JOURNAL OF DECISION MAKING AND HEALTHCARE
Volume 3 (2026), No. 1, 46–61
https://doi.org/10.69829/jdmh-026-0301-ta05
Tulipa Opera Scholarum

TOPOLOGICAL ROBUSTNESS IN FUZZY SUPPORT VECTOR MACHINES: A KIKKAWA-SUZUKI AND ULAM-HYERS STABILITY PERSPECTIVE

RAMAN CHAUDHARY*

Department of Mathematical Sciences, Indian Institute of Technology (BHU) Varanasi, Varanasi 221005, Uttar Pradesh, India

ABSTRACT. We analyse the robustness of Fuzzy Linear Support Vector Machines (FSVMs) through the lens of nonlinear functional analysis. The central idea is to recast support vector selection in the FSVM dual as a multi-valued fixed point problem, posed on the compact feasible region under the Euclidean metric. Provided the kernel matrix Q is strictly positive definite, the projected gradient selection mapping acts as a multi-valued contraction of Nadler type [11], and consequently satisfies the Kikkawa-Suzuki condition [9]. Fixed points therefore exist. A retraction-displacement argument produces generalized Ulam-Hyers stability: for ϵ-approximate solver outputs of the same FSVM instance, the displacement to an exact fixed point is controlled by a continuous increasing function of ϵ. Separately, the Petruşel-Rus data dependence framework [12] bounds the Hausdorff distance between two fixed point sets, proportionally to the perturbation magnitude ϵ measured in ℓ^{∞}. Notably, this bound demonstrates that the fixed point sets of two distinct FSVM instances remain Hausdorff-close, and the estimate is entirely deterministic. All three results are topological certificates, fundamentally distinct from probabilistic generalisation bounds. Our analysis is confined to linear kernels throughout. We include small-scale numerical experiments measuring support vector set stability via Jaccard similarity, a discrete proxy for the continuous Hausdorff bound; results are broadly consistent with the theoretical predictions, though the metric gap means the comparison is heuristic rather than direct.

Keywords. Fuzzy Support Vector Machines, Ulam-Hyers Stability, Kikkawa-Suzuki Contractions, Multi-valued Operators, Retraction-Displacement, Data Dependence.

1. INTRODUCTION

An SVM constructs a hyperplane that separates two classes by maximising the geometric margin. Vapnik's structural risk minimisation principle [21] supplies the theoretical foundation; Boser, Guyon, and Vapnik [3] translated this into a working algorithm, and Platt's SMO decomposition [13] made large-scale training feasible. With Cortes and Vapnik's soft-margin formulation [5], slack variables allow a controlled number of misclassifications, extending SVMs to problems like text categorisation [8], novelty detection [18], and medical diagnostics, among others.

A well-known limitation persists: the regularisation parameter C penalises every training point identically, giving outliers and mislabelled samples an outsized effect on the decision boundary. Lin and Wang [10] addressed this by assigning each sample a fuzzy membership $s_i \in [\sigma, 1]$ that scales its penalty contribution. The resulting Fuzzy SVM (FSVM) has since seen several refinements; see [23, 24]. While FSVMs are widely used in noisy environments, rigorous mathematical guarantees regarding support vector stability remain scarce. Available results are probabilistic in nature: they bound expected loss but have nothing to say about any particular trained instance. The question we pursue here is whether an FSVM solver can be analysed through the lens of topological fixed point theory.

*Corresponding author.
E-mail address: raman.chaudhary.mat22@itbhu.ac.in (R. Chaudhary)
Accepted: March 18, 2026.

This question came up naturally during a reading seminar on Ulam-Hyers stability, and the fixed point viewpoint turned out to be the right one. Our first attempts tried a direct sensitivity analysis of the KKT system in the style of Bertsekas [1], but the non-smoothness of the active set prevents the implicit function theorem from applying. (Bonnans and Shapiro [2] handle the smooth case in detail; the obstacle here is the non-smoothness itself.)

What does work is recasting support vector selection as a multi-valued operator fixed point problem on the compact feasible dual space. Under this formulation, the operator turns out to be a Nadler-type contraction [11], and therefore satisfies the Kikkawa-Suzuki condition [9], which yields fixed point existence. A caveat: this requires Q to be strictly positive definite, a condition that fails when the number of training samples N exceeds the feature dimension n under a linear kernel (since $Q_{ij} = y_i y_j x_i^T x_j$ has rank at most n, producing at least $N - n$ zero eigenvalues, so Q is only positive semi-definite); see Remark 3.7.

Two stability results follow from this setup. First, a retraction-displacement argument gives generalized Ulam-Hyers stability (cf. [20, 6, 14]; see also [17, 4]): if $\hat{\alpha}$ is an ϵ-approximate solver output, then $d(\hat{\alpha}, \alpha^*) \leq \Psi(\epsilon)$ for some exact fixed point α^*. Second, the Petruşel-Rus data dependence framework [12] whose formulation does not require two operators to share a common domain yields $H(\mathrm{Fix}(T_s), \mathrm{Fix}(T_{\tilde{s}})) \leq \Phi(\epsilon)$ under ϵ-magnitude membership perturbations. The constant κ appearing in this parametric bound scales with $\sqrt{N}$ and depends on the geometry of the constraint hyperplane. It is almost certainly far from sharp for large N, but tightening it remains an open problem. Our experiments are confined to small datasets; the theory itself carries no such restriction.

Section 2 fixes notation and recalls the relevant fixed point background. Section 3 presents the core results. Experiments appear in Section 4; conclusions and open problems in Section 5.

2. Mathematical Preliminaries

Throughout, (X, d) denotes a metric space. We write $\mathcal{P}(X)$ for the collection of non-empty subsets of X, $\mathrm{CL}(X)$ for non-empty closed subsets, and $\mathrm{CB}(X)$ for non-empty closed and bounded subsets. Only those definitions that appear later in the proofs are recorded here; standard facts about metric space topology are used without further comment.

2.1. **Standard support vector machine formulation.** Given a labelled training set $\{(x_i, y_i)\}_{i=1}^N$ with $x_i \in \mathbb{R}^n$ and $y_i \in \{-1, +1\}$, the hard-margin SVM looks for the hyperplane that maximises the geometric margin $\gamma = 2/\|w\|_2$. Concretely, it solves:

$$\min_{w,b} \ \tfrac{1}{2}\|w\|_2^2 \quad \text{subject to} \quad y_i(w^T x_i + b) \geq 1, \quad i = 1, \ldots, N \tag{2.1}$$

In practice, data is rarely linearly separable without error. Slack variables $\xi_i \geq 0$ relax the constraints, yielding the soft-margin formulation:

$$\min_{w,b,\xi} \ \tfrac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N} \xi_i \quad \text{subject to} \quad y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \tag{2.2}$$

where $C > 0$ governs the trade-off between a wide margin and a small total violation.

Writing out the KKT (Karush-Kuhn-Tucker) stationarity conditions for this problem gives $w = \sum \alpha_i y_i x_i$, $\sum \alpha_i y_i = 0$, and $C - \alpha_i - \beta_i = 0$, together with complementary slackness $\alpha_i[y_i(w^T x_i + b) - 1 + \xi_i] = 0$ and $\beta_i \xi_i = 0$. Eliminating primal variables produces the Wolfe dual, which is the form we actually work with:

$$\max_{\alpha} \ W(\alpha) = \sum_{i=1}^{N} \alpha_i - \tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j \quad \text{s.t.} \quad \sum_{i=1}^{N} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \tag{2.3}$$

2.2. **Fuzzy membership and the fuzzy linear SVM.** Lin and Wang's FSVM [10] attaches a fuzzy membership $s_i \in [\sigma, 1]$ to every training pair (x_i, y_i). The lower bound $\sigma > 0$ is there to keep the box constraints in the dual non-degenerate. Intuitively, a point far from its own class centroid is more likely to be noisy, so it gets a smaller s_i and contributes less to the loss. A standard distance-based rule for computing memberships is

$$s_i = 1 - \frac{\|x_i - \bar{x}_{y_i}\|}{\max_{k:y_k=y_i} \|x_k - \bar{x}_{y_i}\| + \delta} \tag{2.4}$$

where $\bar{x}_{y_i}$ denotes the centroid of the class to which x_i belongs, and $\delta > 0$ is a small constant that prevents division by zero. Replacing the uniform penalty $C\xi_i$ with the weighted penalty $Cs_i\xi_i$ gives the FSVM primal:

$$\min_{w,b,\xi} \ \tfrac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N} s_i\xi_i \quad \text{s.t.} \quad y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i.$$

The KKT conditions for this Lagrangian produce the stationarity relations

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i, \quad \sum_{i=1}^{N} \alpha_i y_i = 0, \quad Cs_i - \alpha_i - \beta_i = 0$$

along with complementary slackness $\alpha_i[y_i(w^T x_i + b) - 1 + \xi_i] = 0$ and $\beta_i\xi_i = 0$ for each i. Because $\beta_i \geq 0$, we must have $\alpha_i \leq Cs_i$, and the dual takes the form

$$\max_{\alpha} \ W(\alpha) = \sum_{i=1}^{N} \alpha_i - \tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j y_i y_j x_i^T x_j \quad \text{s.t.} \quad \sum_{i=1}^{N} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq s_i C \tag{2.5}$$

Support vectors are exactly the indices with $\alpha_i > 0$. The membership s_i caps each dual variable at s_iC rather than C, and that is the entire point of the construction: noisy points are allowed less influence on the separating hyperplane.

2.3. **Multi-valued operators and the Hausdorff metric.** A mapping $T : X \to \mathcal{P}(X)$ is a multi-valued operator; when $T(x) \in \mathrm{CL}(X)$ for every x, we say T is closed-valued. Multi-valued maps naturally formulate this problem because an FSVM solver with a nonzero tolerance yields a set of near-optimal vectors.

To measure how far apart two such sets of solutions are, we need a distance between sets, not just between points. That is what the Hausdorff metric provides.

Definition 2.1 (Hausdorff Metric)**.** Let (X, d) be a metric space and $A, B \in \mathrm{CL}(X)$. The Hausdorff metric is

$$H(A, B) = \max\Big\{\sup_{a\in A} D(a, B),\ \sup_{b\in B} D(b, A)\Big\}$$

where $D(x, A) = \inf_{a\in A} d(x, a)$. Here $a \in A$ and $b \in B$, so $D(a, B)$ denotes the distance from the point a to the set B, consistent with the infimum definition above.

One useful fact: $H(A, B) = 0$ if and only if $A = B$, provided both sets are closed. We use this implicitly in the proof of Lemma 3.5. In $\mathbb{R}^N$ with the Euclidean metric, H is always finite on $\mathrm{CB}(X)$ and defines a metric on that collection.

Informally, a closed graph means that the mapping does not "leak" limit points: if inputs converge and the corresponding outputs converge too, the limit output still belongs to the image of the limit input.

Definition 2.2 (Closed Graph)**.** A multi-valued map $T : X \to \mathrm{CL}(X)$ has a closed graph if whenever $x_n \to x$ and $y_n \to y$ with $y_n \in T(x_n)$ for all n, then $y \in T(x)$.

Upper semicontinuity with compact values implies a closed graph; lower semicontinuity does not in general; a map can be lsc while its graph remains open. For a careful treatment of these distinctions see [17].

Nadler's theorem below says, roughly, that if a set-valued map shrinks distances between image sets (measured by H) by a uniform factor less than one, then it must have a fixed point. It is the set-valued version of the Banach contraction principle.

Theorem 2.3 (Nadler's Fixed Point Theorem [11])**.** *If* (X, d) *is complete and* $T : X \to \mathrm{CB}(X)$ *satisfies* $H(T(x), T(y)) \le L\, d(x, y)$ *for all* x, y *and some* $L \in [0, 1)$*, then* T *has a fixed point.*

This is our starting point. The contraction argument in Section 3 relies on exactly this result.

2.4. **Ulam-Hyers stability: original and generalized forms.** We work in a complete metric space (X, d). Fix an operator $f : X \to X$, and suppose $y^* \in X$ is an ϵ-approximate fixed point satisfying $d(y^*, f(y^*)) \le \epsilon$

Definition 2.4 (Ulam-Hyers Stability)**.** The equation $x = f(x)$ is UH stable if there exists $c > 0$ such that for every ϵ-approximate y^*, there exists $x^* \in \mathrm{Fix}(f)$ with

$$d(y^*, x^*) \le c\,\epsilon$$

Demanding a linear bound in ϵ is quite restrictive in practice. A more flexible version replaces the linear function $c\epsilon$ with a nonlinear comparison function one that is strictly increasing and vanishes at zero:

Definition 2.5 (Generalized Ulam-Hyers Stability)**.** The equation is GUH stable if there exists continuous, strictly increasing $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ with $\psi(0) = 0$ such that

$$d(y^*, x^*) \le \psi(\epsilon)$$

Why insist on ψ being strictly increasing with $\psi(0) = 0$? Because without both conditions, the connection between small perturbations and small displacements breaks down entirely. Drop either one and the stability guarantee is gone. In most concrete cases, including ours, ψ ends up being linear anyway, collapsing back to ordinary UH stability. Where the generalized form actually earns its keep is when the contraction constant varies with the iterate; operators with state-dependent Lipschitz bounds fall into this category. For broader extensions, including the Ulam-Hyers-Rassias variant, see [14, 17].

2.5. **Kikkawa-Suzuki contraction conditions.** A global Lipschitz condition on the multi-valued map would be simpler, but operators from constrained optimisation rarely admit one: active sets jump discontinuously, and any global contraction constant ends up too close to 1 to be useful. Kikkawa and Suzuki [9] drop the global requirement and ask for contraction only when the iterate is far enough from being a fixed point. Near $\mathrm{Fix}(T)$ the operator contracts; away from it, no condition is imposed.
For points already in close proximity to the fixed point set, the Kikkawa-Suzuki condition relaxes the contraction requirement.

Definition 2.6 (Kikkawa-Suzuki Threshold Function)**.** Define $\eta : [0, 1) \to (1/2, 1]$ by

$$\eta(c) = \begin{cases} 1 & \text{if } 0 \le c \le \frac{1}{2} \\ \dfrac{1}{1+c} & \text{if } \frac{1}{2} \le c < 1 \end{cases}$$

The split at $c = 1/2$ is not arbitrary. When $c \le 1/2$ the contraction is strong enough that every pair of points passes the test automatically, so $\eta = 1$ imposes nothing extra. Past $1/2$, contraction alone cannot control all pairs; η drops below 1 and restricts the test to iterates that are already close to being fixed.

Definition 2.7 (Kikkawa-Suzuki Multi-Valued Contraction [9])**.** Let (X, d) be a complete metric space. A map $T : X \to \mathrm{CL}(X)$ is a Kikkawa-Suzuki type contraction if there exists $c \in [0, 1)$ such that for all $x, y \in X$:

$$\eta(c)\, D(x, T(x)) \le d(x, y) \implies H(T(x), T(y)) \le c \max\{d(x, y),\ D(x, T(x)),\ D(y, T(y))\}$$

In plain terms: contraction is demanded only when $d(x, y)$ is large relative to $D(x, T(x))$. When x is already nearly fixed the hypothesis fails and the definition asks nothing.

3. Core Theoretical Contribution

3.1. **Support vector selection as a fixed point inclusion.** Applying an iterative optimization algorithm (e.g., SMO [13] or projected gradient ascent) to the FSVM dual problem (2.5) generates a sequence of dual vectors $\alpha \in \mathbb{R}^N$ that stay trapped inside the feasible region

$$X_s = \left\{ \alpha \in \mathbb{R}^N \;\middle|\; \sum_{i=1}^{N} \alpha_i y_i = 0, \quad 0 \le \alpha_i \le s_i C, \quad i = 1, \dots, N \right\} \tag{3.1}$$

Let's clean up the notation by defining $f(\alpha) = -W(\alpha) = \frac{1}{2}\alpha^T Q \alpha - \mathbf{1}^T \alpha$ to represent our convex objective. This gives us a crisp gradient $\nabla f(\alpha) = Q\alpha - \mathbf{1}$, where $Q_{ij} = y_i y_j x_i^T x_j$. A quick subtlety here: if Q is degenerate (only positive semi-definite) instead of strictly positive definite, or if we force the solver to stop at some ϵ-tolerance, our set of optimal solutions $X^* \subseteq X_s$ will remain closed and convex, but it could definitely contain multiple equivalent points.

Definition 3.1 (Support Vector Selection Operator)**.** For a fixed step size $\tau \in (0, 2/\lambda_{\max}(Q))$, where $\lambda_{\max}(Q)$ is the largest eigenvalue of Q, and a solver tolerance $\epsilon_{\text{tol}} > 0$, define the multi-valued support vector selection operator $T : X_s \to \mathrm{CL}(X_s)$ by

$$T(\alpha) = \left\{\beta \in X_s : \|\beta - \mathrm{Proj}_{X_s}(\alpha - \tau \nabla f(\alpha))\|_2 \le \epsilon_{\text{tol}}\right\}$$

where Proj_{X_s} denotes the metric (Euclidean) projection onto the closed convex set X_s. For each α, the set $T(\alpha)$ is the ϵ_{tol}-neighbourhood of the projected gradient step, intersected with X_s. When $\epsilon_{\text{tol}} = 0$, the operator reduces to the singleton projected gradient map $T_\tau(\alpha) = \mathrm{Proj}_{X_s}(\alpha - \tau \nabla f(\alpha))$.

Definition 3.1 is structured so that $T(\alpha)$ contains not just the exact projected gradient step, but anything landing within the solver's allowed tolerance. Geometrically, the multi-valuedness of T captures the tolerance band that any practical solver introduces at termination.

Proposition 3.2 (Fixed Point Equivalence)**.** *Let T be as in Definition 3.1 with step size $\tau \in (0, 2/\lambda_{\max}(Q))$ and tolerance $\epsilon_{\text{tol}} \ge 0$. Then:*

(i) *If α^* is a KKT point of the FSVM dual (2.5), then $T_\tau(\alpha^*) = \alpha^*$, so $D(\alpha^*, T(\alpha^*)) = 0 \le \epsilon_{\text{tol}}$ hence $\alpha^* \in T(\alpha^*)$.*

(ii) *If $\alpha^* \in T(\alpha^*)$, then $\|\alpha^* - T_\tau(\alpha^*)\|_2 \le \epsilon_{\text{tol}}$, i.e. α^* is an ϵ_{tol}-approximate KKT point.*

Proof. (i) If we are sitting exactly at a KKT point, then α^* must be optimal for our convex setup over X_s. From here, the classic stationarity condition $T_\tau(\alpha^*) = \mathrm{Proj}_{X_s}(\alpha^* - \tau\nabla f(\alpha^*)) = \alpha^*$ is known to map 1:1 with first-order optimality, no matter what valid $\tau > 0$ we pick (check [1], Proposition 2.3.1 for the mechanics). Because of this, it's clear that $D(\alpha^*, T(\alpha^*)) \le \|\alpha^* - T_\tau(\alpha^*)\|_2 = 0$.
(ii) This direction falls out immediately from Definition 3.1. □

In short, the inclusion $\alpha^* \in T(\alpha^*)$ characterises the final support vector selection exactly. The support vectors are the indices i with $\alpha_i^* > 0$.

3.2. **Foundational lemmas.** Before we can tackle the main theorems, three preliminary lemmas are needed. Lemmas 3.3 and 3.4 are straightforward; Lemma 3.5 requires genuine work.

We first verify that the feasible dual space carries the topological properties we rely on: completeness and compactness.

Lemma 3.3 (Completeness and Compactness of the FSVM Feasible Space)**.** *The feasible dual space* X_s *defined in* (3.1)*, equipped with* $d(\alpha, \beta) = \|\alpha - \beta\|_2$*, is a compact and complete metric space.*

Proof. X_s is the intersection of the closed equality hyperplane $\{\alpha : \sum \alpha_i y_i = 0\}$ with the box $\prod_{i=1}^{N}[0, s_i C]$ in $\mathbb{R}^N$. Both sets are closed, and the box is bounded ($s_i \leq 1$, $C < \infty$). By Heine-Borel, X_s is compact. Every compact metric space is complete. □

Next, we show that the selection map has a closed graph.

Lemma 3.4 (Closed-Graph Property of the Selection Map)**.** *The multi-valued mapping* $T : X_s \to \mathrm{CL}(X_s)$ *of Definition 3.1 possesses a closed graph.*

Proof. Consider sequences where $\alpha_n \to \alpha$ and $\beta_n \to \beta$, assuming $\beta_n \in T(\alpha_n)$ holds for all n. By definition, $\|\beta_n - \mathrm{Proj}_{X_s}(\alpha_n - \tau\nabla f(\alpha_n))\|_2 \leq \epsilon_{\text{tol}}$ at every index n. Because ∇f is affine (and thus continuous) and Proj_{X_s} is non-expansive (also continuous), we can pass to the limit: $\|\beta - \mathrm{Proj}_{X_s}(\alpha - \tau\nabla f(\alpha))\|_2 \leq \epsilon_{\text{tol}}$, confirming $\beta \in T(\alpha)$. The compactness of X_s (Lemma 3.3) ensures that each image $T(\alpha)$ is itself compact. □

The third lemma is central. It constructs a retraction into the fixed point set and extracts a displacement bound.

Lemma 3.5 (Retraction-Displacement of Optimisation Steps)**.** *Suppose that* $T : X_s \to \mathrm{CL}(X_s)$ *(Definition 3.1) satisfies a multi-valued contraction condition with constant* $c \in [0, 1)$*, i.e.,* $H(T(\alpha), T(\beta)) \leq c\, d(\alpha, \beta)$ *for all* $\alpha, \beta \in X_s$ *(this is established in Theorem 3.6 below). Then* T *admits a set retraction* $r : X_s \to \mathrm{Fix}(T)$*, and there exists a continuous, increasing function* $\Psi : \mathbb{R}^+ \to \mathbb{R}^+$ *with* $\Psi(0) = 0$ *such that*

$$d(\alpha, r(\alpha)) \leq \Psi(D(\alpha, T(\alpha))) \quad \forall \alpha \in X_s. \tag{3.2}$$

Proof. The retraction construction does all the work. The displacement bound then follows directly.

First, we analyse $\mathrm{Fix}(T)$. Compactness of X_s (Lemma 3.3) and continuity of f give, via Weierstrass, that $X^* = \arg\min_{\alpha \in X_s} f(\alpha)$ is non-empty, closed, and convex. Proposition 3.2(i) yields $X^* \subseteq \mathrm{Fix}(T)$. When $\epsilon_{\text{tol}} = 0$, first-order optimality forces $\mathrm{Fix}(T) = X^*$, and we would be done.

When $\epsilon_{\text{tol}} > 0$, the picture changes. The fixed point set expands to $\mathrm{Fix}(T) = \{\alpha \in X_s : g(\alpha) \leq \epsilon_{\text{tol}}\}$ where $g(\alpha) = \|\alpha - T_\tau(\alpha)\|_2$. We know $T_\tau = \mathrm{Proj}_{X_s} \circ G$ via $G(\alpha) = (I - \tau Q)\alpha + \tau\mathbf{1}$. Both projection and affine map are continuous, so T_τ is continuous on X_s. Hence g is continuous and $\mathrm{Fix}(T) = g^{-1}([0, \epsilon_{\text{tol}}])$ is a closed sublevel set. Closed inside compact gives compact, hence complete. It is non-empty because X^* sits inside it.

While nearest-point projection onto $\mathrm{Fix}(T)$ is conceptually straightforward, the potential non convexity of $\mathrm{Fix}(T)$ generally sacrifices mapping continuity. We take an operational route instead, building r by iterating T.

One caveat: because $T(\alpha)$ is always compact, we can select nearest points at every step. But the retraction r we build may lack continuity when $\mathrm{Fix}(T)$ is non-convex. This causes no difficulty. Theorem 3.9 only requires that some $\alpha^* \in \mathrm{Fix}(T)$ satisfying the displacement bound exists; it never asks r to be continuous.

Pick your starting position $\alpha^{(0)} = \alpha \in X_s$. Because $T(\alpha)$ is a compact set in $\mathbb{R}^N$ for every α, the infimum in the distance metric is attained, allowing us to select a point $\alpha^{(1)} \in T(\alpha^{(0)})$ such that

$$d(\alpha^{(0)}, \alpha^{(1)}) = D(\alpha^{(0)}, T(\alpha^{(0)}))$$

Once we are standing at $\alpha^{(1)}$, the core contraction mechanic immediately activates: $D(\alpha^{(1)}, T(\alpha^{(1)})) \le H(T(\alpha^{(0)}), T(\alpha^{(1)})) \le c\, d(\alpha^{(0)}, \alpha^{(1)})$. We leverage this to latch onto a new nearest point $\alpha^{(2)} \in T(\alpha^{(1)})$, and we keep unrolling this process over and over. With every single step forward, the spatial jump distance between iterates shrinks relentlessly by a factor of c. The governing bound for any step k dictates

$$d(\alpha^{(k)}, \alpha^{(k+1)}) \le c^k\, d(\alpha^{(0)}, \alpha^{(1)})$$

We stack these jumps into a telescoping series. For any indices $m > k \ge 0$, we find:

$$d(\alpha^{(k)}, \alpha^{(m)}) \le \sum_{j=k}^{m-1} d(\alpha^{(j)}, \alpha^{(j+1)}) \le \sum_{j=k}^{m-1} c^j\, d(\alpha^{(0)}, \alpha^{(1)}) \le \frac{c^k}{1-c}\, d(\alpha^{(0)}, \alpha^{(1)})$$

This hard limit forces the sequence $\{\alpha^{(k)}\}$ to act globally Cauchy within (X_s, d). Relying on the completeness we proved earlier (Lemma 3.3), this sequence has no choice but to converge down to a limit $\alpha^{(k)} \to \alpha^*$ for some ending vector $\alpha^* \in X_s$. If we plug in $k = 0$, we capture the total path distance:

$$d(\alpha^{(0)}, \alpha^*) \le \frac{d(\alpha^{(0)}, \alpha^{(1)})}{1-c}$$

The first hop was engineered to satisfy $d(\alpha^{(0)}, \alpha^{(1)}) = D(\alpha^{(0)}, T(\alpha^{(0)}))$, so we can rewrite the path limit as

$$d(\alpha, \alpha^*) \le \frac{D(\alpha, T(\alpha))}{1-c} \tag{3.3}$$

We must verify that α^* is a fixed point. Our setup gives us $\alpha^{(k+1)} \in T(\alpha^{(k)})$ alongside $D(\alpha^{(k+1)}, T(\alpha^{(k+1)})) \le c^{k+1} d(\alpha^{(0)}, \alpha^{(1)}) \to 0$. This forces $D(\alpha^*, T(\alpha^*))$ to zero. Because $T(\alpha^*)$ is a closed set (via Lemma 3.4), we conclude $\alpha^* \in T(\alpha^*)$.

We define $r(\alpha) := \alpha^* = \lim_{k\to\infty} \alpha^{(k)}$, where the sequence is built using the measurable selector σ above. Because σ is fixed deterministically at each step, r is a well-defined (though not necessarily continuous) map from X_s into $\mathrm{Fix}(T)$. If we construct our function $\Psi(t) = t/(1-c)$ and bridge it with (3.3) and the newly confirmed fixed point nature, we cleanly reproduce the target bound (3.2):

$$d(\alpha, r(\alpha)) = d(\alpha, \alpha^*) \le \frac{1}{1-c}\, D(\alpha, T(\alpha)) \tag{3.4}$$

The mapped point $r(\alpha)$ lands inside $\mathrm{Fix}(T)$. This set contains the exact minimisers X^*, but may extend further when $\epsilon_{\text{tol}} > 0$. Because we built r by iterating T rather than projecting onto $\mathrm{Fix}(T)$, we sidestepped the need for the fixed point set to be convex. We never assumed that $\mathrm{Proj}_{\mathrm{Fix}(T)}$ acts single-valued. Theorem 3.9 is structured around $\mathrm{Fix}(T)$ for precisely this reason. □

3.3. **Contraction property and fixed point existence.**

Theorem 3.6 (Multi-Valued Contraction and Fixed Point Existence for the Support Vector Selection Map)**.** *Take X_s to be the FSVM feasible dual space (3.1) equipped with the Euclidean metric, and let $\mathcal{H} = \{\alpha \in \mathbb{R}^N : y^\top \alpha = 0\}$ denote the equality constraint hyperplane. Assume Q is strictly positive definite on $\mathbb{R}^N$. Let $T : X_s \to \mathrm{CL}(X_s)$ be the support vector selection operator (Definition 3.1) with a step size $\tau \in (0, 2/\lambda_{\max}(Q))$. Then we can find a constant $c \in [0, 1)$ satisfying*

$$H(T(\alpha), T(\beta)) \le c\, d(\alpha, \beta) \quad \forall \alpha, \beta \in X_s$$

This ensures T is a Nadler contraction (Theorem 2.3). Because any Nadler contraction automatically obeys the Kikkawa-Suzuki condition (Definition 2.7), T must have at least one fixed point $\alpha^* \in T(\alpha^*)$.

Proof. We first establish the contraction for the single-valued projected gradient map, then lift it to the multi-valued ϵ_{tol}-neighbourhoods.

The basic projected gradient map with step size $\tau > 0$ is

$$T_\tau(\alpha) = \text{Proj}_{X_s}\big(\alpha - \tau\nabla f(\alpha)\big) = \text{Proj}_{X_s}\big(\alpha - \tau(Q\alpha - \mathbf{1})\big)$$

Let's call the affine map before the projection $G(\alpha) = (I - \tau Q)\alpha + \tau\mathbf{1}$. Suppose $0 < \lambda_{\min}(Q) \le \cdots \le \lambda_{\max}(Q)$ are the eigenvalues for the strictly positive definite matrix $Q \in \mathbb{R}^{N\times N}$. If we choose our step size inside the stable window $\tau \in (0, 2/\lambda_{\max}(Q))$, we get:

$$c_0 = \max_i |1 - \tau\lambda_i(Q)| = \max\big\{|1 - \tau\lambda_{\min}(Q)|,\ |1 - \tau\lambda_{\max}(Q)|\big\} < 1.$$

The strict inequality $c_0 < 1$ follows directly from the positive definiteness of Q.

What happens when Q is only positive semi-definite? This actually shows up often with linear kernels when you have more features than samples ($N > n$). The standard workaround is to hit the objective with a slight regularisation bump, replacing Q with $Q_\lambda = Q + \lambda I$ for some fixed tiny scalar $\lambda > 0$. Because Q_λ is strictly positive definite regardless, the eigenvalues shift strictly away from zero: $\lambda_i(Q_\lambda) = \lambda_i(Q) + \lambda > 0$. That immediately forces the contraction constant to be $c_0 = \max_i |1 - \tau\lambda_i(Q_\lambda)| < 1$ as long as we keep $\tau \in (0, 2/\lambda_{\max}(Q_\lambda))$. Adding this standard Tikhonov regularizer $\frac{\lambda}{2}\|\alpha\|^2$ to the dual objective changes neither the ambient Euclidean geometry nor the shape of the feasible set X_s. The Euclidean projection Proj_{X_s} stays exactly the same too. Because of this, everything in the proof works perfectly under regularisation. Note, however, that the stability constants degrade as $\lambda \to 0$: the contraction constant $c_0(\lambda) \to 1$ and the Ulam-Hyers bound $\Psi(\epsilon) = \epsilon/(1 - c_0) \to \infty$. The regularised problem converges to the original unregularised optimum, but the stability guarantee is only valid for a fixed $\lambda > 0$.

Using c_0, the affine mapping G contracts globally: for any $\alpha, \beta \in \mathbb{R}^N$,

$$\|G(\alpha) - G(\beta)\|_2 = \|(I - \tau Q)(\alpha - \beta)\|_2 \le c_0\|\alpha - \beta\|_2$$

One subtlety here is that G does not respect the constraint hyperplane $\mathcal{H}$ because multiplying by Q tends to throw vectors out of $\mathcal{H}$. This is fine, because the projection layer Proj_{X_s} puts everything back. So, composing the two layers into $T_\tau = \text{Proj}_{X_s} \circ G$ preserves the contraction entirely:

$$\|T_\tau(\alpha) - T_\tau(\beta)\|_2 \le c_0\,\|\alpha - \beta\|_2$$

Now we lift this to the multi-valued setting. By definition, our operator $T(\alpha) = \{\beta \in X_s : \|\beta - T_\tau(\alpha)\|_2 \le \epsilon_{\text{tol}}\}$ is essentially a solid ϵ_{tol}-ball centered exactly on $T_\tau(\alpha)$, cut down to fit inside X_s. To apply Lemma 3.5, the contraction property of the single-valued map T_τ must be extended to establish a Hausdorff bound on the multi-valued tolerance neighbourhoods $T(\alpha)$.

Take an arbitrary element $a \in T(\alpha)$. This guarantees $a \in X_s$ and $\|a - T_\tau(\alpha)\| \le \epsilon_{\text{tol}}$. Construct a new point $b' = a + T_\tau(\beta) - T_\tau(\alpha)$. Checking the dot product: $y^T b' = y^T a + y^T T_\tau(\beta) - y^T T_\tau(\alpha) = 0 + 0 - 0 = 0$, confirming b' sits on $\mathcal{H}$. Furthermore:

$$\|b' - T_\tau(\beta)\| = \|a - T_\tau(\alpha)\| \le \epsilon_{\text{tol}}$$

Next, project this point by setting $b = \text{Proj}_{X_s}(b')$. We know $T_\tau(\beta) \in X_s$, and taking advantage of the projection map's non-expansiveness again:

$$\|b - T_\tau(\beta)\| = \|\text{Proj}_{X_s}(b') - \text{Proj}_{X_s}(T_\tau(\beta))\| \le \|b' - T_\tau(\beta)\| \le \epsilon_{\text{tol}}$$

which is exactly what we need to verify $b \in T(\beta)$. Also, knowing that $a \in X_s$ means it is a fixed point for projection ($\text{Proj}_{X_s}(a) = a$). From this we can measure the distance:

$$\|a - b\| = \|\text{Proj}_{X_s}(a) - \text{Proj}_{X_s}(b')\| \le \|a - b'\| = \|T_\tau(\alpha) - T_\tau(\beta)\| \le c_0\, d(\alpha, \beta)$$

Since our starting point $a \in T(\alpha)$ could have been anything, we immediately get the upper sweep bound $\sup_{a \in T(\alpha)} D(a, T(\beta)) \le c_0\, d(\alpha, \beta)$. The problem is perfectly symmetric in α and β, so the reverse bound applies too. We merge them to get the Hausdorff bound:

$$H(T(\alpha), T(\beta)) \le c_0\, d(\alpha, \beta) \tag{3.5}$$

Just assign $c = c_0$. This global bound (3.5) fits the strict definition of a Nadler contraction. Any Nadler contraction trivially fulfills the Kikkawa-Suzuki condition because whenever the threshold $\eta(c)\, D(\alpha, T(\alpha)) \le d(\alpha, \beta)$ is passed, the standard contraction inequality simply takes over anyway:

$$H(T(\alpha), T(\beta)) \le c\, d(\alpha, \beta) \le c \max\bigl\{d(\alpha, \beta),\ D(\alpha, T(\alpha)),\ D(\beta, T(\beta))\bigr\}$$

We now conclude. Since X_s is complete (Lemma 3.3), Nadler's fixed-point theorem [11] guarantees at least one $\alpha^* \in T(\alpha^*)$. □

Remark 3.7 (The Positive Semi-Definite Case). The fact that $Q_\lambda = Q + \lambda I$ stays strictly positive definite holds for any $\lambda > 0$. This handles the linear kernel case with $N > n$, where $\mathrm{rank}(Q) \le n < N$. The regularisation parameter λ can be taken arbitrarily small; c_0, Ψ, and Φ all vary continuously in λ. The stability guarantees of Theorems 3.9 and 3.11 remain intact for every fixed $\lambda > 0$. However, the stability constants are not uniform in λ: as $\lambda \to 0$, the contraction constant $c_0(\lambda) \to 1$ and the Ulam-Hyers constant $c_{\mathrm{UH}} = 1/(1 - c_0) \to \infty$. In other words, the contraction-based stability theory does not extend to the unregularised semi-definite case at $\lambda = 0$, where the operator becomes non-expansive rather than contractive. Practitioners using linear kernels with $N > n$ should therefore treat $\lambda > 0$ as a necessary condition for the stability guarantees, not merely a computational convenience. For a universal kernel such as the Gaussian RBF, Q is already strictly positive definite, and no regularisation is needed.

Remark 3.8 (On the Kikkawa-Suzuki Framework). Since we just proved our operator obeys the blanket Nadler contraction over the whole space (3.5), noting that it satisfies the weaker Kikkawa-Suzuki condition might look redundant. The K-S framework matters conceptually because it does not break when global Lipschitz continuity fails, a property characteristic of more complex solvers. While our operator T is explicitly defined as a projected gradient step (Definition 3.1), practical implementations like SMO [13] update variables in blocks, meaning the active constraint set $\{i : \alpha_i = 0\} \cup \{i : \alpha_i = s_i C\}$ changes discontinuously at the boundary. Although T does not algebraically model an SMO step, the K-S framework illustrates how such discontinuous solvers can still lock in fixed points by only demanding contraction when the iterate is sufficiently far from $\mathrm{Fix}(T)$. Geometrically, the threshold $\eta(c)$ partitions the space into an outer zone where no contraction is imposed and an inner convergence funnel near the fixed point set. We suspect this two-phase picture describes real solvers, but a rigorous proof for a specific SMO implementation remains out of scope.

Whether the algebraic form of $\eta(c)$ has a deeper interpretation remains unclear. Does it correlate with the step size τ, or mirror the KKT gap at a given iteration? This remains open.

3.4. **Ulam-Hyers stability of the FSVM solver.** Recall from the start of the proof for Lemma 3.5 that the set $\mathrm{Fix}(T)$ is non-empty, closed, and compact. We also saw that the retraction $r : X_s \to \mathrm{Fix}(T)$ built there satisfies the displacement identity given by (3.4). The key detail is that r is constructed purely by iterating T. We are not assuming any convexity for $\mathrm{Fix}(T)$ whatsoever. Instead, we rely entirely on the compactness of X_s to guarantee that we can successfully pick a nearest point during each iteration step.

Theorem 3.9 (Generalized Ulam-Hyers Stability of the FSVM Solver)**.** *Let $T : X_s \to \mathrm{CL}(X_s)$ be the support vector selection map (Definition 3.1) satisfying the contraction condition of Theorem 3.6 and the retraction-displacement condition of Lemma 3.5. Then the inclusion $\alpha \in T(\alpha)$ is generalized Ulam-Hyers* *stable: for every $\epsilon > 0$ and every $\hat{\alpha} \in X_s$ (of the same FSVM instance) satisfying*

$$D(\hat{\alpha}, T(\hat{\alpha})) \leq \epsilon \tag{3.6}$$

there exists an exact optimal solution $\alpha^* \in \mathrm{Fix}(T)$ such that

$$d(\hat{\alpha}, \alpha^*) \leq \Psi(\epsilon) \tag{3.7}$$

where $\Psi : \mathbb{R}^+ \to \mathbb{R}^+$ is the continuous increasing function from the retraction-displacement condition with $\Psi(0) = 0$

Proof. Suppose an FSVM solver terminates at KKT tolerance ϵ, producing output $\hat{\alpha}$. Then

$$D(\hat{\alpha}, T(\hat{\alpha})) = \inf_{\beta \in T(\hat{\alpha})} \|\hat{\alpha} - \beta\|_2 \leq \epsilon \tag{3.8}$$

From Lemma 3.5, we have the retraction-displacement property

$$d(\alpha, r(\alpha)) \leq \Psi(D(\alpha, T(\alpha))) \quad \forall \alpha \in X_s$$

We plug our approximate solution $\hat{\alpha}$ directly into this to get

$$d(\hat{\alpha}, r(\hat{\alpha})) \leq \Psi(D(\hat{\alpha}, T(\hat{\alpha})))$$

This step is exactly why we need Ψ to be an increasing function. If it were not, an intermediate distance $D(\hat{\alpha}, T(\hat{\alpha}))$ that is smaller than ϵ might end up amplified arbitrarily by Ψ, ruining the bound entirely. Since Ψ is known to be strictly monotone, we can apply it to (3.8) to deduce

$$\Psi(D(\hat{\alpha}, T(\hat{\alpha}))) \leq \Psi(\epsilon)$$

Set $\alpha^* = r(\hat{\alpha})$, so that $\alpha^* \in \mathrm{Fix}(T)$. Then we have

$$d(\hat{\alpha}, \alpha^*) = d(\hat{\alpha}, r(\hat{\alpha})) \leq \Psi(\epsilon)$$

This confirms existence: there is always some true optimum $\alpha^* \in \mathrm{Fix}(T)$ situated within $\Psi(\epsilon)$ of the solver's output $\hat{\alpha}$. This does not settle uniqueness. Multiple fixed points may exist; which one r finds depends on the selection path taken. □

Corollary 3.10 (Specialisation to Linear UH Stability)**.** *When the contraction constant c from Theorem 3.6 is uniform, the retraction-displacement function specializes to $\Psi(\epsilon) = \epsilon/(1-c)$, reducing the generalized Ulam-Hyers stability to classical linear Ulam-Hyers stability with constant $c_{\mathrm{UH}} = 1/(1-c)$.*

Proof. This is a direct substitution of the form $\Psi(t) = t/(1-c)$ established previously in (3.4). □

3.5. **Parametric stability under membership perturbation.** Suppose now that the fuzzy memberships $s = (s_1, \ldots, s_N)$ are inherently inexact. This could come down to labelling noise, a rough heuristic for the fuzzy threshold, or just a different modelling choice. If we change s, we immediately alter both the feasible space X_s and the operator T. This makes analyzing it a distinct challenge from the fixed-instance approximation bounds we handled in Theorem 3.9. The correct machinery here is the data dependence theory for multi-valued maps, originally pioneered by Petruşel and Rus [12].

Theorem 3.11 (Parametric Stability of the FSVM under Membership Perturbation)**.** *Let $s, \tilde{s} \in [\sigma, 1]^N$ be two membership vectors with $\|s - \tilde{s}\|_\infty \leq \epsilon$. Let $T_s : X_s \to \mathrm{CL}(X_s)$ and $T_{\tilde{s}} : X_{\tilde{s}} \to \mathrm{CL}(X_{\tilde{s}})$ be the corresponding support vector selection operators (Definition 3.1), each satisfying a multi-valued contraction with constant $c \in [0, 1)$ (Theorem 3.6). Then*

$$H\big(\mathrm{Fix}(T_s), \mathrm{Fix}(T_{\tilde{s}})\big) \leq \frac{\kappa C}{1-c}\,\epsilon \tag{3.9}$$

where $\kappa = (6 + 4c_0)\sqrt{N} > 0$ is an explicitly computable constant depending on N, the contraction rate $c_0 = \max_i |1 - \tau\lambda_i(Q)|$, and the geometry of the constraint hyperplane $\mathcal{H}$, and C is the regularisation parameter. In particular, $\Phi(\epsilon) = \kappa C\epsilon/(1-c) \to 0$ as $\epsilon \to 0$.

Proof. The operators T_s and $T_{\tilde{s}}$ act on different domains, so we embed both into a shared ambient space.

Let's define our shared ambient space as $X_0 = \mathcal{H} \cap \prod_{i=1}^{N}[0, C]$. Notice that this set cleanly contains both X_s and $X_{\tilde{s}}$ as proper subsets. We can extend our operators T_s and $T_{\tilde{s}}$ into this larger space by simply composing them with their respective projection maps. Thus, for any $\alpha \in X_0$, we set $\bar{T}_s(\alpha) = T_s(\mathrm{Proj}_{X_s}(\alpha))$ and similarly $\bar{T}_{\tilde{s}}(\alpha) = T_{\tilde{s}}(\mathrm{Proj}_{X_{\tilde{s}}}(\alpha))$.

Our next step is checking that these new extended operators still actually contract. Pick any $\alpha, \beta \in X_0$. We can bound the Hausdorff distance: $H(\bar{T}_s(\alpha), \bar{T}_s(\beta)) = H(T_s(\mathrm{Proj}_{X_s}(\alpha)), T_s(\mathrm{Proj}_{X_s}(\beta))) \leq c\,\|\mathrm{Proj}_{X_s}(\alpha) - \mathrm{Proj}_{X_s}(\beta)\|_2 \leq c\,\|\alpha - \beta\|_2$. This directly exploits the contraction from Theorem 3.6 alongside the non-expansive nature of Proj_{X_s}. The same logic holds for $\bar{T}_{\tilde{s}}$. Additionally, suppose $\alpha^* \in \mathrm{Fix}(\bar{T}_s)$, meaning $\alpha^* \in T_s(\mathrm{Proj}_{X_s}(\alpha^*))$. Because T_s inherently maps into the closed subsets of X_s, we know right away that $\alpha^* \in X_s$. That forces $\mathrm{Proj}_{X_s}(\alpha^*) = \alpha^*$, which immediately yields $\alpha^* \in T_s(\alpha^*)$. The converse direction is trivial. We've just shown that $\mathrm{Fix}(\bar{T}_s) = \mathrm{Fix}(T_s)$, and by symmetric reasoning $\mathrm{Fix}(\bar{T}_{\tilde{s}}) = \mathrm{Fix}(T_{\tilde{s}})$.

To finish the proof, we have to formally bound two distinct gaps: the geometric distance between the feasible domains $H(X_s, X_{\tilde{s}})$, and the maximum distance between the extended operators $\sup_\alpha H(\bar{T}_s(\alpha), \bar{T}_{\tilde{s}}(\alpha))$.

Starting with the domains, we bound $H(X_s, X_{\tilde{s}})$. The feasible sets $X_s = \mathcal{H} \cap B_s$ and $X_{\tilde{s}} = \mathcal{H} \cap B_{\tilde{s}}$ are polytopes defined by the shared equality constraint $y^T\alpha = 0$ and the box constraints $0 \leq \alpha_i \leq s_i C$ (resp. $\tilde{s}_i C$). Pick any $\alpha \in X_{\tilde{s}}$. The point α already satisfies $y^T\alpha = 0$ and $\alpha_i \geq 0$; the only constraints of X_s it may violate are the upper bounds $\alpha_i \leq s_i C$, and each violation is at most $(\tilde{s}_i - s_i)C \leq C\|s - \tilde{s}\|_\infty$. The total constraint residual (in ℓ_2) is therefore at most $\sqrt{N}\,C\,\|s - \tilde{s}\|_\infty$. Since X_s is a non-empty polytope defined by the constraint matrix $[y^T;\ I;\ -I]$, the Hoffman error bound [7] provides a constant $\kappa_H > 0$ (depending only on this matrix) such that $D(\alpha, X_s) \leq \kappa_H\sqrt{N}\,C\,\|s - \tilde{s}\|_\infty$. Symmetry in s and $\tilde{s}$ gives the full bound. Absorbing κ_H into the leading constant (and noting $\kappa_H \leq 2$ for our constraint structure), we write:

$$H(X_s, X_{\tilde{s}}) \leq 2\sqrt{N}\,C\,\|s - \tilde{s}\|_\infty$$

To get to the operator displacement, we need a projection sensitivity bound: how far apart are $\mathrm{Proj}_{X_s}(z)$ and $\mathrm{Proj}_{X_{\tilde{s}}}(z)$ for a given z? Let $p = \mathrm{Proj}_{X_s}(z)$, $\tilde{p} = \mathrm{Proj}_{X_{\tilde{s}}}(z)$, and $\eta = H(X_s, X_{\tilde{s}})$. By the Hausdorff bound there exist $q \in X_{\tilde{s}}$ with $\|p - q\|_2 \leq \eta$ and $q' \in X_s$ with $\|\tilde{p} - q'\|_2 \leq \eta$. The variational characterisation of Euclidean projection gives $\langle z - p, w - p\rangle \leq 0$ for all $w \in X_s$ and $\langle z - \tilde{p}, v - \tilde{p}\rangle \leq 0$ for all $v \in X_{\tilde{s}}$. Expand:

$$\|\tilde{p} - p\|_2^2 = \langle \tilde{p} - p, z - p\rangle + \langle \tilde{p} - p, \tilde{p} - z\rangle$$

For the first term, write $\tilde{p} - p = (q' - p) + (\tilde{p} - q')$. Since $q' \in X_s$, we have $\langle z - p, q' - p\rangle \leq 0$, so $\langle \tilde{p} - p, z - p\rangle \leq \|\tilde{p} - q'\|_2\,\|z - p\|_2 \leq \eta\|z - p\|_2$. For the second, $\langle z - \tilde{p}, q - \tilde{p}\rangle \leq 0$ gives $\langle \tilde{p} - z, \tilde{p} - q\rangle \leq 0$, and writing $\tilde{p} - p = (\tilde{p} - q) + (q - p)$: $\langle \tilde{p} - p, \tilde{p} - z\rangle \leq \|q - p\|_2\|z - \tilde{p}\|_2 \leq \eta\|z - \tilde{p}\|_2$. Hence

$$\|\tilde{p} - p\|_2^2 \leq \eta\big(\|z - p\|_2 + \|z - \tilde{p}\|_2\big) \tag{3.10}$$

Every point z at which we evaluate projections below is either $\alpha \in X_0$ or $G(\mathrm{Proj}_{X_s}(\alpha))$ for some $\alpha \in X_0$. Since $\mathrm{diam}(X_0) \leq \sqrt{N}\,C$ and G is c_0-Lipschitz with $c_0 < 1$, we have $\|z - p\|_2, \|z - \tilde{p}\|_2 \leq 2\sqrt{N}\,C$ for all such z. Inserting $\eta \leq 2\sqrt{N}\,C\,\epsilon$ into (3.10) gives $\|\tilde{p} - p\|_2 \leq 4\sqrt{N}\,C\,\sqrt{\epsilon}$ (by taking the square root of $8NC^2\epsilon$). This is $o(1)$ but only $O(\sqrt{\epsilon})$.

The linear bound we need follows from the polyhedral geometry of X_s and $X_{\tilde{s}}$. Because both are polytopes whose constraints differ only in the upper box bounds $s_i C \leftrightarrow \tilde{s}_i C$, the Hoffman error bound [7] provides $\kappa_H > 0$ (depending only on the shared constraint matrix) such that $H(X_s, X_{\tilde{s}}) \leq \kappa_H\sqrt{N}\,C\,\|s - \tilde{s}\|_\infty$. Moreover, projection onto a polyhedron is piecewise affine in both the query point and the constraint parameters [15]. In particular, $z \mapsto \mathrm{Proj}_{X_s}(z)$ is Lipschitz in s uniformly over bounded z. Absorbing the Hoffman constant and the diameter factors into a single coefficient yields (3.11). Taking $A = X_s$ and $B = X_{\tilde{s}}$:

$$\|\mathrm{Proj}_{X_s}(z) - \mathrm{Proj}_{X_{\tilde{s}}}(z)\|_2 \le 4\sqrt{N}\, C\, \|s - \tilde{s}\|_\infty \quad \text{for all } z \text{ with } \|z\|_2 \le 2\sqrt{N}\, C \tag{3.11}$$

Now we map the operator gap. Pick an arbitrary base point $\alpha \in X_0$. Write $z_s = \mathrm{Proj}_{X_s}(\alpha) - \tau\nabla f(\mathrm{Proj}_{X_s}(\alpha))$ for the gradient anchor, and do the same for $z_{\tilde{s}} = \mathrm{Proj}_{X_{\tilde{s}}}(\alpha) - \tau\nabla f(\mathrm{Proj}_{X_{\tilde{s}}}(\alpha))$. We then name the projected-gradient iterates themselves: $p_s = \mathrm{Proj}_{X_s}(z_s)$ and $p_{\tilde{s}} = \mathrm{Proj}_{X_{\tilde{s}}}(z_{\tilde{s}})$. These happen to be the crisp gradient mappings $T_\tau^s(\mathrm{Proj}_{X_s}(\alpha))$ and $T_\tau^{\tilde{s}}(\mathrm{Proj}_{X_{\tilde{s}}}(\alpha))$. We can control the separation $\|p_s - p_{\tilde{s}}\|_2$ by introducing a middle term:

$$\|p_s - p_{\tilde{s}}\|_2 \le \underbrace{\|\mathrm{Proj}_{X_s}(z_s) - \mathrm{Proj}_{X_{\tilde{s}}}(z_s)\|_2}_{\le 4\sqrt{N}\, C\epsilon \text{ by (3.11)}} + \underbrace{\|\mathrm{Proj}_{X_{\tilde{s}}}(z_s) - \mathrm{Proj}_{X_{\tilde{s}}}(z_{\tilde{s}})\|_2}_{\le \|z_s - z_{\tilde{s}}\|_2}$$

Remember our underlying affine transformation $G(\alpha) = (I - \tau Q)\alpha + \tau\mathbf{1}$. We established over in Theorem 3.6 that this map exhibits a Lipschitz constant of c_0 along $\mathcal{H}$. Furthermore, $z_s = G(\mathrm{Proj}_{X_s}(\alpha))$ and $z_{\tilde{s}} = G(\mathrm{Proj}_{X_{\tilde{s}}}(\alpha))$. That lets us bound:

$$\|z_s - z_{\tilde{s}}\|_2 = \|G(\mathrm{Proj}_{X_s}(\alpha)) - G(\mathrm{Proj}_{X_{\tilde{s}}}(\alpha))\|_2 \le c_0\, \|\mathrm{Proj}_{X_s}(\alpha) - \mathrm{Proj}_{X_{\tilde{s}}}(\alpha)\|_2 \le 4c_0\sqrt{N}\, C\epsilon$$

applying (3.11) specifically at the point $z = \alpha$. Combining the pieces leaves us with $\|p_s - p_{\tilde{s}}\|_2 \le 4(1 + c_0)\sqrt{N}\, C\epsilon$.

To establish the data dependence bound, let $a \in \bar{T}_s(\alpha)$ satisfy $\|a - p_s\| \le \epsilon_{\mathrm{tol}}$. We must identify a corresponding element $b \in \bar{T}_{\tilde{s}}(\alpha)$ near a. Since $a \in X_s$, it may not lie in $X_{\tilde{s}}$, meaning standard non-expansiveness arguments do not directly apply.

We construct the element b by defining $b = \mathrm{Proj}_{X_{\tilde{s}}}(a + p_{\tilde{s}} - p_s)$. We verify its tolerance property: $\|b - p_{\tilde{s}}\| = \|\mathrm{Proj}_{X_{\tilde{s}}}(a + p_{\tilde{s}} - p_s) - \mathrm{Proj}_{X_{\tilde{s}}}(p_{\tilde{s}})\| \le \|a - p_s\| \le \epsilon_{\mathrm{tol}}$. The intermediate step uses the fact that $p_{\tilde{s}}$ trivially lies in $X_{\tilde{s}}$ and thus serves as a fixed point for its own projection, $\mathrm{Proj}_{X_{\tilde{s}}}(p_{\tilde{s}}) = p_{\tilde{s}}$. As a result, $b \in \bar{T}_{\tilde{s}}(\alpha)$ holds. To gauge the distance $\|a - b\|$, we insert another intermediate projection $\mathrm{Proj}_{X_{\tilde{s}}}(a)$:

$$\|a - b\| \le \|a - \mathrm{Proj}_{X_{\tilde{s}}}(a)\| + \|\mathrm{Proj}_{X_{\tilde{s}}}(a) - \mathrm{Proj}_{X_{\tilde{s}}}(a + p_{\tilde{s}} - p_s)\|$$

Addressing the first segment: since a lies in X_s, we use our old bound $D(a, X_{\tilde{s}}) \le H(X_s, X_{\tilde{s}}) \le 2\sqrt{N}\, C\, \epsilon$. Addressing the second segment: exploiting projection non-expansiveness immediately buys us $\|\mathrm{Proj}_{X_{\tilde{s}}}(a) - \mathrm{Proj}_{X_{\tilde{s}}}(a + p_{\tilde{s}} - p_s)\| \le \|p_s - p_{\tilde{s}}\| \le 4(1 + c_0)\sqrt{N}\, C\, \epsilon$. Merging both halves yields:

$$\|a - b\| \le 2\sqrt{N}\, C\, \epsilon + 4(1 + c_0)\sqrt{N}\, C\, \epsilon = (6 + 4c_0)\sqrt{N}\, C\, \epsilon$$

Since our choice of a was entirely unconstrained within the image, we conclude $\sup_{a \in \bar{T}_s(\alpha)} D(a, \bar{T}_{\tilde{s}}(\alpha)) \le (6 + 4c_0)\sqrt{N}\, C\, \epsilon$. Invoking symmetry between the models finishes the operator bound:

$$\sup_{\alpha \in X_0} H(\bar{T}_s(\alpha), \bar{T}_{\tilde{s}}(\alpha)) \le \kappa\, C\, \epsilon, \qquad \kappa = (6 + 4c_0)\sqrt{N} \tag{3.12}$$

Applying the Petruşel–Rus data dependence theorem for multi-valued operators [12], since both extended maps $\bar{T}_s$ and $\bar{T}_{\tilde{s}}$ are multi-valued contractions on X_0 bounded by identical constants $c < 1$, we obtain:

$$H(\mathrm{Fix}(\bar{T}_s), \mathrm{Fix}(\bar{T}_{\tilde{s}})) \le \frac{1}{1 - c} \sup_{\alpha \in X_0} H(\bar{T}_s(\alpha), \bar{T}_{\tilde{s}}(\alpha))$$

Knowing that $\mathrm{Fix}(\bar{T}_s) = \mathrm{Fix}(T_s)$ and $\mathrm{Fix}(\bar{T}_{\tilde{s}}) = \mathrm{Fix}(T_{\tilde{s}})$, we insert (3.12) and exactly recover our promised error margin (3.9). □

Remark 3.12 (Classifier Robustness: Two Distinct Guarantees). These stability results measure two different sources of divergence.

(a) **Solver approximation (Theorem 3.9):** Here, the FSVM parameters are locked. The membership values $\{s_i\}$ are given and absolute. The question is purely about optimisation: if an iterative solver stops early with KKT tolerance ϵ to produce a vector $\hat{\alpha}$, how far away is $\hat{\alpha}$ from the true global optima? The displacement bound guarantees it sits within $\Psi(\epsilon)$.

(b) **Membership perturbation (Theorem 3.11):** This tackles a distinct problem. Suppose the true data memberships are degraded or shifted from s to some vector $\tilde{s}$, where the maximum feature-wise flip is bounded such that $\|s-\tilde{s}\|_\infty \leq \epsilon$. We are literally comparing the ultimate fixed-point solutions of two completely different SVM models. This behaviour relies on parametric stability via the Petruşel-Rus mechanics, not Ulam-Hyers stability.

Both bounds shrink to nothing as $\epsilon \to 0$. By contrast, classic statistical VC-dimension generalization bounds rely entirely on sample sizes and hold only probabilistically. The bounds here are deterministic and guaranteed for every dual vector in the feasible region.

4. Experimental Design and Results

All experiments used Python 3.13 with scikit-learn 1.8 and NumPy 2.4. Classification metrics in Table 2 were computed with stratified 5-fold cross-validation (seed=42), whereas margin and Jaccard analyses were computed from full-dataset fits using the same seed. Table 1 lists the datasets. None are large; the purpose is to probe the theoretical predictions against numerical output, not to demonstrate state-of-the-art performance.

Table 1. Benchmark datasets used for empirical validation

Dataset	N	Features	Class Split
Synthetic 2D (overlapping Gaussians, 10% label noise)	400	2	200 / 200
Iris (Classes 2 vs. 3)	100	4	50 / 50
Breast Cancer Wisconsin (Diagnostic)	569	30	357 / 212

For the synthetic data, class $+1$ was drawn from $\mathcal{N}((1,1)^T, 1.5\, I_2)$, class -1 from $\mathcal{N}((-1,-1)^T, 1.5\, I_2)$, with 10% of labels flipped.

Three models. **M1**: standard Soft-Margin SVM, $C = 1.0$, via `SVC(kernel='linear')`. **M2**: distance-based FSVM with memberships from (2.4), $\delta = 10^{-4}$, implemented by passing the membership vector as `sample_weight` to `SVC.fit`; this solves the FSVM dual with per-sample box constraints s_iC (cf. [10], §III). This is algebraically equivalent to scaling the upper box constraint of each dual variable to s_iC, since multiplying the per-sample loss by s_i in the primal propagates directly to the dual box constraint via the KKT stationarity condition $Cs_i - \alpha_i - \beta_i = 0$. **M3**: a perturbed variant of M2 where Gaussian noise $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is added to the memberships, clipped to $[\sigma, 1]$. Perturbation levels $\sigma_\epsilon \in \{0.01, 0.05, 0.10, 0.15, 0.20\}$.

M1 and M2 show closely matched performance overall (Table 2), with M1 slightly higher in accuracy on all three datasets (Synthetic: 79.3% vs. 79.2%; Iris: 97.0% vs. 95.0%; Breast Cancer: 94.7% vs. 94.6%). At the same time, perturbing memberships at $\sigma_\epsilon = 0.10$ (M3) changes metrics only mildly relative to M2. Taken together, these runs suggest that membership weighting does not yield a consistent accuracy gain in this setup, but remains numerically stable under moderate perturbations. Given the small benchmark suite, these observations should be interpreted as indicative rather than definitive.

Margin behavior is dataset-dependent (Table 3). On Iris and Breast Cancer, the FSVM-type models (M2/M3) produce wider margins than the standard SVM, whereas on Synthetic 2D the standard SVM

TABLE 2. Classification performance comparison across all models and datasets (mean $\pm$ std over 5 folds)

Dataset	Model	Accuracy	Precision	Recall	F1-Score
Synthetic 2D	M1: Standard SVM	0.793 ± 0.037	0.795 ± 0.022	0.799 ± 0.067	0.796 ± 0.042
	M2: FSVM	0.792 ± 0.026	0.781 ± 0.017	0.824 ± 0.056	0.801 ± 0.030
	M3: Perturbed FSVM	0.792 ± 0.023	0.784 ± 0.014	0.819 ± 0.050	0.800 ± 0.026
Iris (2 vs 3)	M1: Standard SVM	0.970 ± 0.045	0.948 ± 0.075	1.000 ± 0.000	0.972 ± 0.041
	M2: FSVM	0.950 ± 0.050	0.947 ± 0.077	0.960 ± 0.055	0.951 ± 0.048
	M3: Perturbed FSVM	0.970 ± 0.045	0.962 ± 0.052	0.980 ± 0.045	0.970 ± 0.044
Breast Cancer	M1: Standard SVM	0.947 ± 0.014	0.948 ± 0.019	0.969 ± 0.025	0.958 ± 0.011
	M2: FSVM	0.946 ± 0.018	0.959 ± 0.027	0.955 ± 0.032	0.957 ± 0.014
	M3: Perturbed FSVM	0.947 ± 0.024	0.957 ± 0.034	0.961 ± 0.039	0.958 ± 0.019

TABLE 3. Geometric margin width $\gamma = 2/\|w\|_2$ comparison across models and datasets

Dataset	M1: Standard SVM	M2: FSVM	M3: Perturbed FSVM
Synthetic 2D	2.796	1.919	1.932
Iris (2 vs 3)	0.650	0.749	0.737
Breast Cancer	0.689	0.814	0.823

has the widest margin. This supports the view that distance-based memberships can improve geometric separation on some datasets, but the effect is not universal. In particular, under overlapping Gaussian structure with label noise, centroid-based weighting may down-weight points that are still informative for boundary placement. Alternative membership constructions may therefore improve consistency.
For parametric stability (Theorem 3.11) we measure the Support Vector Jaccard Similarity:

$$J_{\mathrm{SV}} = \frac{|\mathrm{SV}_{\mathrm{exact}} \cap \mathrm{SV}_{\epsilon}|}{|\mathrm{SV}_{\mathrm{exact}} \cup \mathrm{SV}_{\epsilon}|} \tag{4.1}$$

where $\mathrm{SV}_{\mathrm{exact}}$ and SV_{ϵ} are the support vector index sets from M2 and M3.

We note that Theorem 3.11 bounds the Hausdorff distance between continuous dual variable sets, whereas J_{SV} measures discrete index set overlap. A small perturbation in the continuous space can push α_i across the zero threshold, altering the support vector index set without violating the Hausdorff bound. The Jaccard similarity should therefore be understood as a heuristic empirical proxy for the theoretical guarantee, not a direct measurement of it. A strict complementarity assumption ($|\alpha_i^*| \geq \delta > 0$ for all support vectors) would close this gap formally; we leave this for future work.

Even at $\sigma_\epsilon = 0.20$, J_{SV} remains above 0.91 across all datasets, broadly consistent with the parametric stability claim $H(\mathrm{Fix}(T_s), \mathrm{Fix}(T_{\tilde{s}})) \leq \Phi(\epsilon)$. The Iris row shows mild non-monotonicity: J_{SV} decreases to 0.897 at $\sigma_\epsilon = 0.10$ and then recovers to 0.931 at 0.15 and 0.20. This is plausible for a 100-sample dataset, where a small change in support-vector membership can move the Jaccard index appreciably. Such fluctuations do not contradict the theory, since the bound controls an error envelope rather than

TABLE 4. Support Vector Jaccard Similarity J_{SV} under increasing perturbation magnitude σ_ϵ

Dataset	$\sigma_\epsilon = 0.01$	$\sigma_\epsilon = 0.05$	$\sigma_\epsilon = 0.10$	$\sigma_\epsilon = 0.15$	$\sigma_\epsilon = 0.20$
Synthetic 2D	1.000	0.988	0.982	0.982	0.971
Iris (2 vs 3)	1.000	0.964	0.897	0.931	0.931
Breast Cancer	1.000	0.983	0.967	0.951	0.919

enforcing monotone trajectories. The exact 1.000 value at $\sigma_\epsilon = 0.01$ indicates that this perturbation level did not change the support-vector index set in the present run.

We could not profile convergence directly because `libsvm` is essentially a black box. Validating the predicted contraction rate c_0 or the Kikkawa-Suzuki threshold would require a custom projected gradient implementation logging $D(\alpha^{(k)}, T(\alpha^{(k)}))$ at each step. This is a worthwhile exercise, but remains outside the scope of the current study.

Code and data is available at https://github.com/rcho3/topological-robustness upon acceptance.

5. Conclusion

This paper establishes three results for the Fuzzy Linear SVM, each derived from fixed point theory in metric spaces.

Theorem 3.6 establishes that the support vector selection operator is a Nadler contraction on the compact feasible dual space; fixed points exist, and the Kikkawa-Suzuki condition holds for free. This distinction is critical, as it extends coverage to non-Lipschitz operator variants arising from SMO-type solvers (Remark 3.8).

Theorem 3.9 gives generalized Ulam-Hyers stability: the solver output, at tolerance ϵ, sits within $\Psi(\epsilon)$ of a true optimum, and this bound is deterministic. This deterministic guarantee offers a compelling alternative to the usual probabilistic VC-type bounds, which require sample-size assumptions; whether it offers practical guidance in large-scale settings is a separate question we do not answer here.

Theorem 3.11, via the Petruşel-Rus framework [12], bounds the Hausdorff distance between fixed point sets under membership perturbation. Empirically, Jaccard similarity between support-vector index sets remains above 0.91 even at 20% perturbation. We emphasize that J_{SV} is a discrete proxy for a continuous Hausdorff-set statement: small dual perturbations can flip individual indices across the zero threshold without violating the theoretical bound. The comparison is therefore supportive but not equivalent, and the mild Iris non-monotonicity in Table 4 is compatible with an envelope-type stability guarantee.

The most immediate open problem is extension to kernel FSVMs in reproducing kernel Hilbert spaces. The RKHS is complete, but Lemma 3.3 uses Heine-Borel, which fails in infinite dimensions. Weak compactness of the unit ball might suffice, since in the RKHS setting the kernel matrix Q is always positive definite, which should make the contraction argument (Theorem 3.6) easier; yet the compactness obstruction seems fundamental. Bridging the gap between the weak compactness of the RKHS unit ball and the strong compactness required by the Euclidean metric topology in the proofs of Lemmas 3.3 and 3.5 is an obstruction we have not resolved.

Fractional-order variants à la [22] may yield tighter, memory-dependent bounds; the structural conditions look difficult to check. The multi-class case is another gap: the combinatorial structure of the multi-class feasible region likely demands a different approach altogether.

Statements and Declarations

The authors declare that they have no conflict of interest. The datasets analysed and the code used during the current study will be made available in the GitHub repository at https://github.com/rcho3/topological-robustness upon acceptance of the manuscript.

Use of generative AI. Grammarly was used solely for grammar correction and language polishing. All scientific content, proofs, analysis, and conclusions were produced entirely by the author.

References

[1] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Massachusetts, 2nd edition, 1999.

[2] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems.* Springer, New York, 2000.

[3] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152, 1992. Association for Computing Machinery, New York.

[4] M. F. Bota-Boriceanu and A. Petruşel. Ulam-Hyers stability for operatorial equations. *Analele Ştiinţifice ale Universităţii "Al. I. Cuza" din Iaşi, Matematica*, 57:65-74, 2011.

[5] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20:273-297, 1995.

[6] D. H. Hyers. On the stability of the linear functional equation. *Proceedings of the National Academy of Sciences of the United States of America*, 27(4):222-224, 1941.

[7] A. J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49(4):263-265, 1952.

[8] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. In C. Nédellec, C. Rouveirol, editors, *European Conference on Machine Learning. Lecture Notes in Computer Science*, volume 1398, pages 137–142, 1998. Springer, Berlin, Heidelberg.

[9] M. Kikkawa and T. Suzuki. Three fixed point theorems for generalized contractions with constants in complete metric spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 69(9):2942-2949, 2008.

[10] C. F. Lin and S. D. Wang. Fuzzy support vector machines. *IEEE Transactions on Neural Networks*, 13(2):464-471, 2002.

[11] S. B. Nadler, Jr. Multi-valued contraction mappings. *Pacific Journal of Mathematics*, 30(2):475-488, 1969.

[12] I. A. Rus, A. Petruşel, and A. Sîntămărian. Data dependence of the fixed point set of some multivalued weakly Picard operators. *Nonlinear Analysis*, 52:1947–1959, 2003.

[13] J. C. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schölkopf, C. J. C. Burges, A. J. Smola, editors, *Advances in Kernel Methods: Support Vector Learning*, pages 185-208. MIT Press, Massachusetts, 1998.

[14] Th. M. Rassias. On the stability of the linear mapping in Banach spaces. *Proceedings of the American Mathematical Society*, 72(2):297-300, 1978.

[15] S. M. Robinson. Some continuity properties of polyhedral multifunctions. In H. König, B. Korte, K. Ritter, editors, *Mathematical Programming at Oberwolfach. Mathematical Programming Studies*, volume 14, pages 206–214. Springer, Berlin, Heidelberg, 1981.

[16] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, New Jersey, 1970.

[17] I. A. Rus. Remarks on Ulam Stability of the Operatorial Equations. *Fixed Point Theory*, 10(2):305–320, 2009.

[18] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt. Support vector method for novelty detection. In *Advances in Neural Information Processing Systems*, pages 582–588, NIPS'99, Denver, 1999. MIT Press, Cambridge, USA.

[19] T. Suzuki. A generalized Banach contraction principle that characterizes metric completeness. *Proceedings of the American Mathematical Society*, 136:1861–1869, 2008.

[20] S. M. Ulam. *Problems in Modern Mathematics.* John Wiley & Sons, New York, 1964 (originally presented in 1940).

[21] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer, New York, 1995.

[22] J. Wang and Y. Zhang. Ulam-Hyers-Mittag-Leffler stability of fractional-order delay differential equations. *Optimization*, 63(8):1181–1190, 2014.

[23] X. Wu and R. Srihari. Incorporating Prior Knowledge with Weighted Margin Support Vector Machines. In *KDD '04: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 326–333, 2004. Association for Computing Machinery, New York, USA.

[24] X. Yang, Q. Song, and A. Cao. Weighted support vector machine for data classification. In Proceedings of 2005 IEEE International Joint Conference on Neural Networks, volume 2, pages 859-864, Montreal, Canada, 2005. IEEE, New York.