Augmented Lagrangian Filter Method*

Sven Leyffer†

November 9, 2016

Abstract

We introduce a filter mechanism to force convergence for augmented Lagrangian methods for nonlinear programming. In contrast to traditional augmented Lagrangian methods, our approach does not require the use of forcing sequences that drive the first-order error to zero. Instead, we employ a filter to drive the optimality measures to zero. Our algorithm is flexible in the sense that it allows for equality constraint quadratic programming steps to accelerate local convergence. We also include a feasibility restoration phase that allows fast detection of infeasible problems. We give a convergence proof that shows that our algorithm converges to first-order stationary points.

Keywords: Augmented Lagrangian, filter methods, nonlinear optimization.

AMS-MSC2000: 90C30

1 Introduction

Nonlinearly constrained optimization is one of the most fundamental problems in scientific computing with a broad range of engineering, scientific, and operational applications. Examples include nonlinear power flow [4, 27, 51, 53, 58], gas transmission networks [28, 50, 15], the coordination of hydroelectric energy [21, 17, 55], and finance [26], including portfolio allocation [45, 38, 64] and volatility estimation [23, 2]. Chemical engineering has traditionally been at the forefront of developing new applications and algorithms for nonlinear optimization; see the surveys [11, 12]. Applications in chemical engineering include process flowsheet design, mixing, blending, and equilibrium models. Another area with a rich set of applications is optimal control [10]. Optimal control applications include the control of chemical reactions, the shuttle re-entry problem [16, 10], and the control of multiple airplanes [3]. More important, nonlinear optimization is a basic building block of more complex design and optimization paradigms such as integer and nonlinear optimization [1, 29, 42, 46, 13, 5] and optimization problems with complementarity constraints [47, 56, 48].

*Preprint ANL/MCS-P6082-1116

†Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439, leyffer@mcs.anl.gov.

Nonlinearly constrained optimization has been studied intensely for more than 30 years, resulting in a range of algorithms, theory, and implementations. Current methods fall into two competing classes: active set [39, 40, 30, 18, 22, 34] and interior point [35, 44, 63, 62, 8, 61, 19, 20]. Both classes are Newton-like schemes; and while both have their relative merits, interior-point methods have emerged as the computational leader for large-scale problems.

The Achilles heel of interior-point methods is our inability to efficiently warm-start these methods. Despite significant recent advances [41, 6, 7], interior-point methods cannot compete with active-set approaches when solving mixed-integer nonlinear programs [14]. This deficiency is at odds with the rise of complex optimization paradigms such as nonlinear integer optimization that require the solution of thousands of closely related nonlinear problems and drive the demand for efficient warm-start techniques. On the other hand, active-set methods enjoy excellent warm-starting potential. Unfortunately, current active-set methods rely on pivoting approaches and do not readily scale to multicore architectures (some successful parallel approaches to linear programming (LP) active-set solvers can be found in the series of papers [43, 60, 49]). To overcome this challenge, we study augmented Lagrangian methods, which combine better parallel scalability potential with good warm-starting capabilities.

We consider solving the following nonlinear program (NLP),

$$\underset{x}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x) = 0, \quad x \geq 0, \tag{NLP}$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $c : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable. We use superscripts, $x^{(k)}$, to indicate iterates and evaluation of nonlinear functions, such as $f^{(k)} := f(x^{(k)})$ and $\nabla c^{(k)} = \nabla c(x^{(k)})$. The Lagrangian of (NLP) is defined as $\mathcal{L}_0(x, y) = f(x) - y^T c(x)$, where $y \in \mathbb{R}^m$ is a vector of Lagrange multipliers of $c(x) = 0$.

The first-order optimality conditions of (NLP) can be written as

$$\min \{\nabla_x \mathcal{L}_0(x, y)\,,\; x\} = 0, \tag{1.1a}$$

$$c(x) = 0, \tag{1.1b}$$

where the $\min$ is taken componentwise and (1.1a) corresponds to the usual first-order condition, $0 \leq x \perp \nabla_x \mathcal{L}_0(x, y) \geq 0$.

1.1 Augmented Lagrangian Methods

The augmented Lagrangian is defined as

$$\mathcal{L}_\rho(x, y) = f(x) - y^T c(x) + \tfrac{1}{2}\rho\|c(x)\|^2. \tag{1.2}$$

Augmented Lagrangian methods have been studied by [9, 54, 52, 57]. Recently, researchers have expressed renewed interest in augmented Lagrangian methods because of their good scalability properties, which had already been observed in [24]. The key computational step in augmented

Lagrangian methods is the minimization of $\mathcal{L}_{\rho_k}(x, y^{(k)})$ for fixed $\rho_k, y^{(k)}$, giving rise to the bound-constrained Lagrangian (BCL($y^{(k)}, \rho_k$)) problem:

$$\underset{x \geq 0}{\text{minimize}} \ \mathcal{L}_{\rho_k}(x, y^{(k)}) = f(x) - y^{(k)^T} c(x) + \frac{1}{2}\rho_k \|c(x)\|^2 = \mathcal{L}_0 + \frac{1}{2}\rho_k \|c(x)\|^2 \qquad (\text{BCL}(y^{(k)}, \rho_k))$$

for given $\rho_k > 0$ and $y^{(k)} \in \mathbb{R}^m$. We denote the solution of (BCL($y^{(k)}, \rho_k$)) by $x^{(k+1)}$. A basic augmented Lagrangian method solves (BCL($y^{(k)}, \rho_k$)) approximately and updates the multipliers using the so-called first-order multiplier update:

$$y^{(k+1)} = y^{(k)} - \rho_k c(x^{(k+1)}). \tag{1.3}$$

Traditionally, augmented Lagrangian methods have used two forcing sequences, $\eta_k \searrow 0$ and $\omega_k \searrow 0$, to control the infeasibility and first-order error and enforce global convergence. Sophisticated update schemes for η, ω can be found in [25]. A rough outline of an augmented Lagrangian method is given in Algorithm 1.

Given sequences $\eta_k \searrow 0$ and $\omega_k \searrow 0$; an initial point $(x^{(0)}, y^{(0)})$, and ρ_0, set $k = 0$;
while $(x^{(k)}, y^{(k)})$ *not optimal* **do**
    Find an ω_k-optimal solution, $x^{(k+1)}$, of (BCL($y^{(k)}, \rho_k$)) ;
    if $\|c(x^{(k+1)})\| \leq \eta_k$ **then**
        Update the multipliers: $y^{(k+1)} = y^{(k)} - \rho_k c(\hat{x}^{(k)})$;
    else
        Increase the penalty parameter: $\rho_{k+1} = 2\rho_k$;

Algorithm 1: Bound-Constrained Augmented Lagrangian Method

Our goal is to improve traditional augmented Lagrangian methods in three ways, extending the augmented Lagrangian filter methods developed in [37] for quadratic programs to general NLPs.

1. Replace the forcing sequences, η_k, ω_k by a less restrictive algorithmic construct, namely, a filter (defined in Section 2).

2. Introduce a second-order step to promote fast local convergence, similar to recent sequential linear quadratic programming (SLQP) methods [18, 22, 34].

3. Equip the augmented Lagrangian method with fast and robust detection of infeasible subproblems [32].

Ultimately, we would like to develop new active-set methods that can take advantage of emerging multicore architectures. Traditional sequential quadratic programming (SQP) and newer SLQP approaches [18, 22, 34] that rely on pivoting techniques to identify the optimal active set are not suitable because the underlying active-set strategies are not scalable to multicore architectures. On the other hand, the main computational kernels in augmented Lagrangian methods (bound constrained minimization and the solution of augmented systems) enjoy better scalability properties.

This paper is organized as follows. The next section defines the filter for augmented Lagrangians, and outlines our method. Section 3 presents the detailed algorithm and its components, and Section 4 presents the global convergence proof. We close the paper with some conclusions and outlooks.

2 An Augmented Lagrangian Filter

This section defines the basic concepts of our augmented Lagrangian filter algorithm. We start by defining a suitable filter and related step acceptance conditions. We then provide an outline of the algorithm that is described in more detail in the next section.

The new augmented Lagrangian filter is defined by using the residual of the first-order conditions (1.1), namely,

$$\omega(x,y) := \| \min\{x, \nabla_x \mathcal{L}_0(x,y)\}\| \tag{2.4a}$$

$$\eta(x) := \|c(x)\|. \tag{2.4b}$$

Augmented Lagrangian methods use forcing sequences ω_k, η_k to drive $(\omega(x,y), \eta(x))$ to zero. Here, we instead use the filter mechanism [33, 31] to achieve convergence to first-order points. A filter is formally defined as follows.

Definition 2.1. *A filter $\mathcal{F}$ is a list of pairs $(\eta_l, \omega_l) := \left(\eta(x^{(l)}), \omega(x^{(l)}, y^{(l)})\right)$ such that no pair dominates another pair. A point $(x^{(k)}, y^{(k)})$ is acceptable to the filter $\mathcal{F}$ if and only if*

$$\eta(x^{(k)}) \le \beta \eta_l \quad \text{or} \quad \omega(x^{(k)}, y^{(k)}) \le \omega_l - \gamma \eta(\hat{x}), \quad \forall l \in \mathcal{F}. \tag{2.5}$$

The fact that $(\eta(x), \omega(x,y)) \ge 0$ implies that we have an automatic upper bound on $\eta(x)$:

$$\eta(x) \le U := \max\left(\omega_{\min}/\gamma, \beta\eta_{\min}\right), \tag{2.6}$$

where $\omega_{\min}$ is the smallest first-order error of any filter entry, that is $\omega_{\min} := \min\{\omega_l : l \in \mathcal{F}\}$, and $\eta_{\min}$ is the corresponding value of η.

We note that our filter is based on the Lagrangian and not on the augmented Lagrangian. This choice is deliberate because one can show that the gradient of the Lagrangian after the first-order multiplier update, (1.3), equals the gradient of the augmented Lagrangian, namely,

$$\nabla_x \mathcal{L}_0(x^{(k)}, y^{(k)}) = \nabla_x \mathcal{L}_\rho(x^{(k)}, y^{(k-1)}). \tag{2.7}$$

Thus, by using the Lagrangian, we ensure that filter-acceptable points remain acceptable after the first-order multiplier update. Moreover, (2.7) shows that the filter acceptance can be readily checked during the minimization of the augmented Lagrangian, in which the multiplier is fixed and we iterate over x only.

The filter envelope defined by β and γ ensures that iterates cannot accumulate at points where $\eta > 0$, and it promotes convergence (see Lemma 4.4). A benefit of the filter approach is that we do

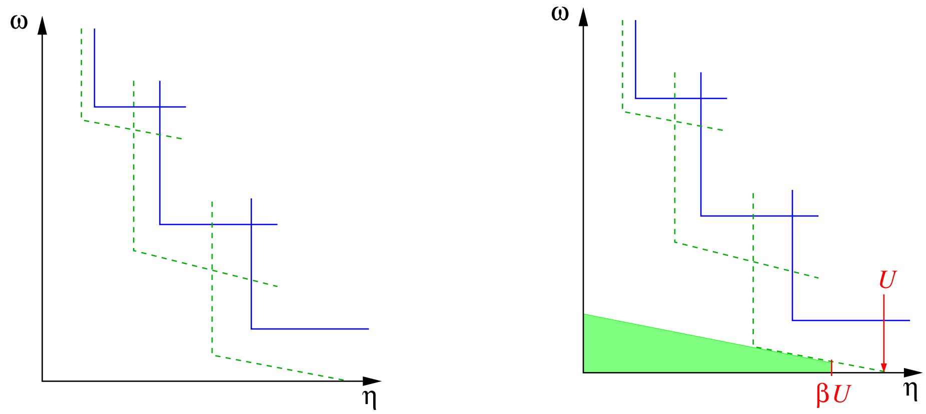

Figure 1: Left: Example of an augmented Lagrangian filter, $\mathcal{F}$, with three entries. The filter is in blue, the dashed green line shows the envelope, and the upper bound U is implied by the sloping envelope condition (2.5) and $\omega \geq 0$. Values above and to the right of the filter are not acceptable. Right: Illustration of proof of Lemma 4.2; the green area shows the set of acceptable filter entries.

> Given $x^{(0)}, y^{(0)}$, and ρ_0, set $\omega_0 = \omega(x^{(0)}, y^{(0)})$, $\eta_0 = \eta(x^{(0)})$, $k = 0$, and $\mathcal{F}_k = \{(\eta_k, \omega_k)\}$;
> **while** $(x^{(k)}, y^{(k)})$ *not optimal* **do**
>   Set $j = 0$ and initialize $\hat{x}^{(j)} = x^{(k)}$;
>   **repeat**
> $\hat{x}^{(j+1)} \leftarrow$ *approximate* $\underset{x \geq 0}{\operatorname{argmin}}\, \mathcal{L}_{\rho_k}(x, y^{(k)})$ from $\hat{x}^{(j)}$;
>   **if** restoration switching condition holds **then**
>   Increase penalty: $\rho_{k+1} = 2\rho_k$ & switch to restoration to find acceptable $(\hat{\eta}_j, \hat{\omega}_j)$;
>   Update multipliers: $\hat{y}^{(j+1)} = \hat{y}^{(j)} - \rho_k c(\hat{x}^{(j+1)})$, and set $j = j + 1$;
>   **until** $(\hat{\eta}_j, \hat{\omega}_j)$ *acceptable to* $\mathcal{F}_k$;
>   Set $(x^{(k+1)}, y^{(k+1)}) = (\hat{x}^{(j)}, \hat{y}^{(j)})$;
>   **if** $\eta_{k+1} > 0$ **then** add $(\eta_{k+1}, \omega_{k+1})$ to $\mathcal{F}_k$;
>   Set $k = k + 1$;

Algorithm 2: Outline of Augmented Lagrangian Filter Method

not need to assume that the multipliers remain bounded or that the iterates remain in a compact set. We outline the main algorithmic ideas in Algorithm 2; in the next section we provide a detailed description of the algorithm and its main components.

The algorithm has an inner iteration in which we minimize the augmented Lagrangian until a filter-acceptable point is found. Inner iterates are distinguished by a "hat", that is $\hat{x}^{(j)}$. Outer iterates are denoted as $x^{(k)}$. The algorithm has a restoration phase that is invoked if the iterates fail to make progress toward feasibility. The outline of our algorithm is deliberately vague to convey the main ideas. Details of the conditions of switching to restoration, termination of the inner iteration, and increasing the penalty parameter are developed in the next section. The algorithm allows for an optional penalty increase condition, which allows us to use a heuristic to estimate the penalty parameter. In addition, our algorithm has an optional second-order step on the set of active constraints. Our analysis, however, concentrates on the plain augmented Lagrangian approach.

3 Detailed Algorithm Statement

We start by describing the four algorithmic components not presented in our outline: the penalty update, the restoration switching condition, the termination condition for the inner iteration, and the second-order step. We then discuss the complete algorithm.

3.1 Optional Penalty Update Heuristic

Augmented Lagrangian methods can be shown to converge provided that the penalty parameter is sufficiently large and the multiplier estimate is sufficiently close to the optimal multiplier; see, for example, [9]. Here, we extend the penalty estimate from [37] to nonlinear functions. We stress that this step of the algorithm is not needed for global convergence, although it improves the behavior of our method. We will show in Section 4 that the penalty update is bounded, so that our heuristic does not harm the algorithm.

Consider the Hessian of the augmented Lagrangian:

$$\nabla^2 \mathcal{L}_\rho = \nabla^2 \mathcal{L}_0 + \rho \nabla c \nabla c^T + \rho \sum_{i=1}^{m} c_i \nabla^2 c_i. \tag{3.8}$$

Ideally, we would want to ensure $\nabla^2 \mathcal{L}_\rho \succeq 0$. Instead, we drop the $\nabla^2 c_i$ terms and consider

$$\nabla^2 \mathcal{L}_\rho \approx \nabla^2 \mathcal{L}_0 + \rho \nabla c \nabla c^T. \tag{3.9}$$

Now, we use the same ideas as in [37] to develop a penalty estimate that ensures that the augmented Lagrangian is positive definite on the null-space of the active inequality constraints. We define the active set and the inactive set as

$$\mathcal{A}^k := \mathcal{A}(x^{(k)}) := \left\{ i : x_i^{(k)} = 0 \right\} \text{ and } \mathcal{I}^k := \{1, 2, \ldots, n\} - \mathcal{A}^k, \tag{3.10}$$

respectively. One can show that a sufficient condition for $\nabla^2 \mathcal{L}_\rho \succeq 0$ on the active set is

$$\rho \geq \rho_{\min}(\mathcal{A}^k) := \frac{\max\{0, -\lambda_{\min}(\tilde{W}_k)\}}{\sigma_{\min}(\tilde{A}_k)^2}, \tag{3.11}$$

where $\tilde{W}_k$ is the reduced Hessian of the Lagrangian; $\tilde{A}_k$ is the reduced Jacobian of the equality constraints; and $\lambda_{\min}(\cdot)$ and $\sigma_{\min}(\cdot)$ denote the smallest eigenvalue and singular value, respectively. Computing (3.11) directly would be prohibitive for large-scale problems, and we use the following estimate instead:

$$\rho_{\min}(\mathcal{A}^k) := \max\left\{1,\ \frac{\|H_k\|_1}{\max\left\{\frac{1}{\sqrt{|\mathcal{I}^k|}}\|A_k\|_\infty\ ,\ \frac{1}{\sqrt{m}}\|A_k\|_1\right\}}\right\}, \tag{3.12}$$

where $|\mathcal{I}^k|$ is the number of free variables and m is the number of general equality constraints. If $\rho_k < \rho_{\min}(\mathcal{A}^k)$, then we increase the penalty parameter to $\rho_{k+1} = 2\rho_{\min}(\mathcal{A}^k)$. We could further improve this estimate by taking the terms $\rho c_i \nabla^2 c_i$ into account, which would change the numerator in (3.12).

3.2 Switching to Restoration Phase

In practice, many NLPs are not feasible, a situation that happens frequently especially when solving MINLPs. In this case, it is important that the NLP solver quickly and reliably finds a minimum of the constraint violation, $\eta(x)^2$. To converge quickly to such a point, we either have to drive the penalty parameter to infinity or switch to the minimization of $\eta(x)$. We prefer the latter approach because it provides an easy escape if we determine that the NLP appears to be feasible after all (unlike linear programming, there cannot be a phase I/phase II approach for NLPs).

We define a set of implementable criteria that force the algorithm to switch to the minimization of the constraint violation. Recall that the augmented Lagrangian filter implies an upper bound, $U = \max\{\omega_{\min}/\gamma, \beta\eta_{\min}\}$ from (2.6). Thus any inner iteration that generates

$$\hat{\eta}_{j+1} = \eta(\hat{x}^{(j+1)}) \geq \beta U \tag{3.13}$$

triggers the restoration phase. The second test to trigger the restoration phase is connected to the first-order conditions of a minimum of the constraint violation. We note that $\hat{x}^{(j)}$ is an approximate minimizer of the infeasibility, $\eta(x)^2$, if

$$\left\| \min\left(\nabla c(\hat{x}^{(j)}) c(\hat{x}^{(j)}), \hat{x}^{(j)}\right) \right\| \leq \epsilon \ \text{ and } \ \|c(\hat{x}^{(j)})\| \geq \beta\eta_{\min}, \tag{3.14}$$

where $\epsilon > 0$ is a constant and $\eta_{\min}$ is the smallest constraint violation of any filter entry, namely, $\eta_{\min} := \min\{\eta_l : l \in \mathcal{F}\} > 0$. In our algorithm, we switch to restoration if (3.13) or (3.14) hold.

Each time we do this switch,we increase the penalty parameter and start a new major iteration. The outcome of the restoration phase is either a (local) minimum of the infeasibility or a new point that is filter acceptable. The mechanism of the algorithm ensures that we can always find such a point because $\eta_{\min} > 0$, which is formalized in the following lemma.

Lemma 3.1. *Either the restoration phase converges to a minimum of the constraint violation, or it finds an acceptable $x^{(k+1)}$ in a finite number of steps.*

Proof. The restoration phase minimizes $\eta(x)^2$ and hence either converges to a local minimum of the constraint violation or generates a sequence of iterates $x^{(j)}$ with $\eta(x^{(j)}) \to 0$. Because $\eta_l > 0$ for all $l \in \mathcal{F}^k$, it follows that we must find a filter-acceptable point in a finite number of iterations in the latter case. □

3.3 Termination of Inner Minimization

The filter introduced in Section 2 ensures convergence only to *feasible* limit points; see Lemma 4.4. Thus, we need an additional condition that ensures that the limit points are also first-order optimal. We introduce a sufficient reduction condition that will ensure that the iterates are stationary. A sufficient reduction condition is more natural (since it corresponds to a Cauchy-type condition, which holds for all reasonable optimization routines) than is a condition that explicitly links the progress in first-order optimality ω_k to progress toward feasibility η_k.

In particular, we require that the following condition be satisfied on each inner iteration:

$$\Delta\mathcal{L}_{\rho_k}^{(j)} := \mathcal{L}_{\rho_k}(\hat{x}^{(j)}, y^{(k)}) - \mathcal{L}_{\rho_k}(\hat{x}^{(j+1)}, y^{(k)}) \geq \sigma\hat{\omega}_j, \tag{3.15}$$

where $\sigma > 0$ is a constant. This condition can be satisfied, for example, by requiring Cauchy decrease on the augmented Lagrangian for fixed ρ_k and $y^{(k)}$.

We will show that this sufficient reduction condition of the inner iterates in turn implies a sufficient reduction condition of the outer iterates as we approach feasibility; see (4.18). This outer sufficient reduction leads to a global convergence result. To the best of our knowledge, this is the first time that a more readily implementable sufficient reduction condition has been used in the context of augmented Lagrangians.

3.4 Optional Second-Order Step

Our algorithm allows for an additional second-order step. The idea is to use the approximate minimizers of the augmented Lagrangian, $x^{(k)}$, identify the active inequality constraints, $x_i^{(k)} = 0$, and then solve an equality-constrained QP (EQP) on those active constraints, similar to popular SLQP approaches. Given a set of active and inactive constraints, (3.10), our goal is to solve an EQP with $x_i^{(k)} = 0, \forall i \in \mathcal{A}(x^{(k)})$. Provided that the EQP is convex, its solution can be obtained by solving an augmented linear system. For a set of row and column indices $\mathcal{R}, \mathcal{C}$ and a matrix M, we define the submatrix $M_{\mathcal{R},\mathcal{C}}$ as the matrix with entries M_{ij} for all $(i,j) \in \mathcal{R} \times \mathcal{C}$ (we also use the Matlab notation ":" to indicate that all entries on a dimension are taken).

With this notation, the convex EQP is equivalent to the following augmented system,

$$\begin{bmatrix} H_{\mathcal{I},\mathcal{I}}^{(k+1)} & A_{:,\mathcal{I}}^{(k+1)} \\ A_{:,\mathcal{I}}^{(k+1)^T} & \end{bmatrix} \begin{pmatrix} \Delta x_{\mathcal{I}} \\ \Delta y \end{pmatrix} = \begin{pmatrix} -\nabla f_{\mathcal{I}}^{(k+1)} \\ -c(x^{(k+1)}) \end{pmatrix}, \tag{3.16}$$

where $\mathcal{A} = \mathcal{A}^{(k+1)}, \mathcal{I} = \mathcal{I}^{(k+1)}$, that is $\Delta x_{\mathcal{A}} = 0$. We note that, in general, we cannot expect that the solution $x^{(k+1)} + \Delta x_{\mathcal{I}}$ is acceptable to the filter (or even can be a descent direction). Hence, we add a back-tracking line search to our algorithm to find an acceptable point. We note that because $(x^{(k+1)}, y^{(k+1)})$ is known to be acceptable, we can terminate the line search if the step size is less that some $\alpha_{\min} > 0$ and instead accept $(x^{(k+1)}, y^{(k+1)})$.

Given $x^{(0)}, y^{(0)}$ and ρ_0,set $\omega_0 = \omega(x^{(0)}, y^{(0)})$, $\eta_0 = \eta(x^{(0)})$, $k = 0$, and $\mathcal{F}_k = \{(\eta_k, \omega_k)\}$;
while $(x^{(k)}, y^{(k)})$ *not optimal* **do**
  Set $j = 0$, *RestFlag* = `false`, and initialize $\hat{x}^{(j)} = x^{(k)}$;
  repeat
    Approximately minimize the augmented Lagrangian for $\hat{x}^{(j+1)}$ starting at $\hat{x}^{(j)}$:

$$\underset{x \geq 0}{\text{minimize}} \ \mathcal{L}_{\rho_k}(x, y^{(k)}) = f(x) - {y^{(k)}}^T c(x) + \tfrac{1}{2}\rho_k \|c(x)\|^2$$

    such that the sufficient reduction condition (3.15) holds.
    if *restoration switching condition (3.13) or (3.14) hold* **then**
      Set *RestFlag* = `true`, and increase penalty parameter: $\rho_{k+1} = 2\rho_k$;
      Switch to restoration phase to find $(\hat{x}^{(j+1)}, \hat{y}^{(j+1)})$ acceptable to $\mathcal{F}$;
      ... or find nonzero minimizer of $\|c(x)\|$ subject to $x \geq 0$;
    else
      Provisionally update multipliers: $\hat{y}^{(j+1)} = y^{(k)} - \rho_k c(\hat{x}^{(j+1)})$;
      Compute $\hat{\omega}_{j+1} = \omega(\hat{x}^{(j+1)}, \hat{y}^{(j+1)})$ and $\hat{\eta}_{j+1} = \eta(\hat{x}^{(j+1)})$.
    Set $j = j + 1$;
  until $(\hat{\eta}_j, \hat{\omega}_j)$ *acceptable to* $\mathcal{F}_k$;
  Set $(\hat{x}^{(k+1)}, \hat{y}^{(k+1)}) = (\hat{x}^{(j)}, \hat{y}^{(j)})$ and solve EQP (3.16) for $(\Delta x^{(k+1)}, \Delta y^{(k+1)})$;
  Line-search: Find $\alpha_k \in \{0\} \cup [\alpha_{\min}, 1]$ such that

$$(x^{(k+1)}, y^{(k+1)}) = (\hat{x}^{(k+1)}, \hat{y}^{(k+1)}) + \alpha_{k+1}(\Delta x^{(k+1)}, \Delta y^{(k+1)}) \quad \text{acceptable to } \mathcal{F}_k$$

  Compute $\omega_{k+1} = \omega(x^{(k+1)}, y^{(k+1)})$, $\eta_{k+1} = \eta(x^{(k+1)})$;
  if $\eta_{k+1} > 0$ **then** add $(\eta_{k+1}, \omega_{k+1})$ to filter: $\mathcal{F}_{k+1} = \mathcal{F}_k \cup \{(\eta_{k+1}, \omega_{k+1})\}$;
  if *(not RestFlag) and ($\rho_k < \rho_{\min}(\mathcal{A}^k)$ in (3.12))* **then**
    Increase penalty $\rho_{k+1} = 2\rho_{\min}(\mathcal{A}^k)$;
  else
    Leave penalty parameter unchanged: $\rho_{k+1} = \rho_k$;
  Set $k = k + 1$;

Algorithm 3: Augmented Lagrangian Filter Method.

3.5 Complete Algorithm

The complete algorithm is stated next. It has an inner loop that minimizes the augmented Lagrangian for a fixed penalty parameter and multiplier until a filter-acceptable point is found. Quantities associated with the inner loop are indexed by j and have a "hat." The outer loop corresponds to major iterates and may update the penalty parameter. The inner iteration also terminates when we switch to the restoration phase. Any method for minimizing $\eta(x)^2$ (or any measure of constraint infeasibility) can be used in this phase. Note that the penalty parameter is also increased every time we switch to the restoration phase, although we could use a more sophisticated penalty update in that case, too. A complete description of the algorithm is given in Algorithm 3.

We note, that Algorithm 3 uses a flag, *RestFlag*, to indicate whether we entered the restoration phase or not. If we enter the restoration phase, then we increase the penalty parameter in the outer loop iterates, k. Two possible outcomes for the restoration phase exist: either we find a nonzero (local) minimizer of the constraint violation indicating that problem (NLP) is infeasible, or we find a filter-acceptable point and exit the inner iteration. In the latter case, *RestFlag* is true and ensures that we do not update the penalty parameter using (3.12), which does not make sense in this situation.

4 Convergence Proof

This section establishes a global convergence result for Algorithm 3, without the second-order step for the sake of simplicity. We make the following assumptions throughout this section.

Assumption 4.1. *Consider problem (NLP), and assume that the following hold:*

A1 *The problem functions f, c are twice continuously differentiable.*

A2 *The constraint norm satisfies $\|c(x)\| \to \infty$ as $\|x\| \to \infty$.*

Assumption **A1** is standard. Assumption **A2** implies that our iterates remain in a compact set (see Lemma 4.1). This assumption could be replaced by an assumption that we optimize over finite bounds $l \le x \le u$. Both assumptions together imply that $f(x)$ and $c(x)$ and their derivatives are bounded for all iterates.

Algorithm 3 has three distinct outcomes.

1. There exists an infinite sequence of restoration phase iterates, $x^{(k_l)}$, indexed by $\mathcal{R} := \{k_1, k_2, \ldots\}$, whose limit point $x^* := \lim x^{(k_l)}$ is a minimum of the constraint violation $\eta(x^*) > 0$.

2. There exists an infinite sequence of successful major iterates, $x^{(k_l)}$, indexed by $\mathcal{S} := \{k_1, k_2, \ldots\}$, and the linear independence constraint qualification fails to hold at the limit $x^* := \lim x^{(k_l)}$, which is a Fritz-John point of (NLP).

3. There exists an infinite sequence of successful major iterates, $x^{(k_l)}$, indexed by $\mathcal{S} := \{k_1, k_2, \ldots\}$, and the linear independence constraint qualification holds at the limit $x^* := \lim x^{(k_l)}$, which is a Karush-Kuhn-Tucker point of (NLP).

Outcomes 1 and 3 are normal outcomes of NLP solvers in the sense that we cannot exclude the possibility that (NLP) is infeasible without making restrictive assumptions such as Slater's constraint qualification. Outcome 2 corresponds to the situation where a constraint qualification fails to hold at a limit point.

Outline of Convergence Proof. We start by showing that all iterates remain in a compact set. Next, we show that the algorithm is well defined by proving that the inner iteration is finite, which implies the existence of an infinite sequence of outer iterates $x^{(k)}$, unless the restoration phase fails or the algorithm converges finitely. We then show that the limit points are feasible and stationary. Finally, we show that the penalty estimate (3.12) is bounded.

We first show that all iterates remain in a compact set.

Lemma 4.1. *All major and minor iterates, $x^{(k)}$ and $\hat{x}^{(j)}$, remain in a compact set, C.*

Proof. The upper bound, U, on $\eta(x)$ implies that $\|c(x^{(k)})\| \leq U$ or all k. The switching condition (3.13) implies that $\|c(\hat{x}^{(j)})\| \leq U$ or all j. The feasibility restoration minimizes $\eta(x)$, implying that all its iterates in turn satisfy $\|c(x^{(k)})\| \leq U$. Assumptions **A1** and **A2** now imply that the iterates remain in a bounded set, C. □

The next lemma shows that the mechanism of the filter ensures that there exists a neighborhood of the origin in the filter that does not contain any filter points, as illustrated in Figure 1 (right).

Lemma 4.2. *There exists a neighborhood of $(\eta, \omega) = (0, 0)$ that does not contain any filter entries.*

Proof. The mechanism of the algorithm ensures that $\eta_l > 0, \ \forall l \in \mathcal{F}_k$. First, assume that $\omega_{\min} := \min\{\omega_l : l \in \mathcal{F}\} > 0$. Then it follows that there exist no filter entries in the quadrilateral bounded by $(0, 0), (0, \omega_{\min}), (\beta\eta_{\min}, \omega_{\min} - \gamma\beta\eta_{\min}), (\beta\eta_{\min}, 0)$, illustrated by the green area in Figure 1 (right). Next, if there exists a filter entry with $\omega_l = 0$, then define $\omega_{\min} := \min\{\omega_l > 0 : l \in \mathcal{F}\} > 0$, and observe that the quadrilateral $(0, 0), (0, \omega_{\min}), (\beta\eta_{\min}, \omega_{\min}), (\beta\eta_{\min}, 0)$ contains no filter entries. In both cases, the area is nonempty, thus proving that there exists a neighborhood of $(0, 0)$ with filter acceptable points. □

Next, we show that the inner iteration is finite and the algorithm is well defined.

Lemma 4.3. *The inner iteration is finite.*

Proof. If the inner iteration finitely terminates with a filter acceptable point or switches to the restoration phase, then there is nothing to prove. Otherwise, there exists an infinite sequence of inner iterates $\hat{x}^{(j)}$ with $\hat{\eta}_j \leq \beta U$. Lemma 4.1 implies that this sequence has a limit point, $x^* = \lim \hat{x}^{(j)}$. We note that the penalty parameter and the multiplier are fixed during the inner iteration, and we consider the sequence $\mathcal{L}_\rho(\hat{x}^{(j)}, y)$ for fixed $\rho = \rho_k$ and $y = y^{(k)}$. The sufficient reduction condition (3.15) implies that

$$\Delta\mathcal{L}_\rho^{(j)} := \mathcal{L}_\rho(\hat{x}^{(j)}, y) - \mathcal{L}_\rho(\hat{x}^{(j+1)}, y) \geq \sigma\hat{\omega}_j.$$

If $\hat{\omega}_j \to 0$, we would exit the inner iteration according to Lemma 4.2, because $\hat{\eta}_j \leq \beta U$. We now assume that $\hat{\omega}_j \geq \bar{\omega} > 0$ and seek a contradiction. However, $\hat{\omega}_j \geq \bar{\omega} > 0$ implies that $\mathcal{L}_\rho(\hat{x}^{(j)}, y)$ is unbounded below, which contradicts the fact that $f(x), \|c(x)\|$ are bounded by Assumption

A1 and Lemma 4.1. Thus, it follows that $\hat{\omega}_j \to 0$, and we exit the inner iteration according to Lemma 4.2. □

The next lemma shows that all limit points of the outer iteration are feasible.

Lemma 4.4. *Assume that there exist an infinite number of outer iterations. Then it follows that* $\eta(x^{(k)}) \to 0$.

Proof. Every outer iteration for which $\eta_k > 0$ adds an entry to the filter. The proof follows directly from [22, Lemma 1]. □

The next two lemmas show that the first-order error, ω_k, also converges to zero. We split the argument into two parts depending on whether the penalty parameter remains bounded or not.

Lemma 4.5. *Assume that the penalty parameter is bounded,* $\rho_k \leq \bar{\rho} < \infty$, *and consider an infinite sequence of outer iterations. Then it follows that* $\omega(x^{(k)}) \to 0$.

Proof. Because the penalty parameter is bounded, it is updated only finitely often. Hence, we consider the tail of the sequence $x^{(k)}$ for which the penalty parameter has settled down, namely, $\rho_k = \bar{\rho}$. We assume that $\omega_k \geq \bar{\omega} > 0$ and seek a contradiction. The sufficient reduction condition of the inner iteration (3.15) implies that

$$\Delta \mathcal{L}^{\text{in}}_{\bar{\rho},k} := \mathcal{L}_{\bar{\rho}}(x^{(k)}, y^{(k)}) - \mathcal{L}_{\bar{\rho}}(x^{(k+1)}, y^{(k)}) \geq \sigma \omega_k \geq \sigma \bar{\omega} > 0. \tag{4.17}$$

We now show that this "inner" sufficient reduction (for fixed $y^{(k)}$) implies an "outer" sufficient reduction. We combine (4.17) with the first-order multiplier update (1.3) and obtain

$$\Delta \mathcal{L}^{\text{out}}_{\bar{\rho},k} := \mathcal{L}_{\bar{\rho}}(x^{(k)}, y^{(k)}) - \mathcal{L}_{\bar{\rho}}(x^{(k+1)}, y^{(k+1)}) \; = \; \Delta \mathcal{L}^{\text{in}}_{\bar{\rho},k} - \bar{\rho} \|c(x^{(k+1)})\|_2^2 \; \geq \; \sigma \bar{\omega} - \bar{\rho} \eta_{k+1}^2. \tag{4.18}$$

Lemma 4.4 implies that $\eta_k \to 0$; hence, as soon as $\eta_{k+1} \leq \sigma \frac{\bar{\omega}}{2\bar{\rho}}$ for all k sufficiently large, we obtain the following sufficient reduction condition for the outer iteration:

$$\Delta \mathcal{L}^{\text{out}}_{\bar{\rho},k} \geq \sigma \frac{\bar{\omega}}{2},$$

for all k sufficiently large. Thus, if $\omega_k \geq \bar{\omega} > 0$ is bounded away from zero, then it follows that the augmented Lagrangian must be unbounded below. However, because all $x^{(k)} \in C$ remain in a compact set, it follows from Assumption **A1** that $f(x)$ and $\|c(x)\|$ are bounded below and hence that $\mathcal{L}_{\bar{\rho}}(x, y)$ can be unbounded below only if $-y^T c(x)$ is unbounded below.

We now show that $c(x^{(k)})^T y^{(k)} \leq M$ for all major iterates. The first-order multiplier update implies that $y^{(k)} = y^{(0)} - \bar{\rho} \sum c^{(l)}$ and hence that

$$\begin{aligned} c(x^{(k)})^T y^{(k)} &= \left(y^{(0)} - \bar{\rho} \sum_{l=1}^{k} c^{(l)} \right)^T c^{(k)} \\ &\leq \left(\|y^{(0)}\| + \bar{\rho} \sum_{l=1}^{k} \|c^{(l)}\| \right) \|c^{(k)}\| \\ &= \left(y_0 + \bar{\rho} \sum_{l=1}^{k} \eta_l \right) \eta_k \;, \end{aligned} \tag{4.19}$$

where $y_0 = \|y^{(0)}\|$, and we have assumed without loss of generality that $\bar\rho$ is fixed for the whole sequence. Now define $E_k := \max_{l\geq k} \eta_l$, and observe that $E_k \to 0$ from Lemma 4.4. The definition of the filter then implies that $E_{k+1} \leq \beta E_k$, and we obtain from (4.19) that

$$c(x^{(k)})^T y^{(k)} \leq \left(y_0 + \bar\rho \sum_{l=1}^{k} E_l \right) E_k = \left(y_0 + \bar\rho \sum_{l=1}^{k} \beta^l E_0 \right) \beta^k E_0 = \left(y_0 + \bar\rho\beta \frac{1-\beta^k}{1-\beta} E_0 \right) \beta^k E_0 < M.$$

Moreover, because $E_0 < \infty$, $\bar\rho < \infty$ and $0 < \beta < 1$ it follows that this expression is uniformly bounded as $k \to \infty$. Hence it follows that $c(x^{(k)})^T y^{(k)} \leq M$ for all k, and $\mathcal{L}_{\bar\rho}(x,y)$ must be bounded below, which contradicts the assumption that $\omega_k \geq \bar\omega > 0$ is bounded away from zero. Hence, it follows that $\omega_k \to 0$. □

We now consider the case where $\rho_k \to \infty$. In this case, we must assume that the linear independence constraint qualification (LICQ) holds at every limit point. If LICQ fails at a limit point, then we cannot guarantee that the limit is a KKT point; it may be a Fritz-John point instead. The following lemma formalizes this result.

Lemma 4.6. *Consider the situation where $\rho_k \to \infty$. Then any limit point, $x^{(k)} \to x^*$, is a Fritz-John point. If in addition LICQ holds at x^*, then it is a KKT point, and $\omega_k \to 0$.*

Proof. Lemma 4.4 ensures that the limit point is feasible. Hence, it is trivially a Fritz-John point. Now assume that LICQ holds at x^*. We use standard augmented Lagrangian theory to show that this limit point also satisfies $\omega(x^*) = 0$. Following Theorem 2.5 of [36], we need to show that for all restoration iterations, $\mathcal{R} := \{k_1, k_2, k_3, \ldots\}$ on which we increase the penalty parameter, it follows that the quantity

$$\sum_{l=1}^{\infty} \eta_{k_\nu + l}$$

remains bounded as $\nu \to \infty$. Similar to the proof of Lemma 4.5, the filter acceptance ensures that $\eta_{k_\nu+l} \leq \beta^l \eta_{k_\nu}$, which gives the desired result. Thus, we can invoke Theorem 2.5 of [36], which shows that the limit point is a KKT point. □

The preceding lemmas are summarized in the following result.

Theorem 4.1. *Assume that Assumptions* **A1** *and* **A2** *hold. Then either Algorithm 3 terminates after a finite number of iterations at a KKT point, that is, for some finite k, $x^{(k)}$ is a first-order stationary point with $\eta(x^{(k)}) = 0$ and $\omega(x^{(k)}) = 0$, or there exists an infinite sequence of iterates $x^{(k)}$ and any limit point, $x^{(k)} \to x^*$, satisfies one of the following.*

1. *The penalty parameter is updated finitely often, and x^* is a KKT point.*
2. *There exists an infinite sequence of restoration steps on which the penalty parameter is updated. If x^* satisfies an LICQ, then it is a KKT point. Otherwise, it is a Fritz-John point.*
3. *The restoration phase converges to a minimum of the infeasibility.*

Remark 4.1. *We seem to be able to show that the limit point is a KKT point without assuming a constraint qualification, as long as the penalty parameter remains bounded. On the other hand, without a constraint qualification, we would expect the penalty parameter to be unbounded. It would be interesting to test these results in the context of mathematical program with equilibrium constraints (MPECs). We suspect that MPECs that satisfy a strong-stationarity condition would have a bounded penalty but that those that do not have strongly stationary points would require for the penalty to be unbounded.*

Remark 4.2. *The careful reader may wonder whether Algorithm 3 can cycle, because we do not add iterates to the filter for which $\eta_k = 0$. We can show, however, that this situation cannot happen. If we have an infinite sequence of iterates for which $\eta_k = 0$, then the sufficient reduction condition (3.15) implies that we must converge to a stationary point, similar to the arguments in Lemma 4.5. If we have a sequence that alternates between iterates for which $\eta_k = 0$ and iterates for which $\eta_k > 0$, then we can never revisit any iterates for which $\eta_k > 0$ because those iterates have been added to the filter. By Lemma 4.4, any limit point is feasible. Thus, if LICQ holds, then the limit is a KKT point; otherwise, it may be a Fritz-John point. We note that these conclusions are consistent with Theorem 4.1.*

5 Conclusions

We have introduced a new filter method for augmented Lagrangian methods that removes the need for the traditional forcing sequences. We prove convergence of our method to first-order stationary points of nonlinear programs under mild conditions, and we present a heuristic for adjusting the penalty parameter based on matrix-norm estimates. We show that second-order steps are readily integrated into our method to accelerate local convergence.

The proposed method is closely related to Newton's method in the case of equality constraints only. If no inequality constraints exist, that is, $x \in \mathbb{R}^n$, then our algorithm reverts to standard Newton/SQP for equality constrained optimization with a line-search safeguard. In this case, we only need to compute the Cauchy point to the augmented Lagrangian step that is acceptable to the filter. Of course, a more direct implementation would be preferable.

Our proof leaves open a number of questions. First, we did not show second-order convergence, but we believe that such a proof follows directly if we use second-order correction steps as suggested in [62] or if we employ a local non-monotone filter similar to [59]. A second open question is how the proposed method performs in practice. We have some promising experience with large-scale quadratic programs [37], and we are working on an implementation for nonlinear programs.

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under Contract DE-AC02-06CH11357. This work was also supported by the U.S. Department of Energy through grant DE-FG02-05ER25694.

References

[1] Kumar Abhishek, Sven Leyffer, and Jeffrey T. Linderoth. FilMINT: An outer-approximation-based solver for nonlinear mixed integer programs. *INFORMS Journal on Computing*, 22:555–567, 2010.

[2] A. Altay-Salih, M.C. Pinar, and S. Leyffer. Constrained nonlinear programming for volatility estimation with GARCH models. *SIAM Review*, 45(3):485–503, 2003.

[3] J.J. Arrieta-Camacho, L.T. Biegler, and D. Subramanian. NMPC-based real-time coordination of multiple aircraft. In R.F. Allgower and L.T. Biegler, editors, *Assessment and Future Directions of Nonlinear Model Predictive Control*, pages 629–639. Springer, Heidelberg, 2007.

[4] Guillermo Bautista, Miguel F. Anjos, and Anthony Vannelli. Formulation of oligopolistic competition in AC power networks: An NLP approach. *IEEE Transaction on Power Systems*, 22(1):105–115, 2007.

[5] Pietro Belotti, Christian Kirches, Sven Leyffer, Jeff Linderoth, James Luedtke, and Ashutosh Mahajan. Mixed-integer nonlinear optimization. *Acta Numerica*, 22:1–131, 5 2013.

[6] Hande Benson and David Shanno. An exact primal-dual penalty method approach to warmstarting interior-point methods for linear programming. *Computational Optimization and Applications*, 38:371–399, 2007. 10.1007/s10589-007-9048-6.

[7] Hande Y. Benson and David F. Shanno. Interior-point methods for nonconvex nonlinear programming: regularization and warmstarts. *Computational Optimization and Applications*, 40(2):143–189, 2008.

[8] H.Y. Benson, D.F. Shanno, and R.J. Vanderbei. Interior-point methods for nonconvex nonlinear programming: filter-methods and merit functions. *Computational Optimization and Applications*, 23(2):257–272, 2002.

[9] D.P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Athena Scientific, New York, 1982.

[10] J.T. Betts. *Practical Methods for Optimal Control Using Nonlinear Programming*. Advances in Design and Control. SIAM, Philadelphia, 2001.

[11] L.T. Biegler and I.E. Grossmann. Challenges and research issues for product and process design optimization. In C.A. Floudas and R. Agrawal, editors, *Proceedings of Foundations of Computer Aided Process Design*. CACHE Corporation, Austin, TX, 2004.

[12] L.T. Biegler and I.E. Grossmann. Part I: Retrospective on optimization. *Computers and Chemical Engineering*, 28(8):1169–1192, 2004.

[13] P. Bonami, L.T. Biegler, A.R. Conn, G. Cornuejols, I.E. Grossmann, C.D. Laird, J. Lee, A. Lodi, F. Margot, N. Sawaya, and A. Wähter. An algorithmic framework for convex mixed integer nonlinear programs. Research Report RC 23771, IBM T. J. Watson Research Center, Yorktown, NY, 2005.

[14] Pierre Bonami, Jon Lee, Sven Leyffer, and Andreas Waechter. More branch-and-bound experiments in convex nonlinear integer programming. Technical Report ANL/MCS-P1949-0911, Argonne National Laboratory, Mathematics and Computer Science Division, Lemont, IL, September 2011.

[15] J.F. Bonnans and J. André. Optimal structure of gas transmission trunklines. Technical Report 6791, INRIA Saclay, 91893 ORSAY Cedex, France, 2009.

[16] B. Bonnard, L. Faubourg, G. Launay, and E. Trélat. Optimal control with state constraints and the space shuttle re-entry problem. *Journal of Dynamical and Control Systems*, 9(2):155–199, 2003.

[17] A. Borghetti, A. Frangioni, F. Lacalandra, and C.A. Nucci. Lagrangian heuristics based on disaggregated bundle methods for hydrothermal unit commitment. *IEEE Transactions on Power Systems*, 18:1–10, 2003.

[18] R.H. Byrd, N.I.M. Gould, J. Nocedal, and R.A. Waltz. An algorithm for nonlinear optimization using linear programming and equality constrained subproblems. *Mathematical Programming, Series B*, 100(1):27–48, 2004.

[19] R.H. Byrd, M.E. Hribar, and J. Nocedal. An interior point algorithm for large scale nonlinear programming. *SIAM J. Optimization*, 9(4):877–900, 1999.

[20] R.H. Byrd, J. Nocedal, and R.A. Waltz. KNITRO: An integrated package for nonlinear optimization. In G. di Pillo and M. Roma, editors, *Large-Scale Nonlinear Optimization*, pages 35–59. Springer-Verlag, New York, 2006.

[21] J. Castro and J.A. Gonzalez. A nonlinear optimization package for long-term hydrothermal coordination. *European Journal of Operational Research*, 154(1):641–658, 2004.

[22] C.M. Chin and R. Fletcher. On the global convergence of an SLP-filter algorithm that takes EQP steps. *Mathematical Programming*, 96(1):161–177, 2003.

[23] T.F. Coleman, Y. Li, and A. Verman. Reconstructing the unknown volatility function. *Journal of Computational Finance*, 2(3):77–102, 1999.

[24] A.R. Conn, N.I.M. Gould, and Ph.L. Toint. *LANCELOT: A Fortran package for large-scale nonlinear optimization (Release A)*. Springer Verlag, Heidelberg, New York, 1992.

[25] A.R. Conn, N.I.M. Gould, and Ph.L. Toint. *Trust-Region Methods*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2000.

[26] G. Cornuejols and R. Tütüncü. *Optimization Methods in Finance*. Cambridge University Press, 2007.

[27] V. Donde, V. Lopez, B. Lesieutre, A. Pinar, C. Yang, and J. Meza. Identification of severe multiple contingencies in electric power networks. In *Proceedings 37th North American Power Symposium*, 2005. LBNL-57994.

[28] Klaus Ehrhardt and Marc C. Steinbach. Nonlinear optimization in gas networks. In H.G. Bock, E. Kostina, H.X. Phu, and R. Ranacher, editors, *Modeling, Simulation and Optimization of Complex Processes*, pages 139–148. Springer, Berlin, 2005.

[29] R. Fletcher and S. Leyffer. Solving mixed integer nonlinear programs by outer approximation. *Mathematical Programming*, 66:327–349, 1994.

[30] R. Fletcher and S. Leyffer. User manual for filterSQP. Numerical Analysis Report NA/181, University of Dundee, April 1998.

[31] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming*, 91:239–270, 2002.

[32] R. Fletcher and S. Leyffer. Filter-type algorithms for solving systems of algebraic equations and inequalities. In G. di Pillo and A. Murli, editors, *High Performance Algorithms and Software for Nonlinear Optimization*, pages 259–278. Kluwer, Dordrecht, 2003.

[33] R. Fletcher, S. Leyffer, and Ph.L. Toint. On the global convergence of a filter-SQP algorithm. *SIAM J. Optimization*, 13(1):44–59, 2002.

[34] Fletcher, R. and Sainz de la Maza, E. Nonlinear Programming and nonsmooth Optimization by successive Linear Programming. *Mathematical Programming*, 43:235–256, 1989.

[35] A. Forsgren, P.E. Gill, and M.H. Wright. Interior methods for nonlinear optimization. *SIAM Review*, 4(4):525–597, 2002.

[36] Michael P. Friedlander. *A Globally Convergent Linearly Constrained Lagrangian Method for Nonlinear Optimization*. Ph.D. thesis, Stanford University, August 2002.

[37] M.P. Friedlander and S. Leyffer. Global and finite termination of a two-phase augmented lagrangian filter method for general quadratic programs. *SIAM J. Scientific Computing*, 30(4):1706–1729, 2008.

[38] L.E. Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4), 2003.

[39] P.E. Gill, W. Murray, and M.A. Saunders. User's guide for SQOPT 5.3: A Fortran package for large-scale linear and quadratic programming. Tech. Rep. NA 97-4, Department of Mathematics, University of California, San Diego, 1997.

[40] P.E. Gill, W. Murray, and M.A. Saunders. SNOPT: An SQP algorithm for large–scale constrained optimization. *SIAM Journal on Optimization*, 12(4):979–1006, 2002.

[41] Jacek Gondzio and Andreas Grothey. Reoptimization with the primal-dual interior point method. *SIAM Journal on Optimization*, 13(3):842–864, 2002.

[42] J.-P. Goux and S. Leyffer. Solving large MINLPs on computational grids. *Optimization and Engineering*, 3:327–346, 2002.

[43] Q. Huangfu and J.A.J. Hall. Novel update techniques for the revised simplex method. Technical Report ERGO-13-001, The School of Mathematics, University of Edinburgh, Edinburgh, UK, 2013.

[44] Y. Kawayir, C. Laird, and A. Wächter. Introduction to Ipopt: A tutorial for downloading, installing, and using Ipopt. Manual, IBM T.J. Watson Research Center, Yorktown Heights, NY, 2009.

[45] H. Konno and H. Yamazaki. Mean-absolute deviation portfolio optimization model and its application to Tokyo stock market. *Management Science*, 37:519–531, 1991.

[46] S. Leyffer. Integrating SQP and branch-and-bound for mixed integer nonlinear programming. *Computational Optimization & Applications*, 18:295–309, 2001.

[47] S. Leyffer, G. Lopez-Calva, and J. Nocedal. Interior methods for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 17(1):52–77, 2006.

[48] S. Leyffer and T.S. Munson. A globally convergent filter method for MPECs. Preprint ANL/MCS-P1457-0907, Argonne National Laboratory, Mathematics and Computer Science Division, September 2007.

[49] Miles Lubin, J.A.J. Hall, Cosmin G. Petra, and Mihai Anitescu. Parallel distributed-memory simplex for large-scale stochastic LP problems. *Computational Optimization and Applications*, 55(3):571–596, 2013.

[50] A. Martin, M. Möller, and S. Moritz. Mixed integer models for the stationary case of gas network optimization. *Mathematical Programming*, 105:563–582, 2006.

[51] J.A. Momoh, R.J. Koessler, M.S. Bond, B. Stott, D. Sun, A. Papalexopoulos, and P. Ristanovic. Challenges to optimal power flow. *IEEE Transaction on Power Systems*, 12:444–455, 1997.

[52] Bruce A. Murtagh and Michael A. Saunders. A projected lagrangian algorithm and its implementation for sparse nonlinear constraints. In A.G. Buckley and J.-L. Goffin, editors, *Algorithms for Constrained Minimization of Smooth Nonlinear Functions*, volume 16 of *Mathematical Programming Studies*, pages 84–117. Springer Berlin Heidelberg, 1982.

[53] P. Penfield, R. Spence, and S. Duinker. *Tellegen's theorem and electrical networks*. M.I.T. Press, Cambridge, London, 1970.

[54] M.J.D. Powell. Algorithms for nonlinear constraints that use Lagrangian functions. *Mathematical Programming*, 14(1):224–248, 1978.

[55] Gad Rabinowitz, Abraham Mehrez, and Gideon Oron. A nonlinear optimization model of water allocation for hydroelectric energy production and irrigation. *Management Science*, 34(8):973–990, 1988.

[56] A. Raghunathan and L. T. Biegler. An interior point method for mathematical programs with complementarity constraints (MPCCs). *SIAM Journal on Optimization*, 15(3):720–750, 2005.

[57] S.M. Robinson. A quadratically-convergent algorithm for general nonlinear programming problems. *Mathematical Programming*, 3(1):145–156, 1972.

[58] G.B. Sheble and G.N. Fahd. Unit commitment literature synopsis. *IEEE Transactions on Power Systems*, 9:128–135, 1994.

[59] Chungen Shen, Sven Leyffer, and Roger Fletcher. A nonmonotone filter method for nonlinear optimization. *Computational Optimization and Applications*, 52(3):583–607, 2012.

[60] E. Smith and J.A.J. Hall. A high performance primal revised simplex solver for row-linked block angular linear programming problems. Technical Report ERGO-12-003, The School of Mathematics, University of Edinburgh, Edinburgh, UK, 2012.

[61] R.J. Vanderbei and D.F. Shanno. An interior point algorithm for nonconvex nonlinear programming. *COAP*, 13:231–252, 1999.

[62] A. Wächter and L.T. Biegler. Line search filter methods for nonlinear programming: Local convergence. *SIAM Journal on Optimization*, 16(1):32–48, 2005.

[63] A. Wächter and L.T. Biegler. Line search filter methods for nonlinear programming: Motivation and global convergence. *SIAM Journal on Optimization*, 16(1):1–31, 2005.

[64] R.S. Womersley and K. Lau. Portfolio optimisation problems. In R.L. May and A.K. Easton, editors, *Computational Techniques and Applications: CTAC95*, pages 795–802. World Scientific Publishing Co., 1996.