The Solution Path of the Slab Support Vector Machine

Michael Eigensatz* Joachim Giesen† Madhusudan Manjunath‡

Abstract

Given a set of points in a Hilbert space that can be separated from the origin. The slab support vector machine (slab SVM) is an optimization problem that aims at finding a slab (two parallel hyperplanes whose distance—the slab width—is essentially fixed) that encloses the points and is maximally separated from the origin. Extreme cases of the slab SVM include the smallest enclosing ball problem and an interpolation problem that was used (as the slab SVM itself) in surface reconstruction with radial basis functions. Here we show that the path of solutions of the slab SVM, i.e., the solution parametrized by the slab width is piecewise linear.

1 Introduction

Data structures used in fields like graphics, visualization and learning often have many free parameters. In most cases a good choice of these parameters is not obvious. Computational geometry was facing similar problems: for example when using alpha shapes [Ede95] for surface reconstruction or in bio-geometric modeling the question arises as to what value to choose for alpha. Computational geometry [Ede95, ELZ02, GCPZ06] gave an answer to this question that can be seminal also for the aforementioned areas of computer science, namely, do not compute the solution for a fixed more or less well chosen value of the parameter, but compute the whole spectrum of structures and then look for good solutions in this spectrum. One method to determine a good structure is topological persistence pioneered by Edelsbrunner, Harer and Zomorodian [ELZ02].

Here we investigate an optimization problem that has its roots in machine learning and was also applied in various forms to the surface reconstruction problem. The problem is called slab support vector machine (slab SVM) [SGS04] and takes as input a set of data points in a Hilbert space that can be separated from the origin and aims at finding a slab (two parallel hyperplanes whose width is essentially fixed as $\delta > 0$) that encloses the points and is maximally separated from the origin. The slab SVM has found applications in surface reconstruction [SGS04], and quantile estimation and novelty detection [SS02]. In these applications the data points reside in d-dimensional Euclidean space but are mapped by a feature map into another (often infinite dimensional) Hilbert space. The structure of the slab SVM is such that the feature map does not have to be given explicitly, but only implicitly through a positive kernel: the dual optimization problem of the slab SVM depends only on the pairwise inner products of the data points. A positive kernel can be used to replace these inner products without changing the nature (convex quadratic program) of the optimization problem.

The parameter we are interested in is δ, which essentially fixes the width of the slab. In the applications, it is difficult to tell beforehand what a good choice of δ is. Hence in the spirit of the computational geometry approach we want to compute the solution to the slab SVM for all values of δ. Once we have this spectrum of solutions other methods can be employed to find good choices for δ. Here we do not want to discuss how such methods could look like, but focus on computing the solution spectrum. We show that the solution path of the slab SVM, i.e., the solution parametrized by δ is piecewise linear. Our arguments provide a complete geometric characterization of the turning points (nodes) of the solution path.

Our results are in spirit similar to results of Hastie et al. [HRTZ04] who obtained the piecewise linearity of the solution of the classification support vector machine [SS02]. Though both results give piecewise linear solution paths, the parameters are different in nature and so are the means to establish the results. Our proof is of geometric nature, whereas Hastie et al. use algebraic arguments.

2 The slab SVM

Given *data points* $X = \{x_1, \dots, x_n\} \subset \mathcal{H}$, where $\mathcal{H}$ is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$, such that the data points can be separated from the origin by a hyperplane, i.e., there exists $w \in \mathcal{H} \setminus \{0\}$ and $\rho \neq 0$ such that

$$\langle w, x_i \rangle \geq \rho \text{ for all } i = 1, \dots, n.$$

The distance of the hyperplane $\{x \in \mathcal{H} : \langle w, x \rangle = \rho\}$ to the origin of $\mathcal{H}$ is given as $\rho / \|w\|$, where the norm of

*Applied Geometry Group, ETH Zürich, eigensatz@inf.ethz.ch

†Institut für Informatik, Friedrich-Schiller-Universität Jena, giesen@minet.uni-jena.de

‡Max-Planck Institut für Informatik, manjun@mpi-inf.mpg.de

w in $\mathcal{H}$ is defined as usual by $\|w\| = \sqrt{\langle w, w\rangle}$.

The *slab SVM* is the following convex quadratic optimization problem that aims at finding the slab (the space between two parallel hyperplanes) with width $\delta/\|w\|$ that contains all the data points and minimizes $\frac{1}{2}\|w\|^2 - \rho$, i.e., essentially maximizes the distance of the slab to the origin (see also Figure 1):

$$\begin{array}{ll} \min_{w,\rho} & \frac{1}{2}\|w\|^2 - \rho \\ \text{s.t.} & \rho \leq \langle w, x_i\rangle \leq \rho + \delta \text{ for all } i = 1, \ldots, n \end{array}$$

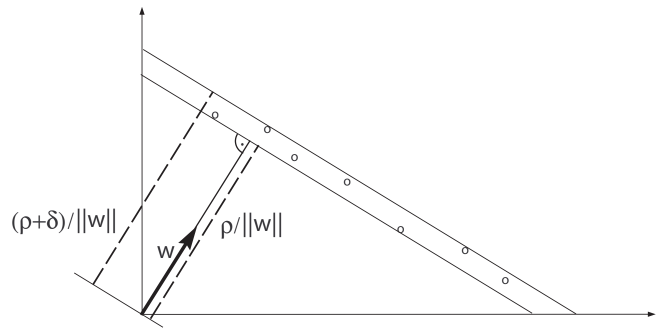

Figure 1: The geometric set-up for the slab SVM.

Note that the slab SVM problem is always feasible since $(w, \rho) = (0, 0)$ is always contained in the constraint polytope.

The Lagrangian dual to this problem can be derived from the saddle point condition for the Lagrangian

$$\begin{aligned} L(w, \rho, \alpha, \beta) &= \frac{1}{2}\|w\|^2 - \rho - \sum_{i=1}^{n} \alpha_i(\langle w, x_i\rangle - \rho) \\ &\quad + \sum_{i=1}^{n} \beta_i(\langle w, x_i\rangle - \rho - \delta), \end{aligned}$$

where $\alpha_i, \beta_i \geq 0$. The saddle point condition gives $\partial L/\partial w = 0$ which implies $w = \sum_{i=1}^{n}(\alpha_i - \beta_i)x_i$ and $\partial L/\partial \rho = 0$ which implies $\sum_{i=1}^{n}(\alpha_i - \beta_i) = 1$ from which the dual follows

$$\begin{array}{ll} \min_{\alpha,\beta} & \frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \beta_i)(\alpha_j - \beta_j)\langle x_i, x_j\rangle + \delta\sum_{i=1}^{n}\beta_i \\ \text{s.t.} & \alpha_i, \beta_i \geq 0 \text{ for all } i = 1, \ldots, n. \\ & \sum_{i=1}^{n}(\alpha_i - \beta_i) = 1 \end{array}$$

In most applications [SS02] the data points are obtained from applying a feature map ϕ to input data points $y_1, \ldots, y_n \in \mathbb{R}^d$, i.e., $x_i = \phi(y_i) \in \mathcal{H}$, where the feature map is not given explicitly, but implicitly in form of a positive kernel function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, i.e.,

$$\langle x_i, x_j\rangle = \langle \phi(y_i), \phi(y_j)\rangle = k(x_i, x_j).$$

and $\mathcal{H}$ is the kernel reproducing Hilbert space. Since the dual of the slab SVM only depends on the inner products of the data points, we can replace $\langle x_i, x_j\rangle$ by $k(x_i, x_j)$. A popular positive kernel is the Gaussian

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right),$$

which is an example of a so called radial basis function kernel, i.e., a kernel that only depends on the distance $\|x_i - x_j\|$. The data points $x_i = \phi(y_i)$ are linearly independent and the Gram matrix $\big(k(x_i, x_j)\big)$ associated with the Gaussian kernel is positive, i.e., it has full rank and thus is invertible. In the following, we always assume that the data points x_i are linear independent.

3 Two extreme cases

The objective function of the slab SVM might look somewhat arbitrary at a first glance. Considering the extreme cases $\delta = \infty$ and $\delta = 0$ helps to get a better understanding of the geometry behind it.

3.1 The open slab SVM

We denote as open slab SVM the special case $\delta = \infty$ of the slab SVM, see Figure 2 for the geometric set-up. The open slab SVM optimization problem reads as

$$\begin{array}{ll} \min_{w,\rho} & \frac{1}{2}\|w\|^2 - \rho \\ \text{s.t.} & \rho \leq \langle w, x_i\rangle \text{ for all } i = 1, \ldots, n \end{array}$$

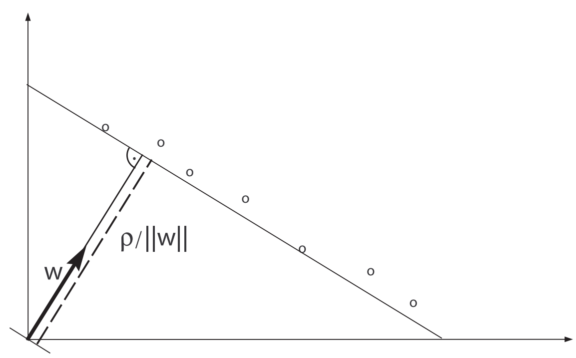

Figure 2: The geometric set-up for the open slab SVM.

As in the general case, we can derive the Lagrangian dual of the open slab SVM and get the following optimization problem in the dual variables $\alpha_i, i = 1, \ldots, n$,

$$\begin{array}{ll} \min_{\alpha} & \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j\langle x_i, x_j\rangle \\ \text{s.t.} & \alpha_i \geq 0 \text{ for all } i = 1, \ldots, n. \\ & \sum_{i=1}^{n}\alpha_i = 1 \end{array}$$

Observe that the optimal value of this problem is the distance of the origin to the convex hull of the data points $x_1, \ldots, x_n$. The saddle point condition $\partial L/\partial w = 0$ implies

$$w = \sum_{i=1}^{n} \alpha_i x_i.$$

Hence the optimal vector w is the shortest vector from the origin to the convex hull of the data points and $\|w\|$ is the distance of origin to this convex hull.

If the data points $x_i \in \mathcal{H}$ are obtained from data points $y_i \in \mathbb{R}^d$ that were mapped to $\mathcal{H}$ by a feature map implicitly given by a radial basis function kernel $r(\cdot)$, i.e., by replacing inner products $\langle y_i, y_j\rangle$ by $r(\|y_i - y_j\|)$, then the open slab SVM is equivalent to computing the smallest enclosing ball of the data points x_i, see [SS02].

3.2 The zero slab SVM

As zero slab SVM we denote the slab SVM for the case $\delta = 0$, see Figure 3 for the geometric set-up. The zero slab SVM optimization problem reads as

$$\begin{array}{ll} \min_{w,\rho} & \frac{1}{2}\|w\|^2 - \rho \\ \text{s.t.} & \langle w, x_i\rangle = \rho \text{ for all } i = 1, \ldots, n \end{array}$$

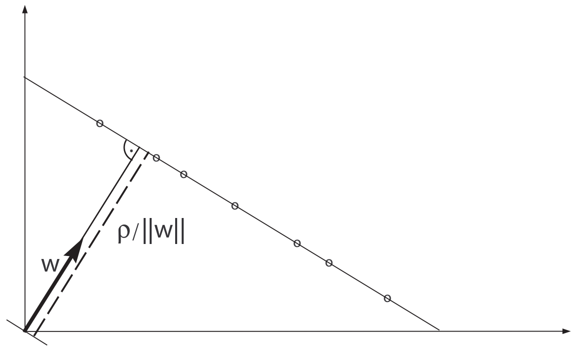

Figure 3: The geometric set-up for the zero slab SVM.

The optimization problem reads in dual variables α_i as

$$\begin{array}{ll} \min_{\alpha} & \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j \langle x_i, x_j\rangle \\ \text{s.t.} & \sum_{i=1}^{n} \alpha_i = 1 \end{array}$$

Note that the optimal value of this problem is the distance of the origin to the affine hull of the data points $x_1, \ldots, x_n$. Again we have $w = \sum_{i=1}^{n} \alpha_i x_i$, and thus w is the shortest vector from the origin to the affine hull of the data points and $\|w\|$ is the distance of origin to this affine hull.

It is worthwhile to note that the dual optimization problem boils down to solving a linear system. We derive from the saddle point condition

$$0 = \frac{\partial L(w, \rho, \alpha)}{\partial \alpha_j} = \langle w, x_j\rangle - \rho \text{ for all } j = 1, \ldots, n,$$

which together with $w = \sum_{i=1}^{n} \alpha_i x_i$ implies

$$\sum_{i=1}^{n} \alpha_i \langle x_i, x_j\rangle = \rho \text{ for all } j = 1, \ldots, n,$$

which is a linear system with unknown right-hand side ρ for the dual variables α_i. Substituting $\gamma_i = \alpha_i/\rho$ gives the linear system

$$\sum_{i=1}^{n} \gamma_i \langle x_i, x_j\rangle = 1 \text{ for all } j = 1, \ldots, n,$$

which can be solved for the γ_i (assuming the Gram matrix $\big(\langle x, x_j\rangle\big)$ has full rank). From the γ_i we can compute ρ as $\rho = \left(\sum_{i=1}^{n} \gamma_i\right)^{-1}$, all the α_i as $\alpha_i = \rho\gamma_i$ and finally w as

$$w = \sum_{i=1}^{n} \alpha_i x_i = \rho \sum_{i=1}^{n} \gamma_i x_i.$$

The linear system for the γ_i was studied extensively in computer graphics for implicit surface reconstruction [CBC+01].

4 Surface reconstruction

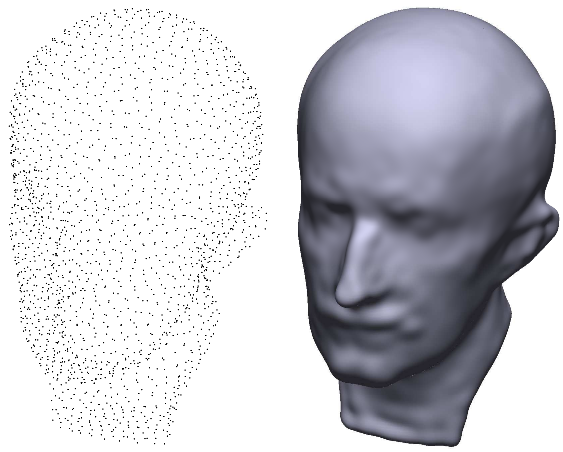

Figure 4: An example surface reconstruction (Max-Planck Head: 2022 points) using the slab SVM for a fixed (small) value of δ.

Let us briefly recapitulate how the slab SVM can be used directly for surface reconstruction [SGS04]. Given are sample points $y_1, \ldots, y_n \in \mathbb{R}^3$ from a smooth surface embedded into $\mathbb{R}^3$. These sample points are mapped into the feature space associated with the Gaussian kernel. The reconstruction is given implicitly as $f^{-1}(0)$, where $f : \mathbb{R}^3 \to \mathbb{R}$ is the kernel expansion

$$f(x) = \langle w, \phi(x)\rangle - \rho = \sum_{i=1}^{n} (\alpha_i - \beta_i) \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right) - \rho,$$

where $x \in \mathbb{R}^3$, $\phi(\cdot)$ is the feature map associated with the Gaussian kernel, and α and β are the solutions to the dual SVM. Note that ρ can also be computed from the solution to the slab SVM (or its dual). See Figure 4 for an example and also note that especially in the presence of noise one probably does not want to have an interpolating solution (as one gets it from the zero slab SVM and the related method proposed in [CBC+01]), but would like to allow small slack in terms of a small value of $\delta > 0$. Note that the slab SVM works the same for surface reconstruction in dimensions beyond three.

5 States and events

For a given value of $\delta \in (0, \infty)$ let (w, ρ) be the optimal solution of the slab SVM. We associate *states* with the data points $x_i, i = 1, \ldots, n$:

(1) lower supporting, if $\langle w, x_i\rangle = \rho$

(2) upper supporting, if $\langle w, x_i\rangle = \rho + \delta$

(3) non-supporting, if neither lower- nor upper supporting

An *event* occurs when while decreasing δ the state of any data point changes. We distinguish two types of events: a supporting data point becomes non-supporting, or a non-supporting data point becomes supporting. We call the first type of event a *lose event* and the second type of event a *gain event*.

6 The Solution Path

From the constraints $\sum_{i=1}^{n}(\alpha_i - \beta_i) = 1$ and $\alpha_i, \beta_i \geq 0$ of the dual of the slab SVM we can conclude that there exists $\alpha_i > 0$. This in turn allows us to conclude using the Karuhn-Kuhn-Tucker condition $\alpha_i\big(\langle w, x_i\rangle - \rho\big) = 0$ that for any δ there always exists a lower supporting data point. For a given δ', let x_i be a lower supporting data point. The continuous dependence of the coefficient α_i on the parameter δ implies that $\alpha_i > 0$ for some neighborhood of $U(\delta') \subset (0, \infty)$. Hence x_i is a lower supporting data point for all $\delta \in U(\delta')$. We use this insight to locally, i.e., for $\delta \in U(\delta')$, transform the slab SVM into an equivalent distance problem. Note that we have $\rho = \langle w, x_i\rangle$. Thus we can write the objective function of the slab SVM as

$$\frac{1}{2}\|w\|^2 - \rho = \frac{1}{2}\|w\|^2 - \langle w, x_i\rangle = \frac{1}{2}\|w - x_i\|^2 - \frac{1}{2}\|x_i\|^2.$$

Since $\frac{1}{2}\|x_i\|^2$ is constant, i.e., does not depend on w or ρ, we can drop it from the objective function. This gives if we set $w' = w - x_i$ and reformulate the constraints in the new variable w' accordingly the following version of the slab SVM:

$$\begin{array}{ll} \min_{w,\rho} & \frac{1}{2}\|w'\|^2 \\ \text{s.t.} & 0 \leq \langle w', x_j - x_i\rangle + \langle x_i, x_j\rangle - \|x_i\|^2 \leq \delta \\ & \text{for } j \neq i \end{array}$$

This problem asks for the shortest vector w' in the constraint polytope or equivalently the distance of the constraint polytope to the origin. Note that this distance problem is also always feasible, i.e., the constraint polytope does not become empty. To see this observe that $w' = -x_i$ is always in the polytope. The gain and lose events can be nicely illustrated for the distance problem, see Figure 5.

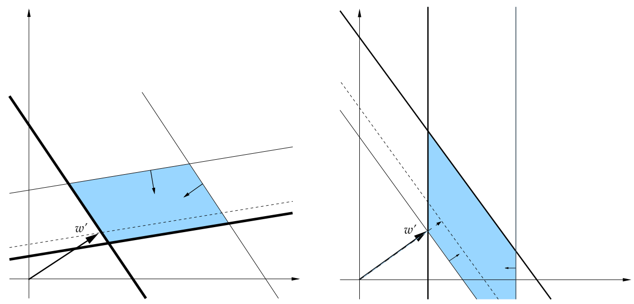

Figure 5: The lower (non-moving) constraints are shown by thick solid lines and the upper (moving) constraints are shown by thin solid lines. On the left: when the moving constraint hits w' this constraint becomes binding (gain event) and the solution is no longer stationary. On the right: once the moving constraint becomes orthogonal to w' we lose the non-moving constraint (lose event).

The formulation of the slab SVM as a distance problem allows to make some observations.

Lemma 1 *The solution to the slab SVM is unique.*

Proof. There is always a unique point in the convex constraint polytope of an equivalent distance problem that realizes the distance of the polytope to the origin. □

Lemma 2 *There exists a δ_0 such that for all $\delta > \delta_0$ the solution to the slab SVM is stationary, i.e. does not vary with δ.*

Proof. The proof is via the distance problem. Let x_i be one of the (lower) supporting data points of the open slab SVM. We use this x_i to formulate the distance problem. The solution of the distance problem at $\delta = \infty$ is finite (we can conclude this from the properties of the open slab SVM). Coming from small values of δ the constraint polytopes of the distance problem for these values of δ sweep the constraint polytope of the distance problem at $\delta = \infty$. Since the solution to the latter is finite the sweep needs to hit the point that realizes this finite distance at some finite value δ_0 of δ. That is, for all $\delta > \delta_0$ the point x_i is lower supporting for the slab SVM and we can conclude that the solution of the slab SVM can be derived from this stationary solution of the distance problem as $w = w' + x_i$ and $\rho = \langle w, x_i \rangle$. □

Lemma 3 *For all $0 < \delta < \delta_0$ the slab SVM has an upper supporting data point.*

Proof. By the proof of Lemma 2 we have that at δ_0, the slab SVM needs to have an upper supporting data point, because only the upper constraints sweep the constraint polytope of the distance problem at $\delta = \infty$. Assume there exists $0 < \delta < \delta_0$ such that at δ the slab SVM has no upper supporting data point. Let Δ be the set of all δ with this property and let $\delta' = \sup \Delta$. At δ' the slab SVM needs to have an upper supporting data point. To see this note that there exists a data point x_j that is upper supporting at $\delta + \varepsilon$ for all sufficiently small $\varepsilon > 0$. At δ' we can derive a distance problem that is equivalent to the slab SVM for some neighborhood of δ'. The data point x_j needs to be upper supporting also for this distance problems at $\delta' + \varepsilon$ for all sufficiently small $\varepsilon > 0$. The constraint hyperplane given by

$$\langle w', x_j - x_i \rangle + \langle x_i, x_j \rangle - \|x_i\|^2 = \delta' \qquad (1)$$

for the data point x_j has all the constraint hyperplanes given by

$$\langle w', x_j - x_i \rangle + \langle x_i, x_j \rangle - \|x_i\|^2 = \delta' + \varepsilon \qquad (2)$$

on one side. The latter hyperplanes all contain a point that realizes the solution of the corresponding distance problem. By the continuity of the distance problem in δ any sequence in the latter point set converges to the solution of the distance problem at δ'. Hence this solution needs to be contained in the constraint hyperplane given by Equation (1) and x_j is an upper supporting data point for both the distance- and the slab SVM problem at δ'. By our assumption there needs to exist some neighborhood U of δ' such that the distance problem does not have an upper supporting data point for all $\delta \in U \cap (0, \delta')$. This means that the family of hyperplanes given by Equation (2) sweeps with $\varepsilon \to 0$, i.e., at δ', out of the constraint polytope given by the constraints

$$\langle w', x_j - x_i \rangle + \langle x_i, x_j \rangle - \|x_i\|^2 = 0.$$

But this can only happen if the constraint polytope of the distance problem grows while sweeping the hyperplane given by Equation (2) from $\delta' + \varepsilon$ to $\delta' - \varepsilon$, which is a contradiction. □

Corollary 1 *For all $0 < \delta < \delta_0$ the solution to the slab SVM is non-stationary.*

We can conclude that the solution path of the slab SVM is piecewise linear (since w' the point that realizes the distance of the constraint polytope to the origin is a piecewise linear curve parametrized by δ).

Theorem 4 *The solution path of the slab SVM, i.e., the optimal coefficients α_i and β_i (in the dual) and w and ρ (in the primal) are piecewise linear functions of δ.*

Corollary 2 *The optimal solution w to the slab SVM is a piecewise linear path that connects the point closest to the origin on the convex hull (solution at $\delta = \infty$) of the data points with the point closest to the origin on the affine hull (solution at $\delta = 0$) of the data points.*

7 Conclusions

Theorem 4 characterizes the solution path, but does not immediately suggest an algorithm to compute it. But algorithms for parametrized convex quadratic programs (such as the slab SVM) are known, see for example [Rit81]. Another interesting open question is about the complexity of the solution path, i.e., the number of bends. We conjecture that this complexity can be exponential in the number of data points.

Acknowledgments. Joachim Giesen wants to thank Edgar Ramos and Bardia Sadri for valuable discussions on the slab SVM.

References

[CBC+01] Jonathan C. Carr, Richard K. Beatson, Jon B. Cherrie, Tim J. Mitchell, W. Richard Fright, Bruce C. McCallum, and Tim R. Evans. Reconstruction and representation of 3d objects with radial basis functions. In *SIGGRAPH*, pages 67–76, 2001.

[Ede95] Herbert Edelsbrunner. The union of balls and its dual shape. *Discrete & Computational Geometry*, 13:415–440, 1995.

[ELZ02] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28(4):511–533, 2002.

[GCPZ06] Joachim Giesen, Frédéric Cazals, Mark Pauly, and Afra Zomorodian. The conformal alpha shape filtration. *The Visual Computer*, 22(8):531–540, 2006.

[HRTZ04] Trevor Hastie, Saharon Rosset, Robert Tibshirani, and Ji Zhu. The entire regularization path for the support vector machine. *Journal of Machine Learning Research*, 5:1391–1415, 2004.

[Rit81] Klaus Ritter. On parametric linear and quadratic programming problems. In *Mathematical Programming: Proceedings of the International Congress on Mathematical Programming*, pages 307–335, 1981.

[SGS04] Bernhard Schölkopf, Joachim Giesen, and Simon Spalinger. Kernel methods for implicit surface modeling. In *NIPS*, 2004.

[SS02] Bernhard Schölkopf and Alex Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, Cambridge, MA, 2002.