we have*

$$\|p^{(n+1)} - p\| \leq \frac{\epsilon}{\mu + \epsilon} \|p^{(n)} - p\|, \quad n \geq 0.$$

Remark 1 In the special case (4), we have $\mu = \lambda_{\min}(S)$, where S was defined in (5).

The convergence rate for the primal sequence $\{u^{(n)}\}$ can also be derived; see [15, Theorem A.4] for details.

4 Characterization as a gradient descent method

Although the convergence analyses for the quadratic and general convex cases presented in Sections 2 and 3 yield the same convergence rate $\frac{\epsilon}{\mu+\epsilon}$, their derivations differ: in the quadratic case, we used the equivalence of the augmented Lagrangian method with the Richardson iteration applied to (6), while in the general convex case, we exploited its equivalence with the proximal point method for solving the dual problem (11). In this section, we present an alternative convergence analysis for the general convex case, which directly generalizes the analysis of the quadratic case in Section 2.

Observe that the problem (1) is equivalent to the following augmented problem:

$$\min_{v \in V} F_\epsilon(v) \quad \text{subject to} \quad Bv = g, \tag{13}$$

where $\epsilon > 0$ and

$$F_\epsilon(v) = F(v) + \frac{1}{2\epsilon}\|Bv - g\|^2, \quad v \in V.$$

By invoking Fenchel–Rockafellar duality, the dual problem associated with (13) is given by

$$\min_{q \in W} \left\{ E_\epsilon(q) := F_\epsilon^*(-B^t q) + (g, q) \right\}. \tag{14}$$

By the inverse property of subdifferentials given in [6, Equation (2.7)], for any $w, z \in V$, we have

$$\nabla F_\epsilon^*(w) = z \quad \Leftrightarrow \quad \nabla F_\epsilon(z) = w \quad \Leftrightarrow \quad z = \arg\min_{v \in V} \left\{ F_\epsilon(v) - (w, v) \right\}.$$

A related result can also be found in [10, Lemma 3.3]. This shows that Algorithm 1 is equivalent to the gradient descent method with step size ϵ^{-1}, presented in Algorithm 3, applied to (14). We summarize this equivalence in Lemma 2.

Algorithm 3 Gradient descent method for solving (14)

Given $\epsilon > 0$:
Choose $p^{(0)} \in W$.
for $n = 0, 1, 2, \dots$ **do**

$$p^{(n+1)} = p^{(n)} - \epsilon^{-1} \nabla E_\epsilon(p^{(n)})$$

end for

Lemma 2 *The augmented Lagrangian method presented in Algorithm 1 is equivalent to the gradient descent method presented in Algorithm 3 in the sense that they generate the same sequence* $\{p^{(n)}\}$.

From the definition of the Legendre–Fenchel conjugate, we can derive the following identity, which generalizes the auxiliary space lemma introduced in [16] to the convex setting (see also [1]):

$$\tilde{F}^*(q) = F^*(-B^t q), \quad \text{where} \quad \tilde{F}(q) = \inf_{Bv = -q} F(v). \tag{15}$$

Hence, similarly to (7), we obtain, for any $q \in W$,

$$\begin{aligned} (F_\epsilon^* \circ (-B^t))^*(q) &= \inf_{Bv=-q} F_\epsilon(v) \\ &= \inf_{Bv=-q} F(v) + \frac{1}{2\epsilon}\|q + g\|^2 = (F^* \circ (-B^t))^*(q) + \frac{1}{2\epsilon}\|q + g\|^2, \end{aligned} \tag{16}$$

where the first and last equalities follow from (15). Using the well-known relationship between smoothness and strong convexity, we deduce that E_ϵ, defined in (14), is ϵ-smooth, which ensures the convergence of Algorithm 3 [4]. Furthermore, (16) implies the following equivalence:

$$
\begin{aligned}
F^* \circ (-B^t) \text{ is } \mu\text{-strongly convex} &\Leftrightarrow (F^* \circ (-B^t))^* \text{ is } \mu^{-1}\text{-smooth,} \\
&\Leftrightarrow (F^*_\epsilon \circ (-B^t))^* \text{ is } (\mu^{-1} + \epsilon^{-1})\text{-smooth,} \\
&\Leftrightarrow E_\epsilon \text{ is } (\mu^{-1} + \epsilon^{-1})^{-1}\text{-strongly convex.}
\end{aligned}
$$

Having established the strong convexity and smoothness properties of E_ϵ, the convergence rate of Algorithm 3 follows from the standard convergence theory of the gradient descent method (see, e.g., [4]), leading to a result analogous to Theorem 1.

Remark 2 The discussion in this section also indicates an equivalence between Algorithms 2 and 3. More generally, one may consider the connection between the gradient descent method and the proximal point method. Such an equivalence indeed holds, as discussed in [7, Section 7].

5 Nearly semicoercive primal subproblems

The discussions so far indicate that the augmented Lagrangian method achieves arbitrarily fast convergence as the parameter ϵ becomes small. Consequently, in practical implementations, the efficient solution of the primal subproblem (2a) becomes the most critical component of the overall algorithm. Since the operator B generally has a nontrivial kernel, the subproblem (2a) becomes nearly semicoercive [11] as $\epsilon \to 0$. In the special quadratic case, the problem becomes nearly singular [13], causing the condition number to blow up, i.e., grow rapidly as ϵ tends to zero, and leading standard iterative solvers to perform poorly. Therefore, ϵ-robust numerical strategies are essential for maintaining computational efficiency.

In [13], it was shown that subspace correction methods [20] based on a stable space decomposition

$$V = \sum_{j=1}^{J} V_j,$$

where each V_j is a subspace of V, converge robustly with respect to ϵ provided that the decomposition satisfies

$$\mathcal{N} = \sum_{j=1}^{J} (V_j \cap \mathcal{N}), \quad \text{where} \quad \mathcal{N} = \ker B. \tag{17}$$

In particular, domain decomposition and multigrid methods designed to respect the kernel decomposition (17) exhibit uniform convergence independent of ϵ [12, 19].

More recently, the theory of ϵ-robust subspace correction methods was extended to general convex variational problems in [11]. This generalization holds under additional but mild assumptions on the solution spaces and problems, which are typically satisfied for a broad class of PDE models. As an illustrative example, [15] presents an application to the Darcy–Forchheimer flow, a nonlinear system of PDEs describing fluids in porous media.

In summary, the characterization of arbitrarily fast convergence of the augmented Lagrangian method for convex variational problems presented in this paper, combined with the development of ϵ-robust subspace correction methods in [11], offers a new theoretical and computational perspective for designing efficient domain decomposition and multigrid solvers for large-scale nonlinear variational problems. Future research directions include the application of these ideas to a wider range of nonlinear PDEs and variational formulations arising in complex physical systems.

References

1. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)
2. Belenki, L., Berselli, L.C., Diening, L., Růžička, M.: On the finite element approximation of p-Stokes systems. SIAM J. Numer. Anal. **50**(2), 373–397 (2012)
3. Benzi, M., Golub, G.H., Liesen, J.: Numerical solution of saddle point problems. Acta Numer. **14**, 1–137 (2005)
4. Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. Acta Numer. **25**, 161–319 (2016)
5. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**(5), 303–320 (1969)
6. Jiang, B., Park, J., Xu, J.: Connections between convex optimization algorithms and subspace correction methods. arXiv preprint arXiv:2505.09765 (2025)
7. Kim, D.: Accelerated proximal point method for maximally monotone operators. Math. Program. **190**(1), 57–87 (2021)
8. Lee, C.O., Park, E.H.: A dual iterative substructuring method with a penalty term. Numer. Math. **112**(1), 89–113 (2009)
9. Lee, C.O., Park, E.H.: A dual iterative substructuring method with a small penalty parameter. J. Korean Math. Soc. **54**(2), 461–477 (2017)
10. Lee, C.O., Park, E.H., Park, J.: A finite element approach for the dual Rudin–Osher–Fatemi model and its nonoverlapping domain decomposition methods. SIAM J. Sci. Comput. **41**(2), B205–B228 (2019)
11. Lee, Y.J., Park, J.: Parallel subspace correction methods for semicoercive and nearly semicoercive convex optimization with applications to nonlinear PDEs. Math. Comp. To appear (2026)
12. Lee, Y.J., Wu, J., Chen, J.: Robust multigrid method for the planar linear elasticity problems. Numer. Math. **113**, 473–496 (2009)
13. Lee, Y.J., Wu, J., Xu, J., Zikatanov, L.: Robust subspace correction methods for nearly singular systems. Math. Models Methods Appl. Sci. **17**(11), 1937–1963 (2007)
14. Park, J.: Unified analysis of saddle point problems via auxiliary space theory. arXiv preprint arXiv:2509.11434 (2025)
15. Park, J., Hassanizadeh, S.M.: Parallel multilevel methods for solving the Darcy–Forchheimer model based on a nearly semicoercive formulation. arXiv preprint arXiv:2507.03192 (2025)
16. Park, J., Xu, J.: Auxiliary space theory for the analysis of iterative methods for semidefinite linear systems. arXiv preprint arXiv:2509.07179 (2025)
17. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, NJ (1970)
18. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**(5), 877–898 (1976)
19. Wu, J., Zheng, H.: Parallel subspace correction methods for nearly singular systems. J. Comput. Appl. Math. **271**, 180–194 (2014)
20. Xu, J.: Iterative methods by space decomposition and subspace correction. SIAM Rev. **34**(4), 581–613 (1992)