



CONVERGENCE ANALYSIS OF TWO METHODS OF LAGRANGE MULTIPLIERS USING ROCKAFELLAR'S THEORY ON MONOTONE OPERATORS AND PROXIMAL-POINT ALGORITHMS

IGOR GRIVA

Department of Mathematical Sciences, George Mason University, Fairfax, VA, USA

Dedicated to Professor Terry Rockafellar and his important contribution to the convergence analysis of Lagrange multiplier methods

Abstract. In the mid-1970s, in his seminal article "Monotone Operators and the Proximal Point Algorithm," Rockafellar developed a convergence analysis for a generalized proximal-point algorithm, which is related to maximal monotone operators. That article thus built a general framework for convergence analysis for methods of Lagrange multipliers, since they can be viewed as proximal-point algorithms. This paper is concerned with two well-known methods of Lagrange multipliers. One method for solving optimization problems with equality constraints is an augmented Lagrangian (AL) method introduced by Hestenes and Powell in the late 1960s. Another Lagrange multiplier method for optimization problems with inequality constraints is based on nonlinear rescaling of the functions that define the constraints with properly selected transformations. This approach was introduced by Polyak in the 1980s. Rescaling the constraints, however, leads to a dynamic rescaling of the corresponding monotone operators, which, in turn, leads to the need for an extension of Rockafellar's methodology to include such a case. This manuscript reviews the relation among the two Lagrange multiplier methods, proximal-point algorithms, and monotone operators. It also outlines the basic steps of convergence analysis of the augmented Lagrangian and nonlinear rescaling methods as proximal-point methods, discussing similarities and differences between them. Thus the manuscript pays tribute to Rockafellar's methodology that made such an analysis possible. It also provides some numerical results demonstrating the value of proximal-point methods.

Keywords. Augmented Lagrangian; Proximal point; Monotone operators; Nonlinear rescaling.

2020 Mathematics Subject Classification. 90C30, 90C25.

1. INTRODUCTION

An augmented Lagrangian method (ALM) with a quadratic penalty term for solving constrained optimization problems with equality constraints was introduced by Hestenes and Powell in 1969 [8, 14]. The ALM consists of minimizing the augmented Lagrangian function followed by an update of the Lagrange multipliers.

E-mail address: igriva@gmu.edu (I. Griva).

Received January 15, 2026; Accepted March 10, 2026.

The ALM was an alternative to the penalty-type methods introduced a decade earlier by Courant [3]. The penalty parameter must increase unboundedly to guarantee convergence of penalty-type algorithms. At the same time, a large value of the penalty parameter results in their inability to find accurate solutions due to the unavoidable ill-conditioning effect.

Therefore, the significance of the introduction of the ALM is hard to overstate. The ALM converged with a fixed penalty parameter and thus avoided the ill-conditioning issues. The ALM decreased the total number of iterations required to reach the same level of solution accuracy by orders of magnitude. Also, the ALM can solve problems with much better accuracy. Section 2 below provides a few examples to illustrate the difference in performance between the penalty and augmented Lagrangian methods.

There are two major views on the convergence analysis of the ALM: local and global. The local convergence analysis is usually done under the standard second-order optimality conditions, which results in the existence of a primal-dual contractive map implicitly implemented at each step of the ALM in a neighborhood of the primal-dual solution. This approach also allows one to establish bounds on the linear rate of convergence in that neighborhood.

The global convergence view requires convexity of the problem and can be analyzed with the help of monotone operators introduced and analyzed by Minty [10]. In the mid-1970s, R. T. Rockafellar showed how the convergence of the augmented Lagrangian method can be established via a generalized proximal-point algorithm, which, in turn, is based on monotone operators [15, 16]. He developed a general methodology that allows one to analyze the convergence of Lagrange multiplier methods using duality, proximal-point methods, and monotone operators. One of the goals of this paper is to pay tribute to the legacy of Rockafellar's work and demonstrate how his theory can be used to analyze the convergence of two Lagrange multiplier methods: one for solving an optimization problem with equality constraints, and one for solving a problem with inequality constraints.

Historically, optimization problems with inequality constraints had to wait to be analyzed until after their equality-constrained counterparts. Perhaps the largest known such time gap happened between Lagrange's letters to Euler in 1755 (or his published work of 1788), where he introduced the first-order optimality conditions for an optimization problem with equality constraints, and the Karush-Kuhn-Tucker first-order optimality conditions for an optimization problem with inequality constraints developed in 1951 [9]. The introduction of penalty functions for equality constraints by Courant in 1943 preceded the introduction of barrier functions for inequality constraints by Frisch in 1950 [5]. The latter were studied extensively by Fiacco and McCormick in the 1960s [4].

As mentioned above, the augmented Lagrangian methods introduced in 1969 offered an innovative approach to dealing with ill-conditioning when solving problems with equality constraints. However, their inequality-constrained counterpart had to wait again until the mid-1970s, when Rockafellar developed a methodology for analyzing the convergence of augmented Lagrangian methods using proximal-point algorithms and monotone operators. The developed methodology allowed analyzing the convergence of the augmented Lagrangian method by showing its equivalence to the proximal-point method in the dual space with a quadratic proximal term.

However, it was immediately recognized that handling inequality constraints required further consideration and overcoming essential challenges. The first known recipe for handling

problems with inequality constraints was offered by Rockafellar in [16], where he modified the augmented Lagrangian to handle inequality constraints. Such a modification of the augmented Lagrangian, however, did not preserve the smoothness of the original functions involved in the problem statement, so methods that require higher-order smoothness, such as Newton's method, could not be applied.

Help came in the 1980s with the development of nonlinear rescaling (NR) theory and methods introduced by R. Polyak [11, 12]. In NR methods, the functions describing inequality constraints are rescaled with appropriately selected transformations. The rescaled problem is equivalent to the original one. Then the NR method consists of sequential minimization of the Lagrangian for the rescaled problem followed by an update of the Lagrange multipliers. The transformations may be selected smooth enough to work with a wide range of techniques of unconstrained minimization. Particular realizations of the NR method include the modified barrier function method introduced by R. Polyak [11] and the exponential multiplier method introduced by D. Bertsekas [2] in the 1980s.

NR methods do not suffer from the ill-conditioning effect, as they can keep the value of the scaling parameter bounded and preserve the original smoothness of the problem. Therefore, NR methods are appropriate for solving inequality-constrained problems, similarly to how augmented Lagrangian methods are appropriate for solving problems with equality constraints.

The numerical performance of both augmented Lagrangian and nonlinear rescaling methods is similar. Examples in the following section illustrate that observation. The theoretical analysis of NR methods can be done with an extension of the original methodology developed by Rockafellar for proximal-point methods and monotone operators. The monotone operators related to NR methods are rescaled by factors that can approach zero as the trajectory of NR iterates approaches the solution.

This manuscript outlines the basic steps of the convergence analysis of both the augmented Lagrangian and NR methods using the theory of proximal-point methods and monotone operators developed by Rockafellar, and discusses the key similarities and differences in the convergence analysis of the methods. In particular, the convergence analysis of the NR method is performed with the help of rescaled monotone operators following [6]. In that work, the authors used the theory developed by Rockafellar to guide the steps of the proofs, but modified the steps to apply them to the rescaled case.

The rest of the manuscript is organized as follows. The following section shows a few numerical examples that illustrate some points described in the introduction. Section 3 describes both augmented Lagrangian and nonlinear rescaling methods and their connection to monotone operators. It also outlines the basic steps of the convergence analysis for both the augmented Lagrangian and nonlinear rescaling methods, overcoming some challenges of using rescaled monotone operators in the NR analysis. Section 4 provides some concluding remarks.

2. A FEW NUMERICAL EXAMPLES

First, let us demonstrate the power of the Lagrange multipliers methods using a problem with equality constraints. We consider the following optimization problem:

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} && f(x) \\ & \text{subject to} && g(x) = 0, \end{aligned} \tag{2.1}$$

Maj. it.	k	fval	$\ \nabla P_k\ $	infeas	Newton steps
0	1.0e+03	2.00e+07	2.00e+05	2.00e+02	0
1	1.0e+03	3.85e+03	7.36e-11	3.85e-02	1
2	1.0e+04	3.86e+03	8.10e-10	3.86e-03	1
3	1.0e+05	3.86e+03	2.39e-09	3.86e-04	1
4	1.0e+06	3.86e+03	1.84e-08	3.86e-05	1
5	1.0e+07	3.86e+03	1.42e-07	3.86e-06	1

TABLE 1. Solving problem (2.2) with a penalty method with different k ($n = 100$)

where $f : \mathbb{R}^n \rightarrow \mathbb{R}^1$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are smooth enough functions. We assume that f is convex, vector-function g is linear, and the solution of (2.1) exists.

The penalty method introduced in 1940s by Courant treats constraints with the penalty term:

$$P_k(x) = f(x) + \frac{k}{2}g(x)^T g(x),$$

where $k > 0$ is the penalty parameter, which needs to increase to ∞ to find the solution of problem (2.1) via minimization of P_k . When k is large, the Hessian of P_k becomes ill-conditioned, making it challenging to find a minimizer of P_k accurately.

To demonstrate this challenge, let us consider the following optimization problem that fits (2.1).

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}x^T D x \\ & \text{subject to} \quad \sum_{i=1}^n x_i = 2n, \end{aligned} \tag{2.2}$$

where $D = \text{diag}(d)$ is a diagonal matrix with $d = (1, \dots, n)$ on the main diagonal. Note that the Hessian D of the objective function also becomes ill-conditioned for large dimensions n , making it difficult to solve the problem.

Since the level sets of P_k become “stretched” for large n and k , the trajectories of first-order gradient-based methods would experience a zigzagging pattern, making first-order methods inefficient. Therefore, second-order methods, which rescale the level sets with the Hessian would be a better choice. The penalty P_k for (2.2) is a strongly convex quadratic function. Therefore, in theory, for example, Newton’s method must find the solution to the minimization of P_k in one iteration for any k and n . However, in practice, this is not the case. Since the Hessian of P_k is ill-conditioned for large k and n , the Newton direction cannot be accurately found, and Newton’s method cannot find accurate solutions with the desired accuracy. The following experiment demonstrates that effect.

We will start with $k = 1000$ and increase k ten times until we cannot find a minimum of P_k any longer. We consider problem (2.2) with $n = 100$ and $n = 1000$. The results of the experiment are described in Tables 1 and 2. Each line of the tables corresponds to the minimization of P_k with a fixed value of k , which is given in the second column. The tables also show the objective function value, the norm of the gradient $\|\nabla P_k\|$ that the algorithm was able to obtain, the value

of the constraint violation, and the number of Newton steps per unconstrained minimization of P_k , which is 1, because the minimized function is quadratic.

As one can see from the tables, the problem of size $n = 100$ can be solved up to the accuracy level of $\approx 10^{-6}$, while the problem of size $n = 1000$ can be solved up to the accuracy level of $\approx 10^{-5}$ because the latter problem is more ill-conditioned and thus more difficult.

Also, one can see from Table 2, for example, why the computations stall at that particular point of the optimization trajectory. As k increases, the infeasibility violation decreases, while the ability of Newton's method to minimize P_k accurately gets worse: the larger k , the larger the final $\|\nabla P_k\|$ that the method can attain. To reduce the infeasibility violation to the level of $\approx 10^{-6}$, we need to find the minimum of P_k to at least that level of accuracy, which the penalty method fails to do.

Another observation following from Tables 1 and 2 is that if the algorithm can keep the value of the penalty parameter $k \leq 1000$ for those values of n , the minimization of a quadratic function can be done accurately. Therefore, when Hestenes [8] and Powell [14] introduced the augmented Lagrangian method in the late 1960s, it was a significant milestone, because the ALM converged under a smaller and fixed penalty parameter k .

The Lagrangian for (2.1) is given by $L(x, \lambda) = f(x) - \lambda^T g(x)$, and the augmented Lagrangian is as follows:

$$\mathcal{L}_k(x, \lambda) = f(x) - \lambda^T g(x) + \frac{k}{2} g(x)^T g(x), \quad (2.3)$$

where $\lambda \in \mathbb{R}^m$ is a vector of Lagrange multipliers that corresponds to the equality constraints and $k > 0$ is the scaling parameter.

The augmented Lagrangian method consists of a sequence of inexact minimizations of $\mathcal{L}_k(x, \lambda)$ in x on $\mathbb{R}^n$

$$\hat{x} \approx \hat{x}(\lambda) = \operatorname{argmin}_{x \in \mathbb{R}^n} \mathcal{L}_k(x, \lambda). \quad (2.4)$$

followed by updating the Lagrange multipliers: $\hat{\lambda} = \lambda - kg(\hat{x})$.

To demonstrate the advantage of the ALM over the penalty type method, we employ the ALM to solve problem (2.2) for $n = 100$ and $n = 1000$. Tables 3 and 4 show the results.

The scaling parameter $k = 60$ is kept fixed so the unconstrained minimization (2.4) is found up to the accuracy of 10^{-9} . Therefore the solution can be found to that level of accuracy in just 5 iterations for both $n = 100$ and $n = 1000$.

Maj. it.	k	fval	$\ \nabla P_k\ $	infeas	Newton steps
0	1.0e+03	2.00e+09	2.00e+06	2.00e+03	0
1	1.0e+03	2.67e+05	8.38e-10	2.67e-01	1
2	1.0e+04	2.67e+05	1.49e-08	2.67e-02	1
3	1.0e+05	2.67e+05	5.09e-09	2.67e-03	1
4	1.0e+06	2.67e+05	8.79e-07	2.67e-04	1
5	1.0e+07	2.67e+05	1.11e-05	2.67e-05	1

TABLE 2. Solving problem (2.2) with a penalty method with different k ($n = 1000$)

Maj. it.	k	fval	$\ \nabla \mathcal{L}_k\ $	infeas	Newton steps
0	6.0e+01	1.20e+06	1.20e+04	2.00e+02	0
1	6.0e+01	3.84e+03	7.57e-12	6.41e-01	1
2	6.0e+01	3.86e+03	2.37e-12	2.05e-03	1
3	6.0e+01	3.86e+03	1.49e-12	6.57e-06	1
4	6.0e+01	3.86e+03	3.01e-12	2.10e-08	1
5	6.0e+01	3.86e+03	2.03e-12	6.73e-11	1

TABLE 3. Solving problem (2.2) with a augmented Lagrangian method with $k = 60$ ($n = 100$)

Maj. it.	k	fval	$\ \nabla \mathcal{L}_k\ $	infeas	Newton steps
0	6.0e+01	1.20e+08	1.20e+05	2.00e+03	0
1	6.0e+01	2.67e+05	2.25e-10	4.44e+00	1
2	6.0e+01	2.67e+05	3.42e-11	9.87e-03	1
3	6.0e+01	2.67e+05	1.50e-10	2.19e-05	1
4	6.0e+01	2.67e+05	1.19e-11	4.87e-08	1
5	6.0e+01	2.67e+05	1.21e-10	1.06e-10	1

TABLE 4. Solving problem (2.2) with a augmented Lagrangian method with $k = 60$ ($n = 1000$)

Now, let us consider the optimization problem with inequality constraints:

$$\begin{aligned}
 & \underset{x \in \mathbb{R}^n}{\text{minimize}} && f(x) \\
 & \text{subject to} && c(x) \geq 0,
 \end{aligned} \tag{2.5}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}^1$ and $c = (c_1, \dots, c_m) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are smooth enough functions. We assume that f is convex, c_i , $i = 1, \dots, m$ are concave, and the solution of (2.5) exists.

We also assume that

- A.** The optimal primal set $X^* = \{x \in \Omega : f(x) = f(x^*)\}$ is not empty and bounded.
- B.** The Slater's condition holds, i.e. there exists $\hat{x} \in \mathbb{R}^n : c_i(\hat{x}) > 0$, $i = 1, \dots, m$.

The algorithms for solving optimization problems with inequality constraints were developed in a similar way. First, Frisch [5] introduced a barrier function in the 1950s, which was popularized by Fiacco and McCormick in the 1960s [4]. The most famous example is the log-barrier function

$$B_k(x) = f(x) - \frac{1}{k} \sum_{i=1}^m \ln c_i(x), k > 0.$$

The solution to (2.5) can be found by finding an unconstrained minimum of B_k for $k > 0$ large enough. In practice, in the neighborhood of the primal-dual solution, $c_i(x) \approx 0$ for the active

Maj. it.	k	fval	$\ \nabla B_k\ $	infeas	comp	NS
0	1.00e+00	2.27e+04	2.99e+02	0.00e+00	1.00e+00	0
1	2.00e+00	7.25e+03	1.24e+02	0.00e+00	5.00e-01	5
2	4.00e+00	4.85e+03	5.66e+01	0.00e+00	2.50e-01	3
3	8.00e+00	4.53e+03	2.38e+01	0.00e+00	1.25e-01	4
4	1.60e+01	4.46e+03	1.10e+01	0.00e+00	6.25e-02	2
5	3.20e+01	4.44e+03	2.27e+00	0.00e+00	3.12e-02	2
6	6.40e+01	4.42e+03	2.39e-01	0.00e+00	1.56e-02	7
7	1.28e+02	4.42e+03	5.26e-04	0.00e+00	7.81e-03	12
8	2.56e+02	4.41e+03	8.53e-04	0.00e+00	3.91e-03	3
9	5.12e+02	4.41e+03	7.00e-04	0.00e+00	1.95e-03	3
10	1.02e+03	4.41e+03	5.76e-04	0.00e+00	9.77e-04	3
11	2.05e+03	4.41e+03	3.16e-04	0.00e+00	4.88e-04	3
12	4.10e+03	4.41e+03	1.00e-04	0.00e+00	2.44e-04	3
13	8.19e+03	4.41e+03	1.79e-05	0.00e+00	1.22e-04	3
14	1.64e+04	4.41e+03	2.06e-06	0.00e+00	6.10e-05	3

TABLE 5. Solving problem (2.6) with a barrier function method, $n = 100$. The barrier function method failed to improve the accuracy beyond 10^{-5} .

constraints, and minimization of B_k cannot be performed accurately, as the numerical examples below demonstrate. This approach is conceptually similar to the one using the penalty function P_k for equality constraints. The parameter k must be large to find the solution accurately; however, finding an unconstrained minimum for large k is problematic due to the ill-conditioning of the Hessian of B_k in the neighborhood of the solution of (2.5). Therefore, an analog of the augmented Lagrangian method was needed to avoid the necessity of pushing k to large values. A few attempts were made to adjust the quadratic augmented Lagrangian (2.3) for problems with inequality constraints (see [16]); however, the resulting augmented Lagrangians lacked the higher-order smoothness needed to use Newton's method. The remedy came with the introduction of a general nonlinear rescaling theory and methods in the 1980s [11, 12]. The NR methodology suggested a general class of smooth transformations ψ used to rescale the constraint function c_i in such a way that the rescaled problem is equivalent to (2.5), and the corresponding Lagrangian is smooth. The nonlinear rescaling method also allowed the scaling parameter to be kept bounded, which, in turn, allowed an unconstrained minimizer to be found accurately.

Maj. it.	k	fval	$\ \nabla B_k\ $	infeas	comp	NS
0	1.00e+00	2.25e+06	3.00e+03	0.00e+00	1.00e+00	0
1	2.00e+00	8.12e+05	1.07e+03	0.00e+00	5.00e-01	7
2	4.00e+00	3.66e+05	3.27e+02	0.00e+00	2.50e-01	13
3	8.00e+00	3.64e+05	1.03e+02	0.00e+00	1.25e-01	1
4	1.60e+01	3.64e+05	2.20e+01	0.00e+00	6.25e-02	2
5	3.20e+01	3.63e+05	1.53e+00	0.00e+00	3.12e-02	13
6	6.40e+01	3.63e+05	2.27e-01	0.00e+00	1.56e-02	2
7	1.28e+02	3.63e+05	3.64e-03	0.00e+00	7.81e-03	3
8	2.56e+02	3.63e+05	2.65e-03	0.00e+00	3.91e-03	3
9	5.12e+02	3.63e+05	1.94e-03	0.00e+00	1.95e-03	3
10	1.02e+03	3.63e+05	7.24e-06	0.00e+00	9.77e-04	4
11	2.05e+03	3.63e+05	2.01e-05	0.00e+00	4.88e-04	4
12	4.10e+03	3.63e+05	3.73e-05	0.00e+00	2.44e-04	4
13	8.19e+03	3.63e+05	2.35e-05	0.00e+00	1.22e-04	4
14	1.64e+04	3.63e+05	3.35e-05	0.00e+00	6.10e-05	4

TABLE 6. Solving problem (2.6) with a barrier function method, $n = 1000$. The barrier function method failed to improve the accuracy beyond 10^{-5} .

To demonstrate this effect, we modify (2.2) as follows:

$$\begin{aligned}
& \underset{x \in \mathbb{R}^n}{\text{minimize}} && \frac{1}{2} x^T D x \\
& \text{subject to} && \sum_{i=1}^n x_i \geq 2n, \\
& && x_i \geq 1, i = 1, \dots, n
\end{aligned} \tag{2.6}$$

where $D = \text{diag}(d)$, $d = (1, \dots, n)$.

The first constraint $\sum_{i=1}^n x_i \geq 2n$ and some of the bounds $x_i \geq 1$ are active, i.e.

$$\sum_{i=1}^n x_i^* = 2n,$$

$$x_i^* = 1, i \in A \subset \{1, \dots, n\},$$

where A is the set of active constraints. The rest bounds are passive, i.e.

$$x_i^* > 1, i \in P = \{1, \dots, n\} \setminus A.$$

Therefore both active and passive sets are not empty.

From our experimentation with solving (2.6) by minimizing B_k , we found that the best strategy is to start with a small k and increase it by a factor of 1.5 after each minimization. Tables 5

Maj. it.	k	fval	$\ \nabla \mathcal{L}_k\ $	infeas	comp	NS
0	9.00e+02	1.01e+04	1.99e+02	-0.00e+00	1.00e+00	0
1	9.00e+02	5.42e+03	5.49e+01	2.09e-02	1.57e+00	1
2	9.00e+02	3.94e+03	1.75e+01	5.16e-01	5.50e+00	2
3	9.00e+02	4.40e+03	1.32e+00	1.18e-01	1.18e-01	1
4	9.00e+02	4.41e+03	5.79e-01	4.07e-02	7.79e-02	3
5	9.00e+02	4.41e+03	1.09e-01	3.77e-03	1.72e-03	2
6	9.00e+02	4.41e+03	8.67e-03	2.78e-04	1.30e-04	3
7	9.00e+02	4.41e+03	1.37e-03	2.29e-05	1.09e-05	2
8	9.00e+02	4.41e+03	2.26e-04	9.65e-10	4.01e-07	1
9	9.00e+02	4.41e+03	3.18e-08	2.96e-12	2.62e-07	1
10	9.00e+02	4.41e+03	1.01e-07	4.49e-12	1.88e-08	1
11	9.00e+02	4.41e+03	3.61e-10	2.84e-13	1.46e-09	1

TABLE 7. Solving problem (2.6) with the nonlinear rescaling method, $n = 100$.

and 6 provide the results of the experiment. In addition to the indicators shown in the previous tables, we also show the complementarity violation:

$$comp = \max_{1 \leq i \leq m} |c_i(x)\lambda_i|.$$

For the barrier function method $\lambda_i = 1/(kc_i(x))$, $i = 1, \dots, m$. Such a choice of calculating λ_i is justified by the condition

$$\nabla B_k(x) = \nabla f(x) - \sum_{i=1}^m \frac{1}{kc_i(x)} \nabla c_i(x) = \nabla f(x) - \sum_{i=1}^m \lambda_i \nabla c_i(x) = \nabla L(x, \lambda).$$

One can verify that $\lambda_i c_i(x) = 1/k$, and $c_i(x) > 0$, $i = 1, \dots, m$ always for the barrier method. Therefore, displaying infeasibility and complementarity is a pure formality. Infeasibility is always 0, and complementarity is always $1/k$. Therefore, the accuracy is defined solely by the ability of the method to minimize B_k accurately for large k , which becomes increasingly more difficult to do as the trajectory of the method approaches the solution, where the barrier function does not exist because $c_i(x^*) = 0$, $i \in A$. In fact, as one can see from Tables 5 and 6, problem (2.6) cannot be solved beyond an accuracy of $\approx 10^{-5}$ with IEEE double-precision arithmetic, since Newton's method cannot find a minimizer with that accuracy for $k \approx 10^5$.

Similar to the augmented Lagrangian method, the nonlinear rescaling method (described in the following section) does not require an unbounded increase of k for convergence. Therefore the problem can be accurately solved. Tables 7 and 8 demonstrate that. It takes a smaller number of Newton steps per unconstrained minimization and thus reduces the total number of iterations, yet solves the problem more accurately.

Maj. it.	k	fval	$\ \nabla \mathcal{L}_k\ $	infeas	comp	NS
0	9.00e+02	1.00e+06	2.00e+03	-0.00e+00	1.00e+00	0
1	9.00e+02	3.45e+05	6.24e+02	1.74e-01	1.09e+02	2
2	9.00e+02	3.63e+05	2.93e+02	2.10e-02	3.21e-01	2
3	9.00e+02	3.63e+05	5.62e+01	9.70e-04	2.53e-01	1
4	9.00e+02	3.63e+05	9.78e+00	2.02e-04	6.01e-02	3
5	9.00e+02	3.63e+05	1.27e+00	3.20e-05	1.10e-02	3
6	9.00e+02	3.63e+05	1.53e-01	5.66e-06	1.42e-03	2
7	9.00e+02	3.63e+05	2.90e-02	1.15e-06	1.65e-04	1
8	9.00e+02	3.63e+05	7.41e-04	2.22e-07	3.18e-05	1
9	9.00e+02	3.63e+05	1.46e-05	4.48e-08	8.41e-07	1
10	9.00e+02	3.63e+05	2.69e-07	9.28e-09	1.29e-07	1
11	9.00e+02	3.63e+05	7.49e-09	1.89e-09	7.84e-08	1

TABLE 8. Solving problem (2.6) with the nonlinear rescaling method, $n = 1000$.

3. AUGMENTED LAGRANGIAN, NONLINEAR RESCALING, PROXIMAL-POINT METHODS AND MAXIMAL MONOTONE OPERATORS

Both augmented Lagrangian and nonlinear rescaling methods are proximal-point methods. Their convergence analysis can be analyzed using maximal monotone operators [15]

An operator $T : \mathbb{R}^p \rightarrow \mathbb{R}^p$, which maps $z \in \mathbb{R}^p$ into a closed bounded convex set $T(z)$ is a *monotone operator* if

$$(z - z')^T (w - w') \geq 0 \quad \text{whenever} \quad w \in T(z), w' \in T(z').$$

Let $F : X \times \Lambda \rightarrow \mathbb{R}^1$ be convex in $x \in X$ and concave in $\lambda \in \Lambda$, $X \subset \mathbb{R}^n$, $\Lambda \subset \mathbb{R}^m$. The subdifferential of F at (x, λ) is defined as follows

$$\partial F(x, \lambda) = \{(u, v) : F(x', \lambda) - F(x, \lambda) \geq u^T (x' - x),$$

$$F(x, \lambda') - F(x, \lambda) \leq v^T (\lambda' - \lambda), \forall x' \in X, \lambda' \in \Lambda\}.$$

It is known (see [17]) that the mapping $T : (x, \lambda) \rightarrow \{(u, v) : (u, -v) \in \partial F(x, \lambda)\}$ is a maximal monotone operator. The Lagrangians $L(x, \lambda)$ for convex problems (2.1) and (2.5) are maximal monotone operators. Let $R = \text{diag}(r_i)_{i=1}^{n+m}$ be a diagonal matrix with $r_i \geq 0$, $i = 1, \dots, n+m$. We consider a mapping $RT : (x, \lambda) \rightarrow \{R(u, v)^T : (u, -v) \in \partial F(x, \lambda)\}$, which one obtains by multiplying an image of T by the diagonal matrix R .

Generally speaking, RT is not necessarily a monotone operator for a convex in x and concave in λ function F . However, if F is smooth enough then RT is a monotone operator. The following lemma justifies that (see [6] for the proof).

Lemma 3.1. *Let twice continuously differentiable function $F : \mathbb{R}^n \times \mathbb{R}_+^m \rightarrow \mathbb{R}^1$ be convex in x and concave in λ . Then the rescaled pseudo-gradient*

$$R\nabla F(z) = R \begin{pmatrix} \nabla_x F(x, \lambda) \\ -\nabla_\lambda F(x, \lambda) \end{pmatrix}$$

is a monotone operator.

Both the augmented Lagrangian and nonlinear rescaling methods are related to proximal-point methods in the dual space. Adding a proximal term in the primal space allows for a primal-dual “symmetric treatment” of the convergence analysis. Let us describe both the augmented Lagrangian and nonlinear rescaling methods in more detail.

The augmented Lagrangian method consists of a sequence of inexact minimizations of $\mathcal{L}_k(x, \lambda)$ in x on the $\mathbb{R}^n$ set

$$\hat{x} \approx \hat{x}(\lambda) = \operatorname{argmin}_{x \in \mathbb{R}^n} \mathcal{L}_k(x, \lambda). \quad (3.1)$$

followed by updating the Lagrange multiplier:

$$\hat{\lambda} = \lambda - kg(\hat{x}).$$

We add the quadratic proximal term $\frac{1}{2k}(x-a)^T(x-a)$ to $\mathcal{L}_k$:

$$\mathcal{L}_k(x, a, \lambda) = f(x) - \lambda^T g(x) + \frac{k}{2} g(x)^T g(x) + \frac{1}{2k} (x-a)^T (x-a).$$

Adding the proximal term makes the augmented Lagrangian strongly convex with a modulus of at least $1/k$, so Newton’s method (NM) can be applied. Also, for an iterative method used for solving (3.1), we can define an explicit stopping criterion for the unconstrained minimization based on the gradient of $\mathcal{L}_k$. On the other hand, adding the proximal term $\frac{1}{2k}(x-a)^T(x-a)$ does not spoil convergence and allows a symmetric primal-dual treatment, as can be seen below.

Note that the function $\operatorname{accur}(x, a, \lambda) =: \max\{\|\nabla \mathcal{L}_k(x, a, \lambda)\|, \|g(x)\|\}$ measures the violation of the optimality conditions for problem (2.1), and the condition $\operatorname{accur}(x, x, \lambda) = 0$ is equivalent to satisfying the first-order optimality conditions for problem (2.1).

Figure 1 describes the augmented Lagrangian method for solving problem (2.1).

1. Set $x \in \mathbb{R}^n$, $\lambda = 0$, $\operatorname{rec} = \operatorname{accur}(x, x, \lambda)$.
Select $k > 0$, $\varepsilon > 0$, $0 < \theta < 1$, $\delta \geq 1$.
2. Find $\hat{x} \approx \operatorname{argmin}_{v \in \mathbb{R}^n} \mathcal{L}_k(v, x, \lambda)$ with NM such that $\|\nabla \mathcal{L}_k(\hat{x}, x, \lambda)\| \leq \varepsilon/k$
3. Set $\operatorname{rec} := \operatorname{accur}(\hat{x}, x, \lambda)$
4. Find $\hat{\lambda} = \lambda - kg(\hat{x})$.
5. Set $x := \hat{x}$, $\lambda := \hat{\lambda}$, $\varepsilon := \theta\varepsilon$, $k = k\delta$.
6. If $\operatorname{rec} > \operatorname{RequiredAccuracy}$ then Goto 2.
7. Stop.

FIGURE 1. Augmented Lagrangian Method

The inexact minimization on $\mathbb{R}^n$ performed by Newton’s method in Step 2 of Figure 1.

The convergence analysis of the ALM in Figure 1 relies on the convergence properties of Newton's method and the augmented Lagrangian method. The AL method falls into the category of proximal-point algorithms that have been extensively investigated (see [15]). The convergence of Newton's method for strongly convex functions can be found in many standard optimization texts (see, e.g., [13]).

This paper provides a sketch of the convergence proof. The details can be found elsewhere (see, e.g., [7]). It is worth mentioning that the theory developed by Rockafellar in [15] can be used as is, without modifications, to prove the convergence of the augmented Lagrangian method. The convergence of the general proximal-point algorithm is analyzed in [15]. We will show briefly how the ALM described in Figure 1 can be viewed as a general proximal-point method.

Note that $L(x, \lambda)$ and $\mathcal{L}_k(x, a, \lambda)$ are convex continuous proper functions of x . As functions of x they are differentiable everywhere in $\mathbb{R}^n$, so that subdifferentials exist for any $x \in \mathbb{R}^n$ and $\partial_x L(x, \lambda) = \nabla_x L(x, \lambda)$. Since problem (2.1) is convex, for the primal-dual pair $z^* = (x^*, \lambda^*)$ to be the solution to problem (2.1), it is necessary and sufficient that (x^*, λ^*) satisfies the following conditions

$$\nabla_x L(x^*, \lambda^*) = 0, \quad g(x^*) = 0. \quad (3.2)$$

Consider a primal-dual pair $z = (x, \lambda)$, and the maximal monotone operator

$$T(z) = (\partial_x L(x, \lambda), -\partial_\lambda L(x, \lambda)) = (\nabla_x L(x, \lambda), g(x)).$$

Therefore problem (2.1) is equivalent to finding z^* such that $0 \in T(z^*)$. Despite the fact the subdifferentials of L consist of a single point, we still use the subdifferential notations to stay close to the notations introduced in [15].

We will show that the primal-dual sequence $\{z_s = (x_s, \lambda_s)\}$ generated by the AL algorithm in Figure 1 converges to z^* by invoking Theorem 1 in [15].

Note that the condition $0 \in \partial_x L(x^*, \lambda^*)$ is equivalent to $\nabla_x L(x^*, \lambda^*) = 0$, while the condition $0 \in \partial_\lambda L(x^*, \lambda^*)$ is equivalent to $g(x^*) = 0$.

We define the distance between 0 and a set Z as

$$\text{dist}(0, Z) = \min_{z \in Z} \max_{1 \leq i \leq m} |z_i|.$$

The convergence analysis of the ALM described in Figure 1 relies on the following lemma.

Lemma 3.2. *The augmented Lagrangian method in Figure 1 is equivalent to the following proximal point method*

$$\text{Find } z_{p+1} : \text{dist}(0, S(z_{p+1})) \leq \frac{\varepsilon_p}{k} \quad (3.3)$$

where $z_p = (x_p, \lambda_p)$, $\sum \varepsilon_p < \infty$,

$$S_p(z) = T(z) + k^{-1}(z - z_p),$$

$$T(z) = \partial_z L(z) = (\partial_x L(x, \lambda), -\partial_\lambda L(x, \lambda)),$$

$$L(x, \lambda) = f(x) - \lambda^T g(x).$$

Clearly, $L(x, \lambda)$ is the Lagrangian of problem (2.1) and $T(z)$ is its pseudogradient.

Proof. Let x_{p+1}^* denote an exact minimizer of $\mathcal{L}_k(v, x_p, \lambda_p) = \mathcal{L}_k(v, z_p)$ in v :

$$x_{p+1}^* = \operatorname{argmin}_{v \in \mathbb{R}^n} \mathcal{L}_k(v, z_p).$$

Then we have $\nabla_x \mathcal{L}(x_{p+1}^*, z_p) = 0$ and

$$0 \in \partial_x \mathcal{L}_k(x_{p+1}^*, z_p) = \nabla f(x_{p+1}^*) - \lambda_p^T \nabla g(x_{p+1}^*) + k g(x_{p+1}^*)^T \nabla g(x_{p+1}^*) + \frac{1}{k} (x_{p+1}^* - x_p).$$

Using $\lambda_{p+1}^* = \lambda_p - k g(x_{p+1}^*)$, the above expression can be rewritten as

$$\begin{aligned} 0 \in \partial_x \mathcal{L}_k(x_{p+1}^*, z_p) &= \nabla f(x_{p+1}^*) - \lambda_{p+1}^{*T} \nabla g(x_{p+1}^*) + \frac{1}{k} (x_{p+1}^* - x_p) \\ &= \nabla_x L(x_{p+1}^*, \lambda_{p+1}^*) + \frac{1}{k} (x_{p+1}^* - x_p) \end{aligned}$$

Therefore x_{p+1}^* is also the minimizer of $L(x, \lambda_{p+1}^*) + \frac{1}{2k} (x - x_p)^T (x - x_p) + C$ in x , where C is some constant.

Now consider the maximization of a smooth and strongly concave function

$$\max_{\lambda \in \mathbb{R}^m} d(\lambda),$$

where

$$d(\lambda) = L(x_{p+1}^*, \lambda) + 0.5k(x_{p+1}^* - x_p)^T (x_{p+1}^* - x_p) - \frac{1}{2k} (\lambda - \lambda_p)^T (\lambda - \lambda_p)$$

is smooth and strongly concave in λ . Therefore the maximizer $\lambda_{\max}$ satisfies the following equation

$$\nabla d(\lambda_{\max}) = -g(x_{p+1}^*)^* - \frac{1}{k} (\lambda_{\max} - \lambda_p) = 0.$$

Rewriting the above equation yields $\lambda_{\max} = \lambda_p - k g(x_{p+1}^*)$. Therefore $\lambda_{p+1}^* = \lambda_{\max}$ and thus λ_{p+1}^* maximizes $d(\lambda)$ in λ .

To summarize, what we obtained so far is that for the function

$$Y_p(x, \lambda) = L(x, \lambda) + \frac{1}{2k} (x - x_p)^T (x - x_p) - \frac{1}{2k} (\lambda - \lambda_p)^T (\lambda - \lambda_p),$$

which is strongly convex in x and strongly concave in λ we have

$$Y_p(x_{p+1}^*, \lambda_{p+1}^*) = \min_{x \in \mathbb{R}^n} Y_p(x, \lambda_{p+1}^*) = \max_{\lambda \in \mathbb{R}^m} Y_p(x_{p+1}^*, \lambda) = \min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^m} Y_p(x, \lambda).$$

In other words, the pair $(x_{p+1}^*, \lambda_{p+1}^*)$ is the solution of the minmax problem with the convex-concave objective function $Y_p(x, \lambda)$ and therefore

$$0 \in \partial_x L(x_{p+1}^*, \lambda_{p+1}^*) + \frac{1}{k} (x_{p+1}^* - x_p),$$

and

$$0 \in -\partial_\lambda L(x_{p+1}^*, \lambda_{p+1}^*) + \frac{1}{k} (\lambda_{p+1}^* - \lambda_p),$$

or, after combining the last two inclusions, we have

$$0 \in (\partial_x L(z_{p+1}^*), -\partial_\lambda L(z_{p+1}^*)) + \frac{1}{k} (z_{p+1}^* - z_p) = T(z_{p+1}^*) + \frac{1}{k} (z_{p+1}^* - z_p) = S_p(z_{p+1}^*).$$

The above condition is equivalent to $\nabla_x \mathcal{L}(x_{p+1}^*, x_p, \lambda_p) = 0$.

Now we show that the stopping condition for an inexact minimizer

$$\|\nabla_x \mathcal{L}_k(x_{p+1}, x_p, \lambda_p)\| \leq \varepsilon_p/k$$

is equivalent to $\text{dist}(0, S(z_{p+1})) \leq \varepsilon_p/k$, where $\varepsilon_p = \varepsilon \theta^p$. Since $Y_p(x, \lambda)$ is the smooth and strongly concave in λ , its maximization is solved precisely by

$$\lambda_{p+1} = \lambda_p - kg(x_{p+1}).$$

Then $\text{dist}(0, S(z_{p+1})) = \text{dist}(0, \partial_v \mathcal{L}_k(x_{p+1}, x_p, \lambda_p))$. Therefore an iteration of the augmented Lagrangian algorithm is equivalent to (3.3). $\square$

The following lemma establishes convergence of the Newton's method to a minimizer of $\mathcal{L}_k(x, x_p, \lambda_p)$ in x and is proven in [13].

Lemma 3.3. *For any $\varepsilon_p > 0$ the sequence $\{v_s\}$ generated by Newton's method with Armijo rule for unconstrained minimization of $\mathcal{L}_k(v, x_p, \lambda_p)$ terminates with a finite number of steps with x_{p+1} satisfying*

$$\text{dist}(0, S(z_{p+1})) = \text{dist}(0, \partial_v \mathcal{L}_k(x_{p+1}, x_p, \lambda_p)) \leq \varepsilon_p/k.$$

The proof follows from the convergence of Newton's method with an Armijo rule for the stepsize, applied to the minimization of a strongly convex function. It follows from Lemma 3.3 that Line 2 in the augmented Lagrangian method described in Figure 1 is well defined.

We are ready to formulate the main convergence result.

Theorem 3.4. *The sequence $\{(x_p, \lambda_p)\}$ generated by the ALM in Figure 1 has a unique cluster primal-dual pair (x^*, λ^*) that satisfies the first order optimality conditions $\nabla_x L(x^*, \lambda^*) = 0$ and $g(x^*) = 0$, i.e. $\{(x_p, \lambda_p)\}$ converges to the optimal solution of problem (2.1) in the weak sense.*

Proof. Consider the following claims.

1) The sequence $\{z_p\} = \{(x_p, \lambda_p)\}$ generated by the ALM is well defined in Lines 2 and 4 of Figure 1. The primal sequence $\{x_p\}$ is well defined by Lemma 3.3. Once x_p is available then λ_p is calculated by the explicit formulas in Line 4, so the dual sequence $\{\lambda_p\}$ is also well defined. Therefore the sequence $\{z_p\}$ generated in (3.3) is also well defined.

2) Since $\delta \geq 1$, the sequence of the scaling parameters $k_p = k_0 \delta^p$ is strictly positive and nondecreasing if $k_0 > 0$.

3) Since $0 < \theta < 1$, the sequence $\{\varepsilon_p = \varepsilon_0 \theta^p\}$ satisfies $\sum_{p=0}^{\infty} \varepsilon_p < \infty$.

4) The solution x^* to (2.1) exists by assumption and such that the first-order optimality conditions hold.

5) The above optimality condition is equivalent to the following optimality conditions for the problem (2.1)

$$0 \in (\partial_x L(x^*, \lambda^*), -\partial_\lambda L(x^*, \lambda^*))$$

or

$$0 \in T(z^*).$$

6) The existence of the solution to the equivalent problem

$$\text{Find } z^* : 0 \in T(z^*)$$

guarantees the boundedness of the sequence $\{z_p\}$ generated in (3.3) (see [15]).

7) All the conditions of Rockafellar's Theorem 1 in [15] are satisfied, therefore the sequence $\{z_p\}$ generated in (3.3) converges to some z^* in a weak sense, i.e. there exists a unique cluster

point z^* that satisfies $0 \in T(z^*)$. The above inclusion also means that the primal-dual sequence generated by the ALM in Figure 1 converges to the primal-dual solution (x^*, λ^*) in a weak sense. $\square$

Remark 3.5. The proof above is based on showing that the sequence generated by the augmented Lagrangian method fits a general proximal-point algorithm analyzed by Rockafellar in [15], and converges to a zero of the maximal monotone operator, which, in turn, is the solution to the convex optimization problem with equality constraints (2.1). It is important to emphasize that the maximal monotone operator is defined by the problem data and does not depend on the primal-dual trajectory $\{(x_s, \lambda_s)\}$.

Remark 3.6. In the presentation of the material, subdifferential notation can be avoided since the subdifferentials consist of a single vector. However, we choose to stick to subdifferential notation to stay closer to the notation of [15].

Remark 3.7. For optimization problems with inequality constraints

$$0 \notin (\partial_x L(x^*, \lambda^*), -\partial_\lambda L(x^*, \lambda^*)),$$

since for the passive constraints $c_i(x^*) > 0$, $i \in P$. Therefore, Theorem 1 from [15] cannot be applied directly.

For solving convex problems with inequality constraints, we employ nonlinear rescaling methods. The nonlinear rescaling principle was developed by Polyak [13]. It is a generalization of the modified barrier function method [11]. The functions that define the inequality constraints are rescaled with a transformation that has certain characteristics described below. As a result, the rescaled problem is equivalent to the original problem yet possesses additional useful properties that can be exploited.

The rescaling of the functions that define the constraints is accomplished using a transformation $\psi : \mathbb{R} \rightarrow \mathbb{R}$ with the following properties:

$$1^0 \psi'(t) > 0, \psi''(t) < 0 \text{ for all } t \in \mathbb{R}.$$

$$2^0 \psi(0) = 0, \psi'(0) = 1.$$

$$3^0 \psi'(t) \leq a/(t+1) \text{ for all } t \geq 0 \text{ and some } a > 0.$$

$$4^0 \alpha e^{-\beta t} \leq -\psi''(t) \leq \gamma/(t+1)^2 \text{ for all } t \geq 0 \text{ and some } \alpha, \beta, \gamma > 0.$$

One can verify that well known transformations $\psi_1(t) = 1 - e^{-t}$ [2], $\psi_2(t) = \ln(t+1)$, $\psi_3(t) = t/(t+1)$ [11] belong to Ψ . Other transformations $\psi \in \Psi$ can be found in [12].

An example of the transformations ψ that satisfy properties (1)-(4) and are used for numerical experiments in Section 2 is the modified barrier function $\psi(t) = \ln(1+t)$, smoothly “glued” with a parabola at $t = -0.5$ so that it is defined on $\mathbb{R}$.

$$\psi(t) = \begin{cases} \ln(t+1), & t > -0.5 \\ -2t^2 + \ln(.5) + .5, & t \leq -0.5. \end{cases} \quad (3.4)$$

The transformation ψ rescales the functions c_i , $i = 1, \dots, m$ to form an equivalent problem. For any given vector $k = (k_1, \dots, k_m) \in \mathbb{R}_{++}^m$ due to 1^0 and 2^0 we have

$$c_i(x) \geq 0 \Leftrightarrow k_i^{-1} \psi(k_i c_i(x)) \geq 0, \quad i = 1, \dots, m. \quad (3.5)$$

Therefore, the problem

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \\ & \text{subject to} \quad k_i^{-1} \psi(k_i c_i(x)) \geq 0, \quad i = 1, \dots, m \end{aligned} \quad (3.6)$$

is equivalent to the primal problem (2.5).

The Lagrangian $\mathcal{L}_k : \mathbb{R}^n \times \mathbb{R}_{++}^m \rightarrow \mathbb{R}$ for the equivalent problem, is given by formula

$$\mathcal{L}_k(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i k_i^{-1} \psi(k_i c_i(x)).$$

We add the primal proximal term $\frac{1}{2k} \|x - a\|^2$ to the Lagrangian so that it becomes a strongly convex function of x .

Let $x^0 \in \mathbb{R}^n$ and $\lambda^0 \in \mathbb{R}_{++}^m$ be initial primal and dual approximations, $k > 0$.

The NR method generates three sequences $\{k^s\} \subset \mathbb{R}_{++}^m$, $\{x^s\} \subset \mathbb{R}^n$, $\{\lambda^s\} \subset \mathbb{R}_{++}^m$, by formulas

$$k^s = (k_i^s = k(\lambda_i^s)^{-1}, \quad i = 1, \dots, m), \quad (3.7)$$

$$x^{s+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ \mathcal{L}_{k^s}(x, \lambda^s) + \frac{1}{2k} \|x - x^s\|^2 \mid x \in \mathbb{R}^n \right\} \quad (3.8)$$

$$\lambda^{s+1} = (\lambda_i^{s+1} = \lambda_i^s \psi'(k_i^s c_i(x^{s+1})), \quad i = 1, \dots, m.) \quad (3.9)$$

The numerical realization of the NR method (3.7)-(3.9) requires replacing the primal minimizer with its approximation. To simplify the presentation we use the same notation for the NR method (3.7)-(3.9) and its modification (3.10)-(3.12).

Let $\{\varepsilon_s > 0\}$ be a sequence such that $\sum_{s=0}^{\infty} \varepsilon_s < \infty$. Then the modified NR method generates the following three sequences $\{k^s\} \subset \mathbb{R}_{++}^m$, $\{x^s\} \subset \mathbb{R}^n$, $\{\lambda^s\} \subset \mathbb{R}_{++}^m$,

$$k^s = (k_i^s = k(\lambda_i^s)^{-1}, \quad i = 1, \dots, m), \quad (3.10)$$

$$x^{s+1} : \|\nabla_x \mathcal{L}_{k^s}(x^{s+1}, \lambda^s) + \frac{1}{k}(x^{s+1} - x^s)\| \leq \frac{\varepsilon_s}{k} \quad (3.11)$$

$$\lambda^{s+1} = (\lambda_i^{s+1} = \lambda_i^s \psi'(k_i^s c_i(x^{s+1})), \quad i = 1, \dots, m.) \quad (3.12)$$

Remark 3.8. Due to the convergence of the trajectory generated by Newton's method with Armijo's rule for minimizing a strongly convex function, condition (3.11) is well defined.

The minimizer x^{s+1} given by (3.8) or its approximation (3.11) exists and it is uniquely defined for any given $\lambda^s \in \mathbb{R}_{++}^m$ and $k^s \in \mathbb{R}_{++}^m$ due to the convexity of f , the concavity of c_i , the properties 1⁰ – 4⁰ of $\psi \in \Psi$, and the strong convexity of the proximal term $\frac{1}{2k_s} \|x - x^s\|^2$ in (3.8). Therefore the modified NR method (3.10)-(3.12) is well defined. The formula (3.7) with that particular “dynamic” update of the scaling parameter has been introduced in [18] for the exponential multipliers method and used on several occasions (see [1, 12]) in the framework of NR methods. The convergence analysis, described in detail in [6], extends the convergence scheme, developed in [15, 16] for quadratic augmented Lagrangians, to a class of nonquadratic augmented Lagrangians.

The main idea behind the NR convergence analysis is also the equivalence of the NR method (3.10)-(3.12) to a quadratic proximal point method for finding a saddle point of the classical Lagrangian for the original problem. The primal proximal term $\frac{1}{2k} \|x - x^s\|^2$ is standard while

the dual proximal term turned out to be a quadratic term in the rescaled dual space. The rescaling factors are dynamic and vary from one iteration to another.

In the following part of the section we outline the basic steps of the convergence analysis of the NR method (3.10)–(3.12) based on the following several lemmas.

Lemma 3.9. *One step of the NR method (3.7)–(3.9) is equivalent to finding a saddle point (x^{s+1}, λ^{s+1}) of the following function*

$$M(x, \lambda, x^s, \lambda^s) = L(x, \lambda) + \frac{1}{2k} \|x - x^s\|^2 - \frac{1}{2k} \|\lambda - \lambda^s\|_{R_{s+1}^{-1}}^2$$

where $\|\lambda\|_{R_{s+1}^{-1}}^2 = \lambda^T R_{s+1}^{-1} \lambda$, and $R_{s+1} = \text{diag}(r_i^{s+1})_{i=1}^m$, $r_i^{s+1} = -[\psi''(\theta_i^s k_i^s c_i(x^{s+1}))] > 0$ and $0 < \theta_i^s < 1$, i.e.

$$(x^{s+1}, \lambda^{s+1}) : \max_{\lambda \in \mathbb{R}_+^m} \min_{x \in \mathbb{R}^n} M(x, \lambda, x^s, \lambda^s) = \min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}_+^m} M(x, \lambda, x^s, \lambda^s). \quad (3.13)$$

Keeping in mind that $\nabla_\lambda L(x, \lambda) = -c(x)$, the solution of the proximal minmax problem (3.13) one can view as an application of the proximal operator $P_R : \mathbb{R}^n \times \mathbb{R}_{++}^m \rightarrow \mathbb{R}^n \times \mathbb{R}_{++}^m$ defined by a formula

$$w = P_R(z), \quad (3.14)$$

where $P_R(z) = (I + kT_R)^{-1}(z)$, $z = (x, \lambda)$ and the operator $T_R : \mathbb{R}^n \times \mathbb{R}_{++}^m \rightarrow \mathbb{R}^n \times \mathbb{R}_{++}^m$ is defined by the formula

$$T_R(z) = \begin{pmatrix} \nabla_x L(x, \lambda) \\ -R \nabla_\lambda L(x, \lambda) \end{pmatrix}.$$

Therefore the sequence generated by the NR method (3.7)–(3.9) can be described as

$$z^{s+1} = P_{R_{s+1}}(z^s), \quad s = 0, 1, \dots$$

with a starting point $z^0 = (x^0, \lambda^0)$, $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}_{++}^m$ and $R_{s+1} = \text{diag}(r_i^{s+1})_{i=1}^m$, $r_i^{s+1} = -[\psi''(\theta_i^s k_i^s c_i(x^{s+1}))] > 0$ and $0 < \theta_i^s < 1$.

It is important to emphasize that the operators R_s depend on the trajectory generated by the NR algorithm (3.7)–(3.9). Also, as will follow from the convergence proof, the elements r_i^s that correspond to the passive constraints approach 0 as the NR trajectory approaches the solution. Therefore, we cannot directly invoke Theorem 1 by Rockafellar [15] to show the convergence of the NR method based on a fixed monotone operator $T(z)$.

However, by following the steps developed by Rockafellar to prove Theorem 1 [15], we can modify them to prove the convergence based on the rescaled monotone operator $T_R(z)$.

The operators T_R and P_R are defined on $\mathbb{R}^n \times \mathbb{R}_{++}^m$. For convergence analysis, however, we need these operators to be defined on $\mathbb{R}^n \times \mathbb{R}_+^m$. We define the generalized operators $T_R(z)$ and consequently P_R in the following way:

$$T_R(z) = \begin{pmatrix} \nabla_x L(x, \lambda) \\ -R \nabla_\lambda L(x, \lambda) \end{pmatrix}, \quad P_R(z) = (I + kT_R)^{-1}(z)$$

where $R = \text{diag}(r_i)_{i=1}^m$, and $\lambda_i = 0 \Rightarrow r_i = 0$.

The modified operators described above satisfy the following properties at any primal-dual solution $z^* = (x^*, \lambda^*) \in X^* \times \Lambda^*$:

$$T_R(z^*) = 0,$$

and

$$P_R(z^*) = z^*. \quad (3.15)$$

The convergence proof of the NR method (3.10)–(3.12) relies on properties of the family of proximal operators $P_{R_{s+1}}$, $s = 1, 2, \dots$, where $R_{s+1} = \text{diag}(r_i^{s+1})_{i=1}^m$ and

$$r_i^{s+1} = \begin{cases} -[\psi''(\theta_i^s k_i^s c_i(x^{s+1}))], & 0 < \theta_i^s < 1, \text{ if } \lambda_i > 0 \\ 0, & \text{if } \lambda_i = 0 \end{cases}$$

The latter, in turn, are closely related to properties of monotone operators $T_{R_{s+1}}$.

The Largangian $L(x, \lambda)$ satisfies the conditions of Lemma 3.1. Therefore it follows from Lemma 3.1 that for a given vector $r \in \mathbb{R}_+^m$ and $R = \text{diag}(r_i)_{i=1}^m$ the operator $T_R : \mathbb{R}^n \times \mathbb{R}_+^m$ defined by a formula

$$T_R(z) = \begin{pmatrix} \nabla_x L(x, \lambda) \\ -R \nabla_y L(x, \lambda) \end{pmatrix},$$

is a monotone operator. i.e. the inequality

$$(T_R(z_1) - T_R(z_2))^T (z_1 - z_2) \geq 0 \quad (3.16)$$

holds for any pair $(z_1, z_2) : z_1 \in \mathbb{R}^n \times \mathbb{R}_+^m, z_2 \in \mathbb{R}^n \times \mathbb{R}_+^m$.

The following lemma shows that the proximal operator $P_{R_{s+1}}$ associated with the monotone operator $T_{R_{s+1}}$ is nonexpansive (see [6] for the proof).

Lemma 3.10. *For any vectors $a, b \in \mathbb{R}^n \times \mathbb{R}_+^m$ and the operator $P_{R_{s+1}}$, the following estimation holds*

$$\|P_{R_{s+1}}(a) - P_{R_{s+1}}(b)\| \leq \|a - b\|, \quad \forall s \geq 0.$$

The next lemma justifies using an approximate minimizer in (3.11).

Lemma 3.11. *For the NR method (3.10)–(3.12), the following estimation holds*

$$\|z^{s+1} - P_{R_{s+1}}(z^s)\| \leq \varepsilon_s.$$

The next lemma establishes the boundedness of the sequence generated by the NR method (3.10)–(3.12) (see [6] for the proof).

Lemma 3.12. *The sequence $\{z^s = (x^s, \lambda^s)\}$ generated by the NR method (3.10)–(3.12) is bounded, i.e. there exists $M > 0$ that*

$$\|z^s\| \leq M < \infty. \quad (3.17)$$

Let us consider the mapping $Q_{R_{s+1}} = I - P_{R_{s+1}}$, then for any $s \geq 0$ we have $Q_{R_{s+1}}(z^*) = 0$, where z^* is the solution to the problem (2.5).

The following auxiliary statements, which will be used later can be proved by the arguments identical to Proposition 1 in [15].

Lemma 3.13. *For the operators $P_{R_{s+1}}$ and $Q_{R_{s+1}}$, $s = 0, 1, 2, \dots$ the following is true*

- a) $z = P_{R_{s+1}}(z) + Q_{R_{s+1}}(z)$ for all z .
- b) $(k_s)^{-1} Q_{R_{s+1}}(z) = (k_s)^{-1} (z - P_{R_{s+1}}(z)) = T_{R_{s+1}}(P_{R_{s+1}}(z))$ for all z .
- c) $(P_{R_{s+1}}(z) - P_{R_{s+1}}(z'))^T (Q_{R_{s+1}}(z) - Q_{R_{s+1}}(z')) \geq 0$ for all z, z' .
- d) $\|P_{R_{s+1}}(z) - P_{R_{s+1}}(z')\|^2 + \|Q_{R_{s+1}}(z) - Q_{R_{s+1}}(z')\|^2 \leq \|z - z'\|^2$ for all z, z' .

The following two lemmas justifies the convergence of the rescaled proximal-point method (see [6] for the proof).

Lemma 3.14. *The sequence $\{z^s = (x^s, \lambda^s)\}$ generated by the NR method (3.10)–(3.12) satisfies the following condition*

$$\lim_{s \rightarrow \infty} T_{R_{s+1}}(z^s) = 0. \quad (3.18)$$

Lemma 3.15. *The sequence $\{z^s = (x^s, \lambda^s)\}$ generated by the NR method (3.10)–(3.12) satisfies the following conditions*

$$\lim_{s \rightarrow \infty} \|z^{s+1} - z^s\| = 0. \quad (3.19)$$

Finally the next lemma shows that KKT conditions are hold for the limit of the NR trajectory (see [6] for the proof).

Lemma 3.16. *Let $\{k_s\}$ be a nondecreasing bounded sequence of positive parameters. Then the sequence $\{z^s = (x^s, \lambda^s)\}$ generated by the NR method (3.10)–(3.12) satisfies the following conditions*

- a) $\lim_{s \rightarrow \infty} \nabla_x L(x^s, \lambda^s) = 0,$
- b) $\lim_{s \rightarrow \infty} \lambda_i^s \geq 0, i = 1, \dots, m.$
- c) $\lim_{s \rightarrow \infty} c_i(x^s) \geq 0, i = 1, \dots, m.$
- d) $\lim_{s \rightarrow \infty} c_i(x^s) \lambda_i^s = 0, i = 1, \dots, m,$

Thus the following theorem takes place.

Theorem 3.17. *If assumptions **A** and **B** are satisfied and $k > 0$, then any limit point $\bar{z} = (\bar{x}, \bar{\lambda})$ of the sequence $\{z^s = (x^s, \lambda^s)\}$ generated by the modified NR method (3.10)–(3.12) is the primal-dual solution, i.e. $(\bar{x}, \bar{\lambda}) \in X^* \times \Lambda^*$. $\square$*

Proof. By Lemma 3.16 $\bar{z} = (\bar{x}, \bar{\lambda})$ satisfies the KKT's optimality conditions. Therefore for the convex optimization problem (2.5), we have $\bar{x} \in X^*$ and $\bar{\lambda} \in \Lambda^*$, i.e. $\bar{z} = (\bar{x}, \bar{\lambda})$ is the primal-dual solution. $\square$

The convergence of the method (3.7)–(3.9) also follows from Theorem 3.17.

Remark 3.18. Under the standard second-order optimality conditions, both the augmented Lagrangian and nonlinear rescaling methods converge with a Q-linear rate (see [15, 6], for example).

Remark 3.19. For a problem with both equality and inequality constraints, the augmented Lagrangian and nonlinear rescaling methods can be combined. The convergence analysis for the resulting nonlinear rescaling-augmented Lagrangian method can be done using arguments similar to those described in this manuscript.

4. CONCLUDING REMARKS

This manuscript outlines the basic steps for the convergence analysis of two Lagrange multiplier methods for convex optimization problems with equality and inequality constraints respectively: the augmented Lagrangian method and the nonlinear rescaling method,

The convergence analysis for both methods relies on the theory developed by Rockafellar in [15] for the proximal-point method based on maximal monotone operators. The trajectory generated by the augmented Lagrangian method satisfies the conditions of Theorem 1 in [15], so the convergence follows from that theorem. For the nonlinear rescaling method, Rockafellar's theory can be extended to work for rescaled monotone operators. But in either case, the significance of Rockafellar's theory for convergence analysis is hard to overstate.

Looking at the convergence analysis for both the AL and NR methods, one can observe again that algorithms for solving optimization problems with inequality constraints are more difficult to analyze than those for a problem with equality constraints only, due to the presence of both active and passive constraints in a problem with inequalities. This fact is in agreement with other similar examples from the past, including the additional challenges of proving the KKT conditions as compared to the Lagrange first-order optimality conditions for a problem with equations only.

The significance of the theory developed by Rockafellar for the proximal-point method spans much beyond the convergence analysis of the methods of Lagrange multipliers described in this manuscript. Proximal-point methods in general serve as an excellent alternative to homotopy-like methods based on pushing penalty, barrier, smoothing, or regularization parameters to the extreme to achieve convergence of the trajectories generated by those methods.

Proximal-point methods may avoid some computational issues and obtain solutions with high accuracy, as numerical experiments suggest. Therefore, we cautiously predict that the scientific and engineering communities will keep relying on them. At the same time, we believe that the analysis of new proximal-point methods to be developed in the future will find some help in Rockafellar's theory.

REFERENCES

- [1] A. Auslender, M. Teboulle, and S. Ben-Tiba, Interior Proximal and Multipliers methods based on second-order homogeneous kernels, *Mathematics of Operations Research*, 24 (1999) 645-668.
- [2] D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
- [3] R. Courant, Variational methods for the solution of problems of equilibrium and vibrations, *Bull. Amer. Math. Soc.* 49 (1943) 1-23.
- [4] A. V. Fiacco, and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley, New York, 1968.
- [5] K. Frisch, The logarithmic potential method for solving linear programming problems, *Memorandum*, University Institute of Economics, Oslo, 1955.
- [6] I. Griva, R. Polyak, Proximal point nonlinear rescaling method for convex optimization, *Numerical Algebra, Control and Optimization* 1(2) (2011) 283-299.
- [7] I. Griva, Convergence analysis of augmented Lagrangian - fast projected gradient method for convex quadratic problems, *Pure and Applied Functional Analysis* 3(3) (2018) 417-428.
- [8] M. R. Hestenes, Multiplier and gradient methods, *J. Optimization Theory and Appl.* 4 (1969) 303-320.
- [9] H. Kuhn, A. Tucker, *Nonlinear Programming*, In: J. Neyman (ed.), *Proceeding of the Second Berkley Symposium on Mathematical Statistics and Probability*, pp.481-492, University of California Press, Berkley, 1951.
- [10] G. J. Minty, Monotone (nonlinear) operators in Hilbert space, *Duke Math. J.* 29 (1962) 341-346.
- [11] R. A. Polyak, Modified barrier functions (theory and methods), *Mathematical Programming* 54(2) (1992) 177-222.
- [12] R. A. Polyak, Nonlinear rescaling multipliers method as interior quadratic prox, *Computational Optimization and Applications* 35 (2006) 347-373.
- [13] R. A. Polyak, *Introduction to Continuous Optimization*, Springer, Switzerland, 2021.

- [14] M. J. D. Powell, A method for nonlinear optimization in minimization problems, In: R. Fletcher, (ed), Optimization, pp. 283-298, Academic Press, New York, 1969.
- [15] R. T. Rockafellar, Monotone Operators and The Proximal Point Algorithm, SIAM Journal of Control and Optimization 14 (5) (1976) 88-898.
- [16] R. T. Rockafellar, Augmented Lagrangians and Applications of the Proximal Point algorithm in convex programming, Mathematics of Operations Research 1 (1976) 97-116.
- [17] R. T. Rockafellar, Convex analysis, Princeton University Press, Princeton, 1996.
- [18] P. Tseng, D. Bertsekas, On the convergence of exponential multipliers method for convex programming, Mathematical Programming 76 (1993) 1-19.