Inner point methods: On necessary optimality conditions of various reformulations of a constrained optimization problem

Marco Rozgiè, Manuel Jaraczewski and Marcus Stiemer

Helmut Schmidt University - University of the Federal Armed Forces Hamburg, Holstenhofweg 85, 22034 Hamburg, Germany,
marco.rozgic@hsu-hh.de

Abstract

Primal-dual inner point algorithms are known to be efficient in solving non-linear constrained optimization problems. Modern implementations are capable of solving optimization problems with a huge number of non-linear constraints. To do this efficiently it is crucial, that necessary optimality conditions are formulated such that they can be easily implemented into a computer program. Favourable is a formulation as a system of equations that can be linearized. The Karush-Kuhn-Tucker conditions represent such a set. This work gives a rigours proof for the equivalence of the necessary conditions of the reformulations of a non-linear constrained optimization problem as they are used in inner point methods.

1. Motivation and Introduction

Optimization plays a prominent role in a vast number of applications. In many applications from physics some energy term has to be minimized (see, e.g., Sheppard et al., 2008), in certain engineering problems optimal shapes are desired (see, e.g., Schwarz et al., 2001), in finance certain costs need to be minimized, where on the other hand profit needs to be maximized (see, e.g., Schaible, 1981), to just mention a few. Mostly the figure of merit that has to be optimized is subject to constraints that are expressed as equalities or inequalities. Further the participating functions are often non-linear, which adds more difficulty to finding a solution of the resulting non-linear constrained optimization problems. Since these problems appear in many different disciplines, an accessible and comprehensive theory as well as easy applicable algorithms were developed in the last decades, like ALGENCAN (see Andreani et al., 2008, 2007), KNITRO (see Byrd et al., 2000, 1999, 2006), LANCELOT (see Conn et al., 2010), filterSQP (see Fletcher and Leyffer, 1998), NPSOL (see Gill et al., 1984), SNOPT (see Gill et al., 2005), PENNON (see Kočvara and Stingl, 2003), MINOS (see Murtagh and Saunders, 1983), LOQO (see Vanderbei and Shanno, 1999) or IPOPT (see Wächter and Biegler, 2006). Generally it can be distinguished between two branches of optimization methods, derivative free methods and gradient based methods. Whereas derivative free methods based on surrogates for the objective function or heuristics for the descent direction, like Kriging (see, e.g., Jones et al., 1998), simulated annealing (Van Laarhoven and Aarts, 1987) or the Nelder-Mead method proposed by Nelder and Mead (1965), require no further information about the participating functions, the number of - under certain circumstances expensive - objective calculations is higher than when applying gradient based methods. Inner point (IP) methods as can be found in Nocedal and Wright (2006) represent a very efficient class of gradient based methods. They are well understood and studied also efficient implementations are available (see, e.g., Wächter and Biegler, 2006; Byrd et al., 2000). However, the efficiency of these implementations depends on the availability of computable linerizations as well as on suitable formulations of necessary optimality conditions, i.e., the optimality conditions are expressed as solvable equation systems. Their theoretic foundation depends on the equivalence of the first order necessary conditions for constrained optimization, the so called Karush-Kuhn-Tucker (KKT) conditions introduced by Karush (1939) and Kuhn and Tucker (1951), for several problem (re-) formulations. Although the presented result seems to be well known folklore and rather basic (see, e.g., Gould et al., 2001), to the best of the authors knowledge no formal proof of the presented theorem has been published. With this work the authors would like to fill the gap for researchers who are not so familiar with the calculus of optimization problems. The authors also would like to refer to the many outstanding text books that are available about optimization theory (see, e.g., Nocedal and Wright, 2006; Bazaraa and Shetty, 1979; Fletcher, 2000), which may provide a guideline to find an efficient and reliable method tailored to the particular optimization problem.

2. Notations and Definitions

Before the main theorem is formulated, some functions and notations have to be introduced. For this purpose let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable real valued function. Further let the functions $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^k$ also be continuously differentiable (see, e.g., Rudin, 1987). With these functions the following (non-linear) optimization problem for n decision variables constrained by m inequality and k equality constraints is formulated:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x), \\ \text{subject to} \quad & g(x) \leq 0, \\ & h(x) = 0, \end{aligned} \tag{1}$$

where all inequalities have to be understood component wise. In the context of IP methods problem (1) is reformulated into a problem with only (non-linear) equalities and box constraints on the variables. In order to do so a component wise positive slack variable $s \in \mathbb{R}^m$ is introduced and added to the inequalities. Further $x \in \mathbb{R}^n$ is decomposed as $x = x^+ - x^-$, with component wise positive vectors $x^+, x^- \in \mathbb{R}^n$. This yields the following optimization problem:

$$\begin{aligned} \min_{x^+, x^- \in \mathbb{R}^n, s \in \mathbb{R}^m} \quad & \tilde{f}(x^+, x^-, s), \\ \text{subject to} \quad & \tilde{c}(x^+, x^-, s) = 0, \\ & x^+, x^-, s \geq 0. \end{aligned} \tag{2}$$

Where the participating functions $\tilde{f} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, and $\tilde{c} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{m+k}$ are defined by

$$\begin{aligned} \tilde{f}(x^+, x^-, s) &= f(x^+ - x^-), \\ \tilde{c}(x^+, x^-, s) &= \begin{pmatrix} g(x^+ - x^-) + s \\ h(x^+ - x^-) \end{pmatrix}. \end{aligned}$$

The Lagrangian of problem (2) is given by

$$\mathcal{L}(x^+, x^-, s, \lambda, \vartheta) = \tilde{f}(x^+, x^-, s) + \lambda^{\mathrm{T}} \tilde{c}_s(x^+, x^-, s) - \vartheta^{\mathrm{T}} \begin{pmatrix} x^+ \\ x^- \\ s \end{pmatrix}, \tag{3}$$

with Lagrangian multipliers $\lambda \in \mathbb{R}^{m+k}$ and $\vartheta \in \mathbb{R}^{\widetilde{m}}$, where $\widetilde{m} = n + n + m$. Further the projections of the multipliers to the respective contributions are denoted by $\lambda_g, \vartheta_s \in \mathbb{R}^m$, $\lambda_h \in \mathbb{R}^k$, $\vartheta_{x^+}, \vartheta_{x^-} \in \mathbb{R}^n$ yielding $\lambda = (\lambda_g, \lambda_h)^{\mathrm{T}}$ and $\vartheta = (\vartheta_{x^+}, \vartheta_{x^-}, \vartheta_s)^{\mathrm{T}}$. The gradient of the Lagrangian with respect to the primal variables x^+, x^-, s, denoted by $\nabla \mathcal{L}(x^+, x^-, s, \lambda, \vartheta)$, is used to define the parametric function $\Psi_\mu : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{2\widetilde{m}+m+k}$ by

$$\Psi_\mu(x^+, x^-, s) = \begin{pmatrix} \nabla \mathcal{L}(x^+, x^-, s, \lambda, \vartheta) \\ \tilde{c}(x^+, x^-, s) \\ \boldsymbol{X}\vartheta - \mu \mathbb{1} \end{pmatrix}, \tag{4}$$

where $\boldsymbol{X} \in \mathbb{R}^{\widetilde{m} \times \widetilde{m}}$ is a diagonal matrix with the target variables of problem (2) on the diagonal, $\boldsymbol{X} = \operatorname{diag}(x^+, x^-, s)$ and $\mathbb{1} \in \mathbb{R}^{\widetilde{m}}$ is a vector of ones of appropriate size, $\mathbb{1} = (1, \ldots, 1)^{\mathrm{T}}$. With these declarations the first order optimality theorem can be formulated.

3. First Order Necessary Conditions

Theorem 1 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^k$ be continuously differentiable functions. Let further $\Psi_\mu : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{2\tilde{m}+m+k}$ be defined as in (4). Then the following holds:*

a) Let $\Psi_{\mu=0}\left(x^+, x^-, s\right) = 0$ and $\vartheta \geq 0$ as well as $\left(x^+, x^-, s\right)^{\mathrm{T}} \geq 0$ hold, then $\left(x^+, x^-, s\right)^{\mathrm{T}}$ is a KKT point of problem (2) with multipliers (λ, ϑ).

b) Let $\left(x^+, x^-, s\right)^{\mathrm{T}}$ be a KKT point of problem (2) with multipliers $(\lambda, \vartheta)^{\mathrm{T}}$, then $x = x^+ - x^-$ is a KKT point of problem (1) with multiplier λ. □

Proof a) By definition, (x^+, x^-, s) is a KKT point of (2), if $(x^+, x^-, s, \lambda, \vartheta)$ fulfils

$$\nabla \tilde{f}\left(x^+, x^-, s\right) + \boldsymbol{J}_{\tilde{c}}^{\mathrm{T}} \lambda - \vartheta = 0, \tag{5a}$$

$$\tilde{c}\left(x^+, x^-, s\right) = 0, \tag{5b}$$

$$x^+, x^-, s \geq 0 \tag{5c}$$

$$\boldsymbol{X}\vartheta = 0, \tag{5d}$$

$$\vartheta \geq 0. \tag{5e}$$

The assumptions

$$\Psi_{\mu=0}\left(x^+, x^-, s\right) = \begin{pmatrix} \nabla \mathcal{L}\left(x^+, x^-, s, \lambda, \vartheta\right) \\ \tilde{c}\left(x^+, x^-, s\right) \\ \boldsymbol{X}\vartheta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{6}$$

yield the stationarity (5a), primal feasibility (5b) and complementary slackness conditions (5d) of problem (2). Dual feasibility (5e) holds, since $\vartheta \geq 0$ is explicitly demanded. Further (5c) follows since $x^+, x^-, s \geq 0$ is assumed.

b) The KKT conditions at a point $x^* \in \mathbb{R}^n$ for problem (1) are given by

$$\nabla f\left(x^*\right) + \nabla g\left(x^*\right)^{\mathrm{T}} \lambda_g + \nabla h\left(x^*\right)^{\mathrm{T}} \lambda_h = 0, \tag{7a}$$

$$g\left(x^*\right) \leq 0, \tag{7b}$$

$$h\left(x^*\right) = 0, \tag{7c}$$

$$\boldsymbol{G}\left(x^*\right) \lambda_g = 0, \tag{7d}$$

$$\lambda_g \geq 0, \tag{7e}$$

where $\boldsymbol{G}\left(x^*\right) = \operatorname{diag}\left(g\left(x^*\right)\right) \in \mathbb{R}^{m \times m}$ is the diagonal matrix with the components of the inequality constraint at x^* on its diagonal. To prove that indeed the KKT conditions (7) of problem (1) hold, the components of the equations given in (5) have to be investigated closer. The gradient of the objective of problem (2) in Eq. (5a) is given by

$$\nabla \tilde{f}\left(x^+, x^-, s\right) = \begin{pmatrix} \nabla f(x) \\ -\nabla f(x) \\ \boldsymbol{0}_{\mathbb{R}^m} \end{pmatrix}, \tag{8}$$

where $\nabla f(x)$ is the gradient of the objective in problem (1) and $\boldsymbol{0}_{\mathbb{R}^m}$ is the vector of zeros in $\mathbb{R}^m$. The second contribution in (5a) is given by

$$\boldsymbol{J}_{\tilde{c}}^{\mathrm{T}} \lambda = \begin{pmatrix} \nabla g_1(x) & \dots & \nabla g_m(x) & \nabla h_1(x) & \dots & \nabla h_k(x) \\ -\nabla g_1(x) & \dots & -\nabla g_m(x) & -\nabla h_1(x) & \dots & -\nabla h_k(x) \\ \boldsymbol{e}_1 & \dots & \boldsymbol{e}_m & \boldsymbol{0}_{\mathbb{R}^m} & \dots & \boldsymbol{0}_{\mathbb{R}^m} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \\ \lambda_{m+1} \\ \vdots \\ \lambda_{m+k} \end{pmatrix},$$

where $\boldsymbol{e}_j$ corresponds to the $j-$th unit vector in $\mathbb{R}^m$, and $\nabla g_j(x)$ denotes the gradient of the $j-$th inequality constraint function in problem (1), and $\nabla h_i(x)$ denotes the gradient of the $i-$th equality constraint function respectively. The first line of the above equation can be written in more detail as

$$\begin{aligned}
&\begin{pmatrix} \nabla g_1(x) & \dots & \nabla g_m(x) & \nabla h_1(x) & \dots & \nabla h_k(x) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \\ \lambda_{m+1} \\ \vdots \\ \lambda_{m+k} \end{pmatrix} \\
&= \begin{pmatrix} (\nabla g_1(x))_1 & \dots & (\nabla g_m(x))_1 & (\nabla h_1(x))_1 & \dots & (\nabla h_k(x))_1 \\ (\nabla g_1(x))_2 & \dots & (\nabla g_m(x))_2 & (\nabla h_1(x))_2 & \dots & (\nabla h_k(x))_2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ (\nabla g_1(x))_n & \dots & (\nabla g_m(x))_n & (\nabla h_1(x))_n & \dots & (\nabla h_k(x))_n \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \\ \lambda_{m+1} \\ \vdots \\ \lambda_{m+k} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^m (\nabla g_i(x))_1 \lambda_i + \sum_{j=1}^k \left(\nabla h_j(x)\right)_1 \lambda_{m+j} \\ \vdots \\ \sum_{i=1}^m (\nabla g_i(x))_n \lambda_i + \sum_{j=1}^k \left(\nabla h_j(x)\right)_n \lambda_{m+j} \end{pmatrix} = \boldsymbol{J}_c^{\mathrm{T}} \lambda,
\end{aligned}$$

where $\boldsymbol{J}_c$ denotes the Jacobian of the function $c : \mathbb{R}^n \to \mathbb{R}^{m+k}$, given by

$$c(x) = \begin{pmatrix} g(x) \\ h(x) \end{pmatrix}.$$

Note that $c : \mathbb{R}^n \to \mathbb{R}^{m+k}$ is representing constraints of problem (1). The product $\boldsymbol{J}_{\tilde{c}}\lambda$ in (5a) can thus be written as

$$\boldsymbol{J}_{\tilde{c}}^{\mathrm{T}} \lambda = \begin{pmatrix} \boldsymbol{J}_c^{\mathrm{T}} \lambda \\ -\boldsymbol{J}_c^{\mathrm{T}} \lambda \\ \boldsymbol{M} \lambda \end{pmatrix}, \tag{9}$$

where the matrix $\boldsymbol{M} \in \mathbb{R}^{m \times m+k}$ is given by

$$M = \begin{pmatrix} \boldsymbol{e}_1 & \dots & \boldsymbol{e}_m & \boldsymbol{0}_{\mathbb{R}^m} & \dots & \boldsymbol{0}_{\mathbb{R}^m} \end{pmatrix}.$$

Finally, Eq. (5a) can be written as

$$\nabla \tilde{f}(x^+, x^-, s) + \boldsymbol{J}_{\tilde{c}}^{\mathrm{T}} \lambda - \vartheta = \begin{pmatrix} \nabla f(x) \\ -\nabla f(x) \\ \boldsymbol{0}_{\mathbb{R}^m} \end{pmatrix} + \begin{pmatrix} \boldsymbol{J}_c^{\mathrm{T}} \lambda \\ -\boldsymbol{J}_c^{\mathrm{T}} \lambda \\ \boldsymbol{M} \lambda \end{pmatrix} - \begin{pmatrix} \vartheta_{x^+} \\ \vartheta_{x^-} \\ \vartheta_s \end{pmatrix} = \begin{pmatrix} \boldsymbol{0}_{\mathbb{R}^n} \\ \boldsymbol{0}_{\mathbb{R}^n} \\ \boldsymbol{0}_{\mathbb{R}^m} \end{pmatrix}.$$

Next, the stationarity condition (7a) for problem (1) can be concluded, since from

$$\nabla f(x) + \boldsymbol{J}_c^{\mathrm{T}} \lambda = \vartheta_{x^+} \geq 0$$

and

$$\nabla f(x) + \boldsymbol{J}_c^{\mathrm{T}} \lambda = -\vartheta_{x^-} \leq 0$$

the equation

$$\nabla f(x) + J_c^{\mathrm{T}} \lambda = 0 \tag{10}$$

follows. Further from $M\lambda = \lambda_g$ the identity $\lambda_g = \vartheta_s$ can be deduced and, hence, $\lambda_g \geq 0$ holds, which is the dual feasibility condition (7e) of problem (1). To obtain the complementary slackness condition (7d) Eq. (5d) is exploited:

$$\begin{aligned} X\vartheta = 0 \Rightarrow s_i \vartheta_{i+2n} = 0 \underset{g(x)+s=0}{\Longleftrightarrow} (-g(x))_i \vartheta_{i+2n} = 0 \\ \underset{\lambda_g = \vartheta_s}{\Longleftrightarrow} (g(x))_i \lambda_i = 0, \end{aligned}$$

for all $i = 1, \ldots, m$. Th last equation can be summed up in matrix notation by

$$G(x) \lambda_g = 0.$$

The primal feasibility conditions (7b) and (7c) for problem (1) can be deduced by exploiting (5b) and (5c). This concludes the proof. ∎

Remark 1 Primal-dual inner point methods as are implemented in Wächter and Biegler (2006) are known to have super linear convergence, as was already shown by Gould et al. (2001). The robust behaviour and convergence rate of this class of methods base on the fact, that in every iteration step a logarithmic barrier problem is solved to a certain extend. The logarithmic barrier problem to problem (2) is given by the optimization program

$$\begin{aligned} \min_{x^+, x^- \in \mathbb{R}^n, s \in \mathbb{R}^m} \quad & \tilde{f}(x^+, x^-, s) - \mu \left(\sum_{i=1}^{n} \ln(x_i^+) + \sum_{i=1}^{n} \ln(x_i^-) + \sum_{i=1}^{m} \ln(s_i) \right), \\ \text{subject to} \quad & \tilde{c}(x^+, x^-, s) = 0. \end{aligned} \tag{11}$$

Here $\mu \in \mathbb{R}$, $\mu \geq 0$ is the barrier parameter. To see how a solution of the barrier problem (11) fits into the above discussed scheme the KKT conditions of (11) have to be analyzed. □

Corollary 1 *Let all prequisits of theorem 1 hold. If* $\Psi_\mu(x^+, x^-, s) = 0$ *and* $(x^+, x^-, s)^{\mathrm{T}} > 0$ *hold, then* $(x^+, x^-, s)^{\mathrm{T}}$ *is a KKT point of the barrier problem (11).* □

Proof The KKT conditions of the barrier problem (11), as can be found in, e.g., Nocedal and Wright (2006), are given by the firstly stationarity condition

$$\nabla \tilde{f}(x^+, x^-, s) - \mu X^{-1} \mathbb{1} + J_{\tilde{c}}^{\mathrm{T}} \lambda = 0, \tag{12}$$

where the Jacobian of $\tilde{c}(x^+, x^-, s)$ denoted by $J_{\tilde{c}}$ and $\lambda \in \mathbb{R}^{m+k}$ is the corresponding Lagrangian multiplier, and secondly the primal feasibility condition

$$\tilde{c}(x^+, x^-, s) = 0. \tag{13}$$

The assumption $\Psi_\mu(x^+, x^-, s) = 0$ is equivalent to

$$\begin{pmatrix} \nabla \mathcal{L}(x^+, x^-, s, \lambda, \vartheta) \\ \tilde{c}(x^+, x^-, s) \\ X\vartheta - \mu \mathbb{1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{14}$$

Particularly, primal feasibility is fulfilled. The matrix X is regular since (x^+, x^-, s) are strictly away from zero. Hence, ϑ can be written as

$$\vartheta = \mu X^{-1} \mathbb{1}. \tag{15}$$

This identity is plugged into the first line of (14) yielding

$$\nabla \tilde{f}\left(x^{+}, x^{-}, s\right) - \mu \boldsymbol{X}^{-1}\mathbb{1} + \boldsymbol{J}_{\tilde{c}}^{\mathrm{T}}\lambda = 0, \tag{16}$$

which is the stationarity condition to be proven. ■

Remark 2 As pointed out in Nocedal and Wright (2006) the inner point method and finding roots of the parametric function Ψ_μ combined with reducing μ to 0 are equivalent. With this in mind it can be concluded that solutions found by the inner point method are at least KKT-points of the underlying constrained minimization problem and therefore fulfil necessary optimality conditions for non-linear constrained optimization problems. □

Acknowledgement

This work has been funded by the German Research Foundation (DFG) under the Contract PAK 343/2. The authors wish to express their gratitude for the support.

Bibliography

Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L., 2007. On augmented lagrangian methods with general lower-level constraints. SIAM Journal on Optimization 18, 1286–1309.

Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L., 2008. Augmented lagrangian methods under the constant positive linear dependence constraint qualification. Mathematical Programming 111, 5–32.

Bazaraa, M.S., Shetty, C., 1979. Nonlinear Programming. John Wiley & Sons.

Byrd, R.H., Gilbert, J.C., Nocedal, J., 2000. A trust region method based on interior point techniques for nonlinear programming. Mathematical Programming 89, 149–185.

Byrd, R.H., Hribar, M.E., Nocedal, J., 1999. An interior point algorithm for large-scale nonlinear programming. SIAM Journal on Optimization 9, 877–900.

Byrd, R.H., Nocedal, J., Waltz, R.A., 2006. Knitro: An integrated package for nonlinear optimization, in: Large-scale nonlinear optimization. Springer, pp. 35–59.

Conn, A.R., Gould, G., Toint, P.L., 2010. LANCELOT: a Fortran package for large-scale nonlinear optimization (Release A). Springer Publishing Company, Incorporated.

Fletcher, R., 2000. Practical Methods of Optimization. Wiley.

Fletcher, R., Leyffer, S., 1998. User manual for filtersqp. University of Dundee Numerical Analysis Report NA-181 .

Gill, P.E., Murray, W., Saunders, M.A., 2005. Snopt: An sqp algorithm for large-scale constrained optimization. SIAM review 47, 99–131.

Gill, P.E., Murray, W., Saunders, M.A., Wright, M.H., 1984. User's guide for NPSOL (Version 2. 1): a Fortran package for nonlinear programming. Technical Report. Stanford Univ., CA (USA). Systems Optimization Lab.

Gould, N.I., Orban, D., Sartenaer, A., Toint, P.L., 2001. Superlinear convergence of primal-dual interior point algorithms for nonlinear programming. SIAM Journal on Optimization 11, 974–1002.

Jones, D.R., Schonlau, M., Welch, W.J., 1998. Efficient global optimization of expensive black-box functions. Journal of Global optimization 13, 455–492.

Karush, W., 1939. Minima of functions of several variables with inequalities as side constraints. Ph.D. thesis. Masters thesis, Dept. of Mathematics, Univ. of Chicago.

Kočvara, M., Stingl, M., 2003. Pennon: A code for convex nonlinear and semidefinite programming. Optimization methods and software 18, 317–333.

Kuhn, H.W., Tucker, A.W., 1951. Nonlinear programming, in: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, University of California Press, Berkeley and Los Angeles. pp. 481–492.

Murtagh, B.A., Saunders, M.A., 1983. MINOS 5. 0 user's guide. Technical Report. Stanford Univ., CA (USA). Systems Optimization Lab.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. The Computer Journal 7, 308–313.

Nocedal, J., Wright, S., 2006. Numerical optimization, series in operations research and financial engineering. Springer, New York.

Rudin, W., 1987. Real and complex analysis. McGraw-Hill.

Schaible, S., 1981. Generalized concavity in optimization and economics. Academic Press, New York.

Schwarz, S., Maute, K., Ramm, E., 2001. Topology and shape optimization for elastoplastic structural response. Computer Methods in Applied Mechanics and Engineering 190, 2135–2155.

Sheppard, D., Terrell, R., Henkelman, G., 2008. Optimization methods for finding minimum energy paths. The Journal of Chemical Physics 128, –.

Van Laarhoven, P.J., Aarts, E.H., 1987. Simulated annealing. Springer.

Vanderbei, R.J., Shanno, D.F., 1999. An interior-point algorithm for nonconvex nonlinear programming. Computational Optimization and Applications 13, 231–252.

Wächter, A., Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming 106, 25–57.