An interior point filter line search method: main convergence results

M. Fernanda P. Costa* and Edite M.G.P. Fernandes†

*Departamento de Matemática para a Ciência e Tecnologia
†Departamento de Produção e Sistemas
Universidade do Minho, 4710-057 Braga, Portugal

Abstract. Here we present the main convergence results of an interior point filter line search method for nonlinear programming. Each entry in the filter relies on three components, feasibility, centrality and optimality, that are present in the optimality conditions. Preliminary numerical experiments show that the method is promising.

Keywords: nonlinear programming, interior point method, filter method
PACS: 02.60.Pn

INTRODUCTION

In this paper we use the filter technique of Fletcher and Leyffer [1] to globalize a primal-dual interior point method for nonlinear optimization. Filter techniques incorporate the concept of non-dominance to build a filter that is able to reject poor trial iterates and enforce global convergence from arbitrary starting points. For easy of presentation, we consider the formulation of a constrained nonlinear optimization problem as follows:

$$\min_{x \in \mathbb{R}^n} F(x)$$
$$\text{s.t. } h(x) \geq 0$$

where $$h_i : \mathbb{R}^n \to \mathbb{R}$$ for $$i = 1, \ldots, m$$ and $$F : \mathbb{R}^n \to \mathbb{R}$$ are nonlinear and twice continuously differentiable functions.

THE INTERIOR POINT FILTER LINE SEARCH METHOD

This section briefly describes an infeasible primal-dual interior point method for solving (1). We refer to [2] for details. Adding nonnegative slack variables $$w$$ to transform the inequality constraints into equality constraints, and incorporating the constraints $$w \geq 0$$ in logarithmic barrier terms in the objective function, (1) is transformed into

$$\min_{x \in \mathbb{R}^n, w \in \mathbb{R}^m} \phi_\mu(x, w) \equiv F(x) - \mu \sum_{i=1}^{m} \log(w_i)$$
$$\text{s.t. } h(x) - w = 0,$$

where $$\phi_\mu(x, w)$$ is the barrier function and $$\mu$$ is a positive barrier parameter. The first-order KKT conditions for a minimum of (2) are

$$\nabla F(x) - A^T y = 0$$
$$-\mu W^{-1} e + y = 0$$
$$h(x) - w = 0$$

where $$y$$ is the dual variable. $$\nabla F$$ is the gradient vector of $$F$$, $$A$$ is the Jacobian matrix of the constraints $$h$$, $$W = \text{diag}(w_i)$$ is a diagonal matrix, and $$e$$ is a $$m$$ vector of all ones. Applying the Newton’s method to solve (3), a reduced KKT system is obtained for $$(\Delta x, \Delta y)$$

$$\begin{bmatrix} -H(x,y) & A^T \\ A & \mu^{-1} W^2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \sigma \\ \rho + \mu^{-1} W^2 \gamma_w \end{bmatrix},$$

and the vector $$\Delta w$$ can be recuperated from

$$\Delta w = \mu^{-1} W^2 (\gamma_w - \Delta y).$$
where $H(x,y) = \nabla^2 F(x) - \sum_{i=1}^{m} y_i \nabla^2 h_i(x)$ and

$$\sigma = \nabla F(x) - A^T y, \quad \gamma_c = \mu W^{-1} e - y, \quad \rho = w - h(x).$$

**Theorem 1** If the dual normal matrix $N(x,w,y) = H(x,y) + A^T \mu W^{-2} A$ is nonsingular, then (4) has a unique solution.

From an initial point $x_0$, $w_0 > 0$ and $y_0 > 0$ the algorithm proceeds iteratively choosing a step length $\alpha_k$ at each iteration, and defining a new estimate to the optimal solution by $x_{k+1} = x_k + \alpha_k \Delta x_k$, $w_{k+1} = w_k + \alpha_k \Delta w_k$, $y_{k+1} = y_k + \alpha_k \Delta y_k$. The step length $\alpha_k$ should guarantee the nonnegativity of the slack and dual variables. The procedure that decides which trial step size is accepted is a filter line search method.

This algorithm is a quasi-Newton based method in the sense that a symmetric positive definite quasi-Newton BFGS approximation, $B_k$, is used to approximate the Hessian of the Lagrangian $H$, at each iteration $k$.

To simplify the notation consider the vectors

$$u = (x,y). \quad \Delta = (\Delta x, \Delta w, \Delta y), \quad u^1 = (x,w), \quad \Delta^1 = (\Delta x, \Delta w),$$

To adapt the methodology of a filter, as described in [1], to this interior point framework, each entry in the filter has three components. Based on the optimality conditions (3) the first component corresponds to feasibility, the second corresponds to centrality and the third to optimality:

$$\theta_{feas}(u^1) = \|\rho\|_2, \quad \theta_c(u^2) = \|y_0\|_2, \quad \theta_{op}(u^3) = \frac{1}{2} \|\sigma\|_2^2.$$

After a search direction $\Delta_k$ has been computed, a backtracking line search procedure is implemented, where a decreasing sequence of step sizes $\alpha_{l+1} \in (0, \alpha_{l}^{\text{max}}] \ (l = 0,1,\ldots)$, with $\lim \alpha_{l} = 0$, is tried until an acceptance criterion is satisfied. Here $l$ denotes the iteration counter for the inner loop. $\alpha_{l}^{\text{max}}$ is the longest step length that can be taken along these directions before violating the nonnegativity conditions $w \geq 0$ and $y \geq 0$ with an upper bound of 1. Using a filter method, the trial point $u_k(\alpha_{l,k}) = u_k + \alpha_{l,k} \Delta_k$ is accepted by the filter if it improves feasibility $\theta_{feas}(u_k^l(\alpha_{l,k})) < \theta_{feas}(u_k^l)$, or if it improves the centrality $\theta_c(u_k^3(\alpha_{l,k})) < \theta_c(u_k^3)$, or if it improves the optimality $\theta_{op}(u_k^3(\alpha_{l,k})) < \theta_{op}(u_k^3)$. However, global convergence cannot be guaranteed with these simple conditions. Stronger convergence have to be imposed in order to consider a trial point to be acceptable. These are closely related to those proposed in [3]. Thus a trial point $u_k(\alpha_{l,k})$ during the backtracking line search is acceptable if it leads to sufficient progress in one of the three measures compared to the current iterate, i.e., if

$$\theta_{feas}(u_k^1(\alpha_{l,k})) \leq (1 - \gamma_0) \theta_{feas}(u_k^1) \quad \text{or} \quad \theta_c(u_k^3(\alpha_{l,k})) \leq (1 - \gamma_0) \theta_c(u_k^3)$$

holds for fixed constants $\gamma_0, \eta_0, \gamma_c \in (0,1)$. However, if for the current trial step size $\alpha_{l,k}$, the following switching conditions

$$m_k(\alpha_{l,k}) < 0 \quad \text{and} \quad [-m_k(\alpha_{l,k})]^{1/2} > \delta \left[\theta_{feas}(u_k^1)\right]^{1/2} \quad \text{and} \quad [-m_k(\alpha_{l,k})]^{1/2} > \delta \left[\theta_c(u_k^3)\right]^{1/2}$$

hold with fixed constants $\delta > 0$, $s_f > 1$, $s_c > 1$, $s_{op} \geq 1$, where $m_k(\alpha) = \alpha \nabla \theta_{op}(u_k^3)^T \Delta_k$, then the trial point $u_k(\alpha_{l,k})$ has to satisfy the Armijo condition

$$\theta_{op}(u_k^3(\alpha_{l,k})) \leq \theta_{op}(u_k^3) + \eta_o m_k(\alpha_{l,k}),$$

instead of (5), in order to be acceptable, where $\eta_o \in (0,0.5)$ is a constant.

In order to prevent cycling between iterates that improve either the feasibility, or the centrality, or the optimality, at each iteration $k$, the algorithm maintains a filter defined by a set $F_k$ that contains values of $\theta_{feas}$, $\theta_c$ and $\theta_{op}$, that are prohibited for a successful trial point in iteration $k$ [3, 4, 5, 6]. Thus, during the line search, a trial point $u_k(\alpha_{l,k})$ is acceptable, if $(\theta_{feas}(u_k^1(\alpha_{l,k})), \theta_c(u_k^3(\alpha_{l,k})), \theta_{op}(u_k^3(\alpha_{l,k}))) \not\in F_k$.

The filter is initialized to $F_0 = \{(\theta_{feas}, \theta_c, \theta_{op}) \in R^3 : \theta_{feas} > (1 - \gamma_0) \theta_{feas}^k \land \theta_c > (1 - \gamma_0) \theta_c^k \land \theta_{op} > \theta_{op}^k - \gamma_0 \theta_{feas}^k\}$ after every iteration in which the accepted trial step size
satisfies (5). On the other hand, if (6) and (7) hold for the accepted step size, the filter remains unchanged. Finally, when the trial step size $\alpha_{k,l}$ becomes too small, the algorithm reverts to a restoration phase that aims to find a new iterate $u_{k+1}$ that is acceptable to the current filter for which (5) holds, by decreasing either the feasibility or the centrality measure:

$$\theta_{z,\text{feas}}(u^1) = \frac{1}{2} \| \rho \|_2^2, \quad \theta_{z,c}(u^2) = \frac{1}{2} \| \gamma \|_2^2.$$

The reader is referred to [7] for details.

The following theorem shows that the search directions $(\Delta x, \Delta w, \Delta y)$ are descent directions for the barrier function $\phi$, whenever the problem is strictly convex, and are descent directions for the optimality measure $\theta_{op}$ as well as for the feasibility and centrality measures, $\theta_{z,\text{feas}}$ and $\theta_{z,c}$, that are present in the restoration algorithm.

**Theorem 2** The search directions have the following properties:

(i) If the dual matrix $N$ is positive definite and $\rho = 0$, then $\nabla \phi^T \Delta^1 \leq 0$.

(ii) Furthermore, $\nabla \theta_{z,\text{feas}}^T \Delta^1 \leq 0$, $\nabla \theta_{z,c}^T \Delta^2 \leq 0$, $\nabla \theta_{op}^T \Delta^3 \leq 0$.

In all cases, the equality holds if and only if $(x,w)$ satisfies (3) for some $y$.

## GLOBAL CONVERGENCE TO STATIONARY POINTS

In this section, we will present the main results of the global convergence properties of the method, that will be stated under the following main assumptions:

**A1** There exists an open set $\mathcal{C} \subseteq \mathbb{R}^d$ with $[u_k, u_k + \Theta^\max_k \Delta_k] \subseteq \mathcal{C}$ for all $k$ so that $\rho$, $\gamma_k$ and $\sigma$ are differentiable on $\mathcal{C}$, and their function values, as well as their first derivatives, are bounded and Lipschitz continuous over $\mathcal{C}$, with $\pi = n + 2m$.

**A2** The matrices $B_k$ that approximate the Hessian of the Lagrangian used in (4) are uniformly bounded for all $k$.

**A3** The matrices $N_k = B_k + \mu_k A_k W_k^{-2} A_k$ are uniformly positive definite on the null space of the Jacobian of the active constraints $A_k^T$.

**A4** The sequence $\{u_k\}$ is bounded.

The next lemma measures the decrease on the centrality obtained by the new iterate $u(\alpha)$.

**Lemma 1** For $\alpha \in (0,1]$ and all $i = 1, \ldots, m$ it holds

$$w_i(\alpha) y_i(\alpha) \leq (1-\alpha) w_i y_i + \alpha \mu + \alpha^2 \| \Delta^2 \|^2,$$

$$\mu(\alpha) \leq \mu + \alpha^2 \| \Delta^2 \|^2,$$

$$\mu(\alpha) \geq \mu - \alpha^2 \| \Delta^2 \|^2.$$

The following result shows that the search direction is a direction of sufficient descent for the optimality measure $\theta_{op}(x^3)$ for all points that are non-optimal.

**Lemma 2** Suppose Assumptions A1-A3 hold. If $\{u_k\}$ is a subsequence of iterates for which $\theta_{op}(u_k^3) \geq \varepsilon$ with a constant $\varepsilon > 0$ independent of $i$, then there exists a constant $\varepsilon_\sigma > 0$, such that for all $i$ and $\alpha \in (0,1]$

$$m_k(\alpha) \leq -\varepsilon_\sigma \alpha.$$

The following lemma provides upper bounds for the three components of the filter, $\theta_{\text{feas}}$, $\theta_c$ and $\theta_{op}$, at the new iterate $u_k(\alpha)$ in terms of $\alpha \Delta_k$ and their corresponding values at the current iterate $u_k$.

**Lemma 3** Suppose Assumption A1 holds. There exist positive constants $M_f$, $M_c$, and $M_o$, depending on an upper bound for $\| \sigma \|$ and on the Lipschitz constants of $\nabla \rho$, $\nabla \gamma_k$, and $\nabla \sigma$, such that, for all $\alpha \leq 1$,

$$\theta_{\text{feas}}(u_k^1(\alpha)) \leq (1-\alpha) \theta_{\text{feas}}(u_k^1) + M_f \alpha^2 \| \Delta_k^1 \|^2,$$

$$\theta_{c}(u_k^2(\alpha)) \leq (1-\alpha) \theta_{c}(u_k^2) + M_c \alpha^2 \| \Delta_k^2 \|^2,$$

$$\theta_{op}(u_k^3(\alpha)) \leq (1-\alpha) \theta_{op}(u_k^3) + M_o \alpha^2 \| \Delta_k^3 \|^2.$$

The following result comes directly from the Assumptions A1 and A4 and Lemmas 1 and 3.

**Lemma 4** Suppose Assumptions A1 and A4 hold. Then

1. The sequences $\{ \theta_{\text{feas}}(u_k^1) \}, \{ \theta_c(u_k^2) \}, \{ \theta_{op}(u_k^3) \}$ are bounded.
2. The constants $M_f, M_e, M_o$ in Lemma 3 are bounded for all $k$.

Finally we state the main global convergence result.

**Theorem 3** Suppose Assumptions A1-A4 hold. Then

$$\lim_{k \to \infty} \theta_{feas}(u_k^1) = 0, \lim_{k \to \infty} \theta_{c}(u_k^2) = 0, \lim_{k \to \infty} \inf \theta_{op}(u_k^3) = 0.$$  

In other words, all limit points are (primal) feasible and satisfy the centrality condition, and there exists a limit point $u_\ast$ of $\{u_k\}$ which is a first-order KKT point for the problem (1).

### NUMERICAL EXPERIMENTS

To test this interior point framework with a filter line search method we used 111 constrained problems from the Hock and Schittkowski test set [8]. These are small and medium scale problems. The tests were done in double precision arithmetic with a Pentium 4 and C programming language. The chosen values for the constants are similar to the ones used in [5]: $\theta_{feas}^{max} = 10^4 \max \{1, \theta_{feas}(u_0^1)\}$, $\theta_{c}^{max} = 10^3 \max \{1, \theta_{c}(u_0^2)\}$, $\theta_{op}^{max} = 10^4 \max \{1, \theta_{op}(u_0^3)\}$, $\gamma_0 = \gamma_1 = 10^{-5}$, $s_f = s_c = 1.1$, $s_o = 2.3$, $\eta_max = 10^{-4}$ and $\delta = 1$.

The stopping criteria are based on relative primal and dual feasibility measures similar to the ones in [5]. These preliminary tests aim to analyze the effect of different starting procedures on the solution. Three experiments were carried out. First, with the initial approximation $x_0$ given in [8], the algorithm recomputes a better approximation, say $\tilde{x}_0$, as well as $y_0$, by solving a simplified reduced KKT (see (4)). The initial matrix $B_0$ is a positive definite modification of $\nabla^2 F(\tilde{x}_0)$. In the second experience $B_0$ is set to the identity matrix as for 6 problems the Hessian of $F$ is singular at the initial approximation. Finally, the third experience considers different initial primal and dual variables. It uses the given $x_0$ and takes $y_0 = 1$. In this situation, the algorithm solves 18 more problems that were not solved in the two previous experiences for an iteration limit of 100. Combining all implementations and considering the best results obtained so far, the algorithm was able to solve 101 problems. This corresponds to a success rate of 91%. The average number of iterations is 17.2 and the registered average number of $\theta_{op}$ evaluations is 40.3.

### CONCLUSIONS

A primal-dual interior point filter line search method that considers three components for each entry in the filter has been studied. Global convergence to stationary points has been shown. The preliminary numerical results seem to show that the algorithm is promising in practice. In the future, besides enlarging the numerical experiments to other problems set, a predictor-corrector strategy, as outline in [2], will be implemented to improve the robustness of the algorithm.

### REFERENCES