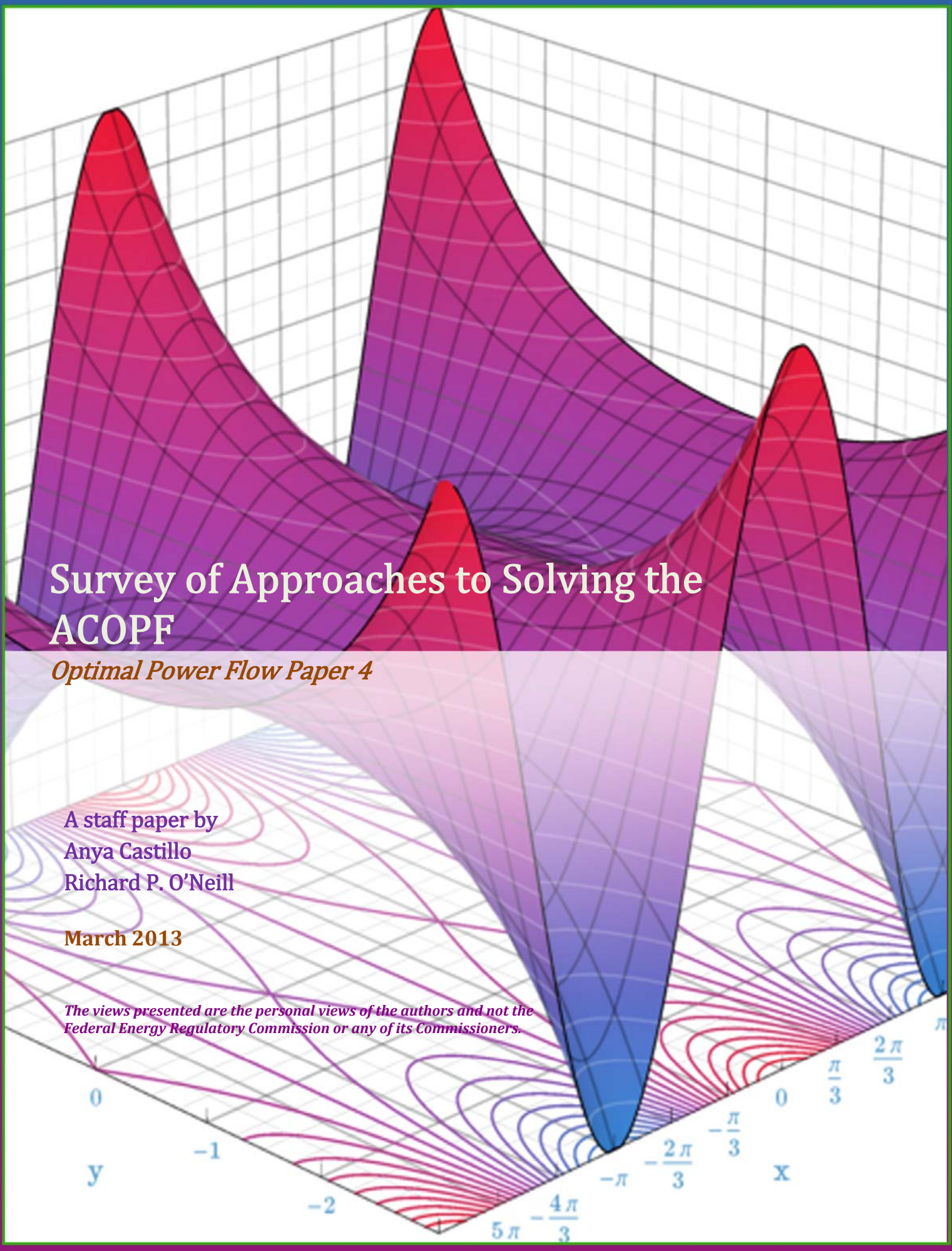
Survey of Approaches to Solving the ACOPF

Optimal Power Flow Paper 4



A staff paper by
Anya Castillo
Richard P. O'Neill

March 2013

The views presented are the personal views of the authors and not the Federal Energy Regulatory Commission or any of its Commissioners.

Survey of Approaches to Solving the ACOPF

Optimal Power Flow Paper 4

Anya Castillo and Richard P O'Neill

anya.castillo@ferc.gov; richard.oneill@ferc.gov

March 2013

Abstract:

We present a background on approaches historically applied to solve the ACOPF, many which are used in our following companion study on testing solution techniques (Castillo, 2013). In this paper we present an introduction on the associated theory in nonlinear optimization, and then discuss the solvers and published algorithms that have been applied to the ACOPF, dating initially from Carpentier in 1962 to current day approaches. We provide insight into the major contributions in solution methods applied to the ACOPF to date.

Table of Contents

1. Introduction

Recently, mathematical optimization has become a powerful tool in many real-life applications along with the more traditional uses in engineering and science. The history is driven, and will continue to be driven, by the demand to solve increasingly larger nonlinear optimization problems under the limitations of storage and computing time.

Here, we present background on approaches that are used in a companion study and are often applied to solve the ACOPF. We are focused on understanding the reliability of such approaches and whether there is good reason to believe that convergence of the ACOPF would be guaranteed at a reasonable rate. Therefore, we introduce and discuss some of the associated theory.

In section 2, we present definitions. In section 3, we review unconstrained optimization for nonlinear programming followed by a review of constrained optimization for nonlinear programming in section 4; there are numerous nonlinear programming resources and, for example, the reader can refer to (Nocedal, 1999), (Panos, 2006) for a more detailed discussion. In section 5, we review problem decomposition techniques and popular approximations to the ACOPF. In section 6, we review the nonlinear commercial solvers applied to the ACOPF in our companion computational study (Castillo, 2013). Then in section 7, we review the literature on solving the ACOPF followed by a brief discussion in section 8.

2. Definitions

Definition 1. If $x \in X \subset R^n$, x is called a feasible solution.

Definition 2. The function $f(x)$ with domain R^n and codomain R is denoted by

$f(x)$: $R^n \rightarrow R$.

Definition 3. If $f(x^*) \leq f(x)$ for all $x \in X$, x^* is called an optimal solution.

Definition 4. If $f(x^*) \leq f(x)$ for all $x \in X' \subset X$, x^* is called a local optimal solution.

Definition 5. The set $TR \subset X$ is called a trust region where an approximation of $f(x)$ is 'trusted.'

Definition 6. The set X is convex, if and only if $\alpha x_1 + (1 - \alpha) x_2 \in X$ for all $\alpha \in [0,1]$ and all $x_1, x_2 \in X$.

Definition 7. The function $f(x)$ is convex in domain X if and only if $f(\alpha x_1 + (1 - \alpha) x_2) \leq \alpha f(x_1) + (1 - \alpha) f(x_2)$ for any $\alpha \in [0,1]$ and all $x_1, x_2 \in X$.

Definition 8. If $d \in R^n$ and $x + sd \in X$ for some $s > 0$, d is called a feasible direction.

Definition 9. The function $F(x_k) = x_{k+1}$ is the result of iteration k and the approach is a descent algorithm if $f(x_{k+1}) < f(x_k)$ (the reduction condition).

Defintion 10. The L_2-norm is $| |$.

Definition 11. If $f(x)$ is once differentiable, the vector of first derivatives of $f(x)$ with respect to x is $\nabla f(x)$; $\nabla f(x)$ is also called the Jacobian or gradient vector.

Definition 12. If $f(x)$ is twice differentiable, the matrix of second derivatives of $f(x)$ with respect to x is $H(x) = \nabla^2 f(x)$; $H(x)$ is called the Hessian.

Definition 13. If Q is a symmetric matrix and $d_1^T Q d_2 = 0$, d_1 and d_2 are Q-orthogonal, also known as conjugate directions.

Definition 14. If $(A - eI)v = 0$, e is an eigenvalue and v is an eigenvector of A.

Definition 15. For unconstrained optimization we solve $\inf \{f(x) | \ x \in X\}$ where the objective function f(x): $R^n \rightarrow R$

Definition 16. For equality constrained optimization we solve $\inf \{f(x) | \ g(x) = 0, x \in X\}$ where the constraints $g(x)$: $R^n \rightarrow R^m (m \leq n)$.

Definition 17. For inequality constrained optimization we solve $\inf \{f(x) | \ g(x) \leq 0, x \in X\}$ where the constraints $g(x)$: $R^n \rightarrow R^m$ (m may be larger than n).

Definition 18. The index set AS is the active set of constraints where $AS(x) = \{j | \ g_j(x) = 0\}$.

Definition 19. The Lagrangian function is $L(x, \lambda) = f(x) + \lambda^T g(x)$.

Definition 20. The Lagrange multiplier λ exists such that (x^*, λ^*) is a stationary point of the Lagrangian function where $\nabla_\lambda L(x^*, \lambda^*) = 0$, which implies $g(x^*) = 0$.

3. Unconstrained Nonlinear Optimization

We start with the 'unconstrained' nonlinear optimization problem. We want to minimize the nonlinear function $f(x)$:

$$f^* = \inf \{f(x) | \ x \in X\}$$

where x^* is a solution to the above problem (P). Methods for solving this problem date back to Newton and Gauss. In unconstrained optimization, the set X is R^n. If X is a subset of R^n, X could be the set of integers Z^n or denote a constraint set of equations. If P is infeasible, we define $f^* = \infty$; if P is unbounded, we define $f^* = -\infty$. The function f is typically called the objective function. Equivalently, the problem can be stated as a maximization of $-f(x)$. We present the following theorems without proof. For more details see Mangasarian (1969) and Luenberger (1984).

Theorem 1. If $f(x)$ is differentiable, x^* is a relative or local minimum if for any feasible direction d, $\nabla f(x^*) d \geq 0$.

Theorem 2. If $f(x)$ is twice differentiable, x^* is a relative or local minimum if for any direction feasible d, $\nabla f(x^*) d \geq 0$; and if $\nabla f(x^*) d = 0$, $d^T H(x^*) d \geq 0$.

Theorem 3. If X is a convex set and $f(x)$ is convex function on X, a relative or local minimum is a global minimum.

Theorem 4. If X is a convex set and $f(x)$ is convex function on X and twice differentiable, $H(x^*)$ is positive semidefinite on X.

Theorem 5. If A is symmetric, all eigenvalues are real.

Theorem 6. If all eigenvalues are positive (negative), A is positive (negative) definite.

Theorem 7. If one or more eigenvalues are 0, A is singular.

Theorem 8. If A is symmetric, V is the eigenvector matrix and E is diag(e) where $AV = VE$ and V can be chosen orthonormal ($V^{-1} = V^T$), thus $A = VEV^T$.

Iterative Methods. The nonlinear function $f(x)$ cannot be solved directly through factorization methods. The algorithms for solving unconstrained nonlinear optimization problems can be broadly defined as derivative-free methods, methods using first derivatives and methods using second-order derivatives. Since the standard ACOPF is continuous and differentiable, we focus on derivative-based methods.

Most of the methods described are iterative methods which generate a sequence of $X_K = \{x_1, x_2, \ldots x_k, \ldots x_K\}$ for K iterations. We call any specific process a solver or an algorithm. Generally, all iterative search methods have a five step process:

Step 1. Choose a function to optimize; set $k = 0$. Choose an initial point: x_0.

Step 2. Choose a search direction: d_k.

Step 3. Choose step size s_k where s_k is a positive scalar and calculate a new point: $x_{k+1} = x_k + s_k d_k$.

Step 4. Test for stopping: If x_{k+1} satisfies the convergence criteria or exceeds the time allotted, then set $K = k+1$ and stop. Otherwise, if x_{k+1} does not satisfy the convergence criteria, then set $k = k + 1$ and go to Step 2.

There are numerous approaches for each step. Whereas line search methods compute a search direction d_k and then a step size s_k, trust region methods define a

region $TR \subset X$, typically an elipse around the current iterate x_k, to choose a step size $s_k \in TR$ and direction d_k simultaneously. The choice of the step size and then direction in line search methods, or in the case of trust region methods the choice of the trust region and then determining $s_k d_k$, is important in promoting convergence from remote starting points.

In Step 1, the user can supply the starting point. If the initial point is infeasible, finding a feasible solution can be an optimization problem in itself. For parallel algorithms, the process may start with many initial points. The simplest parallel algorithm is the horse race where the approaches are started in parallel and the first to converge terminates the algorithm. More complex parallel algorithms interact.

Steps 2 and 3 are often defined as a single step $x_{k+1} = F(x_k)$ where F is the descent algorithm. Whether a linear search or trust region method is applied, the result is where

$$f(x_k + s_k d_k) \leq f(x_k)$$

and hopefully

$$f(x_k + s_k d_k) << f(x_k).$$

Typically d_k is a descent direction where

$$d_k{}^T \nabla f(x_k) < 0.$$

For instance we can determine the search direction d_k by Newton-Raphson's method:

$$\nabla^2 f(x_k)\, d_k = -\nabla f(x_k).$$

We will discuss Netwon-Raphson in further detail below. Note that we may replace the above Hessian $H(x_k) = \nabla^2 f(x_k)$ with any symmetric and nonsingular matrix B_k. If $B_k = I$, the descent algorithm is called steepest descent. If $B_k \approx H(x_k)$, the method is called quasi-Newton. For example, the restricted step varies from a Newton step to a small steepest decent step. The Newton step approach has quadratic convergence, but can fail to converge. The steepest decent approach has slower linear convergence, but fails less frequently.

Step 4 tests for convergence. Numerous strategies assume Lipschitz continuity of the gradient (Nocedal, 1999). In theory, to prove convergence an infinite sequence is shown to converge to a stationary point. A subsequence has a limiting property that satisfies first-order necessary or second-order necessary conditions. In practice, a solver stops in a finite number of iterations when the solution is 'close enough' to the optimal solution. For example, for some user defined convergence criteria $\delta > 0$, stop if

$$|(x_k - x_{k+1})|/|x_k| \leq \delta \text{ or } f_k - f_{k+1} \leq \delta \text{ or } |\nabla f(x_k)| < \delta.$$

Furthermore, the globally convergent algorithms are those which converge from any starting point to a local optima, and therefore have the property that the gradient norms converge to zero:

$$\lim_{k \to \infty} ||\nabla f_k|| = 0$$

Rate of Convergence. The error term is defined as

$$\varepsilon_k = x_k - x^*$$

If $0 \leq \beta = \lim \sup_{k \in \infty} |\varepsilon_{k+1}|/|\varepsilon_k|^p < \infty$, β is called the convergence ratio and p is the order of convergence. If $p = 1$ and $\beta > 0$, the convergence rate is called linear. If $p = 1$ and $\beta = 0$, the convergence rate is called superlinear. If $p = 2$ and $\beta > 0$, the convergence rate is called quadratic.

Feasible Direction Methods. Feasible direction methods require that for a given iterate $x_k \in X$, we can find a descent direction d_k which is also a feasible direction at x_k. Therefore at each iteration there must exist a new feasible point of the form $x_k + s_k d_k$ that meets the reduction condition.

Feasible direction methods include both derivative free and derivative based methods. Line search without derivatives, such as the Fibonacci search, only requires function evaluations. The Fibonacci sequence is the basis for this approach in which sequential points are chosen such that the discrepancy $f(x_{k+1}) - f(x_k)$ is minimized. However if f is differentiable, we solve $\nabla_s f(x_k + sd_k) = 0$. This is a standard algorithm in the initial stage of a line search.

First-derivative gradient approaches use the first derivative as the feasible direction, $d_k = -\nabla f(x_k)$. This is the direction at which f decreases most quickly. Below we summarize the first-derivative methods Gauss-Seidel, steepest descent, conjugate gradient, and Quasi-Newton, and then also the second-derivative based Newton's method.

Gauss-Seidel Method. The Gauss-Seidel approach chooses the unit vector as a coordinate direction $d_k = (0, \ldots 1_k, \ldots 0)$ and a line search in one variable is performed for some $s > 0$:

$$\min_s f(x_k + sd_k)$$

This problem is sequentially solved in each coordinate direction and repeated until the convergence criteria is met. This approach has linear convergence rate.

Steepest Descent Method. The steepest descent method, which is fundamentally

$$x_{k+1} = F(x_k) = x_k - \nabla f(x_k)$$

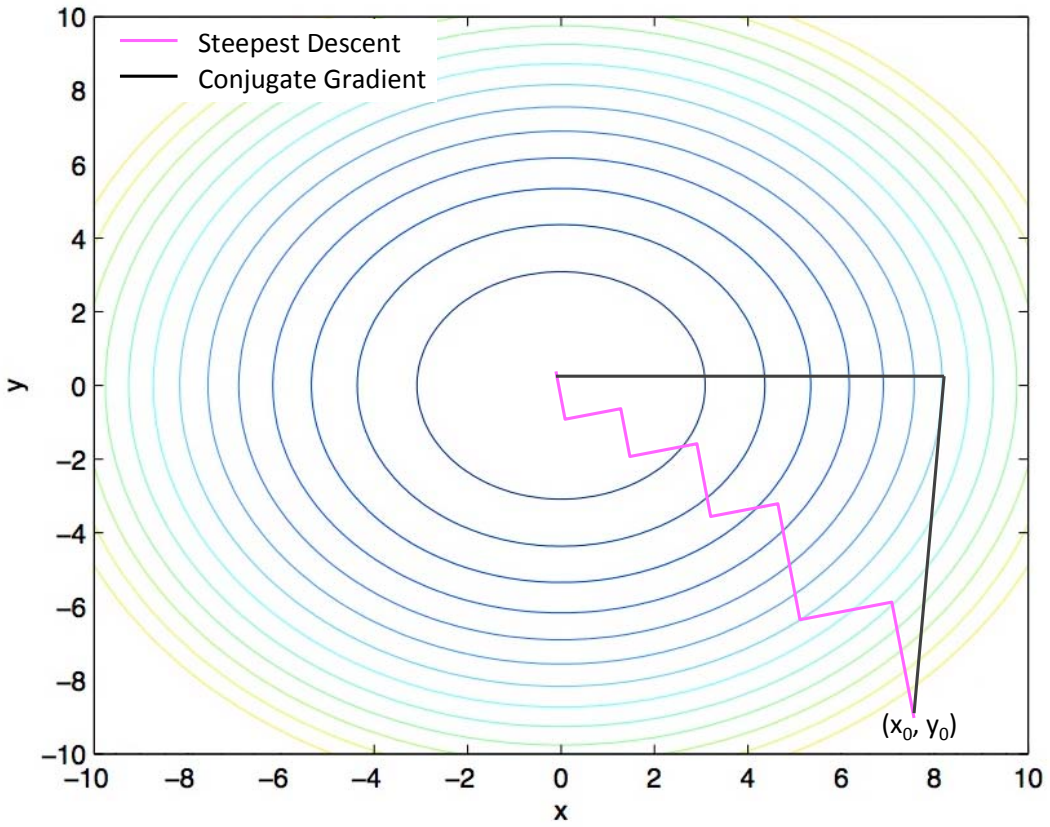
for unit step size and more specifically,

$$x_{k+1} = F(x_k) = x_k + s_k d_k$$

for s_k, $d_k > 0$ suffers from slow convergence due to zigzagging, as shown in Figure 1. Numerous methods have been proposed to avoid zigzagging.

Ideally in choosing the step size s, we find the local minimizer to the above objective function in order to obtain a substantial reduction in f. However this may require too many evaluations of f, and in the case if f is differentiable, then $\nabla_s f$. Therefore strategies that perform inexact line search identify an adequate reduction in f by trying candidate values for s. As we shall see, exact line searches are not needed and may not be the best approach because algorithms that achieve rapid convergence can sometimes conflict with the global convergence requirements, and vice versa.

Figure 1. Zigzagging with steepest descent and lack of zigzag with conjugate directions.



Conjugate Gradient Method. The introduction of the conjugate gradient method by Fletcher and Reeves in 1964 was the inception of large-scale nonlinear optimization. The conjugate gradient method determines conjugate directions to the Hessian

through evaluating f and ∇f, but without directly evaluating the Hessian $\nabla^2 f$. The procedure is as follows:

Step 1. Let $g_0 = \nabla f(x_0)$ and $d_0 = -g_0$; set $k = 0$.

Step 2. Solve for $x_{k+1} = x_k + s_k d_k$ where $s_k = \text{argmin}_s f(x_k + s d_k)$.

Step 3. Determine direction $d_{k+1} = g_{k+1} + r_k d_k$ where
$r_k = g_{k+1}^{\mathrm{T}} g_{k+1} / g_k^{\mathrm{T}} g_k$, and
$g_{k+1} = \nabla f(x_{k+1})$.

Step 4. Go to Step 2 and repeat for k = 1 …, $n-1$.

Quasi-Newton Method.

In the late 1950's, Davidon introduced quasi-Newton methods, based on Newton's method and also known as secant methods, which use a sequence of positive definite matrices to approximate the Hessian (or inverse Hessian). These methods are inexact line search methods and have a superlinear convergence rate. Fletcher and Powell demonstrated that Davidon's approach is equivalent to conjugate gradient method when applied with exact line searches to convex quadratic functions. Quasi-Newton methods only require first derivatives, approximate the Hessian and adhere to the descent property due to the positive definiteness.

For example, if $f(x)$ is twice differentiable, the second-order Taylor's series expansion of $f(x)$ at x_k is:

$$f(x) = f(x_k) + \nabla f(x_k)^{\mathrm{T}}(x - x_k) + (x - x_k)^{\mathrm{T}} H(x_k)(x - x_k)/2 + O(|x - x_k)|^2)$$

Dropping $O(|x - x_k)|^2)$, $\underline{f}(x)$ is a quadratic approximation of $f(x)$

$$\underline{f}(x) = f(x_k) + \nabla f(x_k)^{\mathrm{T}}(x - x_k) + (x - x_k)^{\mathrm{T}} H(x_k)(x - x_k)/2$$

If $f(x)$ is quadratic, then

$f(x) = b^{\mathrm{T}} x + x^{\mathrm{T}} Q x/2$,
$\nabla f(x) = b^{\mathrm{T}} + x^{\mathrm{T}} Q$,
$H(x) = Q$, and
$O(|x - x_k)|^2) = 0$.

The above expansion is exact and x^* is a solution to

$$Qx^* = -b.$$

If Q is positive semidefinite at x^*, we have that

$\nabla f(x) = 0$,
$x^{*\mathrm{T}} Q x^* \geq 0$, and

$f(x) \geq f(x^*)$.

Therefore in quasi-Newton methods the Hessian approximation $\underline{H}_k$ is chosen to satisfy

$$\nabla f(x_k + \Delta x) = \nabla f(x_k) + \underline{H}_k \Delta x$$

which is called the *secant equation* and is the Taylor series of the gradient itself.

Steps in the Quasi-Newton Method:

Step 1. An approximate initial value of $\underline{H}_0 = I$ is frequently sufficient for convergence. *Set* $k = 0$.

Step 2. $\underline{H}_k$ is positive definite. Let $d_k = -\underline{H}_k^{-1} \nabla f(x_k)$ and compute $s_k = \text{argmin}_s f(x_k + sd_k)$. Set $x_{k+1} = x_k + s_k d_k$, where s_k satisfies Wolfe conditions:

$$f(x_k + s_k d_k) \leq f(x_k) + \alpha_1 s_k \, d_k{}^{\mathrm{T}} \nabla f(x_k)$$

$$d_k{}^{\mathrm{T}} \nabla f(x_k + s_k d_k) \geq \alpha_2 s_k d_k{}^{\mathrm{T}} \nabla f(x_k)$$

where $0 < \alpha_2 < \alpha_2 < 1$. The above first inequality is sometimes referred to as the Armijo rule, which ensures that the step length s_k results in a sufficient decrease in f, and the second is the curvature condition, which ensures that the slope has been reduced sufficiently (Nocedal, 1999). Another way of stating the curvature condition is as follows:

$$v_k{}^{\mathrm{T}} (\nabla f(x_k + s_k d_k) - \nabla f(x_k)) = v_k{}^{\mathrm{T}} \underline{H}_k v_k > 0$$

where $v_k = s_k d_k$. The Wolfe conditions can be applied in most line search methods and are important in implementing quasi-Newton approaches.

Step 3. Compute $\underline{H}_{k+1}$ by a rank-two-update, where the inverse of the Hessian $\underline{B}_k = \underline{H}_k^{-1}$ is updated by the sum of two symmetric rank 1 matrices:

$$\underline{B}_{k+1} = \underline{B}_k + \beta_k v_k v_k{}^{\mathrm{T}} - \gamma_k \underline{H}_k y_k y_k{}^{\mathrm{T}} \underline{H}_k$$

where $v_k = s_k d_k$, $y_k = \nabla f(x_k + s_k d_k) - \nabla f(x_k)$, $\beta_k = 1 / v_k{}^{\mathrm{T}} y_k$, and $\gamma_k = 1 / y_k{}^{\mathrm{T}} H_k y_k$. The Hessian approximation is chosen to satisfy the quasi-Newton condition:

$$\underline{H}_{k+1} (\nabla f(x_{k+1}) - \nabla f(x_k)) = x_{k+1} - x_k.$$

Set $k = k + 1$; go to Step 2.

The above algorithm is known as the Davidon-Fletcher-Powell (DFP); the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is the dual of DFP because whereas the DFP converges to the inverse of the Hessian, the BFGS method converges to the Hessian in itself and therefore is a more direct approach.

In 1970, Broyden, Fletcher, Goldfarb, and Shanno independently introduced a secant approach now known as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. In this direct approach,

$$\underline{H}_{k+1} = \underline{H}_k + \beta_k y_k y_k^{\mathrm{T}} - \gamma_k \underline{H}_k v_k v_k^{\mathrm{T}} \underline{H}_k$$

where instead $\gamma_k = 1 / v_k^{\mathrm{T}} H_k v_k$.

The Levenberg-Marquardt method uses a trust-region to solve a system. The method determines a value $\nu \geq 0$ such that $H(x_k) + \nu I$ is positive definite and then solves

$$s_k \Delta x_k = -(H(x_k) + \nu I)^{-1} \nabla f(x_k).$$

An iteration radius r_k may impose a limit on the step size s_k, or alternatively ν controls the length. However by not controlling the length r_k, degeneracy (i.e. non-invertible $(H(x_k) + \nu I)$) and ill-conditioning issues can result in a step size and direction that are badly determined by ν.

Newton's Method. Newton's method has been applied with both line search and trust region approaches. Non-monotone methods in which the function values are allowed to increase at some iterations have been useful in solving highly nonconvex problems.

The basic Newton's algorithm is:

Step 1. Guess x_0, and set $k = 0$.

Step 2. At x_k, evaluate $\nabla f(x_k)$.

Step 3. Evaluate $B_k = H(x^k)$ or an approximation $B_k = \underline{H}(x^k)$. For example, when $H(x_k)$ is not positive definite, then modifications could be applied to determine $\underline{H}(x^k)$ and guarantee a decrease in $f(x_k)$. Such methods may modify $H(x_k)$ so that

$$\nabla f(x_k)^{\mathrm{T}} B_k \nabla f(x_k) > 0,$$

or such methods may compute a negative search direction which satisfies

$$\Delta x_k^{\mathrm{T}} B_k \Delta x_k > 0 \text{ and}$$

$$\nabla_x f(x_k)^{\mathrm{T}} \Delta x_k \leq 0.$$

Step 4. Solve: $B_k d_k = -\nabla f(x_k)$.

Step 5. If the convergence error is less than the tolerance, (e.g., $||\nabla f(x_k)|| \leq \varepsilon$ and $||d_k|| \leq \varepsilon$) then stop; else continue.

Step 6. Find s_k so that $0 < s_k \leq 1$ and $f(x_k + s_k d_k) < f(x_k)$.

Step 7. $x_{k+1} = x_k + s_k d_k$. Set $k = k + 1$ and go to Step 2.

Note that recalculating $\nabla f(x_k)$, B_k, and $s_k d_k = x_{k+1} - x_k$ at each iteration may be very costly in terms of computational time. Let $d_k = \nabla f(x_k)$, $x_{k+1} = x_k + s d_k$. If $H(x_k)$ is positive semidefinite,

$$u = d_k^{\mathrm{T}} H(x_k) d_k \geq 0 \text{ and } v = d_k^{\mathrm{T}} d_k \geq 0$$

$$f(x_{k+1}) - f(x_k) = d_k^{\mathrm{T}} s d_k + s d_k^{\mathrm{T}} H(x_k) s d_k = vs + us^2$$

for some $s < 0$, where we have $f(x_{k+1}) < f(x_k)$.

For globally convergent Newton and quasi-Newton methods, B_k must be positive definite and also have a bounded condition number because a high condition number can result in ill-conditioned problems (Nocedal, 1999). Intuitively, the search direction d_k is not near orthogonal to the gradient in order to promote convergence. However for problems that are ill-conditioned, it may be necessary to search along directions that are nearly orthogonal to the gradient.

Numerical Efficiency and Stability. Applying the appropriate techniques results in computationally efficient and numerically stable procedures. For example the core of solving a system of linear algebraic equations, $Ax = b$, is decomposing or factoring the coefficient matrix. Through this process, the solution can be obtained with much less computational effort, and better methods are solved faster with less numerical error. Matrix inversion is usually unstable and results in dense matrices and unnecessary floating-point calculations. Sometimes for simplicity and sometimes out of ignorance, an algorithm is presented as inverting a matrix at each iteration: $x = A^{-1} b$. This approach is not only computationally expensive, but also numerically unstable.

For large problems, A is usually sparse, but the inverse is usually dense. Sparse matrices are stored efficiently by storing only nonzeroes and their row and column index. When one of the operands in a floating-point operation is zero, it can be skipped, saving time. Problems are referred to as ill-conditioned when small perturbations to the coefficients result in an unpredictably large change of the solution. Certain conditions such as positive definiteness can be checked, and corrections to an indefinite matrix can be applied.

Accumulating floating point round-off and also cancellation errors cause search vectors to lose orthogonality in conjugate search methods. Similarly, steepest descent methods can accumulate floating point round-off, which may cause x_k to converge to some point near x. This effect can be deterred by occasionally recomputing the correct residual.

Scaling. Numerical stability starts with scaling the problem. Scaling is part theory and part art form. Proper consistent scaling means primal, dual solution values, constants and derivatives of nonlinear terms (Jacobian elements) in absolute value be around 1, e.g. from 0.01 to 100. Most solvers have automatic scaling options.

Factoring Linear Equations. In practice, at each iteration $d_k = -B_k \nabla f(x_k)$ is solved by sparse matrix factorization techniques often a matrix is represented as the product or sum of matrices; such a representation is termed as a decomposition or factorization of the original matrix. The Gaussian elimination product form of the inverse of A stores the inverse as a product of elementary column matrices, where

$$A^{-1} = E_k E_{k-1} \cdots E_2 E_1.$$

and E is an elementary column operation where all but one column is the identity matrix. Therefore we have that

$$x = A^{-1} b = E_k E_{k-1} \cdots E_2 E_1 b.$$

For iterative procedures, as k become large, numerical errors build-up and excessive floating-point operations occur. Therefore the matrix needs to be refactored into $k = n$ factors.

Another technique is LU factorization, where L is a lower triangular matrix and U is an upper triangular matrix. For any matrix A,

$$Ax = LUx = L(Ux) = b$$

We can solve $Ly = b$ by substitution and then $Ux = y$ by substitution. For additional details, see Markowitz LU decomposition (Reid, 1976; 1982) and Bartels-Golub update (Bartels and Golub, 1969; 1971).

Factoring Linear Equations with Symmetric Matrices. Symmetric matrices have more efficient factorizations. The Cholesky decomposition of a Hermitian, positive-definite matrix, A, is the product of a lower triangular matrix and its conjugate transpose. $A = LL^*$ where L^* is the conjugate transpose. Cholesky decomposition is roughly twice as efficient as the LU decomposition for solving systems of linear equations. With QR factorization, A results in an orthogonal matrix Q that is a basis for the column space of A and an upper triangular matrix R where $A = QR$. These approaches lose certain invariance properties associated with Newton's method and Newton's method with line search, and therefore may not transform correctly, when the Hessian is ill-conditioned.

4. Constrained Nonlinear Optimization

The constrained nonlinear problem or program in canonical primal (P) form is:

$$f^* = \inf \{f(x) \mid g(x) \leq 0,\ h(x) = 0,\ x \in X\} \quad \text{(P)}$$

This formulation explicitly introduces an additional set of inequality, $g(x) \leq 0$, and equality constraints, $h(x) = 0$, that are linear or nonlinear. If $x \in X$, $h(x) = 0$ and $g(x) \leq 0$, x is said to be a feasible solution. Nonlinear equality constraints can be represented as $g(x) \leq 0$ and $-g(x) \leq 0$. We define $c(x) = 0$ as $(g(x)+s=0,\ h(x)=0)$.

Lagrangian Function. The algorithms for solving constrained nonlinear optimization problems typically start with a transformation of the Lagrangian. The Lagrangian dual (D) formulation is

$$L^* = \sup_{\lambda \geq 0}\{\inf L(x, \lambda) \mid x \in X, \mathrm{h}(x) = 0\}$$

where $L(x, \lambda) = f(x) + \lambda^{\mathrm{T}} g(x)$ is the Lagrangian function and λ is the Lagrangian multiplier vector. The variable λ is sometimes called the dual or slack variable. Let (x^*, λ^*) be the solution to (D). If $L(x, \lambda)$ is unbounded for $x \in X$, then $L^* = \infty$. Furthermore, if the feasible region to (D) is an empty set, then $L^* = -\infty$.

Iterative methods such as penalty and augmented Lagrangian methods, barrier or interior point methods, variable metric methods, sequential linear programming methods and sequential quadratic programming methods often solve a sequence of generalized Lagrangian functions. The generalized Lagrangian function is:

$$\underline{L}(x, \lambda) = f(x) + \Lambda(g(x), \lambda, \mu)$$

that includes a positive penalty parameter $\mu > 0$ with a penalty function Λ.

Karush-Kuhn-Tucker (KKT) Conditions. If $f(x)$ and $g(x)$ are differentiable the KKT necessary (first-order) conditions are

$$\nabla f(x) + \lambda^{\mathrm{T}} \nabla g(x) = 0,$$
$$\lambda \geq 0,$$
$$\lambda^{\mathrm{T}} \perp g(x) = 0, \text{ and}$$
$$h(x) = 0.$$

Under certain constraint qualifications, for example, if there is $g(x) < 0$ and $h(x) = 0$, or $[\nabla g(x), \nabla h(x)]$ has full rank where the active set AS denotes an active constraint, the KKT are necessary for a local optimum. If $\underline{x}$ is a local optimal solution to P and also satisfies the KKT sufficiency conditions with strict complementarity to $\underline{\lambda}$, then $(\underline{x}, \underline{\lambda}) \in X \times R^{m+}$ is a saddle point of the Lagrangian:

$$L(\underline{x}, \lambda) \leq L(\underline{x}, \underline{\lambda}) \leq L(x, \underline{\lambda})$$

for all $(\underline{x}, \underline{\lambda}) \in X \times R^{m+}$. Therefore, the $\underline{x}$ and $\underline{\lambda}$ are optimal solutions to the primal and dual problems, respectively and

$$\inf\{f(x) \mid g(x) \leq 0,\ h(x) = 0,\ x \in X\} \geq \sup_{\lambda \geq 0} \inf\{L(x, \lambda) \mid h(x) = 0,\ x \in X\}.$$

This property is known as weak duality when x is feasible in P and λ is feasible in D. $f^* - L^* \geq 0$ is called the duality (complementarity) gap. Strong duality results in a zero duality gap:

$$f^* - L^* = 0.$$

Furthermore, if x is feasible and there exists λ such that strong duality holds under certain constraint qualifications where x is a local optima. Certain stability properties lead to necessary and sufficient conditions which hold with equality for the existence of a global saddle point of the Lagrangian function with respect to $X \times R^{m+}$. If $g(x)$ and $f(x)$ are convex functions, then the KKT are sufficient for global optimality. If $f(x)$ and $g(x)$ are twice differentiable, and the Hessian

$$H(x, \lambda) = \nabla^2 L(x, \lambda) = \nabla^2 f(x) + \lambda^{\mathrm{T}} \nabla^2 g(x)$$

is positive definite, then we have a local minima, or if H(x,λ) is negative definite, then we have a local maxima. The result is analogous to strong duality.

Often the KKT may not hold at the optimal solution if the problem is nonconvex, and therefore such KKT necessary conditions are not sufficient to prove global optimality. When the primal problem is nonconvex, there may be local optima that are not globally optimal.

The following subsections detail the dominant approaches, which are based directly on solving the KKT conditions for constrained optimization.

Augmented Lagrangian and Penalty Methods. In 1969 in separate papers, Hestenes and Powell introduced the augmented Lagrangian. In the 1970s, it gained a strong following and is still applied today. The augmented Lagrangian is the Lagrangian function with a quadratic penalty term:

$$\Lambda(g(x), \lambda, \mu) = \lambda^{\mathrm{T}} g(x) + \mu^{\mathrm{T}} g(x)^2$$

The basic algorithm for the augmented Lagrangian and general penalty function methods:

Step 1. Choose λ_0, μ_0 in iteration $k = 0$.

Step 2. Find $x_{k+1} = \operatorname{argmin}_x \{f(x_k) + \lambda_k^{\mathrm{T}} g(x_k) + \mu_k^{\mathrm{T}} g(x_k)\}$.

Step 3. Update $\mu_{k+1} > \mu_k$ and $\lambda_{k+1} = \lambda_k + \mu_k^{\mathrm{T}} g(x_k)$.

Step 4. If convergence criteria is met, then stop; else go to Step 2.

Barrier/Interior-Point Methods. Barrier methods require an interior point x_0, where $g(x_0) < 0$, and construct a sequence of unconstrained problems with $\mu > 0$

controlling the weight of the barriers. Barrier terms cannot be defined for equality constraints because μ increases tends to zero as the algorithm approaches the boundary to the feasible region. For example, the most popular barrier functions are of the form

$$\Lambda(g(x), \lambda, \mu) = \mu^{\mathrm{T}}\log(-g(x)) \text{ and } \Lambda(g(x), \lambda, \mu) = -1/\mu^{\mathrm{T}}g(x).$$

Originally proposed by Frisch (1955) and later developed by Fiacco and McCormick (1968), the logarithmic barrier method was applied extensively in the 1960s and 1970s and has become foundational for nonlinear interior-point methods. In 1984, interior point methods gained momentum with Karmarkar's polynomial-time LP. in the log-barrier approach, steps are generated by iteratively solving an unconstrained optimization problem similar to the form:

$$\min \quad f(x) - \mu_k^{\mathrm{T}}\log(-g(x))$$

where

$$\mu_k > 0, \text{ and}$$

$$\partial(\mu_k^{\mathrm{T}}\log(-g_i(x_k)))/\partial x_k = \mu_k^{\mathrm{T}}\nabla g_i(x_k)/g_i(x_k).$$

As $k \to \infty$, we have $\mu_k \to 0$ and $\mu_k/g_i(x_k) \to \lambda_i$ for all i.

Since the logarithm of a small number is negative, the $-\mu_k\log(-\mathrm{g_i}(\mathrm{x}_k))$ term is positive for any variable near the bound; this effectively creates barriers pushing the variables into the interior of the feasible region. Barrier methods suffer from ill-conditioning in the Hessian as the optimum is approached. More generally, interior point methods are often sensitive to the initialization and reduction of the barrier (or centering) parameter μ. Such barrier methods are also sensitive to the step size when a global convergence strategy is employed. However applying factorization and global convergence strategies specifically designed for the given barrier function can address some of the above complications.

In 1992, Polyak introduced a modified barrier method that scales the barrier function so that initial infeasibility is not a problem. Also in 1992, Mehrorta introduced a predictor-corrector mechanism to promote convergence by reducing the number of matrix factorizations when determining search directions. The method uses the Cholesky decomposition to find two different directions: a predictor and a corrector. The predictor computes an optimizing search direction based on a first order term. The corrector step uses the same Cholesky decomposition found during the predictor step. The search direction is the sum of the predictor direction and the corrector direction. By determining the centering parameter adaptively (rather than prior to the affine-scaling direction) and enabling corrections in the predicted direction, Mehrotra's approach introduces extra computation per iteration but significantly reduces the number of iterations

required. In 1996, Gondzio introduced multiple centrality corrections strategy that maximizes the possible step size increase in Mehrotra's approach by computing up to several corrections, which leads to an even larger decrease in the residuals. Most interior point approaches have leveraged the predictor-corrector method. Although there is no theoretical complexity bound, the method is widely used in practice. It appears to be computationally efficient and to converge very fast when close to the optimum.

Primal-Dual Methods. Primal-dual approaches apply (quasi-)Newton's method, and similar to other iterative methods we've covered thus far, includes a search step and direction, and a convergence criteria.
Consider the primal program

$$\min_{x,y} f(x)$$
subject to
$$g(x) + y = 0$$
$$y \geq 0$$

with the Lagrangian

$$L(x, s, \lambda, \mu) = \mathrm{f}(x) + \lambda^{\mathrm{T}}(g(x) + y) - \mu^{\mathrm{T}} y.$$

The first order conditions are

$$\nabla f(x) + \lambda^{\mathrm{T}} \nabla g(x) = 0\,,$$

$$\mu - \nu = 0,$$

$$g(x) + y = 0\,,$$

$$y \geq 0\,,$$

$$\mu \geq 0,$$

$$\nu_i y_i - 1/t = 0, \text{ and}$$

$$\nu \perp y = 0$$

where $\perp$ is element-by-element multiplication and $t \to 0$.
The optimization problem is

$$\min \nu y$$
subject to
$$\nabla f(x) - \lambda^{\mathrm{T}} \nabla g(x) = 0$$
$$\mu - \nu = 0$$
$$g(x) - y = 0$$
$$y \geq 0$$
$$\mu \geq 0.$$

Therefore, x, s, λ, μ and t are updated after each Newton step. Primal-dual interior point methods allow for primal and dual iterates to be infeasible, e.g. $g(x) + s \neq 0$.

Sequential Linear Programming (SLP) Methods. The SLP method determines the solution to a nonlinear problem through solving a sequence of linear approximations, mostly using first-order Taylor series expansions. For each iteration k, we solve the linear program

$$x_{k+1} = \text{argmin } \{\nabla f(x_k)^T d \mid \nabla g(x_k)^T d + g(x_k) \geq 0,\ d^l \leq d \leq d^u\}$$

where $\nabla f(x_k)$ can be replaced with linearized generalized Lagrangians. Since the linearizations are not necessarily bounded, trust regions where $x_{k+1} \in TR_k$ can be used to ensure convergence. Furthermore, this approach may be used in conjunction with a penalty or merit function and step restrictions. If the problem is convex, the linear constraints are always outside the feasible region to the nonlinear problem, but are guaranteed to achieve convergence as $k \rightarrow 0$.

The SLP can handle very large problems, benefits from using LP solvers, but may converge slowly and may violate nonlinear constraints. SLP methods are suitable for solving large-scale nonlinear programming problems, since large linear programs can be solved efficiently. Because of the linearity of the approximation, special scarcity patterns of the Jacobian matrix $\nabla g(x_k)$ are passed to the constraint matrix of the linear program directly. A particular advantage of sequential linear programming methods is that although second order information is not used, convergence is linear even in the case of highly nonconvex problems.

Sequential Quadratic Programming (SQP) Methods. This method is also referred to as a projected Lagrangian or Lagrange-Newton approach because the accompanying subproblem minimizes a quadratic approximation to the Lagrange function subject to a linearized constraint set.

SQP methods solve a sequence of quadratic subproblems to determine the active-set and a search direction d as the solution for each iteration. The key idea is to approximate second order information to get a fast final convergence speed. Thus, we define a quadratic approximation of the generalized Lagrangian function $\underline{L}(x, \lambda, \mu)$ and an approximation of the Hessian matrix by a quasi-Newton Newton where $B_k = \nabla^2{}_x \underline{L}(x_k, \lambda_k, \mu_k)$. Then we obtain the subproblem:

$$\begin{aligned}
&\min\ d_k^T B_k d_k/2 + \nabla f(x_k)^T d_k \\
&\text{subject to} \\
&\nabla g(x_k)^T d_k + g(x_k) = 0 \\
&d_k \in TR_k
\end{aligned}$$

The KKT conditions are applied to the objective of the subproblem to solve for d_k. A merit function is typically the sum of the objective function and the amount of infeasibility of the constraints.

A SQP usually requires fewer functions and gradient evaluations than a SLP, but may violate nonlinear constraints and is harder to solve than a SLP. Another SQP based approach, the sequential linear-quadratic program (SLQP), decouples the active-set identification and the step computation in the QP subproblem. SLQP omits the quadratic term in the objective function and solves the subproblem for a trust region. With only the active-set of constraints, a subsequent linear program is solved to determine d_k.

Generalized Reduced Gradient Methods. In 1961, Rosen introduced a gradient projection method. In 1969, Abadie and Carpentier introduced the generalized reduced gradient (GRG) for nonlinear constraints combining quasi-Newton methods and Wolfe's reduced gradient method. The actual implementation has many modifications to make it efficient for large models see in Drud (1985 and 1992). There are several GRG variations including choices of Lagrangians and line searches. We start with the problem:

$$\min f(x)$$
$$\text{subject to}$$
$$g(x) = 0$$
$$0 \le x \le x^u$$

At major iteration k, we solve the following nonlinear program:

$$\text{GRG}_k\text{: } \min F(x, \lambda_k, \mu_k)$$
$$\text{subject to}$$
$$g_k = g(x_k) + \nabla g(x_k)^T(x - x_k) = 0 \qquad \text{(dual variable } \lambda_{k+1})$$

where

$$F(x, \lambda_k, \mu_k) = f(x) - \lambda_k^T(g(x) - g_k) + \mu_k(g(x) - g_k)^T(g(x) - g_k), \text{ and}$$

μ_k is a scalar penalty parameter.

For the set of basic (dependent) variables B, the set of superbasic (independent) variables S, and the set of nonbasic (dependent, fixed at a bound) variables N we can partition $\nabla g(x_k)$ as

$$\nabla g(x_k) = [B, S, N]$$

where B and S are nonsingular, and we can partition x_k as

$$x_B = (x_k)_B,$$

$x_S = (x_k)_S$, and

$x_N = (x_k)_N$

where

$$B^T x_B + S^T x_S + N^T x_N = 0.$$

At a solution, the basic and superbasic variables are between their bounds, $0 \leq x_B \leq x^u_B$ and $0 \leq x_S \leq x^u_S$ while nonbasic variables will normally be equal to their lower or upper bound. At a solution, S will be no more than the number of nonlinear variables and x_s is regarded as a set of independent variables that are allowed to move in a direction that will decrease the objective function value. The basic variables are easily adjusted to satisfy the linear constraints.

If no improvement can be made with the current definition of B, S and N, some of the nonbasic variables are selected to be added to S, and the process is repeated with an increased value of S. If a basic or superbasic variable encounters one of its bounds, the variable is made nonbasic and the value of S is reduced by one. A step of the reduced-gradient method is called a minor iteration. For linear problems, simplex method is the reduced-gradient method with the number of superbasic variable oscillating between 0 and 1.
The steps in a GRG algorithm are:

Step 1. Set $k = 0$ and start with a feasible solution, x_0.

Step 2. Compute $\nabla g(x_k) = J = [B, S, N]$, $\nabla f(x_k)$, and the reduced gradient, g_r. Select the set of superbasic variables, x_s, as a subset of the nonbasic variables that can be gainfully changed. For the superbasic variables

$$g_r(x_S) = \nabla f(x_k)^T - J^T \lambda_k$$

and for the basic variables

$$g_r(x_B) = 0 \ (B^T \lambda = \partial f / \partial x_b).$$

Step 3. Find a search direction, d_s, for the superbasic variables that is based on g_r and possibly some second order information. The reduced Hessian of f(x) may be approximated by solving a system of the form

$$R^T R q + Z \nabla f(x) = 0$$

where

$Z = [-B^{-1}S, I, 0]$,

$Z \nabla f(x)$ is the reduced gradient, and

$R^T R \approx Z^T H Z$.

The dense upper triangular matrix R is updated in various approaches in order to

$$\begin{bmatrix} (c_i^{\mathrm{T}}x + d_i)I & A_ix + b_i \\ (A_ix + b_i)^{\mathrm{T}} & (c_i^{\mathrm{T}}x + d_i)I \end{bmatrix} \succeq 0, \quad i = 1, \ldots, m$$

approximate $\nabla^2 f(x)$. Therefore, $d_s = Zq$ where g_r is applied to solve the line search $\min_{0<\alpha<\beta} F(x + \alpha d_s, \lambda_k, \mu_k)$.

Step 4. Solve the subproblem GRG_k; let $(x_{k+1}, \lambda_{k+1}\, \mu_{k+1}) = \text{argmin}\ \{GRG_k\}$.

Step 5. If g_r projected on the bounds is smaller than the convergence criteria, then stop; if not, increment $k = k + 1$ go to Step 2.

Generalized reduced gradient methods can be extended easily to very large problems and problems with special structure. GRG is probably more robust than SLP and SQP methods, and once a feasible solution is found in GRG methods, it remains feasible. Moreover, GRG approaches are related to SQP methods and therefore there exists combinations of both approaches.

Conic and Semidefinite Programming Methods. These methods are forms of quadratic programming, and are similar to linear programming because they are convex programs. For linear program

$\min_x c^{\mathrm{T}}x$
subject to
$Ax = b$
$x \geq 0$

the equivalent conic program is

$\min_x c^{\mathrm{T}}x$
subject to
$||A_ix + b_i||_2 \leq c_i^{\mathrm{T}}x + d_i,\ i = 1, \ldots, m$

and the equivalent semidefinite program is of the form

$\min_x c^{\mathrm{T}}x$
subject to

Geometrically, linear, conic, and semidefinite programs are of the form

$\min_x c^{\mathrm{T}}x$
subject to
$Ax + b \in K$

where K is a pointed[1] convex cone with a nonempty interior. Typically these methods are applied to avoid local minima by relaxing the objective function and domain to be convex.

5. Decomposition Techniques

Decomposition methods are applied for two purposes, to fit the decomposed problem into high-speed memory and to decompose the problem into easier to solve subproblems. Decomposition methods usually employ either column generation or cutting planes, where one approach is the dual of the other.

Bender's decomposition is a cutting plane method where the primal nonlinear problem is

$f^* = \min_{x,y} f(x) + c^T y$
subject to — dual variable
$g(x) + Ay \leq b$ — λ

and its dual

$L^* = \max_{\lambda} \min_{x,y} f(x) + c^T y + \lambda(g(x) + Ay - b)$.

The problem can be partitioned where, given x, we can solve for variable y using the following inner minimization of the primal

$\min_y c^T y$
subject to
$Ay \leq b - g(x)$

and maximization of the dual

$\max_{\lambda} (b - g(x))^T \lambda$
subject to
$\lambda A \geq c^T$.

The above inner minimization is a linear program such that

$\lambda_k = argmax_{\lambda} \{(b - g(x_k))^T \lambda \mid \lambda A \leq b, \lambda \geq 0\}$,
$f^* = \text{minimize}_x f(x) + y_0$, and
$y_0 \geq (b - g(x))^T \lambda_k \quad k = 0, 1, \ldots$
for $k = 0, 1, \ldots K$.

Bender's Algorithm:

Step 1. Set k =0; choose a feasible solution x_0.

[1] K is pointed if it does not contain any subspace except the origin.

Step 2. Solve $\lambda_{k+1} = \text{argmin}_{\lambda}\{(b - g(x_k))^T\lambda \mid \lambda A \geq c^T\}$.

Step 3. Solve $x_{k+1} = \text{argmin}_x\{f(x_k) + y_0 \mid y_0 \geq (b - g(x_k))^T\lambda_k\}$ where the inner minimization solves for y_0.

Step 4. If x_{k+1} satisfies the convergence criteria, stop; otherwise, set $k = k + 1$ and go to Step 2.

Dantzig-Wolfe solves a simpler or smaller subproblem that generates a column for the master problem:

$\min_{x,y} f(x) + c^T y$

subject to — dual variable

$g(x) + Ay \leq b$. — λ

$y \in Y$

Therefore the Dantzig-Wolfe algorithm is as follows:

Step 1. Set $k = 0$; choose a feasible solution (x_0, y_0) where $g(x_0) + Ay_0 \leq b$.

Step 2. Generate a column and solve the modified master problem for y and u_k:

$\min_{uk,y} \sum_k f(x_k)u_k + c^T y$

subject to — dual variable

$\sum_k g(x_k)u_k + Ay \leq \lambda_k$

$\sum_k u_k = 1$ — β_k

$u_k \geq 0$

$y \in Y$

Step 3. Solve the subproblem $x_{k+1} = \text{argmin}_x\{f(x) + \lambda_k g(x)\}$.

Step 4. If $\underline{x}_k = \sum_k x_k u_k$ satisfies the convergence criteria, stop; else $k = k + 1$ and go to Step 2.

Equivalent to both steps 2 and 3, the dual can be solved by generating a cutting plane:

$\max_{\lambda k,y} \lambda_k b + \beta_k$
subject to
$\lambda_k A = c^T$
$\lambda_k g(x_k) + \beta_k \geq f(x_k)$
$\lambda_k \geq 0$.

Then in Step 4, If $f(x)$ and $g(x)$ are convex, $\sum_k f(x_k)\, u_k \geq \sum_k f(\underline{x}_k)$ and $\sum_k g(x_k)\, u_k \geq \sum_k g(\underline{x}_k)$. Furthermore, if $u_0^*, \ldots u_k^*$ and y_k^* is a feasible solution then $\underline{x}_k$ is a feasible solution and as $k \to \infty$, we have that $\underline{x}_k$ and y_k^* is an optimal to the NLP.

Approximations of the ACOPF. The ACOPF can be formulated in several ways (see Cain et al. 2012). The canonical ACOPF in polar coordinates is:

$$\min \sum_n c_n(p_n) + c_n(q_n)$$
subject to
$$p_n - p^d{}_n = \sum_m v_n v_m (g_{nm} cos\theta_{nm} + b_{nm} sin\theta_{nm})$$
$$q_n - q^d{}_n = \sum_m v_n v_m (g_{nm} sin\theta_{nm} - b_{nm} cos\theta_{nm})$$
$$p_n^{min} \leq p_n \leq p_n^{max}$$
$$q_n^{min} \leq q_n \leq q_n^{max}$$
$$v_n^{min} \leq v_n \leq v_n^{max}$$
$$\theta^{min}{}_{nm} \leq \theta_{nm} = \theta_n - \theta_m \leq \theta^{max}{}_{nm}$$

Please refer to Cain et al. (2012) for notation definitions and the equivalent ACOPF formulation in rectangular coordinates. For the real and reactive power flow equations, the first partial derivatives on the power injections p_n and q_n are:

$$\partial p_n/\partial v_m = v_n(g_{nm} cos\theta_{nm} + b_{nm} sin\theta_{nm}) \qquad n \neq m$$
$$\partial p_n/\partial \theta_m = v_n v_m(g_{nm} sin\theta_{nm} - b_{nm} cos\theta_{nm}) \quad n \neq m$$
$$\partial p_n/\partial v_n = \sum_{m(\neq n)} v_m(g_{nm} cos\theta_{nm} + b_{nm} sin\theta_{nm}) + 2v_n g_{nn}$$
$$\partial p_n/\partial \theta_n = \sum_{m(\neq n)} v_n v_m(-g_{nm} sin\theta_{nm} + b_{nm} cos\theta_{nm}) - v_n^2 b_{nn}$$
$$\partial q_n/\partial v_m = v_n(g_{nm} sin\theta_{nm} - b_{nm} cos\theta_{nm}) \quad n \neq m$$
$$\partial q_n/\partial \theta_m = -v_n v_m(g_{nm} cos\theta_{nm} + b_{nm} sin\theta_{nm}) \quad n \neq m$$
$$\partial q_n/\partial v_n = \sum_{m(\neq n)} v_m(g_{nm} sin\theta_{nm} - b_{nm} cos\theta_{nm}) - 2v_n b_{nn}$$
$$\partial q_n/\partial \theta_n = \sum_{m(\neq n)} v_n v_m(g_{nm} cos\theta_{nm} + b_{nm} sin\theta_{nm}) - v_n^2 g_{nn}$$

The first partial derivatives are used in approximation methods. Note that in these derivatives the magnitude of the cosine and sine terms are generally scaled by the multiple of the conductance or susceptance. Since reactance of a transmission line is much larger than resistance, the conductance (real part of admittance matrix, B) is much smaller than the susceptance (imaginary part of admittance matrix, G). In most cases the sine terms tend to be small whereas the cosine term is near unity. Upon inspection, the result of these physical properties results in real power that is highly sensitive to changes in voltage angle and reactive power that is highly sensitive to changes in voltage magnitude.

Decoupled Power Flow Model. Initially proposed by Stott and Alsac in 1974, the Decoupled Power Flow (DPF) is based on the principle that when the Jacobian of the power flow equations

$$\begin{bmatrix} \partial p/\partial\theta & \partial p/\partial v \\ \partial q/\partial\theta & \partial q/\partial v \end{bmatrix}$$

is evaluated numerically, the off-diagonal submatrices are much smaller in magnitude than the diagonal submatrices:

$\partial p/\partial\theta >> \partial q/\partial\theta$ and
$\partial q/\partial v >> \partial p/\partial v$.

We therefore set the off-diagonal entries of the Jacobian to zero

$\partial q/\partial\theta \approx 0$ and
$\partial p/\partial v \approx 0$,

and then decompose the problem into a pair of subproblems where the p-θ real power model minimizes the system costs and the q-v reactive power model minimizes the real power transmission losses. In this approach we assume that $\theta_{nm} \approx 0$, $sin\theta_{nm} \approx \theta_{nm}$, $cos\theta_{nm} \approx 1$, and $g_{nm} << b_{nm}$. Let

$$b'_{nm} = [\partial p_n / \partial\theta_m] = -v_n v_m b_{nm} \quad n \neq m,$$
$$b'_{nn} = 0,$$
$$b''_{nm} = [\partial q_n / \partial v_m] = -v_n b_{nm} \quad n \neq m, \text{ and}$$
$$b''_{nn} = -2v_n b_{nn}$$

where b' is an approximation of the matrix of partial derivatives of the real power flow equations with respect to the bus voltage phase angles and b'' is an approximation of the matrix of partial derivatives of the reactive power flow equations with respect to the bus voltage magnitudes.

The Decoupled Power Flow model is:

$\min \sum_n c_n(p_n) + c_n(q_n)$
subject to
$p_n - p^d_n = \sum_m p_{nm} + s^p_n$
$q_n - q^d_n = \sum_m q_{nm} + s^q_n$
$p^{min}_{nm} \leq p_{nm} = b'_{nm}\theta_{mn} \leq p^{max}_{nm}$
$q^{min}_{nm} \leq q_{nm} = b''_{nm}v_m \leq q^{max}_{nm}$
$p^{min}_n \leq p_n \leq p^{max}_n$
$q^{min}_n \leq q_n \leq q^{max}_n$
$v^{min}_n \leq v_n \leq v^{max}_n$
$\theta^{min}_{nm} \leq \theta_{nm} = \theta_n - \theta_m \leq \theta^{max}_{nm}$

where

$$p_n = \sum_m v_n v_m b_{nm} \theta_{nm}$$
$$q_n = \sum_m -v_n v_m b_{nm}.$$

The Decoupled Power Flow algorithm:

Step 1. Set $k = 0$, and initialize $\underline{v}_{nk} = 1$.

Step 2. Solve DPF for p_{nk}, θ_{nk}, q_{nk}, and v_{nk}.

Step 3. If p_{nk}, θ_{nk}, q_{nk}, and v_{nk} satisfy the convergence criteria, e.g., ac feasibility, then stop. If not, set $b'_{nm} = \underline{v}_{nk}\underline{v}_{mk} b_{nm}$ and $b''_{nm} = \underline{v}_{nk} b_{nm}$ and $k = k + 1$, and go to Step 2.

Since the underlying problem is nonconvex, there is no guarantee of convergence in the original ACOPF problem.

Bθ Model. A further simplification drops reactive power completely. If $g_{ij} = 0$, $\sin\theta = \theta$, $\cos\theta = 1$, and $v_m = 1$, the Btheta models solves:

$$\min \sum_n c_n(p_n)$$
subject to
$$p_n - p^d_n = \sum_m p_{nm}$$
$$p^{min}_{nm} \le p_{nm} = b_{nm}\theta_{mn} \le p^{max}_{nm}$$
$$p^{min}_n \le p_n \le p^{max}_n$$
$$\theta^{min}_{nm} \le \theta_{nm} = \theta_n - \theta_m \le \theta^{max}_{nm}$$

Distribution Factor Model. A further simplification is the distribution factor (DF) model where all transactions are decomposed into a 'sale' to a reference node and 'purchase' from a reference node.

$$\min \sum_n c_n p_n$$
subject to
$$\sum_n p_n - \sum_n p^d_n = 0$$
$$p^{min}_n \le p_n \le p^{max}_n$$
$$p_{nmk} = \sum_n df_{kn}(p_n - p^d_n) \le p^{max}_{nmk}$$

Since we have assumed losses are zero, superpositioning makes DF roughly equivalent to the Bθ formulation if θ^{max}_{nm} is equivalent to p^{max}_{nmk}. The difference is the size of the problem; whereas Bθ has N + 2K constraints, the distribution factor model has 1 + K^M constraints where M is the subset of lines monitored for binding thermal limits.

6. Commercial Solvers

Commercial optimization solvers for nonlinear problems vary in algorithmic techniques and implementation. We survey five well-known nonlinear solvers: MINOS, IPOPT, SNOPT, KNITRO, and CONOPT. Please refer to (Castillo, 2013) for a computational study on the performance of these solvers applied to the ACOPF.

MINOS. MINOS is a GRG method and is designed to solve large-scale optimization problems (Murtagh and Saunders, 1982 and 2003). MINOS partitions variables into linear and nonlinear elements and then iteratively solves the subproblems with linearized constraints and an augmented Lagrangian objective function (Robinson, 1972). Instead of a nonlinear conjugate-gradient approach, a quasi-Newton approach is applied to certain subspaces. Sparse LU basis factors are maintained by LUSOL (Gill et at, 1987). The nonlinear constraints may be satisfied only in the limit, so that feasibility and optimality may occur simultaneously. An important feature is a stable implementation of a quasi-Newton algorithm for optimizing the superbasic variables.

IPOPT (Interior Point OPTimizer). IPOPT converts the problem to a barrier problem. Ideally the functions f(x) and g(x) are twice continuously differentiable. IPOPT uses line filtered searches and includes a feasibility restoration phase. Filter methods promote global convergence through measuring decreases in the objective function and infeasibility as two separate criteria that are controlled simultaneously. IPOPT has options in line search strategies for globalization, including an exact penalty merit function, augmented Lagrangian merit function, filter method (R. Fletcher, S. Leyffer, and P. Toint), Hessian and several Hessian approximation methods.

SNOPT (Sparse Nonlinear OPTimizer). SNOPT implements a sparse active-set sequential quadratic programming (SQP) method that employs quasi-Newton approximations to determine the Hessian in the quadratic programming subproblem; then an augmented Lagrangian merit function guides the line search direction. Sparse basis factors are maintained by LUSOL. If only the objective is nonlinear, the problem is linearly constrained (LC) and tends to solve more easily than the general case with nonlinear constraints (NC). SNOPT uses limited-memory quasi-Newton approximations to the Hessian of the Lagrangian. The merit function for step-length control is an augmented Lagrangian, as in the dense SQP solver NPSOL (Gill et at, 1992). In general, SNOPT requires less matrix computation than NPSOL and fewer evaluations of the functions than the nonlinear algorithms in MINOS. It is most efficient if only some of the variables enter nonlinearly, or there are relatively few degrees of freedom at a solution (i.e., many constraints are active).

KNITRO. KNITRO implements both interior-point and active-set methods for solving nonlinear optimization problems. The variables can be continuous, binary, or integer. In the barrier/interior method, KNITRO solves a series of barrier sub-problems controlled by a barrier parameter. The algorithm uses trust regions and a merit function to promote convergence. The algorithm performs one or more minimization steps on each barrier problem, then decreases the barrier parameter, and repeats the process until the original problem has been solved to the desired accuracy. KNITRO provides two procedures for computing the steps. One version where each step is computed using a projected conjugate gradient iteration. This approach factors a projection matrix to approximately minimize a quadratic model of the barrier problem. The other procedure always attempts to compute a new iterate by solving the primal-dual KKT matrix using direct linear algebra, but if the step quality cannot be guaranteed or if negative curvature is detected, then the new iterate is computed by the first procedure. KNITRO also implements an active-set sequential linear-quadratic programming (SLQP) algorithm that uses linear programming sub-problems to estimate the active-set at each iteration. KNITRO includes an active-set SLQP with trust region to promote convergence.

CONOPT (CONstrained OPTimization). CONOPT is a generalized reduced gradient (GRG) method that searches along the steepest descent direction in the superbasic variables. The generalized projection method projects the search direction into the subspace tangent to the active-set of constraints. The active-set is the subset of equality and inequality constraints at a point that satisfy with equality; for example, equality constraints are active at all feasible points. Therefore, the search direction is projected into the null space of the gradients for the equality and binding inequality constraints. The projected gradient could be infeasible, which then requires a correction step.

7. History of ACOPF Solution Techniques

In 1962, Carpentier introduced the alternating current optimal power flow (ACOPF) for economic dispatch based upon Karush-Kuhn-Tucker (KKT) conditions. Carpentier employed the Gauss-Seidel method representing the load flows as power injections in the voltage polar form. Carpentier included operational constraints on real power control, generator bus voltage magnitude limits, reactive power control of switchable VAR sources, and transformer tap setting. This formulation has nonconvex equality constraints with quadratic and trigonmetric functions (see Cain et al, 2012). The complex variables can be expressed in polar or rectangular coordinates and results in different types of nonconvex constraints. Since the ACOPF problem is a nonconvex mathematical program, the KKT conditions yield

only a local optimal solution; even feasibility cannot be guaranteed by most nonlinear programming (NLP) algorithms.

In 1969, Carpentier and Abadie published the generalized reduced gradient method, which is a generalization of Carpentier's 'differential injections' method that was originally conceived in 1964 to solve the optimal power flow problem (Carpentier, 1979). Since Carpentier's contributions, there has been a wealth of research done on algorithmic methods to solve the ACOPF. In 1968, Dommel and Tinney presented an approach based on power flow solution by Newton's method with a gradient adjustment factor for the penalty function to account for dependent constraints. In 1969, Sasson used the Fletcher-Powell method; then in 1973, Sasson directly computed the Hessian and utilized its sparsity. In 1973, Alsac and Stott incorporated exact outage-contingency constraints and used an augmented Lagrangian. In 1977, Biggs and Laughton used a recursive equality quadratic program. In the above studies, test problems with less than 30-buses were used. During this time the advancements of gradient techniques with Newton's method were widely adopted in the early ACOPF algorithms, see Happ (1977). In 1979, Wu et al. used a modified reduced gradient with penalty function methods and included the largest test case published to-date, a 1410-bus system.

In the following table we summarize the published optimal power flow studies by the test problems, convergence criteria, initialization, and performance metrics reported. The studies vary in all the metrics listed in Table 1, and very few of the studies included a comparison against commercial solvers or previously published data. For example out of the studies listed, Burchett et al. (1982), Aoki and Kanezashi (1985), Huneault and Galiana (1990), Ponrajah and Galiana (1990), Momoh and Zhu (1999), and Wang et al. employed MINOS in their algorithmic implementation or compared computational results to that of MINOS.

The abbreviations in Table 1 are as follows: N/A (Not Applicable), NR (Not Reported), FS (Flat Start), LFS (Load Flow Start), RS (Random Start), HS (Hot Start), WS(Warm Start), US (User Specified Start), C.G. (Complementarity Gap), Opt. (Optimality Tolerance), and Feas. (Feasibility Tolerance). Note for the SDP methods in Bai (2008) and Lavaei (2012), initialization is not required.

The studies with US (User Specified) listed as the initialization method employed a non-random, selective approach. Numerous studies reported additional metrics to those listed in Table 1. However, we identify this handful of metrics as basic information needed for fair testing and comparison of proposed algorithms by other researchers.

The following Table 2 summarizes the test problems published on in ACOPF studies to-date. Numerous of the test problems have a network origin that is not described well. Although many of the test problems are used across multiple studies, as denoted in the "Source Data" column a small percentage of the test

problems are readily available in the public domain to use; these test problems are posted online by the University of Washington Electrical Engineering or by MATPOWER. The raw data for test problems 23 and 39 are available in Biggs (1977) and Pai (1989), respectively. We list the referencing study if the reader is interested in further information on the test problem or to contact the author for the source data.

Table 1. ACOPF Studies that include Numerical Analysis

Year	Author	Test Problems (No. of Nodes)	Convergence Criteria	Initial-ization	Work-station	Code Lang.	CPU Time	Iter-ations
1969	Sasson	30	10^-5 Opt.	NR	x	x	x	x
1973	Alsac	30	10^-3 Opt.	US				x
1974	Mukherjee	25	NR	US	x		x	x
1977	Biggs	23	NR	US				x
1979	Wu	11, 136, 333, 1410	NR	LFS	x		x	x
1982	Burchett	118, 597	NR	NR	x		x	x
1982	Divi	9, 10, 11	10^-5 Feas.	FS		x		x
1982	Shoults	5, 30, 962	NR	US	x	x	x	x
1984	Burchett	350, 1100, 1600, 1900	NR	RS	x		x	x
1984	Sun	912	NR	US	x			x
1985	Aoki	14, 135	NR	NR	x	x	x	x
1988	Santos	118, 129	NR	NR	x	x		x
1989	Nanda	14, 30, 89	10^-4 Feas.	LFS			x	x
1989	Ponrajah	6, 10, 30, 118	NR	FS	x	x	x	x
1990	Alsac	1330, 1200, 700	NR	HS	x		x	x
1990	Huneault	30, 118	NR	US			x	x
1990	Salgado	14, 30, 39, 57, 89, 118	10^-3 Feas.	LFS	x	x	x	
1993	Almeida	14, 30, 34	NR	FS,LFS			x	x
1994	Momoh	9, 14, 30, 118	NR	US	x	x	x	x
1994	Wu	9, 30, 39, 118, 244	10^-6 C.G.	FS	x	x	x	x
1995	Chebbo	706	50^-5 Opt., 50^-4 Feas.	US	x		x	x
1997	Lai	30	NR	US	x		x	
1998	Torres	30, 57, 118, 300	10^-4 Opt.	FS,LFS		x		x
1998	Wei	14, 30, 57, 118, 344, 703, 1047	10^-6 Opt.	US	x	x	x	x

Year	Author	Test Problems (No. of Nodes)	Convergence Criteria	Initial-ization	Work-station	Code Lang.	CPU Time	Iter-ations
1999	Momoh	14	10^-6 Opt.	US			x	x
1999	Yan	118, 1062	10^-8 Opt., Feas.	US	x		x	x
2000	Nejdawi	30, 57, 118, 300	NR	FS				x
2000	Torres	30, 57, 118, 300	10^-4 Opt.	US	x	x	x	x
2001	Castronuovo	118, 352, 750, 1704	10^-6 Opt., Feas.	NR	x		x	x
2001	Lima	810, 2256	10^-3 Feas.	US			x	
2001	Torres	118, 256, 300, 555, 2098	10^-4 Opt.	US	x	x	x	
2001	Xie	30, 57, 118	10^-5 C.G., Feas.	WS				x
2002	Jabr	14, 24, 30, 57, 118, 175, 300	10^-4 Opt.	US	x	x	x	x
2002	Lin	118, 244	10^-3 Opt.	US			x	
2002	Torres	30, 57, 118, 256, 30, 555, 2098	10^-4 Opt.	US	x	x	x	x
2003	Oliveira	30, 118, 1564, 1732, 1993	Sqrt of machine eps.	US	x	x	x	x
2004	Lin	118, 244	10^-3 Opt.	US			x	
2005	Min	14, 30, 57, 118, 254, 300, 662	10^-3 Opt.	LFS	x		x	x
2005	Tate	2, 118, 10274	10^-3 Feas.	FS				x
2007	Capitanescu	60, 118, 300	10^-6 C.G.	RS	x		x	x
2007	Lin	9, 14, 30, 57, 118	10^-4 C.G., Feas.	FS	x	x	x	
2007	Sousa	30, 57, 118, 300	10^-4 Opt.	US		x		x
2007	Wang	30, 57, 118, 300, 2383, 2935	NR	FS	x	x	x	x
2008	Bai	4, 14, 57, 118, 300	10^-5 Opt.	N/A	x		x	x
2008	Jabr	9, 14, 30, 39, 57, 118, 300, 2383	10^-8 Opt.	FS,LFS	x	x	x	x
2008	Lin	9, 14, 30, 57, 118, 300	10^-4 C.G., Feas.	FS	x	x	x	x
2009	Bedrinana	2, 14, 57	NR	NR				x
2009	Chiang	678, 2052, 2383	10^-6 C.G., Feas.	FS	x		x	x

Year	Author	Test Problems (No. of Nodes)	Convergence Criteria	Initial-ization	Work-station	Code Lang.	CPU Time	Iter-ations
2009	Jiang	14, 30, 39, 57, 118, 300, 701, 2052, 2383, 2736, 2746	10^-5 C.G., Feas.	FS	x	x	x	x
2009	Sousa	30, 118, 300, 2256	10^-3 Feas.	NR	x	x	x	x
2009	Yang	30, 118, 300	NR	NR		x	x	
2010	Jiang	5, 6, 9, 14, 30, 39, 57, 118, 2383	NR	FS	x	x	x	x
2010	Jiang et al.	118, 300, 678, 2052, 2383, 2746	NR	NR	x	x	x	x
2010	Xie	57, 118, 300, 2052, 2790	10^-6 C.G.	US	x	x	x	x
2011	Chung	14, 118	10^-4 Feas.	FS	x	x	x	x
2011	Phan	6, 9, 14, 30, 39, 57, 118, 300, 2746	10^-5 Opt.	US	x	x	x	
2011	Sousa	30, 57, 118, 300, 1211	10^-4 Opt.	FS,LFS,RS		x		x
2011	Zimmerman	9, 30, 36, 118, 300, 2383, 2736, 3120, 2935, 21000, 42000	NR	NR	x	x	x	
2012	Lavaei	14, 30, 57, 118, 300	NR	N/A		x		

Table 2. Test Problems used in ACOPF Studies

Test Problem	Network Origin	Source Data	Referencing Study
14	Midwestern US System	UW Electrical Eng	Multiple studies
23	N/A	Biggs, 1977	Biggs, 1977
24	IEEE Reliability Test System	UW Electrical Eng	Multiple studies
25	N/A	N/A	Mukherjee, 1974
30	Midwestern US System	UW Electrical Eng	Multiple studies
34	N/A	N/A	Almeida, 1993
39	New England US System	Pai, 1989	Phan, 2011
57	Midwestern US System	UW Electrical Eng	Multiple studies
60	Nordic32 System	N/A	Capitanescu, 2007
89	N/A	N/A	Salgado, 1990
118	Midwestern US System	UW Electrical Eng	Multiple studies
129	Companhia Energetica de Sao Paulo	N/A	Santos, 1988
135	Chugoku Electric Power Co., Japan	N/A	Aoki, 1985
136	N/A	N/A	Wu, 1979
175	N/A	N/A	Jabr, 2002
244	N/A	N/A	Wu, 1994 (Initial study)
254	N/A	N/A	Min, 2005
256	N/A	N/A	Torres, 2001 and 2002
300	Midwestern US System	UW Electrical Eng	Multiple studies
333	N/A	N/A	Wu, 1979
344	Japan System	N/A	Wei, 1998
350	Northeastern US Utility	N/A	Burchett, 1984
352	South-Southeastern Brazil System	N/A	Castronuovo, 2001
555	N/A	N/A	Torres, 2001 and 2002
597	Interconnection of Several Utilities	N/A	Burchett, 1982
662	N/A	N/A	Min, 2005
678	N/A	N/A	Chiang, 2009

Test Problem	Network Origin	Source Data	Referencing Study
700	Interconnection of Several Utilities	N/A	Alsac, 1990
701	Undisclosed Real Power System	N/A	Jiang, 2009
703	China System	N/A	Wei, 1998
706	N/A	N/A	Chebbo, 1995
750	South-Southeastern Brazil System	N/A	Castronuovo, 2001
810	South-Southeastern Brazil System	N/A	Lima, 2001
912	Northeastern US System	N/A	Sun, 1984
962	16 Interconnected Areas	N/A	Shoults, 1982
1047	Simulation Power System	N/A	Wei, 1998
1062	N/A	N/A	Yan, 1999
1100	Eastern United States Pool	N/A	Burchett, 1984
1200	Utility Company	N/A	Alsac, 1990
1211	Undisclosed Real Power System	N/A	Sousa, 2011
1330	Interconnection of Several Utilities	N/A	Alsac, 1990
1410	N/A	N/A	Wu, 1979
1564	South-Southeastern Brazil System	N/A	Oliveira, 2003
1600	Western United States Utility	N/A	Burchett, 1984
1704	South-Southeastern Brazil System	N/A	Castronuovo, 2001
1732	South-Southeastern Brazil System	N/A	Oliveira, 2003
1900	Northeastern US Utility	N/A	Burchett, 1984
1993	South-Southeastern Brazil System	N/A	Oliveira, 2003
2052	N/A	N/A	Multiple studies
2098	Modified Brazil System	N/A	Torres, 2001 and 2002
2256	South-Southeastern Brazil System	N/A	Lima, 2001
2383	Polish System	Zimmerman, 2011	Zimmerman, 2011
2736	Polish System	Zimmerman, 2011	Zimmerman, 2011
2746	Polish System	Zimmerman, 2011	Zimmerman, 2011

Test Problem	Network Origin	Source Data	Referencing Study
2790	Undisclosed Real Power System	N/A	Xie, 2010
2935	Polish System	Zimmerman, 2011	Zimmerman, 2011
3120	Polish System	Zimmerman, 2011	Zimmerman, 2011
10274	N/A	N/A	Tate, 2005
21000	N/A	N/A	Zimmerman, 2011
42000	N/A	N/A	Zimmerman, 2011

Sequential quadratic programming (SQP) methods applied to the OPF were either linear programming (LP)-based QP methods or QP methods based on the concept of an active-set of linearly independent constraints. The LP approaches, based on Wolfe's or Beale's algorithm, solve a linear program by a revised simplex technique. Exploiting many of the strengths of linear programming, these approaches permit the use of artificial variables to guide feasibility, parametric programming, and ease of equality constraints into the formulation. Furthermore, some of the successive quadratic programs apply Newton's method to the subproblem. In 1984 by comparison, Burchett et al. present a method where the QP subproblems are equivalent to a sequence of Newton steps to the optimal solution.

In 1982 Shoults and Sun decomposed the problem into active and reactive subproblems. In 1984, Sun applied Newton's method to the SQP and advanced sparsity techniques to the decomposition, but the method exhibited problems in initialization and ill-conditioning. In 1986, Contaxis et al. and in 1989, Nanda et al. applied Fletcher's method to the decoupled subproblems.

In 1981, Giras et al. applied Powell's quasi-Newton method to the ACOPF, which performs a Broyden-Fletcher-Goldfarb-Shanno (BFGS) update and shows convergence from infeasible starting points. In 1982, Divi et al. applied Fletcher's quasi-Newton method with a shifted penalty function, and the Broyden-Fletcher-Shanno (BFS) updating formula to numerically stabilize the Hessian as a positive definite matrix and promote global convergence by means of a penalty function that ensures sufficient progress along the line-search towards the active-set of constraints. In 1982, Burchett et al. and in 1985 Aoki et al. solve a sequence of sparse, linearly constrained subproblems based on MINOS. In 1988, Santos proposed the dual augmented Lagrangian method that attempts to address ill-conditioning of the Hessian by applying a quasi-Newton method to the dual function. In 1992, Monticelli et al., present an improvement upon Maria et al. (1987) in which an adaptive penalty strategy is used to ensure positive definiteness of the Hessian, without negatively affecting the quadratic convergence characteristic of Newton's method.

These methods build upon Newton's method for unconstrained optimization, Lagrange's method for optimization with equalities, and Fiacco and McCormick's barrier method. Meliopoulos and Xia (1993), Vargas et al. (1992), Momoh et al. (1994), Momoh and Zhu (1999), and Nejdawi et al. (2000) applied an interior point algorithm (IPM) to a LP or QP, where the constraints are linearized. Moreover, primal-dual IPMs have been successful in solving the ACOPF by introducing a logarithmic barrier function in place of the inequality constraints. Wu et al. (1994), Wei et al. (1998), Torres and Quintana (1998), Yan and Quintana (1999), Castronuovo et al. (2001), Xie and Song (2001), Wang et al. (2007), Xie and Chiang

(2010 and 2011), Sousa et al. (2011), and Chung et al. (2011) present a primal-dual interior point method to solving the ACOPF.

In 1994, Granville solves the primal-dual without Mehrotra's predictor-corrector mechanism, and notes that the contribution of barrier terms into the diagonal of the Hessian matrix are very effective in bringing positive definiteness to the problem, therefore making auxiliary penalty functions unnecessary. In 1998, Torres et al. exploit the rectangular formulation of the power flow constraints, which is quadratic, using an interior point method. This results in a constant Hessian and a Taylor expansion terminating at the second-order term that reduces the computational burden and iterations to convergence. Furthermore, Torres et al. (1998) perturb the boundary to deal with numerical ill-conditioning that may occur with binding constraints. In 1998, Wei et al. use an interior-point method based on applying Newton's method to the nonlinear system of perturbed KKT conditions, which is similar to the approach by Torres and Quintana (2001); this method promotes global convergence. In 1998, Wei et al. present a data structure that further reduces the nodal block fill-in elements. In comparison to storing the augmented system in compact blocks, as noted by Sun et al. (1984) and Wei et al. (1998), Castronuovo et al. (2001) propose a vectorization technique that only considers nonzero terms in order to decrease computational cost per operation. In 1999, Yan and Quintana (1999) present a dynamic adjustment of the step size and tolerance to improve convergence speed.

Wu et al. (1994), Torres and Quintana (1998), Yan and Quintana (1999), Castronuovo et al. (2001), Lin et al. (2007), Wang et al. (2007), Lin et al. (2008), Xie and Chiang (2010 and 2011), Sousa et al. (2011) and Chung et al. (2011) use Mehrotra's predictor-corrector. In 2009, Sousa et al. present a predictor-corrector modified barrier approach, based on Polyak's modified barrier method, to address ill-conditioning. However, the corrector step can lead to very slow convergence or failure. In 2001, Torres and Quintana apply Gondzio's multiple centrality corrections strategy. Furthermore, Xie and Song (2001) and Chung et al. (2011) leverage a nodal block data structure within the interior point method iterates in order to reduce the correction equation.

Due to the nonlinearities and ill-conditioning of the ACOPF, recent research has focused on applying methods with particularly robust global convergence properties. Jabr et al. (2002) present a primal-dual interior point method that replaces the Hessian with a "2-norm positive approximant" with a "watch-dog" strategy. Chiang et al. (2009) present a two-stage solution algorithm where an active-set quotient gradient method is applied in stage one to induce global convergence, and an interior-point method is applied in stage two to obtain a local optimal solution. Min and Shengsong (2005), Pajic and Clements (2005), Zhou et al. (2005), Sousa and Torres (2007), Wang et al. (2007), Chiang et al. (2009), and

Sousa et al. (2011) apply a trust region method. Similar to Torres and Quintana (2001), Min and Shengsong (2005) also apply Gondzio's multiple centrality corrections strategy to the primal-dual IPM, but instead incorporate the trust-region method in solving the subproblems. The infinity norm, which forms a closed, compact, convex hypercube in n-dimensional space, is used to define the trust region, and a merit function is used to determine whether to accept or reject a trial step to the trust region subproblem. The authors also introduce a feasibility restoration variable, which controls the feasibility requirements of the trust region subproblem.

Also known as parametric continuation methods, homotopy methods are path-following approaches which explicitly progress towards a solution to the original nonlinear problem. In doing so these methods construct and solve a new simpler problem compared to the original one and then gradually reconstruct the original system of equations in order to solve for the unknown solution. Ponrajah and Galiana (1989), Huneault and Galiana (1990), Almeida et al. (1993), Lima et al. (2001), and Jiang et al. (2010) apply homotopy methods to solve the ACOPF.

Most recently, convex optimization techniques have been applied to the ACOPF. In 2007, Jabr reformulated the ACOPF as a second-order conic program and applied an interior point method in MOSEK that is specific to conic quadratic optimization that has polynomial convergence. In 2012, Lavaei and Low apply semidefinite programming optimization to solve the ACOPF, and prove global optimality under a sufficient zero-duality gap condition.

Another class of optimization, derivative free optimization, is typically applied when first- and second-derivatives are not available or are expensive to compute. Lai et al. (1997), Iba (1994), Abido (2002) and many others apply derivative free approaches to solve the OPF.

8. Summary and Conclusions

For the last fifty years, the latest developments in nonlinear optimization have been applied to the ACOPF in hopes of better solution techniques for large-scale, practical network operations and planning. Until recently, the dominant formulation has been in polar coordinates. Although the research to-date presents a relatively positive picture, clearly the published experimental results have been limited. The lack of reported metrics and available test problems makes it difficult to perform a comparative assessment of proposed solution techniques to-date. Furthermore, there is a significant lack of independent studies that compare the numerous approaches in a systematic manner; for example, refer to the benchmarking study completed by Mittelman (2012) where he reports on solver performance for optimizing general nonlinear problems. In a companion study

Castillo and O'Neill (2013) report numerical results from testing CONOPT, IPOPT, KNITRO, MINOS, and SNOPT on various sized test problems in which we apply various mathematically equivalent AC-OPF formulations with numerous starting points.

Although an approximation of the ACOPF is solved in practice, current market operations and planning do not co-optimize real power dispatch with voltage and reactive power management. A recent white paper by Stott and Alsac (2012) argue that the ACOPF is still very much a 'work in progress.' In fact, the authors contend that most of the work completed on ACOPF solution techniques to-date is not adaptable to real world applications. The ACOPF would often be embedded within bigger calculations that make convergence and computational effort even more difficult. Therefore in developing and testing future solution techniques, it is important to understand how market operations and planning would leverage the ACOPF in real world applications.

References

J. Abadie, J. Carpentier. "Generalization of the Wolfe Reduced Gradient Method to the Case of Nonlinear Constraints" in Optimization by R. Fletcher, Pp. 37-49, Academic Press, London, 1969.

M.A. Abido. "Optimal Power Flow Using Particle Swarm Optimization." International Journal of Electrical Power & Energy Systems, Vol. 24, No. 7, Pp. 563-571, 2002.

K.C. Almeida, F.D. Galiana, S. Soares. "A General Parametric Optimal Power Flow." Power Industry Computer Application Conference Proceedings, 1993.

O. Alsac, B. Stott. "Fast Decoupled Load Flow." IEEE Transactions on Power Apparatus and Systems, PAS-93, Pp. 859-869, May 1974.

O. Alsac, B. Stott. "Optimal Load Flow with Steady-State Security." IEEE PES Summer Meeting, July 15-20, 1973.

O. Alsac, J. Bright, M. Prais, B. Stott. "Further Developments in LP-Based Optimal Power Flow." IEEE Transactions on Power Systems, Vol. 5, No. 3, August 1990.

K. Aoki, M. Kanezashi. "A Modified Newton Method for Optimal Power Flow Using Quadratic Approximated Power Flow." IEEE Transactions on Power Apparatus and Systems, Vol. PAS-104, No. 8, August 1985.

X. Bai, H. Wei, K. Fujisawa, Y. Wang. "Semidefinite Programming for Optimal Power Flow Problems." Electrical Power and Energy Systems, Vol. 30, Pp. 383-392, 2008.

R.H. Bartels. "A Stabilization of the Simplex Method." Numerische Mathematik, Vol. 16, Pp. 441-434, 1971.

R.H. Bartels and G.H. Golub. "The Simplex Method of Linear Programming Using the LU Decomposition." Communications of the ACM, Vol. 12, Pp. 266-268.

E.M.L. Beale. "An Introduction to Beale's Method of Quadratic Programming" in Nonlinear Programming, J. Abadie, Amsterdam: North-Holland, 1967.
M. F. Bedrinana, M.J. Rider, C.A. Castro. "Ill-Conditioned Optimal Power Flow Solutions and Performance of Non-Linear Programming Solvers." IEEE Bucharest Power Tech Conference, June 28th-July 2nd, 2009.
L. T. Biegler, "Nonlinear Programming: Concepts and Algorithms for Process Optimization." Chemical Engineering Department, Carnegie Mellon University
M.C. Biggs, M.A. Laughton. "Optimal Electric Power Scheduling: A Large Nonlinear Programming Test Problem Solved by Recursive Quadratic Programming." Mathematical Programming, Vol. 13, Pp. 167-182, 1977.
R.C. Burchett, H.H. Happ, K.A. Wirgau. "Large Scale Optimal Power Flow." IEEE Transactions on Power Apparatus and Systems, Vol. PAS-101, No. 10, October 1982.
R.C. Burchett, H.H. Happ, D.R. Vierath. "Quadratically Convergent Optimal Power Flow." IEEE Transaction on Power Apparatus and Systems, Vol. PAS-103, No. 11, November 1984.
R. Byrd, J.C. Gilbert, J. Nocedal. "A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming." Math Prog A, Vol. 89, Pp. 149-185, 2000.
R. H Byrd, R. B. Schnabel, and G. A. Schultz. "A Trust Region Algorithm for Nonlinearly Constrained Optimization", SIAM J. Numer. Anal., Vol. 24, Pp. 1152–1170, 1987.
M.B. Cain, R.P. O'Neill, A. Castillo. "Optimal Power Flow Papers: Paper 1. History of Optimal Power Flow and Formulations." Federal Energy Regulatory Commission (FERC) website, 2012.
J. Carpentier. "Contribution a l'Etude du Dispatching Economique." Bull. Soc. Francaise Elect., Vol. 3, Pp. 431-447, 1962.
J. Carpentier. "Optimal Power Flows." International Journal of Electrical Power & Energy Systems, Vol. 1, Iss. 1, Pp. 3-15, April 1979.
A. Castillo, R.P. O'Neill. "Optimal Power Flow Papers: Paper 5. Computational Performance of Solution Techniques Applied to the ACOPF." Federal Energy Regulatory Commission (FERC) website, 2013.
E.D. Castronuovo, J.M. Campagnolo, R. Salgado. "On the Application of High Performance Computation Techniques to Nonlinear Interior Point Methods." IEEE Transactions on Power Systems, Vol. 16, No. 3, August 2001.
M. Celis, J. E. Dennis, and R. A. Tapia. "A Trust Region Strategy for Nonlinear Equality Constrained Optimization" in Numerical Optimization (P. Boggs, R. Byrd and R. Schnabel, eds), Philadelphia: SIAM, Pp. 71–82, 1984.
R.M. Chamberlain, M.J.D. Powell, C. Lemarechal, H.C. Pedersen. "The Watchdog Technique for Forcing Convergence in Algorithms for Constrained Optimization." Math. Prog. Study 16, Pp. 1-17, 1982.

A.M. Chebbo, M.R. Irving, N.H. Dandachi. “Combined Active Reactive Dispatch.” IEEE Proc. Gener. Transm. Distrib., Vol. 142, No. 4, July 1995.
H.-D. Chiang, B. Wang, Q.-Y. Jiang. “Applications of TRUST-TECH Methodology in Optimal Power Flow of Power Systems” in Energy Systems: Optimization in the Energy Industry, Chapter 13, 2009.
C.Y. Chung, W. Yan, F. Liu. “Decomposed Predictor-Corrector Interior Point Method for Dynamic Optimal Power Flow.” IEEE Transactions on Power Systems, Vol. 26, No. 3, August 2011.
A. R. Conn, Nicholas I. M. Gould, Philippe L. Toint. "Trust-Region." MPS-SIAM Series on Optimization, 1987.
N. Demirovic, S. Tesnjak, A. Tokic. “Hot Start and Warm Start in LP Based Interior Point Method and Its Application to Multiperiod Optimal Power Flows.” Power Systems Conference and Exposition (PSCE), 2006.
R. Divi, H.K. Kesavan. “A Shifted Penalty Function Approach for Optimal Load-Flow.” IEEE Transactions on Power Apparatus and Systems, Vol. PAS-101, No. 9, September 1982.
H.W. Dommel, W.F. Tinney. ”Optimal Power Flow Solutions.” IEEE Transactions on Power Apparatus and Systems, Vol. PAS-87, No. 10, October 1968.
A.V. Fiacco, G.P. McCormick. Nonlinear Programming: Sequential Unconstrained Minimization Techniques. John Wiley & Sons, 1968.
R. Fletcher, M.J.D. Powell. “A Rapidly Convergent Descent Method for Minimization.” Computer J., Vol. 6, Pp. 163-168, 1963.
R. Fletcher, C.M. Reeves. “Function Minimization by Conjugate Gradients.” Computer J, Vol. 7, Pp 149-154, 1964.
R. Fletcher. “Fortran Subroutines for Minimization by Quasi-Newton Methods.” AERE Report, Harwell, England, 1972.
R. Fletcher. Practical Methods to Optimization, Volume 1: Unconstrained Optimization. John Wiley & Sons, 1980.
R. Fletcher. Practical Methods to Optimization, Volume 2: Constrained Optimization. John Wiley & Sons, 1981.
R. Fletcher, S. Leyffer, and P. Toint, “A Brief History of Filter Methods.” ANL/MCS-P1372-0906, September 26, 2006 revised October 9, 2006
K.R. Frisch. “The Logarithmic Potential Method for Convex Programming.” Unpublished manuscript. Institute of Economics, University of Oslo, Oslo, Norway, 1955.
P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. “Some Theoretical Properties of an Augmented Lagrangian Merit Function.” Advances in Optimization and Parallel Computing (P. M. Pardalos, ed), Pp. 101-128, North Holland, 1992.

P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. “Maintaining LU Factors of a General Sparse Matrix, Linear Algebra and its Applications.” Vol. 88/89, Pp. 239-270, 1987.
T.C. Giras and S.N. Talukdar. “Quasi-Newton Method for Optimal Power Flows.” International Journal of Electrical Power and Energy Systems, Vol. 3, No. 2, Pp. 59-64, April 1981.
J. Gondzio. “Multiple Centrality Corrections in a Primal-Dual Method for Linear Programming.” Computational Optimization and Applications, Vol. 6, Pp. 137-156, 1996.
S. Granville. “Optimal Reactive Dispatch Through Interior Point Methods.” IEEE Transactions on Power Systems, Vol. 9, No. 1, February 1994.
H.H. Happ. “Optimal Power Dispatch – A Comprehensive Survey.” IEEE Transactions on Power Apparatus and Systems, Vol. PAS-96, No. 3, May/June 1977.
M. R. Hestenes. “Multiplier and Gradient Methods.” J. Optim. Theory Appl., Vol. 4, Pp. 303–320, 1969.
N.J. Higham. “Computing a Nearest Symmetric Positive Semidefinite Matrix.” Lin. Alg. Appl., Vol. 103, Pp. 103-118, 1988.
M. Huneault, F.D. Galiana. “An Investigation of the Solution to the Optimal Power Flow Problem Incorporating Continuation Methods.” IEEE Transactions on Power Systems, Vol. 5, No. 1, February 1990.
K. Iba. “Reactive Power Optimization by Genetic Algorithm.” IEEE Transactions on Power Systems, Vol. 9, No. 2, Pp. 685-692, 1994.
R. A. Jabr, A. H. Coonick, B. J. Cory. “A Primal-Dual Interior Point Method for Optial Power Flow Dispatching.” IEEE Transactions on Power Systems, Vol. 17, No. 3, August 2002.
R.A. Jabr. “A Conic Quadratic Format for the Load Flow Equations of Meshed Networks.” IEEE Transactions on Power Systems, Vol. 22, No. 4, November 2007.
R.A. Jabr. “Optimal Power Flow Using an Extended Conic Quadratic Formulation.” IEEE Transactions on Power Systems, Vol. 23, No. 3, August 2008.
Q.Y. Jiang, H.-D. Chiang, C.X. Guo, Y.J. Cao. “Power-Current Hybrid Rectangular Formulation for Interior-Point Optimal Power Flow.” IET Gener. Transm. Distrib., Vol. 3, Iss. 8, Pp. 748-756, 2009.
Q.Y. Jiang, H.-D. Chiang. “Sequential Feasible Optimal Power Flow: Theoretical Basis and Numerical Implementation.” European Transactions on Electrical Power, Vol. 20, Pp. 695-709, 2010.
Q. Jiang, G. Geng, C. Guo, Y. Cao. “An Efficient Implementation of Automatic Differentiation in Interior Point Optimal Power Flow.” IEEE Transactions on Power Systems, Vol. 25, No. 1, February 2010.
D.S. Johnson. “A Theoreticians Guide for Experimental Analysis of Algorithms.” In Proceedings of the 5th and 6th DIMACS Implementation Challenges, 2002.

N. Karmarker. "A New Polynomial-Time Algorithm for Linear Programming." Combinatorica, Vol. 4, Pp. 373-395, 1984.
L. L. Lai, J. T. Ma. "Improved Genetic Algorithms for Optimal Power Flow under both Normal and Contingent Operation States." Electrical Power & Energy Systems, Vol. 19, No. 5, Pp. 287-292, 1997.
J. Lavaei, S. H. Low. "Zero Duality Gap in Optimal Power Flow Problem." IEEE Transactions on Power Systems, Vol. 27, Iss. 1, Pp. 92-107, 2012.
S. Leyffer, A. Mahajan. Foundations of Constrained Optimization. Argonne National Laboratory, 2010.
F. G. M. Lima, S. Soares, A. Santos Jr., K. C. Almeida, F. D. Galiana. "Numerical Experiments with an Optimal Power Flow Algorithm Based on Parametric Techniques." IEEE Transactions on Power Systems, Vol. 16, No. 3, August 2001.
C.-H. Lin, S.-Y. Lin, S.-S. Lin. "Improvements on the Daulity Based Method Used in Solving Optimal Power Flow Problems." IEEE Transactions on Power Systems, Vol. 17, No. 2, May 2002.
S.-Y. Lin, Y.-C. Ho, C.-H. Lin. "An Ordinal Optimization Theory-Based Algorithm for Solving the Optimal Power Flow Problem With Discrete Control Variables." IEEE Transactions on Power Systems, Vol. 19, No. 1, February 2004.
W.-M. Lin, C.-H. Huang, T.-S. Zhan. "Predictor-Corrector Interior Point Algorithm for Current-Based Optimal Power Flow." IEEE Power Engineering Society General Meeting, Pp. 1-6, 2007.
W.-M. Lin, C.-H. Huang, T.-S. Zhan. "A Hybrid Current-Power Optimal Power Flow Technique." IEEE Transactions on Power Systems, Vol. 23, No. 1, February 2008.
David G. Luenberger, Linear and Nonlinear Programming, Addison-Wesley, 1984.
Olvi L. Mangasarian, Nonlinear Programming, McGraw-Hill, 1969.
A.G. Maria, J.A. Findlay. "A Newton Optimal Power Flow Program for Ontario Hydro EMS." IEEE Transactions on Power Systems, Vol. 2, Iss. 3, August 1987.
S. Mehrotra. "On the Implementation of a Primal-Dual Interior Point Method." SIAM Journal on Optimization, Vol. 2, Iss. 4, Pp. 575-601, 1992.
A.P.S. Meliopoulos, F. Xia. "Optimal Power Flow Application to Composite Power System Reliability Analysis," IEEE/NTUA Athens Power Tech Conference, Athens, 1993.
W. Min, L. Shengsong. "A Trust Region Interior Point Algorithm for Optimal Power Flow Problems." Electrical Power and Energy Systems, Vol. 27, Pp. 293-300, 2005.
H. Mittelman. "AMPL-NLP Benchmark, IPOPT, KNITRO, LOQO, PENNLP, SNOPT, & CONOPT." Benchmarks for Optimization Software. Posted: http://plato.asu.edu/bench.html, Last Accessed: January, 2013.
J.A. Momoh, S.X. Guo, E.C. Ogbuobiri, R. Adapa. "The Quadratic Interior Point method Solving Power System Optimization Problems." IEEE Transactions on Power Systems, Vol. 9, No. 3, August 1994.

J.A. Momoh, J.Z. Zhu. "Improved Interior Point Method for OPF Problems." IEEE Transactions of Power Systems, Vol. 14, No. 3, August 1999.
A. Monticelli, W-H E. Liu. "Adaptive Movement Penalty Method for the Newton Optimal Power Flow." IEEE Transactions on Power Systems, Vol. 7, No. 1, February 1992.
P. K. Mukherjee, R. N. Dhar. "Optimal Load-Flow Solution by Reduced-Gradient Method." Proc. IEEE, Vol. 121, No. 6, June 1974.
B.A. Murtagh, M.A. Saunders. "A Projected Lagrangian Algorithm and its Implementation for Sparse Nonlinear Constraints." Mathematical Programming Study, Vol. 16, Pp. 84-117, 1982.
B.A. Murtagh, M.A. Sunders. "MINOS 5.51 User's Guide." Stanford University, 2003.
J. Nanda, D.P. Kothari, S.C. Srivastava. "New Optimal Power-Dispatch Algorithm Using Fletcher's Quadratic Programming Method." IEE Proceedings C: Generation, Transmission & Distribution, Vol. 136, Iss. 3, 1989.
I.M. Nejdawi, K.A. Clements, P.W. Davis. "An Efficient Interior Point Method for Sequential Quadratic Programming Based Optimal Power Flow." IEEE Transactions on Power Systems, Vol. 15, No. 4, November 2000.
J. Nocedal, S. Wright. Numerical Optimization. Springer-Verlag, 2nd Edition, 1999.
J. Nocedal. "Conjugate Gradient Methods and Nonlinear Optimization." in Linear and Nonlinear Conjugate Gradient-Related Methods (L.M. Adams, J.L. Nazareth, eds). Proceedings of the AMS-IMS-SIAM Summer Research Conference held at the University of Washington, July 1995.
A. R. L. Oliveira, S. Soares, L. Nepomuceno. "Optimal Active Power Dispatch Combining Network Flow and Interior Point Approaches." IEEE Transactions on Power Systems, Vol. 18, No. 4, November 2003.
M. Pai. Energy Function Analysis for Power System Stability. Kluwer Academic Publishers, Boston, MA, 1989.
S. Pajic, K.A. Clements. "Power System State Estimation via Globally Convergent Methods." IEEE Transactions on Power Systems, Vol. 20, No. 4, November 2005.
P. Pardalos, D.-Z. Du. Optimization Theory and Methods: Nonlinear Programming. Springer, 2006.
D.T. Phan. "Lagrangian Duality-Based Branch and Bound Algorithms for Optimal Power Flow." IBM Research Report. April 2011.
R. Polyak. "Modified Barrier Functions." Math. Prog., Vol. 54, No. 2, Pp. 177-222, 1992.
B.T. Polyak. "Newton's Method and Its Use in Optimization." European Journal of Operational Research. Vol. 181, Pp. 1086-1096, 2007.
R.A. Ponrajah, F.D. Galiana. "The Minimum Cost Optimal Power Flow Problem Solved Via the Restart Homotopy Continuation Method." IEEE Transactions on Power Systems, Vol. 4, No. 1, February 1989.

M. J. D. Powell. “A Method for Nonlinear Constraints in Minimization Problems” in Optimization (R. Fletcher), Academic Press, Pp. 283–298, London and New York , 1969.
J.K. Reid. “Fortran Subroutines for Handling Sparse Linear Programming Bases.” Report R8269. Atomic Energy Research Establishment. Harwell, England, 1976.
J.K. Reid. “A Sparsity-Exploiting Variant of the Bartels-Golub Decomposition for Linear Programming Bases.” Mathematical Programming Vol. 24, Pp 55-69, 1982.
S. M. Robinson. “A Quadratically Convergent Algorithm for General Nonlinear Programming Problems.” Mathematical Programming 3, Pp. 145-156, 1972.
J.B. Rosen. “The Gradient Projection Method for Nonlinear Programming. Part II. Nonlinear Constraints.” Journal of the Society for Industrial and Applied Mathematics, Vol. 9, No. 4, Pp. 514-532, Dec. 1961.
R. Salgado, A. Brameller, P. Aitchison. “Optimal Power Flow Solutions Using the Gradient Projection Method. Part 2: Modelling of the Power System Equations.” IEE Proceedings, Vol. 137, Pt. C, No. 6, November 1990.
A. Santos, Jr., S. Deckmann, S. Soares. “A Dual Augmented Lagrangian Approach for Optimal Power Flow.” IEEE Transactions on Power Systems, Vol. 3, No. 3, August 1988.
A.M. Sasson. “Combined Use of the Powell and Fletcher-Powell Nonlinear Programming Methods for Optimal Load Flows.” IEEE Transactions on Power Apparatus and Systems, Vol. PAS-88, No. 10, October 1969.
A.M. Sasson, F. Viloria, F. Aboytes. “Optimal Load Flow Solution Using the Hessian Matrix.” IEEE Transactions on Power Apparatus and Systems. Vol. PAS-92, Iss. 1, 1973.
R. R. Shoults, D. T. Sun. “Optimal Power Flow Based Upon P-Q Decomposition.” IEEE Transactions on Power Apparatus and Systems, Vol. PAS-101, No. 2, February 1982.
A. A. Sousa, G. L. Torres. “Globally Convergent Optimal Power Flow by Trust-Region Interior-Point Methods.” IEEE Lausanne Power Tech, 2007.
V.A. de Sousa, E.C. Baptista, G.R.M. da Costa. “Loss Minimization by the Predictor-Corrector Modified Barrier Approach.” Electric Power Systems Research, Vol. 79, Pages 803-808, 2009.
A.A. Sousa, G.L. Torres, C.A. Cañizares. “Robust Optimal Power Flow Solution Using Trust Region and Interior-Point Methods.” IEEE Transactions on Power Systems, Vol. 26, No. 2, May 2011.
B. Stott, O. Alsac. “Optimal Power Flow—Basic Requirements for Real-Life Problems and their Solutions.” Posted: http://www.ieee.hr/_download/repository/Stott-Alsac-OPF-White-Paper.pdf, July 1, 2012. Last Accessed: January, 2013.
D.I. Sun, B. Ashley, B. Brewer, A. Hughes, W.F. Tinney. “Optimal Power Flow by Newton Approach.” IEEE Transaction on Power Apparatus and Systems, Vol. PAS-103, No. 10, October 1984.

J. E. Tate, T. J. Overbye. "A Comparison of the Optimal Multiplier in Polar and Rectangular Coordinates." IEEE Transactions on Power Systems, Vol. 20, No. 4, November 2005.
G.L.Torres, V.H. Quintana. "Nonlinear Optimal Power Flow by a Noninterior-Point Method Based on Chen-Harker-Kanzow NCP-functions." In Proceeding of Electrical and Computer Engineering, Vol. 2, 1998.
G. L. Torres, V. H. Quintana. "Optimal Power Flow by a Nonlinear Complementarity Method." IEEE Transactions on Power Systems, Vol. 15, No. 3, August 2000.
G.L. Torres, V.H. Quintana. "On a Nonlinear Multiple-Centrality-Correction Interior-Point Method for Optimal Power Flow." IEEE Transactions on Power Systems, Vol. 16, No. 2, May 2001.
G. L. Torres, V. H. Quintana. "A Jacobian Smoothing Nonlinear Complementarity Method for Solving Nonlinear Optimal Power Flows." 14th PSCC, Sevilla, 24-28 June 2002.
G.L. Torres, M.A. De Carvalho. "On Efficient Implementation of Interior-Point Based Optimal Power Flows in Rectangular Coordinates." IEEE Power Systems Conference and Exposition, Pp. 1747-1752, 2006.
University of Washington (UW) Electrical Engineering. Power Systems Test Case Archive. Posted: http://www.ee.washington.edu/research/pstca/. Last accessed: January, 2013.
L.S. Vargas, V.H. Quintana, A. Vannelli. "A Tutorial Description of an Interior Point Method and Its Application to Security-Constrained Economic Dispatch." IEEE Summer Meeting, Paper SM 92 416-8 PWRS, 1992.
H. Wang, C.E. Murillo-Sánchez, R.D. Zimmerman, R.J. Thomas. "On Computational Issues of Market-Based Optimal Power Flow." IEEE Transactions on Power Systems, Vol. 22, No. 3, August 2007.
A. Wächter and L. Biegler, "Line Search Filter Methods for Nonlinear Programming: Motivation and Global Convergence." SIAM Journal on Optimization, Vol. 16, No. 1, Pp. 1-31, 2005.
H. Wei, H. Sasaki, J. Kubokawa, R. Yokoyama. "An Interior Point Nonlinear Programming for Optimal Power Flow Problems with A Novel Data Structure." IEEE Transactions on Power Systems, Vol. 13, No. 3, August 1998.
P. Wolfe. "The Simplex Method for Quadratic Programming," Econometrica, Vol. 27, Pp. 382-398, 1959.
P. Wolfe. "Methods of nonlinear programming." Nonlinear Programming, J. Abadie, Ed., North Holland, Amsterdam, Pp. 97-131, 1967.
Y.-C. Wu, A.S. Debs, R.E. Marsten. "A Direct Nonlinear Predictor-Corrector Primal-Dual Interior Point Algorithm for Optimal Power Flows." IEEE Transactions on Power Systems, Vol. 9, No. 2, May 1994.

F.F. Wu, G. Gross, J.F. Luini, P.M. Look. "A Two-Stage Approach to Solving Large-Scale Optimal Power Flows." IEEE Conference Proceedings , Power Industry Computer Applications Conference, PICA-79, Pp. 126-136, 1979.
L. Xie, H.-D. Chiang. "A Enhanced Multiple Predictor-Corrector Interior Point Method for Optimal Power Flow." IEEE Power and Energy Society General Meeting, 2010.
L. Xie, H.-D. Chiang. "Weighted Multiple Predictor-Corrector Interior Point Method for Optimal Power Flow." Electric Power Components and Systems, Vol. 39, Pp. 99-112, 2011.
K. Xie, Y.H. Song. "Dynamic Optimal Power Flow by Interior Point Methods." IEE Proceedings: Generation, Transmission, & Distribution, Vol. 148, Iss. 1, Pp. 76-84, 2001.
X. Yan, V.H. Quintana. "Improving An Interior-Point-Based OPF by Dynamic Adjustments of Step Sizes and Tolerances." IEEE Transactions on Power Systems, Vol. 14, No. 2, May 1999.
B.-Y. Yang, J.-C. Peng. "A Novel Optimal Power Flow Model and Algorithm Based on the Imputation of an Impedance-Branch Dissipation Power." Asian-Pacific Power and Energy Engineering Conference, 2009.
R. Zhou, Y. Zhang, H. Yang. "A Trust-Region Algorithm Based on Global SQP for Reactive Power Optimization." 2nd International Conference on Electrical and Electronics Engineering (ICEEE) and XI Conference on Electrical Engineering (CIE 2005). Mexico City, Mexico. September 2005.
R.D. Zimmerman, C. E. Murillo-Sanchez, R.J. Thomas. "MATPOWER: Stead-State Operations, Planning, and Analysis Tools for Power Systems Research and Education." IEEE Transactions on Power Systems, Vol. 26, No. 1, February 2011.