Systems & Control Letters, Volume 52, Issues 3–4, pp. 297–304, July 2004.

Continuous dynamical systems that realize discrete optimization on the hypercube*

P.-A. Absil† R. Sepulchre‡

Last modified: 11 Feb 2004

PREPRINT

Abstract

We study the problem of finding a local minimum of a multilinear function E over the discrete set $\{0,1\}^n$. The search is achieved by a gradient-like system in $[0,1]^n$ with cost function E. Under mild restrictions on the metric, the stable attractors of the gradient-like system are shown to produce solutions of the problem, even when they are not in the vicinity of the discrete set $\{0,1\}^n$. Moreover, the gradient-like system connects with interior point methods for linear programming and with the analog neural network studied by Vidyasagar [Vid95] in the same context.

AMS subject classifications. 34D20, 90C51.

Key words. Gradient-like systems, multilinear functions, 0-1 programming, interior point methods, Lyapunov stability.

1 Introduction

Several problems, such as the Maximum Likelihood Decoding problem [BB89], or finding a k-dominating set in a graph [HPV99], can be formulated as the local minimization of a *multilinear function* E over the discrete set $\{0,1\}^n$. A multilinear function is a polynomial in which the variables appear with power 0 or 1. It has the characteristic property of being affine in each variable when the other variables are fixed; the term "multilinear" is thus an abuse of language and some authors prefer "multiaffine". The local minimization problem considered here can be formalized as follows.

Problem 1 *Let* $E : [0,1]^n \to \mathbb{R}$ *be multilinear, i.e.*

$$E(\boldsymbol{x}) = \sum_{i=1}^{m} w_i \prod_{j \in S_i} x_j \tag{1}$$

where $w_i \in \mathbb{R}$ *and* S_i *are subsets of* $\{1,\ldots,n\}$, $i = 1,\ldots,m$. *Let*

$$\hat{E} = E|_{\{0,1\}^n} \tag{2}$$

*This paper presents research partially supported by the Belgian Programme on Inter-university Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. This work was done while the first author was a Research Fellow with the Belgian National Fund for Scientific Research (Aspirant du F.N.R.S.) at the University of Liège.

†School of Computational Science and Information Technology, Florida State University, Tallahassee FL 32306-4120, USA (`www.csit.fsu.edu/~absil/`).

‡Department of Electrical Engineering and Computer Science, Université de Liège, Institut Montefiore (B28), Grande Traverse 10, 4000 Liège, Belgium. (`www.montefiore.ulg.ac.be/systems`)

be the restriction of E *on* $\{0,1\}^n$*. The problem is to find a* local minimum of $\hat{E}$*, i.e. a point* $\boldsymbol{v}^* \in \{0,1\}^n$ *such that* $\hat{E}(\boldsymbol{v}^*) \leq \hat{E}(\boldsymbol{v})$ *for all* $\boldsymbol{v} \in N(\boldsymbol{v}^*)$ *where*

$$N(\boldsymbol{v}^*) = \{\boldsymbol{v} \in \{0,1\}^n : \|\boldsymbol{v} - \boldsymbol{v}^*\| = 1\}$$

is the set of all neareast neighbours of $\boldsymbol{v}^*$ *in* $\{0,1\}^n$ *(with respect to the Hamming distance).*

This problem is equivalent to finding a local minimum of a pseudo-boolean function [HR68] or finding a local solution of the unconstrained 0-1 nonlinear programming problem [BM84]. It can be solved by *local search* algorithms [SY91] which search among adjacent corners of the hypercube $[0,1]^n$ whether the cost can be decreased until it reaches a local minimum (see e.g. Hopfield's two-state neural network [Hop82]). Such algorithms can however be very inefficient: in the worst case, they visit all the set $\{0,1\}^n$ before convergence occurs.

In the spirit of interior-point methods, one way to speed up the process is to relax the problem and consider algorithms evolving in the solid hypercube $[0,1]^n$. In the neural networks community this idea led to analog neural networks such as Hopfield's [Hop84]. In [Vid95] Vidyasagar studies the possibility of solving Problem 1 by means of a generalization of Hopfield's neural network described by the differential equations

$$\dot{u}_i = -\frac{u_i}{\alpha_i} - \frac{\partial E}{\partial x_i}(\boldsymbol{x}) + b_i,\ x_i = g(\lambda u_i), \quad i = 1, \ldots, n, \tag{3}$$

where g is a sigmoidal nonlinearity satisfying some mild assumptions. Stated in an informal way, the main result in [Vid95] is that "for almost all sufficiently small bias vector $\boldsymbol{b}$, for all sufficiently large sigmoidal gains λ, solution trajectories of (3) starting from almost all initial conditions converge to a corner of the hypercube $[0,1]^n$. Moreover, each such corner is a local minimum of the [multilinear] objective function E over the discrete set $\{0,1\}^n$". Vidyasagar suggests in practice to use a "small" bias vector $\boldsymbol{b}$ and a "large" sigmoidal gain.

In the present paper, the minimization of $\hat{E}$ (Problem 1) is realized by means of the gradient-like system

$$\dot{x}_i = -h(x_i)\frac{\partial E}{\partial x_i}(\boldsymbol{x}), \tag{4}$$

which, for a suitable choice of h, can be viewed as (3) in the limit $\boldsymbol{b} = 0$ and $\boldsymbol{\alpha} = +\infty$. The function h, which we term the *temporizing function*, is assumed throughout to verify the following assumption.

Assumption 1 *The function* $h : [0,1] \to \mathbb{R}$ *is* C^1 *on* $[0,1]$*, strictly positive on* $(0,1)$ *and satisfies* $h(0) = h(1) = 0$*.*

The purpose of the function h is to confine the solutions of (4) in the closed hypercube $[0,1]^n$. An interest of system (4) is that it encompasses several classical interior point methods of mathematical programming. For instance, as explained in [Vid95], an approach proposed by Faybusovich in [Fay91] yields $h(x) = 2x(1-x)$, and the affine-scaling vector field [BL89, section VIII] yields $h(x) = x^2(1-x)^2/(x^2 + (1-x)^2)$.

As shown in Section 2, the stable equilibria of (4) are local minima of the function E. A difficulty however is that stable points of (4) need not belong to $\{0,1\}^n$ and are therefore not necessarily solutions of Problem 1. The main result of the paper, established in Section 3, is that under an extra assumption on the temporization function, at least one solution of Problem 1 is directly deduced from any (Lyapunov) stable point of (4). This result provides an alternative to the use of a bias vector as in (3). Its practical implications are illustrated with some simulations in Section 4.

2 Gradient-like systems in the hypercube

In this section, we introduce some definitions, state basic convergence properties of the system (4), and study the relationship between the solutions of the discrete minimization problem (minima of $\hat{E}$) and those of the relaxed continuous problem (minima of E).

The set $[0,1]^n$ will be termed the *(closed) hypercube*, and the elements of $\{0,1\}^n$ its *corners*. We will use the notation $\frac{\partial E}{\partial \boldsymbol{x}}(\boldsymbol{x}) = (\frac{\partial E}{\partial x_1}(\boldsymbol{x}), \ldots, \frac{\partial E}{\partial x_n}(\boldsymbol{x}))^T$.

Definition 2 *Let I be a proper subset of $\{1, \dots, n\}$ and let $\boldsymbol{v_0} \in \{0,1\}^n$. The* open face *attached to the corner $\boldsymbol{v_0}$ with constraint set I is the set $\{\boldsymbol{x} \in [0,1]^n : x_i = (\boldsymbol{v_0})_i$ if $i \in I, x_i \in (0,1)$ if $i \notin I\}$.*

The hypercube $[0,1]^n$ is the disjoint union of the corner set $\{0,1\}^n$ and all the open faces. For example, the three-dimensional cube $[0,1]^3$ is the disjoint union of the corners $\{0,1\}^3$, the edges ($|I| = 2$), the "classical" faces ($|I| = 1$) and the interior of the cube ($I = \emptyset$).

Definition 3 *A point $\boldsymbol{x}^* \in [0,1]^n$ is a* local minimum of E *if*

$$\exists \epsilon > 0 : \forall \boldsymbol{x} \in B_{\epsilon, \boldsymbol{x}^*} \cap [0,1]^n : E(\boldsymbol{x}^*) \le E(\boldsymbol{x})$$

where $B_{\epsilon,\boldsymbol{x}^*} = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon\}$.

Consider the differential equation

$$\dot{x}_i = -h(x_i) \frac{\partial E}{\partial x_i}(\boldsymbol{x}) \tag{5}$$

in $[0,1]^n$, where the cost function E is the multilinear function (1), and h verifies Assumption 1. The solution $\boldsymbol{x}(t)$ of (5) for any initial condition in $[0,1]^n$ exists and is unique for all t, and each open face is invariant.

The solution of (5) generates a descent flow for the function E because

$$\dot{E}(\boldsymbol{x}) := \sum_{i=1}^{n} \dot{x}_i \frac{\partial E}{\partial x_i}(\boldsymbol{x}) = -\sum_{i=1}^{n} h(x_i) \left(\frac{\partial E}{\partial x_i}(\boldsymbol{x}) \right)^2 \le 0.$$

Moreover, $\dot{E}(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{x}$ is an equilibrium point ($\dot{\boldsymbol{x}} = 0$) of (5). Let $\mathcal{F}$ be an open face with constraint set I (Definition 2). A point $\boldsymbol{x}^* \in \mathcal{F}$ is an equilibrium if and only if $\frac{\partial E}{\partial x_i}(\boldsymbol{x}^*) = 0$ for all $i \notin I$, i.e. $\frac{\partial E}{\partial \boldsymbol{x}}$ is orthogonal to the face $\mathcal{F}$. By an application of LaSalle's invariance principle (see e.g. [Kha96]), since the solutions of (5) are confined to the closed hypercube and since $\dot{E} \le 0$, each solution converges to a connected component of the set of equilibrium points.

The next proposition shows that the stable points of (5) are local minima of the function E (in the sense of Definition 3). We first need the following lemma.

Lemma 4 *Let $\boldsymbol{y}$ verify $\dot{E}(\boldsymbol{y}) = 0$ and let $\mathcal{S} = \{\boldsymbol{x} \in [0,1]^n : \dot{E}(\boldsymbol{x}) = 0, E(\boldsymbol{x}) \ne E(\boldsymbol{y})\}$. Then* $\operatorname{dist}(\mathcal{S}, \boldsymbol{y}) > 0$.

Proof. Let $\mathcal{F}$ be an open face (Definition 2) such that $\boldsymbol{y}$ belongs to the closure of $\mathcal{F}$, and let I be the constraint set of $\mathcal{F}$. Let $J = \{1, \dots, n\} \setminus I$. Define

$$\begin{aligned} \mathcal{S}_{\mathcal{F}} &= \mathcal{S} \cap \mathcal{F} \\ &= \{\boldsymbol{x} \in \mathcal{F} : \frac{\partial E}{\partial x_J}(\boldsymbol{x}) = 0 \text{ and } E(\boldsymbol{x}) \ne E(\boldsymbol{y})\}, \end{aligned}$$

where $\frac{\partial E}{\partial x_J} = \left(\frac{\partial E}{\partial x_i}\right)_{i \in J}$. If $\frac{\partial E}{\partial x_J}(\boldsymbol{y}) \ne 0$, then $\operatorname{dist}(\mathcal{S}_{\mathcal{F}}, \boldsymbol{y}) > 0$ by continuity of $\frac{\partial E}{\partial x_J}$. If $\frac{\partial E}{\partial x_J}(\boldsymbol{y}) = 0$, then by Łojaciewicz's inequality [Loj65, KMP00] applied to $E|_{\mathcal{F}}$, there exists a neighbourhood U of $\boldsymbol{y}$ such that, for all $\boldsymbol{x} \in U \cap \mathcal{F}$, one has $\left|\frac{\partial E}{\partial x_J}(\boldsymbol{x})\right| \ge c|E(\boldsymbol{x}) - E(\boldsymbol{y})|^\rho$ for some $\rho < 1$ and $c > 0$. Therefore $U \cap \mathcal{S}_{\mathcal{F}}$ is empty whence $\operatorname{dist}(\mathcal{S}_{\mathcal{F}}, \boldsymbol{y}) > 0$. Since the open faces partition $[0,1]^n$ it comes $\operatorname{dist}(\mathcal{S}, \boldsymbol{y}) > 0$. $\square$

Proposition 5 *Let $\boldsymbol{x}^* \in [0,1]^n$ be a stable equilibrium of* (5)*. Then $\boldsymbol{x}^*$ is a local minimum of E.*

Proof. Let $\boldsymbol{y} \in [0,1]^n$ be an equilibrium point of (5) and suppose that $\boldsymbol{y}$ is not a local minimum of E. We prove that $\boldsymbol{y}$ is not a stable equilibrium point of (5). Suppose w.l.o.g. that $E(\boldsymbol{y}) = 0$. Let $\mathcal{S} = \{\boldsymbol{x} \in [0,1]^n : \dot{E}(\boldsymbol{x}) = 0, E(\boldsymbol{x}) \ne 0\}$ and let $r = \frac{1}{2}\operatorname{dist}(\boldsymbol{y}, \mathcal{S})$. By Lemma 4, $r > 0$. Let $U = \{\boldsymbol{x} \in \bar{B}_{r,\boldsymbol{y}} \cap [0,1]^n : E(\boldsymbol{x}) < 0\}$ where $\bar{B}_{r,\boldsymbol{y}} = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{y}\| \le r\}$. Since $\boldsymbol{y}$ is not a local minimum of E, there exists some $\boldsymbol{x}_0$ arbitrarily close to $\boldsymbol{y}$ such that $\boldsymbol{x}_0 \in U$. Moreover, r has been chosen such that $\dot{E} < 0$ on U. The rest of the proof follows Chetaev's argument [Kha96, p. 112]. Let $\boldsymbol{x}(t)$ be the trajectory of (5) starting at $\boldsymbol{x}(0) = \boldsymbol{x}_0$. One has $E(\boldsymbol{x}_0) < 0$, $\mu := \max\{\dot{E}(\boldsymbol{x}) : \boldsymbol{x} \in U$ and $E(\boldsymbol{x}) \le E(\boldsymbol{x}_0)\} < 0$ as the maximum of a strictly negative function on a compact set, and $E(\boldsymbol{x}(t)) \le E(\boldsymbol{x}_0) + \mu t$ as long as $\boldsymbol{x}(t)$ remains in U. Therefore $\boldsymbol{x}(t)$ must eventually leave U through the boundary of the ball $\bar{B}_{r,\boldsymbol{y}}$. This means that $\boldsymbol{y}$ is not a stable equilibrium point of (5). $\square$

By Proposition 5, stable equilibria of (5) identify local minima of E. However, they do not necessarily correspond to local minima of $\hat{E}$, i.e. solutions of the discrete optimization problem. The next proposition nevertheless identifies important properties of the local minima of E. These properties follow from the multilinearity of E.

Proposition 6 *Let $\boldsymbol{w}$ be a local minimum of E.*
a) If $\boldsymbol{w} \in \{0,1\}^n$, then $\boldsymbol{w}$ is a local minimum of $\hat{E}$.
b) If $\boldsymbol{w}$ belongs to an open face $\mathcal{F}$, then E is constant on $\mathcal{F}$. This also holds true under the weaker condition that $\boldsymbol{w}$ is a local minimum of $E|_{\mathcal{F}}$.
c) $\boldsymbol{w}$ is a strict local minimum of E if and only if $\boldsymbol{w}$ is a strict local minimum of $\hat{E}$.

Proof. a) It is a direct consequence of the multilinear nature of E.

b) Let $\boldsymbol{w} \in (0,1)^n$. We use the shorthand notations $x_{\le p} = x_1, \ldots, x_p$ and $x_{\ge p} = x_p, \ldots, x_n$. For $p = 1, \ldots, n$, let $E_p : [0,1]^p \to \mathbb{R} : x_{\le p} \mapsto E(x_{\le p}, w_{\ge p+1})$. Since E_p is multilinear, it admits the following development

$$E_p(x_{\le p}) = E(\boldsymbol{w}) + \frac{\partial E}{\partial x_1}(w_{\ge 1})(x_1 - w_1) + \ldots + \frac{\partial E}{\partial x_p}(x_{\le p-1}, w_{\ge p})(x_p - w_p). \tag{6}$$

Suppose $\frac{\partial E}{\partial x_i}(x_{\le i-1}, w_{\ge i}) \equiv 0$, $i = 1, \ldots, p$, which trivially holds true for $p = 1$. We will show that $\frac{\partial E}{\partial x_{p+1}}(x_{\le p}, w_{\ge p+1}) \equiv 0$, and the fact that E_n (i.e. E) is identically equal to the constant $E(\boldsymbol{w})$ will follow by induction. One has $E_{p+1}(x_{\le p+1}) = E(\boldsymbol{w}) + \frac{\partial E}{\partial x_{p+1}}(x_{\le p}, w_{\ge p+1})(x_{p+1} - w_{p+1})$. Since $(w_{\le p+1})$ is a local minimum of E_{p+1}, $\frac{\partial E}{\partial x_{p+1}}(x_{\le p}, w_{\ge p+1})$ must vanish on a neighbourhood of $(x_{\le p}) = (w_{\le p})$, whence it vanishes for every $(x_{\le p})$. The general case where $\boldsymbol{w}$ is a local minimum of $E|_{\mathcal{F}}$ follows from the same argument applied to the multilinear function $E|_{\mathcal{F}}$.

c) ($\Rightarrow$) If $\boldsymbol{w}$ is a strict local minimum of E, then it is a corner, because if it were on a face then E would be constant on that face and $\boldsymbol{w}$ would not be a strict local minimum of E. Moreover, $\boldsymbol{w}$ must be a strict local minimum of $\hat{E}$ because of the affine properties of E.
($\Leftarrow$) Let $\boldsymbol{w}$ be a strict local minimum of $\hat{E}$. Without loss of generality, let $\boldsymbol{w} = (0, \ldots, 0)$. Then for all i, $\frac{\partial E}{\partial x_i}(\boldsymbol{w}) > 0$. By continuity of the partial derivatives of E, there exists $\epsilon > 0$ such that $\frac{\partial E}{\partial x_i}(\boldsymbol{y}) > 0$ for all i and for all $\boldsymbol{y} \in B_{\epsilon,\boldsymbol{w}}$. Therefore, for all $\boldsymbol{x}$ in $[0,1]^n \cap B_{\epsilon,\boldsymbol{w}} \setminus \{\boldsymbol{w}\}$, each term $\frac{\partial E}{\partial x_i}(x_{\le i-1}, w_{\ge i})(x_i - w_i)$ in the development (6) of E (i.e. E_n) is nonnegative, one of them at least being strictly positive, and thus $E(\boldsymbol{x}) > E(\boldsymbol{w})$. This means that $\boldsymbol{w}$ is a strict local minimum of E. $\square$

As a consequence of the Proposition 6, the next result shows that *asymptotically* stable equilibria of (5) identify *strict* local minima of $\hat{E}$ (and of E).

Proposition 7 *A point $\boldsymbol{x}^*$ in $[0,1]^n$ is asymptotically stable for (5) if and only if it is a strict local minimum of $\hat{E}$.*

Proof. ($\Rightarrow$) Let $\boldsymbol{x}^* \in [0,1]^n$ be asymptotically stable for (5). Then $\boldsymbol{x}^*$ is a local minimum of E, otherwise it is unstable (Proposition 5). Suppose $\boldsymbol{x}^* \notin \{0,1\}^n$. Then $\boldsymbol{x}^*$ belongs to an open face $\mathcal{F}$, and E is constant on $\mathcal{F}$ (Lemma 6), whence each point of $\mathcal{F}$ is an equilibrium point and $\boldsymbol{x}^*$ is not asymptotically stable. Now suppose that $\boldsymbol{x}^*$ belongs to $\{0,1\}^n$ and is not a strict local minimum of $\hat{E}$. W.l.o.g, let $\boldsymbol{x}^* = \boldsymbol{0}$. Then there exists a canonical vector $\boldsymbol{e}_i$ such that $E(\boldsymbol{x}^* + \boldsymbol{e}_i) \le E(\boldsymbol{x}^*)$. For λ arbitrarily small, the solution $\boldsymbol{x}(t)$ of (5) with $\boldsymbol{x}(0) = \boldsymbol{x}^* + \lambda\boldsymbol{e}_i$ does not converge to $\boldsymbol{x}^*$, which means that $\boldsymbol{x}^*$ is not an asymptotically stable equilibrium point.

($\Leftarrow$) Let $\boldsymbol{x}^* \in \{0,1\}^n$ be a strict local minimum of $\hat{E}$ and suppose w.l.o.g. that $\boldsymbol{x}^* = \boldsymbol{0}$. Then $\boldsymbol{x}^*$ is also a strict local minimum of E (Proposition 6). Moreover, for $i = 1, \ldots, n$, $\frac{\partial E}{\partial x_i}(\boldsymbol{x}^*) \neq 0$, and therefore by continuity of the derivatives there is an $\epsilon > 0$ such that $\dot{E} < 0$ in $B_{\epsilon,\boldsymbol{x}^*} \setminus \{\boldsymbol{x}^*\}$. One concludes by Lyapunov's stability theorem [Kha96, p. 100]. $\square$

3 Main result

The analysis of Section 2 has shown that the descent flow (5) may possess stable equilibria on an open face of the hypercube. As a consequence, solutions of (5) may converge to an open face of stable equilibria rather than to a corner. This situation is a potential failure of the descent flow to identify local minima of the discrete optimization problem. We will now show that convergence to a face of stable equilibria rather than to a corner still allows to identify local minima of the discrete optimization problem, provided some further restrictions are imposed on the temporizing function.

Assumption 2

$$h(1-x) = h(x) = x^{\gamma}(1+\beta(x)) \tag{7}$$

where $\lim_{x\to 0}\beta(x) = 0$. *Since* h *is* C^1 *on* $[0,1]^n$, *one has* $\gamma \geq 1$. *The temporization is termed* linear *when* $\gamma = 1$, quadratic *when* $\gamma = 2$.

Assumption 2 controls the slope of h near the boundary of the interval $[0,1]$. The functions $h(x) = 2x(1-x)$ and $h(x) = x^2(1-x)^2/(x^2+(1-x)^2)$ cited in the introduction are examples of linear and quadratic temporizations, respectively. The following proposition shows that, provided the temporizing function h is at least quadratic, if an open face $\mathcal{F}$ contains a stable equilibrium of (5), then all the corners of $\mathcal{F}$ are solutions of Problem 1.

Proposition 8 *Under Assumption 2 with* $\gamma \geq 2$, *if an open face* $\mathcal{F}$ *contains a stable equilibrium point* $\boldsymbol{x}^*$ *of the system (5), then the corners of* $\mathcal{F}$ *are all local minima of* $\hat{E}$.

Proof. Consider a face $\mathcal{F}_0$ containing a stable equilibrium point $\boldsymbol{x}^*$ for (5). Suppose w.l.o.g. that $\mathcal{F}_0 = \{\boldsymbol{x} \in (0,1)^n : x_{m+1} = \ldots = x_n = 0\}$. Since $\boldsymbol{x}^*$ is stable, it is a local minimum of E (Proposition 5) and E is constant on $\mathcal{F}_0$ (Proposition 6). Suppose w.l.o.g. that $E = 0$ on $\mathcal{F}_0$.

The rest of the proof goes by contradiction. Suppose one of the corners of $\mathcal{F}_0$ is not a local minimum of $\hat{E}$. We can consider, w.l.o.g., that $E(\boldsymbol{e}_{m+1}) = a < 0$. Denote by $\mathcal{F}_1$ the face parallel to $\mathcal{F}_0$ that contains $\boldsymbol{e}_{m+1}$, that is $\mathcal{F}_1 = \{\boldsymbol{x} \in (0,1)^n : x_{m+1} = 1, x_{m+2} = \ldots = x_n = 0\}$.

Either E is constant on $\mathcal{F}_1$. Then for all $y \in [0,1]$, $E(\boldsymbol{x}^* + y\boldsymbol{e}_{m+1}) = E(\boldsymbol{x}^*) + yE(\boldsymbol{x}^* + 1\boldsymbol{e}_{m+1}) = ya$ with $a < 0$. Hence $\boldsymbol{x}^*$ is not a local minimum of E, which is a contradiction.

Or, E is not constant on $\mathcal{F}_1$. Then the contradiction comes by showing that $\boldsymbol{x}^*$ is not an stable equilibrium point, and the proof will be complete.

Let us restrict to the $(m+1)$-dimensional face $\mathcal{F} = \{\boldsymbol{x} : x_{m+2} = \ldots = x_n = 0\}$ that contains both $\mathcal{F}_0$ and $\mathcal{F}_1$ in its closure. As a face, $\mathcal{F}$ is invariant for the gradient-like system (5). Denote $\boldsymbol{z} = (x_1, \ldots, x_m)$, $y = x_{m+1}$ and $\boldsymbol{h}(\boldsymbol{z}) = \operatorname{diag}(h(x_1), \ldots, h(x_m))$. The system (5), restricted to $\mathcal{F}$, reads

$$\dot{\boldsymbol{z}} = -\boldsymbol{h}(\boldsymbol{z})\,\frac{\partial E}{\partial \boldsymbol{z}}(\boldsymbol{z}, y) = y\boldsymbol{F}(\boldsymbol{z}) \tag{8a}$$

$$\dot{y} = -h(y)\,\frac{\partial E}{\partial y}(\boldsymbol{z}, y) = h(y)k(\boldsymbol{z}), \tag{8b}$$

In (8a), $\frac{\partial E}{\partial \boldsymbol{z}}(\boldsymbol{z}, y) = y\frac{\partial E}{\partial \boldsymbol{z}}(\boldsymbol{z}, 1)$ because $\frac{\partial E}{\partial \boldsymbol{z}}$ is multilinear and vanishes on $\mathcal{F}_0$, whence $\boldsymbol{F}(\boldsymbol{z}) = -\boldsymbol{h}(\boldsymbol{z})\,\frac{\partial E}{\partial \boldsymbol{z}}(\boldsymbol{z}, 1)$. In (8b), $k(\boldsymbol{z}) = -\frac{\partial E}{\partial y}(\boldsymbol{z}, y)$ does not depend on y because E is multilinear.

We now introduce quantities ϵ, $\boldsymbol{\delta}_\eta$, ϵ' and K that will be useful shortly. Restricted to the invariant face $\mathcal{F}_1$, the system (8) reads $\dot{\boldsymbol{z}} = \boldsymbol{F}(\boldsymbol{z})$ and $y = 1$. Since E is not constant on $\mathcal{F}_1$, $(\boldsymbol{x}^* + \boldsymbol{e}_{m+1})$ is not a local minimum of $E|_{\mathcal{F}_1}$ (Proposition 6). So it is not a stable equilibrium point (Proposition 5). That is, there exists an $\epsilon > 0$ such that, for all $\eta > 0$, there exists $\boldsymbol{\delta}_\eta$ with $\|\boldsymbol{\delta}_\eta\| < \eta$ and a corresponding $\hat{t} > 0$ such that the solution $\boldsymbol{z}(t)$ of $\dot{\boldsymbol{z}} = \boldsymbol{F}(\boldsymbol{z})$ with initial condition $\boldsymbol{z}^* + \boldsymbol{\delta}_\eta$ is outside the ball $B_{\epsilon,\boldsymbol{z}^*} = \{\boldsymbol{z} \in (0,1)^m : \|\boldsymbol{z} - \boldsymbol{z}^*\| < \epsilon\}$ at $t = \hat{t}$. Since $h(y) \simeq y^\gamma$ (Assumption 2) and $k(\boldsymbol{z})$ is continuous on $[0,1]^m$ thus bounded, there exist positive constants ϵ' and K such that, for all $0 < y < \epsilon'$ and for all $\boldsymbol{z} \in [0,1]^m$, $h(y)k(\boldsymbol{z}) > -Ky^\gamma$.

Let $\eta > 0$ and let $(\boldsymbol{z}(t), y(t))$ be the solution of (8) with initial conditions $0 < y(0) < \epsilon'$ and $\boldsymbol{z}(0) = \boldsymbol{z}^* + \boldsymbol{\delta}_\eta$. We now show that either $y(t)$ is greater than ϵ' at some t or $\boldsymbol{z}(t)$ is outside $B_{\epsilon,\boldsymbol{z}^*}$ at some t. This will complete the proof of instability of $\boldsymbol{x}^*$. So we suppose that $y(t)$ verifies $0 < y(t) < \epsilon'$ for all t and show that $\boldsymbol{z}(t)$ leaves $B_{\epsilon,\boldsymbol{z}^*}$. Consider the new time variable $\tau = \tau(t)$ with $d\tau = y(t)dt$. Then (8a) becomes $d\boldsymbol{z}/d\tau = \boldsymbol{F}(\boldsymbol{z})$. By the argument above, $\boldsymbol{z}(\tau)$ leaves $B_{\epsilon,\boldsymbol{z}^*}$ in finite $\tau = \hat{\tau}$. It remains to show that $\hat{t} := \tau^{-1}(\hat{\tau}) < \infty$. Let $\gamma > 1$. The comparison principle [Kha96, p. 84] applied to (8b) with respect to the above inequality $h(y)k(\boldsymbol{z}) > -Ky^\gamma$ yields $y(t) > \left(y(0)^{1-\gamma} + (\gamma - 1)Kt\right)^{-1/(\gamma-1)}$. Therefore, if $\gamma \geq 2$, then $y(t) \notin L^1$. Since $y(t) \notin L^1$, the relation $t \mapsto \tau(t) = \int_0^t y(s)ds$ has a well-defined inverse τ^{-1} on $[0, \infty)$, whence the time $\hat{t}$ at which $\boldsymbol{z}(t)$ leaves the ball $B_{\epsilon,\boldsymbol{z}^*}$ is finite. □

Thanks to the result of Proposition 8, each stable equilibrium of (5) produces solutions of Problem 1. This result is summarized in the following theorem.

Theorem 9 *Consider system* (5) *restricted to* $[0,1]^n$ *where the multilinear function* E *is given by* (1) *and the temporizing function* $h : [0,1] \to [0,\infty)$ *verifies Assumption 1. Let* $\hat{E} = E|_{\{0,1\}^n}$. *Then each solution of* (5) *converges to a connected set of equilibrium points. If (and only if) an equilibrium point is asymptotically stable, it is a strict local minimum of* $\hat{E}$. *If an equilibrium point is stable, then it is a local minimum of* $\hat{E}$, *or it belongs to an open face (Definition 2) each corner of which is a local minimum of* $\hat{E}$, *provided that Assumption 2 holds with* $\gamma \geq 2$.

A stronger conclusion would be that each solution of (5) converges to a single equilibrium point. This behaviour is indeed observed in practice (see Section 4). Łojasiewicz's theorem [Loj84, KMP00] implies that if $\boldsymbol{x}^* \in (0,1)^n$ belongs to the ω-limit set of a trajectory of (5), then the trajectory converges to $\boldsymbol{x}^*$. However, Łojasiewicz's argument does not apply when $\boldsymbol{x}^*$ belongs to the boundary of $[0,1]^n$ because the metric represented by h is degenerate.

4 Numerical simulations

Figure 1 (left) compares the behaviours of linear and quadratic temporizations in a case where there is no local minimum of E on open faces. Here $n = 2$ and the initial point is chosen at the center of the cube. The simulations suggest that quadratic temporization is more efficient for keeping the solution of (5) away from the faces.

The simulations on Figure 1 (right) have been carried out using the Large Step-size Method (see [Vid95, p. 1372]). In this method, starting from the initial condition, one identifies the search direction, takes a large step leading as far as the boundary of $[0,1]^n$ and then comes 5% back. This gives a new interior point from which the procedure is repeated. Convergence is seen to be faster with quadratic temporization, because stronger temporization improves the parallelism between the nearby face and the direction of the large step.

The next simulations deal with cases where E admits a local minimum on an open face. This situation is frequent when the expression (1) of E contains few terms.

Figure 2 shows a situation where local minima of E lie on an open face. The points on the right half of the lower edge are local minima of E, and one of the adjacent corners of this edge is not a local minimum of $\hat{E}$. With linear temporization, these local minima of E are stable. But with quadratic (or higher) temporization, they are unstable in accordance with Proposition 8.

Figure 3 illustrates a case where a stable point of (5) belongs to an open face and the temporization is quadratic. Since the solution did not converge to a corner, in the spirit of the approach in [Vid95] one would restart the simulation with a "sufficiently small" bias vector. As shown on Figure 4 this destroys the stability of the former limit point and picks out a local minimum in a corner. However, the choice of an adequate bias is not an easy matter: if the bias is very small then the sliding motion along the face is very slow (see Figure 4) but if the bias is not small enough then spurious solutions may appear. In contrast, Proposition 8 certifies that the adjacent corners of the face containing the stable point are local minima of $\hat{E}$, i.e. solutions of Problem 1. The bias is thus superfluous.

5 Conclusion

We addressed the issue of finding a local minimum over $\{0,1\}^n$ of a multilinear function E. The minimization is achieved by means of gradient-like systems defined on the hypercube $[0,1]^n$. We showed that for a general class of gradient-like systems, each stable equilibrium point provides a solution of the minimization problem, even if this equilibrium point does not belong to $\{0,1\}^n$.

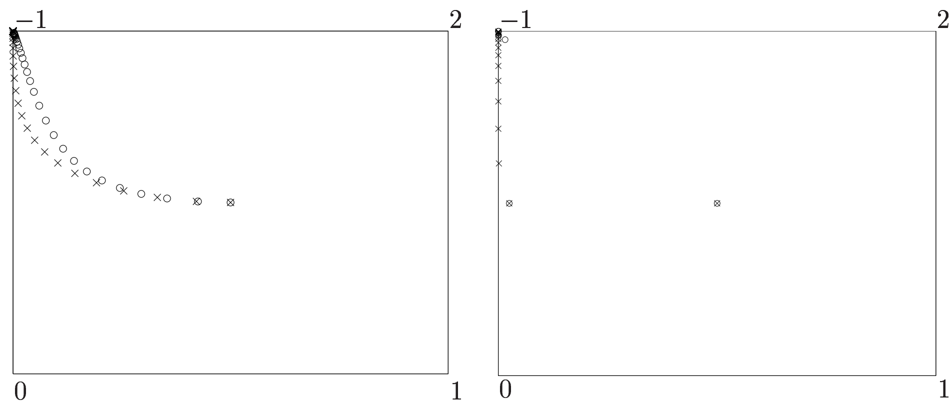

Figure 1: Left: numerical simulation of the trajectory of the gradient-like system (5) with linear (x) and quadratic (o) temporizations. The values at the corners define the function $\hat{E}$ and hence the multilinear function E. Right: gradient-like system (5) with linear (x) and quadratic (o) temporizations, solved using Large Step-size Method.

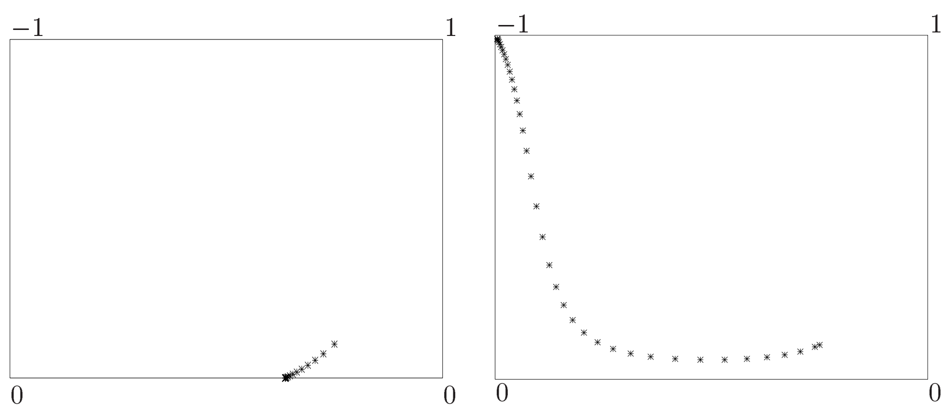

Figure 2: Local minima on an open face, a corner of which is *not* a local minimum. Left: a trajectory of the gradient-like system (5) with linear temporization ($\gamma = 1$). Right: the same trajectory with quadratic temporization can no longer converge to the face.

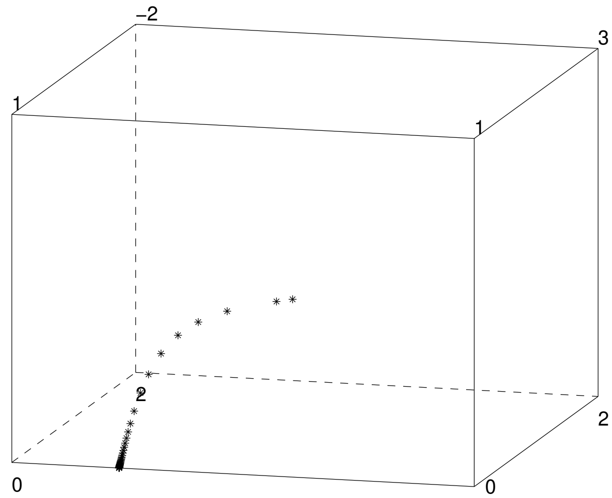

Figure 3: Numerical simulation of the trajectory of the gradient-like system (5) with quadratic temporization ($\gamma = 2$).

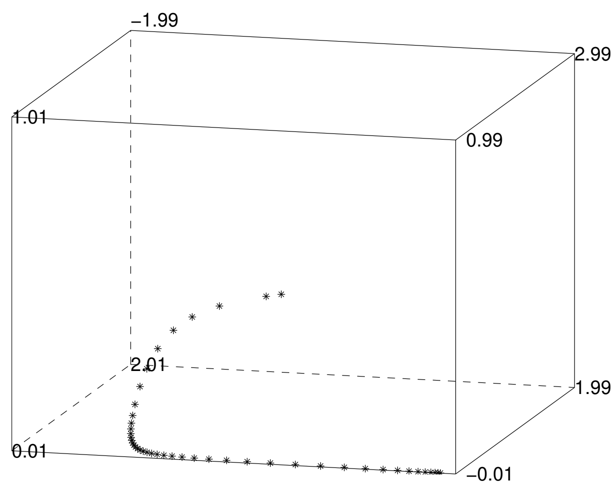

Figure 4: Same as Figure 3 with a small bias vector.

Acknowledgements

The authors thank M. Vidyasagar for having brought to their attention the problem addressed in this paper. They also thank A. Campbell and V. Blondel for their help regarding the result in Lemma 4, and S. Yoshizawa for suggesting reference [KMP00]. The suggestions of an anonymous reviewer improved the quality of this paper considerably and are gratefully acknowledged.

This work was performed in part while the first author was a guest in the Department of Mathematics, University of Würzburg, under a grant from the European Nonlinear Control Network. The hospitality of the members of the department is gratefully acknowledged.

References

[BB89] J. Bruck and M. Blaum. Neural networks, error-correcting codes, and polynomials over the binary n-cube. *IEEE Transactions on Information Theory*, 35(5):976–987, 1989.

[BL89] D. A. Bayer and J. C. Lagarias. The nonlinear geometry of linear programming. II. Legendre transform coordinates and central trajectories. *Transactions of the American Mathematical Society*, 314(2):527–581, 1989.

[BM84] E. Balas and J. B. Mazzola. Nonlinear 0-1 programming: I. Linearization techniques. *Mathematical Programming*, 30:1–21, 1984.

[Fay91] L. Faybusovich. Hamiltonian structure of dynamical systems which solve linear programming problems. *Physica D*, 53:217–232, 1991.

[Hop82] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences USA*, 79:2554–2558, 1982.

[Hop84] J. J. Hopfield. Neurons with graded response have collective computational cababilities like those of two-state neurons. *Proceedings of the National Academy of Sciences USA*, 81:3088–3092, 1984.

[HPV99] J. Harant, A. Pruchnewski, and M. Voigt. On dominating sets and independent sets of graphs. *Combin. Probab. Comput.*, 8(6):547–553, 1999.

[HR68] P. L. Hammer and S. Rudeanu. *Boolean Methods in Operation Research and Related Areas.* Springer, 1968.

[Kha96] H. K. Khalil. *Nonlinear systems, second edition.* Prentice Hall, 1996.

[KMP00] K. Kurdyka, T. Mostowski, and A. Parusiński. Proof of the gradient conjecture of R. Thom. *Ann. of Math. (2)*, 152(3):763–792, 2000.

[Loj65] S. Lojasiewicz. Ensembles semi-analytiques. Inst. Hautes Études Sci., Bures-sur-Yvette, 1965.

[Loj84] S. Lojasiewicz. Sur les trajectoires du gradient d'une fonction analytique. In *Seminari di Geometria 1982-1983*, pages 115–117, Università di Bologna, Istituto di Geometria, Dipartimento di Matematica, 1984.

[SY91] A. A. Schäffer and M. Yannakakis. Simple local search problems that are hard to solve. *SIAM J. Comput.*, 20(1):56–87, 1991.

[Vid95] M. Vidyasagar. Minimum-seeking properties of analog neural networks with multilinear objective functions. *IEEE Trans. Automatic Control*, 40(8):1359–1375, 1995.