# Set Containment Characterization

O. L. Mangasarian*

## Abstract

Characterization of the containment of a polyhedral set in a closed halfspace, a key factor in generating knowledge-based support vector machine classifiers [7], is extended to the following:

(i) Containment of one polyhedral set in another.

(ii) Containment of a polyhedral set in a reverse-convex set defined by convex quadratic constraints.

(iii) Containment of a general closed convex set, defined by convex constraints, in a reverse-convex set defined by convex nonlinear constraints.

The first two characterizations can be determined in polynomial time by solving $m$ linear programs for $(i)$ and $m$ convex quadratic programs for $(ii)$, where $m$ is the number of constraints defining the containing set. In $(iii)$, $m$ convex programs need to be solved in order to verify the characterization, where again $m$ is the number of constraints defining the containing set. All polyhedral sets, like the *knowledge sets* of support vector machine classifiers, are characterized by the intersection of a finite number of closed halfspaces.

**Keywords** ***set containment, knowledge-based classifier, linear programming, quadratic programming***

## 1 Introduction

Support vector machine classifiers [15, 1, 12] generate separating planes or surfaces by training on labeled data, that is data for which the class of each point is given. Knowledge-based classifiers [13, 14] on the other hand utilize prior knowledge, e.g. an expert's experience such as a doctor's knowledge in diagnosing a certain disease. Recently [7] a precise incorporation of prior knowledge into a linear support vector machine classifier was achieved by placing nonempty polyhedral sets representing such knowledge in the correct halfspace determined by a separating plane classifier. Key to this approach was a dual characterization, using the nonhomogeneous Farkas theorem [11, Theorem 2.4.8], of the containment of a polyhedral set in a closed halfspace. This characterization was then used as a constraint in a linear program that determined the linear classifier thereby incorporating prior knowledge into the classifier. In Section 2 we extend this characterization to the containment of one polyhedral set in another (Figure 1). In Section 3 we characterize the containment of a polyhedral set in a reverse-convex constraint [11, Definition 7.3.5] set determined

*Computer Sciences Department, University of Wisconsin, 1210 West Dayton Street, Madison, WI 53706, *olvi@cs.wisc.edu*.

by quadratic constraints (Figure 2), and in Section 4 we characterize the containment of a convex set determined by nonlinear convex functions in a reverse-convex constraint set determined by nonlinear functions (Figure 3). An interesting aspect of the present results is that, despite the nonconvexity of the containing set of Sections 3 and 4, the containment problems of these sections can be solved by a finite number of polynomial-time convex quadratic programs and by a finite number of convex programs respectively. The case of Section 2, containment of one polyhedral set in another, can be solved by a finite number of linear programs.

There have been other set containment studies which emphasize the complexity issue of the problem. Notable among those is the work of Freund and Orlin [6] regarding the containment of polyhedral sets in balls and vice-versa, and the inner and outer radii of convex bodies by Grtizmann and Klee [8]. In [4, Lemma, p 140] the nonhomogeneous Farkas theorem was also used for the optimal scaling of balls and polytopes.

We now describe our notation. All vectors will be column vectors unless transposed to a row vector by a prime $'$. The scalar (inner) product of two vectors $x$ and $y$ in the $n$-dimensional real space $R^n$ will be denoted by $x'y$. The notation $A \in R^{m\times n}$ will signify a real $m \times n$ matrix. For such a matrix, $A'$ will denote the transpose of $A$, $A_i$ will denote the $i$-th row of $A$ and $A_{\cdot j}$ will denote the $j$-th column of $A$. The identity matrix of arbitrary dimension will be denoted by $I$. For simplicity, the dimensionality of some vectors will not be explicitly given. For a vector function $h : R^n \longrightarrow R^k$, $\nabla h(x)$ will denote the $k \times n$ Jacobian matrix of first partial derivatives, and $h$ is said to be convex on $R^n$ if each of its $k$ components are convex on $R^n$.

## 2 Polyhedral Set Containment

In this section we generalize the nonhomogeneous Farkas theorem [11, Theorem 2.4.8], a result that we used earlier to generate a knowledge-based support vector machine classifier [7]. The nonhomogeneous Farkas Theorem gives a dual characterization of the containment of a nonempty polyhedral knowledge set in a closed halfspace. Proposition 1 below generalizes this latter result, using linear programming duality, to containment of a nonempty polyhedral set in an arbitrary polyhedral set depicted in Figure 1, instead of containment in a closed halfspace.

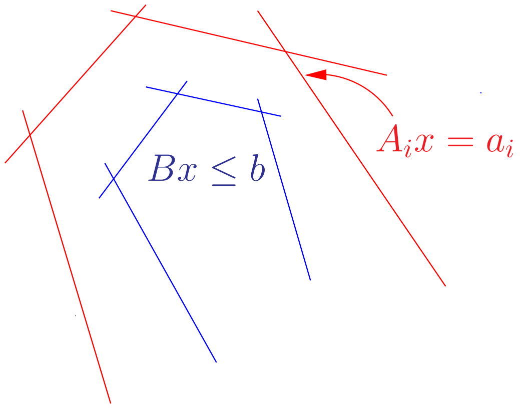


Figure 1: **Containment of the polyhedral set $\mathcal{B} := \{x \mid Bx \leq b\}$ in another polyhedral set $\mathcal{A} := \{x \mid Ax \geq a\}$.**

**Proposition 2.1 Polyhedral Set Containment** *Let the set* $\mathcal{A} := \{x \mid Ax \geq a\}$ *and let* $\mathcal{B} := \{x \mid Bx \leq b\}$*, where* $A \in R^{m \times n}$*,* $B \in R^{k \times n}$ *and let* $\mathcal{B}$ *be nonempty. Then the following are equivalent:*

*(i)* $\mathcal{B} \subseteq \mathcal{A}$*, that is:*

$$Bx \leq b \implies Ax \geq a. \tag{1}$$

*(ii) There exists a matrix* $U \in R^{m \times k}$ *such that:*

$$A + UB = 0, \ a + Ub \leq 0, \ \ U \geq 0. \tag{2}$$

*(iii) For* $i = 1, \ldots, m$*, the* $m$ *linear programs are solvable and satisfy:*

$$\min_x \{(A_i x - a_i) \mid Bx \leq b\} \geq 0. \tag{3}$$

**Proof** (($i$) $\implies$ ($iii$)) For $i \in \{1, \ldots, m\}$, the $m$ linear programs of (3) are feasible because $\mathcal{B} \neq \emptyset$ and their objective functions are bounded below by zero and hence attain their nonnegative minima as asserted by (3).
(($iii$) $\implies$ ($ii$)) By linear programming duality [3, 11], for $i = 1, \ldots, m$, each of the $m$ linear programs that are dual to the $m$ linear programs of (3) are solvable and satisfy:

$$\max_u \{(-b'u - a_i) \mid -B'u = A_i', \ u \geq 0\} \geq 0. \tag{4}$$

Calling the solution of each of these $m$ dual linear programs $u^i \in R^k$, $i = 1, \ldots, m$, and defining the $m \times k$ matrix $U$ as $U' = [u^1 \ldots u^m]$, we obtain:

$$-b'U' - a' \geq 0, \ -B'U' = A', \ U \geq 0, \tag{5}$$

which is equivalent to (2).
(($ii$) $\implies$ ($i$))

$$Bx \leq b \implies Ax = -UBx \geq -Ub \geq a. \tag{6}$$

□

**Remark 2.2** *It is interesting to note that even though the validity of the polyhedral set containment implication of problem (1) can be resolved by the above proposition in polynomial time by solving the* $m$ *linear programs (3), it can also be characterized by solving the following* **minimization** *of a piecewise linear* **concave** *function on a polyhedral set:*

$$\min_x \{ \min_{i=1,\ldots,m} \{A_i x - a_i\} \mid Bx \leq b\} \geq 0. \tag{7}$$

*General piecewise linear concave minimization on a polyhedral set is NP-hard because the general linear complementarity problem, which is NP-complete, [2] can be formulated as such a problem [10, Lemma 1].*

We turn now to the characterization of the containment of a polyhedral set in a quadratically determined nonconvex set.

## 3 Containment of a Polyhedral Set in a Nonconvex Set Determined by Quadratic Constraints

We characterize now the containment of a polyhedral set in a nonconvex set determined by convex quadratic quadratic constraints generating a reverse-convex [11, Definition 7.3.5] set as depicted in Figure 2. An interesting aspect of this nonconvex problem is that it is solvable in polynomial time as a consequence of the following characterization result.

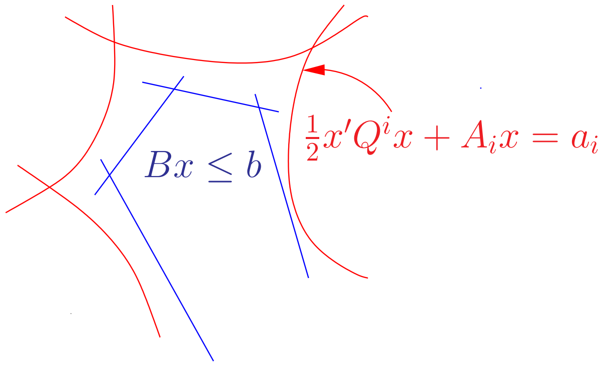


Figure 2: **Containment of the polyhedral set $\mathcal{B} := \{x \mid Bx \leq b\}$ in the reverse-convex quadratic set $\mathcal{A} := \{x \mid \frac{1}{2}x'Q^ix + A_ix \geq a_i,\ i = 1, \ldots, m\}$, where $Q^i$ are positive semidefinite symmetric matrices.**

**Proposition 3.1 Polyhedral Set Containment in Reverse Convex Quadratic Set** *Let the set $\mathcal{B} := \{x \mid Bx \leq b\}$ be nonempty and let $\mathcal{A} := \{x \mid \frac{1}{2}x'Q^ix + A_ix \geq a_i,\ i = 1, \ldots, m\}$, where $Q^i \in R^{n\times n}$, $i = 1, \ldots, m$ are symmetric positive semidefinite matrices. Then the following are equivalent:*

*(i)* $\mathcal{B} \subseteq \mathcal{A}$, *that is:*

$$Bx \leq b \implies \frac{1}{2}x'Q^ix + A_ix \geq a_i,\ i = 1, \ldots, m. \tag{8}$$

*(ii) There exist matrices $U \in R^{m\times k}$, $X \in R^{m\times n}$ such that:*

$$A_i + U_iB + X_iQ^i = 0,\ a_i + U_ib + \frac{1}{2}X_iQ^iX_i' \leq 0,\ U_i \geq 0,\ i = 1, \ldots, m. \tag{9}$$

*(iii) For $i = 1, \ldots, m$, the $m$ convex quadratic programs are solvable and satisfy:*

$$\min_x \{(\frac{1}{2}x'Q^ix + A_ix - a_i) \mid Bx \leq b\} \geq 0. \tag{10}$$

**Proof** ($(i) \implies (iii)$) For $i \in \{1, \ldots, m\}$, the $m$ quadratic programs of (10) are feasible because $\mathcal{B} \neq \emptyset$ and their objective functions are bounded below by zero and hence attain [5] their nonnegative minima as asserted in (10).
($(iii) \implies (ii)$) By quadratic programming duality [11, Section 8.2], for $i = 1, \ldots, m$, the $m$ quadratic programs that are dual to the $m$ quadratic programs (10) are solvable and satisfy:

$$\max_{x,u} \{(-\frac{1}{2}x'Q^ix - b'u - a_i) \mid Q^ix + B'u + A_i' = 0,\ u \geq 0\} \geq 0. \tag{11}$$

Calling the solution of each of these $m$ dual quadratic programs $x^i \in R^n, u^i \in R^k$, $i = 1, \ldots, m$, and defining the $m \times n$ matrix $X$ as $X' = [x^1 \ldots x^m]$, and the $m \times k$ matrix $U$ as $U' = [u^1 \ldots u^m]$, we obtain that for $i = 1, \ldots, m$:

$$-\frac{1}{2}X_i Q^i X_i' - b'U_i' - a_i \geq 0,\ Q^i X_i' + B'U_i' + A_i' = 0,\ U_i \geq 0, \tag{12}$$

which is equivalent to (9).
($(ii) \implies (i)$) For $i = 1, \ldots, m$:

$$\begin{array}{rcl} Bx \leq b & \implies & \frac{1}{2}x'Q^i x + A_i x - a_i \geq -\frac{1}{2}X_i Q^i X_i' - b'U_i' - a_i \\ & \implies & \frac{1}{2}x'Q^i x + A_i x - a_i \geq 0 \end{array} \tag{13}$$

where the first implication follows from the weak duality theorem [11, Theorem 8.23.], since $x$ is feasible for each of the $m$ primal quadratic programs of (10), while $(X_i, U_i)$, $i = 1, \ldots, m$, are feasible for the $m$ dual programs (11). The second implication above follows because $-\frac{1}{2}X_i Q^i X_i' - b'U_i' - a_i \geq 0$ by (9) of $(ii)$. The two implications of (13) above result in (8) of $(i)$. $\square$

We note that an interesting consequence of this proposition, is that the the containment of a polyhedral set in a nonconvex set determined by quadratic constraints can be solved in polynomial time by solving the $m$ convex quadratic programs (10) [9].

We turn finally to the containment of a general convex set in a general nonlinear reverse-convex set.

## 4 Containment of a General Convex Set in a Nonconvex Set Determined by Nonlinear Constraints

We consider a general closed convex set $\mathcal{B}$ in $R^n$ and and characterize its containment in a general nonconvex set $\mathcal{A}$ depicted in Figure 3 as follows.

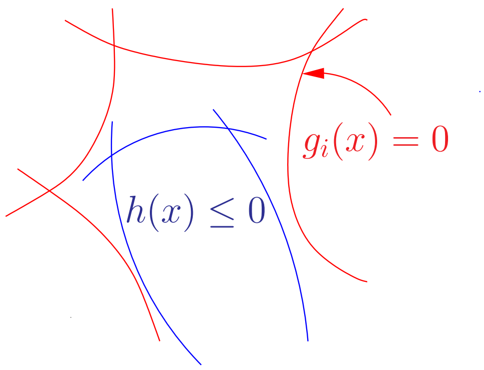


Figure 3: **Containment of the convex set $\mathcal{B} := \{x \mid h(x) \leq 0\}$ in the reverse-convex nonlinear set $\mathcal{A} := \{x \mid g(x) \geq 0\}$, where $g : R^n \longrightarrow R^m$ and $h : R^n \longrightarrow R^k$ are convex functions on $R^n$.**

**Proposition 4.1 Convex Set Containment in Reverse Convex Set** *Let be $\mathcal{B}$ be a nonempty closed convex set in $R^n$ defined as $\mathcal{B} := \{x \mid h(x) \leq 0\}$, where $h : R^n \longrightarrow R^k$ is a differentiable* *convex function on $R^n$, and let the the nonconvex set $\mathcal{A}$ in $R^n$ be defined as $\mathcal{A} := \{x \mid g(x) \geq 0\}$, where $g : R^n \longrightarrow R^m$ is a differentiable convex function on $R^n$. Then,*

$$(i) \iff (iii) \impliedby (ii), \tag{14}$$

*where:*

*(i) $\mathcal{B} \subseteq \mathcal{A}$, that is:*

$$h(x) \leq 0 \implies g(x) \geq 0. \tag{15}$$

*(ii) For $i = 1, \ldots, m$, there exist $x^i \in R^n$ and $u^i \in R^k$, such that:*

$$\nabla g_i(x^i) + u^{i'}\nabla h(x^i) = 0,\ g_i(x^i) + u^{i'}h(x^i) \geq 0,\ u^i \geq 0. \tag{16}$$

*(iii) For $i = 1, \ldots, m$, the $m$ convex programs satisfy:*

$$\inf_x \{g_i(x) \mid h(x) \leq 0\} \geq 0. \tag{17}$$

*If in addition $g_i$, $i = 1, \ldots, m$, have bounded level sets on $\mathcal{B}$, that is:*

$$\{x \mid g_i(x) \leq \alpha,\ h(x) \leq 0\},\ i = 1, \ldots, m,\ \textit{are bounded for each } \alpha, \tag{18}$$

*and*

$$\{x \mid h(x) < 0\} \neq \emptyset, \quad \textit{or } h(x) \textit{ is linear}, \tag{19}$$

*then,*

$$(i) \iff (iii) \iff (ii). \tag{20}$$

**Proof** ($(i) \implies (iii)$) If not, then for some $i \in \{1, \ldots, m\}$, there exists an $x$ such that:

$$g_i(x) < 0,\ h(x) \leq 0, \tag{21}$$

which contradicts the implication (15).
($(i) \impliedby (iii)$) $h(x) \leq 0 \implies g_i(x) \geq 0, i = 1, \ldots, m$, which is implication (15) of (i).
($(ii) \implies (iii)$) For $i \in \{1, \ldots, m\}$, the $m$ points $(x^i, u^i)$ given by (16) of $(ii)$ are feasible for the dual problems to (17):

$$\sup_{(x,u)\in R^{n+k}} \{g_i(x) + u'h(x) \mid \nabla g_i(x) + u'\nabla h(x) = 0,\ u \geq 0\} \geq 0,\ i = 1, \ldots, m, \tag{22}$$

with dual objective function values that are nonnegative. Hence by the weak duality theorem of convex programming [11, Theorem 8.1.3], the corresponding $m$ primal problems (17) with the nonempty feasible region $\mathcal{B}$ have infima bounded below by zero which implies $(iii)$.
($(iii) \implies (ii)$) Let $\alpha^i \geq 0$, $i = 1, \ldots, m$, be the infima of each of the $m$ problems of (17). Hence for each $i = 1, \ldots, m$, there exists a sequence $\{\epsilon^i_j\} \downarrow 0$, $\{x^i_j\} \in \mathcal{B}$, such that:

$$\alpha^i \leq g_i(x^i_j) < \epsilon^i_j + \alpha^i. \tag{23}$$

Since the sequence $\{x^i_j\}$ lies in the closed bounded set $\mathcal{B} \cap \{g_i(x) \leq \epsilon^i_0 + \alpha^i\}$, it must have an accumulation point $x^i \in \mathcal{B}$ such that $\alpha_i = g_i(x^i) = \inf_x \{g_i(x) \mid h(x) \leq 0\}$. Hence, for $i = 1, \ldots, m$, $g_i(x^i)$ is an attained infimum $\alpha_i$ of (17), and since a constraint qualification (19) is satisfied, it follows by Wolfe's duality theorem of convex programming [11, Theorem 8.1.4] that the supremum $\alpha_i$ of the dual problem (22) is attained at $x_i$ and some $u^i$. Hence $(x^i, u^i)$, $i = 1, \ldots, m$, satisfy (16) of (ii).$\square$

## 5 Conclusion

We have proposed computationally tractable characterizations of set containment properties for both polyhedral and nonlinear sets. Polyhedral set containment in another polyhedral set is characterized by the solution of a finite number of linear programs. Containment of a polyhedral set in a reverse-convex set, defined by convex quadratic constraints, is characterized by polynomial time solution of a finite number of convex quadratic programs. Containment of a general closed convex set, defined by convex constraints, in a reverse-convex set defined by convex constraints, is characterized by solving a finite number of convex programs. These results, motivated by knowledge-based linear classification, may possibly lead to general methods of incorporating more complex knowledge into both linear and nonlinear classifiers and merit further study.

## Acknowledgements

The research described in this Data Mining Institute Report 01-10, November 2001, was supported by National Science Foundation Grants CCR-9729842 and CDA-9623632, by Air Force Office of Scientific Research Grant F49620-00-1-0085 and by the Microsoft Corporation.

## References

[1] V. Cherkassky and F. Mulier. *Learning from Data - Concepts, Theory and Methods*. John Wiley & Sons, New York, 1998.

[2] S.-J. Chung. NP-completeness of the linear complementarity problem. *Journal of Optimization Theory and Applications*, 60:393–399, 1989.

[3] G. B. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, New Jersey, 1963.

[4] B. C. Eaves and R. M. Freund. Optimal scaling of balls and polyhedra. *Mathematical Programming*, 23:138–147, 1982.

[5] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3:95–110, 1956.

[6] R. M. Freund and J. B. Orlin. On the complexity of four polyhedral set containment problems. *Mathematical Programming*, 33:139–145, 1985.

[7] G. Fung, O. L. Mangasarian, and J. Shavlik. Knowledge-based suppor vector machine classifiers. Technical Report 01-09, Data Mining Institute, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, November 2001. ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/01-09.ps.

[8] P. Gritzmann and V. Klee. Inner and outer j-radii of convex bodies in finite-dimensional normed spaces. *Discrete and Computational Geometry*, 7:255–280, 1992.

[9] M. Kojima, S. Mizuno, and A. Yoshise. A polynomial–time algorithm for a class of linear complementarity problems. *Mathematical Programming*, 44:1–26, 1989.

[10] O. L. Mangasarian. Characterization of linear complementarity problems as linear programs. *Mathematical Programming Study*, 7:74–87, 1978.

[11] O. L. Mangasarian. *Nonlinear Programming.* SIAM, Philadelphia, PA, 1994.

[12] O. L. Mangasarian. Generalized support vector machines. In A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, pages 135–146, Cambridge, MA, 2000. MIT Press. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-14.ps.

[13] G. G. Towell and J. W. Shavlik. The extraction of refined rules from knowledge-based neural networks. *Machine Learning*, 13:71–101, 1993.

[14] G. G. Towell and J. W. Shavlik. Knowledge-based artificial neural networks. *Artificial Intelligence*, 70:119–165, 1994.

[15] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer, New York, second edition, 2000.