Local Saddle Point for the Hyperbolic Augmented Lagrangian

Lennin Mallma Ramirez, Nelson Maculan, Adilson Elias Xavier and Vinicius Layter Xavier

Systems Engineering and Computer Science Program (PESC), Graduate School of Enginnering (COPPE), Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil. Rio de Janeiro State University, Institute of Mathematics and Statistics, Graduate Program in Computational Science. Rio de Janeiro, Brazil

ARTICLE HISTORY
Compiled March 24, 2022

ABSTRACT
In this work, we present an approach to guarantee that the hyperbolic augmented Lagrangian function (HALF) has local saddle points. This result is obtained under the second-order sufficient condition.

KEYWORDS
Nonconvex problem; constrained optimization; hyperbolic augmented Lagrangian; local saddle point.

1. Basic Results

Throughout this paper, we are interested in studying the following inequality constrained nonconvex optimization problem

$$(P) \qquad \min\{f(x) \mid x \in S\},$$

where

$$S = \{x \in X \subseteq I\!R^n \mid g_i(x) \geq 0, \;\; i = 1, ..., m\},$$

is the feasible set of the problem (P), where $f : I\!R^n \to I\!R$ and $g_i : I\!R^n \to I\!R, \; i = 1, ..., m$ are continuously differentiable functions and where X is a nonempty closed set in $I\!R^n$.

The Lagrangian function of the problem (P) is $L : I\!R^n \times I\!R^m_+ \to I\!R$, defined as

$$L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x), \tag{1.1}$$

where $\lambda_i \geq 0, \; i = 1, ..., m$.

CONTACT L. Mallma Ramirez. Email: lenninmr@gmail.com

The dual function of problem (P) is $\Phi : \mathbb{R}_+^m \rightarrow \mathbb{R}$, is defined as follows

$$\Phi(\lambda) = \min_{x \in X} L(x, \lambda), \tag{1.2}$$

and the corresponding Lagrangian dual problem of (P) is

$$(D) \qquad \max\{\Phi(\lambda) \mid \lambda \in \mathbb{R}_+^m\}.$$

A pair (x^*, λ^*) is said to be a local saddle point of $L(x, \lambda)$ if there exists a $\delta > 0$ such that

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*), \tag{1.3}$$

holds for all $\lambda \geq 0$ and $x \in X \cap N(x^*, \delta)$, where $N(x^*, \delta) = \{x \in \mathbb{R}^n \mid \|x - x^*\| \leq \delta\}$.

Let $x^* \in X$ be a local solution to problem (P). Let's consider the following classic assumption:

A. (Second-order sufficiency condition) Let $x^* \in X$ be a local solution to (P). Assume that x^* is a regular point of problem (P) and satisfies the second-order sufficient conditions; that is, $\nabla g_i(x^*)$, $i \in I(x^*)$, are linearly independent and there exists $\lambda^* \geq 0$ such that

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0, \tag{1.4}$$

$$\lambda_i^* g_i(x^*) = 0, \; i = 1, ..., m, \tag{1.5}$$

and the Hessian matrix

$$\nabla_{xx}^2 L(x^*, \lambda^*) = \nabla^2 f(x^*) - \sum_{i \in J(x^*)} \lambda_i^* \nabla^2 g_i(x^*), \tag{1.6}$$

is positive definite on the cone

$$M(x^*) = \left\{d \in \mathbb{R}^n, d \neq 0 \mid d^T \nabla g_i(x^*) = 0, \; for\ any \;\; i \in J(x^*)\right\}, \tag{1.7}$$

where $J(x^*) = \{i \mid \lambda_i^* > 0, i = 1, ..., m\}$.

Assumption **A** is widely used in different types of augmented Lagrangian functions, see, Arrow et al. (1973), Sun et al. (2005) and Echebest et al. (2016). The following result will be important, to guarantee the existence of local saddle points.

Corollary 1.1. *(See, Corollary 12.9 of Avriel (1976)) Let A be an $n \times n$ matrix and let B be an $m \times n$ matrix. Then $z^T A z > 0$ for every $z \neq 0$ satisfying $Bz = 0$ if and only if there exists a number $c^* > 0$ such that, for all $c > c^*$, it follows that,*

$$z^T \left(A + cB^T B\right) z > 0,$$

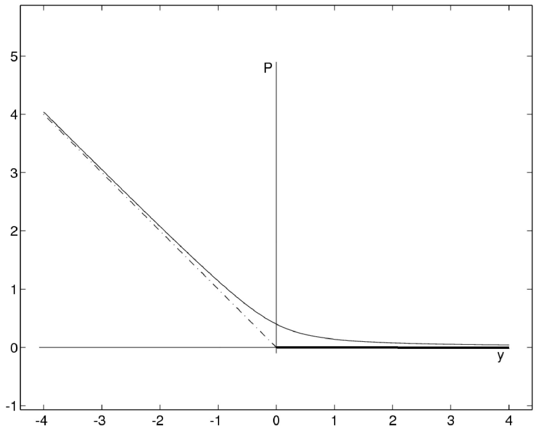

Figure 1. Hyperbolic Penalty Function

for all $z \neq 0$.

1.1. Hyperbolic Penalty

The hyperbolic penalty was introduced in Xavier (1982) and is meant to solve the problem (P). The penalty method adopts the hyperbolic penalty function (HPF)

$$P(y, \lambda, \tau) = -\lambda y + \sqrt{(\lambda y)^2 + \tau^2}, \tag{1.8}$$

where $P : (-\infty, +\infty) \times I\!R_+ \times I\!R_{++} \to I\!R$. The graphic representation of $P(y, \lambda, \tau)$, is as shown in Figure 1.

Remark 1.1. *The HPF is originally proposed in Xavier (1982) and studied in Xavier (2001). In these works, the following properties are important for HPF:*

(a) $P(y, \lambda, \tau)$ *is asymptotically tangent to the straight lines* $r_1(y) = -2\lambda y$ *and* $r_2(y) = 0$ *for* $\tau > 0$.

(b)
- $P(y, \lambda, 0) = 0, \quad$ *for* $\quad y \geq 0$.
- $P(y, \lambda, 0) = -2\lambda y, \quad$ *for* $\quad y < 0$.

Due to the properties (a) *and* (b) *the HPF is equivalent to a smoothing of the penalty studied in Zangwill (1967).*

In particular we use the following properties:

P0) $P(y, \lambda, \tau)$ is $k-$times continuously differentiable for any positive integer k for $\tau > 0$.

P1) $P(0, \lambda, \tau) = \tau$, for $\tau > 0$ and $\lambda \geq 0$.

P2) $P(y, \lambda, \tau)$ is strictly decreasing function of y, i.e.,

$$\nabla_y P(y, \lambda, \tau) = -\lambda \left(1 - \frac{\lambda y}{\sqrt{(\lambda y)^2 + \tau^2}}\right) < 0,$$

for $\tau > 0$ and $\lambda > 0$.

P3) $P(y, \lambda, \tau)$ is a convex function equal to τ for $\lambda = 0$, i.e., $P(y, 0, \tau) = \tau$.

Remark 1.2. *For any $\lambda \geq 0$, $y \geq 0$ and $\tau > 0$. We have $\tau^2 > 0$, so we can obtain the following inequalities*

$$(\lambda y)^2 < (\lambda y)^2 + \tau^2,$$

follow,

$$-\lambda y - \sqrt{(\lambda y)^2 + \tau^2} < 0 < -\lambda y + \sqrt{(\lambda y)^2 + \tau^2}, \tag{1.9}$$

considering the definition of the function P in (1.9), we have

$$P(y, \lambda, \tau) > 0. \tag{1.10}$$

2. Hyperbolic Augmented Lagrangian Function

The hyperbolic augmented Lagrangian function (HALF) $L_H : \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}_{++} \to \mathbb{R}$, is given as follows

$$L_H(x, \lambda, \tau) = f(x) + \sum_{i=1}^{m} P(g_i(x), \lambda_i, \tau).$$

The function L_H is introduced and studied in Xavier (1992), Mallma-Ramirez et al. (2021) and Mallma-Ramirez (2022).

Definition 2.1. *A pair (x^*, λ^*) is said to be a local saddle point of L_H for $\tau > 0$ fixed. If there exists a $\delta > 0$ such that*

$$L_H(x^*, \lambda, \tau) \leq L_H(x^*, \lambda^*, \tau) \leq L_H(x, \lambda^*, \tau), \tag{2.11}$$

holds for all $\lambda \geq 0$ and $x \in X \cap N(x^*, \delta)$.

The following sections are mainly based on the work of Xavier (1992).

2.1. Saddle Point Theory

Theorem 2.1. *If (x^*, λ^*) is a local saddle point of $L_H(x, \lambda, \tau)$ for some $\tau > 0$, then x^* is a local optimal solution to the problem (P).*

Proof. Let (x^*, λ^*) be a local saddle point, see Definition 2.1, then there exists $\delta > 0$ such that

$$L_H(x^*, \lambda, \tau) \leq L_H(x^*, \lambda^*, \tau) \leq L_H(x, \lambda^*, \tau), \tag{2.12}$$

for all $x \in X \cap N(x^*, \delta)$ and $\lambda \geq 0$, with $N(x^*, \delta) = \{x \in I\!R^n \mid \|x - x^*\| \leq \delta\}$.

First, we claim that x^* is a feasible solution to the problem (P). We will show by contradiction, that is, suppose that $g_i(x^*) < 0$ for some i. In this inequality, we apply the property P2, so it follows that

$$P(0, \lambda_i, \tau) < P(g_i(x^*), \lambda_i, \tau), \tag{2.13}$$

by P1 and definition of P in (2.13), we have

$$0 < P(g_i(x^*), \lambda_i, \tau) = -\lambda_i g_i(x^*) + \sqrt{(\lambda_i g_i(x^*))^2 + \tau^2} \to \infty, \quad \lambda_i \to \infty, \tag{2.14}$$

which contradicts the first inequality of (2.12). So, $g_i(x^*) \geq 0$ for all $i = 1, ..., m$.

Henceforth, for any feasible x (i.e., $g_i(x) \geq 0,\ i = 1, ..., m$) and $\lambda_i \geq 0$, we have the following: From (2.12), we get

$$f(x^*) + \sum_{i=1}^{m} P(g_i(x^*), \lambda_i^*, \tau) = L_H(x^*, \lambda^*, \tau)$$

$$\leq L_H(x, \lambda^*, \tau) = f(x) + \sum_{i=1}^{m} P(g_i(x), \lambda_i^*, \tau), \tag{2.15}$$

then by (2.15), it follows that

$$f(x^*) \leq f(x) + \sum_{i=1}^{m} \left(P(g_i(x), \lambda_i^*, \tau) - P(g_i(x^*), \lambda_i^*, \tau)\right), \tag{2.16}$$

from (2.16), let us define

$$W = \sum_{i=1}^{m} \left(P(g_i(x), \lambda_i^*, \tau) - P(g_i(x^*), \lambda_i^*, \tau)\right). \tag{2.17}$$

On the other hand, of (2.12), we have, $L_H(x^*, \lambda, \tau) \leq L_H(x^*, \lambda^*, \tau)$, i.e.,

$$f(x^*) + \sum_{i=1}^{m} P(g_i(x^*), \lambda_i, \tau) \leq f(x^*) + \sum_{i=1}^{m} P(g_i(x^*), \lambda_i^*, \tau),$$

it follows that

$$\sum_{i=1}^{m} P(g_i(x^*), \lambda_i^*, \tau) \geq \sum_{i=1}^{m} P(g_i(x^*), \lambda_i, \tau), \tag{2.18}$$

since that $\lambda_i \geq 0, i = 1, ..., m$, in particular taking $\lambda_i = 0, i = 1, ..., m$ in (2.18), then we obtain the following

$$\sum_{i=1}^{m} P(g_i(x^*), \lambda_i^*, \tau) \geq \sum_{i=1}^{m} P(g_i(x^*), 0, \tau), \tag{2.19}$$

and by P2 in (2.19), we have

$$\sum_{i=1}^{m} P(g_i(x^*), \lambda_i^*, \tau) \geq \sum_{i=1}^{m} \tau. \tag{2.20}$$

We apply the property P2 in $g_i(x) \geq 0, i = 1, ..., m$, so we get

$$\sum_{i=1}^{m} P(0, \lambda_i, \tau) \geq \sum_{i=1}^{m} P(g_i(x), \lambda_i, \tau), \tag{2.21}$$

for $\lambda_i \geq 0,\ i = 1, ..., m$. Now let us consider in particular $\lambda_i = \lambda_i^*,\ i = 1, ..., m$. Thus, it follows from (2.21), that

$$\sum_{i=1}^{m} P(0, \lambda_i^*, \tau) \geq \sum_{i=1}^{m} P(g_i(x), \lambda_i^*, \tau), \tag{2.22}$$

now, we appliying the property P1 in (2.22), we will have

$$\sum_{i=1}^{m} \tau \geq \sum_{i=1}^{m} P(g_i(x), \lambda_i^*, \tau). \tag{2.23}$$

We replace (2.20) in (2.23), thus we obtain

$$0 \geq W = \sum_{i=1}^{m} \left(P(g_i(x), \lambda_i^*, \tau) - P(g_i(x^*), \lambda_i^*, \tau) \right). \tag{2.24}$$

Therefore, replacing (2.24) in (2.16), hence we get $f(x^*) \leq f(x)$, whenever $x \in X \cap N(x^*, \delta)$ is feasible. Therefore, x^* is a local optimal solution of (P). ■

2.1.1. Existence of Local Saddle Point for the HALF

Theorem 2.2. *Let x^* be a local optimal solution to the problem (P). The assumption* **A** *is satisfied at x^*. Then there exist $\bar{\tau} > 0$ and $\delta > 0$ such that for all $\tau < \bar{\tau}$,*

$$L_H(x^*, \lambda, \tau) \leq L_H(x^*, \lambda^*, \tau) \leq L_H(x, \lambda^*, \tau), \tag{2.25}$$

for any $x \in N(x^*, \delta)$ *and* $\lambda \geq 0$.

Proof. By P2 and feasibility of x^* (i.e., $g_i(x^*) \geq 0, \ i = 1, ..., m$), we have

$$P(g_i(x^*), \lambda_i, \tau) \leq P(0, \lambda_i, \tau), \ \ i = 1, ..., m,$$

for any $\tau > 0$ and $\lambda_i \geq 0, \ i = 1, ..., m$. We rewrite the above, as follows

$$\sum_{i=1}^{m} P(g_i(x^*), \lambda_i, \tau) \leq \sum_{i=1}^{m} P(0, \lambda_i, \tau). \tag{2.26}$$

We apply P1 in (2.26), thus we obtain

$$\sum_{i=1}^{m} P(g_i(x^*), \lambda_i, \tau) \leq \sum_{i=1}^{m} \tau,$$

by the above inequality, then we can get

$$f(x^*) + \sum_{i=1}^{m} P(g_i(x^*), \lambda_i, \tau) \leq f(x^*) + \sum_{i=1}^{m} \tau = f(x^*) + m\tau. \tag{2.27}$$

Now, we will prove that, $L_H(x^*, \lambda^*, \tau) = f(x^*) + m\tau$, thus, the first inequality in (2.25) will be verified. Indeed, for $i \in J(x^*)$ we have $\lambda_i^* > 0$ and $g_i(x^*) = 0$. By definition of P and by P1 we have

$$P(g_i(x^*), \lambda_i^*, \tau) = \tau, \ \ i = 1, ..., m,$$

next, we can get

$$\sum_{i=1}^{m} P(g_i(x^*), \lambda_i^*, \tau) = \sum_{i=1}^{m} \tau = m\tau. \tag{2.28}$$

Now, if $i \notin J(x^*)$, we have $\lambda_i^* = 0$. Then by P3, we have

$$P(g_i(x^*), \lambda_i^*, \tau) = \tau, \ \ i = 1, ..., m,$$

next, we can get

$$\sum_{i=1}^{m} P(g_i(x^*), \lambda_i^*, \tau) = \sum_{i=1}^{m} \tau = m\tau. \tag{2.29}$$

From (2.28) and (2.29), we have

$$L_H(x^*, \lambda^*, \tau) = f(x^*) + \sum_{i=1}^{m} P(g_i(x^*), \lambda_i^*, \tau) = f(x^*) + m\tau. \tag{2.30}$$

We replace (2.30) in (2.27), so we get

$$L_H(x^*,\lambda,\tau) \le f(x^*) + m\tau = L_H(x^*,\lambda^*,\tau),$$

so, the first inequality of (2.25) holds, for any $\tau > 0$.

In what follows, we will prove the second inequality of (2.25). By the definition of L_H, by (1.4) and (1.5), we have that

$$\nabla_x L_H(x^*,\lambda^*,\tau) = \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^{m} \frac{(\lambda_i^*)^2 g_i(x^*)}{\sqrt{(\lambda_i^* g_i(x^*))^2 + \tau^2}} \nabla g_i(x^*),$$

we rewrite the above as follows

$$\nabla_x L_H(x^*,\lambda^*,\tau) = \nabla f(x^*) - \sum_{i\in J(x^*)} \lambda_i^* \nabla g_i(x^*) - \sum_{i\notin J(x^*)} \lambda_i^* \nabla g_i(x^*)$$

$$+ \sum_{i\in J(x^*)} \frac{(\lambda_i^*)^2 g_i(x^*)}{\sqrt{(\lambda_i^* g_i(x^*))^2 + \tau^2}} \nabla g_i(x^*) + \sum_{i\notin J(x^*)} \frac{(\lambda_i^*)^2 g_i(x^*)}{\sqrt{(\lambda_i^* g_i(x^*))^2 + \tau^2}} \nabla g_i(x^*),$$

so, we have

$$\nabla_x L_H(x^*,\lambda^*,\tau) = \nabla f(x^*) - \sum_{i\in J(x^*)} \lambda_i^* \nabla g_i(x^*) = \nabla_x L(x^*,\lambda^*) = 0. \tag{2.31}$$

Now, let's prove that there exists $\bar{\tau} > 0$ such that $\nabla^2_{xx} L_H(x^*,\lambda^*,\bar{\tau})$ is positive definite. Indeed, the Hessian of $L_H(x,\lambda,\tau)$ in the point (x^*,λ^*), is

$$\nabla^2_{xx} L_H(x^*,\lambda^*,\tau) = \nabla^2_{xx} f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla^2_{xx} g_i(x^*)$$

$$+ \sum_{i=1}^{m} \left(\frac{(\lambda_i^*)^2}{\sqrt{(\lambda_i^* g_i(x^*))^2 + \tau^2}} - \frac{(\lambda_i^*)^4 g_i^2(x^*)}{((\lambda_i^* g_i(x^*))^2 + \tau^2)^{\frac{3}{2}}} \right) (\nabla_x g_i(x^*))^2$$

$$+ \sum_{i=1}^{m} \frac{(\lambda_i^*)^2 g_i(x^*)}{\sqrt{(\lambda_i^* g_i(x^*))^2 + \tau^2}} \nabla^2_{xx} g_i(x^*),$$

we rewrite the above as follows,

$$\nabla^2_{xx} L_H(x^*,\lambda^*,\tau) = \nabla^2_{xx} f(x^*) - \sum_{i\in J(x^*)} \lambda_i^* \nabla^2_{xx} g_i(x^*) - \sum_{i\notin J(x^*)} \lambda_i^* \nabla^2_{xx} g_i(x^*)$$

$$+\sum_{i\in J(x^*)}\left(\frac{(\lambda_i^*)^2}{\sqrt{(\lambda_i^* g_i(x^*))^2+\tau^2}}\right)(\nabla_x g_i(x^*))^2$$

$$+\sum_{i\notin J(x^*)}\left(\frac{(\lambda_i^*)^2}{\sqrt{(\lambda_i^* g_i(x^*))^2+\tau^2}}\right)(\nabla_x g_i(x^*))^2$$

$$-\sum_{i\in J(x^*)}\left(\frac{(\lambda_i^*)^4(g_i(x^*))^2}{((\lambda_i^* g_i(x^*))^2+\tau^2)^{\frac{3}{2}}}\right)(\nabla_x g_i(x^*))^2$$

$$-\sum_{i\notin J(x^*)}\left(\frac{(\lambda_i^*)^4(g_i(x^*))^2}{((\lambda_i^* g_i(x^*))^2+\tau^2)^{\frac{3}{2}}}\right)(\nabla_x g_i(x^*))^2$$

$$+\sum_{i\in J(x^*)}\frac{(\lambda_i^*)^2 g_i(x^*)}{\sqrt{(\lambda_i^* g_i(x^*))^2+\tau^2}}\nabla_{xx}^2 g_i(x^*)+\sum_{i\notin J(x^*)}\frac{(\lambda_i^*)^2 g_i(x^*)}{\sqrt{(\lambda_i^* g_i(x^*))^2+\tau^2}}\nabla_{xx}^2 g_i(x^*), \quad (2.32)$$

we immediately consider (1.5) and (1.6) in (2.32), thus one has

$$\nabla_{xx}^2 L_H(x^*,\lambda^*,\tau)=\nabla_{xx}^2 L(x^*,\lambda^*)+\frac{1}{\tau}\sum_{i\in J(x^*)}(\nabla g_i(x^*)\lambda_i^*)(\nabla g_i(x^*)\lambda_i^*)^T. \quad (2.33)$$

In (2.33), defining $A=\nabla_{xx}^2 L(x^*,\lambda^*)$ and $B=\nabla g_i(x^*)\lambda_i^*$, and by assumption **A**, we have, $y^T\nabla_{xx}^2 L(x^*,\lambda^*)y>0$ for every $y\neq 0$ such that $By=0$. Then by Corollary 1.1,

$$\nabla_{xx}^2 L(x^*,\lambda^*)+cB^TB,$$

is positive definite for some scalar c sufficiently large. For our case, we take

$$\frac{1}{\tau}=c>c^*=\frac{1}{\bar{\tau}},$$

i.e., there exists $\bar{\tau}$ such that for all $\tau<\bar{\tau}$ the HALF, $L_H(x,\lambda,\tau)$ is positive definite. Now combined with (2.31), we have that x^* is a (strict) local minimizer of $L_H(x,\lambda^*,\bar{\tau})$ and then the second inequality of (2.25) holds. ■

Acknowledgement(s)

This work was partially supported by *Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq*; FAPERJ E-26/201.176/2020 (258168) and COPPETEC Foundation, Brazil.

References

Arrow, K. J., Gould, F. J., Howe, S. M. (1973). A general saddle point result for constrained optimization. *Mathematical Programming*, *5*, 225–234.

Avriel, M. (1976). *Nonlinear Programming: Analysis and Methods.* Prentice Hall, Englewood Cliffs.

Echebest, N., Sánchez, M. D., Schuverdt, M. L. (2016). Convergence results of an augmented Lagrangian method using the exponential penalty function. *J. Optim. Theory Appl.*, *168*, 92–108.

Mallma-Ramirez, L., Maculan, N., Xavier, A.E. (2021). Convergence analysis of the hyperbolic augmented Lagrangian algorithm. *Technical Report.* PESC-COPPE-UFRJ. ES-2991/21 (https://www.cos.ufrj.br/index.php/pt-BR/publicacoes-pesquisa/details/15/2991)

Mallma-Ramirez, L. (2022). *The Hyperbolic Augmented Lagrangian Algorithm* (Ph.D. thesis). Federal University of Rio de Janeiro/COPPE, Rio de Janeiro, Brazil.

Sun, X. L., Li, D., Mckinnon, K.I. M. (2005). On saddle points of augmented Lagrangians for constrained nonconvex optimization. *SIAM J. OPTIM.*, *15*, 1128–1146.

Xavier, A. E. (1982a). *Penalização Hiperbólica- Um Novo Método para Resolução de Problemas de Otimização* (master's thesis). Federal University of Rio de Janeiro \ COPPE, Rio de Janeiro, Brazil.

Xavier, A. E. (1992). *Penalização Hiperbólica* (Ph.D. thesis). Federal University of Rio de Janeiro/COPPE, Rio de Janeiro, Brazil.

Xavier, A. E. (2001). Hyperbolic penalty: a new method for nonlinear programming with inequalities. *Intl. Trans. in Op. Res.*, *8*, 659–671.

Zangwill, W. I. (1967). Non-linear programming via penalty functions. *Management Sciences*, *13*, 344–358.